Internet Supplement

to

Stochastic-Process Limits

An Introduction to Stochastic-Process And their Application to Queues

Ward Whitt

AT&T Labs - Research
The Shannon Laboratory
Florham Park, New Jersey

Draft
August 1, 2002

Preface

Why is there an Internet Supplement?

This Internet Supplement has three purposes: First, it is intended to maintain a list of corrections for errors found after the book has been published. Second, it is intended to provide supporting details, such as proofs, for material in the book. Third, it is intended to provide supplementary material related to the subject of the book. Needless to say, prior to publication of the book no corrections will appear.

As indicated in the Preface to the book, in order to avoid excessive length, material was deleted from the book and placed in this Internet Supplement. The first choice for cutting was the more technical material. Thus, the Internet Supplement contains many proofs for theorems in the book. Specifically, missing proofs for results stated in the book are contained here in Chapter 1 (all but Section 1.4), Section 5.3 and Chapters 6–8.

It was also considered necessary to cut some entire discussions. Hence the book also contains supplementary material related to, but going beyond, what is in the book. Such material is contained here in Section 1.4, Chapters 2–5 (all but Section 5.3) and Chapter 9.

In addition to making corrections as errors are discovered, the Internet Supplement provides an opportunity to add other material after the book has been published. We would like to add additional material on the spaces E and F, going beyond the brief introduction in Chapter 15 of the book.

Organization

We now indicate how the Internet Supplement is organized.

Chapter 1 here complements Chapter 3 of the book on the framework for stochastic-process limits. Sections 1.2 and 1.3 provide proofs for the

Prohorov metric properties and the Skorohod representation theorem from Section 3.2 of the book. Section 1.4 explains the adjective "weak" in "weak convergence" from a Banach-space perspective. Finally, Section 1.5 gives proofs of the continuous-mapping theorems and the Lipschitz-mapping theorem in Section 3.4 of the book.

Chapter 2 here complements Chapter 4 of the book on basic stochastic-process limits. Section 2.2 complements Section 4.3 of the book on Donsker's theorem by providing an introduction to strong approximations and their application to establish rates of convergence in the setting of Donsker's theorem, using the Prohorov metric on the space of probability measures $\mathcal{P}$ on the function space D. Section 2.3 complements Section 4.4 of the book on Brownian limits with weak dependence by presenting FCLT's exploiting Markov, regenerative and martingale structure. Section 2.4 complements Section 4.5 in the book on convergence to stable Lévy motion by discussing FCLT's in the framework of double sequences (or triangular arrays) of random variables; with an IID assumption, the scaled partial sums converge to general Lévy processes. Finally, Section 2.5 complements Section 4.6 of the book on strong dependence by showing that the linear-process representation in equation (6.6) of the book arises naturally in the framework of time-series models.

Chapter 3 here complements Chapter 13 of the book on useful functions that preserve convergence by showing how pointwise convergence in $\mathbb{R}$ is preserved under mappings. Section 3.2 shows that in some settings pointwise convergence directly implies uniform convergence over bounded intervals. As a consequence, an ordinary strong law of large numbers (SLLN) directly implies the more general functional strong law of large numbers (FSLLN). The remaining sections in Chapter 3 discuss the preservation of pointwise convergence under the supremum, inverse and composition maps. With the inverse map, attention is focused on counting processes, with and without centering.

Chapter 4 here complements Sections 5.9 and 10.4.4 of the book by discussing another application of stochastic-process limits to simulation. Sections 5.9 and 10.4.4 of the book show how heavy-traffic stochastic-process limits for queues can be used to help plan queueing simulations. In particular, they determine the approximate required simulation run length, as a function of model parameters, in order to achieve desired statistical precision. Drawing upon and extending Glynn and Whitt (1992a), Chapter 4 shows how FCLT's and the continuous-mapping approach can be used to establish general criteria for sequential stopping rules for simulations to be asymptotically valid.

Chapter 5 here complements Chapters 5, 8 and 9 of the book on single-server queues. Section 5.2 here discusses general reflected-Lévy-process approximations for queues that arise when there is a sequence of queueing models with net-input processes satisfying the FCLT's discussed here in Section 2.4. Section 5.3 here provides the proof of Theorem 8.3.1 in the book, which establishes a FCLT for the cumulative busy time of a single on-off source. Finally, following Puhalskii (1994), Section 5.4 here shows how the continuous-mapping approach with the inverse map and nonlinear centering in Theorem 13.7.4 of the book can be used to convert stochastic-process limits for arrival, departure and queue-length processes into associated stochastic-process limits for waiting-time and workload processes in quite general queueing models.

Chapters 6, 7 and 8 here provide proofs for theorems in Chapters 12, 13 and 14, respectively, in the book. The numbering within the chapters here closely parallels the numbering within the corresponding chapter in the book, so the desired proof here should be easy to find. In addition, there is an extra section in Chapter 8 here on queueing networks. Drawing on and extending Kella and Whitt (1996), Section 8.9 establishes general conditions for a multidimensional reflected process to have a limiting stationary version.

Chapter 9 here continues the study of useful functions begun in Chapter 13 of the book. In particular, drawing upon and extending Mandelbaum and Massey (1995), Chapter 9 here studies convergence preservation of the supremum, (one-sided, one-dimensional) reflection and inverse maps with nonlinear centering. Under regularity conditions, the limit for the scaled functions after applying these maps can be identified with an appropriate "directional" derivative of the map.

Finally, Chapter 10 here is intended to contain corrections for errors found after the book has been published.

Contents

Chapter 1

Fundamentals

1.1. Introduction

In this chapter we present material supplementing the book on fundamental topics. In Sections 1.2 and 1.3 we give detailed proofs of the Prohorov metric properties and the Skorohod representation theorem, stated in Theorems 3.2.1 and 3.2.2 of the book. In Section 1.4 we explain the adjective "weak" in weak convergence from a Banach-space perspective. In Section 1.5 we provide proofs of the continuous mapping theorems, stated in Section 3.4 of the book.

1.2. The Prohorov Metric

In this section we prove Theorem 3.2.1 in the book, establishing that the Prohorov (1956) metric is indeed a metric inducing weak convergence $P_n \Rightarrow P$.

Recall that we are considering probability measures on a separable metric space (S, m). In that setting, $P_n \Rightarrow P$ if

$$\lim_{n\to\infty} \int_S f dP_n = \int_S f dP \qquad (2.1)$$

for all functions f in $C(S)$, the space of all continuous bounded real-valued functions on S. Recall that the Prohorov metric π is defined on the space $\mathcal{P} \equiv \mathcal{P}(S)$ of all probability measures on the separable metric space (S, m) by

$$\pi(P_1, P_2) \equiv \inf\{\epsilon > 0 : P_1(A) \le P_2(A^\epsilon) + \epsilon \quad \text{for all} \quad A \in \mathcal{B}(S)\} \ , \qquad (2.2)$$

for P_1, $P_2 \in \mathcal{P}(S)$, where A^ϵ is the *open ϵ-neighborhood* of A, i.e.,

$$A^\epsilon \equiv \{y \in S : m(x,y) < \epsilon \quad \text{for some} \quad x \in A\} \; . \tag{2.3}$$

Here is the result that we wish to prove:

Theorem 1.2.1. (the Prohorov metric on $\mathcal{P}$) *For any separable metric space (S,m), the function π on $\mathcal{P}(S)$ in* (2.2) *is a separable metric. There is convergence $\pi(P_n, P) \to 0$ in $\mathcal{P}(S)$ if and only if $P_n \Rightarrow P$, as defined in* (2.1). *Moreover, in* (2.2) *it suffices to let the sets A be closed.*

To carry out the proof, we show that weak convergence $P_n \Rightarrow P$ implies uniform convergence of integrals $\int g dP_n$ for an appropriate class of functions g.

Consider a class $\mathcal{G}$ real-valued functions on S. We say that $\mathcal{G}$ is *uniformly bounded* if

$$\sup_{g \in \mathcal{G}, x \in S} \{|g(x)|\} < \infty \; .$$

We say that $\mathcal{G}$ is *equicontinuous at x* if, for all $\epsilon > 0$, there is a $\delta > 0$ such that

$$\sup_{g \in \mathcal{G}} |g(x) - g(y)| < \epsilon \;\; \text{when} \;\; d(x,y) < \delta \; .$$

We say that $\mathcal{G}$ is *equicontinuous* if it is equicontinuous at all $x \in S$.

Lemma 1.1. (uniform convergence for a class of integrals) *Suppose that $P_n \Rightarrow P$ on a separable metric space (S,m). Let $\mathcal{G}$ be a uniformly bounded class of measurable real-valued functions on S that is equicontinuous at all $x \in E^c$. If $P(E) = 0$, then*

$$\sup_{g \in \mathcal{G}} |\int g dP_n - \int g dP| \to 0 \quad \textit{as} \quad n \to \infty \, . \tag{2.4}$$

Proof. If (2.4) were to fail, then there must exist $\epsilon > 0$ and a sequence $\{g_n : n \geq 1\}$ of functions in $\mathcal{G}$ for which $|\int g_n dP_n - \int g_n dP| > \epsilon$ infinitely often. We will show that cannot happen. Given $P_n \Rightarrow P$, we can apply the Skorohod representation theorem to construct S-valued random elements X_n and X with probability laws P_n and P such that $X_n \to X$ w.p.1. By the almost-sure equicontinuity of $\mathcal{G}$ with respect to P,

$$\sup_n |g_n(X_n) - g_n(X)| \to 0 \quad \text{w.p.1} \; .$$

By the uniform-boundedness condition and the bounded convergence theorem,

$$\sup_n |Eg_n(X_n) - Eg_n(X)| \le E\left[\sup_n |g_n(X_n) - g_n(X)|\right] \to 0 \quad \text{as} \quad n \to \infty$$

or, equivalently,

$$\sup_n \left|\int g_n dP_n - \int g_n dP\right| \to 0 \quad \text{as} \quad n \to \infty .$$

Since that is a contradiction, (2.4) must actually hold. ∎

We now define a generalization of the Prohorov metric on the space $\mathcal{P}(S)$ of all probability measures on (S, m). We define a family of metrics indexed by the scalar γ; the standard Prohorov metric is the special case with $\gamma = 1$. For any P_1, $P_2 \in \mathcal{P}(S)$ and $\gamma > 0$, let

$$\pi_\gamma(P_1, P_2) \equiv \inf\{\epsilon > 0 : P_1(F) \le P_2(F^\epsilon) + \gamma\epsilon \quad \text{for all closed } F \text{ in } S\} , \tag{2.5}$$

where F^ϵ is the open ϵ-neighborhood of F, as in (2.3).

Here is our main result.

Theorem 1.1. (generalized Prohorov metric) *Let (S, m) be a separable metric space. For each $\gamma > 0$, $(\mathcal{P}(S), \pi_\gamma)$ for π_γ in* (2.5) *is a separable metric space. The definition is unchanged if the closed sets F in (2.5) are replaced by general measurable sets A. There is convergence $\pi_\gamma(P_n, P) \to 0$ as $n \to \infty$ if and only if $P_n \Rightarrow P$.*

In preparation for the proof, we first establish some preliminary results. We first show that $\pi_\gamma(P_2, P_1) = \pi_\gamma(P_1, P_2)$. For that purpose, use the following elementary lemma. Recall that A^- is the closure of the set A.

Lemma 1.2. *For any subset A of S and $\alpha > 0$,*

$$A^- = S - (S - A^\alpha)^\alpha . \tag{2.6}$$

Lemma 1.3. *If $P_1(F) \le P_2(F^\alpha) + \beta$ for all closed F for $\alpha, \beta > 0$, then $P_2(F) \le P_1(F^\alpha) + \beta$ for all closed F.*

Proof. Since F^α is open, $S - F^\alpha$ is closed. Under the condition,

$$P_1(S - F^\alpha) \le P_2((S - F^\alpha)^\alpha) + \beta \ ,$$

so that

$$P_2(S - (S - F^\alpha)^\alpha) \le P_1(F^\alpha) + \beta \ .$$

By Lemma 1.2, $F = S - (S - F^\alpha)^\alpha$. hence

$$P_2(F) = P_2(S - (S - F^\alpha)^\alpha) \le P_1(F^\alpha) + \beta \ . \quad \blacksquare$$

We now show that closed sets and measurable sets are interchangeable in (2.5).

Lemma 1.4. (closed sets suffice) *For any constants $\alpha > 0$ and $\beta > 0$, the inequality $P_1(A) \le P_2(A^\alpha) + \beta$ holds for all $A \in \mathcal{S}$ if and only if it holds for all $A = F$, where F is closed.*

Proof. One direction is immediate. For the nontrivial direction, given any measurable set A, choose a sequence of closed sets $\{F_n : n \ge 1\}$ such that $F_n \subseteq F_{n+1}$ and $F_n \uparrow A$. Then $F_n^\alpha \uparrow F^\alpha$, $P_1(F_n) \uparrow P_1(A)$ and $P_2(F_n^\alpha) \uparrow P_2(A^\alpha)$. Hence we have $P_1(A) \le P_2(A^\alpha) + \beta$ when we have $P_1(F_n) \le P_2(F_n^\alpha) + \beta$ for all n. $\blacksquare$

Proof of Theorem 1.1. Lemma 1.3 establishes the symmetry property. if $\pi_\gamma(P_1, P_2) = 0$, then $P_1(F) = P_2(F)$ for each closed subset F. Since the closed sets form a determining class, $P_1 = P_2$. To establish the triangle inequality, suppose that $\pi_\gamma(P_1, P_2) < \epsilon_1 < \pi_\gamma(P_1, P_2) + \delta$ and $\pi_\gamma(P_2, P_3) < \epsilon_2 < \pi_\gamma(P_2, P_3) + \delta$ for some $\delta > 0$. Then for any closed F,

$$\begin{aligned} P_1(F) &\le P_2(F^{\epsilon_1}) + \gamma\epsilon_1 \\ &\le P_2((F^{\epsilon_1})^-) + \gamma\epsilon_1 \\ &\le P_3(F^{\epsilon_1+\epsilon_2}) + \gamma(\epsilon_1 + \epsilon_2) \ , \end{aligned}$$

so that

$$\pi_\gamma(P_1, P_3) \le \epsilon_1 + \epsilon_2 \le \pi_\gamma(P_1, P_2) + \pi_\gamma(P_2, P_3) + 2\delta \ .$$

Since δ was arbitrary, the triangle inequality is established, completing the proof of the metric property.

If $\pi_\gamma(P_n, P) \to 0$, then for any $\epsilon > 0$ there exists n_0 such that $P_n(F) \le P(F^\epsilon) + \gamma\epsilon$ for all closed F and $n \ge n_0$. Hence

$$\limsup_{n\to\infty} P_n(F) \le P(F^\epsilon) + \gamma\epsilon \ .$$

However, $F^\epsilon \downarrow F$ as $\epsilon \downarrow 0$, so that $P(F^\epsilon) \downarrow P(F)$ as $\epsilon \downarrow 0$. Hence,

$$\limsup_{n\to\infty} P_n(F) \le P(F) ,$$

which implies $P_n \Rightarrow P$ by Theorem 11.3.1 in the book.

Next we show that $\pi_\gamma(P_n, P) \to 0$ if $P_n \Rightarrow P$. For each $A \in \mathcal{S}$, define

$$g_A(x) \equiv [1 - \epsilon^{-1} m(x, A)]^+ . \tag{2.7}$$

Notice that $I_A(x) \le g_A(x) \le I_{A^\epsilon}(x)$ for all x, where I_B is the indicator function of the set B. Moreover,

$$|g_A(x) - g_A(y)| \le \epsilon^{-1}|m(x, A) - m(y, A)| \le \epsilon^{-1} m(x, y)$$

for all A, so that the class of all such g_A defined in (2.7) is uniformly bounded and equicontinuous. By Lemma 1.1,

$$\Delta_n \equiv \sup_{A\in\mathcal{S}} \left| \int g_A dP_n - \int g_A dP \right| \to 0 \quad \text{as} \quad n \to \infty .$$

Then

$$P(A^\epsilon) \ge \int g_A dP \ge \int g_A dP_n - \Delta_n \ge P_n(A) - \Delta_n$$

so that

$$P_n(A) \le P(A^\epsilon) + \epsilon$$

when $\Delta_n < \epsilon$.

Finally, we want to show that $(\mathcal{P}(S), \pi_\gamma)$ is separable. For that purpose, let S_0 be a countable dense subset of (S, m), which exists because we have assumed that (S, m) is separable. We will show that the countable family of rational-valued probability measures with finite support in S_0 are dense in $\mathcal{P}(S)$.

Given any $P_1 \in \mathcal{P}(S)$ and any $\epsilon > 0$, we show how to construct P_2 with finite support in S_0 such that $P_1(A) \le P_2(A^\epsilon)$ for all $A \in \mathcal{S}$, so that $\pi_\gamma(P_1, P_2) \le \epsilon$. Let the sequence $\{x_n : n \ge 1\}$ enumerate the elements of S_0. We construct a partition of S containing subset of ϵ-balls about points in S_0. We start by letting $C_1 = B_m(x_1, \epsilon)$. For $C_1, \ldots, C_n$ given, let k_{n+1} be the index of the first point from $\{x_n : n \ge 0\}$ not contained in $\cup_{i=1}^n C_i$. Then let

$$C_{n+1} = B_m(x_{k_{n+1}}, \epsilon) - \cup_{i=1}^n C_i .$$

Let $k_1 = 1$. Now let P_2 attach mass $P_1(C_n)$ to point x_{k_n} (in C_n) for $n \ge 1$. To give P_2 finite support, stop when $P_1(\cup_{i=1}^k C_i) > 1 - \gamma\epsilon$ and let P_2

assign the mass $P_1(\cup_{k+1}^{\infty} C_i)$ to x_1. Hence $P_2(\{x_1\}) = P_1(C_1) + P_1(\cup_{k+1}^{\infty} C_i)$. Now consider an arbitrary measurable set A. Note that $C_i \subseteq A^{\epsilon}$ whenever $A \cap C_i \neq \phi$. Since $\{C_i\}$ is a partition of S,

$$P_1(A) = \sum_{i=1}^{\infty} P_1(A \cap C_i) \leq \sum_{i=1}^{k} P_1(A \cap C_i) + \gamma\epsilon \leq P_2(A^{\epsilon}) + \gamma\epsilon . \quad \blacksquare$$

1.3. The Skorohod Representation Theorem

In this section we prove the Skorohod representation theorem, Theorem 3.2.2 in the book. We restate it here:

Theorem 1.3.1. (Skorohod representation theorem) *If $X_n \Rightarrow X$ in a separable metric space (S, m), then there exist other random elements of (S, m), $\tilde{X}_n, n \geq 1$, and $\tilde{X}$, defined on a common underlying probability space, such that*

$$\tilde{X}_n \stackrel{\mathrm{d}}{=} X_n, n \geq 1, \quad \tilde{X} \stackrel{\mathrm{d}}{=} X$$

and

$$P(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}) = 1 .$$

We start by giving an elementary proof for the case in which the space S is the real line. Then we give Skorohod's (1956) original proof for the case in which S is a complete separable metric space. Finally, we give a proof for general separable metric spaces due to Wichura (1970). Dudley (1968) first showed that the completeness condition is not needed.

1.3.1. Proof for the Real Line

Suppose that $S = \mathbb{R}$. Then we can characterize the probability laws of X and X_n, $n \geq 1$, by their cumulative distribution functions (cdf's), i.e.,

$$F(t) \equiv P(X \leq t), \quad t \in \mathbb{R} . \tag{3.1}$$

For any cdf F, let F be its right-continuous inverse, defined as in Chapter I by

$$F^{-1}(t) = \inf\{s : F(s) > t\}, \quad 0 < t < 1 . \tag{3.2}$$

The representation is achieved by letting $\Omega = [0, 1]$ with Lebesgue measure (the uniform probability distribution), $\tilde{X}(\omega) = F^{-1}(\omega)$ and $\tilde{X}_n(\omega) =$

$F_n^{-1}(\omega)$, $n \geq 1$, with an arbitary definition for $\omega = 0$ and $\omega = 1$. The proof is based on the following four basic lemmas, the first two of which have been discussed in Sections 1.3 and 1.4 of the book.

Lemma 1.5. *If F is a cdf on $\mathbb{R}$ and U is a random variable uniformly distributed on $[0, 1]$, then $F^{-1}(U)$ is a random variable with cdf F.*

Lemma 1.6. (weak convergence criterion in terms of cdf's) *Let X and X_n be real-valued random variables with cdf's F and F_n for $n \geq 1$. Then $X_n \Rightarrow X$ as $n \to \infty$ if and only if $F_n(t) \to F(t)$ as $n \to \infty$ for all t that are continuity points of F.*

Lemma 1.7. *Let F and F_n, $n \geq 1$, be cdf's on $\mathbb{R}$. Then $F_n(t) \to F(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ that are continuity points of F if and only if $F_n^{-1}(t) \to F^{-1}(t)$ for all $t \in (0, 1)$ that are continuity points of F^{-1}.*

Lemma 1.8. *For any cdf F on $\mathbb{R}$, the set of discontinuities of F^{-1} in* (3.2) *is at most countably infinite.*

1.3.2. Proof for Complete Separable Metric Sspaces

The proof of Theorem 1.3.1 will be based on constructing a special family of subsets of (S, m) and relating these subsets to associated subintervals of the interval $[0, 1)$. The length of the subinterval in $[0, 1)$ (probability with respect to Lebesgue measure) will match the probability of the corresponding subset of S. The proof is a combination of Lemma 1.9 below, which shows the existence of the subsets with the required properties, and Lemma 1.10 below, which shows how to exploit such subsets to establish the Skorohod representation. Lemma 1.9 uses the separability; Lemma 1.10 uses the completeness.

A *partition* of a set A is a collection of disjoint subsets of A whose union is A. A *nested family of countable partitions* of a set A is a collection of subsets $A_{i_1,\ldots,i_k}$ of A indexed by k-tuples of positive integers such that $\{A_i : i \geq 1\}$ is a partition of A and $\{A_{i_1,\ldots,i_{k+1}} : i_{k+1} \geq 1\}$ is a partition of $A_{i_1,\ldots,i_k}$ for all $k \geq 1$ and $(i_1, \ldots, i_k) \in \mathbb{N}_+^k$. We allow $A_{i_1,\ldots,i_k}$ to be empty for some $(i_1, \ldots, i_k)$. For each $x \in A$, there is one and only one sequence $\{i_k : k \geq 1\}$ such that $x \in A_{i_1,\ldots,i_k}$ for all k.

Example 1.1. Suppose that $S = \mathbb{R}^+$. We can obtain a nested family of countably partitions of S by letting A_i be $[i-1, i)$ and $A_{i_1,\ldots,i_k}$ be the set of all positive numbers with decimal expansion beginning $(i_1 - 1).(i_2 - 1), (i_3 - 1), \ldots, (i_k - 1)$. Let $A_{i_1,\ldots,i_k} = \phi$ if $i_j > 10$ for any $j \geq 2$. ■

We say that the *radius* of a set A in S is less than r, and write $rad(A) < r$ if $A \subseteq B_m(x,r)$ for some $x \in S$, where $B_m(x,r)$ is the open ball of radius r about x in (S,m). As before, let ∂A be the boundary of A.

Lemma 1.9. *If P is a probability measure on a separable metric space (S,m), then there exists a nested family of countably partitions $\{S_{i_1,\ldots,i_k}\}$ of S such that, for all k and $(i_1,\ldots,i_k)$,*

$$(i) \quad rad(S_{i_1,\ldots,i_k}) < 2^{-k} \tag{3.3}$$

and

$$(ii) \quad P(\partial S_{i_1,\ldots,i_k}) = 0 \ . \tag{3.4}$$

Proof. Since (S,m) is a separable metric space, there exists a countable dense subset, which we can express as a sequence $\{x_i : i \geq 1\}$. For each k, we can choose an r_k such that $2^{-(k+1)} < r_k < 2^{-k}$ and

$$P(\partial B_m(x_i, r_k)) = 0 \quad \text{for all} \quad i \ , \tag{3.5}$$

because there are at most countably many (r,i) such that $P(\partial B_m(x_i,r) > 0)$. Now write

$$D_i^k = B_m(x_i, r_k) - \cup_{j=1}^{i-1} B_m(x_j, r_k) \tag{3.6}$$

and

$$S_{i_1,\ldots,i_k} = D_{i_1}^1 \cap D_{i_2}^2 \cap \cdots \cap D_{i_k}^k \ . \tag{3.7}$$

Since

$$S_{i_1,\ldots,i_k} \subseteq D_{i_k}^k \subseteq B_m(x_{i_k}, r_k) \subseteq B_m(x_{i_k}, 2^{-k}) \ , \tag{3.8}$$

(3.3) holds. Since

$$\partial D_i^k \subseteq \cup_{j=1}^i \partial B_m(x_j, r_k) \tag{3.9}$$

and

$$\partial S_{i_1,\ldots,i_k} \subseteq \partial D_{i_1}^1 \cup \cdots \cup \partial D_{i_k}^k \subseteq \cup_{j=1}^k \cup_{l=1}^{i_j} \partial B_m(x_l, r_j) \ , \tag{3.10}$$

(3.5) implies that (3.4) holds. ∎

Lemma 1.10. *Suppose that P_0 is a probability measure on a complete metric space (S,m) with a nested family of countable partitions $\{S_{i_1,\ldots,i_k}\}$ satisfying* (3.3) *and* (3.4). *If $P_n \Rightarrow P_0$ as $n \to \infty$ on (S,m), then there exist $\tilde{X}_n$, $n \geq 0$, defined on $[0,1]$ with Lebesgue measure, denoted by P, such that $P\tilde{X}_n^{-1} = P_n$, $n \geq 0$, and*

$$P\left(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}_0\right) = 1 \ . \tag{3.11}$$

Proof. We construct nested sequences of countably partitions of $[0,1)$ corresponding to the given nested sequence $\{S_{i_1,\ldots,i_k}\}$ of (S,m). For $n \geq 0$, we construct subintervals $I^n_{i_1,\ldots,i_k}$ corresponding to X_n. We make each subinterval closed on the left and open on the right. Let $I^n_1 = [0, P_n(S_1))$ and

$$I^n_i = \left[\sum_{j=1}^{i-1} P_n(S_j), \sum_{j=1}^{i} P_n(S_j)\right], \quad i > 1 \ . \tag{3.12}$$

Let $\{I^n_{i_1,\ldots,i_{k+1}} : i_{k+1} \geq 1\}$ be a countable partition of subintervals of $I^n_{i_1,\ldots,i_k}$. If $I^n_{i_1,\ldots,i_k} = [a_n, b_n)$, then

$$I^n_{i_1,\ldots,i_{k+1}} = \left[a_n + \sum_{j=1}^{i_{k+1}-1} P_n(S_{i_1,\ldots,i_k,j}), a_n + \sum_{j=1}^{i_{k+1}} P_n(S_{i_1,\ldots,i_k,j})\right) \ . \tag{3.13}$$

The length of each subinterval $I^n_{i_1,\ldots,i_k}$ is the probability $P_n(S_{i_1,\ldots,i_k})$. Now from each nonempty subset $S_{i_1,\ldots,i_k}$ we choose one point $x_{i_1,\ldots,i_k}$. For each $n \geq 0$ and $k \geq 1$, we define functions $x^k_n : [0,1) \to S$ by letting $x^k_n(\omega) = x_{i_1,\ldots,i_k}$ for $\omega \in I^n_{i_1,\ldots,i_k}$. By the nested partition property and (3.3),

$$m(x^k_n(\omega), x^{k+j}_n(\omega)) < 2^{-k} \quad \text{for all} \quad j,k,n \tag{3.14}$$

and $\omega \in [0,1)$. Since (S,m) is a complete metric space, (3.14) implies that there is $x_n \in S$ for all $n \geq 0$ such that

$$m(x^k_n(\omega), x_n(\omega)) \to 0 \quad \text{as} \quad k \to \infty \ . \tag{3.15}$$

We let $\tilde{X}_n = x_n$ on $[0,1)$ for $n \geq 0$. Since $P_n \Rightarrow P_0$ as $n \to \infty$, $P_n(A) \to P_0(A)$ as $n \to \infty$ for all A for which $P_0(\partial A) = 0$ by Theorem 11.3.1 of the book. Hence, $P_n(S_{i_1,\ldots,i_k}) \to P_0(S_{i_1,\ldots,i_k})$ by (3.4). Consequently, the length of the intervals $I^n_{i_1,\ldots,i_k}$ converge to the length of the intervals $I^0_{i_1,\ldots,i_k}$ as $n \to \infty$. Since

$$\begin{aligned} m(\tilde{X}_n(\omega), \tilde{X}_0(\omega)) &\leq m(\tilde{X}_n(\omega), x^k_n(\omega)) + m(x^k_n(\omega), x^k_0(\omega)) \\ &\qquad + m(x^k_0(\omega), \tilde{X}_0(\omega)) \\ &\leq 2^{-(k-1)} + m(x^k_n(\omega), x^k_0(\omega)) \ , \end{aligned} \tag{3.16}$$

for all ω in the interior of $I^0_{i_1,\ldots,i_k}$,

$$\lim_{n\to\infty} m(\tilde{X}_n(\omega), \tilde{X}_0(\omega)) \leq 2^{-(k-1)} \ . \tag{3.17}$$

Since k is arbitrary, we must have $\tilde{X}_n(\omega) \to \tilde{X}_0(\omega)$ as $n \to \infty$ for all but at most countably many $\omega \in [0, \infty)$.

It remains to show that $\tilde{X}_n$ has the probability law P_n for $n \geq 0$. It suffices to show that $P(\tilde{X}_n \in A) = P_n(A)$ for each A such that $P_n(\partial A) = 0$. Let A be such a set. Let A^k be the union of the sets $S_{i_1,\ldots,i_k}$ such that $S_{i_1,\ldots,i_k} \subseteq A$ and let A'^k be the union of the sets $S_{i_1,\ldots,i_k}$ such that $S_{i_1,\ldots,i_k} \cap A \neq \phi$. Then $A^k \subseteq A \subseteq A'^k$ and, by construction above,

$$P(\tilde{X}_n \in A^k) = P_n(A^k) \quad \text{and} \quad P(\tilde{X}_n \in A'^k) = P_n(A'^k) \; . \tag{3.18}$$

Now let

$$C^k = \{x \in S : m(x, \partial A) \leq 2^{-k}\} \; . \tag{3.19}$$

Then $A'^k - A^k \subseteq C^k \downarrow \partial A$ as $k \to \infty$. Since $P_n(\partial A) = 0$ by assumption, $P_n(C^k) \downarrow 0$ as $k \to \infty$. Hence

$$P(\tilde{X}_n \in A) = \lim_{k\to\infty} P(\tilde{X}_n \in A^k) = \lim_{k\to\infty} P_n(A^k) = P_n(A) \; . \quad \blacksquare \tag{3.20}$$

1.3.3. Proof for Separable Metric Spaces

We now do the proof of Theorem 1.3.1 without assuming completeness. Start by letting P_n be the probability distribution of X_n on S for $n \geq 0$. Let the underlying probability space be the product space $\Omega \equiv S^\infty$ with elements $\omega \equiv \{s_k : k \geq 0\}$. Let $\tilde{X}_n$ be the coordinate mapping, e.g., $\tilde{X}_n(\{s_k : k \geq 0\}) = s_n$, $n \geq 0$. To quickly get the idea, first suppose that $P_n(\{s\}) = 1$ for all $n \geq 0$. In this special case we can let the probability measure P on Ω be the product measure $P = \delta_s \times \delta_s \times \cdots$, where δ_s is the Dirac measure assigning probability 1 to the point $s \in S$. Then P assigns probability 1 to the sequence $\{s_n : n \geq 0\}$ where $s_n = s$ for all n. Since $P(\tilde{X}_n = s) = 1$ for all n,

$$P(\tilde{X}_n = \tilde{X}_0 \quad \text{for all} \quad n) = P\left(\cap_{n=0}^{\infty}\{\tilde{X}_n = s\}\right) = 1 \; . \tag{3.21}$$

To continue to develop the idea of the approach, now suppose that each probability measure P_n, $n \geq 0$, concentrates all probability on a common finite subset of S. Thus it suffices to assume that S is finite. For a sequence $\{k_n : n \geq 1\}$ with $k_n \to \infty$ as $n \to \infty$ to be defined later, let

$$U_k = \cap_{n:k_n \geq k}\{\tilde{X}_n = \tilde{X}_0\} \; . \tag{3.22}$$

(Note that we have a strong form of convergence on U_k.) Also let $\{Q_n : n \geq 1\}$ be a sequence of probability measures on S to be defined later. Now let

$P_{j,s}$ be the product measure

$$P_{j,s} = \delta_s \times \prod_{n=1}^{\infty} P_{j,s,n} \ , \tag{3.23}$$

where $P_{j,s,n}$ is a probability measure on S defined by

$$P_{j,s,n} = \begin{cases} Q_n & \text{if} \quad 0 \le k_n < j \\ \delta_s & \text{if} \quad j \le k_n \le \infty \ . \end{cases} \tag{3.24}$$

Then let P'_j be a mixture of the probabilities $P_{j,s}$ in (3.23) with respect to P_0, in particular,

$$P'_j = \sum_{s \in S} P_0(\{s\}) P_{j,s} \ . \tag{3.25}$$

Next let $\{w_k : k \ge 1\}$ and $\{q_k : k \ge 0\}$ be sequences of numbers with

$$w_k \ge 0, \quad \sum_{k=1}^{\infty} w_k = 1, \quad q_0 = 0, \quad q_k = \sum_{j=1}^{k} w_j < 1, \quad 1 \le k < \infty \ . \tag{3.26}$$

Then let P be a mixture of the probabilities P'_j in (3.25) using the weights w_j in (3.26), i.e.,

$$P = \sum_{j=1}^{\infty} w_j P'_j \ . \tag{3.27}$$

We will show that this construction does the job with an appropriate choice of the sequences $\{k_n : n \ge 1\}$ and $\{Q_n : n \ge 1\}$. (The weights w_k in (3.26) can be arbitrary subject to the conditions in (3.26).)

Note that $P_{j,s}$ in (3.23) attaches positive probability only to sets of sequences $\{s_n : n \ge 0\}$ such that $s_n = s$ for all but a finite number of n (those n for which $0 \le k_n < j$). Thus even though S^∞ is uncountably infinite, $P_{j,s}$ has finite support. Since S is finite, P'_j in (3.25) also has finite support. All sequences $\{s_n : n \ge 0\}$ in S^∞ with positive P-measure have $s_n = s$ for all sufficiently large n for some s.

By (3.23), $P_{j,s}(\tilde{X}_0 = s) = 1$. Thus, by (3.25) and (3.27), $P(\tilde{X}_0 = s) = P'_j(\tilde{X}_0 = s) = P_0(\{s\})$ for all $s \in S$. Hence $P\tilde{X}_0^{-1} = P_0$ or, equivalently, $\tilde{X}_0 \stackrel{\mathrm{d}}{=} X_0$.

Next $P_{j,s}(\tilde{X}_n = s) = P_{j,s,n}(\{s\})$ for $n \ge 1$. Note that $P_{j,s}(U_k) = 1$ for $j \le k$, where U_k is given in (3.22), so that $P'_j(U_k) = 1$ if $j \le k$ and $P(U_k) \ge q_k$. Since $q_k \to 1$ as $k \to \infty$ by (3.26), $\tilde{X}_n \to \tilde{X}_0$ as $n \to \infty$ almost uniformly on Ω with respect to P, i.e., for any $\epsilon > 0$, there exists a subset

U_k of S with $P(U_k) > 1 - \epsilon$ such that $\tilde{X}_n$ converges uniformly to $\tilde{X}_0$ as $n \to \infty$ on U_k. (In our finite-state-space setting, we actually have $\tilde{X}_n = \tilde{X}_0$ on U_k for all n such that $k_n \geq k_0$ by (3.22) and (3.26).) For ϵ given, choose k so that $q_k > 1 - \epsilon$. By Egoroff's theorem, p. 89 of Halmos (1956), that implies that

$$P\left(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}_0\right) = 1 \ . \tag{3.28}$$

The difficult part is to obtain $\tilde{X}_n \stackrel{\mathrm{d}}{=} X_n$ for $n \geq 1$. The construction above yields

$$P(\tilde{X}_n = s) = q_{k_n} P_0(\{s\}) + (1 - q_{k_n}) Q_n(\{s\}) \quad \text{for all} \quad n \ . \tag{3.29}$$

We now choose the sequences $\{k_n : n \geq 1\}$ and $\{Q_k : k \geq 1\}$ to achieve $P\tilde{X}_n^{-1} = P_n$ for all n. Note that (3.29) is equivalent to

$$Q_n(\{s\}) = P_n(\{s\}) + \frac{q_{k_n}}{1 - q_{k_n}} (P_n(\{s\}) - P_0(\{s\}) \tag{3.30}$$

provided $k_n < \infty$. If $k_n = \infty$, then $q_{k_n} = 1$, so that we must have $P_n(\{s\}) = P_0(\{s\})$, and then any $Q_n(\{s\})$ will do.

Thus, let

$$\begin{aligned} Q(k,s,n) &\equiv P_n(\{s\}) + \frac{q_k}{1-q_k}(P_n(\{s\}) - P_0(\{s\}) \ , & (3.31)\\ m_{k,n} &\equiv \min_{s\in S} Q(k,s,n) \ , & (3.32)\\ k_n &= \sup\{j \geq 0 : m_{j,n} \geq 0\} & (3.33) \end{aligned}$$

and

$$Q_n(\{s\}) = Q(k_n, s, n) \quad \text{for} \quad k_n < \infty \ . \tag{3.34}$$

Note that $m_{0,n} \geq 0$, so that $k_n \leq \infty$ is well defined in (3.33). Note that $\sum_{s\in S} Q(k,s,n) = 1$ for all k, $0 \leq k < \infty$, and $Q(k_n, s, n) \geq 0$ by (3.32) and (3.33). Thus, under (3.31)–(3.34), Q_n is a probability measure on S satisfying (3.29) provided that $k_n < \infty$.

Since $\sum_{s\in S} Q(k,s,n) = 1$, we must have $0 \leq Q(k,s,n) \leq 1$ for $Q(k,s,n)$ in (3.31). Since $q_k \to 1$ as $k \to \infty$, $q_k/(1-q_k) \to \infty$ as $k \to \infty$. Hence, we must have $P_n(\{s\}) = P_0(\{s\})$ for all n if $k_n = \infty$, under which (3.29) has been shown to hold for any probability measure Q_n.

We now show that $k_n \to \infty$ as $n \to \infty$. Since $P_n(\{s\}) \to P_0(\{s\})$ as $n \to \infty$ for each s, $Q(k,s,n) \to P_0(\{s\})$ as $n \to \infty$ for each s and k, $1 \leq k < \infty$. This, together with the fact that $Q(k,s,n) \geq 0$ if $P_0(\{s\}) = 0$, implies that $Q(k,s,n)$ is ultimately nonnegative for all sufficiently large

n depending upon k. Thus, for each k, there is an index n_k such that $Q(k,s,n) \geq 0$, and thus $m_{k,n} \geq 0$, for all $n \geq n_k$. Since $m_{k,n} \geq 0$ implies $k_n \geq k$, we can conclude that, for all $n \geq n_k$, $k_n \geq k$. Hence, $k_n \to \infty$ as $n \to \infty$.

We now turn to the general case: We now assume that S is a separable metric space. We start by constructing a finite collection of subsets appropriately approximating S. This step is a minor modification of Lemma 1.9.

Lemma 1.11. *If P is a probability measure on separable metric space (S,m), then for any δ, $\epsilon > 0$ there exist disjoint subsets $S_{i_1,\ldots,i_k}$ of S, $1 \leq i_j \leq i'_j$, $1 \leq j \leq k$, such that, for all k and $(i_1,\ldots,i_k)$, (3.3) holds for $2^{-k} < \delta$, (3.4) holds and*

$$P\left(\cup_{i_1=1}^{i'_1} \cdots \cup_{i_k=1}^{i'_k} S_{i_1,\ldots,i_k}\right) > 1-\epsilon\ . \tag{3.35}$$

Proof. We use the construction in Lemma 1.9. Choose i'_1 such that $P(S_1 \cup \cdots \cup S_{i'_1}) > 1-\epsilon 2^{-1}$; choose i'_2 such that

$$P(S_{i_1,1} \cup \cdots \cup S_{i_1,i'_2}) > 1 - P(S_{i_1})\epsilon 2^{-2} \tag{3.36}$$

for all i_1, $1 \leq i_1 \leq i'_1$; choose i'_{j+1} such that

$$P(S_{i_1,\ldots,i_j,1} \cup \cdots \cup S_{i_1,\ldots,i_j,i'_{j+1}}) > 1 - P(S_{i_1,\ldots,i_j})\epsilon 2^{-j} \tag{3.37}$$

for all $(i_1,\ldots,i_j) \leq (i'_1,\ldots,i'_j)$. Stop at k with $2^{-k} < \delta$, so that (3.3) holds. Then

$$P\left(\cup_{i_1=1}^{i'_1} \cdots \cup_{i_k=1}^{i'_k} S_{i_1,\ldots,i_k}\right) > 1-\epsilon(2^{-1}+\cdots+2^{-k}) > 1-\epsilon\ , \tag{3.38}$$

so that (3.35) holds. ∎

We now return to the proof of the theorem. Let $\{\delta_k : k \geq 1\}$ and $\{\epsilon_k : k \geq 1\}$ be sequences of positive numbers such that $\delta_k \to 0$, $\epsilon_k \to 0$ and $\sum_{k=1}^{\infty} \epsilon_k < \infty$. For each k, let $\{C_{k,j} : 0 \leq j \leq n_k\}$ be the finite collection of subsets $S_{i_1,\ldots,i_k}$ in Lemma 1.11 constructed with respect to P_0, where δ and ϵ for k are required to be δ_k and ϵ_k. Let $C_{k,0} = S - \cup_{j=1}^{n_k} C_{k,j}$. By (3.35), $P_0(C_{k,0}) < \epsilon_k$.

With $\tilde{X}_n$ the coordinate projections on S^∞ as before, instead of (3.22), let

$$U_k = \cap_{n:k_n \geq k}\{m(\tilde{X}_n,\tilde{X}_0) \leq \delta_{k_n}\} \tag{3.39}$$

where $\delta_\infty = 0$. (The separability of (S, m) is used to have $\{m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}\}$ and thus U_k be measurable.) Given that $k_n \to \infty$ as $n \to \infty$, $\tilde{X}_n \to \tilde{X}_0$ uniformly on U_k. To apply Egoroff's theorem, we will need to show that $P(U_k) \to 1$ as $k \to \infty$.

Let Π_k be the collection of sets $C_{k,j}$, $1 \leq j \leq n_k$, and let $\Pi_0 = S$. We now modify the finite-state-space proof above, letting $C_{k,j}$ play the role of s. Let the weights w_k and their partial sums q_k be defined by (3.26). Paralleling (3.31)–(3.34), for $0 \leq k < \infty$, let

$$\begin{aligned} Q(k, C, n) &= P_n(C) + \frac{q_k}{1 - q_k}(P_n(C) - P_0(C)) \ , & (3.40) \\ m_{k,n} &= \min_{C \in \Pi_k} \{Q(k, C, n)\} \ , & (3.41) \\ k_n &= \sup\{j \geq 0 : m_{j,n} \geq 0\} & (3.42) \end{aligned}$$

and

$$Q_n(C) = Q(k_n, C, n) \ . \tag{3.43}$$

Since $P_n(C) \to P_0(C)$ as $n \to \infty$ for all $C \in \Pi_k$, $k_n \to \infty$ as $n \to \infty$ by the same argument as before.

Paralleling (3.23), let $P_{j,s}$ be the product measure

$$P_{j,s} = \delta_s \times \prod_{n=1}^{\infty} P_{j,s,n} \ , \tag{3.44}$$

where $P_{j,s,n}$ is a probability measure on S defined by

$$P_{j,s,n} = \begin{cases} Q_n & \text{if } \ 0 \leq k_n < j \\ P_n(\cdot | C_{k_n,s}) & \text{if } \ j \leq k_n < \infty \\ \delta_s & \text{if } \ k_n = \infty \ , \end{cases} \tag{3.45}$$

where $P_n(\cdot | C_{k_n,s})$ is the conditional probability measure with $C_{k,s}$ being the element of Π_k containing $s \in S$. Note that $P_{j,s,n}$ in (3.45) has three possibilities instead of only the two in (3.24). Unlike the case of finite S, $P_{j,s}$ in (3.44) does not have finite support, but if $s \in C_{k_n,i}$, then $P_{j,s}$ has support on the set of sequences $\{s_n : n \geq 0\}$ such that $s_n \in C_{k_n,i}$ for all but finitely many n, in particular, for all n such that $k_n \geq j$. On this subset of sequences, $m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}$ for all n such that $k_n \geq j$.

Paralleling (3.25), let

$$P'_j(A) = \int_S P_0(ds) P_{j,s}(A) \ . \tag{3.46}$$

The integral in (3.46) is well defined since $P_{j,s}(A)$ is a measurable function on S for each A a cylinder set with finite base in the σ-field on S^∞; see pp. 74–76 of Neveu (1965). Note that P'_j has support on the set of sequences $\{s_n : n \geq 0\}$ such that $s_n \in C_{k_n,i}$ for all but finitely many n, for some i. Thus

$$P'_j(m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}) \geq 1 - P(C_{k_n,0}) > 1 - \epsilon_{k_n} \ . \tag{3.47}$$

Paralleling (3.27), let

$$P = \sum_{j=1}^{\infty} w_j P'_j \ . \tag{3.48}$$

As before, the construction yields $P\tilde{X}_0^{-1} = P_0$. The probability distribution of $\tilde{X}_n$ is

$$P\tilde{X}_n^{-1} = \begin{cases} q_{k_n} \sum_{C \in \Pi_{k_n}} P_n(\cdot|C) P_0(C) + (1 - q_{k_n}) Q_n & \text{if} \quad k_n < \infty \\ P_0 & \text{if} \quad k_n = \infty \ . \end{cases} \tag{3.49}$$

For n such that $k_n < \infty$, let

$$Q_n = \sum_{C \in \Pi_k} Q(k_n, C, n) P_n(\cdot|C) \ . \tag{3.50}$$

Combining (3.40), (3.49) and (3.50), we see that $P\tilde{X}_n^{-1} = P_n$ if $k_n < \infty$. On the other hand, as before, if $k_n = \infty$, then we are forced to have $P_n(C) = P_0(C)$ for all $C \in \Pi_k$ for any $k \geq 1$, but that implies that $P_n = P_0$. (We can apply the reasoning in the proof of Lemma 1.10 using (3.18) and (3.19).)

Finally, it remains to show that $P(U_k) \to 1$ as $k \to \infty$ for U_k in (3.39). However,

$$\begin{aligned} P(U_k) &= \sum_{j=1}^{\infty} w_j P'_j(U_k) \geq \sum_{j=1}^{k} w_j P'_j(U_k) \\ &\geq \left(\sum_{j=1}^{k} w_k \right) \left(1 - \sum_{j=k}^{\infty} \epsilon_j \right) \to 1 \quad \text{as} \quad k \to \infty \ , \end{aligned} \tag{3.51}$$

since, for $j \leq k \leq k_n$,

$$1 - P'_j(U_k) \leq P_0\left(\cup_{l=k}^{\infty} C_{l,0}\right) \leq \sum_{l=k}^{\infty} \epsilon_l \ , \tag{3.52}$$

because $P'_{j,s}$ assigns probability 1 to product sets in which all coordinates are in common sets $C_{i,k}$.

1.4. The "Weak" in Weak Convergence

This section is devoted, not to a proof of a theorem, but to an expanation of a term – the adjective "weak" in "weak convergence." The term "weak" can be understood from a Banach-space perspective.

The starting point is the definition of convergence $P_n \Rightarrow P$; i.e., $P_n \Rightarrow P$, if

$$\lim_{n\to\infty} \int_S f dP_n = \int_S f dP \tag{4.1}$$

for all functions f in $C(S)$, the space of all continuous bounded real-valued functions on S.

The space $C(S)$ of continuous bounded real-valued functions h on S used in definition (4.1) is a Banach space (a complete normed linear topological space) with the uniform norm

$$\|h\| \equiv \sup_{s\in S} |h(s)| \ .$$

The adjoint or conjugate space of $C(S)$, the space of all continuous linear real-valued functions L on $C(S)$, denoted by $C^*(S)$, turns out to be the space $Z(S)$ of all finite signed measures μ on S, defined via

$$L(h) \equiv \int_S h d\mu \ ;$$

e.g., see pp. 262, 419 of Dunford and Schwartz (1958) or Chapter 9 of Simmons (1963).

The adjoint space B^* of any Banach space B is itself a Banach space with the norm

$$\|L\| \equiv \sup\{\|L(b)\| : b \in B, \quad \|b\| \leq 1\} \ .$$

Since B^* is a Banach space, one can consider its adjoint space B^{**}. There is a natural embedding of B in B^{**} so that we can regard B as a subset of B^{**}. (Just let $L_b(f) = f(b)$ for $b \in B$ and $f \in B^*$.) When $B = B^{**}$, B is said to be *reflexive.* However, $C(S)$ is reflexive only when S is finite. So, in our setting with infinite S, $C(S)$ is a proper subset of $C^{**}(S)$.

Instead of the topology induced on a Banach space B by its norm, it is sometimes of interest to consider a weaker topology on B called the *weak topology*, which is the weakest topology such that all the functions in B^* remain continuous; i.e., $b_n \to b$ in B with the weak topology if and only if $L(b_n) \to L(b)$ for all L in B^*. Furthermore, on the adjoint space B^* one can also consider a still weaker topology called the weak* topology, which is the weakest topology such that all the functions in B, regarded as a subset of B^{**}, remain continuous. Thus the weak* topology on $Z(S) = C^*(S)$ relativized to the subset $P(S)$ is what is characterized by (4.1). (The discussion also implies that the weak topology on $Z(S)$ is stronger than the weak* topology on $Z(S)$, so the terminology "weak convergence" is something of a misnomer. From this Banach-space perspective, we should actually call weak convergence $P_n \Rightarrow P$ *weak** convergence.) ■

1.5. Continuous-Mapping Theorems

In this section we supplement the discussion of the continuous-mapping approach in Section 3.4 of the book by providing proofs for the unproved theorems. We first prove the Lipschitz mapping theorem, which comes from Whitt (1974).

1.5.1. Proof of the Lipschitz Mapping Theorem

We now prove the Lipschitz mapping theorem, Theorem 2.4.2 in the book. First suppose that (S, m) is a separable metric space and $B = S$. Then we can employ the Strassen representation theorem, Theorem 11.3.5 in the book. It is elementary that the Lipschitz property is inherited by the in-probability distance p: Given $P(m(X,Y) > \delta) < \delta$, the Lipschitz property of g implies that $P(m'(g(X), g(Y)) > K\delta) < \delta$, so that $p(g(X), g(Y)) \leq (K \vee 1)p(X,Y)$. By the Strassen representation theorem, for X, Y and positive ϵ given, we can find $\tilde{X}, \tilde{Y}$ on a common probability space so that $\tilde{X} \stackrel{\mathrm{d}}{=} X$, $\tilde{Y} \stackrel{\mathrm{d}}{=} Y$ and

$$p(\tilde{X}, \tilde{Y}) \leq \pi(X, Y) + \epsilon \ .$$

Hence,

$$\pi(g(X), g(Y)) = \pi(g(\tilde{X}), g(\tilde{Y})) \leq p(g(\tilde{X}), g(\tilde{Y}))$$

and

$$p(g(\tilde{X}), g(\tilde{Y})) \leq (K \vee 1)p(\tilde{X}, \tilde{Y}) \leq (K \vee 1)(\pi(X, Y) + \epsilon) \ .$$

Since ϵ was arbitrary, we have the desired conclusion.

Now we consider the general case, for which we argue directly. Let B be the subset for which $P(Y \in B) = 1$. The Lipschitz property implies that

$$B \cap g^{-1}(A)^{\delta} \subseteq g^{-1}(A^{\epsilon}) \quad \text{in} \quad S \quad \text{for} \quad \delta \leq \epsilon/K$$

and any $A \in \mathcal{S}'$. Hence,

$$
\begin{aligned}
&\pi(g(X), g(Y)) \\
&= \inf\{\epsilon > 0 : P(g(X) \in A) \leq \epsilon + P(g(Y) \in A^{\epsilon}) \text{ for all } A \in \mathcal{S}'\} \\
&= \inf\{\epsilon > 0 : P(X \in g^{-1}(A)) \leq \epsilon + P(Y \in g^{-1}(A^{\epsilon})) \text{ for all } A \in \mathcal{S}'\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in g^{-1}(A) \leq \epsilon + P(Y \in B \cap g^{-1}(A)^{\delta}) \text{ for all } A \in \mathcal{S}'\right\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in g^{-1}(A) \leq \epsilon + P(Y \in g^{-1}(A)^{\delta}) \text{ for all } A \in \mathcal{S}'\right\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in A) \leq \epsilon + P(Y \in A^{\delta}) \text{ for all } A \in \mathcal{S}\right\} \\
&\leq (1 \vee K)\pi(X, Y) \ . \quad \blacksquare
\end{aligned}
$$

Example 1.5.1. *The advantage of the Prohorov metric on* $\mathcal{P}(\mathbb{R})$. Even on the real line $\mathbb{R}$, the Prohorov metric is useful to establish rate of convergence results, because the Lipschitz mapping theorem does not apply to two other metrics commonly used. On $\mathcal{P}(\mathbb{R})$ one often uses the *Lévy metric* λ, which is defined just as the Prohorov metric π in (2.2) except that only sets of the form $A = (-\infty, x]$ are used. The *uniform metric for cdf's* is also sometimes used; i.e.,

$$\|F_1 - F_2\| \equiv \mu(P_1, P_2) \equiv \sup\{(|P_1(A) - P_2(A)| : A = (-\infty, x]\} \ ,$$

where $F_i(x) \equiv P((-\infty, x])$. The uniform-cdf metric μ also induces weak convergence at limiting probability measures without atoms. However, the Lipschitz theorem is not valid for λ and μ. To see that, for $n \geq 1$, let

$$P(X_n = 2j) = P(Y_n = 2j + 1) = 1/n \quad \text{for} \quad 1 \leq j \leq n \ ,$$

and let $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$g(t) = \sin(\pi x/2) \quad \text{for} \quad t \in \mathbb{R} \ .$$

Clearly, g is Lipschitz with Lipschitz constant 1, but

$$\lambda(X_n, Y_n) \leq \mu(X_n, Y_n) = 1/n \ ,$$

while

$$P(g(X_n) = 0) = P(g(Y_n) \in \{-1, 1\}) = 1 \quad \text{for all} \quad n \ ,$$

so that

$$\mu(g(X_n), g(Y_n)) \geq \lambda(g(X_n), g(Y_n)) = 1/2 \quad \text{for all} \quad n \ .$$

Given a bound with the Prohorov metric π in $\mathcal{P}(\mathbb{R})$, we can obtain corresponding bounds with the metrics λ and μ. First we use the inequality $\lambda \le \pi$. In many cases we can relate λ and μ: When a probability measure P_1 on $\mathbb{R}$ has a Lipschitz cdf F_1 with Lipschitz constant c, i.e., when

$$|F_1(t_1) - F_1(t_2)| \le c|t_1 - t_2| ,$$

then we have the ordering

$$\mu(P_1, P_2) \le (1 + c)\lambda(P_1, P_2) \quad \text{for all} \quad P_2 \in \mathcal{P}(\mathbb{R}) . \quad \blacksquare \tag{5.1}$$

1.5.2. Proof of the Continuous-Mapping Theorems

We now turn to Theorem 3.4.3 of the book, following Billingsley (1968, Section 5), which we restate here. Let $Disc(g)$ be the set of discontinuity points of the function g.

Theorem 1.5.1. (continuous-mapping theorem) *If $X_n \Rightarrow X$ in (S, m) and $g : (S, m) \to (S', m')$ is measurable with $P(X \in Disc(g)) = 0$, then $g(X_n) \Rightarrow g(X)$.*

We first establish the measurability of $Disc(g)$ (even if g is not measurable).

Lemma 1.5.1. (measurability of the set of discontinuity points) *For $g : (S, m) \to (S', m')$, $Disc(g) \in \mathcal{S}$.*

Proof. For any $y, z \in S$ with $m'(g(y), g(z)) \ge \epsilon$ and $\epsilon > 0$, let

$$A_{\epsilon,\delta}(y, z) \equiv \{x \in S : m(x, y) < \delta \text{ and } m(x, z) < \delta\} .$$

Then the complement is

$$A^c_{\epsilon,\delta}(y, z) \equiv \{x \in S : m(x, y) \ge \delta \text{ or } m(x, z) \ge \delta\} .$$

It is easy to see that $A^c_{\epsilon,\delta}(y, z)$ is closed, so that $A_{\epsilon,\delta}(y, z)$ is open, as necessarily is

$$A_{\epsilon,\delta} \equiv \bigcup_y \bigcup_z A_{\epsilon,\delta}(y, z) .$$

Since

$$Disc(g) = \bigcup_\epsilon \bigcap_\delta A_{\epsilon,\delta} ,$$

where ϵ and δ run over the positive rationals, $Disc(g)$ is a $G_{\delta\sigma}$, implying that $Disc(g) \in \mathcal{S}$. $\blacksquare$

Proof of Theorem 1.5.1. By Theorem 11.3.4 (iii) in the book, it suffices to show that

$$\overline{\lim_{n\to\infty}} \; P(g(X_n) \in F) \le P(g(X) \in F)$$

for each closed subset $F \in \mathcal{S}'$. Given that $X_n \Rightarrow X$, we have

$$\begin{aligned} \overline{\lim_{n\to\infty}} \; P(g(X_n) \in F) &= \overline{\lim_{n\to\infty}} \; P(X_n \in g^{-1}(F)) \\ &\le \overline{\lim_{n\to\infty}} \; P(X_n \in g^{-1}(F)^-) \\ &\le P(X \in g^{-1}(F)^-) \; . \end{aligned}$$

However, $P(X \in g^{-1}(F)^-) = P(X \in g^{-1}(F))$ because $P(Disc(g)) = 0$ and $g^{-1}(F)^- \subseteq g^{-1}(F) \cup Disc(g)$. ∎

Finally, we treat Theorem 3.4.4 of the book, involving a sequence of measurable mappings:

Theorem 1.5.2. (generalized continuous-mapping theorem) *Let g and g_n, $n \ge 1$, be measurable functions mapping (S, m) into (S', m'). Let the range (S', m') be separable. Let E be the set of x in S such that $g_n(x_n) \to g(x)$ fails for some sequence $\{x_n : n \ge 1\}$ with $x_n \to x$ in S. If $X_n \Rightarrow X$ in (S, m) and $P(X \in E) = 0$, then $g_n(X_n) \Rightarrow g(X)$ in (S', m').*

Here we need to assume that the range is a separable metric space. Again we follow Billingsley (1968, Section 5).

Lemma 1.5.2. (measurability of the bad set) *Suppose that g_n, $n \ge 1$, and g are measurable functions from a metric space (S, m) into a separable metric space (S', m'). Let E be the set of x in S such that $g_n(x_n) \to g(x)$ fails for some sequence $\{x_n : n \ge 1\}$ with $m(x_n, x) \to 0$ as $n \to \infty$. Then E is a measurable subset of S.*

Proof. Let $B_{\epsilon,\delta,i}$ be the set of x in S such that $m'(g(x), g_i(y)) \ge \epsilon$ for some y with $m(x, y) < \delta$. Note that

$$E = \cup_{\epsilon} \cap_{\delta} \cap_{k\ge 1} \cup_{i\ge k} B_{\epsilon,\delta,i} \tag{5.2}$$

where ϵ and δ range over the positive rationals. We would be done if we could conclude that $B_{\epsilon,\delta,i}$ is measurable, but we do not know that. Note that $B_{\epsilon,\delta,i}$ is decreasing in ϵ. Hence (5.2) remains true if $B_{\epsilon,\delta,i}$ is replaced by

$B_{\epsilon/2,\delta,i}$. It thus suffices to show that, for all (ϵ, δ, i), there are sets $C_{\epsilon,\delta,i} \in \mathcal{S}$ such that

$$B_{\epsilon,\delta,i} \subseteq C_{\epsilon,\delta,i} \subseteq B_{\epsilon/2,\delta,i} \ . \tag{5.3}$$

Since (S', m') is separable, we can find a sequence $\{u_k : k \geq 1\}$ dense in S'. Let $A_{\epsilon,k} = \{x : m'(g(x), u_k) < \epsilon/4\}$ and note that $A_{\epsilon,k} \in \mathcal{S}$ and $S = \cup_k A_{\epsilon,k}$. Then (5.3) holds if

$$C_{\epsilon,\delta,i} = \cup_k (A_{\epsilon,k} \cap J_{\epsilon,\delta,i,k}) \ ,$$

where $J_{\epsilon,\delta,i,k}$ is the set of x such that $m'(g_i(y), g(z)) \geq \epsilon$ for some pair of points y, z in S with $m(x,y) < \delta$, $m(x,z) < \delta$ and $z \in A_{\epsilon,k}$. It is not difficult to see that $J^c_{\epsilon,\delta,i,k}$ is closed, so that $J_{\epsilon,\delta,i,k}$ is open and $C_{\epsilon,\delta,i} \in \mathcal{S}$. ∎

Proof of Theorem 1.5.2. By Lemma 1.5.2, $E \in \mathcal{S}$. From Theorem 11.3.4 (iv) in the book, it suffices to show that

$$P(g(X) \in G) \leq \varliminf_{n\to\infty} P(g_n(X_n) \in G)$$

for every open G in S'. If $x \in E^c$ and $g(x) \in G$, then there must exist k and δ such that $g_i(y) \in G$ if $i \geq k$ and $m(x,y) < \delta$, so that $x \in T_k^o$, the interior of T_k, where

$$T_k = \cap_{i \geq k} g_i^{-1}(G) \ .$$

Consequently,

$$g^{-1}(G) \subseteq E \bigcup \cup_{k=1}^{\infty} T_k^o \ .$$

Since $P(X \in E) = 0$ and $T_k^o \subseteq T_{k+1}^o$, for any given ϵ there is a k such that

$$P(X \in g^{-1}(G)) \leq P(X \in \cup_k T_k^o) \leq P(X \in T_k^o) + \epsilon$$

for $k \geq k_0$. Since $X_n \Rightarrow X$ and $T_k \subseteq g_n^{-1}(G)$ for $n \geq k$,

$$P(X \in T_k^o) \leq \varliminf_{n\to\infty} P(X_n \in T_k^\circ) \leq \varliminf_{n\to\infty} P(X_n \in g_n^{-1}(G)) \ .$$

Since ϵ was arbitrary, the proof is completed by combining these two strings of inequalities. ∎

The continuous-mapping approach to stochastic-process limits leads us to focus on the underlying sample paths of the stochastic processes. Thus the continuous-mapping approach is a *sample-path method.* In recent years, probabilists have tended to favor sample-path methods over more traditional analytic methods, because they are less removed from the phenomenon under study. However, the two approaches often can be fruitfully combined.

Many traditional analytic results are based on transforms, such as the characteristic function (version of the Fourier transform), probability generating function (or z transform) and the Laplace transform, as can be seen from Feller (1971). Fortunately, the analytic approach has been applied with great success over the years to yield explicit expressions for many probability distributions of interest in the form of transforms. That is true for many of the limit processes that we will consider. Thus we can use previous analytic results to obtain explicit transforms for approximating distributions. We then can apply numerical transform inversion to compute the probability distribution itself; e.g., see Abate and Whitt (1992a, 1995), Choudhury, Lucantoni and Whitt (1994) and Abate, Choudhury and Whitt (1999).

For example, as shown in Section 8.5 of the book and Section 5.2 here, the heavy-traffic limit for a queue with heavy-tailed distributions is often a reflection of a stable Levy motion or more general Levy process without negative jumps. These limit processes are somewhat complicated, but fortunately the analytic approach has shown that the steady-state distribution has a relatively simple expression via its Laplace transform, which is known as the generalized Pollaczek-Khintchine transform. Thus we can calculate the steady-state distribution of the limit process by applying numerical transform inversion.

Chapter 2

Stochastic-Process Limits

2.1. Introduction

Chapters 4 and 7 of the book present a panorama of stochastic-process limits. In this chapter we present even more material. In Section 2.2 we present an introduction to strong approximations and the rates of convergence in the setting of Donsker's theorem that they imply using the Prohorov metric. In Section 2.3 we present additional Brownian limits under weak dependence; here we focus on Markov and regenerative structure.

In Section 2.4 we briefly discuss the convergence to general Lévy processes that holds when we have a sequence of random walks (based on a double sequence of random walk steps). Finally, in Section 2.5 we point out that the linear-process representation assumed with strong dependence in Sections 4.6 and 4.7 of the book arises naturally from modelling when we take a time-series perspective.

2.2. Strong Approximations and Rates of Convergence

In Sections 1.4 and 4.3 of the book we noted that the CLT and FCLT are invariance principles, meaning that the same limits occur in great generality. In the IID case we only need the summands X_n to have finite variance. However, the quality of the approximation for any given n is affected by the distribution of X_n. Indeed, that is obvious for the CLT: If $X_n \stackrel{\mathrm{d}}{=} N(0, \sigma^2)$, then the limit can be replaced by equality in distribution. Moreover, the closer the distribution of X_n is to the normal distribution, the better the normal approximation for the scaled partial sum should be. More generally, the advantage of extra structure in the distribution of X_n can be seen from more refined results giving bounds on the rate of convergence and asymptotic expansions. We review some of these results in this section.

2.2.1. Rates of Convergence in the CLT

A bound on the rate of convergence in the basic CLT, given a finite third absolute moment of a summand, is provided by the Berry-Esseen theorem; see p. 542 of Feller (1971). To state it, we use the uniform metric on cdf's, defined by

$$\|F_1 - F_2\| \equiv \sup_x |F_1(x) - F_2(x)| \ . \tag{2.1}$$

As before, let Φ be the standard normal cdf.

Theorem 2.2.1. (Berry-Esseen theorem) *Let $\{X_n\}$ be a sequence of IID random variables with $EX_1 = 0$, $E[X_1^2] = \sigma^2$ and $E[|X_1|^3] = \delta_3 < \infty$. Then*

$$\|F_n - \Phi\| \le 3\delta_3/\sigma^3\sqrt{n} \quad \textit{for all} \quad n \ ,$$

where $F_n(x) \equiv P((n\sigma^2)^{-1/2}(X_1 + \cdots + X_n) \le x)$.

Theorem 2.2.1 implies that for given n and σ^2, the bound on the distances decreases as the third absolute moment δ_3 decreases. We now describe the Edgeworth expansion, which shows how further regularity conditions can improve the quality of the normal approximation; see p. 535 of Feller (1971). We also get convergence of pdf's.

Theorem 2.2.2. (Edgeworth expansion) *If, in addition to the assumptions of Theorem 2.2.1 above, moments $E[X_1^k]$ exist for $3 \le k \le r$ and $|E[exp(itX_1)|^\nu$ is integrable for some $\nu \ge 1$, then $(n\sigma^2)^{-1/2}(X_1 + \cdots + X_n)$ has a pdf f_n for all n and*

$$f_n(x) = n(x)[1 + \sum_{k=3}^{r} n^{-(k-2)/2} P_k(x) + o(n^{-(r-2)/2})]$$

as $n \to \infty$, uniformly in x, where n is the standard normal pdf and $P_k(x)$ is a real polynomial depending on the first k moments of X_1, with the property that $P_k(x) = 0$ if the first k moments of X_1 agree with those of the standard normal distribution.

Note that the rate of convergence in Theorem 2.2.2 is $O(n^{-1/2})$ if $E[X_1^3] \neq 0$, but is $O(n^{-1})$ or better if $E[X_1^3] = 0$. When $E[X_1^3] \neq 0$, the refinement provided by the second term can be useful.

2.2.2. Rates of Convergence in the FCLT

We now turn to Donsker's FCLT. From the Lipschitz mapping theorem, Theorem 3.4.2 in the book, we can deduce a bound on the rate of convergence in the CLT from a bound on a rate of convergence in the FCLT. Hence, we can see in advance that the rate of convergence in the FCLT, given a finite third absolute moment, can be no better than the $O(n^{-1/2})$ bound provided by the Berry-Esseen theorem. In fact, the best possible bound for the FCLT, under an even stronger regularity condition, is somewhat worse, being larger by a factor of $\log n$. From a practical perspective, though, the difference is not great.

We now give the final rate-of-convergence result, expressed in terms of the Prohorov metric π from Section 3.2 of the book; see (2.2) here. For this application, it is convenient to let the underlying function space be the set $D_Q \equiv D_Q([0,1],\mathbb{R})$ of functions in $D \equiv D([0,1],\mathbb{R})$ with discontinuities only at rational points in the domain $[0,1]$, endowed with the uniform metric $\|\cdot\|$; we refer to the space as (D_Q, U). The space (D_Q, U) is a separable metric space and the stochastic processes considered here all have sample paths in this space. Thus, the Prohorov metric π is defined on the space $\mathcal{P}((D_Q, U))$, the space of all probability measures on (D_Q, U). Since

$$d_{M_1}(x_1, x_2) \le d_{J_1}(x_1, x_2) \le \|x_1 - x_2\| \quad \text{for} \quad x_1, x_2 \in D \ ,$$

the result also holds for the spaces (D, d_{J_1}) and (D, d_{M_1}).

The following combines Theorems 1.16 and 1.17 in Csörgő and Horváth (1993).

Theorem 2.2.3. (bounds on the rate of convergence in Donsker's FCLT) *Let $\{X_n\}$ be a sequence of IID random variables with $EX_n = 0$ and $E[X_1^2] = \sigma^2$. If, in addition, $E[exp(tX_1)] < \infty$ for t in a neighborhood of the origin, then there exist positive constants C_1 and C_2 such that*

$$C_1 \log n/\sqrt{n} \le \pi(\mathbf{S}_n, \sigma\mathbf{B}) \le C_2 \log n/\sqrt{n} \tag{2.2}$$

for all n, where π is the Prohorov metric on the space $\mathcal{P}((D_Q, U))$, $\mathbf{B}$ is standard Brownian motion and $\mathbf{S}_n(t) \equiv n^{-1/2} S_{\lfloor nt \rfloor}$, $0 \le t \le 1$. If, instead, only $E[|X_1|^p < \infty$ for some $p > 2$, then there is a constant C such that

$$\pi(\mathbf{S}_n, \sigma\mathbf{B}) \le C n^{-(p-2)/2(p+1)} \tag{2.3}$$

for all n. Moreover, for any sequence $\{a_n\}$ with $a_n \to \infty$ as $n \to \infty$, there is a random variable X_1 with $E[|X_i|^p] < \infty$ such that

$$\overline{\lim_{n\to\infty}} \ a_n n^{(p-2)/2(p+1)} \pi(\mathbf{S}_n, \sigma\mathbf{B}) = \infty \ . \tag{2.4}$$

The lower bound in (2.2) and the limit in (2.4) show that the upper bounds in Theorem 2.2.3 are indeed best possible. Note that the rate $O(\log n/\sqrt{n})$ in (2.2) exceeds the Berry-Esseen bound $O(1/\sqrt{n})$ by a factor of $\log n$. We regard that difference as negligible.

However, there is a big difference between the bounds in (2.3) and in Theorem 2.2.2. When there is only a finite third absolute moment, we have (2.3) with $p = 3$, which only yields the rate $O(n^{-1/8})$. For finite p^{th} moment with $p > 2$, (2.3) gives a rate that can be substantially worse than $O(n^{-1/2})$, while Theorem 2.2.2 gives rates that can be much better than $O(n^{-1/2})$. It should be recognized that the conditions are quite different though.

By the Lipschitz mapping theorem, Theorem 3.4.2 of the book, the rate of convergence in Theorem 2.2.3 is inherited by Lipschitz functions. For real-valued Lipschitz functions, we then can obtain bounds on the uniform metric for cdf's.

Corollary 2.2.1. (bounds on the uniform metric for cdf's of the images of real-valued Lipschitz maps) *Suppose that* $g : (D_Q, U) \to \mathbb{R}$ *is a Lipschitz function and that* $g(\mathbf{B})$ *has a bounded pdf. If the conditions of Theorem 2.2.3 hold with* $E exp(tX_1) < \infty$ *for* t *in a neighborhood of the origin, then there is a positive constant* C *such that*

$$\sup_x |P(g(\mathbf{S}_n) \leq x) - P(g(\sigma \mathbf{B}) \leq x)| \leq C \log n/\sqrt{n} \qquad (2.5)$$

for all $n \geq 1$.

We can apply Corollary 2.2.1 to obtain a bound on the rate of convergence in the CLT; we use the projection map $\pi_1(x) \equiv x(1)$, which is easily seen to be Lipschitz. However, the bound is not as good as provided by the Berry-Esseen theorem, so the bound may no longer be best possible when we consider the image measure associated with a single Lipschitz map.

We can also apply Theorem 2.2.3 to establish bounds on the rate of convergence in heavy-traffic FCLTs for queues. We illustrate by stating a result for the queueing model in Section 1.6. We use the fact that the two-sided reflection map $\phi_K : D \to D$ is Lipschitz; see Theorem 13.10.1. An early result of this kind is Kennedy (1973). That served as motivation for the Lipschitz mapping theorem in Whitt (1974).

Corollary 2.2.2. (bounds on the rate of convergence in a heavy-traffic stochastic-process limit for queues) *Consider the queueing model in Section* 2.3 *of the book with IID inputs* V_k *with mean* m_v *and variance* σ^2*.*

If, in addition, $K_n = n^{1/2}K$ *and* $\mu_n = m_v + mn^{-1/2}$ *for all* n *and with* $E[exp(tV_1)] < \infty$ *for some* $t > 0$*, then there exists a constant* C *such that*

$$\pi(\mathbf{W}_n, \phi_K(\sigma\mathbf{B} - m\mathbf{e})) \leq C \log n / n^{1/2} \ ,$$

where $\mathbf{W}_n$ *is the scaled workload process in equation* (2.3.6) *of the book and* ϕ_K *is the two-sided reflection map.*

2.2.3. Strong Approximations

Theorem 2.2.3 can be extablished by applying *strong approximations.* Like the Skorohod and Strassen representation theorems in Chapters 3 and 11 of the book, strong approximations are special constructions of random objects on the same underlying probability space, often called couplings; see Lindvall (1992).

We start by stating the Komlós, Major and Tusnády (1975, 1976) strong approximation theorems for partial sums of IID random variables; see Chapter 2 of Csörgő and Révész (1981) and Chapter 1 of Csörgő and Horváth (1993). See Philipp and Stout (1975) for extensions to the weakly dependent case and Einmahl (1989) for extensions to the multivariate case. See Csörgő and Horvath (1993) for strong approximations of renewal processes and random sums. For applications of strong approximations to queues, see Zhang et al. (1990), Horváth (1990), Glynn and Whitt (1991a,b) and Chen and Mandelbaum

Theorem 2.2.4. (strong approximation with finite moment generating function) *Let* $\{X_n : n \geq 1\}$ *be a sequence of IID random variables with* $EX_1 = 0$, $EX_1^2 = 1$ *and* $Ee^{tX_1} < \infty$ *for* t *in a neighborhood of the origin. Let* $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, *with* $S_0 \equiv 0$*. Then there exists a standard Brownian motion* $\mathbf{B} \equiv \{\mathbf{B}(t) : t \geq 0\}$ *such that, for all real* x *and every* $n \geq 1$,

$$P\left(\max_{1 \leq k \leq n} |S_k - \mathbf{B}(k)| > C_1 \log n + x\right) < C_2 e^{-\lambda x} \ , \tag{2.6}$$

where C_1, C_2 *and* λ *are positive constants depending upon the distribution of* X_1.

As a consequence of Theorem 2.2.4, we can deduce that

$$S_n - \mathbf{B}(n) = O(\log n) \quad \text{as} \quad n \to \infty \quad \text{w.p.1} \ ; \tag{2.7}$$

i.e., there is a constant C such that

$$P(|S_n - \mathbf{B}(n)| > C \log n \quad \text{infinitely often}) = 0 \ . \tag{2.8}$$

Note that (2.8) follows from (2.6) by substituting $C' \log n$ for x in (2.6) for suitably large C' and then applying the Borel-Cantelli theorem.

We now relax the extra condition on the tail of the ccdf $P(|X_1| > t)$, at the expenses of obtaining a slower rate.

Theorem 2.2.5. (strong approximation with p^{th} moment) *Let $\{X_n : n \geq 1\}$ be a sequence of IID random variables with $EX_1 = 0$, $EX_1^2 = 1$ and $E|X_1|^p < \infty$ for some $p > 2$. Let $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, with $S_0 \equiv 0$. Then there exists a standard Brownian motion $\mathbf{B}$ such that*

$$n^{-1/p}|S_n - \mathbf{B}(n)| \to 0 \quad \text{w.p.1} \tag{2.9}$$

To apply Theorems 2.2.4 and 2.2.5 to establish Theorem 2.2.3, we need to relate Brownian motion $\mathbf{B}$ to the associated processes

$$\mathbf{B}_n^1(t) \equiv n^{-1/2}\mathbf{B}(\lfloor nt \rfloor), \quad \mathbf{B}_n^2(t) \equiv \mathbf{B}(\lfloor nt \rfloor / n), \quad \mathbf{B}_n^3(t) \equiv n^{-1/2}\mathbf{B}(nt)$$

for $0 \leq t \leq 1$. By the self-similarity property, $\mathbf{B} \stackrel{\text{d}}{=} \mathbf{B}_n^3$ and $\mathbf{B}_n^1 \stackrel{\text{d}}{=} \mathbf{B}_n^2$ for all $n \geq 1$. We can relate $\mathbf{B}_n^2$ to $\mathbf{B}$ by bounding the fluctuations of Brownian motion. The following is Lemma 1.1.1 of Csörgő and Révész (1981).

Theorem 2.2.6. (uniform bound on the fluctuations of Brownian motion) *For any $\epsilon > 0$, there exists a constant $C = C(\epsilon)$ such that*

$$P(\sup_{0 \leq t \leq T-h} \sup_{0 \leq s \leq h} |\mathbf{B}(t+s) - \mathbf{B}(t)| \geq \nu\sqrt{h}) \leq (CT/h)exp(-\nu^2/(2+\epsilon)) \tag{2.10}$$

for all positive ν, T, and h, $0 < h < T$.

Theorem 2.2.6 can be applied to determine the precise modulus of continuity of Brownian sample paths (originally determined by Lévy); see Theorem 1.1 of Csörgő and Révész (1981).

Theorem 2.2.7. (modulus of continuity of Brownian paths) *If $\mathbf{B}$ is Brownian motion, then*

$$\lim_{h \to 0} \sup_{0 \leq s \leq 1} \sup_{0 \leq t \leq h} \frac{|\mathbf{B}(s+t) - \mathbf{B}(s)|}{\sqrt{2h \log h^{-1}}} = 1 \quad \textit{w.p.1} .$$

From Theorem 2.2.7, we see that the sample paths of Brownian motion are continuous but not differentiable; the largest increment of length h is almost surely of order $O(\sqrt{2h \log h^{-1}})$. We can also apply Theorem 2.2.6 to determine the following bound on the in-probability distance $p(\mathbf{B}, \mathbf{B}_n^2)$ and the Prohorov distance $\pi(\mathbf{B}, \mathbf{B}_n^1)$, where π is defined on the space $\mathcal{P}((C, U))$.

Corollary 2.2.3. *There exists a constant* C_1 *such that*

$$\pi(\mathbf{B}, \mathbf{B}_n^1) \leq p(\mathbf{B}, \mathbf{B}_n^2) \leq C_1 \sqrt{\log n / n}$$

for all $n \geq 1$.

Proof. The first inequality holds because $\mathbf{B}_n^1 \stackrel{\mathrm{d}}{=} \mathbf{B}_n^2$ and $\pi \leq p$. For the second inequality, let $\nu = \sqrt{c \log n}$ for $c > 4$ in (2.10). Then the right hand side of (2.10) for $T = 1$ becomes $C' n^{-(1+\delta)}$ for $\delta > 0$ and constant C'. ■

Partial proof of Theorem 2.2.3. For the upper bound in (2.1), let $x = C_3 \log n$ in (2.6) to obtain

$$\pi(\mathbf{S}_n, \mathbf{B}_n^1) \leq p(\mathbf{S}_n, \mathbf{B}_n^1) \leq C \log n / \sqrt{n} \ .$$

Then use the triangle inequality with Corollary 2.2.3. ■

Theorem 2.2.4 can be applied to obtain a strong approximation for a Lévy process, i.e., a random element of D with stationary and independent increments; see Corollary 5.5 on p. 359 of Ethier and Kurtz (1986).

Theorem 2.2.8. (strong approximation for a Lévy process) *Let* $\{\mathbf{L}(t) : t \geq 0\}$ *be a real-valued Lévy process. Assume that*

$$Ee^{\alpha \mathbf{L}(1)} < \infty \tag{2.11}$$

for all α *with* $|\alpha| \leq \alpha_0$ *for some* $\alpha_0 > 0$. *Then there exist versions of the Lévy process* $\mathbf{L}$ *and a standard Brownian motion* $\mathbf{B}$ *on a common probability space such that*

$$|\mathbf{L}(t) - mt - \sigma \mathbf{B}(t)| = O(\log t) \quad as \quad t \to \infty \quad w.p.1 \ , \tag{2.12}$$

where $m = E\mathbf{L}(1)$ *and* $\sigma^2 = Var\, \mathbf{L}(1)$.

A precursor to the strong approximation theorems, of interest in its own right, is the Skorohod (1961) embedding theorem; see p. 88 of Csörgő and Révész (1981).

Theorem 2.2.9. (Skorohod embedding theorem) *Let* $\{X_n : n \geq 1\}$ *be a sequence of IID real-valued random variables with* $EX_1 = 0$ *and* $EX_1^2 = 1$. *Let* $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, *with* $S_0 \equiv 0$. *There exists a probability space supporting a standard Brownian motion* $\mathbf{B}$ *and a sequence* $\{T_n : n \geq 1\}$ *of nonnegative IID random variables such that* *(i)* $\{\mathbf{B}(T_1 + \cdots T_n) : n \geq 1\} \stackrel{\mathrm{d}}{=} \{S_n : n \geq 1\}$ *in* $\mathbb{R}^\infty$*;*

(ii) $\{T_1 + \cdots + T_n : n \geq 1\}$ *is a sequence of stopping times, i.e., the event* $\{T_1 + \cdots + T_n \leq t\}$ *is contained in the* σ*-field generated by* $\{\mathbf{B}(s) : 0 \leq s \leq t\}$ *for all* $t \geq 0$*;*

(iii) $ET_1 = 1$*;*

(iv) $ET_1^k < \infty$ *if, in addition,* $EX^{2k} < \infty$ *for positive integer* k*.*

As a consequence of Theorem 2.2.9,

$$\{n^{-1/2}S_{\lfloor nt \rfloor} : t \geq 0\} \stackrel{\mathrm{d}}{=} \{n^{-1/2}B(T_1 + \cdots + T_{\lfloor nt \rfloor}) : t \geq 0\}$$
$$\stackrel{\mathrm{d}}{=} \{B(n^{-1}(T_1 + \cdots + T_{\lfloor nt \rfloor}) : t \geq 0\} .$$

By the FSLLN,

$$\sup_{0 \leq t \leq u} |n^{-1}(T_1 + \cdots + T_{\lfloor nt \rfloor}) - t| \to 0 \quad \text{w.p.1} ,$$

so that Donsker's theorem again is a consequence. Rate of convergence results follow too.

2.3. Weak Dependence from Regenerative Structure

This section is a sequel to Section 4.4 in the book, in which we showed that many Brownian limits still hold for random walks $\{S_n : n \geq 0\}$ when the IID condition on the sequence of steps $\{X_n : n \geq 1\}$ is relaxed, with the finite-second-moment condition $EX_n^2 < \infty$ remaining in place. We now obtain results for stochastic-processes with regenerative structure.

This new setting allows us to abandon the assumption of stationarity and obtain explicit expressions for the asymptotic variance σ^2, defined by

$$\sigma^2 \equiv \lim_{n \to \infty} \frac{Var(S_n)}{n} . \tag{3.1}$$

For a stationary sequence $\{X_n\}$, the asymptotic variance has the representation

$$\sigma^2 = Var\, X_n + 2 \sum_{k=1}^{\infty} Cov(X_1, X_{1+k}) . \tag{3.2}$$

We now obtain more explicit representations for the asymptotic variance in terms of basic model elements.

2.3.1. Discrete-Time Markov Chains

We start by stating results for finite-state Markov chains. We first consider discrete-time chains and then we consider continuous-time chains. Afterwards, we state results for general regenerative processes, which cover more general Markov processes and nonMarkov processes. The first result for DTMC's extends Theorem 4.4.2 in the book. An important point is that an explicit expression can be given for the asymptotic variance σ^2. It is expressible as a function of the fundamental matrix of the DTMC. The most effective way to calculate the asymptotic variance is usually to solve a system of equations, collectively known as the *Poisson equation.*

Let P be the transition matrix of an irreducible k-state DTMC and let Π be a matrix with each row being the steady-state vector π. (We will work with row vectors; let A^t be the transpose of a matrix A, so that the column vector associated with a row vector x is x^t.) Then the *fundamental matrix* of the DTMC is

$$Z \equiv (I - P + \Pi)^{-1} \; ; \tag{3.3}$$

see pp. 75, 100 of Kemeny and Snell (1960). (The matrix $I - P + \Pi$ is nonsingular.)

Theorem 2.3.1. (FCLT for a DTMC with explicit asymptotic variance) *Let $\{Y_n : n \geq 1\}$ be an irreducible k-state DTMC and let $X_n = f(Y_n)$ for a real-valued function f. Then the FCLT*

$$\mathbf{S}_n \Rightarrow \sigma \mathbf{B} \quad \textit{in} \quad (D, J_1) \; , \tag{3.4}$$

where $\mathbf{B}$ *is standard Brownian motion and*

$$\mathbf{S}_n(t) = n^{-1/2}(S_{\lfloor nt \rfloor} - mnt), \quad t \geq 0 \; , \tag{3.5}$$

holds with

$$m \equiv \sum_{i=1}^{k} \pi_i f(i) \; ,$$

$$\sigma^2 \equiv 2 \sum_{i=1}^{k} \sum_{j=1}^{k} (f(i) - m) \pi_i Z_{i,j} (f(j) - m) - \sum_{i=1}^{k} \pi_i (f(i) - m)^2 \; , \tag{3.6}$$

π the steady-state vector and $Z \equiv (Z_{i,j})$ the fundamental matrix in (3.3).

As a quick sanity check on (3.6), note that in the IID case we have $P = A$, $Z = I$ and, from (3.6),

$$\sigma^2 = \sum_{i=1}^{k} \pi_i (f(i) - m)^2 \ ,$$

as we should.

It is significant that we can calculate π, m, Z and σ^2 in Theorem 2.3.1 by solving the Poisson equation(s). We state both row-vector and column-vector versions. Let $\mathbf{1} \equiv (1, \ldots, 1)$ be a vector of 1's and $\mathbf{0} \equiv (0, \ldots, 0)$ be a vector of 0's.

Theorem 2.3.2. (Poisson equations for a DTMC) *Consider an irreducible finite-state DTMC with transition matrix P. The row-vector version of the Poisson equation*

$$x(I - P) = y \tag{3.7}$$

has a solution x for given y if and only if $y\mathbf{1}^t = 0$. All solutions to (3.7) *are of the form*

$$x = yZ + (x\mathbf{1}^t)\pi \ .$$

The column-vector version of the Poisson equation

$$(I - P)x^t = y^t \tag{3.8}$$

has a solution x^t for given y^t if and only if $\pi y^t = 0$. All solutions to (3.8) *are of the form*

$$x^t = Zy^t + (\pi x^t)\mathbf{1} \ .$$

Proof. We consider only the row-vector form. Clearly $y\mathbf{1}^t = 0$ is necessary, because $(I - P)\mathbf{1}^t = 0^t$. Given (3.7),

$$x(I - P + \Pi) = y + (x\mathbf{1}^t)\pi \ ,$$

but $I - P + \Pi$ is nonsingular with inverse Z, so that

$$x = yZ + (x\mathbf{1}^t)\pi Z = yZ + (x1^t)\pi$$

since $\pi Z = Z$. ∎

Theorem 2.3.3. (Poisson equations for the steady-state vector and the asymptotic variance of a DTMC) *For an irreducible finite-state DTMC, the steady-state vector π is the unique solution x to the Poisson equation* (3.7) *with* $y = (0, \ldots, 0)$ *and* $x\mathbf{1}^t = 1$. *The asymptotic variance can be expressed as*

$$\sigma^2 = 2 \sum_{i=1}^{k} x_i (f(i) - m)$$

where m *is the mean and* x *solves the Poisson equation* (3.7) *with*

$$y_i = (f(i) - m)\pi_i, \quad 1 \le i \le k \ .$$

2.3.2. Continuous-Time Markov Chains

We now turn to the continuous-time processes. There are analogs of the DTMC results in Theorems 2.3.1–2.3.3 for CTMC's. Let $\{(Y(t) : t \ge 0\}$ be an irreducible k-state CTMC. Then the limit is for the integral

$$S(t) \equiv \int_0^t f(Y(s))ds, \quad t \ge 0 \ .$$

The associated normalized processes in D are

$$\mathbf{S}_n(t) \equiv n^{-1/2}(S(nt) - mnt), \quad t \ge 0 \ . \tag{3.9}$$

Given transition matrices $P(t) \equiv (P_{i,j}(t))$, where

$$P_{i,j}(t) \equiv P(Y(t) = j | Y(0) = i) \ ,$$

the *infinitesimal generator matrix* of the CTMC is $Q \equiv (Q_{i,j})$ where

$$Q \equiv \lim_{t \downarrow 0} (P(t) - I)$$

and the *fundamental matrix* is $Z \equiv (Z_{i,j})$ where

$$Z_{i,j} \equiv \int_0^\infty (P_{i,j}(t) - \pi_j)dt$$

and

$$Z = (\Pi - Q)^{-1} - \Pi \tag{3.10}$$

see Kemeny and Snell (1961) and Whitt (1992). A CTMC model is usually specified by giving the infinitesimal generator matrix Q. For an irreducible finite-state CTMC, the steady-state vector π is the unique vector with sum 1 that satisfies

$$\pi Q = 0 \ .$$

Paralleling (3.1) and (3.2) above, the asymptotic variance in this continuous-time framework is

$$\sigma^2 \equiv \lim_{t\to\infty} \frac{Var(S(t))}{t} = 2\int_0^\infty r(t)dt \ ,$$

where $r(t)$ is the (*auto*) *covariance function*, i.e.,

$$r(t) \equiv E[X(0)X(t)] - (E[X(0)])^2$$

for $X(t) \equiv f(Y(t))$, $t \geq 0$.

The following is the continuous-time analog of Theorem 2.3.1.

Theorem 2.3.4. (FCLT for a CTMC with explicit asymptotic variance) *Let $\{Y(t) : t \geq 0\}$ be an irreducible k-state CTMC, and let $X(t) = f(Y(t))$ for a real-valued function f. Then the FCLT* (3.4) *holds for $\mathbf{S}_n$ in* (3.9) *with m the steady-state mean and σ^2 the asymptotic variance, which can be expressed as*

$$\sigma^2 \equiv 2\sum_{i=1}^{k}\sum_{j=1}^{k} f(i)\pi_i Z_{i,j} f(j) \ ,$$

where Z is the fundamental matrix in (3.10).

We can calculate π, m, Z and σ^2 by solving Poisson equations for CTMC's; see Whitt (1992). The following is the continuous-time analog of Theorem 2.3.2.

Theorem 2.3.5. (Poisson equations for a CTMC) *Consider an irreducible finite-state CTMC with infinitesimal generator matrix Q. The row-vector version of the Poisson equation*

$$xQ = y \tag{3.11}$$

has a solution x for given y if and only if $y\mathbf{1}^t = 0$. All solutions to (3.11) *are of the form*

$$x = -yZ + (x\mathbf{1}^t)\pi \ .$$

The column-vector version of the Poisson equation

$$Qx^t = y^t$$

has a solution x^t for given y^t if and only if $\pi y^t = 0$. All solutions are of the form

$$x^t = -Zy^t + (\pi x^t)\mathbf{1}^t \ .$$

The following is the continuous-time analog of Theorem 2.3.3.

Theorem 2.3.6. (Poisson equations for the steady-state vector and the asymptotic variance of a CTMC) *For an irreducible finite-state CTMC, the steady-state vector* π *is the unique solution* x *to the Poisson equation* (3.11) *with* $y = (0, \ldots, 0)$ *and* $x\mathbf{1}^t = 1$. *The asymptotic variance can be expressed as*

$$\sigma^2 = 2 \sum_{i=1}^{k} x_i f_i \ ,$$

where x *is the unique solution to the Poisson equation* (3.11) *with*

$$y_i = (f_i - m)\pi_i \quad \textit{and} \quad \sum_{i=1}^{k} x_i = 0 \ .$$

and m *is the mean.*

We can also obtain even more explicit expressions for the asymptotic variance in Markov chains with additional structure. For example, suppose that the CTMC $\{Y(t) : t \geq 0\}$ is a birth-and-death processes on the integers $\{0, 1, \ldots, n\}$ with positive birth rates λ_i, death rates μ_i and stationary probabilities

$$\pi_i = \frac{\pi_0 \lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \ . \tag{3.12}$$

If the process is irreducible, then the process must be reflecting at 0 and n; i.e., $\lambda_n = \mu_0 = 0$.) The following is Proposition 1 of Whitt (1992). Corresponding results for diffusion processes are also stated there.

Theorem 2.3.7. (asymptotic variance of a birth-and-death process) *Suppose that* $X(t) = f(Y(t))$, *where* f *is a real-valued function and* $\{Y(t) : t \geq 0\}$ *is an irreducible birth-and-death process on the integers* $\{0, 1, \ldots, n\}$ *with birth rates* λ_i *and death rates* μ_i. *Then the asymptotic variance can be expressed as*

$$\sigma^2 = 2 \sum_{j=0}^{n-1} (\lambda_j \pi_j)^{-1} \left[\sum_{i=0}^{j} (f(i) - m)\pi_i \right]^2$$

for m *the mean and* π *in* (3.12) *above.*

We now state a corollary of Theorem 2.3.7 for an elementary queueing model – the $M/M/1$ queue. The queue-length process in an $M/M/1$ queue is a birth-and-death process with $\lambda_i = \lambda$ and $\mu_i = \mu$ when positive. The following would properly be a corollary to Theorem 2.3.7 except for the fact that the state space is infinite. Extensions to countably infinite and more general state spaces are covered by the results for regenerative processes below.

Corollary 2.3.1. (asymptotic variance for the queue-length process in the M/M/1 queue) *For the queue-length (number in system) process in the M/M/1 queue with traffic intensity* $\rho \equiv \lambda/\mu < 1$, *the asymptotic variance is*

$$\sigma^2 = \frac{2\rho(1+\rho)}{(1-\rho)^4} \ . \tag{3.13}$$

The $(1-\rho)^4$ term in the denominator of (3.13) shows that very long simulation runs are required to directly estimate the steady-state mean of the queue-length process by the sample mean when ρ is close to its upper limit 1. That insight is important for related models for which we do not already know the steady-state distribution, so that simulation is actually needed. We discuss applications of stochastic-process limits to obtain insights about simulation in Section 5.9 of the book.

For a birth-and-death process it is also possible, and usually preferable, to recursively solve the Poisson equation, see Remarks 1, 2 and 5 of Whitt (1992). For more on Poisson equations, see Glynn (1994) and Glynn and Meyn (1996).

2.3.3. Regenerative FCLT

Donsker's theorem itself applies quite directly when we have regenerative structure, as in the case of DTMC's and CTMC's in Theorem 2.3.1 and 2.3.4 above. For this discussion, we use the classical definition of regenerative process, meaning that the process splits into IID cycles; see p. 125 of Asmussen (1987). We will present the result in continuous time, following Glynn and Whitt (1993), but corresponding results hold in discrete time, as in Glynn and Whitt (1987). An earlier related Markov chain FCLT is due to Maigret (1978).

Consider a stochastic process $\{Y(t) : t \geq 0\}$ with general state space and a measurable real-valued function f on that state space. We assume that the stochastic process $\{Y(t) : t \geq 0\}$ is regenerative with respect to regeneration times T_i satisfying

$$0 \leq T_0 < T_1 < \cdots$$

with $T_{-1} \equiv 0$. We focus on the associated *cumulative process*

$$C(t) \equiv \int_0^t f(Y(s))ds, \quad t \geq 0 , \tag{3.14}$$

and consider the associated normalized processes

$$\mathbf{C}_n(t) \equiv n^{-1/2}(C(nt) - mnt), \quad t \geq 0 \tag{3.15}$$

where m is a real number yet to be specified. The key random variables associated with the regenerative cycles are

$$\begin{aligned} \tau_i &\equiv T_i - T_{i-1} , \\ X_i &\equiv X_i(m) \equiv \int_{T_{i-1}}^{T_i} [f(Y(u)) - m]du , \\ Z_i &\equiv Z_i(m) \equiv \sup_{0 \leq s \leq \tau_i} \left| \int_0^s [f(Y(T_{i-1} + u)) - m]du \right| . \end{aligned} \tag{3.16}$$

By *regenerative structure* we mean that the three-tuples (τ_i, X_i, Z_i) are IID for $i \geq 1$. We also assume that $E\tau_i < \infty$ and

$$\int_0^t |f(Y(s))|ds < \infty \text{ w.p.1} \quad \text{for each} \quad t ,$$

which implies that the cumulative process has continuous sample paths w.p.1.

The general idea is that the cumulative process C in (3.14) is approximately equal to a random sum. In particular,

$$C(t) = S_{N(t)} + R_1(t) + R_2(t), \quad t \geq 0 ,$$

where

$$S_n \equiv X_1 + \cdots X_n, \quad n \geq 1 ,$$

for X_i in (3.16) with $S_0 \equiv 0$, $N \equiv \{N(t) : t \geq 0\}$ is the (possibly delayed) renewal counting process associated with the regeneration times, i.e.,

$$N(t) \equiv \max\{i : T_i \leq t\}, \quad t \geq 0 ,$$

and $R_i \equiv \{R_i(t) : t \geq 0\}$ are remainder processes, defined by

$$R_1(t) \equiv \int_0^{\min\{t,T_0\}} f(Y(s))ds \tag{3.17}$$

and

$$R_2(t) = \int_{T_{N(t)}}^{t} f(Y(s))ds, \quad t \geq 0 . \tag{3.18}$$

Since $E\tau_1 < \infty$, we have

$$t^{-1}N(t) \to \lambda \equiv 1/E\tau_1, \quad \text{as} \quad t \to \infty \quad \text{w.p.1} . \tag{3.19}$$

Under (3.19), FCLTs for partial sums tend to extend to random sums, as we see in Chapter 13 of the book. The major difficulty here is treating the two remainder terms in (3.17). Since $|R_1(t)| \leq Z_0$, the first remainder term in (3.17) is easily dispensed with in limit theorems. The second remainder term is more complicated; the key bound is

$$|R_2(t)| \leq Z_{N(t)+1}, \quad t \geq 0 .$$

Then we observe that $\{R_2(t) : t \geq 0\}$ is tight without space scaling. Thus, after space scaling, it is asymptotically neglible.

Theorem 2.3.8. (FCLT for regenerative processes) *With the regenerative structure above, there is convergence in distribution*

$$\mathbf{C}_n \Rightarrow \sigma\mathbf{B} \quad \text{in} \quad (D, J_1)$$

for $\mathbf{C}_n$ *in* (3.15) *and* $\mathbf{B}$ *standard BM if and only if there is a constant* m *such that*

$$EX_1(m) = 0, \quad EX_1(m)^2 < \infty$$

and

$$t^2P(Z_1(m) > t) \to 0 \quad \text{as} \quad t \to \infty . \tag{3.20}$$

for $X_1(m)$ *and* $Z_1(m)$ *in* (3.16). *Then the asymptotic variance is*

$$\sigma^2 = EX_1(m)^2 .$$

A sufficient condition for the regularity condition (3.20) is $EZ_1(m)^{2+\epsilon} < \infty$ for some $\epsilon > 0$. (A finite second moment is not enough. We remark that condition (3.20) does not appear in the ordinary CLT; see Glynn and Whitt (1993, 2000).) The role of the regularity condition (3.20) can be understood from the following lemma.

Lemma 2.3.1. (condition for the scaled maximum to be asymptotically negligible) *Let* $\{Z_i : i \geq 1\}$ *be a sequence of IID real-valued random variables and let* $\psi : R_+ \to R_+$ *be a function such that* $\psi(t) \to \infty$ *as* $t \to \infty$. *Then*

$$\psi(n)^{-1} \max_{1 \leq i \leq n} \{|Z_i|\} \Rightarrow 0 \quad \text{as} \quad n \to \infty$$

if and only if

$$tP(|Z_1| > \epsilon\psi(t)) \to 0 \quad as \quad t \to \infty \quad for\ all \quad \epsilon > 0\,. \tag{3.21}$$

Proof. Let $M_n \equiv \max\{|Z_i| : 1 \leq i \leq n\}$ and $F(t) \equiv P(|Z_1| \leq t)$, $t \geq 0$. Note that $\psi(n)^{-1}M_n \Rightarrow 0$ if and only if, for all $\epsilon > 0$, $P(\psi(n)^{-1}M_n > \epsilon) \to 0$ as $n \to \infty$. However,

$$P(M_n > \epsilon\psi(n)) < \delta$$

if and only if

$$P(M_n) \leq \epsilon\psi(n)) \geq 1 - \delta\ ,$$

where

$$\begin{aligned} P(M_n \leq \epsilon\psi(n)) &= F(\epsilon\psi(n))^n \\ &= (1 - n^{-1}n(1 - F(\epsilon\psi(n)))^n \\ &= (1 - n^{-1}nF^c(\epsilon\psi(n)))^n \\ &\to 1 \quad \text{as} \quad n \to \infty \end{aligned} \tag{3.22}$$

if and only if

$$nF^c(\epsilon\psi(n)) \to 0 \quad \text{as} \quad n \to \infty$$

or, equivalently, (3.21). ∎

Corollary 2.3.2. *If the conditions of Lemma* 2.3.1 *hold with* $\psi(t) = t^\alpha$ *for* $\alpha > 0$, *then condition* (3.21) *is equivalent to*

$$t^{1/\alpha}P(|Z_1| > t) \to 0 \quad as \quad t \to \infty\,.$$

Proof. Under the assumption, condition (3.21) becomes

$$tP(|Z_1| > \epsilon t^\alpha) \to 0 \quad \text{as} \quad t \to \infty \quad \text{for all} \quad \epsilon > 0\,,$$

which first is equivalent to

$$\epsilon^\alpha(\epsilon^{-\alpha}t)P(|Z_1| > (\epsilon^{-\alpha}t)^\alpha)$$

and then is equivalent to

$$\epsilon^\alpha tP(|Z_1|) > t^\alpha) \to 0 \quad \text{as} \quad t \to \infty \quad \text{for all} \quad \epsilon > 0\,,$$

which in turn is equivalent to the stated result. ∎

A general application of Theorem 2.3.8 is to obtain a FCLT for the counting processes associated with a batch Markovian arrival process (BMAP) as in Lucantoni (1993) or, equivalently, the virtual Markovian point process in Neuts (1989). An explicit formula for the variance of the number of arrivals in $[0, t]$ in a BMAP, from which the asymptotic variance easily can be obtained, is given on p. 284 of Neuts (1989).

2.3.4. Martingale FCLT

Martingale FCLTs are versatile tools for many applications. We have stated one martingale FCLT in Theorem 4.4.4 of the book, but there are others. We conclude this section by stating another. It is Theorem 18.1 of Billingsley (1999).

We start with the double sequence $\{X_{n,i} : n \geq 1, i \geq 1\}$ and an associated double sequence of σ-fields $\{\mathcal{F}_{n,k} : n \geq 1, k \geq 1\}$. We assume that $X_{n,k}$ is a *martingale difference* with respect to these σ-fields, i.e., $X_{n,k}$ is $\mathcal{F}_{n,k}$-measurable and

$$E[X_{n,k}|\mathcal{F}_{n,k-1}] = 0 \quad \text{for all} \quad n \quad \text{and} \quad k\,.$$

Suppose that $EX_{n,k}^2 < \infty$ and put

$$V_{n,k} \equiv E[X_{n,k}^2|\mathcal{F}_{n,k-1}]\,. \tag{3.23}$$

Note that $V_{n,k}$, being a conditional expectation, is a random variable. If the martingale is originally defined only for $1 \leq k \leq k_n$, let $X_{n,k} = 0$ and $\mathcal{F}_{n,k} = \mathcal{F}_{n,k_n}$ for $k > n$. Assume that $\sum_{k=1}^{\infty} X_{n,k}$ and $\sum_{k=1}^{\infty} V_{n,k}$ converge w.p.1 for each n.

Theorem 2.3.9. (martingale FCLT) *If, in addition to the assumptions above,*

$$\sum_{k=1}^{\lfloor nt \rfloor} V_{n,k} \Rightarrow \sigma^2 t \quad \textit{as} \quad n \to \infty \quad \textit{for every} \quad t > 0 \tag{3.24}$$

with $V_{n,k}$ in (3.23) *and the Lindeberg condition*

$$\sum_{k=1}^{\lfloor nt \rfloor} E[X_{n,k}^2 I_{\{|X_{n,k}| \geq \epsilon\}}] \to 0 \quad \textit{as} \quad n \to \infty$$

holds for every $t > 0$ and $\epsilon > 0$, then

$$\mathbf{S}_n \Rightarrow \sigma \mathbf{B} \quad \textit{in} \quad D\,,$$

where σ is determined by (3.24),

$$\mathbf{S}_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} X_{n,k}, \quad t \geq 0\,,$$

and $\mathbf{B}$ is standard Brownian motion.

Generalizations and other variations of Theorem 2.3.9 are contained on p. 339 of Ethier and Kurtz (1986) and Jacod and Shiryaev (1987).

2.4. Double Sequences and Lévy Limits

We have seen that there are only a few possible limits for normalized partial-sum processes with weak dependence when we work in the framework of a single sequence $\{X_n : n \geq 1\}$. In addition to the Brownian motion limits discussed in Sections 4.3 and 4.4 of the book, there are the stable Lévy motion limits discussed in Sections 4.5 and 4.7 of the book. However, there are many more possible limits for normalized partial-sum processes with weak dependence when we work in the framework of a double sequence $\{X_{n,k} : n \geq 1, k \geq 1\}$. We give a brief account in this section.

Throughout this section we assume that the sequence $\{X_{n,k} : k \geq 1\}$ is IID for each n, so that we are in a classic well-studied setting; e.g., see Gnedenko and Kolmogorov (1968) and Feller (1971). Since there is a different sequence for each n, we can incorporate multiplicative and additive normalization constants directly in the variables $X_{n,k}$. Hence we focus on the partial sums

$$S_{n,n} \equiv X_{n,1} + \cdots + X_{n,n} \tag{4.1}$$

without further normalization and the associated random functions in D defined by

$$\mathbf{S}_n(t) \equiv S_{n,\lfloor nt \rfloor}, \quad t \geq 0 . \tag{4.2}$$

The class of limits processes in FCLTs for $\mathbf{S}_n$ now are all Lévy processes. As indicated in Section 4.5 of the book, a *Lévy process* $\mathbf{L} \equiv \{\mathbf{L}(t) : t \geq 0\}$ is a stochastic process with sample paths in $D \equiv D([0,\infty), \mathbb{R})$, $\mathbf{L}(0) = 0$ and stationary and independent increments. Brownian motion and stable Lévy motion are important examples of Lévy processes, but there are many more; see Bertoin (1996) and Jacod and Shiryaev (1987).

The distribution of $\mathbf{L}(t)$ for any t is an infinitely divisible distribution. A probability distribution is *infinitely divisible* if for each n it is the n-fold convolution of another probability distribution; i.e., a random variable X has an infinitely divisible distribution if, for all n, there are IID random variables $X_1, \ldots, X_n$ (depending upon X and n) such that

$$X \stackrel{\mathrm{d}}{=} X_1 + \cdots + X_n .$$

Lévy processes and infinitely divisible distributions are characterized by their characteristic functions. In particular, the one-dimensional marginal distribution of every Lévy process has characteristic function

$$Ee^{i\theta L(t)} = e^{t\psi(\theta)} ,$$

where the *Lévy exponent* $\psi(\theta)$ can be expressed as

$$\psi(\theta) = ib\theta - \frac{\sigma^2\theta^2}{2} + \int_{-\infty}^{\infty} (\exp(i\theta x) - 1 - i\theta h(x))\mu(dx) \ , \qquad (4.3)$$

with b being the *centering coefficient*, $\sigma^2 \geq 0$ is the *Gaussian coefficient*, μ the *Lévy measure* and h a *truncation function*. There is quite a lot of freedom in the choice of the truncation function h. Following Jacod and Shiryaev (1987, pp. 75) we assume that the truncation function has compact support, is bounded and coincides with x in a neighborhood of the origin. To characterize convergence, we also want h to be continuous. A truncation function with all these properties is

$$h(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2-x, & 1 \leq x \leq 2 \\ -x, & -1 \leq x \leq 0 \\ 2+x, & -2 \leq x \leq 0 \\ 0, & |x| \geq 2\,. \end{cases} \qquad (4.4)$$

Other truncation functions are considered in the literature. Changing the truncation function h typically changes the centering coefficient b, but does not change the Gaussian coefficient σ^2 or the Lévy measure μ. The Lévy measure has support on $\mathbb{R} - \{0\}$; it is a bonafide measure with

$$\int_{-\infty}^{\infty} \min\{1, x^2\}\mu(dx) < \infty \ . \qquad (4.5)$$

Given a specific truncation function, such as h in (4.4), there is a one-to-one correspondence between Lévy processes, infinitely distributions and the triple of characteristics (b, σ^2, μ) appearing in (4.3), with $\sigma^2 \geq 0$ and μ being a measure on $\mathbb{R} - \{0\}$ satisfying (4.5).

Brownian motion is the special Lévy process with null Lévy measure, i.e., $\mu(A) = 0$ for all measurable subsets A. NonGaussian stable Lévy motions with index α are the special cases with $\sigma^2 = 0$ and

$$\mu(dx) = \begin{cases} c^+ x^{-(1+\alpha)}, & x > 0\ , \\ c^- |x|^{-(1+\alpha)}, & x < 0\ , \end{cases} \qquad (4.6)$$

for nonnegative constants c^+ and c^-, where $c^+ + c^- > 0$. From (4.6), we see that the power-tail structure of a stable law is manifested very strongly in the Lévy measure. While the stable law $S_\alpha(\sigma, \beta, \mu)$ has the power-tail asymptotics in equations 4.5.12 and 4.5.13 in the book, the corresponding Lévy measure has simple power densities on $(0, \infty)$ and $(-\infty, 0)$. A stable Lévy motion is totally skewed to the right, so that $\beta = 1$, (left, so that $\beta = -1$) if and only if $c^- = 0$ ($c^+ = 0$).

The Lévy measure μ characterizes the possible jumps of the Lévy process. Indeed, the jump process of the Lévy process is a Poisson random measure on $\mathbb{R} \times \mathbb{R}^+$ with intensity $\mu(dx)dt$; i.e., the number of jumps in the Lévy process falling in any spatial subinterval $[a, b]$ during time subinterval $[c, d]$ for $a < b$ and $0 < c < d$ has a Poisson distribution with mean $\mu([a, b])|d - c|$. As a simple consequence, if the Lévy measure μ has support in $\mathbb{R}^+$, then the Lévy process has no negative jumps. Thus we know that the totally skewed stable Lévy motion with $\beta = 1$ (and thus $c^- = 0$ in (4.6)) has sample paths without negative jumps.

A complication with Lévy processes is the large (in general, infinite) number of very small jumps. For any $c > 0$, a Lévy process has only finitely many jumps of at least size c in any finite interval w.p.1. However, for any $c > 0$, it can have infinitely many jumps of absolute size less than or equal to c in any finite interval. This large number of small jumps is compensated for by deterministic drift built into the final integral in (4.3), in particular, this drift occurs in the region that the truncation function h is positive. Thus the true process drift is the sum of the drift b and the drift associated with h. In general, the total drift may be infinite, which explains why the representation (4.3) does not separate out all the drift.

It is possible to decompose a Lévy process into the independent sum of component Lévy processes by decomposing the exponent $\psi(\theta)$ in (4.3) into separate pieces; see Theorem 1 of p. 13 of Bertoin (1996) and its proof. The first component Lévy process L_1 has Lévy exponent

$$\psi_1(\theta) \equiv ib\theta - \frac{\sigma^2\theta^2}{2}$$

and is Brownian motion with drift coefficient b and diffusion coefficient σ^2. The second component Lévy process L_2 has exponent

$$\psi_2(\theta) = \int_{|x| \geq 2} (\exp(i\theta x) - 1)\mu(dx)$$

and is a compound Poisson process, with jumps of absolute size at least 2, having Poisson intensity $\lambda_2 \equiv \mu((-\infty, -2]) + \mu((2, \infty)) < \infty$ and jump size probability distribution $\mu(dx)/\lambda_2$ on $(-\infty, -2) \cup (2, \infty)$. The complicated component is the third one. The third component Lévy process L_3 has exponent

$$\psi_3(\theta) = \int_{-2}^{2} (\exp(i\theta x) - 1 - i\theta h(x))\mu(dx) \ .$$

It can be shown to be a pure jump martingale with jumps of absolute size at most 2. It includes some deterministic drift to compensate for the jumps. In summary, we can write

$$\psi(\theta) = \psi_1(\theta) + \psi_2(\theta) + \psi_3(\theta)$$

and

$$L \stackrel{\mathrm{d}}{=} L_1 + L_2 + L_3 \ ,$$

where L_1, L_2 and L_3 are the independent Lévy processes with exponents ψ_1, ψ_2 and ψ_3 defined above.

If an infinitely divisible distribution has finite moments, these moments can be derived by differentiating the characteristic function. For example, if $E|L(1)| < \infty$, then

$$EL(1) = \frac{\psi'(\theta)}{i} = b + \int_{-\infty}^{\infty} [x - h(x)]\mu(dx) \ , \tag{4.7}$$

where, because of the definition of the truncation function h, the integrand is nonzero only in $(-\infty, -1] \cup [1, \infty)$.

An important point is that the class of infinitely divisible distributions is remarkably large. An indication is the fact that infinitely divisible distributions are characterized by the triples (b, σ^2, μ), where μ is a measure on $\mathbb{R} - \{0\}$ satisfying (4.5). Two Lévy processes with triples (b_1, σ_1^2, μ_1) and (b_2, σ_2^2, μ_2) reduce to the same process if and only if $b_1 = b_2$, $\sigma_1^2 = \sigma_2^2$ and $\mu_1(A) = \mu_2(A)$ for all measurable sets $A \subseteq \mathbb{R}$. Nevertheless, infinitely divisible distributions may seem very special. However, over the years, many common distributions have been shown to be infinitely divisible. For example, lognormal distributions, Weibull distributions with ccdf's $e^{-(t/a)^c}$ for $c \leq 1$, Pareto distributions, and all mixtures of exponential distributions are infinitely divisible; see Thorin (1977a,b), p. 452 of Feller (1971), Bondesson (1992) and Abate and Whitt (1996). (The Weibull and Pareto distributions actually are mixtures of exponential distributions so infinite divisibility follows from that structure.) Moreover, the class of infinitely divisible distributions is easily seen to be closed under convolutions.

We now consider convergence in distribution of partial sums to infinitely divisible distributions and Lévy processes. First note that each infinitely divisible distribution can serve as a limit, because if X is infinitely divisible then there is a sequence of sequences $\{X_{n,k} : k \geq 1\}$ of IID random variables such that $X \stackrel{\mathrm{d}}{=} S_n$ for all n by the definition of infinite divisibility.

The following characterization of all possible limits is a consequence of Theorem 2, p. 303, of Feller (1971) and Theorem 2.7 of Skorohod (1957).

Theorem 2.4.1. (Lévy process FCLT for double sequences) *Let* $\{X_{n,k} : k \geq 1\}$ *be a sequence of IID random variables for each* n *and let* $S_{n,n}$ *and* $\mathbf{S}_n$ *be defined as in* (4.1) *and* (4.2). *If*

$$S_{n,n} \Rightarrow Z \quad in \quad \mathbb{R} \ ,$$

then Z *has an infinitely divisible distribution and*

$$\mathbf{S}_n \Rightarrow \mathbf{L} \quad in \quad D([0,\infty), J_1) \ ,$$

where $\mathbf{L}$ *is the Lévy process with* $\mathbf{L}(1) \stackrel{\mathrm{d}}{=} Z$.

Necessary and sufficient conditions for the FCLT with convergence to a specific Lévy process are consequences of Theorems 2.35, 2.52 and 3.4 of pp. 362, 368 and 373 of Jacod and Shiryaev (1987). (The partial sum process is both a semimartingale and a process with independent increments (PII) but not a process with stationary independent increments (PIIS).)

Theorem 2.4.2. (criteria for the Lévy-process FCLT) *Let* $\{X_{n,k} : k \geq 1\}$ *be a sequence of IID random variables for each* n, *with* $\{X_{n,1} : n \geq 1\}$ *being infinitesimal, i.e.,*

$$\lim_{n\to\infty} P(|X_{n,1}| > \epsilon) = 0 \quad for\ all \quad \epsilon > 0 \ . \tag{4.8}$$

Then

$$\mathbf{S}_n \Rightarrow \mathbf{L} \quad in \quad D([0,\infty), \mathbb{R}, J_1) \tag{4.9}$$

for $\mathbf{S}_n$ *in (4.2), where* $\mathbf{L}$ *is a Lévy process with characteristics* (b, σ^2, μ), *if and only if*

$$\text{(i)} \ \lim_{n\to\infty} nE[h(X_{n,1}] = b \ , \tag{4.10}$$

$$\text{(ii)} \ \lim_{n\to\infty} nVar[h(X_{n,1})] = \sigma^2 \ , \tag{4.11}$$

$$\text{(iii)} \ \lim_{n\to\infty} nE[g(X_{n,1})] = \int_{-\infty}^{\infty} g(x)\mu(dx) \ , \tag{4.12}$$

for the truncation function h *and all continuous bounded real-valued functions* g *on* $\mathbb{R}$ *with* $g(x) = 0$ *in a neighborhood of 0 and* $g(x) \to y$, $-\infty < y < \infty$, *as* $x \to \pm\infty$.

Note that $h(x) = x$ for $|x| \leq 1$, so that conditions (i) and (ii) above correspond closely to convergence of the scaled means and variances.

Theorem 2.4.2 provides a large class of initial FCLT's to use with the continuous-mapping approach. We have only stated the classical results. Jacod and Shiryaev (1987) go much further, generalizing the characteristics of a Lévy process to define characteristics for semimartingales, allowing for nonstationarity. They also establish conditions for FCLTs in which processes with independent increments converge to other processes with independent increments (Chapter VII), semimartingales converge to processes with independent increments (Chapter VIII) and semimartingales converge to other semimartingales (Chapter IX), all expressed via the process characteristics. Actually verifying these conditions may not be straightforward, however.

2.5. Linear Models

In this section we discuss the linear-process representation in equation 4.6.6 of the book that was critical for obtaining the FCLT with strong dependence. The linear-process representation expresses the basic summands X_n as

$$X_n \equiv \sum_{j=0}^{\infty} a_j Y_{n-j}, \quad n \geq 1\,, \tag{5.1}$$

where $\{Y_n : -\infty < n < \infty\}$ is a two-sided sequence of IID random variables with $EY_n = 0$ and $EY_n^2 = 1$, and $\{a_j : j \geq 0\}$ is a sequence of (deterministic, finite) constants with

$$\sum_{j=0}^{\infty} a_j^2 < \infty\,. \tag{5.2}$$

We now show that the linear-process representation can arise naturally from modeling. First, however, it is important to repeat our earlier disclaimer. It is important to realize that the stochastic-process limits with strong dependence characterized by (5.1) are less universal. Many other forms of strong dependence are possible. And, if the dependence does not approximately correspond to a linear process, then there may appear a very different limit process or there may even be no stochastic-process limit at all.

Nevertheless, the linear-process representation is very natural. It provides a useful concrete model of strong dependence with an associated FCLT. To explain how linear models can arise, we describe some time-series models.

In particular, we show how the Gaussian linear process arises from a fundamental time-series model. We especially want to show how the Gaussian linear process with strong dependence arises from the *fractional autoregressive integrated moving average* (FARIMA) model; e.g., see Section 2.5 of Beran (1994) and Sections 7.12 and 7.13 of Samorodnitsky and Taqqu (1994).

The starting point is the *autoregressive moving average* (ARMA (p,q)) process, where p and q are nonnegative integers. To define the ARMA (p,q) process, let B be the *backshift operator*, defined by $BX_n \equiv X_{n-1}$, so that *differences* can be expressed as $X_n - X_{n-1} \equiv (1-B)X_n$ and $(X_n - X_{n-1}) - (X_{n-1} - X_{n-2}) \equiv (1-B)^2 X_n$. Let ϕ and ψ be polynomials of degree p and q, respectively, of the form

$$\phi(z) \equiv 1 - \sum_{j=1}^{p} \phi_j z^j$$

and

$$\psi(z) \equiv 1 + \sum_{j=1}^{q} \psi_j z^j ,$$

where z is a complex variable and $\phi_1, \ldots, \phi_p$, $\psi_1, \ldots, \psi_q$ are real coefficients. As regularity conditions, assume that the equations $\phi(z) = 0$ and $\psi(z) = 0$ have no common roots and that all solutions of the equation $\phi(z) = 0$ fall outside the unit disk $\{z : |z| \le 1\}$. An ARMA (p,q) process is defined to be the stationary solution to the equation

$$\phi(B)X_n = \psi(B)Y_n \tag{5.3}$$

where $\{Y_n : n \ge 1\}$ is a sequence of IID $N(0,1)$ random variables; e.g., see Chapter 3 of Box, Jenkins and Reinsel (1994). In this setting, the sequence $\{Y_n\}$ is called the *innovation process*. Note that the exponential smoothing in Example 1.4.2 in the book is an ARMA$(1,0)$ process.

Theorem 2.5.1. (the ARMA process) *Under the regularity conditions above, the system of ARMA (p,q) equations* (5.3) *has a unique solution of the form*

$$X_n = \sum_{j=0}^{\infty} w_j Y_{n-j}, \quad n \ge 1 , \tag{5.4}$$

with real constant coefficients w_j satisfying $|w_j| < \delta^j$ for all sufficiently large j, for some δ, $0 < \delta < 1$. The coefficients w_j in (5.4) *are the coefficients of the power series $\psi(z)/\phi(z)$.*

Note that the coefficients w_j in the linear-process representation are available via their generating function $\psi(z)/\phi(z)$. Given the polynomials ψ and ϕ, we can thus calculate the coefficients w_j by numerically inverting the generating function; see Abate and Whitt (1992b).

Also note that the coefficients w_j in (5.4) decay exponentially fast, so that an ARMA process only exhibits weak dependence. To obtain strong dependence, we need the coefficients w_j in (5.4) to decay more slowly. We achieve that by considering fractional differencing. We do so by introducing a generalization of the ARIMA model. If instead $\{X_n\}$ is the solution of the equation

$$\phi(B)(1-B)^d X_n = \psi(B)Y_N \ , \tag{5.5}$$

where d is a nonnegative integer and $\{Y_n\}$ is again a sequence of IID $N(0,1)$ random variables, then $\{X_n\}$ is said to be an ARIMA (p,d,q) process, which was introduced by Box and Jenkins (1970); see Chapter 4 of Box, Jenkins and Reinsel (1994).

The FARIMA process is a generalization of the ARIMA process to fractional differencing. The FARIMA generalization of ARIMA was introduced by Granger and Joyeux (1980) and Hosking (1981). The FARIMA model with strong dependence depends on a parameter triple (p,q,d), where p and q are nonnegative integers and $0 < d < 1/2$. (There also are FARIMA models with $-1/2 < d \le 0$, but we will not consider them.) Given (p,q), there are $p+q$ further parameters.

For any real number d, we define the *fractional difference operator*

$$(1-B)^d \equiv \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \ ,$$

where

$$\binom{d}{k} \equiv \frac{d!}{k!(d-k)!} \equiv \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

with $\Gamma(x)$ the gamma function. A stationary process $\{X_n\}$ that satisfies (5.5) for positive integers p and q and for $0 < d < 1/2$ is a FARIMA (p,d,q) process. (Values of d with $-1/2 < d \le 0$ are also possible, but we are primarily interested in the range $0 < d < 1/2$.)

Theorem 2.5.2. (the FARIMA process) *Under the regularity conditions above, including $0 < d < 1/2$, the system of FARIMA (p,d,q) equations (5.5) has a unique solution of the form*

$$X_n = \sum_{j=0}^{\infty} a_j Y_{n-j}, \quad n \ge 1 \, ,$$

which converges almost surely, where

$$a_j \equiv \sum_{i=0}^{j} w_i b_{j-i}(-d)$$

with $\{w_j\}$ *being the sequence of constant coefficients in* (5.4) *and*

$$b_j(-d) \equiv \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sim \frac{1}{\Gamma(d)} j^{d-1} \quad as \quad j \to \infty\,.$$

As a consequence,

$$a_j \sim a j^{d-1} \quad as \quad j \to \infty\,,$$

where

$$a \equiv \sum_{i=0}^{j} w_i / \Gamma(d)$$

for w_i *in* (5.4), *and*

$$r_j \equiv Cov(X_1, X_{1+j}) \sim r j^{2d-1} \quad as \quad j \to \infty\,,$$

where

$$r \equiv \left(\frac{\psi(1)}{\Gamma(d)\phi(1)}\right)^2 \int_0^\infty g(x)dx$$

for

$$g(x) = x^{2(d-1)} + (1+x)^{2(d-1)} - (x^{d-1} - (1+x)^{d-1})^2\,.$$

The point of this discussion has been to show that a linear process of the form (5.1) and (5.2), with

$$Var(S_n) = n^{2H} L(n) \quad \text{as} \quad n \to \infty\,, \tag{5.6}$$

where $L(t)$ is a slowly varying function and $H > 1/2$, arises naturally from the FARIMA (p, d, q) model with $0 < d < 1/2$. In the FARIMA case the linear process is also a Gaussian process, but the key relations in Theorems 2.5.1 and 2.5.2 here hold for stationary sequences with finite second moments. We also remark that the parameters H and d are related by

$$d = H - \frac{1}{2}\,.$$

It is also significant that the FARIMA model provides a natural framework to exploit the strong dependence in order to make predictions; see

Beran (1994) for a full account of statistics for strongly dependent, light-tailed processes. We only make a few remarks.

In applications, we may have a stochastic sequence $\{X_n\}$ that we are willing to regard as a zero-mean stationary sequence with $\mathrm{Var}(X_n) < \infty$. We can examine the variance $\mathrm{Var}(S_n)$. If we find that

$$Var(S_n) \sim cn^{2H} \quad \text{as} \quad n \to \infty$$

for $1/2 < H < 1$, then we have the Joseph effect. That can be checked by looking for a linear relationship after taking logarithms; i. e.,

$$\log(Var(S_n)) \sim \log(c) + 2H \log(n) \ .$$

We then can invoke Theorem 4.6.1 in the book, without directly verifying the linear-process representation in (5.1) and without identifying the weights a_j in (5.1), to support the approximation (in distribution)

$$\{(cn^{2H})^{-1/2} S_{\lfloor nt \rfloor} : t \geq 0\} \approx \{Z_H(t) : t \geq 0\} \ , \tag{5.7}$$

where Z_H is standard FBM. Note that we obtain a parsimonious approximation, depending only on the two parameters c and H. Attention naturally focuses on ways to estimate the parameters c and H. That can be done simply from a plot of $\log Var(S_n)$ as a function of $\log n$; see Beran (1994).

It is important to remember that the justification of approximation (5.7) from Theorem 4.6.1 in the book actually depends on the linear-process representation. However, we can directly justify the approximation equation 4.6.13 in the book. by checking that the finite-dimensional distributions are approximately Gaussian and that the covariance function is approximately the covariance function of FBM in equation 4.6.13 in the book. The limit theorem explains why the FBM approximation may be appropriate.

We conclude by remarking that there is again a time-series motivation for considering the linear-process representation in the case of heavy tails plus dependence, discussed in Section 4.7 of the book. Specifically, there is a time-series motivation for the linear-process representation in equation 4.7.1 of the book, where the innovation variables Y_n have heavy tails, just as there was for the light-tailed case in Section 4.6 of the book, because there are analogs of the ARMA, ARIMA and FARIMA processes with stable innovations; i.e., there are analogs of Theorems 2.5.1 and 2.5.2 here for the case in which the innovation process $\{Y_n\}$ is a sequence of IID random variables with a stable law $S_\alpha(\sigma, \beta, \mu)$ for $0 < \alpha < 2$; see Sections 7.12 and 7.13 of Samorodnitsky and Taqqu (1994).

Chapter 3

Preservation of Pointwise Convergence

3.1. Introduction

With the continuous-mapping approach to stochastic-process limits, we are concerned about limits $x_n \to x$ and $f_n(x_n) \to f(x)$ for a sequences of functions $\{x_n : n \geq 1\}$ in D and $f_n,\ f : D \to D$; see Section 3.5 and Chapter 13 in the book. However, in many applications we actually are interested in the pointwise limits

$$x(t)/\phi(t) \to \gamma \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad t \to \infty \tag{1.1}$$

and

$$f(x)(t)/\phi(t) \to \eta \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad t \to \infty \tag{1.2}$$

for single functions $x \in D$ and $f : D \to D$, where ϕ is a suitable scaling function. In particular, we may want to show that the pointwise limit (1.1) implies the associated pointwise limit (1.2) and identify the limit η.

It is significant that we often can obtain such limits in $\mathbb{R}$ as consequences of function-space limits by setting

$$y_s(t) \equiv x(st)/\phi(s), \quad s > 0\ . \tag{1.3}$$

As a regularity condition, we will assume that the scaling function ϕ is a homeomorphism of $\mathbb{R}_+$, i.e., $\phi \in \Lambda(\mathbb{R}_+)$. That implies that $\phi(0) = 0$, ϕ is increasing and $\phi(t) \to \infty$ as $t \to \infty$. If we can show that

$$y_s \to y \quad \text{in} \quad D \quad \text{as} \quad s \to \infty \tag{1.4}$$

for y_s in (1.3), where $1 \notin Disc(y)$, then we can apply the projection map π_1 taking x into $x(1)$ to obtain

$$y_s(1) = x(s)/\phi(s) \to y(1) \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad s \to \infty\,, \tag{1.5}$$

which implies the desired convergence in (1.1) and identifies the limit γ in (1.1) as $y(1)$ in (1.5). Moreover, if we can show that

$$f(x)(t)/\phi(t) = g_t(y_t) \quad \text{for each} \quad t > 0\,, \tag{1.6}$$

where $g_s, g : D \to \mathbb{R}$ and

$$g_s(y_s) \to g(y) \quad \text{in} \quad \mathbb{R} \tag{1.7}$$

whenever $y_s \to y$ in D, then we can obtain (1.2) from (1.4) as well, and we identify the limit η in (1.2) as $g(y)$.

For example, this reasoning applies to the supremum function: $f(x)(t) = x^{\uparrow}(t)$ for $t > 0$. Then $g(y) = g_s(y) = f(y)(1) = y^{\uparrow}(1)$ for all $y \in D$ and $s > 0$. As a consequence, the limit in (1.2) holds with $\eta = g(y) = y^{\uparrow}(1)$.

Even though many pointwise limits for single functions can be subsumed as special cases of function-space limits, it is interesting to consider what can be obtained directly without resorting to the function-space construction in (1.3). In particular, it is natural to ask how pointwise limits for single functions are preserved under the composition, supremum, and inverse maps. We investigate that question in this chapter.

For queues and related applied probability models, this convergence-preservation issue for single sample paths corresponds to sample-path analysis, which is commonly associated with the fundamental relations (conservation laws) $L = \lambda W$ and Arrivals See Time Averages (ASTA); see El-Taha and Stidham (1999); see the chapter notes at the end of the chapter.

3.2. From Pointwise to Uniform Convergence

Clearly, the pointwise limit in (1.1) is more elementary than the function-space limit (1.4) but, surprisingly, (1.4) is not much stronger than (1.1). Indeed, under minor regularity conditions, (1.1) actually implies (1.4). Recall that ϕ in $\Lambda(\mathbb{R}_+)$ is regularly varying with index $p > 0$, denoted by $\phi \in \mathcal{R}(p)$, if

$$\phi(tx)/\phi(t) \to x^p \quad \text{as} \quad t \to \infty \tag{2.1}$$

for all $x > 0$; see Appendix A at the end of the book.

Theorem 3.2.1. (from pointwise to uniform convergence) *Let $x \in D$ and $\phi \in \Lambda(\mathbb{R}_+)$ with $\phi \in \mathcal{R}(p)$ for $p > 0$. If the limit* (1.1) *holds in $\mathbb{R}$, then*

$$\|y_s - y\|_T \to 0 \quad as \quad s \to \infty \quad for\ each \quad T > 0 \tag{2.2}$$

for y_s in (1.3) *and*

$$y(t) = \gamma t^p, \quad t \geq 0 \ .$$

Proof. Under the conditions, for any $\epsilon > 0$, there is a t_0 such that

$$|x(t)/\phi(t) - \gamma| < \epsilon \quad \text{for all} \quad t \geq t_0 \ . \tag{2.3}$$

and an s_0 such that

$$\sup_{0 \leq t \leq T} \left| \frac{\phi(st)}{\phi(s)} - t^p \right| < \epsilon \quad \text{for all} \quad s \geq s_0 \ ; \tag{2.4}$$

see Theorem A.5 in Appendix A in the book. For $t \leq t_0/s$,

$$|y_s(t) - y(t)| \leq |y_s(t)| + |y(t)| \leq \frac{\|x\|_{t_0}}{\phi(s)} + \gamma \left(\frac{t_0}{s}\right)^p \ , \tag{2.5}$$

which is less than ϵ for all sufficiently large s, say $s \geq s_1 \geq s_0$. Since

$$y_s(t) - y(t) = \frac{x(st)}{\phi(st)} \left(\frac{\phi(st)}{\phi(s)} - t^p \right) + t^p \left(\frac{x(st)}{\phi(st)} - \gamma \right) \ , \tag{2.6}$$

for $s \geq s_1$,

$$\begin{aligned} \|y_s - y\|_T &\leq \epsilon + \sup_{t \geq t/s} \left\{ \left| \frac{x(st)}{\phi(st)} \right| \left| \frac{\phi(st)}{\phi(s)} - t^p \right| + t^p \left| \frac{x(st)}{\phi(st)} - \gamma \right| \right\} \\ &\leq \epsilon + (\gamma + \epsilon)\epsilon + T^p \epsilon \ , \end{aligned} \tag{2.7}$$

which can be made arbitrarily small with an appropriate choice of ϵ. ∎

For the special case in which $\phi(t) = t$, condition (1.1) corresponds to a strong law of large numbers (SLLN) for a stochastic process, while the conclusion (2.2) corresponds to a functional strong law of large numbers (FSLLN). The following corollary is Theorem 4 from Glynn and Whitt (1988).

Corollary 3.2.1. (from a SLLN to a FSLLN) *Let $\{X(t) : t \geq 0\}$ be a real-valued stochastic process and let*

$$\hat{\mathbf{X}}_n(t) \equiv n^{-1} X(nt), \quad t \geq 0, \quad n \geq 1 \ . \tag{2.8}$$

If a SLLN holds, i.e., if

$$t^{-1}X(t) \to \gamma \ \textit{w.p.1 in } \mathbb{R} \textit{ as } t \to \infty\,, \tag{2.9}$$

then a FSLLN holds, i.e.,

$$\|\hat{\mathbf{X}}_n - \gamma \mathbf{e}\|_T \to 0 \quad \textit{w.p.1 in} \quad D([0,T],\mathbb{R}) \quad \textit{as} \quad n \to \infty \tag{2.10}$$

for all $T > 0$.

3.3. Supremum

In this section we consider the supremum map. The following elementary convergence-preservation result is referred to as the "fundamental lemma of maxima" in Section 2.5 of El-Taha and Stidham (1999).

Proposition 3.3.1. (preservation of pointwise convergence for the supremum) *Suppose that* $x \in D([0,\infty),\mathbb{R})$, ϕ *is an increasing real-valued function on* $\mathbb{R}_+$ *with* $\phi(t) \to \infty$ *as* $t \to \infty$. *If* $x(t)/\phi(t) \to \gamma \geq 0$ *as* $t \to \infty$, *then* $x^{\uparrow}(t)/\phi(t) \to \gamma$ *as* $t \to \infty$.

Proof. Under the condition, for any $\epsilon > 0$, there exists t_0 such that

$$(\gamma - \epsilon)\phi(t) \leq x(t) \leq (\gamma + \epsilon)\phi(t)$$

for all $t \geq t_0$. Hence,

$$(\gamma - \epsilon)\phi(t) \leq x^{\uparrow}(t) \leq x^{\uparrow}(t_0) \vee (\gamma + \epsilon)\phi(t)$$

for all $t \geq t_0$. Since $\gamma \geq 0$ and $\phi(t) \to \infty$, there is $t_1 \geq t_0$ such that $x^{\uparrow}(t_0) \leq (\gamma + \epsilon)\phi(t)$ for all $t \geq t_1$. Thus, for $t \geq t_1$,

$$|\phi(t)^{-1}x^{\uparrow}(t) - \gamma| \leq \epsilon\,. \quad \blacksquare$$

Under the conditions of Theorem 3.2.1, if $\gamma \geq 0$, then we can apply the continuous mapping theorem to deduce that $x^{\uparrow}(t)/\phi(t) \to \gamma$ as $t \to \infty$; i.e., the conclusion of Proposition (3.3.1) holds by virtue of Theorems 3.2.1 here and 13.4.1 in the book. However, Theorem 3.2.1 here has the extra assumption that ϕ is regularly varying.

Paralleling Proposition (3.3.1), we can also establish a pointwise-convergence result for supremum with centering for a single function.

Proposition 3.3.2. (preservation of pointwise convergence with centering for the supremum) *Suppose that ϕ is an increasing real-valued function such that $\phi(t) \to \infty$ and $\phi(t)/t \to 0$ as $t \to \infty$. If*

$$\phi(t)^{-1}[x(t) - \lambda t] \to \gamma \quad as \quad t \to \infty \tag{3.1}$$

for $\lambda > 0$, then

$$\phi(t)^{-1}[x^{\uparrow}(t) - \lambda t] \to \gamma \quad as \quad t \to \infty\,. \tag{3.2}$$

Proof. Under condition (3.1), for any $\epsilon > 0$, there exists t_0 such that

$$\lambda t - \phi(t)(\gamma - \epsilon) \le x(t) \le \lambda t + \phi(t)(\gamma + \epsilon)$$

for all $t \ge t_0$. Then

$$\lambda t - \phi(t)(\gamma - \epsilon) \le x^{\uparrow}(t) \le x^{\uparrow}(t_0) \vee (\lambda t + \phi(t)(\gamma + \epsilon))\,.$$

However, since $\lambda > 0$ and $\phi(t)/t \to 0$, there is a $t_1 > t_0$ such that $x^{\uparrow}(t_0) \le \lambda t + \phi(t)(\gamma + \epsilon)$ for all $t \ge t_1$. Hence

$$|\phi(t)^{-1}[x^{\uparrow}(t) - \lambda t] - \gamma| < \epsilon$$

for all $t \ge t_1$, so that (3.2) holds. ■

3.4. Counting Functions

We now turn to counting functions, as in Section 13.8 of the book. A counting function is defined in terms of a sequence $\{s_n : n \ge 0\}$ of nondecreasing nonnegative real numbers with $s_0 = 0$. We can think of s_n as the partial sum

$$s_n \equiv x_1 + \cdots + x_n, \quad n \ge 1\,, \tag{4.1}$$

by simply writing $x_i \equiv s_i - s_{i-1}$, $i \ge 1$. The associated *counting function* $\{c(t) : t \ge 0\}$ is defined by

$$c(t) \equiv \max\{k \ge 0 : s_k \le t\}, \quad t \ge 0\,. \tag{4.2}$$

To have $c(t)$ finite for all $t > 0$, we assume that $s_n \to \infty$ as $n \to \infty$.

To establish limits for counting functions, we use two scaling functions. We again let the scaling functions be elements of $\Lambda(\mathbb{R}_+)$. Note that if $\phi \in \Lambda(\mathbb{R}_+)$, then $\phi(0) = 0$ and ϕ is strictly increasing. Also, ϕ necessarily has an inverse ϕ^{-1} with $\phi \circ \phi^{-1} = \phi^{-1} \circ \phi = e$. Moreover, $(\phi_1 \circ \phi_2)^{-1} = \phi_2^{-1} \circ \phi_1^{-1}$ for two homeomorphisms ϕ_1 and ϕ_2.

The basis for positive results is the basic inverse relation in Lemma 13.8.1 of the book, which we restate here:

Lemma 3.4.1. (basic inverse relation) *For any nonnegative integer n and nonnegative real number t,*

$$s_n \le t \quad \text{if and only if} \quad c(t) \ge n\,. \tag{4.3}$$

The relation between the limits for s_n as $n \to \infty$ and $c(t)$ as $t \to \infty$ follows easily from the following bounds, which are of independent interest. Let $\lfloor x \rfloor$ be the greatest integer less than or equal to x and let $\lceil x \rceil$ be the least integer greater than or equal to x. One-sided bounds are obtained below by either setting $\epsilon = 1$ or setting $\delta = \infty$. Let $1/0 = \infty$ and $1/\infty = 0$.

Lemma 3.4.2. (one-sided bounds) *Suppose that $\phi_1, \phi_2 \in \Lambda(\mathbb{R}_+)$, $0 < \epsilon \le 1$ and $0 < \delta \le \infty$.*

(a) If

$$1 - \epsilon \le \frac{\phi_2(c(t))}{\phi_1(t)} < 1 + \delta \quad \text{for all} \quad t \ge t_0, \tag{4.4}$$

then

$$\frac{1}{1+\delta} < \frac{\phi_1(s_n)}{\phi_2(n)} \le \frac{1}{1-\epsilon} \quad \text{for all} \quad n \ge n_0 \equiv \lceil \phi_2^{-1}(\phi_1(t_0)(\lambda + \delta)) \rceil. \tag{4.5}$$

(b) If

$$1 - \epsilon < \frac{\phi_1(s_n)}{\phi_2(n)} \le 1 + \delta \quad \text{for all} \quad n \ge n_0\,, \tag{4.6}$$

then

$$\frac{\phi_2(c(t))}{\phi_1(t)} \le \frac{1}{1-\epsilon} \tag{4.7}$$

and

$$\frac{\phi_2(c(t)+1)}{\phi_1(t)} \ge \frac{1}{1+\delta}. \tag{4.8}$$

for all $t \ge t_0 \equiv \lceil \phi_1^{-1}(\phi_2(t_0)(1+\delta)) \rceil$. Moreover, there is a sequence of times $\{t_k\}$ such that $t_k \to \infty$ as $k \to \infty$ and

$$\frac{\phi_2(c(t_k))}{\phi_1(t_k)} \ge \frac{1}{1+\delta} \tag{4.9}$$

for all $t_k \ge t_0$.

Proof. (a) If (4.4) holds, then

$$n_1(t) \equiv \lfloor \phi_2^{-1}(\phi_1(t)(1-\epsilon)) \rfloor \le c(t) < \lceil \phi_2^{-1}(\phi_1(t)(1+\delta)) \rceil \equiv n_2(t)$$

for all $t \ge t_0$ and, by Lemma 3.4.1,

$$s_{n_1(t)} \le t < s_{n_2(t)} \quad \text{for all} \quad t \ge t_0 . \tag{4.10}$$

Let t_1 and t_2 be functions of n defined by

$$t_1(n) \equiv \phi_1^{-1}(\phi_2(n)/(1-\epsilon)) \quad \text{and} \quad t_2(n) \equiv \phi_1^{-1}(\phi_2(n)/(1+\delta)) ,$$

and note that $n_1(t_1(n)) = n_2(t_2(n)) = n$ for all n. Hence, for all $n \ge n_0 \equiv \lceil \phi_2^{-1}(\phi_1(t_0)(1+\delta)) \rceil$, we have $t_1(n_0) \ge t_2(n_0) \ge t_0$ and, by (4.10),

$$t_2(n) < s_{n_2(t_2(n))} = s_n = s_{n_1(t_1(n))} \le t_1(n)$$

or, equivalently,

$$\phi_2(n)\left(\frac{1}{1+\delta} - 1\right) < \phi_1(s_n) - \phi_2(n) \le \phi_2(n)\left(\frac{1}{1-\epsilon} - 1\right)$$

which implies (4.5).

(b) If (4.6) holds, then

$$\tilde{t}_1(n) \equiv \phi_1^{-1}(\phi_2(n)(1-\epsilon)) < s_n \le \phi_1^{-1}(\phi_2(n)(1+\delta)) \equiv \tilde{t}_2(n)$$

for all $n \ge n_0$ and, by Lemma 3.4.1,

$$c(\tilde{t}_1(n)) < n \le c(\tilde{t}_2(n)) \quad \text{for all} \quad n \ge n_0 . \tag{4.11}$$

Let $\tilde{n}_1$ and $\tilde{n}_2$ be functions of t defined by

$$\tilde{n}_1(t) \equiv \lceil \phi_2^{-1}(\phi_1(t)/(1-\epsilon)) \rceil \quad \text{and} \quad \tilde{n}_2(t) \equiv \lfloor \phi_2^{-1}(\phi_1(t)/(1+\delta)) \rfloor$$

and note that

$$\tilde{t}_2(\tilde{n}_2(t)) \le t \le \tilde{t}_1(\tilde{n}_1(t)) .$$

Hence, by (4.11),

$$\tilde{n}_2(t) \le c(\tilde{t}_2(\tilde{n}_2(t))) \le c(t) \le c(\tilde{t}_1(\tilde{n}_1(t))) < \tilde{n}_1(t)$$

and

$$\phi_2^{-1}(\phi_1(t)/(1+\delta)) - 1 \le c(t) \le \phi_2^{-1}(\phi_1(t)/(1-\epsilon))$$

for all $t \ge t_0 \equiv \phi_1^{-1}\ (\phi_2(n_0)(1+\delta))$, because $\tilde{n}_1(t_0) \ge \tilde{n}_2(t_0) = n_0$, which implies (4.7) and (4.8) by the reasoning for part (a). For (4.9), choose the sequence $\{t_k\}$ so that $\phi_2^{-1}(\phi_1(t_k)/(1+\delta))$ is an integer. Then we have the lower bound $c(t_k) \ge \phi_2^{-1}(\phi_1(t_k)/(1+\delta))$ for all k, which implies (4.9). ∎

We now apply Lemma 3.4.2 to characterize the asymptotic behavior.

Theorem 3.4.1. (implications for pointwise convergence) *Suppose that* ϕ_1, $\phi_2 \in \Lambda(\mathbb{R}_+)$ *and* $0 \le \lambda \le \infty$.

(a) If $\phi_2(c(t))/\phi_1(t) \to \lambda$ *as* $t \to \infty$, *then* $\phi_1(s_n)/\phi_2(n) \to \lambda^{-1}$ *as* $n \to \infty$.

(b) If $\phi_1(s_n)/\phi_2(n) \to \lambda^{-1}$ *as* $n \to \infty$, *then*

$$\overline{\lim_{t\to\infty}}\ \phi_2(c(t))/\phi_1(t) = \lambda. \tag{4.12}$$

(c) If, in addition to the condition for (b), either

$$\frac{\phi_2(c(t)+1) - \phi_2(c(t))}{\phi_1(t)} \to 0 \quad \text{as} \quad t \to \infty \tag{4.13}$$

or

$$\frac{\phi_2(n+1)}{\phi_2(n)} \to 1 \quad \text{as} \quad n \to \infty, \tag{4.14}$$

then $\phi_2(c(t))/\phi_1(t) \to \lambda$ *as* $t \to \infty$.

(d) If $\phi_1(s_n)/\phi_2(n) \to 0$ *as* $n \to \infty$ *and either*

$$\overline{\lim_{t\to\infty}}\ \frac{\phi_2(c(t)+1) - \phi_2(c(t))}{\phi_1(t)} < \infty \tag{4.15}$$

or

$$\underline{\lim}_{n\to\infty}\ \frac{\phi_2(n)}{\phi_2(n+1)} > 0, \tag{4.16}$$

then $\phi_2(c(t))/\phi_1(t) \to \infty$ *as* $t \to \infty$.

Proof. (a) First suppose that $0 < \lambda < \infty$. Then incorporate λ into $\phi_1(t)$ by dividing by λ. The condition implies that for all appropriate ϵ and δ there exists t_0 such that (4.4) holds. By Lemma 3.4.2(a), (4.5) holds. Since ϵ and δ are arbitrary in (4.5), it implies the desired conclusion. To treat the cases $\lambda = 0$ and $\lambda = \infty$, use the one-sided bounds in Lemma 3.4.2. For example, if $\phi_2(c(t))/\phi_1(t) \to 0$ as $t \to \infty$, then for all positive ϵ and δ there exists t_0 such that $\phi_2(c(t))/\epsilon\phi_1(t)) < 1 + \delta$ for all $t \ge t_0$. By Lemma 3.4.2(a), $\epsilon\phi_1(s_n)/\phi_2(n) > 1/(1+\delta)$ for all $n \ge n_0$. Since ϵ can be arbitrarily small, $\phi_1(s_n)/\phi_2(n) \to \infty$ as $n \to \infty$.

(b) Reason as in (a) using (4.6), (4.7) and (4.9).

(c) Use (4.8), (4.13) and (4.14), noting that

$$\frac{1}{1-\epsilon} - \frac{\phi_2(c(t)+1) - \phi_2(c(t))}{\phi_1(t)} \le \frac{\phi_2(c(t))}{\phi_1(t)} \le \frac{1}{1-\epsilon} \tag{4.17}$$

and

$$\frac{\phi_2(c(t))}{\phi_2(c(t)+1)(1+\epsilon)} \le \frac{\phi_2(c(t))}{\phi_1(t)} \le \frac{1}{1-\epsilon}. \tag{4.18}$$

(d) Reason as in (c), using (4.15) and (4.16) with (4.17) and (4.18). ∎

Remark 3.4.1. Note that $\phi_2(c(t))/\phi_1(t) \to \lambda$ as $t \to \infty$ if and only if $\phi_2(c(\phi_1^{-1}(t)))/t \to \lambda$ as $t \to \infty$; i.e., the spatial normalization $\phi_1(t)$ is equivalent to the standard normalizing function e after making a time transformation by ϕ^{-1}. ∎

Example 3.4.1. *The need for an extra condition.* To see that an extra condition is needed in Theorem 3.4.1(c), let $s_n = n$ for all n, so that $c(t) = \lfloor t \rfloor$ for all t. Also let $\phi_1(t) = \phi_2(t) = e^t$ for all t. Then $\phi_1(s_n)/\phi_2(n) = 1$ for all n, while

$$\phi_2(c(t))/\phi_1(t) = e^{\lfloor t \rfloor - t},$$

which has limit supremum 1 and limit infimum e^{-1}. Also note that neither (4.13) nor (4.14) is satisfied.

Example 3.4.2. *The extra conditions are not necessary.* To see that the specific extra conditions in Theorem 3.4.1(c) are not necessary, let $s_n = e^n$ for all n, so that $c(t) = \lfloor \log t \rfloor$. Let $\phi_2(t) = e^t$ and $\phi_1(t) = t$ for all t. Then $\phi_1(s_n)/\phi_2(n) = 1$ for all n and

$$\frac{\phi_2(c(t))}{\phi_1(t)} = \frac{e^{\lfloor \log t \rfloor}}{t} \to 1 \quad \text{as} \quad t \to \infty,$$

but $\phi_2(n+1)/\phi_2(n) = e$ for all n and

$$\frac{\phi_2(c(t)+1) - \phi_2(c(t))}{\phi_1(t)} = \frac{(e-1)e^{\lfloor \log t \rfloor}}{t} \to e - 1 \quad \text{as} \quad t \to \infty\,. \quad ∎$$

A special case of interest is when the homeomorphisms are of the form $\phi(t) = t^p$ for $p > 0$. Of course, the case of greatest interest is $p = 1$; then we have simple averages.

Corollary 3.4.1. (the special case of powers) *Suppose that* $0 < p < \infty$ *and* $0 \le \lambda \le \infty$. *The following are equivalent:*

(i) $c(t)/t^p \to \lambda \quad \text{as} \quad t \to \infty,$

(ii) $(c(t))^{1/p}/t \to \lambda^{1/p} \quad \text{as} \quad t \to \infty,$

(iii) $s_n/n^{1/p} \to \lambda^{-1}/p \quad \text{as} \quad n \to \infty,$

(iv) $(s_n)^p/n \to \lambda^{-1} \quad \text{as} \quad n \to \infty\,.$

Proof. Apply Theorem 3.4.1 with $\phi_2(t) = t$ and $\phi_1(t) = t^p$ to relate (i) and (iv). Note that (4.14) holds. To relate (i) and (ii), note that $(c(t))^{1/p}/t = (c(t)/t^p)^{1/p}$, and similarly for (iii) and (iv).∎

We used the property that $\phi(x/y) = \phi(x)/\phi(y)$ for $\phi(x) = x^p$ in Corollary 3.4.1. The following classic lemma shows that this does not hold more generally.

Lemma 3.4.3. *A homeomorphism ϕ of $\mathbb{R}_+$ satisfies $\phi(xy) = \phi(x)\phi(y)$ for all nonnegative x and y if and only if $\phi(t) = t^p$ for some $p > 0$.*

Proof. The sufficiency is immediate. For the necessity, suppose that $\phi(xy) = \phi(x)\phi(y)$ for all nonnegative x and y. If we let $\psi(x) = \log\phi(e^x)$, then $\psi(x+y) = \psi(x) + \psi(y)$ for all real x and y. It is well known and easy to see that $\psi(x) = px$ for some real number p, which implies that $\phi(x) = e^{\psi(\log x)} = e^{p\log x} = x^p$. Since ϕ is strictly increasing, we must have $p > 0$. ∎

The Corollary to Theorem 3.4.1 is useful because it enables us to replace $\phi_2(c(t))/\phi_1(t)$ and $\phi_1(s_n)/\phi_2(n)$ by $c(t)/\phi_2^{-1}(\phi_1(t))$ and $s_n/\phi_1^{-1}(\phi_2(n))$ respectively. The following lemma shows that we can do this more generally.

Lemma 3.4.4. *Suppose that $\phi \in \Lambda(\mathbb{R}_+)$, $a_n \to \infty$ and $a_n/b_n \to 1$ as $n \to \infty$. If there is a t_0 such that $\log\phi(e^t)$ is uniformly continuous in (t_0, ∞), then $\phi(a_n)/\phi(b_n) \to 1$ as $n \to \infty$.*

Proof. Since $a_n \to \infty$ and $a_n/b_n \to 1$ as $n \to \infty$, $\log a_n - \log b_n \to 0$, $\log a_n \to \infty$ and $\log b_n \to \infty$ as $n \to \infty$. If $\log(\phi(e^t))$ is uniformly continuous in (t_0, ∞), then

$$\begin{aligned} &\log\phi(e^{\log a_n}) - \log\phi(e^{\log b_n}) = \log\phi(a_n) - \log\phi(b_n) \\ &= \log(\phi(a_n)/\phi(b_n)) \to 0 \quad \text{as} \quad n \to \infty, \end{aligned}$$

so that $\phi(a_n)/\phi(b_n) \to 1$ as $n \to \infty$. ∎

The following Corollary to Lemma 3.4.4 indicates how Lemma 3.4.4 can be applied in our context.

Corollary 3.4.2. *If $\phi_2(c(t))/\phi_1(t) \to \lambda$ as $t \to \infty$, where $\phi_1,\ \phi_2 \in \Lambda(\mathbb{R}_+)$ and $\log\phi_2^{-1}(e^t)$ is uniformly continuous in (t_0, ∞) for some t_0, then $c(t)/\phi_2^{-1}(\lambda\phi_1(t)) \to 1$ as $t \to \infty$.*

Remark 3.4.2. Lemma 3.4.4 implies Corollary 3.4.1 because $\log\phi(e^t) = \log\lambda + pt$ when $\phi(t) = \lambda t^p$. Another function covered by Lemma 3.4.4 is

$\phi(t) = a \log bt$; then $\log \phi(e^t)) = \log a + \log(\log b + t)$. However, $\log \phi(e^t) = \log a + be^t$ when $\phi(t) = ae^{bt}$, so that the uniform continuity does not hold when $\phi(t) = ae^{bt}$. ■

The following result is also useful to characterize the normalizing functions.

Lemma 3.4.5. *Suppose that $\phi \in \Lambda(\mathbb{R}_+)$, $0 < \lambda < \infty$ and $a_n \to \infty$ as $n \to \infty$. If there is a t_0 such that $\log \phi(e^t)$ is uniformly continuous in (t_0, ∞), then*

$$\overline{\lim_{n\to\infty}} \left| \frac{\phi(a_n)}{\phi(\lambda a_n)} \right| < \infty.$$

Proof. Recall that if a function ψ is uniformly continuous in (t_0, ∞), then

$$\sup\{|\psi(t + x) - \psi(t)| : t \geq t_0\} < \infty$$

for any positive x. Since

$$\log \lambda a_n - \log a_n = \log \lambda \ ,$$

$$\begin{aligned} \overline{\lim_{n\to\infty}} \ \{|\log \phi(e^{\log \lambda a_n}) - \log \phi(e^{a_n})|\} &= \overline{\lim_{n\to\infty}} \ \{|\log \phi(\lambda a_n) - \log \phi(a_n)|\} \\ &= \overline{\lim_{n\to\infty}} \ \{|\log(\phi(\lambda a_n)/\phi(a_n))| < \infty \ , \end{aligned}$$

which implies the desired conclusion.■

We are thinking of $\{s_n : n \geq 1\}$ being the points in a point process sample path, so it is natural to assume that $\{s_n\}$ is nondecreasing. However, we could start with a general sequence of real numbers $\{t_n : n \geq 1\}$ and obtain $\{s_n\}$ as the successive maxima, i.e.,

$$s_n \equiv t_n^{\uparrow} \equiv \max\{t_k : 0 \leq k \leq n\}, \quad n \geq 1, \tag{4.19}$$

where $t_0 = 0$. A similar result holds for $c(t)$. The following result closely parallel Proposition 3.3.1.

Proposition 3.4.1. *Suppose that $\phi_1, \ \phi_2 \in \Lambda(\mathbb{R}_+)$ and $0 \leq \lambda \leq \infty$. If $\phi_1(t_n)/\phi_2(n) \to \lambda^{-1}$ as $n \to \infty$, then $\phi_1(s_n)/\phi_2(n) \to \lambda^{-1}$ as $n \to \infty$ for s_n in (4.19).*

Proof. First assume that $0 < \lambda < \infty$. Given the assumed convergence, for all $\epsilon > 0$, there is an n_0 such that

$$\phi_1^{-1}(\phi_2(n)/\lambda(1+\epsilon)) \le t_n \le \phi_1^{-1}(\phi_2(n)/\lambda(1-\epsilon)) \quad \text{for all} \quad n \ge n_0,$$

which implies

$$\phi_1^{-1}(\phi_2(n)/\lambda(1+\epsilon)) \le s_n \le \max\{s_{n_0}, \phi_1^{-1}(\phi_2(n)/\lambda(1-\epsilon))\} \quad \text{for all} \quad n \ge n_0.$$

Let n_1 be such that

$$\phi_1^{-1}(\phi_2(n)/\lambda(1-\epsilon)) \ge s_{n_0} .$$

Then, for all $n \ge n_1$,

$$\frac{1}{\lambda(1+\epsilon)} \le \frac{\phi_1(s_n)}{\phi_2(n)} \le \frac{1}{\lambda(1-\epsilon)},$$

which implies the conclusion. For $\lambda = 0$ and $\lambda = \infty$ use associated one-sided inequalities. ∎

3.5. Counting Functions with Centering

We now turn to counting functions with centering. Due to the results for the inverse map with centering in Section 13.7 of the book, Theorem 13.8.2 in the book yields FCLTs for stochastic counting processes with centering given FCLTs for associated sequences of nondecreasing nonnegative random variables with centering, by an application of the continuous mapping theorem. We now show that we can also exploit the monotonicity to obtain *ordinary* CLTs for stochastic counting processes from associated *ordinary* CLTs for sequences of nondecreasing nonnegative random variables. The resulting CLT for stochastic counting process is the same as can be obtained from the FCLT by projection, but the condition is weaker. In both cases, we rely on an existing limit rather than specific stochastic assumptions. For this purpose, let $\{S_n : n \ge 0\}$ be a sequence of nondecreasing nonnegative random variables with $S_0 = 0$ and let $\{C(t) : t \ge 0\}$ be the associated stochastic counting process, defined as before by

$$C(t) \equiv \max\{k \ge 0 : S_k \le t\}, \quad t \ge 0 . \tag{5.1}$$

We again use regularly varying functions.

Theorem 3.5.1. (CLT equivalence) *Suppose that* $m > 0$ *and* $\psi \in \mathcal{R}(p)$ *for* $0 < p < 1$*. Then*

$$\psi(n)^{-1}[S_n - nm] \Rightarrow L \quad in \quad \mathbb{R} \quad as \quad n \to \infty\ , \tag{5.2}$$

where $\{S_n : n \geq 0\}$ *is a sequence of nondecreasing nonnegative random variables with* $S_0 = 0$ *if and only if*

$$\psi(t)^{-1}[C(t) - m^{-1}t] \Rightarrow -m^{-(1+p)}L \quad in \quad \mathbb{R} \quad as \quad n \to \infty\,, \tag{5.3}$$

where $\{C(t) : t \geq 0\}$ *is the associated stochastic counting process.*

We obtain Theorem 3.5.1 from a more general theorem which allows more general scalings, which are of value when analyzing nonstationary point processes.

Theorem 3.5.2. (CLT implications with more general scaling functions) *Suppose that* $\phi_1,\ \phi_2 \in \Lambda(\mathbb{R}_+)$, $\psi \in C_\uparrow$ *and*

$$\psi(t)/\psi(t + x\psi(t)) \to 1 \quad as \quad t \to \infty \tag{5.4}$$

for all x.

(a) If

$$X(t) \equiv \psi(\phi_1(t))^{-1}[\phi_2(C(t)) - \phi_1(t)] \Rightarrow L \quad in \quad \mathbb{R} \quad as \quad n \to \infty\,, \tag{5.5}$$

then

$$Y(n) = \psi(\phi_2(n))^{-1}[\phi_1(S_n) - \phi_2(n)] \Rightarrow -L \quad in \quad \mathbb{R} \quad as \quad n \to \infty\,. \tag{5.6}$$

(b) If (5.6) *above holds, then there exists an increasing sequence of positive real numbers* $\{t_n : n \geq 1\}$ *with* $t_n \to \infty$ *as* $n \to \infty$ *such that* $X(t_n) \Rightarrow L$ *for* $X(t)$ *in* (5.5) *above.*

(c) If, in addition to (5.6) *above,*

$$[\phi_2(n+1) - \phi_2(n)]/\psi(\phi_2(n)) \to 1 \quad as \quad n \to \infty \tag{5.7}$$

and

$$\psi(\phi_2(n+1))/\psi(\phi_2(n)) \to 1 \quad as \quad n \to \infty\,, \tag{5.8}$$

then (5.5) *above holds.*

We first apply Theorem 3.5.2 to prove Theorem 3.5.1.

Proof of Theorem 3.5.1. We apply Theorem 3.5.2 with $\phi_1(t) = mt$, $\phi_2(t) = t$ and $\psi \in \mathcal{R}(p)$ for $0 < p < 1$. It is easy to see that ψ satisfies (5.4): For any x, there is a t_0 such that

$$\psi((1-\epsilon)t) \le \psi(t + x\psi(t)) \le \psi((1 = \epsilon)t) \tag{5.9}$$

for all $t \ge t_0$, from which it follows that

$$\left(\frac{1}{1+\epsilon}\right)^p \le \frac{\psi(t)}{\psi(t + x\psi(t))} \le \left(\frac{1}{1-\epsilon}\right)^p \tag{5.10}$$

for all suitably large t. We also apply the regular variation property to deduce that $\psi(\phi_1(t))$ in (5.5) has the asymptotic form

$$\psi(\phi_1(t)) = \psi(mt) \sim m^p \psi(t) \quad \text{as} \quad t \to \infty \ . \tag{5.11}$$

Thus (5.2) is equivalent to (5.6) with the limit in (5.2) changed to $m^p L$. Similarly, (5.3) is equivalent to (5.5) with the limit in (5.3) changed to $-m^{-1}L$. Thus the form of the limits in (5.2) and (5.3) follow from (5.5) and (5.6). Finally, it remains to observe that the assumptions in Theorem 3.5.1 imply that conditions (5.7) and (5.8) hold. ∎

We now turn to the proof of Theorem 3.5.2. For that purpose, we use a basic lemma about cumulative distribution functions (cdf's).

Lemma 3.5.1. *Let F_n, $n \ge 0$, be cdf's. The following are equivalent:*

(i) For each $t \in Disc(F_0)^c$, $F_n(t_n) \to F_0(t)$ as $n \to \infty$ for some sequence $\{t_n : n \ge 1\}$ with $t_n \to t$.

(ii) $F_n \Rightarrow F_0$; i.e., for each $t \in Disc(F_0)^c$, $F_n(t) \to F_0(t)$ as $n \to \infty$.

(iii) For each $t \in Disc(F_0)^c$ and all sequences $\{t_n : n \ge 1\}$ with $t_n \to t$ as $n \to \infty$, $F_n(t_n) \to F_0(t)$ as $n \to \infty$.

Proof. Clearly (iii)→(ii)→(i), so it suffices to show that (i)→(iii). Let $t \in Disc(F_0)^c$. Then, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$F_0(t) - \epsilon \le F_0(t-\delta) \le F_0(t) \le F_0(t+\delta) \le F_0(t) + \epsilon \ . \tag{5.12}$$

Since F_0 is nondecreasing, it has at most countably many discontinuities. Let t', $t'' \in Disc(F_0)^c$ be such that $t - \delta < t' < t < t'' < t + \delta$. Given (i), there exist sequences $\{t'_n : n \ge 1\}$ and $\{t''_n : n \ge 1\}$ such that $t'_n \to t'$, $t''_n \to t''$, $F_n(t'_n) \to F_0(t')$ and $F_n(t''_n) \to F_0(t'')$ as $n \to \infty$. Let $\{t_n : n \ge 1\}$ be any sequence such that $t_n \to t$ as $n \to \infty$. Hence, there is an n_0 such that

$$F_0(t-\delta) < F_n(t'_n) \le F_n(t_n) \le F_n(t''_n) \le F_0(t+\delta) \tag{5.13}$$

for all $n \geq n_0$. Combining (5.12) and (5.13), we see that

$$F_0(t) - \epsilon \leq F_n(t_n) \leq F_0(t) + \epsilon \ . \quad \blacksquare \tag{5.14}$$

Proof of Theorem 3.5.2. (a) Suppose that (5.5) holds. Then

$$F_t(x-) \equiv P(X(t) < x) \to P(L < x) \equiv F(x) \quad \text{as} \quad t \to \infty \tag{5.15}$$

for each $x \in Disc(F)^c$. However,

$$\begin{aligned} F_t(x-) &= P(\phi_2(C(t)) < \phi_1(t) + x\psi(\phi_1(t))) \\ &= P(C(t) < \phi_2^{-1}(\phi_1(t) + x\psi(\phi(t))) \end{aligned}$$

so that, by Lemma 3.4.1, $F_t(x) = P(S_{n(t)} > t)$ for any t such that

$$n(t) \equiv \phi_2^{-1}(\phi_1(t) + x\psi(\phi_1(t))) \tag{5.16}$$

is an integer. For such t,

$$F_t(x-) = P(\psi(\phi_2(n(t)))^{-1}[\phi_1(S_{n(t)}) - \phi_2(n(t))] > -x(t)) \,, \tag{5.17}$$

where

$$\begin{aligned} x(t) &= -[\phi_1(t) - \phi_2(n(t))]/\psi(\phi_2(n(t))) \\ &= x\psi(\phi_1(t))/\psi(\phi_1(t) + x\psi(\phi_1(t))) \to x \quad \text{as} \quad t \to \infty \end{aligned} \tag{5.18}$$

by (5.16) and (5.4). Note that, for each positive integer n, we can find t_n such that $n(t_n) = n$, because ϕ_1, ϕ_2 and ψ are nondecreasing and continuous, and $n(t) \to \infty$ as $t \to \infty$. Hence

$$G_n(x_n) \equiv P(\psi(\phi_2(n))^{-1}[\phi_1(S_n) - \phi_2(n)] < x_n) = F_{t_n}(x_n) \tag{5.19}$$

where $x_n = x(t_n) \to x$ as $n \to \infty$. Since $x \in Disc(F)^c$, $F_{t_n}(x_n) \to F(x)$ as $n \to \infty$. By Lemma 3.5.1, $G_n \Rightarrow F$, so that $Y(n) \Rightarrow -L$.

(b) Let the cdf G_n be defined by (5.19). Then

$$G_n(x) = P(S_n > \phi_1^{-1}(\phi_2(n) - x\psi(\phi_2(n)))) = P(A(t_n) < n)$$

for

$$t_n = \phi_1^{-1}(\phi_2(n) - x\psi(\phi_2(n))) \tag{5.20}$$

by Lemma 3.4.1. Thus, for F_t in (5.15) and t_n in (5.20), $F_{t_n}(x_n-) = G_n(x)$ for

$$\begin{aligned} x_n &= [\phi_2(n) - \phi_1(t_n)]/\psi(\phi_1(t_n)) \\ &= x\psi(\phi_2(n))/\psi(\phi_2(n) - x\psi(\phi_2(n)) \to x \quad \text{as} \quad n \to \infty \end{aligned} \tag{5.21}$$

by (5.20) and (5.4). Hence, if $G_n(x) \to F(x)$ for $x \in Disc(F)^c$, then $F_{t_n}(x_n) \to F(x)$ as $n \to \infty$. By Lemma 3.5.1, $F_{t_n} \Rightarrow F$, so that $X(t_n) \Rightarrow L$.

(c) For any t, let n be such that $t_n \le t < t_{n+1}$ for t_n in (5.20). Since $C(t_n) \le C(t) \le C(t_{n=1})$, it suffices to show that $C(t_n)$ and $C(t_{n+1})$ have the same limits with the normalization. It suffices to show that

$$\psi(\phi_1(t_{n+1}))^{-1}[\phi_2(C(t_n)) - \phi_1(t_{n+1})] \Rightarrow L \ , \tag{5.22}$$

which in turn holds if

$$\psi(\phi_1(t_{n+1}))/\psi(\phi_1(t_n)) \to 1 \tag{5.23}$$

and

$$[\phi_1(t_{n+1}) - \phi_1(t_n)]/\psi(\phi_1(t_n)) \to \quad \text{as} \quad n \to \infty \, . \tag{5.24}$$

By (5.4) and (5.20), (5.23) is equivalent to (5.8). By (5.20) and (5.23), (5.24) is equivalent to (5.7): Applying (5.20) and dividing numerator and denominator by $\psi(\phi_2(n))$, we see that (5.21) becomes A_n/B_n, where

$$B_n = \psi(\phi_2(n) - x\psi(\phi_2(n)))/\psi(\phi_2(n)) \to 1 \quad \text{as} \quad n \to \infty \tag{5.25}$$

by (5.4) and

$$A_n = [\phi_2(n+1) - x\psi(\phi_2(n+1)) - \phi_2(n) + x\psi(\phi_2(n)]/\psi(\phi_2(n)) \tag{5.26}$$

by (5.8) and (5.7). ■

Example 3.5.1. *It is possible that only subsequences converge.* To see that we can have $X(t_n) \Rightarrow L$ as $n \to \infty$ in Theorem 3.5.2 (b) without $X(t) \Rightarrow L$ as $t \to \infty$ for $X(t)$ in (5.5), let $\phi_1(t) = t$, $\phi_2(t) = t^2$ and $\psi(t) = 1$ for $t \ge 0$. Then (5.4) and (5.8) hold, but (5.7) does not:

$$\frac{\phi_2(n+1) - \phi_2(n)}{\psi(\phi_2(n))} = \frac{(n+1)^2 - n^2}{n^2} \to 0 \ . \tag{5.27}$$

Let $P(L = 0) = 1$ and let $S_n = n^2$, so that

$$\psi(\phi_2(n))^{-1}[\phi_1(S_n) - \phi_2(n)] = 0 = -L \quad \text{w.p.1 for all } n. \tag{5.28}$$

However, $C(t) = \sqrt{\lfloor t \rfloor}$ and $\phi_2(C(t)) = \lfloor t \rfloor$, so that

$$\psi(\phi_1(t))^{-1}[\phi_2(C(t)) - \phi_1(t)] = \lfloor t \rfloor - 1 \, , \tag{5.29}$$

from which we see that

$$\varliminf_{t \to \infty} X(t) = -1 < 0 = \varlimsup_{t \to \infty} X(t)$$

for $X(t)$ in (5.5). ■

A major theme here is obtaining probabilistic limits directly from deterministic limits. Thus it is natural to ask if there is a deterministic analog of Theorem 3.5.2 that implies Theorem 3.5.2. We show that there is. In particular, the following result implies parts (a) and (c) of Theorem 3.5.2.

Theorem 3.5.3. (deterministic analog of Theorem 3.5.2) *Let* $\phi_1,\ \phi_2 \in \Lambda(\mathbb{R}_+)$ *and let* ψ *be a continuous positive real-valued function on* $[0,\infty)$ *for which (5.4) holds*

(a) If

$$x(t) \equiv \psi(\phi_1(t))^{-1}[\phi_2(c(t)) - \phi_1(t)] \to \alpha \quad in \quad \mathbb{R} \quad as \quad t \to \infty\ , \tag{5.30}$$

then

$$y(n) \equiv \psi(\phi_2(n))^{-1}[\phi_1(s_n) - \phi_2(n)] \to -\alpha \quad in \quad \mathbb{R} \quad as \quad n \to \infty\ . \tag{5.31}$$

. *(b) If, in addition to* (5.31) *here,* (5.7) *and* (5.8) *above hold, then* (5.30) *here holds.*

Proof. (a) If (5.30) holds, then for all $\epsilon > 0$ there exists t_0 such that $\alpha - \epsilon \le x(t) < \alpha + \epsilon$ for all $t \ge t_0$. Given that $x(t) < \alpha + \epsilon$,

$$\phi_2^{-1}(\phi_1(t) + (\alpha - \epsilon)\psi(\phi_1(t)) \le c(t) < \phi_2^{-1}(\phi_1(t) + (\alpha + \epsilon)\psi(\phi_1(t)))\,. \tag{5.32}$$

Let t be such that

$$n(t) \equiv \phi_2^{-1}(\phi_1(t) + (\alpha + \epsilon)\psi(\phi_1(t))) \tag{5.33}$$

is an integer. By Lemma 3.4.1,

$$s_{n(t)} > t\ . \tag{5.34}$$

Given (5.34),

$$y(n(t)) \equiv \psi(\phi_2(n(t)))^{-1}[\phi_1(s_{n(t)}) - \phi_2(n(t))] > -\alpha(t) \tag{5.35}$$

where

$$\alpha(t) \equiv \frac{\phi_1(t) - \phi_2(n(t))}{\psi(\phi_2(n(t)))} = \frac{-(\alpha+\epsilon)\psi(\phi_1(t))}{\psi(\phi_1(t) + (\alpha+\epsilon)\psi(\phi_1(t)))} \to -(\alpha+\epsilon) \tag{5.36}$$

by (5.4), so that $y(n(t)) > -(\alpha + 2\epsilon)$ for all $t \ge t_1 \ge t_0$. Since for each positive integer n, we can find t such that $n(t) = n$, there is an n_0 such that $y(n) > -(\alpha + 2\epsilon)$ for all $n \ge n_0$. Similarly, from the lower bound in (5.32), we can conclude that $y(t) \leq -(\alpha - 2\epsilon)$ for all $n \geq n_1$. Since ϵ was arbitrary, the proof is complete.

(b) For any $\epsilon > 0$, there exists n_0 such that $-\alpha - \epsilon < y(n) \leq -\alpha + \epsilon$ for $n > n_0$. As a consequence,

$$\phi_1^{-1}(\phi_2(n) - (\alpha + \epsilon)\psi(\phi_2(n))) < s_n \leq \phi_1^{-1}(\phi_2(n) - (\alpha - \epsilon)\psi(\phi_2(n))) \quad (5.37)$$

for all $n \geq n_0$. Now let

$$t_n \equiv \phi_1^{-1}(\phi_2(n) - (\alpha - \epsilon)\psi(\phi_2(n))) \; . \quad (5.38)$$

By Lemma 3.4.1,

$$c(t_n) \geq n \quad (5.39)$$

and

$$\begin{aligned} x(t_n) &= \frac{\phi_2(c(t_n)) - \phi_1(t_n)}{\psi(\phi_1(t_n))} \\ &\geq \frac{\phi_2(n) - \phi_1(t_n)}{\psi(\phi_1(t_n))} \\ &= \frac{(\alpha - \epsilon)\psi(\phi_2(n))}{\psi(\phi_2(n) - (\alpha - \epsilon)\psi(\phi_2(n)))} \geq \alpha - 2\epsilon \end{aligned} \quad (5.40)$$

for $n \geq n_1 \geq n_0$ by (5.4). We now want to show that there is t_0 such that $x(t) \geq \alpha - 3\epsilon$ for all $t \geq t_0$. Consider t with $t_n \leq t < t_{n+1}$. Notice that

$$\phi_2(c(t_n)) - \phi_1(t_{n+1)} \leq \phi_2(c(t)) - \phi_1(t) \leq \phi_2(c(t_{n+1})) - \phi_1(t_n)) \, . \quad (5.41)$$

Since (5.41) holds, (5.40), (5.7) and (5.8) imply that there is t_0 such that $x(t) > \alpha - 3\epsilon$ for $t > t_0$. Similarly, using the lower bound in (5.37), we can deduce that for any $\epsilon > 0$ there exists t_0 such that $x(t) < \alpha + 3\epsilon$ for $t > t_0$. Since ϵ was arbitrary, the proof is complete. ■

3.6. Composition

We now turn to the composition map. We first state a preliminary lemma.

Lemma 3.6.1. *If $\phi \in \mathcal{R}(p)$ with $p > 0$, and $t^{-q}y(t) \to \mu > 0$, then*

$$\phi(y(t))/\phi(t^q) \to \mu^p \quad as \quad t \to \infty \, . \quad (6.1)$$

Proof. For any $\epsilon > 0$, there is t_0 such that $t^q(\mu - \epsilon) < y(t) < t^q(\mu + \epsilon)$. Since ϕ is regularly varying with index p,

$$\overline{\lim_{t\to\infty}} \frac{\phi(y(t))}{\phi(t^q)} \le \lim_{t\to\infty} \frac{\phi(t^q(\mu+\epsilon))}{\phi(t^q)} \le (\mu+\epsilon)^p$$

and

$$\underline{\lim_{t\to\infty}} \frac{\phi(y(t))}{\phi(t^q)} \ge \lim_{t\to\infty} \frac{\phi(t^q(\mu-\epsilon))}{\phi(t^q)} \ge (\mu-\epsilon)^p . \quad \blacksquare$$

Proposition 3.6.1. *Suppose that $\phi \in \mathcal{R}(p)$ with $p > 0$. If*

$$\phi(t)^{-1}X(t) \Rightarrow U \quad \textit{and} \quad t^{-1}Y(t) \Rightarrow \mu \quad \textit{in} \quad \mathbb{R} , \tag{6.2}$$

then

$$\phi(t)^{-1}X(Y(t)) \Rightarrow \mu^p U \quad \textit{in} \quad \mathbb{R} . \tag{6.3}$$

Proof. Since the limit μ in (6.2) is deterministic, we have the joint limit

$$(\phi(t)^{-1}X(t), t^{-1}Y(t)) \Rightarrow (U, \mu) \quad \text{in} \quad R^2 . \tag{6.4}$$

Use the Skorohod representation theorem to replace convergence in distribution in (6.4) with convergence w.p.1 (for special versions). By Lemma 3.6.1, $\phi(Y(t))/\phi(t) \to \mu^p$. Then

$$\frac{X(Y(t))}{\phi(t)} = \frac{\phi(Y(t))}{\phi(t)} \frac{X(Y(t))}{\phi(Y(t))} \to \mu^p U . \tag{6.5}$$

Finally, (6.5) implies (6.3). $\blacksquare$

Proposition 3.6.2. *Suppose that $\phi \in \mathcal{R}(p)$ with $0 < p < 1$. If*

$$\phi(t)^{-1}[X(t) - \lambda t, Y(t) - \mu t] \Rightarrow (U, V) \quad \textit{in} \quad \mathbb{R}^2 \tag{6.6}$$

then

$$\phi(t)^{-1}[X(Y(t)) - \lambda\mu t] \Rightarrow \mu^p U + \lambda V \quad \textit{in} \quad \mathbb{R}^2 . \tag{6.7}$$

Proof. From (6.6) the regular variation condition, we have $t^{-1}Y(t) \Rightarrow \mu$ and $\phi(Y(t))/\phi(t) \to \mu^p$ as $t \to \infty$. Now replace convergence in distribution by convergence w.p.1 for special versions. Then

$$\frac{\phi(Y(t))}{\phi(t)} \frac{X(Y(t)) - \lambda Y(t)}{\phi(Y(t)} + \frac{\lambda Y(t) - \lambda \mu t}{\phi(t)} \to \mu^p U + \lambda V \text{ w.p.1 as } t \to \infty\,, \tag{6.8}$$

which implies (6.7). ∎

There are two difficulties with Propositions (3.6.1) and (3.6.2) for applications. An obvious difficulty is that we may actually need the stronger conclusions giving limits in D in applications. The other difficulty is that it may be difficult to obtain the conditions. The joint limit in (6.6) holds if the component limits hold in $\mathbb{R}$ when $X(t)$ and $Y(t)$ are independent, but in most applications $X(t)$ and $Y(t)$ are actually dependent. A critical step then is to establish condition (6.6).

To illustrate, we may start with the sequence $\{(A_n, B_n) : n \geq 1\}$ of ordered pairs of nonnegative random variables. We may be able to determine that

$$\phi(n)^{-1}[A_n - n\lambda, B_n - n\mu^{-1}] \Rightarrow (U', V') \quad \text{as} \quad n \to \infty \quad \text{in} \quad \mathbb{R}^2 \tag{6.9}$$

and be interested in the asymptotic behavior of $A_{C(t)}$, where

$$C(t) = \max\{k \geq 1 : B_k \leq t\}, \quad t \geq 0\,. \tag{6.10}$$

From the second component of (6.9), we can determine that

$$\phi(t)^{-1}[C(t) - \mu t] \Rightarrow -\mu^{(1+p)} V' \tag{6.11}$$

from Theorem 3.5.1. However, we have difficulty directly expressing the joint limits of

$$\phi(n)^{-1}(A_n - n\lambda) \quad \text{as} \quad n \to \infty \quad \text{and} \quad \phi(t)^{-1}[C(t) - \mu t] \quad \text{as} \quad t \to \infty\,. \tag{6.12}$$

The extension of (6.9) in D offers a resolution. We can hopefully extend (6.9) to

$$(\mathbf{A}_n, \mathbf{B}_n) \Rightarrow (\mathbf{U}, \mathbf{V}) \quad \text{in} \quad D^2 \tag{6.13}$$

where

$$\begin{aligned} \mathbf{A}_n(t) &\equiv \phi(n)^{-1}[A_{\lfloor nt \rfloor} - \lambda nt] \\ \mathbf{B}_n(t) &\equiv \phi(n)^{-1}[B_{\lfloor nt \rfloor} - \mu^{-1} nt] \\ (\mathbf{U}(1), \mathbf{V}(1)) &\stackrel{\mathrm{d}}{=} (U', V')\,. \end{aligned} \tag{6.14}$$

From (6.13), we can get

$$(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n) \Rightarrow (\mathbf{U}, \mathbf{V}, -\mu\mathbf{V} \circ \mu\mathbf{e}) \tag{6.15}$$

where

$$\mathbf{C}_n(t) \equiv \phi(n)^{-1}[C(nt) - \mu nt] \ . \tag{6.16}$$

We can then apply the composition map in D. In particular, letting

$$\mathbf{\Phi}_n(t) \equiv n^{-1}C(nt) \ , \tag{6.17}$$

and

$$\mathbf{X}_n \equiv \phi(n)^{-1}[A_{C(nt)} - \lambda\mu nt] \, , \tag{6.18}$$

we obtain

$$\mathbf{X}_n = \mathbf{A}_n \circ \mathbf{\Phi}_n + \mu\mathbf{C}_n \Rightarrow \mathbf{U} \circ \lambda\mathbf{e} + \mu\mathbf{C} \quad \text{in} \quad D \tag{6.19}$$

under regularity conditions, by Theorem 13.3.2 in the book. As a consequence,

$$\phi(n)^{-1}[A_{C(n)} - \lambda\mu n] \Rightarrow \mathbf{U}(\lambda) + \mu\mathbf{C}(1) \quad \text{in} \quad \mathbb{R}, \tag{6.20}$$

assuming that $P(1 \in Disc(\mathbf{U} \circ \lambda\mathbf{e} + \mu\mathbf{C})) = 0$.

Alternative approaches have been developed for dealing with this problem directly, starting with the Anscombe (1952, 1953) condition, see Gut (1988), but those conditions are essentially equivalent to $\mathbf{A}_n \Rightarrow \mathbf{U}$ with $P(\mathbf{U} \in C) = 1$.

3.7. Chapter Notes

The main results in this chapter are so basic that they no doubt have a long history, but we are unable to trace that history beyond our own work. Much related material, with emphasis on the classical case of partial sums of i.i.d. random variables, appears in Gut (1988).

We have primarily drawn upon Glynn and Whitt (1986, 1988) and Massey and Whitt (1994). Those papers contain further applications to queues related to the conservation law $L = \lambda W$. El-Taha and Stidham (1999) is closely related from that perspective. El-Taha and Stidham demonstrate the far-reaching implications possible from pointwise limits for single functions (sample-path analysis). Baccelli and Bremaud (1994) provide an alternative treatment of many of the same topics in the context of stationary processes. An overview of $L = \lambda W$ appears in Whitt (1991, 1992).

The strengthening of pointwise convergence to uniform convergence in Theorem 3.2.1 extends Theorem 4 of Glynn and Whitt (1988), which was in the form of Corollary 3.2.1. For the case $\phi(t) = t$, Proposition 3.3.1 is implication (iii)$\rightarrow$(v) in Theorem 2(b) of Glynn and Whitt (1986). The more general version appears in Section 2.5 of El-Taha and Stidham (1999).

Theorem 3.5.1 here extends Theorem 4.2 of Massey and Whitt (1994) by allowing the space scaling function ψ to be regularly varying instead of a simple power. Lemma 3.5.1 is an improved statement of Lemma 4.1 of Massey and Whitt (1994). The deterministic basis for Theorem 3.5.2 in Theorem 3.5.3 is new here.

An extensive treatment of the composition map and convergence in distribution under a random time change appears in Gut (1988). The first few sections there provide useful perspective. A related result is the conservation law $Y = \lambda X$ in El-Taha and Stidham (1999).

Chapter 4

An Application to Simulation

4.1. Introduction

In Sections 5.9 and 10.4.4 of the book we showed how heavy-traffic stochastic-process limits for queues can be used to help plan queueing simulations. In this chapter we discuss another application of stochastic-process limits to simulation. We draw on Glynn and Whitt (1992a). In Section 4.2 we show how stochastic-process limits and the continuous-mapping approach can be used to determine general criteria for sequential stopping rules to be asymptotically valid.

Yet another application of stochastic-process limits and the continuous-mapping approach to simulation is contained in Glynn and Whitt (1992b). Glynn and Whitt (1992b) shows how stochastic-process limits and the continuous-mapping approach can be exploited to determine the asymptotic efficiency of simulation estimators. These two applications can be applied to queueing simulations, but they are not limited to queueing simulations.

4.2. Sequential Stopping Rules for Simulations

In this section, following Glynn and Whitt (1992a), we show how FCLTs and the continuous-mapping approach can be used to establish general conditions for the asymptotic validity of sequential stopping rules for stochastic simulations. The general conditions are expressed in terms of FCLTs and FWLLNs. The conditions allow the possibility of limit processes with discontinuous sample paths, but usually the limit process will be related to

Brownian motion, and thus have continuous sample paths. We use the composition and inverse maps to demonstrate the asymptotic validity.

The goal is to estimate a deterministic parameter $\alpha \in \mathbb{R}^k$. We start with an $\mathbb{R}^k$-valued stochastic process $Y \equiv \{Y(t) : t > 0\}$ called the *estimation process*. We think of $Y(t)$ as being the estimate of α based on a simulation with runlength t. The results also apply to statistical estimation more generally, but we are especially concerned with simulation.

With simulation, a common problem is to estimate a steady-state mean vector α. The simulation may be used to generate a stochastic process $X \equiv \{X(t) : t \geq 0\}$, where $X(t) \Rightarrow X(\infty)$ in $\mathbb{R}^k$ as $t \to \infty$. We may then want to estimate the steady-state mean $\alpha \equiv EX(\infty) \equiv [EX^1(\infty), \ldots, EX^k(\infty)]$ by the sample mean

$$Y(t) \equiv t^{-1} \int_0^t X(s)ds, \quad t > 0 . \tag{2.1}$$

That is a common way for the estimation process Y to arise, but not the only way.

The simulator must select a runlength t. The runlength can be selected either in advance or sequentially while the simulation is in process. The principal disadvantage of selecting the runlength in advance is that the posterior precision of the estimator may not be appropriate. Since the volume of the confidence set (the width of a confidence interval in one dimension) is unknown in advance, the volume may be too large to be of practical use (meaning that the preassigned runlength was too small) or too small (meaning that computational resources were wasted in refining the estimator beyond the level of accuracy required).

We are interested in sequential procedures in which we let the simulation run until the volume of a confidence set achieves a prescribed value. That avoids the problems associated with preassigned runlengths, but new difficulties are introduced because the runlength is now randomly determined. The first difficulty is that we no longer have direct control of the amount of simulation time to be generated or the amount of computer time to be expended. Consequently, the runlength may turn out to be much longer than we want. On the other hand, it is possible that the runlength may turn out to be inappropriately short. This creates certain statistical difficulties that can compromise the accuracy of such procedures. For example, it is known that in many statistical settings, the point estimator and the variance estimator are positvely correlated. Since the volume of a confidence set is typially determined by the variance estimator, this suggests that the confidence set volume will tend to be small when the point estimator is small.

Consequently, the resulting sequential procedure will tend to terminate early in situations in which the point estimator is too small, leading to possibly significant coverage problems for the confidence sets. Nevertheless, sequential stopping rules are of interest because of the possiblity of automatically obtaining prescribed precision.

Various sequential stopping rules for simulation estimators have been proposed and investigated empirically. Among these are sequential procedures involving: batch means in Law and Carson (1979) and Law and Kelton (1982), regenerative simulation in Fishman (1977) and Lavenberg and Sauer (1977) and spectral methods in Heidelberger and Welch (1981a, b, 1983); see pages 81, 92, 97 and 103 of Bratley, Fox and Schrage (1987) for an overview. Unfortunately, however, the empirical evidence is not entirely encouraging. Evidently, care must be taken in the design and implementation of sequential procedures to avoid inappropriate early termination. On the positive side, the sequential procedures do tend to perform well when the run lengths are relatively long, which is achieved in part by having a suitably small precribed volume for the confidence set. The observed good performance with small prescribed confidence set volumes is consistent with the asymptotic theory to be developed below. The asymptotic theory for general simulation estimators below is in turn consistent with the classical asymptotic theory associated with the sample mean of i.i.d. random variables; we cite references below.

4.2.1. The Mathematical Framework

To start, we assume that the estimation process Y satisfies a CLT, i.e.,

$$\phi(t)[Y(t)-\alpha] \Rightarrow \Gamma L \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad t \to \infty\,, \tag{2.2}$$

where Γ is a nonsingular $k \times k$ scaling matrix and $\phi(t)$ is a real-valued scaling function with $\phi(t) \to \infty$ as $t \to \infty$. The common case for ϕ is $\phi(t) = t^{1/2}$, in which case the limit L in (2.2) typically is $N(0, I)$, a standard normal random vector with the identity matrix I as its covariance matrix, but we want to allow for other possibilities. With heavy-tailed probability distributions or long-range dependence, we might have $\phi(t) = t^\gamma$ for $\gamma < 1/2$ or, more generally, ϕ regularly varying with index γ. The treatment here generalizes Glynn and Whitt (1992a) by allowing regularly varying scaling functions instead of simple powers.

As a consequence of (2.2),

$$Y(t) \Rightarrow \alpha \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad t \to \infty\,. \tag{2.3}$$

The limit (2.3) says that the estimation process is *weakly consistent.* Of course, weak consistency is a minimal requirement.

We assume that the confidence sets are all based on a bounded measurable subset A of $\mathbb{R}^k$ with $m(A) > 0$, where m is Lebesgue measure on $\mathbb{R}^k$. To obtain approximate $100(1-\delta)\%$ confidence sets for α, we assume that

$$P(L \in A) = 1 - \delta \quad \text{and} \quad P(L \in \partial A) = 0 \ , \tag{2.4}$$

where L is the limiting random variable in (2.2) and ∂A is the boundary of the set A, i.e., $\partial A = A^- - A^o$, where A^- and A^o are the closure and interior of A. Given that we know A and Γ, we can let the *confidence set* be

$$\tilde{C}(t) \equiv Y(t) - \phi(t)\Gamma A \ , \tag{2.5}$$

where

$$z + QA \equiv \{x \in R^d : \text{there exists } y \in A \text{ such that } x = z + Qy\} \, .$$

The confidence set $\tilde{C}(t)$ in (2.5) clearly depends on t. When the runlength t is specified in advance, the confidence set is asymptotically valid, in the sense of the following proposition.

Proposition 4.2.1. *If* (2.2) *and* (2.4) *hold, then*

$$P(\alpha \in \tilde{C}(t)) \to 1 - \delta \quad as \quad t \to \infty$$

for $\tilde{C}(t)$ *in* (2.5).

Proof. Since Γ is nonsingular,

$$P(\alpha \in \tilde{C}(t)) = P(\Gamma^{-1}\phi(t)(Y(t) - \alpha) \in A) \ ,$$

but

$$\Gamma^{-1}\phi(t)(Y(t) - \alpha) \Rightarrow \Gamma^{-1}\Gamma L = L \quad \text{as} \quad t \to \infty$$

by (2.2). Since (2.4) holds,

$$P(\Gamma^{-1}\phi(t)(Y(t) - \alpha) \in A) \to P(L \in A) = 1 - \delta \quad \text{as} \quad t \to \infty$$

by Theorem 11.3.4 (v) in the book. ■

Of course, in applications the scaling matrix Γ is typically unknown, so that it too must be estimated. We assume that there is an estimator $\Gamma(t)$ that is weakly consistent, i.e.,

$$\Gamma(t) \Rightarrow \Gamma \quad \text{in} \quad \mathbb{R}^{k^2} \quad \text{as} \quad t \to \infty \ . \tag{2.6}$$

Given an estimator $\Gamma(t)$, $t > 0$, we can form approximate confidence sets based on $\Gamma(t)$. For that purpose, let

$$C(t) \equiv Y(t) - \phi(t)\Gamma(t)A \ . \tag{2.7}$$

We now extend Proposition 4.2.1 to include $\Gamma(t)$ instead of Γ.

Proposition 4.2.2. *If, in addition to the assumptions of Proposition* 4.2.1 *above,* (2.6) *holds, then*

$$P(\alpha \in C(t)) \to 1 - \delta \quad as \quad t \to \infty$$

for $C(t)$ *in* (2.7).

Proof. By (2.2) above and Theorem 11.4.5 in the book,

$$(\Gamma(t), \phi(t)(Y(t) - \alpha)) \Rightarrow (\Gamma, \Gamma L) \quad \text{as} \quad t \to \infty \ .$$

Then noting that matrix inversion is continuous at all nonsingular limits, we can deduce that $\Gamma(t)$ is nonsingular, and thus invertible, for all sufficiently large t and then apply the continuous mapping theorem to obtain

$$\Gamma(t)^{-1}\phi(t)(Y(t) - \alpha) \Rightarrow \Gamma^{-1}\Gamma L \quad \text{as} \quad t \to \infty \, .$$

The rest of the proof is the same as the last part of the proof of Proposition 4.2.1. ■

We now use the confidence set $C(t)$ in (2.7) to define sequential stopping rules. Recall that, for a generic (measurable) set $B \subseteq \mathbb{R}^k$, $m(B)$ denotes the k-dimensional volume (Lebesgue measure) of the set. Of course, when $k = 1$ and B is an interval, $m(B)$ is just the length of the interval. We first consider the case in which the procedure terminates when the k^{th} root of the volume of the confidence region $C(t)$ drops below a prescribed level ϵ. [It is natural to use the k^{th} root, because $m(cB)^{1/k} = cm(B)^{1/k}$ for a scalar c.] We call such a procedure an *absolute-precision sequential stopping rule.* For such a rule, the time $\tilde{T}(\epsilon)$ at which the simulation terminates execution is defined by

$$\tilde{T}(\epsilon) = \inf\{t \geq 0 : m(C(t))^{1/k} < \epsilon\} \ . \tag{2.8}$$

Actually, this stopping rule needs to be modified, because $\tilde{T}(\epsilon)$ in (2.8) can terminate much too early if the estimator $\Gamma(t)$ is badly behaved for small t. To see this, suppose that $P(\Gamma(1) = 0,\ m(C(t)) = 1,\ 0 \leq t < 1) = 1$. In this case, $\tilde{T}(\epsilon) = 1$ for $\epsilon < 1$, so $C(\tilde{T}(\epsilon)) = Y(1)$ for $\epsilon < 1$. Hence, in this example, $P(\alpha \in C(\tilde{T}(\epsilon))) = P(\alpha = Y(1))$ for $\epsilon < 1$. Hence convergence of the coverage probability of the region $C(T(\epsilon))$ to the nominal level $1 - \delta$ does *not* occur when we let $\epsilon \downarrow 0$.

In order for the asymptotic theory to be relevant to the sequential stopping problem, it is necessary that $T(\epsilon) \to \infty$ as $\epsilon \downarrow 0$. In other words, small values of the precision constant ϵ need to correspond to large values of simulation time. We can force the termination time to behave in this way if we inflate the volume $m(C(t))$ slightly. Let $\alpha(t)$ be a strictly positive function that decreases monotonically to 0 as $t \to \infty$ and satisfies $a(t) = o(\phi(t))$ as $t \to \infty$, where ϕ is the scaling function in the CLT (2.2). Then set

$$T_1(\epsilon) \equiv inf\{t \geq 0 : m(C(t))^{1/k} + a(t) < \epsilon\}\,. \tag{2.9}$$

Note that

$$T_1(\epsilon) \geq t_1(\epsilon) \equiv \inf\{t \geq 0 : a(t) < \epsilon\} \to \infty \quad \text{as} \quad \epsilon \downarrow 0\,. \tag{2.10}$$

Thus the early termination associated with $\tilde{T}(\epsilon)$ in (2.8) is prevented by incorporating the deterministic function $a(t)$ in $T_1(\epsilon)$ in (2.9). For practical purposes, it remains to determine appropriate functions $a(t)$, though.

An alternative to the absolute-precision sequential stopping rule in (2.9) is a *relative-precision sequential stopping rule*. The basic idea here is that the simulation should terminate when the k^{th} root of the volume of the confidence region is less than an ϵ^{th} fraction of the norm of the parameter α, denoted by $\|\alpha\|$, under the additional condition that $\|\alpha\| > 0$. Since $Y(t)$ is an estimator for α, this suggests replacing $T_1(\epsilon)$ with

$$T_2(\epsilon) = \inf\{t \geq 0 : m(C(t))^{1/k} + \alpha(t) < \epsilon\|Y(t)\|\}\ . \tag{2.11}$$

The question now is: When are these sequentially stopping rules asymptotically valid? That is, when can we conclude that

$$P(\alpha \in C(T(\epsilon))) \to 1 - \delta \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.12}$$

for $T(\epsilon)$ being $T_1(\epsilon)$ in (2.9) or $T_2(\epsilon)$ in (2.11)?

It turns out that, unlike in Propositions 4.2.1 and 4.2.2, the assumed convergence in (2.2) and (2.6) is *not* enough to achieve asymptotic validility for the sequential stopping rules. That is for the same reason that CLTs involving random time change require extra conditions. However, we do obtain asymptotic validity if we replace the ordinary CLT in (2.2) by a FCLT and if we replace the ordinary WLLN in (2.6) by a SLLN or FWLLN.

(Recall that the SLLN implies a FSLLN by Corollary 3.2.1 in Chapter 3 here, which in turn implies a FWLLN, so that the SLLN is the stronger condition.)

For that purpose, we form scaled processes indexed by ϵ in the function space $D((0,\infty),\mathbb{R}^k)$. We work with time domain $(0,\infty)$ instead of $[0,\infty)$ in order to avoid having to deal with possible singularities in the estimation process Y at the origin $t=0$. For example, such singularities occur in the special case in (2.1). Recall that $x_n \to x$ in $D((0,\infty),\mathbb{R}^d)$ if the restrictions converge in $D([t_0,t_1],\mathbb{R}^d)$ for all t_0, t_1 with $0 < t_0 < t_1 < \infty$.

Given the estimation process Y, the associated scaled estimation processes are

$$\mathbf{Y}_\epsilon(t) \equiv \phi(\epsilon^{-1})[Y(t/\epsilon) - \alpha], \quad t > 0\,. \tag{2.13}$$

For the results below we need to assume that the scaling function ϕ in (2.13) is regularly varying with index γ, denoted by $\phi \in \mathcal{R}(\gamma)$; see Appendix A in the book. We also assume that ϕ is a homeomorphism of $\mathbb{R}^+$, which implies that $\phi(0) = 0$ and ϕ is strictly increasing.

4.2.2. The Absolute-Precision Sequential Estimator

We first state a result for the absolute-precision sequential estimator $T_1(\epsilon)$ in (2.9).

Theorem 4.2.1. *Let $D \equiv D((0,\infty),\mathbb{R}^k)$ be endowed with the WM_2 or any other Skorohod topology. Suppose that*

$$\mathbf{Y}_\epsilon \Rightarrow \Gamma\mathbf{Z} \quad in \quad D \quad as \quad \epsilon \downarrow 0\,, \tag{2.14}$$

for $\mathbf{Y}_\epsilon$ in (2.13), where (2.4) holds with $L = \mathbf{Z}(1)$, $P(t \in Disc(\mathbf{Z})) = 0$ for all t, ϕ is a homeomorphism of $\mathbb{R}_+$, $\phi \in \mathcal{R}(\gamma)$ for $\gamma > 0$, and Γ is nonsingular. If, in addition,

$$\Gamma(t) \to \Gamma \quad w.p.1 \quad in \quad \mathbb{R}^{k^2} \quad as \quad t \to \infty\,, \tag{2.15}$$

then as $t \to \infty$ or as $\epsilon \downarrow 0$

(a) $\phi(t)[m(C(t))^{1/k} + a(t)] \to m(\Gamma A)^{1/k}$ *w.p.1,*

(b) $\epsilon\phi(T_1(\epsilon)) \to m(\Gamma A)^{1/k}$ *w.p.1,*

(c) $\epsilon^{-1}m(C(T_1(\epsilon)))^{1/k} \to 1$ *w.p.1,*

(d) $\epsilon^{-1}[Y(T_1(\epsilon)) - \alpha] \Rightarrow m(\Gamma A)^{-1/k}\Gamma\mathbf{Z}(1) \quad in \quad \mathbb{R}^k$*,*

(e) $P(\alpha \in C(T_1(\epsilon))) \to 1 - \delta$ *(asymptotic validity).*

In our proof of Theorem 4.2.1, we use the following lemma, which shows the scaling implications for the limit process $\mathbf{Z}$ from having the FCLT in (2.14) hold with the regularly varying scaling function ϕ in (2.13). The result is a consequence of Theorem 5.2.1 in the book, but we give a direct proof here.

Lemma 4.2.1. *If the FCLT* (2.14) *holds with* $\phi \in \mathcal{R}(\gamma)$, $\gamma > 0$, *for* ϕ *in* (2.13), *then*

$$\{\mathbf{Z}(ct) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{c^{-\gamma}\mathbf{Z}(t) : t \geq 0\} \tag{2.16}$$

for any $c > 0$.

Proof. Note that $\mathbf{Y}_\epsilon \circ ce \Rightarrow \mathbf{Z} \circ ce$ as $\epsilon \downarrow 0$. On the other hand,

$$\mathbf{Y}_\epsilon \circ c\mathbf{e} = \frac{\phi(\epsilon^{-1})}{\phi(ce^{-1})}\mathbf{Y}_{\epsilon/c} \Rightarrow c^{-\gamma}\mathbf{Z} \quad \text{as} \quad \epsilon \downarrow 0\,, \tag{2.17}$$

using the regular variation to get $\phi(\epsilon^{-1})/\phi(c/\epsilon^{-1}) \to c^{-\gamma}$ as $\epsilon \downarrow 0$ for every $c > 0$; see Appendix A in the book. ∎

Proof of Theorem 4.2.1. (a) Let

$$V(t) \equiv m(C(t))^{1/k} + a(t), \quad t > 0\,. \tag{2.18}$$

By the spatial invariance and scaling properties of Lebesgue measure m on $\mathbb{R}^k$,

$$\begin{aligned} m(C(t))^{1/k} &= m(Y(t) - \phi(t)^{-1}\Gamma(t)A)^{1/k} \\ &= m(-\phi(t)^{-1}\Gamma(t)A)^{1/k} = \phi(t)^{-1}m(\Gamma(t)A)^{1/k}\,. \end{aligned} \tag{2.19}$$

Since A is a bounded set, $\Gamma(t)A$ is contained in a bounded set for all sufficiently large t w.p.1. Thus, we can apply the bounded convergence theorem to deduce that

$$m(\Gamma(t)A)^{1/k} \to m(\Gamma A)^{1/k} \quad \text{w.p.1 as} \quad t \to \infty\,. \tag{2.20}$$

Since Γ is nonsingular, $m(A) > 0$ and $a(t) = o(\phi(t)^{-1})$ as $t \to \infty$, (2.18) and (2.20) imply that

$$\phi(t)V(t) \to m(\Gamma A)^{1/k} > 0 \quad \text{w.p.1 as} \quad t \to \infty\,. \tag{2.21}$$

(b) By the definition of $T_1(\epsilon)$ in (2.9), $V(T_1(\epsilon)-1) \geq \epsilon$ and there exists a random variable $Z(\epsilon)$ with $0 \leq Z(\epsilon) \leq 1$ such that $V(T_1(\epsilon)+Z(\epsilon)) < \epsilon$. (Note that $V(t)$ is not necessarily monotone.) By (2.21) and the fact that $T_1(\epsilon) \to \infty$ w.p.1 as $\epsilon \downarrow 0$,

$$\limsup_{\epsilon \downarrow 0} \epsilon\phi(T_1(\epsilon)) \leq \limsup_{\epsilon \downarrow 0} \phi(T_1(\epsilon))V(T_1(\epsilon)-1) = m(\Gamma A)^{1/k} \quad \text{w.p.1} \tag{2.22}$$

and

$$\liminf_{\epsilon \downarrow 0} \epsilon\phi(T_1(\epsilon)) \geq \liminf_{\epsilon \downarrow 0} \phi(T_1(\epsilon))(V(T_1(\epsilon))+Z(\epsilon)) = m(\Gamma A)^{1/k} \quad \text{w.p.1.} \tag{2.23}$$

(c) Note that

$$m(C(T_1(\epsilon))^{1/k} = \phi(T_1(\epsilon))^{-1} m(\Gamma(T_1(\epsilon))A)^{1/k} \tag{2.24}$$

and recall that $m(\Gamma(t)A) \to m(\Gamma A)$ w.p.1 as $t \to \infty$, so that $m(\Gamma(T_1(\epsilon)) \to m(\Gamma A)$ w.p.1 as $\epsilon \downarrow 0$. By (b), $\epsilon^{-1}\phi(T_1(\epsilon)) \to m(\Gamma A)^{-1/k}$. Hence

$$\begin{aligned} \epsilon^{-1} m(C(T_1(\epsilon)))^{1/k} &= \epsilon^{-1}\phi(T_1(\epsilon))^{-1} m(\Gamma(T_1(\epsilon)A)^{1/k} \\ &\to m(\Gamma A)^{-1/k} m(\Gamma A)^{1/k} = 1 \text{ w.p.1} \quad \text{as} \quad \epsilon \downarrow 0 \end{aligned} \tag{2.25}$$

(d) From the assumed FCLT (2.14), $\mathbf{Z}_\epsilon \Rightarrow \Gamma\mathbf{Z}$ in $D((0,\infty), M_2)$ as $\epsilon \downarrow 0$, where

$$\mathbf{Z}_\epsilon(t) \equiv \mathbf{Y}_{1/\phi^{-1}(\epsilon^{-1})}(t) \equiv \epsilon^{-1}(Y(\phi^{-1}(\epsilon^{-1})t) - \alpha), \quad t > 0\,. \tag{2.26}$$

Now form the deterministic function

$$\boldsymbol{\psi}_\epsilon(t) = \frac{\phi^{-1}(\epsilon^{-1}t)}{\phi^{-1}(\epsilon^{-1})}, \quad t > 0\,. \tag{2.27}$$

Since ϕ is a homeomorphism of $\mathbb{R}_+$, the inverse ϕ^{-1} exists and is itself an homeomorphism of $\mathbb{R}_+$. Moreover, since $\phi \in \mathcal{R}(\gamma)$, $\phi^{-1} \in \mathcal{R}(\gamma^{-1})$ by Theorem 1.5.12 of Bingham, Goldie and Tengels (1989). Hence

$$\boldsymbol{\psi}_\epsilon \to \mathbf{e}^{1/\gamma} \quad \text{in} \quad D \quad \text{as} \quad \epsilon \downarrow 0\,. \tag{2.28}$$

We can apply the continuous-mapping theorem with the composition map taking $D \times D$ into D with (2.26)–(2.28), using Theorem 13.2.3 in the book, to conclude that

$$\mathbf{Z}'_\epsilon \Rightarrow \Gamma\mathbf{Z} \circ \mathbf{e}^{1/\gamma} \text{ in } (D, M_2) \text{ as } \epsilon \downarrow 0\,, \tag{2.29}$$

where

$$\mathbf{Z}'_\epsilon(t) \equiv (\mathbf{Z}_\epsilon \circ \boldsymbol{\psi}_\epsilon)(t) \equiv \epsilon^{-1} Y(\phi^{-1}(\epsilon^{-1}t) - \alpha), \quad t > 0\,. \tag{2.30}$$

Finally, we can apply the continuous-mapping theorem with the composition map taking $D((0,\infty),\mathbb{R}^k) \times \mathbb{R}$ into $\mathbb{R}^k$ with (2.30), invoking Proposition 13.2.1 in the book and part (b) here, to obtain

$$\epsilon^{-1} Y(T_1(\epsilon) - \alpha) = \mathbf{Z}'_\epsilon(\epsilon\phi(T_1(\epsilon)) \Rightarrow \Gamma(\mathbf{Z} \circ \mathbf{e}^{1/\gamma})(m(\Gamma A)^{1/k}) \quad \text{in} \quad \mathbb{R}^k\,, \tag{2.31}$$

where

$$(\mathbf{Z} \circ \mathbf{e}^{1/\gamma}(m(\Gamma A)^{1/k}) = \mathbf{Z}(m(\Gamma A)^{1/\gamma k}) \stackrel{\mathrm{d}}{=} m(\Gamma A)^{-1/k} \mathbf{Z}(1) \tag{2.32}$$

by Lemma 4.2.1.
(e) Note that

$$\begin{aligned} P(\alpha \in C(T_1(\epsilon)) &= P(Y(T_1(\epsilon)) - \alpha \in \phi(T_1(\epsilon))^{-1}\Gamma(T_1(\epsilon))A) \\ &= P(\Gamma(T_1(\epsilon))^{-1}\phi(T_1(\epsilon))[Y_1(T_1(\epsilon)) - \alpha] \in A, \det(\Gamma(T_1(\epsilon)) \neq 0) \\ &+ P(Y(T_1(\epsilon)) - \alpha \in \phi(T_1(\epsilon))^{-1}\Gamma(T_1(\epsilon))A; \det(\Gamma(T_1(\epsilon)) = 0) \end{aligned} \tag{2.33}$$

Since $T_1(\epsilon) \to \infty$ w.p.1 and $\Gamma(t) \to \Gamma$ w.p.1, where Γ is nonsingular, $P(\det(\Gamma(T_1(\epsilon))) = 0) \to 0$ as $\epsilon \downarrow 0$, so that the second term on the right in (2.33) is negligible. On the other hand, for the first term,

$$\begin{aligned} &\Gamma(T_1(\epsilon))^{-1}\phi(T_1(\epsilon))[Y(T_1(\epsilon)) - \alpha] = \Gamma(T_1(\epsilon))^{-1}\epsilon\phi(T_1(\epsilon))\epsilon^{-1}[Y(T_1(\epsilon) - \alpha] \\ &\Rightarrow \Gamma^{-1} m(\Gamma A)^{1/k} m(\Gamma A)^{-1/k} \Gamma \mathbf{Z}(1) = \mathbf{Z}(1) \end{aligned} \tag{2.34}$$

by parts (b) and (d). Hence, combining (2.33) and (2.34), we get

$$P(\alpha \in C(T_1(\epsilon))) \to P(\mathbf{Z}(1) \in A) = 1 - \delta\,, \tag{2.35}$$

because (2.4) holds with $L = \mathbf{Z}(1)$. ■

4.2.3. The Relative-Precision Sequential Estimator

We now state the analogous result for the relative-precision sequential estimator $T_2(\epsilon)$ in (2.11). Note that $T_2(\epsilon)$ behaves asymptotically like $T_1(\|\alpha\|\epsilon)$, as one would expect. In addition to the conditions in Theorem 4.2.2, we require that $Y(t) \to \alpha$ w.p.1 as $t \to \infty$. This is a reasonable condition, but it does not follow form the FCLT (2.14).

Theorem 4.2.2. *In addition to the conditions of Theorem 4.2.1, suppose that*

$$Y(t) \to \alpha \quad w.p.1 \quad in \quad \mathbb{R}^k \quad as \quad t \to \infty\,,$$

where $\|\alpha\| > 0$*. Then as* $t \to \infty$ *and* $\epsilon \to 0$

(a) $\phi(t)[m(C(t))^{1/k} + a(t)]/\|Y(t)\| \to \|\alpha\|^{-1}m(\Gamma A)^{1/k}$ *w.p.1,*

(b) $\epsilon\phi(T_2(\epsilon)) \to \|\alpha\|^{-1}m(\Gamma A)^{1/k}$ *w.p.1,*

(c) $\epsilon^{-1}m(C(T_2(\epsilon))^{1/k} \to \|\alpha\|$ *w.p.1,*

(d) $\epsilon^{-1}[Y(T_2(\epsilon)) - \alpha] \Rightarrow \|\alpha\|m(\Gamma A)^{-1/k}\Gamma Z(1)$ *in* $\mathbb{R}^k$

(e) $P(\alpha \in C(T_2(\epsilon))) \to 1 - \delta$ *(asymptotic validity).*

Since the proof of Theorem 4.2.2 closely parallels the proof of Theorem 4.2.1, we omit the proof of Theorem 4.2.2.

4.2.4. Analogs Based on a FWLLN

There are analogs of Theorems 4.2.1 and 4.2.2, where the SLLN for $\Gamma(t)$ in (2.15) is replaced by the weaker condition of a FWLLN. The w.p.1 limits in parts (a)–(c) of Theorems 4.2.1 and 4.2.2 are then replaced by FWLLNs and the CLT in (d) becomes a FCLT. Since the two results are similar, we only state the analog of Theorem 4.2.1.

Now we also generalize the framework by allowing a family of estimation processes indexed by ϵ. We start with processes $\{Y_\epsilon(t) : t \geq 0\}$ and $\{\Gamma_\epsilon(t) : t \geq 0\}$ for each $\epsilon > 0$. Then instead of (2.7), (2.9) and (2.13), let

$$\begin{aligned} C_\epsilon(t) &\equiv Y_\epsilon(t) - \phi(t)\Gamma_\epsilon(t)A\,, \\ T_{1\epsilon} &\equiv \inf\{t \geq 0 : m(C_\epsilon(t))^{1/k} + a(t) < \epsilon\}\,, \\ \mathbf{Y}_\epsilon(t) &= \phi(\epsilon^{-1})[Y_\epsilon(t/\epsilon) - \alpha], \quad t \geq 0\,. \end{aligned} \tag{2.36}$$

Then the limit will be for the following processes: For that purpose, we define the following random elements of D:

$$\begin{aligned} \mathbf{\Gamma}_\epsilon(t) &\equiv \Gamma_\epsilon(t/\epsilon), \quad t > 0\,, \\ \mathbf{U}^1_\epsilon(t) &\equiv \phi(\epsilon^{-1})m(C_\epsilon(t/\epsilon))^{1/k}\,, \\ \mathbf{U}^2_\epsilon(t) &\equiv \epsilon T_{1\epsilon}(1/t\phi(\epsilon^{-1}))\,, \\ \mathbf{U}^3_\epsilon(t) &\equiv \epsilon^{-1}m(C_\epsilon(T_{1\epsilon}(\epsilon/t)))^{1/k}\,, \\ \mathbf{Z}_\epsilon(t) &\equiv \epsilon^{-1}[Y_\epsilon(T_{1\epsilon}(\epsilon/t)) - \alpha]\,. \end{aligned} \tag{2.37}$$

Theorem 4.2.3. *Let the topology on D be one of WM_2, SM_2, WM_1, SM_1, WJ_1 or SJ_1 throughout. Suppose that the assumptions of Theorem 4.2.1 hold, except that (2.13) is replaced by (2.36) and condition (2.15) is replaced by*

$$\mathbf{\Gamma}_\epsilon \Rightarrow \Gamma\mathbf{1} \quad in \quad D^{k^2} \quad as \quad \epsilon \downarrow 0\,, \tag{2.38}$$

for Γ_ϵ in (2.37) and $\mathbf{1}(t) = (1, \ldots, 1)$ for all $t > 0$. Then

$$(\mathbf{\Gamma}_\epsilon, \mathbf{U}^1_\epsilon, \mathbf{U}^2_\epsilon, \mathbf{U}^3_\epsilon, \mathbf{Z}_\epsilon) \Rightarrow (\Gamma\mathbf{1}, \mathbf{U}^1, \mathbf{U}^2, \mathbf{U}^3, \mathbf{Z}^1) \quad in \quad D^{4+k} \quad as \quad \epsilon \downarrow 0\,, \tag{2.39}$$

for $(\mathbf{U}^1_\epsilon, \mathbf{U}^2_\epsilon, \mathbf{U}^3_\epsilon, \mathbf{Z}_\epsilon)$ in (2.37), where

$$\begin{aligned} \mathbf{U}^1(t) &\equiv t^{-\gamma} m(\Gamma A)^{1/k}, \quad \mathbf{U}^2(t) \equiv t^{1/\gamma}(\Gamma A)^{1/\gamma k} \\ \mathbf{U}^3(t) &= t^{-1} \quad and \quad \mathbf{Z}'(t) \equiv m(\Gamma A)^{-1/k}\Gamma\mathbf{Z}(t^{1/\gamma})\,. \end{aligned} \tag{2.40}$$

Moreover,

$$P(\alpha \in C_\epsilon(T_{1\epsilon}(\epsilon))) \to 1 - \delta \ (asymptotic\ validity)\,. \tag{2.41}$$

In preparation for the proof of Theorem 4.2.3, we prove a lemma.

Lemma 4.2.2. *If $x_i \in D([a,b], \mathbb{R})$ for $i = 1, 2$, where $x_1(t)$, $x_2(t) \geq c > 0$ for all t, then*

$$\|y_1 - y_2\| \leq c^{-2}\|x_1 - x_2\|$$

for $y_i(t) = 1/x_i(t)$, $a \leq t \leq b$.

Proof. Note that

$$|y_1(t) - y_2(t)| = \frac{|x_2(t) - x_1(t)|}{|x_1(t)| \cdot |x_2(t)|} \leq c^{-2}|x_2(t) - x_1(t)|\,. \quad \blacksquare$$

Corollary 4.2.1. *If $x_i \in D([a,b], \mathbb{R})$, $x_i(t) \geq c > 0$, and $y_i(t) = 1/x_i(t)$, $a \leq t \leq b$, $i = 1, 2$ then*

$$d(y_1, y_2) \leq (c^{-2} \vee 1) d(x_1, x_2)$$

where d is one of the J_1, M_1 or M_2 metrics.

Proof. To illustrate, we do the J_1 case:

$$\begin{aligned} d(y_1, y_2) &= \inf_{\lambda \in \Lambda}\{\|y_1 - y_2 \circ \lambda\| \vee \|\lambda - e\|\} \\ &\leq \inf_{\lambda \in \Lambda}\{c^{-2}\|x_1 - x_2 \circ \lambda\| \vee \|\lambda - e\|\} \\ &\leq (c^{-2} \vee 1) d(x_1, x_2)\,. \quad \blacksquare \end{aligned}$$

Proof of Theorem 4.2.3. First since the limit $\Gamma\mathbf{1}$ in (2.38) is deterministic, the two limits in (2.38) and (2.14) hold jointly (where $\mathbf{Y}_\epsilon$ is defined by (2.36) instead of (2.13)), by virtue of Theorem 11.4.5 in the book. Given those limits, we can apply the Skorohod representation theorem, as $\epsilon \downarrow 0$ through an arbitrary sequence, to replace the convergence in distribution by special versions converging w.p.1. Let the special versions be represented by the same notation. Since $\mathbf{1} \in C$, the convergence $\mathbf{\Gamma}_\epsilon \to \Gamma\mathbf{1}$ in $D((0,\infty), \mathbb{R}^{d^2})$ is equivalent to uniform convergence over bounded intervals. Then, as in the proof of Theorem 4.2.1 (a), apply the bounded convergence theorem to get $m(\Gamma_\epsilon(t/\epsilon)A)^{1/k} \to m(\Gamma A)^{1/k}$ w.p.1 uniformly for $t \in [t_0, t_1]$ for any t_0, t_1 with $0 < t_0 < t_1 < \infty$. This yields w.p.1 convergence in $D((0,\infty), \mathbb{R})$ for the special versions. Since $\phi(t/\epsilon)$ a $(t/\epsilon) \to 0$ as $\epsilon \downarrow 0$ uniformly in t for $t > t_0$, we obtain

$$\phi(t/\epsilon)V_\epsilon(t/\epsilon) \to m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.42}$$

uniformly in $[t_0, t_1]$ for the special versions. Since $a(t/\epsilon) = o(\phi(t/\epsilon)^{-1})$, (2.42) implies that

$$\phi(t/\epsilon)m(C_\epsilon(t/\epsilon))^{1/k} \to m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \to 0 \tag{2.43}$$

uniformly in $[t_0, t_1]$. However, since $\phi \in \mathcal{R}(\gamma)$, $\phi(\epsilon^{-1})/\phi(t/\epsilon) \to t^{-\gamma}$ as $\epsilon \downarrow 0$ uniformly on $[t_0, t_1]$, by Theorem A.5 in Appendix A of the book. Thus,

$$\phi(\epsilon^{-1})V_\epsilon(t/\epsilon) \to t^{-\gamma}m(\Gamma A)^{1/k} \tag{2.44}$$

and

$$\phi(\epsilon^{-1})m(C_\epsilon(t/\epsilon))^{1/k} \to t^{-\gamma}m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.45}$$

uniformly in $[t_0, t_1]$, again for the special versions, which implies the FCLT conclusion for $\mathbf{U}^1_\epsilon$ in $D((0,\infty), \mathbb{R})$. Turning to $\mathbf{U}^2_\epsilon$, we will show for the special versions that

$$\begin{aligned} \epsilon T_{1\epsilon}(1/t\phi(\epsilon^{-1})) &= \inf\{s \geq 0 : \phi(\epsilon^{-1})V_\epsilon(s/\epsilon) < t^{-1}\} \\ &= \inf\{s \geq 0 : \phi(\epsilon^{-1})^{-1}V_\epsilon(s/\epsilon)^{-1} > t\} \\ &\to \inf\{s \geq 0 : s^\gamma m(\Gamma A)^{-1/k} > t\} \\ &= t^{1/\gamma}m(\Gamma A)^{1/\gamma k}\ . \end{aligned} \tag{2.46}$$

uniformly in $[t_0, t_1]$. In the first line of (2.46), without loss of generality, we can replace $\phi(\epsilon^{-1})V_\epsilon(s/\epsilon)$ by $\max\{\phi(\epsilon^{-1})V_\epsilon(s/\epsilon),\ (2t_1)^{-1}\}$. Then we can invoke Corollary 4.2.1 above to show that the third line follows from (2.44). For $\mathbf{U}^3_\epsilon$, apply the continuous-mapping theorem with the composition map, using Theorem 13.2.3 in the book and (2.43) and (2.46) here, to get

$$\phi(\epsilon^{-1})m(C_\epsilon(T_{1\epsilon}(1/t\phi(\epsilon^{-1}))))^{1/k} \to t^{-1} \quad \text{as} \quad \epsilon \to 0 \tag{2.47}$$

uniformly in $[t_0, t_1]$ or, equivalently,

$$\epsilon^{-1} m(C_\epsilon(T_{1\epsilon}(\epsilon/t)))^{1/k} \to t^{-1} \quad \text{as} \quad \epsilon \to 0 \tag{2.48}$$

uniformly in $[t_0, t_1]$. Next, for $\mathbf{Z}_\epsilon$, apply the composition map again with the FCLT in (2.14) and the limit for $\mathbf{U}_\epsilon^2$ in (2.48) and part (b) to get

$$\mathbf{Z}_{\phi(\epsilon^{-1})^{-1}} \to \mathbf{Z}' \quad \text{in} \quad D((0,\infty), \mathbb{R}^k) \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.49}$$

where

$$\mathbf{Z}'(t) \equiv \Gamma \mathbf{Z}(t^{1/\gamma} m(\Gamma A)^{1/\gamma k}) \stackrel{\mathrm{d}}{=} m(\Gamma A)^{-1/k} \Gamma \mathbf{Z}(t^{1/\gamma})$$

and the topology is the same as for (2.14). Clearly, $\mathbf{Z}_\epsilon \to \mathbf{Z}'$ in $D((0,\infty), \mathbb{R}^k)$ as well. Finally, for (2.41), apply the projection map for $t = 1$ with the result $\mathbf{Z}_\epsilon \to \mathbf{Z}'$ just established. Then use the argument for Theorem 4.2.1 (e). ∎

4.2.5. Examples

We conclude this section by giving several examples. We illustrate how the theorems can be applied by discussing a few specific estimation settings. These examples show that FCLT requirement for the estimation process Y in (2.14) is a mild hypothesis that is satisfied in virtually all practical contexts. However, some work may be required to establish the SLLN or FWLLN for the estimators $\Gamma(t)$ of the scaling matrix Γ. Our last example shows that we cannot instead use weak consistency of $\Gamma(t)$.

Example 4.2.1. (*Sample mean of IID random variables*). Suppose that α can be represented as $\alpha = EX$ for some real-valued r.v. X. For example, α might correspond to the expected number of customers served in a queue over the time interval $[0, T]$. Then α can be estimated by generating i.i.d. replicates X_1, $X_2, \ldots$ of the r.v. X; the resulting estimator for α is then the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. The corresponding estimation process is $Y(t) = \bar{X}_{\lfloor t \rfloor}$, where $\lfloor t \rfloor$ is the greatest integer less than t and $\bar{X}_0 = 0$. If $EX^2 < \infty$, then Donsker's theorem, Theorem 4.3.2 in the book, asserts that the FCLT in (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$ in (2.13), $\Gamma = \sigma$, where $\sigma^2 = \operatorname{var} X$, and $\mathbf{Z}(t) = \mathbf{B}(t)/t$, where $\mathbf{B}$ is Brownian motion. Note that $\mathbf{Z}(1) =_d N(0,1)$. The typical choice for the set A in this setting is the interval $[-z(\delta),\, z(\delta)]$, where $z(\delta)$ is chosen to satisfy $P(N(0,1) \le z(\delta)\} = 1 - \delta/2$. Of course, it is well known that

$$\Gamma_n \equiv \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]^{1/2} \to \sigma \quad \text{w.p.1 as} \quad n \to \infty\,. \tag{2.50}$$

Suppose that $\sigma^2 > 0$. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we have the strong consistency required by Theorems 4.2.1 and 4.2.2. Hence both the absolute-precision and relative-precision stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid for this example when the precision-constant ϵ shrinks to 0. In this setting, Theorems 4.2.1 and 4.2.2 reproduce the classical results of Chow and Robbins (1965), Starr (1966) and Nadas (1969); see Chapter 7 of Siegmund (1985) and Section 8.8 of Wetherill and Glazebrook (1986). (See Anscombe (1952, 1953) for related earlier work.) Implementation considerations are discussed in Law, Kelton and Koenig (1981).

Example 4.2.2. (*The sample mean of IID random vectors*). Now we consider the case in which α can be represented as $\alpha = EX$, where X is $\mathbb{R}^k$-valued. Assume that $E\|X\|^2 < \infty$. As in Example 4.2.1, we can estimate α via the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, where X_i's are i.i.d. copies of X. Setting $Y(t) = \bar{X}_{\lfloor t \rfloor}$, we obtain the FCLT (2.14) from the k-dimensional version of Donsker's theorem, Theorem 4.3.5 in the book, where now $\mathbf{Z}(t) = \mathbf{B}(t)/t$, $\mathbf{B}$ is k-dimensional standard Brownian motion (composed of k independent one-dimensional standard Brownian motions) and $\Gamma\Gamma^t$ is the covariance matrix C of X. We assume that C is positive definite. Note that $\mathbf{Z}(1) = \mathbf{B}(1) =_d N(0, I)$, where I is the identity matrix. In this k-dimensional setting, we can assume that A is the k-sphere $\{x : \|x\| \leq w(\delta)\}$, where $w(\delta)$ is chosen so that

$$P\{\|N(0,I)\|^2 \leq w^2(\delta)\} = P\{\mathcal{X}_k^2 \leq w^2(\delta)\} = 1 - \delta \ , \tag{2.51}$$

with $\mathcal{X}_k^2$ being a chi-squared r.v. with k degrees of freedom. Let

$$C_n = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^t - \bar{X}_n \bar{X}_n^t \tag{2.52}$$

(writing all k-vectors as column vectors). Then $C_n \to C$ a.s. as $n \to \infty$. Let Γ_n be obtained by taking the Cholesky factorization of C_n, so that Γ_n is a lower triangular matrix such that $C_n = \Gamma_n \Gamma_n^t$; see pages 164 and 165 of Bratley, Fox and Schrage (1987). It follows that $\Gamma_n \to \Gamma$ w.p.1 as $n \to \infty$, since Cholesky factors are continuous at positive definite matrices. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we again have the strong consistency required by Thoerems 4.2.1 and 4.2.2. Thus we have proved that the absolute-precision and relative-precision stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid for sequential stopping of multiple performance measure stochastic simulations. In this setting, Theorems 4.2.1 and 4.2.2 reproduce results by Gleser (1965), Albert (1966) and Srivastava (1967); see Section 5.5 of Govindarajulu (1987).

Example 4.2.3. (*Functions of sample means*). Let X be an $\mathbb{R}^k$-valued random vector and let $\mu = EX$. Suppose that α can be represented as $\alpha = g(\mu)$ for some (known) real-valued function $g : \mathbb{R}^k \to \mathbb{R}$. An example of this occurs in the ratio estimation setting, in which $k = 2$ and $g(x, y) = x/y$. Because the steady state of a regenerative stochastic process can be expressed as a ratio of two means, this estimation setting subsumes that of regenerative steady-state simulation. Of course, this observation lies at the heart of the regenerative method of steady-state simulation; see, for example, Crane and Lemoine (1977).

In this nonlinear setting, we estimate α via $Y(t) = g(\bar{X}_{\lfloor t \rfloor})$, where X_i are i.i.d. random vectors as in Example 4.2.2. Suppose that $E\|X\|^2 < \infty$ and that g is continuously differentiable in a neighborhood of μ. In addition, we require that $\nabla g(\mu) \neq 0$ and that the covariance matrix C of X is positive definite. Then Theorem 3 of Glynn and Whitt (1992b) implies that the FCLT in (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$, $\mathbf{Z}(t) = \mathbf{B}(t)/t$ and $\Gamma = \sigma$ as in Example 4.2.1, but with

$$\sigma = (\nabla g(\mu)^t C \nabla g(\mu))^{1/2} \; .$$

Let C_n be defined as in Example 4.2.2 and note that

$$[\nabla g(Y(t))^t C_{\lfloor t \rfloor} \nabla g(Y(t))]^{1/2} \to \sigma \text{ w.p.1 as } t \to \infty \, .$$

Hence we have the strong consistency required for the application of Thoerems 4.2.1 and 4.2.2. As a consequence, we are assured that the stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are again asymptotically valid in this estimation setting. In particular, in the regenerative simulation setting, we recover the asymptotic theory developed by Lavenberg and Sauer (1977).

Example 4.2.4. (*The jackknife*). Consider the estimation problem of Example 4.2.3 in which our goal is to estimate $\alpha = g(\mu)$, where μ can be expressed as $\mu = EX$ and g is real-valued. One practical difficulty with the estimator suggested in Example 4.2.3 is that it tends to be significantly affected by bias problems induced by the presence of the nonlinearity in g. One way to address the small-sample bias problem that this nonlinearity creates is to jackknife the estimator. Specifically, let $\alpha(n) = g(\bar{X}_n)$ and, for $1 \leq i \leq n$, let

$$\begin{aligned} \bar{X}_{in} &= \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} X_j, \quad \alpha_i(n) = g(\bar{X}_{in}) \; , \\ \tilde{\alpha}_i(n) &= n\alpha(n) - (n-1)\alpha_i(n) \; . \end{aligned} \tag{2.53}$$

Then the estimator $Y_n = n^{-1}\sum_{i=1}^{n}\tilde{\alpha}_i(n)$ is the *jackknife estimator* of α. Let $Y(t) = Y_{\lfloor t\rfloor}$. It is shown in Glynn and Heidelberger (1989) that if $E\|X\|^3 < \infty$ and g is twice continuously differentiable in a neighborhood of μ, then the FCLT in (2.14) holds where σ and $Z(t)$ are as in Example 4.2.3. Since the form of the FCLT is the same as for Example 4.2.3, the jackknife has the same asymptotic efficiency as the estimator of Example 4.2.3. However, as argued in Miller (1964, 1974), the jackknife estimator typically possesses superior small-sample bias properties.

Two estimators for the scaling constant $\sigma = [\nabla g(\mu)^t C \nabla g(\mu)]^{1/2}$ are possible. One approach is to use the estimator $\sigma(t) = [\nabla g(Y(t))^t C_{\lfloor t\rfloor}\nabla g(Y(t))]^{1/2}$ suggested in Example 4.2.3. Theorem 4(i) of Glynn and Heidelberger (1989) shows that $Y(t) \to \alpha$ w.p.1 as $t \to \infty$, under the conditions stated here. Since $C_n \to C$ w.p.1, it follows that $\sigma(t) \to \sigma$ w.p.1 as $t \to \infty$. Hence sequential stopping procedures based on the jackknife point estimator and the "variance" estimator $\sigma^2(t)$ are asymptotically valid by Theorems 4.2.1 and 4.2.2, provided that $\sigma^2 > 0$.

An alternative estimator for the scaling constant σ is given by the jackknife variance estimator $\sigma_J(t)$:

$$\sigma_J(t) = \left(\frac{1}{\lfloor t\rfloor}\sum_{i=1}^{\lfloor t\rfloor}(\tilde{\alpha}_i(\lfloor t\rfloor) - Y(t))^2\right)^{1/2}. \qquad (2.54)$$

Although it is known that $\sigma_J^2(t) \Rightarrow \sigma^2$ as $t \to \infty$ under suitable regularity conditions, we need convergence w.p.1 in order to satisfy the hypothesis of Theorems 4.2.1 and 4.2.3. However, Theorem 4 of Glynn and Whitt (1992a) establishes the following result.

Theorem 4.2.4. *If g is continuously differentiable in a neighborhood of μ and $E|X|^2 < \infty$, then*

$$\sigma_J^2(t) \to \sigma^2 = \nabla g(\mu)^t C\ \nabla g(\mu) \ \textit{w.p.1 as } t \to \infty \qquad (2.55)$$

for $\sigma_J^2(t)$ in (2.54). *Thus the sequential stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ may be applied to jackknife point estimators in conjunction with the jackknifed variance estimator $\sigma_J^2(t)$.*

Example 4.2.5. (*A steady-state mean*). Suppose that our goal is to estimate the steady-state mean vector α of an $\mathbb{R}^k$-valued stochastic process $X = \{X(t) : t \geq 0\}$. We assume that X satisfies an FCLT, namely,

$$\mathbf{X}_\epsilon \Rightarrow \Gamma\mathbf{B} \quad \text{in} \quad D((0,\infty),\mathbb{R}^k) \quad \text{as} \quad \epsilon \downarrow 0 \qquad (2.56)$$

where

$$\mathbf{X}_\epsilon(t) \equiv \epsilon^{-1}\left(\int_0^{t/\epsilon} X(s)ds - t\alpha\right), \quad t > 0\,. \tag{2.57}$$

and $\mathbf{B}$ is a standard $\mathbb{R}^k$-valued Brownian motion. It is easily shown that (2.56) implies that

$$Y(t) \equiv t^{-1}\int_0^t X(s)ds \Rightarrow \alpha \quad \text{as} \quad t \to \infty\,. \tag{2.58}$$

Hence (2.56) implies that the centering vector α appearing there is indeed the steady-state mean of X. Another easy consequence of (2.56) is that the FCLT (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$ and $\mathbf{Z}(t) = \mathbf{B}(t)/t$.

It turns out that (2.56) is typically satisfied for most "real-world" steady-state simulations. In particular, a great variety of different assumptions on the structure of the process X give rise to FCLTs of the form (2.56); see Section 4.4 in the book and Section 2.3 here.

The primary difficulty in applying Theorems 4.2.1–4.2.3 arises in the construction of a process $\Gamma(t)$ such that $\Gamma(t) \to \Gamma$ w.p.1 as $t \to \infty$ or $\mathbf{\Gamma}_\epsilon \Rightarrow \Gamma\mathbf{1}$ in $D(0,\infty)$ as $\epsilon \downarrow 0$. Since $\Gamma\Gamma^t$ is the covariance matrix of the limiting Brownian motion, this is equivalent to the construction of a strongly consistent estimator $C(t)$ for the *time-average covariance matrix* $C = \Gamma\Gamma^t$ of X. In general, this is known to be a challenging problem.

Suppose that X is regenerative, with regeneration times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$. Suppose that $E(\int_{\tau_1}^{\tau_2} |X(s) - \alpha(ds)^2 < \infty$ and that $E(\tau_2 - \tau_1) < \infty$. Let $N(t) = \max\{n \geq 0 : \tau_n \leq t\}$. Then it is easily proved that

$$C(t) = \frac{1}{t}\sum_{i=1}^{N(t)} \int_{\tau_i - 1}^{\tau_i} [X(s) - Y(t)][X(s) - Y(t)]^t ds \tag{2.59}$$

is strongly consistent for C, where $C = \Gamma\Gamma^t$ and Γ is the scaling matrix appearing in (2.56). Thus when X is regenerative, the sequential stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid. Of course, when X is scalar, we already established this result in Example 4.2.3.

For nonregenerative processes, less is known about the strong consistency of estimators $C(t)$ for the steady-state covariance matrix. However, Glynn and Iglehart (1988) and Damerdji (1991, 1994) have recently used strong approximation techniques to establish strong consistency for a broad class of estimators for C. Thus Theorems 4.2.1 and 4.2.2 prove that these estimators do indeed lead to asymptotically valid sequential procedures.

Our theory for this example provides theoretical support complementing previous work by Fishman (1977), Law and Carson (1979) and Law and Kelton (1982). They develop specific empirically based sequential stopping rules for steady-state simulations.

Example 4.2.6. (*Kiefer-Wolfowitz stochastic approximation*). This example is interesting, in part, because it illustrates that the FCLT (2.14) can hold for the estimator with a subcanonical convergence rate; in particular, here $\phi(\epsilon^{-1}) = \epsilon^{-1/3}$. For other examples of noncanonical estimator convergence rates, see Fox and Glynn (1989) and Sections 5 and 6 of Glynn and Whitt (1992b). Suppose that we are given a real-valued smooth function $\beta(\theta)$, which can be represented as $\beta(\theta) = EZ(\theta)$. Assume that our goal is to compute the parameter $\alpha \equiv \theta^*$ minimizing β. If θ is scalar, we can apply the following Kiefer-Wolfowitz stochastic approximation algorithm:

$$\theta_{n+1} = \theta_n - c_n X_{n+1} , \tag{2.60}$$

where $\{c_n : n \geq 0\}$ is a sequence of (deterministic) nonnegative constants,

$$\begin{aligned} P(X_{n+1} \in A|\theta_0, X_0, \ldots, \theta_n, X_n) = \\ P([Z(\theta_0 + h_{n+1}) - Z(\theta_0 - h_{n+1}]/2h_{n+1} \in A) , \end{aligned} \tag{2.61}$$

$Z(\theta_0 + h_{n+1})$ and $Z(\theta_0 - h_{n+1})$ are generated independently of one another and $\{h_n : n \geq 1\}$ is another sequence of deterministic constants. Suppose that $c_n = cn^{-1}$ and $h_n = hn^{-1/3}$, c, $h > 0$. Let $Y(t) = \theta_{\lfloor t \rfloor}$. For this problem, Ruppert (1982) showed that under suitable regularity conditions, the FCLT in (2.14) holds for $\mathbf{Y}_\epsilon$ in (2.13) with $\phi(\epsilon^{-1}) = \epsilon^{-1/3}$, $\Gamma = \kappa$, $\mathbf{Z}(t) = t^{-b}\mathbf{B}(t^{2\eta+1})$, $\mathbf{B}$ is a standard Brownian motion, $b = c\beta''(\theta^*)$, $\eta = b - 5/6$, $\kappa^2 = c^2\sigma^2/(2\eta + 1)(4h^2)$ and $\sigma^2 = 2var\ \mathbf{Z}(\theta^*)$.

The construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ involves more work. For some directions on how to obtain such an estimator, see page 189 of Venter (1967).

Example 4.2.7. (*Robbins-Monro stochastic approximation*). As in Example 4.2.6, suppose that our goal is to estimate the minimizer θ^* of a smooth function β: $\mathbb{R} \to \mathbb{R}$. However, we assume here that we can represent the derivative β' as an expectation; that is, there exists a process $Z(\theta)$ such that $\beta'(\theta) = EZ(\theta)$. [In Example 4.2.6 we assumed only that the function values $\beta(\theta)$ could be represented as expectations.] To calculate θ^* in this setting, we can use the Robbins-Monro stochastic approximation algorithm, which is based on (2.60), where $\{c_n : n \geq 0\}$ is sequence of (deterministic) nonnegative constants and

$$P(X_{n+1} \in A | \theta_0, X_0, \ldots, \theta_0, X_n) = P(Z(\theta_n) \in A) . \tag{2.62}$$

Suppose that our estimator is $Y(t) = \theta_{\lfloor t \rfloor}$ and $c_n = cn^{-1}$ with $c > 0$. Then Ruppert (1982) showed that under suitable regularity hypotheses, the FCLT in (2.14) holds for $\mathbf{Y}_\epsilon$ in (2.13) with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$, $\Gamma = \kappa$, $\mathbf{Z}(t) = t^{-(D+1)}\mathbf{B}(t^{2D+1})$, $D = c\beta'(\theta^*) - 1$, $\kappa^2 = c^2\sigma^2(2D+1)^{-1}$, $\sigma^2 = var\ \mathbf{Z}(\theta^*)$ and $\mathbf{B}$ is a standard Brownian motion.

Construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ follows from results established by Venter (1967). When this estimator is used, the sequential stopping rule $T_1(\epsilon)$ reduces to one studied by McLeish (1976).

Example 4.2.8. (*The Hill estimator*). The framework of Theorems 4.2.1–4.2.3 has been made quite general, so that there can be many applications. One intended application is to estimation problems associated with heavy-tailed probability distributions and long-range dependence. If we use the direct (naive) estimators, e.g., the time average for the steady-state mean, then we anticipate that the FCLT in (2.14) will typically hold with ϕ in (2.13) satisfying $\phi(\epsilon^{-1})/\epsilon^{-1/2} \to 0$ as $\epsilon \downarrow 0$. A common case would by $\phi(\epsilon^{-1}) = \epsilon^{-\gamma}$ or $\phi \in \mathcal{R}(\gamma)$ for $0 < \gamma < 1/2$. A major new difficulty, however, is that now the scaling exponent γ is typically unknown.

Thus, attention naturally shifts to estimating the scaling parameter γ. Estimating the parameter γ is challenging even from observations of i.i.d. random variables. One approach is via the Hill estimator. Recent results of Resnick and Stărică (1997) show that Theorems 4.2.1–4.2.3 can be applied.

The setting is a sequence $\{X_n : n \geq 1\}$ of i.i.d. positive random variables hoaving cdf F, where $F^c \equiv 1 - F \in \mathcal{R}(-\alpha)$ for $\alpha > 0$, i.e.

$$F^c(tx)/F^c(t) \to x^{-\alpha} \quad \text{as} \quad t \to \infty . \tag{2.63}$$

The goal is to estimate the tail index α. For n given, let $X_{(i)}$ be the i^{th} largest among the first n. The Hill estimator based on the k upper order statistics is

$$H_{k,n} = k^{-1} \sum_{i=1}^{k} \log\left(\frac{X_{(i)}}{X_{(k+1)}}\right) . \tag{2.64}$$

The Hill estimator is known to be consistent if $k \equiv k(n)$ satisfies $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$. Given a specific function $k(n)$, the Hill estimator is a single sequence of random variables $\{H_{k(n),n} : n \geq 1\}$. Resnick and

Stărică (1997) show that the Hill estimator also satisfies a FCLT. To state it, let

$$\mathbf{Y}_n(t) \equiv H_{\lceil k_n t \rceil, n}, \quad t \geq 0 \ , \tag{2.65}$$

where $\lceil x \rceil$ is the least integer greater than or equal to x. The FCLT states that, under regularity conditions, including $k(n) \to \infty$ and $k(n)/n \to 0$,

$$k(n)[\mathbf{Y}_n - \alpha^{-1}\mathbf{1}] \Rightarrow \alpha^{-1}\mathbf{Z} \quad \text{in} \quad D \quad \text{as} \quad n \to \infty \ , \tag{2.66}$$

with $\mathbf{Z}(t) = t^{-1}\mathbf{B}(t)$, where $\mathbf{B}$ is standard Brownian motion. Notice that here we use the more general framework in (2.36) in which there is a family of estimation processes indexed by $\epsilon > 0$. (Here we have used $n \to \infty$ instead of $\epsilon \downarrow 0$.)

Also notice that in this special case the scaling matrix Γ in (2.14) is just α^{-1}. So, with $\mathbf{Y}_n$ in (2.65), we estimate α and Γ simultaneously. As a consequence of the FCLT in (2.66), we have the associated FWLLN

$$\mathbf{Y}_n \Rightarrow \alpha^{-1}\mathbf{1} \quad \text{in} \quad D \quad \text{as} \quad n \to \infty \tag{2.67}$$

needed in Theorem 4.2.3. It is also known that $Y_n(t) \to \alpha^{-1}$ w.p.1 as $n \to \infty$ under regularity conditions.

Given the FCLT in (2.66) and the FWLLN in (2.67), the conditions of Theorem 4.2.3 are satisfied. Hence sequential stopping rules are asymptotically valid for the Hill estimator too.

Example 4.2.9. (*Sample mean with infinite variance*). One can also estimate a mean by the sample mean of i.i.d. random variables when the random variables X_i have finite mean but infinite variance. As in Example 4.2.1, the estimation process can be $Y(t) = \bar{X}_{\lfloor t \rfloor}$, where $\bar{X}_0 = 0$, although it is often better to use alternative robust estimators such as trimmed means or to estimate other quantities such as the median. Under regularity conditions, FCLT (2.14) is valid with $\phi \in \mathcal{R}(1 - \alpha^{-1})$ for some α, $1 < \alpha < 2$, where ϕ is the scaling function in (2.13). The topology on D can be the J_1 topology. The limit process $\mathbf{Z}(t)$ is then $t^{-1}\mathbf{S}_\alpha(t)$, where $\{\mathbf{S}_\alpha(t) : t \geq 0\}$ is a stable process of index α, which depends on two parameters in addition to α: a scale parameter σ and a skewness parameter β, $-1 \leq \beta \leq 1$. Unfortunately, in order to form confidence sets we need to estimate the scaling function ϕ and the parameters σ and β.

Suppose that we consider the special case in which X_i is nonnegative and is assumed to have an asymptotic power tail, i.e.

$$F^c(t) \equiv P(X > t) \sim At^{-\alpha} \quad \text{as} \quad t \to \infty \tag{2.68}$$

for positive constants A and α, $1 < \alpha < 2$. Under condition (2.68), the FCLT (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-(1-\alpha^{-1})}$ and limit process $\Gamma Z(t)$ where $Z(t)$ is a stable process with index α, scale $\sigma = 1$ and skewness 1. Hence, in this special case it suffices to estimate only the two parameters α and Γ.

Suppose that $\hat{\alpha}_\epsilon$ is an estimate of α with the property that

$$(\hat{\alpha}_\epsilon^{-1} - \alpha^{-1}) \log(\epsilon^{-1}) \to 0 \quad \text{w.p.1} \quad \text{as} \quad \epsilon \downarrow 0 . \tag{2.69}$$

Given (2.69),

$$\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1}) \equiv \epsilon^{-(1-\hat{\alpha}_\epsilon^{-1})}/\epsilon^{-(1-\alpha)} , \tag{2.70}$$

and

$$\log[\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1})] = (\alpha^{-1} - \hat{\alpha}_\epsilon^{-1}) \log(\epsilon^{-1}) \to 0 \quad \text{w.p.1} \quad \text{as} \quad \epsilon \downarrow 0, \tag{2.71}$$

so that

$$\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1}) \to 1 \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.72}$$

and the FCLT (2.14) holds with the estimator $\hat{\phi}(\epsilon^{-1}) \equiv \epsilon^{-(1-\hat{\alpha}_\epsilon^{-1})}$ used in place of the scaling function $\phi(\epsilon^{-1}) = \epsilon^{-(1-\alpha^{-1})}$. Hence it only remains to estimate the scale parameter Γ. Given that (2.68) holds, the scale parameter is

$$\Gamma = (A/A_\alpha)^{1/\alpha} \tag{2.73}$$

for A in (2.68) and

$$A_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} . \tag{2.74}$$

Hence it suffices to estimate the asymptotic constant A in (2.68). We can estimate in various ways if we estimate the cdf in (2.68) by the empirical cdf.

Hence, under regularity conditions, the sequential stopping rules will again be asymptotically valid. However, in this situation it is often much better to use different (robust) estimators for the mean or to estimate different quantities, such as the median or other percentiles.

Example 4.2.10. (*A counterexample for weak consistency*). Since the SLLN or FWLLN for $\Gamma(t)$ is relatively difficult to establish, it is natural to ask if the weak consistency $\Gamma(t) \Rightarrow \Gamma$ as $t \to \infty$ in (2.6) might not be enough to ensure asymptotic validity of the sequential stopping rules.

Unfortunately, however, weak consistency of $\Gamma(t)$ is not enough. The difficulty is in establishing the in-probability analog of Theorem 4.2.1 (b).

We now give a direct counterexample. Consider Example 4.2.1 and the process $\Gamma(t)$ defined there. Let N be a unit rate Poisson process independent of $\{X_i : i \geq 1\}$ and let $T_1, T_2, \ldots$ be the jump times of the process N. Suppose that

$$\tilde{\Gamma}_{(t)} = \begin{cases} \Gamma(t), & t \notin \cup_{n=1}^{\infty}[T_n, T_n + 1/n), \\ \\ 0, & t \in \cup_{n=1}^{\infty}[T_n, T_n + 1/n). \end{cases} \tag{2.75}$$

Then

$$\begin{aligned} P(\tilde{\Gamma}(t) \neq \Gamma(t)) &= P\left(t \in \left[T_{N(t)}, T_{N(t)} + \frac{1}{N(t)}\right]\right) \\ &\leq P(t - T_{N(t)} \leq \epsilon) + P\left(N(t) \leq \frac{1}{\epsilon}\right) \end{aligned} \tag{2.76}$$

for ϵ arbitrary. Letting $t \to \infty$, we find that $\limsup_{t\to\infty} P(\tilde{\Gamma}(t) \leq \Gamma(t)) = 1 - \exp(-\epsilon)$ (recall that the equilibrium age distribution of N is exponential with mean 1). Since ϵ was arbitrary, it follows that $P(\tilde{\Gamma}(t) \neq \Gamma(t)) \to 0$ as $t \to \infty$. Then it is evident that $\tilde{\Gamma}(t) \Rightarrow \sigma$ as $t \to \infty$, since $\Gamma(t) \to \sigma$ w.p.1 as $t \to \infty$.

Now, in the setting of Example 4.2.1 using $\tilde{\Gamma}(t)$,

$$\tilde{T}_1(\epsilon) = \inf\left\{t \geq 0 : z(\delta)\left(\frac{\Gamma(t)}{\sqrt{t}} + a(t)\right) \leq \epsilon\right\} . \tag{2.77}$$

Put $a(t) = 1/t$. Then clearly $z(\delta)(\tilde{\Gamma}(s)/\sqrt{s} + 1/s) \geq z(\delta)/t$ and $s \leq t$, so $\tilde{T}_1(z(\delta)/t) \geq t$. On the other hand, $\tilde{\Gamma}(T_{N(t)+1}) = 0$, so $\tilde{T}_1(z(\delta)/t) \leq T_{N(t)+1}$. By the SLLN, $t^{-1}T_{N(t)+1} \to 1$ w.p.1 as $t \to \infty$. Hence $\tilde{T}_1(z(\delta)/t) \sim t$ w.p.1 as $t \to \infty$. Thus the stopping rule is asymptotically independent of the scaling constant Γ. As a consequence, formation of asymptotically valid confidence intervals is impossible. In fact, even the asymptotic scaling of the rule is incorrect. It is well known that for estimation problems of the type described in Example 4.2.1, the amount of simulation time required to obtain an absolute precision of order ϵ is of order ϵ^{-2}, whereas the stopping rule $\tilde{T}_1(\epsilon)$ based on $\tilde{\Gamma}(t)$ in (2.75) yields a termination time of order ϵ^{-1}.

Chapter 5

Heavy-Traffic Limits for Queues

5.1. Introduction

In this chapter we include additional material on heavy-traffic limits for queues. The first two sections below supplement Chapter 8 in the book; the final section supplements Chapter 9 in the book.

In particular, Section 5.2 discusses general Lévy approximations for queues, obtained by considering a sequence of queueing models, exploiting the FCLT in Section 2.4 above and the continuous-mapping approach. Then Section 5.3 provides the missing proof to Theorem 8.3.1 in the book. Finally, Section 5.4, drawing upon Puhalskii (1994), shows how heavy-traffic limits for arrival, queue-length and departure processes can be used to establish associated limits for waiting-time and workload processes in single-server queues.

5.2. General Lévy Approximations

The Brownian and stable-Lévy approximations for queues in Chapters 5 and 8 in the book are robust approximations: The same approximation, characterized by only a few parameters, serves as an approximation for a large class of queueing models. We obtain the Brownian (stable-Lévy) approximation with light-tailed (heavy-tailed) distributions.

We can obtain a larger, more flexible, class of approximating processes if we consider stochastic-process limits based on a sequence of queueing models, where the input processes are allowed to change with the sequence index. Of course, we also can obtain the previous limit processes in this more general framework, but we can obtain new limit processes as well, which may be useful for applications.

Closely paralleling the previous sections, we can apply the continuous-mapping approach with the reflection map and a Lévy-process FCLT for double sequences in Theorem 2.4.1 here to obtain a stochastic-process limit for workload processes associated with a sequence of queueing models. When we allow the input processes to change in the limit, we can obtain stochastic-process limits without requiring heavy traffic.

As noted in Section 2.4, we obtain a large class of limit processes from the stochastic-process limits for partial sums from double sequences of random variables, with the variables in each sequence being IID. Indeed, the limit process for the net inputs can be an arbitrary Lévy process $\{L(t) : t \geq 0\}$. Of course, in applications it remains to determine the appropriate Lévy process. Since the Lévy process has stationary and independent increments, it suffices to specify the distribution of the random variable $L(1)$, which must be infinitely divisible. From (4.3) in Section 2.4, it suffices to specify the triple (b, σ^2, μ), where b is the centering constant, σ^2 is the Gaussian coefficient and μ is the Lévy measure. These limiting characteristics can be specified in approximations by exploiting the asymptotic relations in equations (4.10) – (4.12) in Section 2.4.

For applications, it is significant that there is a large class of reflected Lévy processes that are remarkably tractable. In particular, *a reflected Lévy process, constructed from a one-sided reflection, is tractable if the associated Lévy process has no negative jumps.* For example, the steady-state distribution can be characterized by its Laplace transform, which is often called the *generalized Pollaczek-Khintchine transform*, because the Pollaczek-Khintchine transform of the steady-state distribution of the workload process in the M/G/1 queue is a special case.

The original charachterization of the steady-state distribution of a reflected Lévy process for the case with no negative jumps is due to Zolotarev (1964); also see Section 24 of Takács (1967), Bingham (1975) and Kella and Whitt (1992b), especially Section 4(a). The short martingale proof in Kella and Whitt (1992b) is convenient.

When a Lévy process L has no negative jumps, the Lévy measure μ concentrates on $(0, \infty)$ and the bilateral Laplace-Stieltjes transform of $L(1)$

is well defined, with *Laplace exponent*

$$\begin{aligned}\psi(s) &\equiv \log Ee^{-sL(1)} \\ &= -bs + \frac{\sigma^2 s^2}{2} + \int_0^\infty (\exp(-sx) - 1 + sh(x))\mu(dx) \ . \qquad (2.1)\end{aligned}$$

An important special case is a subordinator (totally skewed stable Lévy motion with $\beta = 1$ plus a negative drift, which is just (2.1) without the second Brownian term. Storage models with such Lévy net-input processes are analyzed directly in Chapter 4 of Prabhu (1998). With (2.1), we can conveniently characterize the Laplace transform of the steady-state distribution. The following is a generalization of Theorems 5.8.2 and 8.5.2 in the book.

Theorem 5.2.1. (generalized Pollaczek-Khintchine transform) *Let $\{\phi_K(L)(t) : t \geq 0\}$ be a reflected Lévy process, where ϕ_K is the two-sided reflection map, $EL(1) < 0$, L has no negative jumps and L has Laplace exponent ψ in (2.1).*

(a) If $K = \infty$, then

$$\lim_{t\to\infty} P(\phi_K(L))(t) \leq x) = H(x) \ , \qquad (2.2)$$

where H is a proper cdf with Laplace-Stieltjes transform

$$\hat{h}(s) \equiv \int_0^\infty e^{-sx} dH(x) = \frac{s\psi'(0)}{\psi(s)} \ , \qquad (2.3)$$

and ψ is the Laplace exponent in (2.1).

(b) If $K < \infty$, then

$$\lim_{t\to\infty} P(\phi_K(L))(t) \leq x) = \frac{H(x)}{H(K)}, \quad 0 \leq x \leq K \ , \qquad (2.4)$$

for H in (2.2).

Example 5.2.1. *The special case of the M/G/1 queue.* The workload in unfinished service time in the M/G/1 queue is a reflected Lévy process. If V is a service time and λ is the arrival rate, then the Laplace exponent of the compound-Poisson net-input process is

$$\psi(s) = s - \lambda(1 - E[\exp(-sV)]) \ .$$

Example 5.2.2. *The gamma process.* A possible subordinator is the gamma process, which can be expressed via the Laplace exponent

$$\psi(s) = \int_0^\infty (e^{-sx} - 1)\frac{e^{-x/\eta}}{x} dx = -\log(1 + \eta s)$$

for constant η; e.g., see p. 111 of Prabhu (1998). (The centering function is not needed in this case.) If we add a constant negative drift to the gamma process then we obtain a Lévy process with negative drift but without negative jumps, having Laplace exponent $\psi(s) = bs - \log(1 + \eta s)$. If $b > \eta$, then $EL(1) < 0$ and we can apply Theorem 5.2.1. In this case, the steady-state ccdf H^c is easy to compute from its Laplace transform $H^c(s) = [1 - h(s)]/s$ by numerical inversion. The gamma process is a Lévy process without Brownian component; i.e., $b = \sigma^2 = 0$. The Lévy measure has density $\mu(dx) = x^{-1}e^{-x/\eta}$, $x > 0$. We can approximate the gamma process by a compound Poisson process by restricting μ to $[\epsilon, \infty)$ for some $\epsilon > 0$. ■

For other properties of Lévy processes without negative jumps, see Takács (1967), Samorodnitsky and Taqqu (1994), Bertoin (1996) and Prabhu (1998). For a numerical inversion algorithm to calculate first-passage probabilities, see Rogers (2000).

5.3. A Fluid Queue Fed by On-Off Sources

This section is devoted to proving Theorem 8.3.1 in the book, which establishes a FCLT for the cumulative busy time of a single on-off source.

We first restate the theorem. Recall that $B_{n,i}$ is the i^{th} busy period and $I_{n,i}$ is the i^{th} idle period in the n^{th} model, in the sequence of models under consideration. Let

$$\begin{aligned}
\mathbf{B}_n(t) &\equiv c_n^{-1}\sum_{i=1}^{\lfloor nt\rfloor}(B_{n,i} - m_{B,n}) \\
\mathbf{I}_n(t) &\equiv c_n^{-1}\sum_{i=1}^{\lfloor nt\rfloor}(I_{n,i} - m_{I,n}) \\
\mathbf{N}_n(t) &\equiv c_n^{-1}[N_n(nt) - \gamma_n nt] \\
\mathbf{B}'_n(t) &\equiv c_n^{-1}[B_n(nt) - \xi_n nt], \quad t \geq 0 ,
\end{aligned} \tag{3.1}$$

where again $\lfloor nt\rfloor$ is the integer part of nt,

$$\xi_n \equiv \frac{m_{B,n}}{m_{B,n} + m_{I,n}} \quad \text{and} \quad \gamma_n \equiv \frac{1}{m_{B,n} + m_{I,n}} . \tag{3.2}$$

We think of $m_{B,n}$ in (3.1) as the mean busy period, $EB_{n,i}$, and $m_{I,n}$ as the mean idle period, $EI_{n,i}$, in the case $\{(B_{n,i}, I_{n,i}) : i \geq 1\}$ is a stationary sequence for each n, but in general that is not required.

Theorem 5.3.1. (FCLT for the cumulative busy time) *If*

$$(\mathbf{B}_n, \mathbf{I}_n) \Rightarrow (\mathbf{B}, \mathbf{I}) \quad in \quad (D, M_1)^2 \tag{3.3}$$

for $\mathbf{B}_n$ *and* $\mathbf{I}_n$ *in* (3.1), $c_n \to \infty$, $c_n/n \to 0$, $m_{B,n} \to m_B$, $m_{I,n} \to m_I$, *with* $0 < m_B + m_I < \infty$, *so that* $\xi_n \to \xi$ *with* $0 \leq \xi \leq 1$ *and* $\gamma_n \to \gamma > 0$ *for* ξ_n *and* γ_n *in* (3.2), *and*

$$P(Disc(\mathbf{B}) \cap Disc(\mathbf{I}) = \phi) = 1 \ , \tag{3.4}$$

then

$$(\mathbf{B}_n, \mathbf{I}_n, \mathbf{N}_n, \mathbf{B}'_n) \Rightarrow (\mathbf{B}, \mathbf{I}, \mathbf{N}, \mathbf{B}') \quad in \quad (D, M_1)^4 \ , \tag{3.5}$$

for $\mathbf{N}_n$, $\mathbf{B}'_n$ *in* (3.1) *and*

$$\begin{aligned} \mathbf{N}(t) &\equiv -\gamma[\mathbf{B}(\gamma t) + \mathbf{I}(\gamma t)] \\ \mathbf{B}'(t) &\equiv (1-\xi)\mathbf{B}(\gamma t) - \xi \mathbf{I}(\gamma t) \ . \end{aligned} \tag{3.6}$$

The possibility of the limit processes having discontinuous sample paths makes the required argument more complicated than what it might otherwise be. To make that clear, before presenting an argument that works, we present two false starts.

5.3.1. Two False Starts

For the first false start, note that the cumulative busy-time process can be bounded above and below by random sums by

$$c_n^{-1} \sum_{i=1}^{N_n(nt)} B_{n,i} \leq c_n^{-1} B_n(nt) \leq c_n^{-1} \sum_{i=1}^{N_n(nt)+1} B_{n,i} \ , \tag{3.7}$$

so let us start by trying to find limits for the outer terms in (3.7). We apply the continuous mapping theorem with addition (Section 12.7 in the book) and the inverse map (Sections 13.7 and 13.8 in the book) to get, first, $\mathbf{B}_n + \mathbf{I}_n \Rightarrow \mathbf{B} + \mathbf{I}$ and then $\mathbf{N}_n \Rightarrow \mathbf{N}$ jointly.

As a consequence, we get $\mathbf{T}_n \Rightarrow \gamma e$, where

$$T_n(t) \equiv n^{-1} N_n(nt), \quad t \geq 0 \ . \tag{3.8}$$

Then we try to treat the term on the left in (3.7) by writing

$$\begin{aligned}
& c_n^{-1}\left(\sum_{i=1}^{N_n(nt)} B_{n,i} - m_{n,2}\gamma_n nt\right) \\
= \quad & c_n^{-1}\left(\sum_{i=1}^{\lfloor nt \rfloor} B_{n,i} - m_{n,2}\right) \circ \frac{N_n(nt)}{n} + m_{n,2}(c_n^{-1}[N_n(nt) - \gamma_n nt]) \\
\Rightarrow \quad & \mathbf{B}(\gamma t) - m_2(\gamma[\mathbf{B}(\gamma t) + \mathbf{I}(\gamma t)]) = (1-\xi)\mathbf{B}(\gamma t) - \xi \mathbf{I}(\gamma t) . \qquad (3.9)
\end{aligned}$$

This argument works fine if $P(\mathbf{B} \in C) = 1$, but not otherwise. This argument is not valid here because we need to apply addition when the limit processes $\mathbf{B} \circ \gamma e$ and $-\gamma(\mathbf{B} \circ \gamma e + \mathbf{I} \circ \gamma e)$ typically have common discontinuities of opposite sign. (If they had the same sign, then we could apply Theorem 12.7.3 in the book.) Hence we need to find a different approach.

For our second false start, instead of (3.7), we find different bounds for the cumulative busy-time process, in particular, note that

$$\begin{aligned}
\mathbf{B}_n'(t) \quad \leq \quad & c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)+1} B_{n,i} - \xi_n \sum_{i=1}^{N_n(nt)} I_{n,i}\right] \\
\leq \quad & c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)+1} (B_{n,i} - m_{n,1}) - \xi_n \sum_{i=1}^{N_n(nt)} (I_{n,i} - m_{n,2})\right] \\
& + c_n^{-1} m_{n,1}
\end{aligned}$$

$$\begin{aligned}
\mathbf{B}_n'(t) \quad \geq \quad & c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)} B_{n,i} - \xi_n \sum_{i=1}^{N_n(nt)+1} I_{n,i}\right] \\
\geq \quad & c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)} (B_{n,i} - m_{n,1}) - \xi_n \sum_{i=1}^{N_n(nt)+1} (I_{n,i} - m_{n,2})\right] \\
& - c_n^{-1} m_{n,2} .
\end{aligned}$$

Note that the deterministic terms $c_n^{-1} m_{n,1}$ and $c_n^{-1} m_{n,2}$ are asymptotically negligible. Thus, let the asymptotically bounding processes be

$$\mathbf{B}_n^u(t) \equiv c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)+1} (B_{n,i} - m_{n,1}) - \xi_n \sum_{i=1}^{N_n(nt)} (I_{n,i} - m_{n,2})\right] \qquad (3.10)$$

and

$$\mathbf{B}_n^l(t) \equiv c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n(nt)}(B_{n,i}-m_{n,1}) - \xi_n\sum_{i=1}^{N_n(nt)+1}(I_{n,i}-m_{n,2})\right]. \tag{3.11}$$

Also let

$$\mathbf{T}_n(t) \equiv \frac{N_n(nt)}{n}, \quad \mathbf{T}_n'(t) \equiv \frac{N_n(nt)+1}{t} \tag{3.12}$$

and

$$\mathbf{N}_n'(t) \equiv c_n^{-1}[N_n(nt)+1-\gamma_n nt], \quad t \geq 0\,. \tag{3.13}$$

As before, we apply the continuous mapping theorem with the addition and the inverse map to get, first $\mathbf{B}_n + \mathbf{I}_n \Rightarrow \mathbf{B} + \mathbf{I}$ and then $\mathbf{N}_n \Rightarrow \mathbf{N}$ and $\mathbf{N}_n' \Rightarrow \mathbf{N}$, all jointly. Given $\mathbf{N}_n \Rightarrow \mathbf{N}$ and $\mathbf{N}_n' \Rightarrow \mathbf{N}$ we obtain $\mathbf{T}_n \Rightarrow \gamma e$ and $\mathbf{T}_n' \Rightarrow \gamma e$ by multiplying by c_n/n. Applying the composition map, we obtain

$$\mathbf{B}_n^u = (1-\xi_n)\mathbf{B}_n \circ \mathbf{T}_n' - \xi_n \mathbf{I}_n \circ \mathbf{T}_n \Rightarrow \mathbf{B}' \tag{3.14}$$

and

$$\mathbf{B}_n^l = (1-\xi_n)\mathbf{B}_n \circ \mathbf{T}_n - \xi_n \mathbf{I}_n \circ \mathbf{T}_n' \Rightarrow \mathbf{B}'\,, \tag{3.15}$$

again jointly with the other limits. Hence we are close to obtaining (3.5). However, even though $(\mathbf{B}_n^l, \mathbf{B}_n^u) \Rightarrow (\mathbf{B}', \mathbf{B}')$ and $\mathbf{B}_n^l \leq \mathbf{B}_n' \leq \mathbf{B}_n^u$, we cannot deduce that $\mathbf{B}_n' \Rightarrow \mathbf{B}'$ in (D, M_1).

5.3.2. The Proof

We can deduce that $\mathbf{B}_n' \Rightarrow \mathbf{B}'$ in the weaker Skorohod M_2 topology by this reasoning, though, by virtue of Corollary 12.11.4 in the book, from which we can deduce convergence of the finite-dimensional distributions. To get the desired M_1 limit, it thus suffices to apply Theorem 12.5.1 (iv) in the book and control the oscillations as in equation (12.5.3) of the book. To do so, we introduce a slightly different approximation. Let

$$\mathbf{B}_n^a(t) = c_n^{-1}\left[(1-\xi_n)\sum_{i=1}^{N_n^B(nt)}(B_{n,i}-m_{n1}) - \xi_n\sum_{i=1}^{N_n^I(nt)}(I_{n,i}-m_{n,2})\right], \tag{3.16}$$

where $N_n^B(t)$ and $N_n^I(t) = N_n(t)$ are the number of complete busy periods and idle periods by time t. Reasoning as with (3.14) and (3.15) we can deduce that $\mathbf{B}_n^a \Rightarrow \mathbf{B}'$. However, we can make a stronger connection between $\mathbf{B}_n^a$ and $\mathbf{B}_n$. Note that

$$N_n^I(t) = N_n(t) \le N_n^B(t) \le N_n(t) + 1$$

and

$$\mathbf{B}_n^a \circ \mathbf{S}_n = \mathbf{B}_n \circ \mathbf{S}_n \quad \text{and} \quad \mathbf{B}_n^a \circ \mathbf{S}_n' = \mathbf{B}_n \circ \mathbf{S}_n'$$

where

$$\mathbf{S}_n(t) \equiv n^{-1}\tau_{n,\lfloor nt \rfloor}, \quad \mathbf{S}_n'(t) \equiv n^{-1}\tau'_{n,\lfloor nt \rfloor},$$

$\tau_{n,0} = 0$,

$$\tau_{n,k} \equiv B_{n,1} + I_{n,1} + \cdots + B_{n,k} + I_{n,k}, \quad k \ge 1,$$

and

$$\tau'_{n,k} \equiv \tau_{n,k} + B_{n,k+1}, \quad k \ge 0.$$

Moreover $\mathbf{B}_n^a$ is piecewise-constant and $\mathbf{B}_n$ is piecewise linear in each of the intervals $[n^{-1}\tau_{n,k}, n^{-1}\tau'_{n,k}]$ and $[n^{-1}\tau'_{n,k}, n^{-1}\tau_{n,k+1}]$. Hence we can relate the oscillation of $\mathbf{B}_n$ to those of $\mathbf{B}_n^a$.

First, we can apply the Skorohod representation theorem to replace convergence in distribution by convergence w.p.1. We obtain $\mathbf{B}_n^a \to \mathbf{B}'$ w.p.1 for new versions of these processes. From the specific structure above, we can construct the corresponding special version of $\mathbf{B}_n'$ associated with $\mathbf{B}_n^a$. (It is the piecewise-linear interpolation of the piecewise-constant function.) Since $\mathbf{B}_n^a \to \mathbf{B}'$, $\mathbf{S}_n \to \gamma^{-1}e$ and $\mathbf{S}_n' \to \gamma^{-1}e$ for the new versions, we can deduce that $\mathbf{B}_n'(t) \to \mathbf{B}'(t)$ w.p.1 for each continuity point t of $\mathbf{B}'$. (We also got this part from the convergence of $\mathbf{B}_n^l$ and $\mathbf{B}_n^u$.) Let w_s be the M_1 oscillation function over the interval $[0,T]$, where T is chosen to be a continuity point of $\mathbf{B}'$, i.e.,

$$w_s(x,\delta) \equiv \sup_{0\vee(t-\delta)\le t_1 < t_2 < t_3 \le (t+\delta)\wedge T} \{|x(t_2) - [x(t_1), x(t_3)]|\}$$

where $[x(t_1), x(t_3)]$ is the line segment connecting $x(t_1)$ and $x(t_2)$. From the properties above, we can deduce that

$$w_s(\mathbf{B}_n', \delta) \le w_s(\mathbf{B}_n^a, 2\delta)$$

for all suitably large n. Since $\mathbf{B}_n^a \to \mathbf{B}'$, we deduce that

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; w_s(\mathbf{B}_n^a, \delta) = 0, \tag{3.17}$$

which implies the same limit with $\mathbf{B}_n^a$ replaced by $\mathbf{B}_n'$ in (3.17). By the characterization of M_1 convergence in Theorem 12.5.1 (iv) in the book, we get $\mathbf{B}_n' \to \mathbf{B}'$ w.p.1 (in D, M_1) for the special versions and thus $\mathbf{B}_n' \Rightarrow \mathbf{B}'$ for the original versions. This can be done jointly with the other processes, so that we get (3.5).

5.4. From Queue Lengths to Waiting Times

In this section, following Puhalskii (1994), we show how the continuous-mapping approach with the inverse map and nonlinear centering term, Theorem 13.7.4 in the book, can be used to convert limits for arrival, departure and queue-length processes into associated limits for waiting-time and workload processes in quite general queueing models. The nonlinear centering enables us to capture nonstationary phenomena.

5.4.1. The Setting

The setting is a family of queueing models indexed by n. Suppose that all arrivals eventually get served and then depart, so that the queue length (number of customers in the system) at time t is just the initial queue length plus the arrivals minus the departures, i.e.,

$$Q_n(t) = Q_n(0) + A_n(t) - D_n(t), \quad t \geq 0\,, \tag{4.1}$$

where $Q_n(t)$ is the queue length at time t, $A_n(t)$ is the number of arrivals in the interval $[0, t]$, and $D_n(t)$ is the number of departures in the interval $[0, t]$, all in model n. To treat customer waiting times (but not the workload), we need to make assumptions about the service mechanism. In particular, we assume that the customers are served one at a time in order of their arrival. Thus, we are again in the setting of the standard single-server queue. Let $A_n(t)$ count the new arrivals, and let $D_n(t)$ counts all departures, including those customers originally in the system at time 0. Note that $\{A_n(t) : t \geq 0\}$ and $\{D_n(t) : t \geq 0\}$ are counting processes. As a regularity condition, we assume that $A_n(0) = D_n(0) = 0$.

5.4.2. The Inverse Map with Nonlinear Centering

We can use the inverse map to define related quantities of interest. Let $A_{n,k}$ be the arrival time of the k^{th} arriving customer, $D_{n,k}$ the departure time of the k^{th} arriving customer and $L_n(t)$ the workload facing the server at time

t, not counting arrivals after time t (the virtual waiting time), all in model n. Then

$$
\begin{aligned}
A_{n,k} &\equiv \inf\{s \ge 0 : A_n(s) > (k-1)^+\}\ , \\
D_{n,k} &\equiv \inf\{s \ge 0 : D_n(s) > (Q_n(0) + k - 1)^+\}\ , \\
L_n(t) &\equiv \inf\{s \ge 0 : D_n(s) > Q_n(0) + A_n(t)\}
\end{aligned} \tag{4.2}
$$

for $k \ge 1$ and $t \ge 0$, where $(x)^+ = \max\{x, 0\}$.

Let $W_{n,k}$ be the waiting time for arriving customer k to begin service and let $W'_{n,k}$ be the waiting time until customer k completes service. Then, under the assumptions about the service mechanism above,

$$W_{n,k} \equiv [D_{n,k-1} - A_{n,k}]^+\ , \tag{4.3}$$

and

$$W'_{n,k} \equiv D_{n,k} - A_{n,k}, \quad k \ge 1\ . \tag{4.4}$$

Suppose that the time scaling is already incorporated in the models indexed by n. We assume that functional weak laws of large numbers (FWLLNs) holds with additional space scaling by n and that FCLTs hold with additional space scaling by c_n after centering. Thus, let

$$
\begin{aligned}
\hat{\mathbf{X}}_n(t) &\equiv n^{-1} D_n(t), \\
\hat{\mathbf{Y}}_n(t) &\equiv n^{-1} A_n(t), \\
\hat{\mathbf{Q}}_n(t) &\equiv n^{-1} Q_n(t), \\
\mathbf{X}_n(t) &\equiv c_n(\hat{\mathbf{X}}_n - \mathbf{x}), \\
\mathbf{Y}_n(t) &\equiv c_n(\hat{\mathbf{Y}}_n - \mathbf{y}), \\
\mathbf{Q}_n(t) &\equiv c_n(\hat{\mathbf{Q}}_n - \mathbf{q}), \quad t \ge 0\,.
\end{aligned} \tag{4.5}
$$

We assume that

$$(\hat{\mathbf{X}}_n, \hat{\mathbf{Y}}_n, \hat{\mathbf{Q}}_n) \Rightarrow (\mathbf{x}, \mathbf{y}, \mathbf{q}) \quad \text{in} \quad (D^3, WM_1) \tag{4.6}$$

where $\mathbf{x}, \mathbf{y} \in D_{\uparrow}$, $\mathbf{q} \in D$ and, by (4.1),

$$\mathbf{q}(t) = \mathbf{q}(0) + \mathbf{y}(t) - \mathbf{x}(t), \quad t \ge 0\ . \tag{4.7}$$

We will also impose smoothness conditions on $\mathbf{x}$ and $\mathbf{y}$. In addition, we assume that $c_n \to \infty$ and

$$(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Q}_n) \Rightarrow (\mathbf{X}, \mathbf{Y}, \mathbf{Q}) \quad \text{in} \quad (D^3, WM_1)\,. \tag{4.8}$$

As a consequence of (4.1) and (4.5)–(4.8),

$$\mathbf{Q}(t) - \mathbf{Q}(0) = \mathbf{A}(t) - \mathbf{D}(t) \quad \text{for} \quad t > 0\,. \tag{4.9}$$

Given the FWLLN (4.6) and the FCLT (4.8), we want to establish related limits for appropriately scaled versions of the random variables $A_{n,k}$, $D_{n,k}$, $L_n(t)$, $W_{n,k}$ and $W'_{n,k}$ in (4.2)–(4.4). For that purpose, let

$$\hat{\mathbf{D}}_n(t) \equiv D_{n,\lfloor nt \rfloor}, \quad \hat{\mathbf{A}}_n(t) \equiv A_{n,\lfloor nt \rfloor}, \quad \hat{\mathbf{L}}_n(t) = L_n(t) \tag{4.10}$$

and

$$\hat{\mathbf{W}}_n(t) \equiv W_{n,\lfloor nt \rfloor} \quad \text{and} \quad \hat{\mathbf{W}}'_n(t) \equiv W'_{n,\lfloor nt \rfloor},\ t \geq 0\,. \tag{4.11}$$

We now form the final scaled random elements of D. Let

$$\begin{aligned}
\mathbf{U}_n(t) &\equiv c_n(\hat{\mathbf{X}}_n^{-1} - \mathbf{x}^{-1}),\\
\mathbf{V}_n(t) &\equiv c_n(\hat{\mathbf{Y}}_n^{-1} - \mathbf{y}^{-1}),\\
\mathbf{A}_n(t) &\equiv c_n(\hat{\mathbf{A}}_n - \mathbf{y}^{-1}),\\
\mathbf{D}_n(t) &\equiv c_n(\hat{\mathbf{D}}_n - \mathbf{x}^{-1} \circ \mathbf{z}_1),\\
\mathbf{L}_n(t) &\equiv c_n(\hat{\mathbf{L}}_n - \mathbf{x}^{-1} \circ \mathbf{z}_2),\\
\mathbf{W}_n(t) &\equiv c_n(\hat{\mathbf{W}}_n - (\mathbf{x}^{-1} \circ \mathbf{z}_1 - \mathbf{y}^{-1}),\\
\mathbf{W}'_n(t) &\equiv c_n(\hat{\mathbf{W}}'_n - (\mathbf{x}^{-1} \circ \mathbf{z}_1 - \mathbf{y}^{-1}), \quad t \geq 0\,.
\end{aligned} \tag{4.12}$$

We now state the theorem.

Theorem 5.4.1. (FCLT for the workload and waiting time given a FCLT for arrivals, departures and queue length) *Suppose that the limit* (4.8) *holds for* $\mathbf{X}_n$, $\mathbf{Y}_n$, $\mathbf{Q}_n$ *in* (4.5), *where* $c_n \to \infty$, $\mathbf{x}, \mathbf{y} \in \Lambda$ *and are absolutely continuous with continuous positive derivatives* $\dot{\mathbf{x}}$, $\dot{\mathbf{y}}$, *and* $P(\mathbf{X}(0) = 0) = P(\mathbf{Y}(0) = 0) = 1$. *Then, jointly with* (4.8),

$$(\mathbf{U}_n, \mathbf{V}_n, \mathbf{A}_n, \mathbf{D}_n) \Rightarrow (\mathbf{U}, \mathbf{V}, \mathbf{A}, \mathbf{D}) \tag{4.13}$$

in (D^4, WM_1) *for* $\mathbf{U}_n$, $\mathbf{V}_n$, $\mathbf{A}_n$ *and* $\mathbf{D}_n$ *in* (4.12), *where*

$$\mathbf{U} = \frac{-\mathbf{X} \circ \mathbf{x}^{-1}}{\dot{\mathbf{x}} \circ \mathbf{x}^{-1}}, \quad \mathbf{V} = \mathbf{A} = \frac{-\mathbf{Y} \circ \mathbf{y}^{-1}}{\dot{\mathbf{y}} \circ \mathbf{y}^{-1}} \tag{4.14}$$

and

$$\mathbf{D} = \frac{-\mathbf{X} \circ \mathbf{x}^{-1} \circ \mathbf{z}_1 + \mathbf{Q}(0)\mathbf{1}}{\dot{\mathbf{x}} \circ \mathbf{x}^{-1} \circ \mathbf{z}_1}, \quad \mathbf{z}_1 = \mathbf{q}(0)\mathbf{1} + \mathbf{e}\,, \tag{4.15}$$

where $\mathbf{e}(t) = t$ *for* $t \geq 0$. *If, in addition,*

$$P(Disc(\mathbf{X} \circ \mathbf{x}^{-1} \circ \mathbf{z}_2) \cap Disc(\mathbf{Y}) = \phi) = 1 \tag{4.16}$$

for

$$\mathbf{z}_2 = \mathbf{q}(0)\mathbf{1} + \mathbf{y} \ , \tag{4.17}$$

then, jointly with (4.8) and (4.13),

$$\mathbf{L}_n \Rightarrow \mathbf{L} \tag{4.18}$$

for $\mathbf{L}_n$ *in* (4.12), *where*

$$\mathbf{L} = \frac{-\mathbf{X} \circ \mathbf{x}^{-1} \circ \mathbf{z}_2 + \mathbf{Y} + \mathbf{Q}(0)\mathbf{1}}{\dot{\mathbf{x}} \circ \mathbf{x}^{-1} \circ \mathbf{z}_2} \ . \tag{4.19}$$

If, in addition,

$$P(Disc(\mathbf{A}) \cap Disc(\mathbf{D}) = \phi) = 1 \ , \tag{4.20}$$

then, jointly with (4.8), (4.13) *and* (4.18),

$$(\mathbf{W}_n, \mathbf{W}_n^{'}) \Rightarrow (\mathbf{D} - \mathbf{A}, \mathbf{D} - \mathbf{A}) \tag{4.21}$$

in (D^2, WM_1) *for* $\mathbf{W}_n$ *and* $\mathbf{W}_n'$ *in* (4.12).

In preparation for the proof, we now restate Theorem 13.7.4 from the book. Recall that $D_\uparrow$ is the subset of all nondecreasing nonnegative functions in D. Recall that D_u is the subset of all functions in $D([0,\infty),\mathbb{R})$ that are unbounded above and satisfy $x(0) \geq 0$.

The following is Puhalskii's (1994) result extended to allow discontinuous limits.

Theorem 5.4.2. *Suppose that* $x_n \in D_u$, $y_n \in D_\uparrow$, $c_n \to \infty$,

$$c_n(x_n - x, y_n - y) \to (u, v) \quad in \quad D \times D \tag{4.22}$$

with one of the J_1, M_1 *or* M_2 *topologies, where* $u(0) = 0$, u *has no positive jumps if the topology is* J_1,

$$Disc(u \circ x^{-1} \circ y) \cap Disc(v) = \phi \ , \tag{4.23}$$

$y \in C_{\uparrow\uparrow}$ *and* x *is absolutely continuous with a continuous positive derivative* $\dot{x}$, *then*

$$c_n(x_n^{-1} \circ y_n - x^{-1} \circ y) \to \frac{v - u \circ x^{-1} \circ y}{\dot{x} \circ x^{-1} \circ y} \quad in \quad D \tag{4.24}$$

with the same topology.

Proof of Theorem 5.4.1. We start by applying the Skorohod representation theorem to replace convergence in distribution by convergence w.p.1. For simplicity, we do not introduce new notation for these special versions of the random functions converging w.p.1. Thus consider a single sample path for which the limit (4.8) holds. Now we can apply the deterministic convergence-preservation results. From (4.2)–(4.11), we see that we can represent $\hat{\mathbf{D}}_n$, $\hat{\mathbf{A}}_n$ and $\hat{\mathbf{L}}_n$ in terms of $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ via the inverse map

$$\begin{aligned} \hat{\mathbf{A}}_n(t) &\equiv \inf\{s \geq 0 : A_n(s) > \lfloor nt \rfloor - 1)\} \\ &= \inf\{s \geq 0 : \hat{\mathbf{Y}}_n(s) > (\lfloor nt \rfloor - 1)/n\} \\ &= (\hat{\mathbf{Y}}_n^{-1} \circ \xi_n)(t), \quad t \geq 0, \end{aligned} \tag{4.25}$$

where

$$\xi_n(t) = (\lfloor nt \rfloor - 1)^+/n, \quad t \geq 0, \tag{4.26}$$

$$\begin{aligned} \hat{\mathbf{D}}_n(t) &\equiv \inf\{s \geq 0 : D_n(s) > (Q_n(0) + \lfloor nt \rfloor - 1)^+\} \\ &= \inf\{s \geq 0 : \hat{\mathbf{X}}_n(s) > \{Q_n(0) + \lfloor nt \rfloor - 1)^+/n\} \\ &= (\hat{\mathbf{X}}_n^{-1} \circ \zeta_n)(t), \quad t \geq 0, \end{aligned} \tag{4.27}$$

where

$$\zeta_n(t) = (Q_n(0) + \lfloor nt \rfloor - 1)^+/n, \quad t \geq 0, \tag{4.28}$$

and

$$\begin{aligned} \hat{\mathbf{L}}_n(t) &\equiv \inf\{s \geq 0 : D_n(s) > Q_n(0) + A_n(nt)\} \\ &= \inf\{s \geq 0 : \hat{\mathbf{X}}_n(s) > \hat{\mathbf{Q}}_n(0) + \hat{\mathbf{Y}}_n(t)\} \\ &= [\hat{\mathbf{X}}_n^{-1} \circ (\hat{\mathbf{Q}}_n(0)\mathbf{1} + \hat{\mathbf{Y}}_n)](t), \quad t \geq 0 \end{aligned} \tag{4.29}$$

where $\mathbf{1}(t) \equiv 1$, $t \geq 0$. Given (4.3)–(4.27),

$$\hat{\mathbf{W}}_n(t) = [(\hat{\mathbf{D}}_n \circ \xi_n)(t) - \hat{\mathbf{A}}_n(t)]^+, \quad t \geq 0 \ , \tag{4.30}$$

for ξ_n in (4.26) and

$$\hat{\mathbf{W}}_n'(t) = (\hat{\mathbf{D}}_n - \hat{\mathbf{A}}_n)(t), \quad t \geq 0 \ . \tag{4.31}$$

We now return to the proof of (4.13). First, for the inverse processes $\hat{\mathbf{X}}_n^{-1}$ and $\hat{\mathbf{Y}}_n^{-1}$, we apply Theorem 13.7.2 from the book. Given those two limits, we treat $\hat{\mathbf{A}}_n$ and $\hat{\mathbf{D}}_n$ by applying the composition result, Theorem 12.3.1. Alternatively, we directly apply Theorem 5.4.2 above, noting that $\xi_n \to \mathbf{e}$,

$\zeta_n \to \mathbf{z}_1$, $c_n(\xi_n - e) \to \mathbf{0}$ and $c_n(\zeta_n - z_1) \Rightarrow \hat{\mathbf{Q}}(0)\mathbf{1}$. To treat $\hat{\mathbf{L}}_n$ we again apply Theorem 12.3.1 or Theorem 5.4.2 above, using the fact that $\hat{\mathbf{Q}}_n(0)\mathbf{1} + \hat{\mathbf{Y}}_n \to \mathbf{z}_2$ and $c_n(\hat{\mathbf{Q}}_n(0)\mathbf{1} + \hat{\mathbf{Y}}_n - \mathbf{z}_2) \to \mathbf{Y} + \mathbf{Q}(0)\mathbf{1}$ in D. Finally, to treat $\hat{\mathbf{W}}_n$ and $\hat{\mathbf{W}}_n'$, we use the subtraction map. We first apply subtraction directly to $\hat{\mathbf{W}}_n'$ in (4.31). Since $\xi_n \Rightarrow \mathbf{e}$, we can conclude that $\mathbf{W}_n$ has the same limit as $\mathbf{W'}_n$. ■

Remark 5.4.1. If Theorem 5.4.1 holds for stationary models, then $\mathbf{x} = \mathbf{y} = \lambda\mathbf{e}$, and $\mathbf{q} = \mathbf{q}(0)\underline{1}$. Suppose in addition that $\mathbf{q}(0) = 0$. By (4.9), if we cannot conclude that the limit processes almost surely have continuous paths, then we should anticipate $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Q}$ can have common discontinuities. Then

$$\mathbf{U} = -\lambda^{-1}\mathbf{X} \circ \lambda^{-1}\mathbf{e} \tag{4.32}$$

and

$$\mathbf{V} = \mathbf{A} = -\lambda^{-1}\mathbf{Y} \circ \lambda^{-1}\mathbf{e}\ . \tag{4.33}$$

Condition (4.16) then becomes

$$P(Disc(\mathbf{X}) \cap Disc(\mathbf{Y}) = \phi) = 1 \tag{4.34}$$

and

$$\mathbf{D} = \lambda^{-1}(\mathbf{Y} - \mathbf{X} + \mathbf{Q}(0)\mathbf{1}) = \lambda^{-1}\mathbf{Q}\,. \tag{4.35}$$

Then the centering terms in (4.21) become

$$\mathbf{x}^{-1} \circ \mathbf{z}_1 - \mathbf{y}^{-1} = \lambda^{-1}\mathbf{e} - \lambda^{-1}\mathbf{e} = \mathbf{0} \tag{4.36}$$

and

$$\begin{aligned} \mathbf{D} - \mathbf{A} &= \lambda^{-1}(\mathbf{Y} - \mathbf{X} + \mathbf{Q}(0)\mathbf{1}) + \lambda^{-1} \circ \mathbf{X} \circ \lambda^{-1}\mathbf{e} \\ &= \lambda^{-1}(\mathbf{Q} + \mathbf{X} \circ \lambda^{-1}\mathbf{e})\,. \end{aligned}$$

■

5.4.3. An Application to Central-Server Models

Following Puhalskii (1994), we illustrate how Theorem 5.4.1 can be applied by considering a limit for a central-server model. Central-server models were originally introduced to model the contention among programs for the processor and input-output devices in a multiprogrammed computer system; e.g., see Section 3.4.2 of Lavenberg and Sauer (1983). The specific model we consider is a closed queueing network with $n + 1$ single-server queues, one of which is called the central-server queue while the others are called peripheral queues. There are n customers (jobs) in the network, one for each peripheral queue. Each customer has a designated distinct peripheral queue. Each customer circulates between the central-server queue and its own designated peripheral queue. The customers are served one at a time in order of arrival at the central-server queue. The service times are assumed to be mutually independent exponential random variables. (That ensures that the closed network has a product-form steady-state distribution.) Let the mean service time at each peripheral queue be λ^{-1}, and let the mean service time at the central-server queue be $(n\mu)^{-1}$.

Since only one customer receives service at each peripheral queue, there is no contention there. Thus, each customer enters service at its peripheral immediately upon arrival. Consequently, the $(n+1)$-queue model is equivalent to a 2-queue model, with one queue being the central-server queue and the other queue being an infinite-server queue. Moreover, the number of customers at the central-server queue evolves as a birth-and-death process with state-dependent transition rates. Let $Q_n(t)$ denote the number of customers at the central-server queue at time t, as a function of n. When $Q_n(t) = k$, the birth (arrival) rate is $(n-k)\lambda$ and the death (service) rate is $n\mu$. Hence the steady-state distribution is easy to calculate.

However, it is also of interest to consider limits as $n \to \infty$ in order to better understand the behavior of such systems with fast central servers and many customers. First a FLLN is quite elementary. For that purpose, let $A_n(t)$ and $D_n(t)$ count the numbers of arrivals and departures, respectively, at the central-server queue in the interval $[0,t]$. Then form the scaled processes $\hat{\mathbf{X}}_n$, $\hat{\mathbf{Y}}_n$ and $\hat{\mathbf{Q}}_n$ as in (4.5). It is then relatively elementary to show that, if $\hat{\mathbf{Q}}_n(0) = \mathbf{q}(0)$, $0 \le q(0) \le 1$, then the FWLLN in (4.6) holds here with

$$\mathbf{x}(t) = \mu t, \quad \mathbf{y}(t) = \lambda \int_0^t [1 - \mathbf{q}(s)] ds \tag{4.37}$$

and $\mathbf{q}$ satisfying the ordinary differential equation

$$\dot{\mathbf{q}}(t) \equiv \frac{d\mathbf{q}}{dt}(t) = \lambda(1 - \mathbf{q}(t)) - \mu \ . \tag{4.38}$$

Kogan, Lipster and Smorodinskii (1986) then established the following result; also see Chapter 8, Section 3, of Liptser and Shiryaev (1989) and Puhalskii (1994).

Theorem 5.4.3. (FCLT for the central-server model) *If*

$$\sqrt{n}[\mathbf{Q}_n(0) - q(0)] \Rightarrow \mathbf{Q}(0) \quad in \quad \mathbb{R}, \tag{4.39}$$

then the joint limit (4.8) *holds with*

$$\mathbf{X}(t) = \sqrt{\mu}\mathbf{B}_2(t) \ , \tag{4.40}$$

$$\mathbf{Y}(t) = \int_0^t \sqrt{\lambda(1-\mathbf{q}(s)}d\mathbf{B}_1(s) - \lambda\int_0^t \mathbf{Q}(s)ds \tag{4.41}$$

and

$$\mathbf{Q}(t) = \mathbf{Q}(0) + \mathbf{X}(t) - \mathbf{Y}(t), \quad t \geq 0 \ . \tag{4.42}$$

The limit process $\mathbf{Q}$ *can be expressed as the solution to*

$$\begin{aligned}\mathbf{Q}(t) = \mathbf{Q}(0) \quad &- \quad \lambda\int_0^t \mathbf{Q}(s)ds \\ &+ \quad \int_0^t \sqrt{\lambda(1-\mathbf{q}(s))}d\mathbf{B}_1(s) - \sqrt{\mu}\mathbf{B}_2(t) \, . \end{aligned} \tag{4.43}$$

We can now combine Theorems 5.4.1 and 5.4.3 to obtain associated limits for the scaled versions of $\hat{\mathbf{A}}_n$, $\hat{\mathbf{D}}_n$, $\hat{\mathbf{L}}_n$ in (4.10) and $\hat{\mathbf{W}}_n$ and $\hat{\mathbf{W}}'_n$ in (4.11), as stated in Theorem 5.4.1. Theorem 5.4.1 is genuinely helpful here, because these limits are not so easy to obtain directly.

Theorem 5.4.1 has also been applied by Mandelbaum, Massey, Reiman and Stolyar (1999).

Chapter 6

The Space D

6.1. Introduction

This chapter contains proofs omitted from Chapter 12 of the book, with the same title. For convenience, the theorems are restated here. The section and theorem numbers parallel Chapter 12 of the book, so the proofs should be easy to find.

Here is how the present chapter is organized: We start in Section 6.2 by discussing regularity properties of the function space D. A key property, which we frequently use, is the fact that any function in D can be approximated uniformly closely by piecewise-constant functions with only finitely many discontinuities.

In Section 6.3 we introduce the strong and weak versions of the M_1 topology on $D([0,T],\mathbb{R}^k)$, referred to as SM_1 and WM_1, and establish basic properties. We also discuss the relation among the nonuniform Skorohod topologies on D. In Section 6.4 we discuss local uniform convergence at continuity points and relate it to oscillation functions used to characterize different forms of convergence.

In Section 6.5 we provide several different alternative characterizations of SM_1 and WM_1 convergence. Some involve parametric representations of the completed graphs and others involve oscillation functions. It is significant that there are forms of the oscillation-function characterizations that involve considering one function argument t at a time. Consequently, the examples in Figure 11.2 of the book tend to be more than illustrative: The topologies are characterized by the local behavior in the neighborhood of single discontinuities.

In Section 6.6 we discuss conditions that allow us to strengthen the mode of convergence from WM_1 to SM_1. The key condition is to have the coordinate limit functions have no common discontinuities. In Section 6.7 we study how SM_1 convergence in $D([0,T],\mathbb{R}^k)$ can be characterized by associated limits of mappings.

In Section 6.8 we exhibit a complete metric topologically equivalent to the incomplete metric inducing the SM_1 topology introduced earlier. As with the J_1 metric d_{J_1} in equation (3.2) of Section 3.3 in the book, the natural M_1 metric is incomplete, but there exists a topologically equivalent complete metric, so that D with the SM_1 topology is Polish (metrizable as a complete separable metric space).

In Section 6.9 we discuss extensions of the SM_1 and WM_1 topologies on $D([0,T],\mathbb{R}^K)$ to corresponding spaces of functions with noncompact domains. The principal example of such a noncompact domain is the interval $[0,\infty)$, but $(0,\infty)$ and $(-\infty,\infty)$ also arise.

In Section 6.10 we introduce the strong and weak versions of the M_2 topology, denoted by SM_2 and WM_2. In Section 6.11 we provide alternative characterizations of these topologies and discuss additional properties.

Finally, in Section 6.12 we discuss characterizations of compact subsets of D using oscillation functions. These characterizations are useful because they lead to characterizations of tightness for sequences of probability measures on D, which is a principal way to establish weak convergence of the probability measures; see Section 11.6 of the book.

6.2. Regularity Properties of D

Recall that $D \equiv D^k \equiv D([0,T],\mathbb{R}^k)$ is the set of all $\mathbb{R}^k$-valued functions $x \equiv (x^1,\ldots,x^k)$ on $[0,T]$ that are right continuous at all $t \in [0,T)$ and have left limits at all $t \in (0,T]$:

We use superscripts to designate coordinate functions, so that subscripts can index different functions in D. For example, x_3^2 denotes the second coordinate function in $D([0,T],\mathbb{R}^1)$ of $x_3 \equiv (x_3^1,\ldots,x_3^k)$ in $D([0,T],\mathbb{R}^k)$, where x_3 is the third element of the sequence $\{x_n : n \geq 1\}$. Let C be the subset of continuous functions in D.

Let $\|\cdot\|$ be the maximum (or l_∞) norm on $\mathbb{R}^k$ and the *uniform norm* on D; i.e., for each $b \equiv (b^1,\ldots,b^k) \in \mathbb{R}^k$, let

$$\|b\| \equiv \max_{1 \leq i \leq k} |b^i| \tag{2.1}$$

and, for each $x \equiv (x^1,\ldots,x^k) \in D([0,T],\mathbb{R}^k)$, let

$$\|x\| \equiv \sup_{0 \leq t \leq T} \|x(t)\| = \sup_{0 \leq t \leq T} \max_{1 \leq i \leq k} |x^i(t)| \ . \tag{2.2}$$

The maximum norm on $\mathbb{R}^k$ in (2.1) is topologically equivalent to the l_p norm

$$\|b\|_p \equiv \left(\sum_{i=1}^{k}(b^i)^p\right)^{1/p} .$$

For $p = 2$, the l_p norm is the Euclidean (or l_2) norm. For $p = 1$, the l_p norm is the sum (or l_1) norm. The uniform norm on D induces the uniform metric on D.

We first discuss regularity properties of D due to the existence of limits. Let $Disc(x)$ be the set of discontinuities of x, i.e.,

$$Disc(x) \equiv \{t \in (0,T] : x(t-) \neq x(t)\} \tag{2.3}$$

and let $Disc(x,\epsilon)$ be the set of discontinuities of magnitude at least ϵ, i.e.,

$$Disc(x,\epsilon) \equiv \{t \in (0,T] : \|x(t-) - x(t)\| \geq \epsilon\} . \tag{2.4}$$

The following is a key regularity property of D.

Theorem 6.2.1. (the number of discontinuities of a given size) *For each $x \in D$ and $\epsilon > 0$, $Disc(x,\epsilon)$ is a finite subset of* $[0,T]$.

Proof. We will show that $Disc(x,\epsilon)$ being infinite contradicts the existence of limits from the left and right. If $Disc(x,\epsilon)$ were infinite, then there would exist $t \in [0,T]$ and a sequence $\{t_n : n \geq 1\}$ with $t_n \in Disc(x,\epsilon)$ for all n and $t_n \downarrow t$ or $t_n \uparrow t$ as $n \to \infty$. Suppose that $t_n \downarrow t$; the other case is treated in the same way. Since $t_n \in Disc(x,\epsilon)$, we must have $\|x(t_n-) - x(t_n)\| \geq \epsilon$ for all n. Hence, there must exist another sequence $\{t'_n : n \geq 1\}$ such that $t_n > t'_n > t_{n+1} > t'_{n+1} > t$ for all n and $\|x(t_n) - x(t'_n)\| > \epsilon/2$ for all n. However, that contradicts the existence of limits from the right at t. ■

Corollary 6.2.1. (the number of discontinuities) *For each $x \in D$, $Disc(x)$ is either finite or countably infinite.*

Proof. Note that

$$Disc(x) = \bigcup_{n=1}^{\infty} Disc(x,n^{-1}) . \quad ■$$

We say that a function x in D is *piecewise-constant* if there are finitely many time points t_i such that $0 \equiv t_0 < t_1 < \cdots < t_{m-1} \leq t_m \equiv T$ and x is constant on the intervals $[t_{i-1}, t_i)$, $1 \le i \le m-1$, and $[t_{m-1}, T]$. Let D_c be the subset of piecewise-constant functions in D. Let $v(x; A)$ be the *modulus of continuity* of the function x over the set A, defined by

$$v(x; A) \equiv \sup_{t_1, t_2 \in A} \{\|x(t_1) - x(t_2)\|\} \tag{2.5}$$

for $A \subseteq [0, T]$. The following is a second important regularity property of D.

Theorem 6.2.2. (approximation by piecewise-constant functions) *For each $x \in D$ and $\epsilon > 0$, there exists $x_c \in D_c$ such that $\|x - x_c\| < \epsilon$.*

Proof. We show how to construct x_c. Given x and ϵ, construct the subset $Disc(x, \epsilon)$, which is finite by Theorem 6.2.1. Due to the existence of limits, for each $t \in Disc(x, \epsilon)$ we can find $t_1 \equiv t_1(t)$ and $t_2 \equiv t_2(t)$ such that $t_1 < t < t_2$, $v(x, [t_1, t)) < \epsilon$, $v(x, [t, t_2]) < \epsilon$,

$$Disc(x, \epsilon) \cap [t_1, t) = \phi \quad \text{and} \quad Disc(x, \epsilon) \cap (t, t_2] = \phi\,.$$

For each $t \in Disc(x, \epsilon)$, let these points t, $t_1(t)$ and $t_2(t)$ all belong to $Disc(x_c)$; let $x_c(t') = x(t-)$ for $t' \in (t_1, t)$ and let $x_c(t') = x(t)$ for $t \in [t, t_2)$. Now let

$$A \equiv [0, T] - \bigcup_{t \in Disc(x, \epsilon)} (t_1(t), t_2(t))\ .$$

The set A is a finite union of closed intervals. Consider any one of these intervals, say $[a, b]$. If $v(x; [a, b]) < \epsilon$, then it suffices to let $x_c(t) = x(t)$ for any $t \in [a, b]$, and not add any points to $Disc(x_c)$. Suppose that $v(x; [a, b]) \ge \epsilon$. For each $t \in [a, b]$, since $t \in Disc(x, \epsilon)^c$, it is possible to find an interval $(t_1(t), t_2(t))$, $[a, t_2(t))$ or $(t_1(t), b]$ containing t such that $v(x, (t_1(t), t_2(t)) < \epsilon$. (The intervals $[a, t)$ and $(t, b]$ are open in the relative topology on $[a, b]$. Thus the collection of all these subintervals form an open cover of $[a, b]$.) Since $[a, b]$ is compact, there is a finite collection of these intervals covering $[a, b]$; i.e., there are points

$$a < t'_1 < t_1 < \cdots < t'_m < t_m < b$$

for $m \ge 1$ such that $[a, t_1)$, (t'_1, t_2), $(t'_2, t_3), \ldots$, (t'_{m-1}, t_m), $(t'_m, b]$ are in the finite collection. Necessarily, $t'_i < t_i$ for all i. It suffices to choose $t''_i \in (t'_i, t_i)$ for each i, $1 \le i \le m$, and let $t''_i \in Disc(x_c)$. We can let $x_c(t''_i) = x(t''_i)$ for each such t''_i. We have thus constructed $x_c \in D_c$ with $\|x - x_c\| < \epsilon$. ∎

6.3. Strong and Weak M_1 Topologies

6.3.1. Definitions

We start by making some definitions, repeating what is in the book. The strong and weak topologies will be based on different notions of a segment in $\mathbb{R}^k$. For $a \equiv (a^1, \dots, a^k)$, $b \equiv (b^1, \dots, b^k) \in \mathbb{R}^k$, let $[a, b]$ be the *standard segment*, i.e.,

$$[a, b] \equiv \{\alpha a + (1 - \alpha) b : 0 \le \alpha \le 1\} \tag{3.1}$$

and let $[[a, b]]$ be the *product segment*, i.e.,

$$[[a, b]] \equiv \bigtimes_{i=1}^{k} [a^i, b^i] \equiv [a^1, b^1] \times \cdots \times [a^k, b^k] \ , \tag{3.2}$$

where the one-dimensional segment $[a^i, b^i]$ coincides with the closed interval $[a^i \wedge b^i, a^i \vee b^i]$, with $c \wedge d = \min\{c, d\}$ and $c \vee d = \max\{c, d\}$ for $c, d \in \mathbb{R}$. Note that $[a, b]$ and $[[a, b]]$ are both subsets of $\mathbb{R}^k$. If $a = b$, then $[a, b] = [[a, b]] = \{a\} = \{b\}$; if $a^i \neq b^i$ for one and only one i, then $[a, b] = [[a, b]]$. If $a \neq b$, then $[a, b]$ is always a one-dimensional line in $\mathbb{R}^k$, while $[[a, b]]$ is a j-dimensional subset, where j is the number of coordinates i for which $a^i \neq b^i$. Always, $[a, b] \subseteq [[a, b]]$.

We now define completed graphs of the functions: For $x \in D$, let the (standard) *thin graph* of x be

$$\Gamma_x \equiv \{(z, t) \in \mathbb{R}^k \times [0, T] : z \in [x(t-), x(t)]\} \, , \tag{3.3}$$

where $x(0-) \equiv x(0)$ and let the *thick graph* of x be

$$\begin{aligned} G_x &\equiv \{(z, t) \in \mathbb{R}^k \times [0, T] : z \in [[x(t-), x(t)]]\} \\ &= \{(z, t) \in \mathbb{R}^k \times [0, T] : z^i \in [x^i(t-), x^i(t)] \quad \text{for each } i\} \end{aligned} \tag{3.4}$$

for $1 \le i \le k$. Since $[a, b] \subseteq [[a, b]]$ for all $a, b \in \mathbb{R}^k$, $\Gamma_x \subseteq G_x$ for each x.

We now define *order relations* on the graphs Γ_x and G_x. We say that $(z_1, t_1) \le (z_2, t_2)$ if either (i) $t_1 < t_2$ or (ii) $t_1 = t_2$ and $|x^i(t_1-) - z_1^i| \le |x^i(t_1-) - z_2^i|$ for all i. The relation $\le$ induces a total order on Γ_x and a partial order on G_x.

It is also convenient to look at the ranges of the functions. Let the *thin range* of x be the projection of Γ_x onto $\mathbb{R}^k$, i.e.,

$$\rho(\Gamma_x) \equiv \{z \in \mathbb{R}^k : (z, t) \in \Gamma_x \quad \text{for some} \quad t \in [0, T]\} \tag{3.5}$$

and let the *thick range* of x be the projection of G_x onto $\mathbb{R}^k$, i.e.,

$$\rho(G_x) \equiv \{z \in \mathbb{R}^k : (z,t) \in G_x \quad \text{for some} \quad t \in [0,T]\} \ . \tag{3.6}$$

Note that $(z,t) \in \Gamma_x$ (G_x) for some t if and only if $z \in \rho(\Gamma_x)$ $(\rho(G_x))$. Thus a pair (z,t) cannot be in a graph of x if z is not in the corresponding range.

We now define strong (standard) and weak parametric representations based on these two kinds of graphs. A *strong parametric representation* of x is a continuous nondecreasing function (u,r) mapping $[0,1]$ onto Γ_x. A *weak parametric representation* of x is a continuous nondecreasing function (u,r) mapping $[0,1]$ into G_x such that $r(0)=0$, $r(1)=T$ and $u(1)=x(T)$. (For the parametric representation, "nondecreasing" is with respect to the usual order on the domain $[0,1]$ and the order on the graphs defined above.) Here it is understood that $u \equiv (u^1,\ldots,u^k) \in C([0,1],\mathbb{R}^k)$ is the spatial part of the parametric representation, while $r \in C([0,1],[0,T])$ is the time (domain) part. Let $\Pi_s(x)$ and $\Pi_w(x)$ be the sets of strong and weak parametric representations of x, respectively. For real-valued functions x, let $\Pi(x) \equiv \Pi_s(x) = \Pi_w(x)$. Note that $(u,r) \in \Pi_w(x)$ if and only if $(u^i,r) \in \Pi(x^i)$ for $1 \le i \le k$.

We use the parametric representations to characterize the strong and weak M_1 topologies. As in (2.1) and (2.2), let $\|\cdot\|$ denote the supremum norms in $\mathbb{R}^k$ and D. We use the definition $\|\cdot\|$ in (2.2) also for the $\mathbb{R}^k$-valued functions u and r on $[0,1]$.

Now, for any $x_1, x_2 \in D$, let

$$d_s(x_1,x_2) \equiv \inf_{\substack{(u_j,r_j)\in\Pi_s(x_j)\\ j=1,2}} \{\|u_1-u_2\| \vee \|r_1-r_2\|\} \tag{3.7}$$

and

$$d_w(x_1,x_2) \equiv \inf_{\substack{(u_j,r_j)\in\Pi_w(x_j)\\ j=1,2}} \{\|u_1-u_2\| \vee \|r_1-r_2\|\} \ . \tag{3.8}$$

Note that $\|u_1-u_2\| \vee \|r_1-r_2\|$ can also be written as $\|(u_1,r_1)-(u_2,r_2)\|$, due to definitions (2.1) and (2.2). Of course, when the range is $\mathbb{R}$, $d_s = d_w = d_{M_1}$ for d_{M_1} defined in equation (3.4) in Section 3.3 of the book.

We say that $x_n \to x$ in D for a sequence or net $\{x_n\}$ in the SM_1 (WM_1) topology if $d_s(x_n,x) \to 0$ $(d_w(x_n,x) \to 0)$ as $n \to \infty$. We start with the following basic result.

6.3.2. Metric Properties

Theorem 6.3.1. (metric inducing SM_1) *d_s is a metric on D.*

Proof. Only the triangle inequality is difficult. By Lemma 6.3.2 below, for any $\epsilon > 0$, a common parametric representation $(u_3, r_3) \in \Pi_s(x_3)$ can be used to obtain

$$\|u_1 - u_3\| \vee \|r_1 - r_3\| < d_s(x_1, x_3) + \epsilon$$

and

$$\|u_2 - u_3\| \vee \|r_2 - r_3\| < d_s(x_1, x_3) + \epsilon$$

for some $(u_1, r_1) \in \Pi_s(x_1)$ and $(u_2, r_2) \in \Pi_s(x_2)$. Hence

$$d_s(x_1, x_2) \leq \|u_1 - u_2\| \vee \|r_1 - r_2\| \leq d_s(x_1, x_3) + d_s(x_3, x_2) + 2\epsilon \ .$$

Since ϵ was arbitrary, the proof is complete. ■

To prove Theorem 6.3.1, we use finite approximations to the graphs Γ_x. We first define an order-consistent distance between a graph and a finite subset. We use the notion of a finite ordered subset.

Definition 6.3.1. (order-consistent distance) *For $x \in D$, let A be a finite ordered subset of the ordered graph $(\Gamma_x, \leq)$, i.e., for some $m \geq 1$, A contains $m + 1$ points (z_i, t_i) from Γ_x such that*

$$(x(0), 0) \equiv (z_0, t_0) \leq (z_1, t_1) \leq \cdots \leq (z_m, t_m) \equiv (x(T), T) \ . \qquad (3.9)$$

The order-consistent distance between A and Γ_x is

$$\hat{d}(A, \Gamma_x) \equiv \sup\{\|(z, t) - (z_i, t_i)\| \vee \|(z, t) - (z_{i+1}, t_{i+1})\|\} \ , \qquad (3.10)$$

where the supremum is over all $(z_i, t_i) \in A$, $1 \leq i \leq m-1$, and all $(z, t) \in \Gamma_x$ such that

$$(z_i, t_i) \leq (z, t) < (z_{i+1}, t_{i+1}) \ ,$$

using the order on the graph. ■

We now show that finite ordered subsets A can be chosen to make $\hat{d}(A, \Gamma_x)$ arbitrarily small.

Lemma 6.3.1. (finite approximations to graphs) *For any $x \in D$ and $\epsilon > 0$, there exists a finite ordered subset A of Γ_x such that $\hat{d}(A, \Gamma_x) < \epsilon$ for $\hat{d}$ in* (3.10).

Proof. First put finitely many points $(x(t_i), t_i)$ in A to meet the requirement on the domain $[0,T]$, i.e., to have $0 = t_1 < t_2 < \cdots < t_m = T$ with $t_{i+1} - t_i < \epsilon$. We add additional points to account for the spatial component. For each $t \in Disc(x, \epsilon)$, choose the points $(x(t-), t)$, $(x(t), t)$ and finitely many points on the segment $[(x(t-), t), (x(t), t)]$ such that the distance between successive points is less than ϵ. Since x has left and right limits everywhere, there are open neighborhoods (t_1, t) and (t, t_2) of each $t \in Disc(x, \epsilon)$ such that

$$\sup\{\|x(t') - x(t'')\| : t_1 < t' < t'' < t\} < \epsilon$$

and

$$\sup\{\|x(t') - x(t'')\| : t < t' < t'' < t_2\} < \epsilon \ .$$

We thus can choose one more point, if needed, in each of the sets $\Gamma_x \cap [R^k \times (t_1, t)]$ and $\Gamma_x \cap [R^k \times (t, t_2)]$ to achieve the desired property over each open interval (t_1, t_2) in $[0,T]$. The complement of the union of these finitely may open intervals in $[0,T]$ is a compact subset of $[0,T]$. Knowing that (i) all remaining discontinuities are of magnitude less than ϵ and (ii) limits exist everywhere from the left and right, we can conclude that there is a closed interval of positive length about each point in the compact set, where x oscillates by less than ϵ, i.e., $\sup\{\|x(t') - x(t'')\| < \epsilon$, where t', t'' are points in the interval. However, by the compactness, only finitely many of these closed intervals cover the compact set. We add points $(x(t), t)$ to A to ensure that there is at least one point (z, t) for which t is in one of these closed intervals. By this construction, A is finite and $\hat{d}(A, \Gamma_x) < \epsilon$. ■

To complete the proof of Theorem 6.3.1, we need the following result, which we prove by applying Lemma 6.3.1.

Lemma 6.3.2. (flexibility in choice of parametric representations) *For any $x_1, x_2 \in D$, $(u_1, r_1) \in \Pi_s(x_1)$ and $\epsilon > 0$, it is possible to find $(u_2, r_2) \in \Pi_s(x_2)$ such that*

$$\|u_1 - u_2\| \vee \|r_1 - r_2\| \le d_s(x_1, x_2) + \epsilon \ .$$

Proof. For $x_1, x_2 \in D$ and ϵ given, choose $(u_1', r_1') \in \Pi_s(x_1)$ and $(u_2', r_2') \in \Pi_s(x_2)$ such that

$$\|u_1' - u_2'\| \vee \|r_1' - r_2'\| < d_s(x_1, x_2) + \epsilon/4 \ . \tag{3.11}$$

Next apply Lemma 6.3.1 to find a finite ordered subsets $A_1 \subseteq \Gamma_{x_1}$ such that $\hat{d}(A_1, \Gamma_{x_1}) < \epsilon/4$. Next find a finite subset S_1' of $[0,1]$ of the same cardinality as A_1 such that $(u_1'(s), r_1'(s)) \in A_1$ for each $s \in S_1'$. Let S_1 be another finite subset of $[0,1]$ of the same cardinality as A_1 such that $(u_1(s), r_1(s)) \in A_1$ for each $s \in S_1$. Let λ be a homeomorphism of $[0,1]$ such that λ maps S_1 onto S_1'. Let $(u_2, r_2) = (u_2' \circ \lambda, r_2' \circ \lambda)$, where $\circ$ is the composition map. Trivially, by (3.11),

$$\|u_1' \circ \lambda - u_2' \circ \lambda\| \vee \|r_1' \circ \lambda - r_2' \circ \lambda\| < d_s(x_1, x_2) + \epsilon/4 \ .$$

Hence, it suffices to show that

$$\|u_1 - u_1' \circ \lambda\| \vee \|r_1 - r_1' \circ \lambda\| < 3\epsilon/4 \ . \tag{3.12}$$

First there is equality $u_1(s) = u_1'(\lambda(s))$ by construction at each $s \in S_1$. However, since $\hat{d}(A_1, \Gamma_x) < \epsilon/4$, (3.12) holds: For each $s \in [0,1]$, there is $s_i \in S_1$ such that $s_i \le s < s_{i+1}$ and

$$\begin{aligned} \|u_1(s) - u_1'(\lambda(s))\| &\le \|u_1(s) - u_1(s_i)\| + \|u_1(s_i) - u_1'(\lambda(s_i))\| \\ &\quad + \|u_1'(\lambda(s_i)) - u_1'(\lambda(s))\| \le \epsilon/2 \ . \ \blacksquare \end{aligned}$$

We will show that the metric d_s induces the standard M_1 topology defined by Skorohod (1956); see Theorem 6.5.1. Since $\Pi_s(x) \subseteq \Pi_w(x)$ for all x, we have $d_w(x_1, x_2) \le d_s(x_1, x_2)$ for all x_1, x_2, so that the WM_1 topology is indeed weaker than the SM_1 topology. However, we show below in Example 12.3.2 of the book that d_w in (3.8) is *not* a metric when $k > 1$.

For $x_1, x_2 \in D([0,T], \mathbb{R}^k)$, let d_p be a metric inducing the product topology, defined by

$$d_p(x_1, x_2) \equiv \max_{1 \le i \le k} d(x_1^i, x_2^i) \tag{3.13}$$

for $x_j \equiv (x_j^1, \ldots, x_j^k)$ and $j = 1, 2$. (Note that $d_s = d_w = d_p$ when the functions are real valued, in which case we use the notation d.) It is an easy consequence of (3.8), (3.13) and the second representation in (3.4) that the WM_1 topology is stronger than the product topology, i.e., $d_p(x_1, x_2) \le d_w(x_1, x_2)$ for all $x_1, x_2 \in D$. In Section 6.5 we will show that actually the WM_1 and product topologies coincide.

Example 12.3.1 of the book shows that SM_1 is strictly stronger than WM_1.

We now relate the metrics $d_{M_1} \equiv d_s$ and d_{J_1} for d_{J_1} in equation 3.2 of Section 3.3 in the book.

Theorem 6.3.2. (comparison of J_1 and M_1 metrics) *For each $x_1, x_2 \in D$,*

$$d_s(x_1, x_2) \le d_{J_1}(x_1, x_2) \ .$$

Proof. For any $x_1, x_2 \in D$ and $\lambda \in \Lambda$, we show how to define parametric representations (u_j, r_j) in $\Pi_s(x_j)$ for $j = 1, 2$ such that

$$\|u_1 - u_2\| \vee \|r_1 - r_2\| = \|x_1 \circ \lambda - x_2\| \vee \|\lambda - e\| \ . \tag{3.14}$$

If, for any $\epsilon > 0$, we first choose $\lambda \in \Lambda$ so that

$$\|x_1 \circ \lambda - x_2\| \vee \|\lambda - e\| \leq d_{J_1}(x_1, x_2) + \epsilon \ ,$$

the associated parametric representation yield

$$d_s(x_1, x_2) \leq \|u_1 - u_2\| \vee \|r_1 - r_2\| \leq d_{J_1}(x_1, x_2) + \epsilon \ .$$

Since ϵ is arbitrary, that will complete the proof. Suppose that

$$t_n \in Disc(x_1, x_2) \equiv Disc(x_1) \cup Disc(x_2) \, , \ n \geq 1 \, ,$$

where t_n is ordered (indexed) first by the norm of the jump and then the location, with values closer to 0 occurring first. Associate with each time point t_n a closed subinterval $[a_n, b_n]$ in $(0, 1)$ such that the subintervals are ordered, i.e., if $t_i < t_j < t_k$ are three points in $Disc(x_1, x_2)$, then $a_i < b_i < a_j < b_j < a_k < b_k$. Then let $r_2(s) = t_n$ for $a_n \leq s \leq b_n$. If $t \notin Disc(x_1, x_2)$ but $t_{n_k} \downarrow t$ as $n_k \to \infty$ for $t_{n_k} \in Disc(x_1, x_2)$, then let $r_2(s) = \lim_{n_k \to \infty} r_2(a_{n_k})$. Similarly, if $t \notin Disc(x_1, x_2)$ but $t_{n_k} \uparrow t$ as $n_k \to \infty$ for $t_{n_k} \in Disc(x_1, x_2)$, then let $r_2(s) = \lim_{n_k \to \infty} r_2(b_{n_k})$. Finally, let $r_2(s)$ be defined by linear interpolation in all remaining gaps. This makes r_2 continuous and nondecreasing. Having defined r_2, let $r_1 = \lambda \circ r_2$, $u_1(s) = (x_1 \circ r_1)(s)$ and $u_2(s) = (x_2 \circ r_2)(s)$ for all s, except $s \in (a_n, b_n)$ for some n. Within each subinterval (a_n, b_n), let u_1 and u_2 be defined by linear interpolation from their values at the endpoints a_n and b_n. This construction makes $(u_j, r_j) \in \Pi_s(x_j)$ for $j = 1, 2$ and yields (3.14), thus completing the proof. ∎

6.3.3. Properties of Parametric Representations

We conclude this section by further discussing strong parametric representations. For $x \in D$, $t \in Disc(x)$ and $(u, r) \in \Pi_s(x)$, there exists a unique pair of points $s_- \equiv s_-(t, x)$ and $s_+ \equiv s_+(t, x)$ such that $s_- < s_+$ and $r^{-1}(\{t\}) = [s_-, s_+]$, i.e.,

$$\begin{aligned} &\text{(i) } r(s) < t \text{ for } s < s_- \\ &\text{(ii) } r(s) = t \text{ for } s_- \leq s \leq s_+ \\ &\text{(iii) } r(s) > t \text{ for } s > s_+ \ . \end{aligned} \tag{3.15}$$

We will exploit the fact that a parametric representation (u, r) in $\Pi_s(x)$ is *jump consistent*: for each $t \in Disc(x)$ and pair $s_- \equiv s_-(t,x) < s_+ \equiv s_+(t,x)$ such that (3.15) holds, there is a continuous nondecreasing function β_t mapping $[0,1]$ onto $[0,1]$ such that

$$u(s) = \beta_t\left(\frac{s-s_-}{s_+-s_-}\right)u(s_+)+\left[1-\beta_t\left(\frac{s-s_-}{s_+-s_-}\right)\right]u(s_-) \quad \text{for} \quad s_- \le s \le s_+ . \tag{3.16}$$

Condition (3.16) means that u is defined within jumps by interpolation from the definition at the endpoints s_- and s_+, consistently over all coordinates. In particular, suppose that $t \in Disc(x^i)$. (Since $t \in Disc(x)$, we must have $t \in Disc(x^i)$ for some coordinate i.) Suppose that $x^i(t-) < x^i(t)$. Then we can let

$$\beta_t(s) = \frac{u^i(s) - u^i(s_-)}{u^i(s_+) - u^i(s_-)} . \tag{3.17}$$

We see that (3.16) and (3.17) are consistent in that

$$u^i(s) = \beta_t\left(\frac{s-s_-}{s_+-s_-}\right)u^i(s_+) + \left[1-\beta_t\left(\frac{s-s_-}{s_+-s_-}\right)\right]u^i(s_-) \tag{3.18}$$

for β_t in (3.17). For another coordinate j, (3.16) and (3.17) imply that

$$u^j(s) = \left(\frac{u^i(s)-u^i(s_-)}{u^i(s_+)-u^i(s_-)}\right)u^j(s_+) + \left(\frac{u^i(s_+)-u^i(s)}{u^i(s_+)-u^i(s_-)}\right)u^j(s_-) . \tag{3.19}$$

It is possible that $t \notin Disc(x^j)$, in which case $u^j(s) = u^j(s_-) = u^j(s_+)$ for all s, $s_- \le s \le s_+$.

We can further characterize the behavior of a strong parametric representation at a discontinuity point. For $x \in D$, $t \in Disc(x)$ and $(u,r) \in \Pi_s(x)$, there exists a unique set of four points $s_- \equiv s_-(t,x) \le s'_- \equiv s'_-(t,x) < s'_+ \equiv s'_+(t,x) \le s_+ \equiv s_+(t,x)$ such that (3.15) holds and

$$\begin{aligned}
&\text{(i) } u(s) = u(s_-) \text{ for } s_- \le s \le s'_-,\\
&\text{(ii) for each } i, \text{ either } u^i(s_-) < u^i(s) < u^i(s_+),\\
&\qquad \text{or } u^i(s_-) > u^i(s) > u^i(s_+) \text{ for } s'_- < s < s'_+,\\
&\text{(iii) } u(s) = u(s_+) \text{ for } s'_+ \le s \le s_+ .
\end{aligned} \tag{3.20}$$

Let D_1 be the subset of D containing functions all of whose jumps occur in only one coordinate, i.e., the set of x such that, for each $t \in Disc(x)$ there exists one and only one $i \equiv i(t)$ such that $t \in Disc(x^i)$. (The coordinate i may depend on t.)

Lemma 6.3.3. (strong and weak parametric representations coincide on D_1) *For each* $x \in D_1$, $\Pi_s(x) = \Pi_w(x)$.

Proof. Since $\Pi_s(x) \subseteq \Pi_w(x)$, we need to show that $(u,r) \in \Pi_w(x)$ is in $\Pi_s(x)$ for x in $D^{(1)}$. Pick any $t \in Disc(x)$ and let i be the coordinate of x with a jump at t. We can then define the β_t needed for (3.16) using (3.17). Since $u^j(s) = u^j(s_-) = u^j(s_+)$ for all j with $j \neq i$, (3.19) and (3.16) are then satisfied. ∎

Corollary. *For each* $x \in D([0,T],\mathbb{R}^1)$, $\Pi_s(x) = \Pi_w(x)$.

We now show that parametric representations are preserved under linear functions of the coordinates when $x \in \Pi_s(x)$. That is *not* true in $\Pi_w(x)$.

Lemma 6.3.4. (linear functions of parametric representations) *If* $(u,r) \in \Pi_s(x)$, *then* $(\eta u, r) \in \Pi_s(\eta x)$ *for any* $\eta \in \mathbb{R}^k$.

Proof. By the Corollary to Lemma 6.3.3, $\Pi_s(\eta x) = \Pi_w(\eta x)$. Hence, it suffices to show that $(\eta u, r) \in \Pi_w(\eta x)$. It is clear that $(\eta u, r)$ is continuous and nondecreasing. For $t \in Disc(\eta x)$, necessarily $t \in Disc(x)$. (We could have $t \in Disc(x)$ but $t \notin Disc(\eta x)$, but that does not concern us.) By (3.16), when $r(s) = t$,

$$\eta u(s) = \beta_t\left(\frac{s - s_-}{s_+ - s_-}\right)\eta u(s_+) + \left[1 - \beta_t\left(\frac{s - s_-}{s_+ - s_-}\right)\right]\eta u(s_-)$$

which completes the proof. ∎

6.4. Local Uniform Convergence at Continuity Points

In this section we provide alternative characterizations of local uniform convergence at continuity points of a limit function. The nonuniform Skorohod topologies on D all imply local uniform convergence at continuity points of a limit function. They differ by their behavior at discontinuity points.

We start by defining two basic *uniform-distance functions.* For $x_1, x_2 \in D$, $t \in [0,T]$ and $\delta > 0$, let

$$u(x_1, x_2, t, \delta) \equiv \sup_{0 \vee (t-\delta) \le t_1 \le (t+\delta) \wedge T} \{\|x_1(t_1) - x_2(t_1)\|\}\ , \qquad (4.1)$$

$$v(x_1, x_2, t, \delta) \equiv \sup_{0 \vee (t-\delta) \le t_1, t_2 \le (t+\delta) \wedge T} \{\|x_1(t_1) - x_2(t_2)\|\}\ , \qquad (4.2)$$

We also define an *oscillation function.* For $x \in D$, $t \in [0,T]$ and $\delta > 0$, let

$$\bar{v}(x, t, \delta) \equiv \sup_{0 \vee (t-\delta) \le t_1 \le t_2 \le (t+\delta) \wedge T} \{\|x(t_1) - x(t_2)\|\}\ . \qquad (4.3)$$

We next define oscillation functions that we will use with the M_1 topologies. They use the distance $\|z - A\|$ between a point z and a subset A in $\mathbb{R}^k$ defined in equation 5.3 in Section 11.5 of the book. The SM_1 and WM_1 topologies use the standard and product segments in (3.1) and (3.2). For each $x \in D$, $t \in [0,T]$ and $\delta > 0$, let

$$w_s(x,t,\delta) \equiv \sup_{0\vee(t-\delta)\le t_1<t_2<t_3\le(t+\delta)\wedge T} \{\|x(t_2) - [x(t_1), x(t_3)]\| \tag{4.4}$$

and

$$w_w(x,t,\delta) \equiv \sup_{0\vee(t-\delta)\le t_1<t_2<t_3\le(t+\delta)\wedge T} \{\|x(t_2) - [[x(t_1), x(t_3)]]\| \tag{4.5}$$

We now turn to the M_2 topology, which we will be studying in Sections 6.10 and 6.11. We define two uniform-distance functions. We use $\bar{w}$ as opposed to w to denote an M_2 uniform-distance function. Just as with the M_1 topologies, the SM_2 and WM_2 topologies use the standard and product segments in (3.1) and (3.2). For x_1, $x_2 \in D$, let

$$\bar{w}_s(x_1,x_2,t,\delta) \equiv \sup_{0\vee(t-\delta)\le t_1\le(t+\delta)\wedge T} \{\|x_1(t_1) - [x_2(t-), x_2(t)]\|\} \tag{4.6}$$

$$\bar{w}_w(x_1,x_2,t,\delta) \equiv \sup_{0\vee(t-\delta)\le t_1\le(t+\delta)\wedge T} \{\|x_1(t_1) - [[x_2(t-), x_2(t)]]\|\} \tag{4.7}$$

It is easy to establish the following relations among the uniform-distance and oscillation functions.

Lemma 6.4.1. (inequalities for uniform-distance and oscillation functions) *For all $x, x_n \in D$, $t \in [0,T]$ and $\delta > 0$,*

$$u(x_n,x,t,\delta) \le v(x_n,x,t,\delta) \le u(x_n,x,t,\delta) + \bar{v}(x,t,\delta) \ ,$$

$$w_w(x_n,t,\delta) \le w_s(x_n,t,\delta) \le \bar{v}(x_n,t,\delta) \le 2v(x_n,x,t,\delta) + \bar{v}(x,t,\delta) \ ,$$

$$\bar{w}_w(x_n,x,t,\delta) \le \bar{w}_s(x_n,x,t,\delta) \le v(x_n,x,t,\delta) \le 2\bar{w}_w(x_n,x,t,\delta) + \bar{v}(x,t,\delta) \ .$$

Since the M_1-oscillation functions $w_s(x_n,t,\delta)$ and $w_w(x_n,t,\delta)$ do not contain the limit x, their convergence to 0 as $n \to \infty$ and then $\delta \downarrow 0$ does not directly imply local uniform convergence at a continuity point of a prospective limit function x.

We relate convergence of $w_s(x)n,t,\delta)$ and $w_w(x_n,t,\delta)$ to 0 as $n \to \infty$ and $\delta \downarrow 0$ to local uniform convergence by requiring pointwise convergence in a neighborhood of t; see (vi) in Theorem 6.4.1 below.

Theorem 6.4.1. (characterizations of local uniform convergence at continuity points) *If $t \notin Disc(x)$, then the following are equivalent:*

$$(i) \qquad \lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; u(x_n, x, t, \delta) = 0 \; , \tag{4.8}$$

$$(ii) \qquad \lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; v(x_n, x, t, \delta) = 0 \; , \tag{4.9}$$

$$(iii) \qquad \lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; \bar{w}_s(x_n, x, t, \delta) = 0 \; , \tag{4.10}$$

$$(iv) \qquad \lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; \bar{w}_w(x_n, x, t, \delta) = 0 \; , \tag{4.11}$$

(v) $x_n(t_1) \to x(t_1)$ for all t_1 in a dense subset of a neighborhood of t (including 0 if $t = 0$ or T if $t = T$) and

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; w_s(x_n, t, \delta) = 0 \; ,$$

(vi) $x_n(t_1) \to x(t_1)$ for all t_1 in a dense subset of a neighborhood of t (including 0 if $t = 0$ or T if $t = T$) and

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; w_w(x_n, t, \delta) = 0 \; . \tag{4.12}$$

Proof. By Lemma 6.4.1, we have the implications (i) $\leftrightarrow$ (ii) $\leftrightarrow$ (iii) $\leftrightarrow$ (iv) and (ii) $\to$ (v) $\to$ (vi). Hence it suffices to show that (vi) $\to$ (i), which we now do. For $x, t \notin Disc(x)$ and $\epsilon > 0$ given, choose $\delta > 0$ so that $\bar{v}(x, t, \delta) < \epsilon$, which is possible since $t \notin Disc(x)$. Also let δ be sufficiently small so that $x_n(t_1') \to x(t_1')$ as $n \to \infty$ for all t_1' in a dense subset of $[0 \vee (t-\delta), (t+\delta) \wedge T]$. Note that we can treat 0 and T directly. For $t_1 \in (0 \vee (t-\delta), T \wedge (t+\delta))$ given, choose t_1', t_2' so that $0 \vee (t-\delta) < t_1' < t_1 < t_2' < (t+\delta) \wedge T$ and $x_n(t_j') \to x(t_j')$ as $n \to \infty$ for $j = 1, 2$. Then choose n_0 so that $\|x_n(t') - x(t'')\| < \epsilon$ for $t'' = 0$, T, t_1' and t_2' and $w_w(x_n, t, \delta) < \epsilon$ for $n \geq n_0$. Then, for $n \geq n_0$,

$$\begin{aligned}
\|x_n(t_1) - x(t_1)\| &\leq \|x_n(t_1) - x_n(t_1')\| + \|x_n(t_1') - x(t_1')\| + \|x(t_1') - x(t_1)\| \\
&\leq \|x_n(t_1) - x_n(t_1')\| + 2\epsilon \\
&\leq \|x_n(t_1) - [[x_n(t_1'), x_n(t_2')]]\| + \|x_n(t_1') - x_n(t_2')\| + 2\epsilon \\
&\leq w_w(x_n, t, \delta) + \|x_n(t_1') - x_n(t_2')\| + 2\epsilon \\
&\leq \|x_n(t_1') - x(t_1')\| + \|x(t_1') - x(t_2')\| \\
&\qquad + \|x(t_2') - x_n(t_2')\| + 3\epsilon \leq 6\epsilon \; .
\end{aligned}$$

It remains to consider $t = 0$ and $t = T$. The reasoning is the same for these two cases, so we consider only $t = 0$. For $t = 0$, note that

$$\|x_n(t_1) - x(t_1)\| \le \|x_n(t_1) - x_n(0)\| + \|x_n(0) - x(0)\| + \|x(0) - x(t)\| \ . \quad (4.13)$$

The third term in (4.13) can be made small using the right continuity of x at 0; the second term in (4.13) can be made small by the assumed convergence at 0; the first term in (4.13) can be made small by (4.12). ∎

We now show that local uniform convergence at all points in a compact interval implies uniform convergence over the compact interval.

Lemma 6.4.2. (local uniform convergence everywhere in a compact interval) *If (4.8) holds for all $t \in [a, b]$, then*

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \sup_{0\vee(a-\delta)\le t\le(b+\delta)\wedge T} \{\|x_n(t) - x(t)\|\} = 0 \ .$$

Proof. By (4.8), for all $\epsilon > 0$ and $t \in [a, b]$, there exists $\delta(t)$ such that

$$\overline{\lim_{n\to\infty}} \ u(x_n, x, t, \delta(t)) < \epsilon \ .$$

For each t, there is thus uniform asymptotic closeness in the intervals $(0 \vee (t - \delta(t)), (t + \delta(t)) \wedge T)$. However, these intervals form an open cover of the interval $[a, b]$. Since $[a, b]$ is compact, there is a finite subcover. Hence, there is a $\delta' > 0$ such that

$$\overline{\lim_{n\to\infty}} \sup_{0\vee(a-\delta'])\le t\le(b+\delta')\wedge T} \{\|x_n(t) - x(t)\|\} < \epsilon \ .$$

Since ϵ was arbitrary, this implies the desired conclusion. ∎

6.5. Alternative Characterizations of M_1 Convergence

We now give alternative characterizations of SM_1 and WM_1 convergence.

6.5.1. SM_1 Convergence

We first establish alternative characterizations of SM_1 convergence or, equivalently, d_s-convergence. One characterization is a minor variant of the original one involving an oscillation function established by Skorohod (1956). Another one – (v) below – involves only the local behavior of the functions.

It helps us establish sufficient conditions to have $d_s((x_n, y_n), (x, y)) \to 0$ in $D([0,T], \mathbb{R}^{k+l})$ when $d_s(x_n, x) \to 0$ in $D([0,T], \mathbb{R}^k)$ and $d_s(y_n, y) \to 0$ in $D([0,T], \mathbb{R}^l)$; see Section 6.6. For the SM_1 topology, we define another oscillation function. For any $x_1, x_2 \in D$ and $\delta > 0$, let

$$w_s(x, \delta) \equiv \sup_{0 \le t \le T} w_s(x, t, \delta) , \tag{5.1}$$

for $w_s(x, t, \delta)$ in (4.4).

The following main result is proved in the book. It only remains to prove the supporting lemmas, which we do here.

Theorem 6.5.1. (characterizations of SM_1 convergence) *The following are equivalent characterizations of convergence $x_n \to x$ as $n \to \infty$ in (D, SM_1):*

(i) For any $(u, r) \in \Pi_s(x)$, there exists $(u_n, r_n) \in \Pi_s(x_n)$, $n \ge 1$, such that

$$\|u_n - u\| \vee \|r_n - r\| \to 0 \quad as \quad n \to \infty . \tag{5.2}$$

(ii) There exist $(u, r) \in \Pi_s(x)$ and $(u_n, r_n) \in \Pi_s(x_n)$ for $n \ge 1$ such that (5.2) holds.

(iii) $d_s(x_n, x) \to 0$ as $n \to \infty$; i.e., for all $\epsilon > 0$ and all sufficiently large n, there exist $(u, r) \in \Pi_s(x)$ and $(u_n, r_n) \in \Pi_s(x_n)$ such that

$$\|u_n - u\| \vee \|r_n - r\| < \epsilon .$$

(iv) $x_n(t) \to x(t)$ as $n \to \infty$ for each t in a dense subset of $[0, T]$ including 0 and T, and

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ w_s(x_n, \delta) = 0 \tag{5.3}$$

for $w_s(x, \delta)$ in (5.1) and $w_s(x, t, \delta)$ in (4.4).

(v) $x_n(T) \to x(T)$ as $n \to \infty$; for each $t \notin Disc(x)$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ v(x_n, x, t, \delta) = 0 \tag{5.4}$$

for $v(x_1, x_2, t, \delta)$ in (4.2); and, for each $t \in Disc(x)$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ w_s(x_n, t, \delta) = 0 \tag{5.5}$$

for $w_s(x, t, \delta)$ in (4.4).

(vi) *For all* $\epsilon > 0$, , *there exist integers* m *and* n_1, *a finite ordered subset* A *of* Γ_x *of cardinality* m *as in* (3.9) and, for all $n \geq n_1$, finite ordered subsets A_n of Γ_{x_n} of cardinality m *such that, for all* $n \geq n_1$, $\hat{d}(A, \Gamma_x) < \epsilon$, $\hat{d}(A_n, \Gamma_{x_n}) < \epsilon$ *for* $\hat{d}$ *in* (3.10) *and* $d^*(A, A_n) < \epsilon$, *where*

$$d^*(A, A_n) \equiv \max_{1 \leq i \leq m} \{\|(z_i, t_i) - (z_{n,i}, t_{n,i})\| : (z_i, t_i) \in A, (z_{n,i}, t_{n,i}) \in A_n\}. \tag{5.6}$$

In preparation for the proof of Theorem 6.5.1, we establish some preliminary results. We first show that SM_1 convergence implies local uniform convergence at all continuity points.

Lemma 6.5.1. (local uniform convergence) *If* $d_s(x_n, x) \to 0$ *as* $n \to \infty$, *then* (4.9) *holds for each* $t \notin Disc(x)$.

Proof. For x, $t \in Disc(x)^c$ and $\epsilon > 0$ given, choose $\delta > 0$ so that $\|x(t') - x(t)\| < \epsilon$ for $|t - t'| < \delta$. Then choose $n_0 \geq 4$, $(u_n, r_n) \in \Pi_s(x_n)$ and $(u, r) \in \Pi_s(x)$ such that

$$\|u_n - u\| \vee \|r_n - r\| < (\delta \wedge \epsilon)/4$$

for all $n \geq n_0$. Let s_1, s_2, s_3 be such that $r(s_1) = t - \delta/2$, $r(s_2) = t$ and $r(s_3) = t + \delta/2$. Then $r_n(s_1) < t < \delta/4$ and $r_n(s_3) > t + \delta/4$ for all $n \geq n_0$. Hence, for all $t' \in (t - \delta/4, t + \delta/4)$ and $n \geq n_0$ there exists s_n, $s_1 < s_n < s_3$, such that $(u_n(s_n), r_n(s_n)) = (x_n(t'), t')$. Hence,

$$\begin{aligned}
\|x_n(t') - x(t')\| &= \|u_n(s_n) - u(s_2)\| + \|x(t) - x(t'')\| \\
&\leq \|u_n(s_n) - u(s_n)\| + \|u(s_n) - u(s_2)\| + \epsilon \\
&\leq (\delta \wedge \epsilon)/2 + 2\epsilon < 3\epsilon . \quad \blacksquare
\end{aligned}$$

We next relate the modulus w_s applied to x and the modulus applied to corresponding points on the graph Γ_x. The following lemma is established in the proof of Skorohod's (1956) 2.4.1.

Lemma 6.5.2. (extending the modulus from a function to its graph) *If* (z_1, t_1), (z_2, t_2), $(z_3, t_3) \in \Gamma_x$ *with* $0 \vee (t - \delta) \leq t_1 < t_2 < t_3 \leq (t + \delta) \wedge T$, *then* $\|z_2 - [z_1, z_3]\| \leq w_s(x, \delta)$.

Proof. Suppose that $w_s(x,\delta) = \epsilon$. It suffices to show: (i) that $\|z_2 - [z_1, z_3]\| \le \epsilon$ when $\|z_2' - [z_1, z_3]\| \le \epsilon$, $\|z_2'' - [z_1, z_3]\| \le \epsilon$ and $z_2 \in [z_2', z_2'']$ and (ii) that $\|z_2 - [z_1, z_3]\| \le \epsilon$ when $\|z_2 - [z_1', z_3]\| \le \epsilon$, $\|z_2 - [z_1'', z_3]\| \le \epsilon$ and $z_1 \in [z_1', z_1'']$. For (i), note that there exist z', $z'' \in [z_1, z_3]$ such that $\|z_2' - z'\| \le \epsilon$ and $\|z_2'' - z''\| \le \epsilon$. Also there exists α, $0 \le \alpha \le 1$ such that $z_2 = \alpha z_2' + (1-\alpha) z_2''$. Hence $\|z_2 - (\alpha z' + (1-\alpha) z'')\| \le \epsilon$, which implies that

$$\|z_2 - [z', z'']\| \le \|z_2 - [z_1, z_3]\| \le \epsilon \ .$$

For (ii), note first that there exist $z' \in [z_1', x_3]$ and $z'' \in [z_1'', z_3]$ such that $\|z_2 - z'\| \le \epsilon$ and $\|z_2 - z''\| \le \epsilon$. Hence, for any $z \in [z', z'']$, $\|z_2 - z\| \le \epsilon$. The desired z lies on the intersection of $[z_1, z_3]$ and $[z', z'']$. That implies the desired conclusion. ∎

Lemma 6.5.3. (asymptotic negligibility of the modulus) *For any* $x \in D$, $w_s(x, \delta) \downarrow 0$ *as* $\delta \downarrow 0$.

Proof. For any $\epsilon > 0$, choose $x_c \in D_c$ such that $\|x - x_c\| < \epsilon/2$, which is always possible by Theorem 6.2.2. Note that, for any $\delta > 0$,

$$w_s(x, \delta) \le w_s(x_c, \delta) + 2\|x - x_c\| \ ,$$

so that

$$w_s(x, \delta) \le w_s(x_c, \delta) + \epsilon \ .$$

Let η be the minimum distance between successive discontinuities in x_c. Since $w_s(x_c, \delta) = 0$ when $\delta < \eta$, $w_s(x, \delta) < \epsilon$ when $\delta < \eta$. ∎

Proof of Theorem 6.5.1. Contained in the book. ∎

6.5.2. WM_1 Convergence

We now establish an analog of Theorem 6.5.1 for the WM_1 topology. Several alternative characterizations of WM_1 convergence will follow directly from Theorem 6.5.1 because we will show that convergence $x_n \to x$ as $n \to \infty$ in WM_1 is equivalent to $d_p(x_n, x) \to 0$. To treat the WM_1 topology, we define another oscillation function. Let

$$w_w(x, \delta) \equiv \sup_{0 \le t \le T} w_w(x, t, \delta) \tag{5.7}$$

for $w_w(x, t, \delta)$ in (4.5). Recall that $w_w(x, t, \delta)$ in (4.5) is the same as $w_s(x, t, \delta)$ in (4.4) except it has the product segment $[[x(t_1), x(t_3)]]$ in (3.2) instead of the standard segment $[x(t_1), x(t_3)]$ in (3.1).

Paralleling Definition 6.3.1, let an ordered subset A of G_x of cardinality m be such that (3.9) holds, but now with the order being the order on G_x. Paralleling (3.10), let the *order-consistent distance* between A and G_x be

$$\hat{d}(A, G_x) \equiv \sup\{\|(z,t) - (z_i, t_i)\| \vee \|(z,t) - (z_{i+1}, t_{i+1})\| : (z,t) \in G_x\} \quad (5.8)$$

with the supremum being over all $(z,t) \in G_x$ such that $(z_i, t_i) \leq (z,t) \leq (z_{i+1}, t_{i+1})$ for all i, $1 \leq i \leq m-1$.

Theorem 6.5.2. (characterizations of WM_1 convergence) *The following are equivalent characterizations of $x_n \to x$ as $n \to \infty$ in (D, WM_1):*

(i) $d_w(x_n, x) \to 0$ *as* $n \to \infty$.

(ii) $d_p(x_n, x) \to 0$ *as* $n \to \infty$.

(iii) $x_n(t) \to x(t)$ *as* $n \to \infty$ *for each* t *in a dense subset of* $[0,T]$ *including* 0 *and* T*, and*

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \; w_w(x_n, \delta) = 0 \; . \quad (5.9)$$

(iv) $x_n(T) \to x(T)$ *as* $n \to \infty$*; for each* $t \notin Disc(x)$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \; v(x_n, x, t, \delta) = 0 \quad (5.10)$$

for $v(x_n, x, t, \delta)$ *in* (4.2)*; and, for each* $t \in Disc(x)$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \; w_w(x_n, t, \delta) = 0 \quad (5.11)$$

for $w_w(x_n, t, \delta)$ *in* (4.5).

(v) for all $\epsilon > 0$ *and all* n *sufficiently large, there exist finite ordered subsets* A *of* G_x *(in general depending on* n*) and* A_n *of* G_{x_n} *of common cardinality such that* $\hat{d}(A, G_x) < \epsilon$, $\hat{d}(A_n, G_{x_n}) < \epsilon$ *and* $d^*(A, A_n) < \epsilon$ *for* $\hat{d}$ *in* (5.8) *and* d^* *in* (5.6).

Proof. (i)$\to$(ii). Since $d_p \leq d_w$, (i)$\to$(ii) is immediate.
(ii)$\leftrightarrow$(iii). The implication (iii)$\to$(ii) is immediate, so we show (ii)$\to$(iii). By Lemma 6.5.1, $x_n^i(t) \to x^i(t)$ as $n \to \infty$ for each $t \in Disc(x^i)^c$, $1 \leq i \leq k$.

That implies that $x_n(t) \to x(t)$ as $n \to \infty$ for each $t \in Disc(x)^c$. From Theorem 6.5.1, $d_p(x_n, x) \to 0$ as $n \to \infty$ also implies that

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; w_s(x_n^i, \delta) = 0$$

for each i, $1 \le i \le k$, but that directly implies (5.9), because

$$\|x_n(t_2) - [[x_n(t_1), x_n(t_3)]]\| = \max_{1\le i\le k} \|x_n^i(t_2) - [x_n^i(t_1), x_n^i(t_3)]\| \; , \qquad (5.12)$$

so that

$$w_w(x_n, \delta) = \max_{1\le i\le k} w_s(x_n^i, \delta) \; . \qquad (5.13)$$

(iii)$\leftrightarrow$(iv). The equivalence between (iii) and (iv) holds by the same reasoning used to establish the equivalence of (iv) and (v) in Theorem 6.5.1.
(iii)$\to$(v). The proof of (iii)$\to$(v) parallels the proof of (iv)$\to$(vi) in Theorem 6.5.1, but requires some modifications. Paralleling the previous beginning, for $\epsilon > 0$ given, find $\eta < \epsilon/16$ and n_0 such that $w_w(x_n, \eta) < \epsilon/32$ for $n \ge n_0$. However, we do not next directly construct $A \in G_x$. Instead, just as with the SM_1 topology, we first construct the finite set A of Γ_x as before with the properties in the proof of Theorem 6.5.1. We denote this subset A' to distinguish it from the desired subset A of G_x. As before, for all $t_i \in S \cap A'$, let $n_1 \ge n_0$ be such that $\|x_n(t_i) - x(t_i)\| < \epsilon/32$ for all i, $1 \le i \le k$, and all $n \ge n_1$. We now want to construct the ordered subset A_n in G_{x_n}. For $t \in S$, the construction is as before: $(z_{n,i}, t_{n,i}) = (x_n(t_i), t_i)$. Next suppose that (**??**) holds. Then $(z_{n,r}, t_{n,r})$ and $(z_{n,r+j+1}, t_{n,r+j+1})$ have been defined with respect to A'. We insert points into A_n from G_{x_n} appropriately spaced in between the two points. By construction specified before (but using the product segments),

$$\begin{aligned} &\|[[(x_n(t_r), t_r), (x_n(t_{r+j+1}), t_{r+j+1})]] \\ &-[[(x(t_r), t_r), (x(t_{r+j+1}), t_{r+j+1})]]\| < \epsilon/32 \end{aligned} \qquad (5.14)$$

and

$$\|[[(x(t_r), t_r), (x(t_{r+j+1}), t_{r+j+1})]] - [[(x(t-), t), (x(t), t)]]\| < \epsilon/32 \; . \qquad (5.15)$$

To simplify the discussion, suppose that $x^i(t-) \le x^i(t)$ for all i. (This is without loss of generality after redefining the order.) Consider an arbitrary nondecreasing (in the order on G_{x_n}) continuous curve in G_{x_n} from $(z_{n,r}, t_{n,r})$ to $(z_{n,r+j+1}, t_{n,r+j+1})$. Let $(z'_{n,r+1}, t'_{n,r+1})$ be the first point on this curve for which the i^{th} coordinate first reaches $z^i_{n,r} + \epsilon/4$ for some i. Given $(z_{n,r+k}, t_{n,r+k})$, let $(z_{n,r+k+1}, t_{n,r_k+1})$ be the next point on the curve at which the i^{th} coordinate first reaches $z^i_{n,r+k} + \epsilon/4$ for some i. Since $x^i(t-) \le x^i(t)$ for all i and since $w_w(x_n, \eta) < \epsilon/32$, no coordinate of the curve in G_{x_n} can decrease by more than $\epsilon/32$ over any subinterval, and thus from one points to the next in A_n. Continue in this manner for at most finitely many steps until the end point $(z_{n,r+j+1}, t_{n,r+j+1})$ is reached. The distance between successive points is $\epsilon/4$, while the distance between the last point inserted and $(z_{n,r+j+1}, t_{n,r+j+1})$ is less than $\epsilon/4$. Delete the first and last point inserted, so that all distances between successive points are between $\epsilon/4$ and $\epsilon/2$. In general, the number of inserted points is some finite number, not necessarily equal to j. These points are ordered, since they lie on the non-decreasing continuous curve through G_{x_n}. For each $t \in Disc(x, \epsilon/2)$, let A_n contain these specified points. This construction yields $\hat{d}(A_n, G_{x_n}) < \epsilon/2$. For $t \notin Disc(x, \epsilon/2)$, let A contain the points already constructed in A'. It remains to construct the points in A for $t \in Disc(x, \epsilon/2)$. For this purpose, we use the points in A_n associated with t. Again, to simplify the discussion, suppose that $x^i(t-) \le x^i(t)$ for all i. With this ordering, we let

$$z^i_{r+k} = x^i(t-) \vee \max_{1 \le l \le k} z^i_{n,r+l} \wedge x^i(t)$$

for each k and i. This definition guarantees that the points (z_{r+k}, t) belong to G_x and are ordered. Moreover, $\hat{d}(A, G_x) < \epsilon$. Finally, we must have $d^*(A, A_n) < \epsilon$, because otherwise the condition $w_w(x_n, \eta) < \epsilon/32$ would be violated.

(v)$\rightarrow$(i). Suppose that the conditions in (v) hold and let $\epsilon > 0$ be given. Construct the finite subsets A and A_n with the specified properties. Let (u, r) and (u_n, r_n) be arbitrary parametric representations of G_x and G_{x_n} such that there are points s_i in $S \subseteq [0, 1]$ such that both $(u(s_i), r(s_i)) = (z_i, t_i) \in A$ and $(u_n(s_i), r_n(s_i)) = (z_{n,i}, t_{n,i}) \in A_n$. Since A and A_n are ordered subsets of G_x and G_{x_n}, respectively that construction is possible. Finally, for any s, $0 < s < 1$, there is $s_i \in S$ such that $s_i \le s < s_{i+1}$ and

$$\begin{aligned} &\|u_n(s) - u(s)\| \vee \|r_n(s) - r(s)\| \le \|(u_n(s), r_n(s)) - (u_n(s_i), r_n(s_i))\| \\ &\quad + \|(u_n(s_i), r_n(s_i)) - u(s_i), r(s_i)\| + \|(u(s_i), r(s_i)) - u(s), r(s)\| \\ &\le \hat{d}(A_n, G_{x_n}) + d^*(A, A_n) + \hat{d}(A, G_x) \le 3\epsilon \ . \quad \blacksquare \end{aligned}$$

6.6. Strengthening the Mode of Convergence

Section 12.6 of the book applies the characterizations of M_1 convergence in previouis sections to establish conditions under which the mode of convergence can be strengthened: We find conditions under which WM_1 convergence can be replaced by SM_1 convergence. Most of the material appears in the book.

We use the following Lemma.

Lemma 6.6.1. (modulus bound for (x_n, y_n)) *For* $x_n \in D([0,T], \mathbb{R}^k)$, $y_n, y \in D([0,T], \mathbb{R}^l)$, $t \in [0,T]$ *and* $\delta > 0$,

$$w_s((x_n, y_n), t, \delta) \leq w_s(x_n, t, \delta) + 2v(y_n, y, t, \delta).$$

Proof. For $(t-\delta) \vee 0 \leq t_1 < t_2 < t_3 \leq (t+\delta) \wedge T$,

$$\begin{aligned}
\|(x_n, y_n)(t_2) &- [(x_n, y_n)(t_1), (x_n, y_n)(t_3)]\| \\
&\leq \|(x_n, y_n)(t_2) - [(x_n(t_1), y(t)), (x_n(t_3), y(t))]\| \\
&\quad + (\|y_n(t_1) - y(t)\| \vee \|y_n(t_3) - y(t)\|) \\
&\leq \|x_n(t_2) - [x_n(t_1), x_n(t_3)]\| \vee \|y_n(t_2) - y(t)\| \\
&\quad + (\|y_n(t_1) - y(t)\| \vee \|y_n(t_3) - y(t)\|) \\
&\leq \|x_n(t_2) - [x_n(t_1), x_n(t_3)]\| + 2v(y_n, y, t, \delta). \quad \blacksquare
\end{aligned}$$

Theorem 6.6.1. (extending SM_1 convergence to product spaces) *Suppose that* $d_s(x_n, x) \to 0$ *in* $D([0,T], \mathbb{R}^k)$ *and* $d_s(y_n, y) \to 0$ *in* $D([0,T], \mathbb{R}^l)$ *as* $n \to \infty$. *If*

$$Disc(x) \cap Disc(y) = \phi.$$

then

$$d_s((x_n, y_n), (x, y)) \to 0 \quad in \quad D([0,T], \mathbb{R}^{k+l}) \quad as \quad n \to \infty.$$

The proof is in the book.

6.7. Characterizing Convergence with Mappings

In this section we focus on alternative characterizations of SM_1 convergence using mappings.

6.7.1. Linear Functions of the Coordinates

The strong topology SM_1 differs from the weak topology WM_1 by the behavior of linear functions of the coordinates. Example **??** shows that linear functions of the coordinates are not continuous in the product topology (there $(x_n^1 - x_n^2) \not\to (x^1 - x^2)$ as $n \to \infty$), but they are in the strong topology, as we now show. Note that there is no subscript on d on the left in (7.1) below because ηx is real valued.

Theorem 6.7.1. (Lipschitz property of linear functions of the coordinate functions) *For any* x_1, $x_2 \in D([0,T], \mathbb{R}^k)$ *and* $\eta \in \mathbb{R}^k$,

$$d(\eta x_1, \eta x_2) \le (\|\eta\| \vee 1) d_s(x_1, x_2) \ . \tag{7.1}$$

Proof. Pick an arbitrary $\epsilon > 0$ and choose $(u_j, r_j) \in \Pi_s(x_j)$ for $j = 1, 2$ such that

$$\|u_1 - u_2\| \vee \|r_1 - r_2\| < d_s(x_1, x_2) + \epsilon \ ,$$

which is possible by the definition (3.7). Because $\eta u_j \in \Pi(\eta x_j)$ for $j = 1, 2$, by Lemma 6.3.4,

$$\begin{aligned} d(\eta x_1, \eta x_2) &\le \|\eta u_1 - \eta u_2\| \vee \|r_1 - r_2\| \\ &\le \|r_1 - r_2\| \vee \|u_1 - u_2\| \|\eta\| \\ &\le (\|\eta\| \vee 1)(d_s(x_1, x_2) + \epsilon) \ . \end{aligned}$$

Since ϵ was arbitrary, (7.1) is established. ∎

We now obtain a sufficient condition for addition to be continuous on $(D, d_s) \times (D, d_s)$, which is analogous to the J_1 result in Theorem 4.1 of Whitt (1980).

Corollary 6.7.1. (SM_1-continuity of addition) *If* $d_s(x_n, x) \to 0$ *and* $d_s(y_n, y) \to 0$ *in* $D([0,T], \mathbb{R}^k)$ *and*

$$Disc(x) \cap Disc(y) = \phi \, ,$$

then

$$d_s(x_n + y_n, x + y) \to 0 \quad in \quad D([0,T], \mathbb{R}^k) \, .$$

Proof. First apply Theorem 6.6.1 to get $d_s((x_n, y_n), (x, y)) \to 0$ in $D([0,T], \mathbb{R}^{2k})$. Then apply Theorem 6.7.1. ∎

Remark 6.7.1. *Measurability of addition.* The measurability of addition on $(D, d_s) \times (D, d_s)$ holds because the Borel σ-field coincides with the Kolmogorov σ-field. It also follows from part of the proof of Theorem 4.1 of Whitt (1980). ∎

In Theorem 6.7.1 we showed that linear functions of the coordinates are Lipschitz in the SM_1 metric. We now apply Theorem 6.5.1 to show that convergence in the SM_1 topology is characterized by convergence of all such linear functions of the coordinates.

Theorem 6.7.2. (characterization of SM_1 convergence by convergence of all linear functions) *There is convergence $x_n \to x$ in $D([0,T],\mathbb{R}^k)$ as $n \to \infty$ in the SM_1 topology if and only if $\eta x_n \to \eta x$ in $D([0,T],\mathbb{R}^1)$ as $n \to \infty$ in the M_1 topology for all $\eta \in \mathbb{R}^k$.*

Proof. One direction is covered by Theorem 6.7.1. Suppose that $x_n \not\to x$ as $n \to \infty$ in SM_1. Then apply part (v) of Theorem 6.5.1 to deduce that $\eta x_n \not\to \eta x$ as $\eta \to \infty$ for some η. Note that $\|a\| > 0$ for $a \in \mathbb{R}^k$ if and only if $|\eta a| > 0$ in $\mathbb{R}$ for some $\eta \in \mathbb{R}^k$. Also, $\|a - A\| > 0$ for $A \subseteq R^k$ if and only if $|\eta a - \eta A| > 0$ in $\mathbb{R}$ for some $\eta \in \mathbb{R}^k$, where $\eta A = \{\eta b : b \in A\}$. ■

We can get convergence of sums under more general conditions than in Corollary 6.7.1. It suffices to have the jumps of x^i and y^i have common sign for all i. We can express this property by the condition

$$(x^i(t) - x^i(t-))(y^i(t) - y^i(t-)) \geq 0 \tag{7.2}$$

for all t, $0 \leq t \leq T$, and all i, $1 \leq i \leq k$.

Theorem 6.7.3. (continuity of addition at limits with jumps of common sign) *If $x_n \to x$ and $y_n \to y$ in $D([0,T],\mathbb{R}^k, SM_1)$ and if condition* (7.2) *above holds, then*

$$x_n + y_n \to x + y \quad in \quad D([0,T],\mathbb{R}^k, SM_1) \ .$$

Proof. The proof is in the book.

6.7.2. Visits to Strips

In Sections (2.2.7)–(2.2.13) of Skorohod (1956), convenient characterizations of convergence in each topology are given for real-valued functions. We can apply Theorem 6.7.2 to develop associated characterizations for $\mathbb{R}^k$-valued functions. For each $x \in D([0,T],\mathbb{R}^1)$, $0 \leq t_1 < t_2 \leq T$ and, for each $a < b$ in $\mathbb{R}$, let $v_{t_1,t_2}^{a,b}(x)$ be the number of visits to the strip $[a,b]$ on the interval $[t_1,t_2]$; i.e., $v_{t_1,t_2}^{a,b}(x) = k$ if it is possible to find k (but not $k+1$) points t_i' such that $t_1 < t_1' < \cdots < t_k' \leq t_2$ such that either

$$x(t_1) \in [a,b], \ x(t_1') \notin [a,b], \ x(t_2') \in [a,b], \ldots,$$

or

$$x(t_1) \notin [a,b],\ x(t_1') \in [a,b],\ x(t_2') \notin [a,b], \ldots$$

We say that $x \in D([0,T],\mathbb{R})$ has a *local maximum* (*minimum*) *value at t relative to* (t_1, t_2) in $(0,T)$ if $t_1 < t < t_2$ and either

$$\text{(i)} \quad \sup\{x(s) : t_1 \le s \le t_2\} \le x(t) \quad (\inf\{x(s) : t_1 \le s \le t_2\} \ge x(t))$$

or

$$\text{(ii)} \quad \sup\{x(s) : t_1 \le s \le t_2\} \le x(t-) \quad (\inf\{x(s) : t_1 \le s \le t_2)\} \ge x(t-)) \ .$$

We say that x has a *local maximum* (*minimum*) *value* at t if it has a local maximum (minimum) value at t relative to some interval (t_1, t_2) with $t_1 < t < t_2$. We call local maximum and minimum values *local extreme values.*

Lemma 6.7.1. (local extreme values) *Any $x \in D([0,T],\mathbb{R})$ has at most countably many local extreme values.*

Proof. For each n, let $\{t_{n,i}\}$ be a finite collection of points in $[0,T]$, including 0 and T. Let $\{t_{n,i}\}$ be a subcollection of $\{t_{n+1}, i\}$ for each n and let the minimum distance between points in $\{t_{n,i}\}$ be ϵ_n, where $\epsilon_n \downarrow 0$ as $n \to \infty$. Note that there is one local maximum value and one local minimum value of x relative to the interval endpoints in each interval $[t_{n,i}, t_{n,i+1})$, where $t_{n,i}$ and $t_{n,i+1}$ are successive points in $\{t_{n,i}\}$. Hence the total number of extreme values of x relative to $\{t_{n,i}\}$ is countably infinite. Next note that any extreme value of x is contained in this set. To see this, suppose that b is an extreme value of x at t relative to the interval (t_1, t_2). Then, for sufficiently large n, there is an interval $(t_{n,i}, t_{n,i+1})$ such that $t_1 \le t_{n,i} < t < t_{n,i+1} \le t_2$, so that b is an extreme value of x within $(t_{n,i}, t_{n,i+1})$. ∎

If b is not a local extreme value of x, then x crosses level b whenever x hits b; i.e., if b is not a local extreme value and if $x(t) = b$ or $x(t-) = b$, then for every t_1, t_2 with $t_1 < t < t_2$ there exist t_1', t_2' with $t_1 < t_1'$, $t_2' < t_2$ such that $x(t_1') < b$ and $x(t_2') > b$. This property implies the following lemma.

Lemma 6.7.2. *Consider an interval $[t_1, t_2]$ with $0 < t_1 < t_2 < T$. If $x(t_i) \notin \{a, b\}$ for $i = 1, 2$ and a, b are not local extreme values of x, then x crosses one of the levels a and b at each of the $v_{t_1,t_2}^{a,b}(x)$ visits to the strip $[a,b]$ in $[t_1, t_2]$.*

Theorem 6.7.4. (characterization of SM_1 convergence in terms of convergence of number of visits to strips) *There is convergence $d_s(x_n, x) \to 0$ as $n \to \infty$ in $D([0,T], \mathbb{R}^k)$ if and only if*

$$v^{a,b}_{t_1,t_2}(\eta x_n) \to v^{a,b}_{t_1,t_2}(\eta x) \quad as \quad n \to \infty$$

for all $\eta \in \mathbb{R}^k$, all points $t_1, t_2 \in \{T\} \cup Disc(x)^c$ with $t_1 < t_2$ and almost all a, b with respect to Lebesgue measure.

Proof. By Theorem 6.7.2, it suffices to establish the result for $\mathbb{R}$-valued functions. First, suppose that $x_n \to x$ as $n \to \infty$ in $D([0,T], \mathbb{R}, M_1)$. Suppose that a and b are not local extreme values of x. Let $t_1, t_2 \in Disc(x)^c$ and suppose that $x(t_1), x(t_2) \notin \{a, b\}$. Then, for sufficiently large n, by Lemma 6.7.2, $v^{a,b}_{t_1,t_2}(x_n) = v^{a,b}_{t_1,t_2}(x)$. Since there are at most countably many "bad" a, b for any x, $v^{a,b}_{t_1,t_2}(x_n) \to v^{a,b}_{t_1,t_2}(x)$ for almost all a, b with respect to Lebesgue measure. On the other hand, suppose that $v^{a,b}_{t_1,t_2}(x_n) \to v^{a,b}_{t_1,t_2}(x)$ for all $t_1, t_2 \in Disc(x)^c$ and for almost all a, b. We will show that characterization (v) of SM_1 convergence in Theorem 6.5.1 holds. For x, t and $\epsilon > 0$ given, find η such that $v(x, [t-\eta, t)) < \epsilon/2$ and $v(x, [t, t+\eta]) < \epsilon/2$. First suppose that $t \in Disc(x)^c$. Then $v^{a,b}_{t_1,t_2}(x) = 0$ for t_1, $t_2 \in Disc(x)^c$, $t - \eta < t_1 < t < t_2 < t + \eta$ and all (a, b) with $a < x(t) - \epsilon/2 < x(t) + \epsilon/2 < b$. By assumption, for all suitably large n, $v^{a'b'}_{t_1,t_2}(x_n) = 0$ for some a', b' with

$$x(t) - \epsilon < a' < x(t) - \epsilon/2 < x(t) + \epsilon/2 < b' < x(t) + \epsilon\ .$$

By the argument above, we can show that, for a time interval before t, x_n and x are first in a neighborhood of $x(t-)$ and then leave. Afterwards, x_n and x enter the neighborhood of $x(t)$ and stay there for a short interval after t. To see this, let t_1 and t_2 be as above and then find a_1, b_1, a_2, b_2 such that

$$\begin{aligned} &x(t-) - \epsilon < a_1 < x(t-) - \epsilon/2,\ x(t-) + \epsilon/2 < b_1 < x(t) + \epsilon \\ &x(t) - \epsilon < a_2 < x(t) - \epsilon/2,\ x(t) + \epsilon/2 < b_2 < x(t) + \epsilon\ , \end{aligned}$$

$v^{a_1,b_1}_{t_1,t_2}(x_n) \to v^{a_1,b_1}_{t_1,t_2}(x) = 1$ and $v^{a_2,b_2}_{t_1,t_2}(x_n) \to v^{a_2,b_2}_{t_1,t_2}(x) = 1$. that implies that $v(x_n, x, t, \delta) < \epsilon$ for $\delta < \min\{|t-t_1|, |t-t_2|\}$. Next suppose that $t \in Disc(x)$. Let t_1, t_2 be as above. Find a_1, b_1, a_2, b_2 such that

$$x(t-) - \epsilon < a_1 < x(t-) - \epsilon/2 < x(t-) + \epsilon/2 < b_1 < x(t-) + \epsilon\,,$$

$$x(t) - \epsilon < a_2 < x(t) - \epsilon/2 < x(t) + \epsilon/2 < b < x(t) + \epsilon\,,$$

$v_{t_1,t_2}^{a_1,b_1}(x_n) \to v_{t_1,t_2}^{a_1,b_1}(x) = 1$ and $v_{t_1,t_2}^{a_2,b_2}(x_n) \to v_{t_1,t_2}^{a_2,b_2}(x) = 1$. It remains to show that x_n cannot fluctuate significantly between $x(t-)$ and $x(t)$. To be definite, suppose that $x(t-) < x(t)$ and suppose that $\epsilon < x(t) - x(t-)$. Then for almost all a, b with

$$x(t-) + \epsilon/2 < a < b < x(t) - \epsilon/2 \ ,$$

$$v_{t_1,t_2}^{a,b}(x_n) \to v_{t_1,t_2}^{a,b}(x) = 2 \quad \text{as} \quad n \to \infty \ .$$

That implies that $w_s(x_n, x, t, \delta) \to 0$ as $n \to \infty$ for $\delta < \min\{|t_1 - t|, |t - t_2|\}$, which completes the proof. ∎

6.8. Topological Completeness

In this section we exhibit a complete metric topologically equivalent to the incomplete metric d_s in (3.7) inducing the SM_1 topology. Since a product metric defined as in (3.13) inherits the completeness of the component metrics, we also succeed in constructing complete metrics inducing the associated product topology. We make no use of the complete metrics beyond showing that the topology is topologically complete. Another approach to topological completeness would be to show that D is homeomorphic to a G_δ subset of a complete metric space, as noted in Section 11.2 of the book.

In our construction of complete metrics, we follow the argument used by Prohorov (1956, Appendix 1) to show that the J_1 topology is topologically complete; we incorporate an oscillation function into the metric. For M_1, we use $w_s(x, \delta)$ in (5.1). Since $w_s(x, \delta) \to 0$ as $\delta \to 0$ for each $x \in D$, we need to appropriately "inflate" differences for small δ. For this purpose, let

$$\hat{w}_s(x, z) \equiv \begin{cases} w_s(x, e^z), & z < 0 \\ \\ w_s(x, 1), & z \geq 1 \ . \end{cases} \tag{8.1}$$

Since $w_s(x, \delta)$ is nondecreasing in δ, $\hat{w}_s(x, z)$ is nondecreasing in z. Note that $\hat{w}_s(x, z)$ as a function of z has the form of a cumulative distribution function (cdf) of a finite measure. On such cdf's, the Lévy metric λ is known to be a complete metric inducing the topology of pointwise convergence at all continuity points of the limit; i.e.,

$$\lambda(F_1, F_2) \equiv \inf\{\epsilon > 0 : F_2(x - \epsilon) - \epsilon \leq F_1(x) \leq F_2(x + \epsilon) + \epsilon\} \ . \tag{8.2}$$

The Helly selection theorem, p. 267 of Feller (1971), can be used to show that the metric λ is complete.

Thus, our new metric is

$$\hat{d}_s(x_1, x_2) \equiv d_s(x_1, x_2) + \lambda(\hat{w}_s(x_1, \cdot), \hat{w}_s(x_2, \cdot)) \ . \tag{8.3}$$

Theorem 6.8.1. (a complete SM_1 metric) *The metric $\hat{d}_s$ on D in (8.3) is complete and topologically equivalent to d_s.*

Proof. To show topological equivalence of $\hat{d}_s$ and d_s, it suffices to show that $\lambda(\hat{w}_s(x_n, \cdot), \hat{w}_s(x, \cdot)) \to 0$ as $n \to \infty$ whenever $d_s(x_n, x) \to 0$ as $n \to \infty$. However, if $d_s(x_n, x) \to 0$ as $n \to \infty$, then $w_s(x_n, \delta) \to w_s(x, \delta)$ as $n \to \infty$ at all δ which are continuity points of $w_s(x, \delta)$. (See Lemma 6.8.1 below.) That in turn implies that $\hat{w}_s(x_n, z) \to \hat{w}_s(x, z)$ as $n \to \infty$ for all z which are continuity points of $\hat{w}_s(x, z)$. However, such convergence is equivalent to convergence under λ. Next, suppose that a sequence $\{x_n\}$ is fundamental under $\hat{d}_s$, i.e., $\hat{d}_s(x_m, x_n) \to 0$ as $m, n \to \infty$. It follows that $\{x_n(t) : 0 \leq t \leq T, n \geq 1\}$ is compact. Hence, there exists a countable dense set N of $[0, T]$, including 0 and T, and a subsequence $\{x_{n_k}\}$ such that $x_{n_k}(t) \to x(t)$ as $n_k \to \infty$ for all $t \in N$, where x is some $\mathbb{R}^k$-valued function on $[0, T]$. At the same time, since λ is known to be a complete metric, there must exist a distribution function F such that

$$\lim_{n\to\infty} \lambda(\hat{w}_s(x_n, \cdot), F) = 0 \ ,$$

which implies that

$$\lim_{\delta\to 0} \ \overline{\lim_{n\to\infty}} \ w_s(x_n, \delta) = 0 \ .$$

However, Theorem **??** and Corollary **??** imply that there exists $\bar{x} \in D$ (with $\bar{x}$ not necessarily x) such that $d_s(x_{n_k}, \bar{x}) \to 0$ as $n_k \to \infty$. Since $d_s(x_n, \bar{x}) \leq d_s(x_n, x_{n_k}) + d_s(x_{n_k}, \bar{x})$ and $d_s(x_m, x_n) \to 0$ as $m, n \to \infty$, $d_s(x_n, \bar{x}) \to 0$ as $n \to \infty$. ■

To complete the proof of Theorem 6.8.1, we need the following lemma.

Lemma 6.8.1. (continuity of SM_1 modulus) *If $d_s(x_n, x) \to 0$ as $n \to \infty$, then $w_s(x_n, \delta) \to w_s(x, \delta)$ as $n \to \infty$ for each δ that is a continuity point of $w_s(x, \delta)$.*

Proof. Let δ be a continuity point of $w_s(x, \delta)$. Then, for each $\epsilon_1 > 0$, there is $\epsilon_2 > 0$ such that $w_s(x, \delta - \epsilon_2) \geq w_s(x, \delta) - \epsilon_1$. For δ, ϵ_1 and ϵ_2 given, it is possible to choose continuity points t, t_1, t_2 and t_3 of x such that

$$(t - \delta) \vee 0 \leq t_1 \leq t_2 \leq t_3 \leq (t + \delta) \wedge T \tag{8.4}$$

and

$$\|x(t_2) - [x(t_1), x(t_3)]\| \ge w_s(x, \delta - \epsilon_2) - \epsilon_1 \ge w_s(x, \delta) - 2\epsilon_1 \ .$$

Since $d_s(x_n, x) \to 0$ as $n \to \infty$, $x_n(t_j) \to x(t_j)$ as $n \to \infty$ for $j = 1, 2, 3$. Hence, there exists n_0 such that, for all $n \ge n_0$,

$$\|x_n(t_2) - [x_n(t_1), x_n(t_3)]\| \ge w_s(x, \delta) - 3\epsilon_2 \ .$$

However,

$$w_s(x_n, \delta) \ge \|x_n(t_2) - [x_n(t_1), x_n(t_3)]\| \ ,$$

so that $w_s(x_n, \delta) \ge w_s(x, \delta) - 3\epsilon_2$. Since ϵ_2 can be made arbitrarily small,

$$\underline{\lim}_{n \to \infty} \ w_s(x_n, \delta) \ge w_s(x, \delta) \ . \tag{8.5}$$

We now establish an inequality in the other direction. Since δ is a continuity point of $w_s(x, \delta)$, for any $\epsilon_1 > 0$ there exists $\epsilon_2 > 0$ so that $w_s(x, \delta + \epsilon_2) \le w_s(x, \delta) + \epsilon_1$. We can choose t_n, t_{n1}, t_{n2} and t_{n3} so that

$$(t_n - \delta) \vee 0 \le t_{n1} \le t_{n2} \le t_{n3} \le (t_n + \delta) \wedge T$$

and

$$\|x_n(t_{n2}) - [x_n(t_{n1}), x_n(t_{n3})]\| \ge w_s(x_n, \delta) - \epsilon_2$$

for all n. There thus exists a subsequence $\{n_k\}$ such that $t_{n_k} \to t$ and $t_{n_k j} \to t_j$, $j = 1, 2, 3$, (8.4) holds and $\|x_{n_k}(t_{n_k j}) - z_j\| \to 0$ as $n_k \to \infty$. Moreover, since x and x_n, $n \ge 1$, are right-continuous for all n, we can have t_1, t_2 and t_3 be continuity points of x with

$$(t - (\delta + \epsilon_2)) \vee 0 \le t_1 \le t_2 \le t_3 \le (t + (\delta + \epsilon_2)) \wedge T \ .$$

Then $\|x_{n_k}(t_{n_k j}) - x(t_j)\| \to 0$ as $n_k \to \infty$. Hence, there is n_0 such that, for all $n_k \ge n_0$,

$$\begin{aligned} \|x(t_2) - [x(t_1), x(t)_3)]\| &\ge \|x_{n_k}(t_{n_k 2}) - [x_{n_k}(t_{n_k 1}), x_{n_k}(t_{n_k 3})]\| - \epsilon_2 \\ &\ge w_s(x_n, \delta) - 2\epsilon_2 \ . \end{aligned} \tag{8.6}$$

However,

$$w_s(x, \delta) + \epsilon_1 \ge w_s(x, \delta + \epsilon_2) \ge \|x(t_2) - [x(t_1), x(t_3)]\| \ . \tag{8.7}$$

Combining (8.6) and (8.7), we obtain

$$w_s(x, \delta) \ge w_s(x_n, \delta) - \epsilon_1 - 2\epsilon_2 \ .$$

Since ϵ_1 and ϵ_2 can be made arbitrarily small,

$$\overline{\lim_{n\to\infty}} \; w_s(x_n, \delta) \le w_s(x, \delta) \; . \tag{8.8}$$

Combining (8.5) and (8.8) completes the proof. ∎

6.9. NonCompact Domains

It is often convenient to consider the function space $D([0,\infty), \mathbb{R}^k)$ with domain $[0,\infty)$ instead of $[0,T]$. More generally, we may consider the function space $D(I, \mathbb{R}^k)$, where I is a subinterval of the real line. Common cases besides $[0,\infty)$ are $(0,\infty)$ and $(-\infty,\infty) \equiv \mathbb{R}$.

Given the function space $D(I, \mathbb{R}^k)$ for any subinterval I, we define convergence $x_n \to x$ with some topology to be convergence in $D([a,b], \mathbb{R}^k)$ with that same topology for the restrictions of x_n and x to the compact interval $[a,b]$ for all points a and b that are elements of I and either boundary points of I or are continuity points of the limit function x. For example, for $I = [c,d)$ with $-\infty < c < d < \infty$, we include $a = c$ but exclude $b = d$; for $I = [c,d]$, we include both c and d.

For simplicity, we henceforth consider only the special case in which $I = [0,\infty)$. In that setting, we can equivalently define convergence $x_n \to x$ as $n \to \infty$ in $D([0,\infty), \mathbb{R}^k)$ with some topology to be convergence $x_n \to x$ as $n \to \infty$ in $D([0,t], \mathbb{R}^k)$ with that topology for the restrictions of x_n and x to $[0,t]$ for $t = t_k$ for each t_k in some sequence $\{t_k\}$ with $t_k \to \infty$ as $k \to \infty$, where $\{t_k\}$ can depend on x. It suffices to let t_k be continuity points of the limit function x; for the J_1 topology, see Lindvall (1973),

Whitt (1980) and Jacod and Shiryaev (1987). We will discuss only the SM_1 topology here, but the discussion applies to the other nonuniform topologies as well. We also will omit most proofs.

As a first step, we consider the case of closed bounded intervals $[t_1, t_2]$. The space $D([t_1,t_2], \mathbb{R}^k)$ is essentially the same as (homeomorphic to) the space $D([0,T], \mathbb{R}^k)$ already studied, but we want to look at the behavior as we change the interval $[t_1,t_2]$. For $[t_3,t_4] \subseteq [t_1,t_2]$, we consider the restriction of x in $D([t_1,t_2], \mathbb{R}^k)$ to $[t_3,t_4]$, defined by

$$r_{t_3,t_4} : D([t_1,t_2], \mathbb{R}^k) \to D([t_3,t_4], \mathbb{R}^k)$$

with $r_{t_3,t_4}(x)(t) = x(t)$ for $t_3 \le t \le t_4$. Let d_{t_1,t_2} be the metric d_s on $D([t_1,t_2], \mathbb{R}^k)$. We want to relate the distance $d_{t_1,t_2}(x_1,x_2)$ and convergence $d_{t_1,t_2}(x_n,x) \to 0$ as $n \to \infty$ for different domains. We first state a result enabling us to go from the domains $[t_1, t_2]$ and $[t_2, t_3]$ to $[t_1, t_3]$ when $t_1 < t_2 < t_3$.

Lemma 6.9.1. (metric bounds) *For* $0 \le t_1 < t_2 < t_3$ *and* x_1, $x_2 \in D([t_1, t_3], \mathbb{R}^k)$,

$$d_{t_1,t_3}(x_1, x_2) \le d_{t_1,t_2}(x_1, x_2) \vee d_{t_2,t_3}(x_1, x_2) \ .$$

We now observe that there is an equivalence of convergence provided that the internal boundary point is a continuity point of the limit function.

Lemma 6.9.2. *For* $0 \le t_1 < t_2 < t_3$ *and* x, $x_n \in D([t_1, t_3], \mathbb{R}^k)$, *with* $t_2 \in Disc(x)^c$, $d_{t_1,t_3}(x_n, x) \to 0$ *as* $n \to \infty$ *if and only if* $d_{t_1,t_2}(x_n, x) \to 0$ *and* $d_{t_2,t_3}(x_n, x) \to 0$ *as* $n \to \infty$.

For $x \in D([0, T], \mathbb{R}^k)$ and $0 \le t_1 < t_2 \le T$, let $r_{t_1,t_2} : D([0, T], \mathbb{R}^k) \to D([t_1, t_2], \mathbb{R}^k)$ be the restriction map, defined by $r_{t_1,t_2}(x)(s) = x(s)$, $t_1 \le s \le t_2$.

Corollary 6.9.1. (continuity of restriction maps) *If* $x_n \to x$ *as* $n \to \infty$ *in* $D([0, T], \mathbb{R}^k, SM_1)$ *and if* $t_1, t_2 \in Disc(x)^c$, *then*

$$r_{t_1,t_2}(x_n) \to r_{t_1,t_2}(x) \quad as \quad n \to \infty \quad in \quad D([t_1, t_2], \mathbb{R}^k, SM_1) \ .$$

Let $r_t : D([0, \infty), \mathbb{R}^k) \to D([0, t], \mathbb{R}^k)$ be the *restriction map* with $r_t(x)(s) = x(s)$, $0 \le s \le t$. Suppose that $f : D([0, \infty), \mathbb{R}^k) \to D([0, \infty), \mathbb{R}^k)$ and $f_t : D([0, t], \mathbb{R}^k) \to D([0, t], \mathbb{R}^k)$ for $t > 0$ are functions with

$$f_t(r_t(x)) = r_t(f(x))$$

for all $x \in D([0, \infty), \mathbb{R}^k)$ and all $t > 0$. We then call the functions f_t *restrictions of the function* f.

Theorem 6.9.1. (continuity from continuous restrictions) *Suppose that* $f : D([0, \infty), \mathbb{R}^k) \to D([0, \infty), \mathbb{R}^l)$ *has continuous restrictions* f_t *with some topology for all* $t > 0$. *Then* f *itself is continuous in that topology.*

Proof. Suppose that $x_n \to x$ as $n \to \infty$ in the specified topology. That means that $r_{t_m}(x_n) \to r_{t_m}(x)$ as $n \to \infty$ for some sequence $\{t_m\}$ with $t_m \to \infty$, possibly depending on x. Since f has continuous restrictions,

$$r_{t_m}(f(x_n)) = f_{t_m}(r_{t_m}(x_n)) \to f_{t_m}(r_{t_m}(x)) = r_{t_m}(f(x))$$

as $n \to \infty$ for all m, which implies that $f(x_n) \to f(x)$ as $n \to \infty$ in the specified topology. ∎

No more material has been deleted from Section 12.9 of the book.

6.10. Strong and Weak M_2 Topologies

We now define strong and weak versions of Skorohod's M_2 topology. In Section 6.11 we will show that it is possible to define the M_2 topologies by a minor modification of the definitions in Section 6.3, in particular, by simply using parametric representations in which only r is nondecreasing instead of (u, r), but now we will use Skorohod's (1956) original approach, and relate it to the Hausdorff metric on the space of graphs.

The weak topology will be defined just like the strong, except it will use the thick graphs G_x instead of the thin graphs Γ_x. In particular, let

$$\mu_s(x_1, x_2) \equiv \sup_{(z_1,t_1)\in\Gamma_{x_1}} \inf_{(z_2,t_2)\in\Gamma_{x_2}} \{\|(z_1, t_1) - (z_2, t_2)\|\} \tag{10.1}$$

and

$$\mu_w(x_1, x_2) \equiv \sup_{(z_1,t_1)\in G_{x_1}} \inf_{(z_2,t_2)\in G_{x_2}} \{\|(z_1, t_1) - (z_2, t_2)\|\} . \tag{10.2}$$

Following Skorohod (1956), we say that $x_n \to x$ as $n \to \infty$ for a sequence or net $\{x_n\}$ in the strong M_2 topology, denoted by SM_2 if $\mu_s(x_n, x) \to 0$ as $n \to \infty$. Paralleling that, we say that $x_n \to x$ as $n \to \infty$ in the weak M_2 topology, denoted by WM_2, if $\mu_w(x_n, x) \to 0$ as $n \to \infty$. We say that $x_n \to x$ as $n \to \infty$ in the product topology if $\mu_s(x_n^i, x^i) \to 0$ (or equivalently $\mu_w(x_n^i, x^i) \to 0$) as $n \to \infty$ for each i, $1 \le i \le k$.

We can also generate the SM_2 and WM_2 topologies using the Hausdorff metric in equation 5.2 of Section 11.5 in the book. As in equation (5.4) in Section 11.5 of the book, for $x_1, x_2 \in D$,

$$m_s(x_1, x_2) \equiv m_H(\Gamma_{x_1}, \Gamma_{x_2}) = \mu_s(x_1, x_2) \vee \mu_s(x_2, x_1) , \tag{10.3}$$

$$m_w(x_1, x_2) \equiv m_H(G_{x_1}, G_{x_2}) = \mu_w(x_1, x_2) \vee \mu_w(x_2, x_1) \tag{10.4}$$

and

$$m_p(x_1, x_2) \equiv \max_{1\le i\le k} m_s(x_1^i, x_2^i) . \tag{10.5}$$

We will show that the metric m_s induces the SM_2 topology.

That will imply that the metric m_p induces the associated product topology. However, it turns out that the metric m_w does *not* induce the WM_2 topology. We will show that the WM_2 topology coincides with the product topology, so that the Hausdorff metric can be used to define the WM_2 topology via m_p in (10.5).

Closely paralleling the d or M_1 metrics, we have $m_p \le m_s$ on $D([0,T],\mathbb{R}^k)$ and $m_p = m_w = m_s$ on $D([0,T],\mathbb{R}^1)$. Just as with d, we use m without subscript when the functions are real valued. Example **??**, which showed that WM_1 is strictly weaker than SM_1 also shows that WM_2 is strictly weaker than SM_2. Example **??** shows that the SM_2 topology is strictly weaker than the SM_1 topology.

Note that μ_s in (10.1) is *not* symmetric in its two arguments. Example 12.10.1 of the book shows that if $\mu_s(x,x_n) \to 0$ as $n \to \infty$, we need not have $\mu_s(x_n,x) \to 0$ as $n \to \infty$.

6.10.1. The Hausdorff Metric Induces the SM_2 Topology

We now show that m_s induces the SM_2 topology.

Theorem 6.10.1. (the Hausdorff metric m_s induces the SM_2 topology) *If $\mu_s(x_n,x) \to 0$ as $n \to \infty$, then $\mu_s(x,x_n) \to 0$ as $n \to \infty$. Hence, $\mu_s(x_n,x) \to 0$ as $n \to \infty$ if and only if $m_s(x_n,x) \to 0$ as $n \to \infty$.*

Proof. Our proof will exploit lemmas below. Suppose that $\mu_s(x_n,x) \to 0$ but $\mu_s(x,x_n) \not\to 0$ as $n \to \infty$. Since $\mu_s(x,x_n) \not\to 0$, there exists $(z,t) \in \Gamma_x$ for which it is not possible to find $(z_n,t_n) \in \Gamma_{x_n}$ for $n \ge 1$ such that $(z_n,t_n) \to (z,t)$ as $n \to \infty$, but that contradicts Lemma 6.10.4 below. ∎

In order to complete the proof of Theorem 6.10.1, we prove the following four lemmas.

Lemma 6.10.1. *Suppose that $\mu_s(x_n,x) \to 0$ as $n \to \infty$. If $(z_n,t_n) \in \Gamma_{x_n}$ for $n \ge 1$, then there exists a subsequence $\{(z_{n_k},t_{n_k})\}$ with $(z_{n_k},t_{n_k}) \to (z,t)$ as $n_k \to \infty$ for some $(z,t) \in \Gamma_x$. Moreover, the limits of all convergent subsequences must be in Γ_x.*

Proof. Suppose that $\mu_s(x_n,x) \to 0$ as $n \to \infty$ and consider any sequence $\{(z_n,t_n)\}$ with $(z_n,t_n) \in \Gamma_{x_n}$ for $n \ge 1$. By the definition of μ_s, there must exist $(z_n',t_n') \in \Gamma_x$ such that $\|(z_n,t_n) - (z_n',t_n')\| \to 0$ as $n \to \infty$. Since Γ_x is compact, there exists a convergent subsequence of the sequence $\{(z_n',t_n')\}$; i.e., there exists $\{(z_{n_k}',t_{n_k}')\}$ such that $(z_{n_k}',t_{n_k}') \to (z,t)$ for some $(z,t) \in \Gamma_x$. By the triangle inequality, we must also have $(z_{n_k},t_{n_k}) \to (z,t)$ as $n_k \to \infty$. Finally, suppose (z_{n_k},t_{n_k}) is an arbitrary convergent subsequence of $\{(z_n,t_n)\}$. By the argument above, there exists $(z,t) \in \Gamma_x$ such that a subsequence $(z_{n_{k_j}},t_{n_{k_j}}) \to (z,t)$ as $n_{k_j} \to \infty$. This implies that (z,t) must be the limit of the convergent subsequence $\{(z_{n_k},t_{n_k})\}$. ∎

Lemma 6.10.2. *Suppose that $\mu_s(x_n, x) \to 0$ as $n \to \infty$, $t \notin Disc(x)$ and $(z_n, t) \in \Gamma_{x_n}$ for $n \geq 1$. Then $z_n \to x(t)$ as $n \to \infty$.*

Proof. By Lemma 6.10.1, there is a subsequence $(z_{n_k}, t) \to (z, t) \in \Gamma_x$, but $z = x(t)$ for $(z, t) \in \Gamma_x$ because $t \notin Disc(x)$. Since all convergent subsequences must have the same limit, $z_n \to z = x(t)$ as $n \to \infty$. ∎

Corollary 6.10.1. *If $t \notin Disc(x)$ and $\mu_s(x_n, x) \to 0$ as $n \to \infty$, then $x_n(t) \to x(t)$ and $x_n(t-) \to x(t)$ in $\mathbb{R}^k$ as $n \to \infty$.*

Lemma 6.10.3. *If $\mu_s(x_n, x) \to 0$ as $n \to \infty$ and $(z, t) \in \Gamma_x$, then for any i, $1 \leq i \leq k$, there exist $(z_n, t_n) \in \Gamma_{x_n}$ for $n \geq 1$ such that $|z_n^i - z^i| \vee |t_n - t| \to 0$.*

Proof. The conclusion follows from Corollary 6.10.1 if $t \notin Disc(x)$, so suppose that $t \in Disc(x)$. Then z belongs to the segment $[x(t-), x(t)]$. First choose $t'_m > t$ with $t'_m \notin Disc(x)$ for all m and $t'_m \downarrow t$ as $m \to \infty$. By Lemma 6.10.2, there exist $(z'_{m,n}, t'_m) \in \Gamma_{x_n}$ such that $z'_{m,n} \to x(t'_m)$ as $n \to \infty$. Next choose $t''_m < t$ with $t''_m \notin Disc(x)$ for all m and $t''_m \uparrow t$ as $m \to \infty$. By Lemma 6.10.2 again, there exist $(z''_{m,n}, t''_m) \in \Gamma_{x_n}$ such that $z''_{m,n} \to x(t''_m)$ as $n \to \infty$. The diagonal sequences $(z'_{n,n}, t'_n)$ and $(z''_{n,n}, t''_n)$ thus belong to Γ_{x_n} and satisfy $t'_n \downarrow t$, $t''_n \uparrow t$, $z'_{n,n} \to x(t)$ and $z''_{n,n} \to x(t-)$ as $n \to \infty$. Since $\Gamma_{x_n^i}$ is a continuous real-valued curve, every value in the segment $[z'^i_{n,n}, z''^i_{n,n}]$ is realized for some t'''_n with $t''_n \leq t'''_n \leq t'_n$. Hence, for any $(z, t) \in \Gamma_x$, there exists $(z'''_n, t'''_n) \in \Gamma_{x_n}$ such that $(z'''^i_n, t'''_n) \to (z^i, t)$ as $n \to \infty$. ∎

Lemma 6.10.4. *If $\mu_s(x_n, x) \to 0$ as $n \to \infty$ and $(z, t) \in \Gamma_x$, then there exist $(z_n, t_n) \in \Gamma_{x_n}$ for $n \geq 1$ such that $\|(z_n, t_n) - (z, t)\| \to 0$ as $n \to \infty$.*

Proof. If $t \notin Disc(x)$, then we can take $(x_n(t), t) \in \Gamma_{x_n}$ or $(x_n(t-), t) \in \Gamma_{x_n}$ by Corollary 6.10.1. Hence it suffices to assume that $t \in Disc(x)$. Then, by the first part of the proof of Lemma 6.10.3, it suffices to consider (z, t) with $z \neq x(t)$ and $z \neq x(t-)$. For at least one coordinate i, either $x^i(t-) < z < x^i(t)$ or $x^i(t) > z > x^i(t)$. Consider one such coordinate. By Lemma 6.10.3, there is $(z_n, t_n) \in \Gamma_{x_n}$ such that $t_n \to t$ and $z_n^i \to z^i$ as $n \to \infty$. Moreover, since $\mu_s(x_n, x) \to 0$, given $(z_n, t_n) \in \Gamma_{x_n}$, we must have $(z'_n, t'_n) \in \Gamma_x$ such that $\|z_n - z'_n\| \vee |t_n - t'_n| \to 0$. Since $t_n \to t$, we must also have $t'_n \to t$. Since $z_n^i \to z^i$ and Γ_x contains the line joining $(x(t-), t)$ and $(x(t), t)$, we must have $z'_n \to z$ as well, which implies that $z_n \to z$, establishing the desired conclusion. ∎

6.10.2. WM_2 is the Product Topology

We now observe that m_p induces the WM_2 topology.

Theorem 6.10.2. (WM_2 is the product topology) *$\mu_w(x_n, x) \to 0$ as $n \to \infty$ for μ_w in (10.2) if and only if $m_p(x_n, x) \to 0$ as $n \to \infty$ for m_p in (10.5), so that the WM_2 topology on $D([0,T], \mathbb{R}^k)$ coincides with the product topology.*

Proof. First, if $\mu_w(x_n, x) \to 0$ as $n \to \infty$, then $\mu_w(x_n^i, x^i) \to 0$ for each i, but $\mu_w(x_n^i, x^i) = \mu_s(x_n^i, x^i)$, so that $\mu_s(x^i, x_n^i) \to 0$ and $m_p(x_n, x) \to 0$ by Theorem 6.10.1. Conversely, suppose that $m_p(x_n, x) \to 0$ as $n \to \infty$. Lemma 6.10.1 implies that $\cup_{n \ge 1} \Gamma_{x_n^i}$ is compact for each i, $1 \le i \le k$. That in turn implies that $\cup_{n \ge 1} G_{x_n}$ is compact. Hence, if $(z_n, t_n) \in G_{x_n}$ for $n \ge 1$, then every subsequence necessarily has a convergent subsubsequence. To have $\mu_w(x_n, x) \not\to 0$, we must have a subsequence of $\{(z_n, t_n)\}$ converge to a limit not in G_x. We will show that is not possible. Consider $(z_n, t_n) \in G_{x_n}$, $n \ge 1$. Since $t_n \in [0, T]$ for all n, there exists a subsequence (z_{n_k}, t_{n_k}) such that $t_{n_k} \to t$ for some t, $0 \le t \le T$. Since $m_p(x_n, x) \to 0$ as $n \to \infty$, there is a subsequence $\{(z_{n_{k_j}}, t_{n_k})\}$ such that $z_{n_{k_j}}^i \to z^i$ for some z^i where $(z^i, t) \in \Gamma_{x^i}$. Moreover, there are such subsequences for all i, $1 \le i \le k$, so that $z_n^i \to z^i$ for all i along the final subsequence. Moreover, $(z^i, t) \in \Gamma_{x^i}$ for all i, but this implies that $(z, t) \in G_x$. Hence every subsequence of (z_n, t_n) has a convergent subsubsequence and every convergent subsequence of $\{(z_n, t_n)\}$ has limit $(z, t) \in G_x$. That implies that $\mu_w(x_n, x) \to 0$ as $n \to \infty$. ∎

6.11. Alternative Characterizations of M_2 Convergence

We now give alternative characterizations of the SM_2 and WM_2 topologies.

6.11.1. M_2 Parametric Representations

We first observe that the SM_2 and WM_2 topologies can be defined just like the SM_1 and WM_1 topologies in Section 6.3. For this purpose, we say that a *strong M_2 (SM_2) parametric representation* of x is a continuous function (u, r) mapping $[0, 1]$ onto Γ_x such that r is nondecreasing. A *weak M_2 (WM_2) parametric representation* of x is a continuous function mapping $[0, 1]$ into G_x such that r is nondecreasing with $r(0) = 0$, $r(1) = T$ and $u(1) = x(T)$. The corresponding M_1 parametric representations are nondecreasing using the order defined on the graphs Γ_x and G_x in Section

2. In contrast, only the component function r is nondecreasing in the M_2 parametric representations. Let $\Pi_{s,2}(x)$ and $\Pi_{w,2}(x)$ be the sets of all SM_2 and WM_2 parametric representations of x.

Paralleling (3.7) and (3.8), define the distance functions

$$d_{s,2}(x_1, x_2) \equiv \inf_{\substack{(u_j,r_j)\in\Pi_{s,2}(x_j)\\ j=1,2}} \{\|u_1 - u_2\| \vee \|r_1 - r_2\|\} \tag{11.1}$$

and

$$d_{w,2}(x_1, x_2) \equiv \inf_{\substack{(u_j,r_j)\in\Pi_{w,2}(x_j)\\ j=1,2}} \{\|u_1 - u_2\| \vee \|r_1 - r_2\|\} \; . \tag{11.2}$$

We then can say that $x_n \to x$ as $n \to \infty$ for a sequence or net $\{x_n\}$ if $d_{s,2}(x_n, x) \to 0$ or $d_{w,2}(x_n, x) \to 0$ as $n \to \infty$. A difficulty with this approach, just as for the WM_1 topology, is that neither $d_{s,2}$ nor $d_{w,2}$ is a metric.

6.11.2. SM_2 Convergence

We now establish the equivalence of several alternative characterizations of convergence in the SM_2 topology. To have a characterization involving the local behavior of the functions, we use the uniform-distance function $\bar{w}_s(x, x_2, t, \delta)$ in (4.6). We also use the related uniform-distance functions

$$\bar{w}_s(x_1, x_2, \delta) \equiv \sup_{0\le t\le T} \bar{w}(x_1, x_2, t, \delta) \; . \tag{11.3}$$

$$\bar{w}_s^*(x_1, x_2, t, \delta) \equiv \|x_1(t) - [x_2((t-\delta)\vee 0), x_2((t+\delta)\wedge T)]\| \tag{11.4}$$

$$\bar{w}_s^*(x_1, x_2, \delta) \equiv \sup_{0\le t\le T} \bar{w}_s^*(x_1, x_2, t, \delta) \; . \tag{11.5}$$

We now define new oscillation functions. The first is

$$\bar{w}_s^*(x, t, \delta) \equiv \sup\{\|x(t) - [x(t_1), x(t_2)]\|\} \; , \tag{11.6}$$

where the supremum is over

$$0 \vee (t-\delta) \le t_1 \le [0 \vee (t-\delta)] + \delta/2 \text{ and } [T \wedge (t+\delta)] - \delta/2 \le t_2 \le (t+\delta) \wedge T.$$

The second is

$$\bar{w}_s^*(x, \delta) \equiv \sup_{0\le t\le T} \bar{w}_s^*(x, t, \delta) \; . \tag{11.7}$$

The uniform-distance function $\bar{w}_s^*(x_1, x_2, \delta)$ in (11.5) and the oscillation function $\bar{w}_s^*(x, \delta)$ in (11.7) were originally used by Skorohod (1956).

As before, T need not be a continuity point of x in $D([0,T],\mathbb{R}^k)$. Unlike for the M_1 topology, we can have $x_n \to x$ in (D, M_2) without having $x_n(T) \to x(T)$.

Let $v(x, A)$ represent the oscillation of x over the set A as in (2.5).

Theorem 6.11.1. (characterizations of SM_2 convergence) *The following are equivalent characterizations of $x_n \to x$ as $n \to \infty$ in (D, SM_2):*

(i) $d_{s,2}(x_n, x) \to 0$ *as* $n \to \infty$ *for* $d_{s,2}$ *in* (11.1); *i.e., for any* $\epsilon > 0$ *and* n *sufficiently large, there exist* $(u, r) \in \Pi_{s,2}(x)$ *and* $(u_n, r_n) \in \Pi_{s,2}(x_n)$ *such that* $\|u_n - u\| \vee \|r_n - r\| < \epsilon$.

(ii) $m_s(x_n, x) \to 0$ *as* $n \to \infty$ *for the metric* m_s *in* (10.3).

(iii) $\mu_s(x_n, x) \to 0$ *as* $n \to \infty$ *for* μ_s *in* (10.1).

(iv) Given $\bar{w}_s(x_1, x_2, \delta)$ *defined in* (11.3),

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ \bar{w}_s(x_n, x, \delta) = 0 \ . \tag{11.8}$$

(v) For each t, $0 \le t \le T$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ \bar{w}_s(x_n, x, t, \delta) = 0 \tag{11.9}$$

for $\bar{w}_s(x_1, x_2, t, \delta)$ *in* (4.6).

(vi) For all $\epsilon > 0$ *and all* n *sufficiently large, there exist finite ordered subsets* A *of* Γ_x *and* A_n *of* Γ_{x_n}, *as in* (3.9) *where* $(z_1, t_1) \le (z_2, t_2)$ *if* $t_1 \le t_2$, *of the same cardinality such that* $\hat{d}(A, \Gamma_x) < \epsilon$, $\hat{d}(A_n, \Gamma_{x_n}) < \epsilon$ *and* $d^*(A, A_n) < \epsilon$ *for* $\hat{d}$ *in* (3.10) *and* d^* *in* (5.6).

(vii) Given $\bar{w}_s^*(x_1, x_2, \delta)$ *defined in* (11.5),

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ \bar{w}_s^*(x_n, x, \delta) = 0 \ .$$

(viii) $x_n(t) \to x(t)$ *as* $n \to \infty$ *for each* t *in a dense subset of* $[0, T]$ *including* 0 *and*

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ \bar{w}_s^*(x_n, \delta) = 0 \tag{11.10}$$

for $\bar{w}_s^*(x, \delta)$ *in* (11.7).

Proof. We already have shown the equivalence (ii)↔(iii) in Theorem 11.10.1.
(i)→(ii). Suppose that (i) holds with ϵ and n given. Since the parametric representations in $\Pi_{s,2}(x)$ map onto the graph Γ_x, for any $(z_n, t_n) \in \Gamma_{x_n}$, we can find $s \in [0,1]$ such that $(u_n(s), r_n(s)) = (z_n, t_n)$. For that s, $(u(s), r(s)) = (z, t)$ for some $(z, t) \in \Gamma_x$ and

$$\|(z_n, t_n) - (z, t)\| \le \|u_n - u\| \vee \|r_n - r\| < \epsilon \ . \tag{11.11}$$

By the same reasoning, for any $(z, t) \in \Gamma_x$, there exists $(z_n, t_n) \in \Gamma_{x_n}$ such that (11.11) holds.
(ii)→(v). For x, t and ϵ given, find δ such that $v(x, [t-\delta, t)) < \epsilon/2$ and $v(x, [t, t+\delta)) < \epsilon/2$ for v in (2.5). Then apply (ii) to find n_0 such that $m_s(x_n, x) < \eta \equiv (\epsilon \wedge \delta)/2$ for $n \ge n_0$. Then, for each t' with $0 \vee (t - \eta) \le t' \le (t + \eta) \wedge T$, there must exist $(\bar{z}, \bar{t}) \in \Gamma_x$ such that

$$\|(x_n(t'), t') - (\bar{z}, \bar{t})\| < \eta \quad \text{for} \quad n \ge n_0 \ .$$

Since $|\bar{t} - t| \le |\bar{t} - t'| + |t' - t| < 2\eta < \delta$,

$$\|(\bar{z}, \bar{t}) - [x(t-), x(t)]\| < \epsilon/2 \ .$$

Consequently, for $n \ge n_0$,

$$\|x_n(t') - [x(t-), x(t)]\| < \eta + \epsilon/2 < \epsilon \ .$$

Since t' was arbitrary,

$$w_s(x_n, x, t, \delta) < \epsilon \ .$$

(v)↔(iv). Characterization (iv) clearly implies (v), so that it suffices to show that (v) implies (iv). We will show that if (iv) fails, then so does (v). Hence suppose that (iv) does not hold. Then there must exist $\epsilon > 0$, such that for any $\delta > 0$ there is a subsequence $\{n_k\}$ such that $n_k \to \infty$ and $\bar{w}_s(x_{n_k}, x, \delta) > \epsilon$ for all n_k. Hence, there is an associated sequence t_{n_k} such that

$$\bar{w}_s(x_{n_k}, x, t_{n_k}, \delta) > \epsilon/2$$

for all n_k. However, $\{t_{n_k}\}$ has a convergent subsequence $\{t_{n_{k_j}}\}$ with $t_{n_{k_j}} \to t$ as $n_{k_j} \to \infty$ for some t. Note that, if $z_n \in [x(t_n-), x(t_n)]$ for all n, where $t_n \to t$, and if $z_n \to z$, then necessarily $z \in [x(t-), x(t)]$. Hence,

$$\bar{w}_s(x_{n_{k_j}}, x, t, 2\delta) > \epsilon/2$$

for all sufficiently large n_{k_j}. That implies that (11.9) does not hold, so that (v) fails.

(iv)$\to$(vi). We construct the desired finite subsets A of Γ_x and A_n of Γ_{x_n} by considering two kinds of points in Γ_x. For $\epsilon > 0$ given, we let A contain at least one point (z,t) for each $t \in Disc(x, \epsilon/2)$. The other points have $t \in Disc(x)^c$. We first construct A for t outside a finite union of neighborhoods of points in $Disc(x, \epsilon/2)$. We then construct A_n and finally we complete the definition of A by adding appropriate points (z,t) for $t \in Disc(x, \epsilon/2)$, which depend on A_n. Thus the set A ultimately depends upon A_n and thus upon x_n and n.

Let $t(A)$ denote the set of t for which there is at least one pair (z,t) from Γ_x in A. We first identify $t(A)$. We include $Disc(x, \epsilon/2)$ in $t(A)$. Use (11.8) to find an η and an n_0 such that $\bar{w}_s(x_n, x, \eta) < \epsilon/4$ for all $n \geq n_0$. Let $t_1 < \cdots < t_m$ be the ordered set of points in $Disc(x, \epsilon/2) - \{T\}$; let $t_0 = 0$ and $t_{m+1} = T$. Use the existence of left and right limits for x to identify points, for $1 \leq i \leq m$, points t_i' and t_i'' in $Disc(x)^c$ such that $t_{i-1}'' < t_i' < t_i < t_i'' < t_{i+1}'$, $|t_i - t_i'| < \eta$, $|t_i - t_i''| < \eta$, $v(x, [t_i', t_i)) < \epsilon/4$ and $v(x, [t_i, t_i'']) < \epsilon/4$ for $v(x, B)$ in (2.5). We include these points t_i' and t_i'' in $t(A)$. We also include in A points t_0'' and t_{m+1}' from $Disc(x)^c$ such that $t_0 = 0 < t_0'' < t_1'$, $t_m'' < t_{m+1}' < t_{m+1} = T$, $v(x, [0, t_0'']) < \epsilon/4$ and $v(x, [t_{m+1}', T]) < \epsilon/4$. We also include the points 0 and T in $t(A)$. Moreover, we include the points $(x(t_i'), t_i')$, $(x(t_i''), t_i'')$, $(x(0), 0)$ and $(x(T), T)$ in A itself. (Except possibly for T, these are the only possibilities since t_i', t_i'', $0 \in Disc(x)^c$.) We next define A for t in the compact set

$$C \equiv [0,T] - \bigcup_{i=1}^{m} (t_i', t_i'') - [0, t_0'') - (t_{m+1}', T] \ . \tag{11.12}$$

The set C is a finite union of the closed intervals $[t_i'', t_{i+1}']$, $0 \leq i \leq m-1$. For each t in C not a boundary point of one of these subintervals, it is possible to find t' and t'' in the same subinterval as t such that $t' < t < t''$, $|t - t'| < \eta/4$, $|t - t''| < \eta/4$ and $v(x, [t', t'']) < \epsilon/2$. (Recall that $C \subseteq Disc(x, \epsilon/2)^c$.) For the boundary points t_i' and t_i'', include intervals $(\bar{t}_i, t_i']$ and $[t_i'', t_i^*)$ with the same properties; these intervals are open in the relative topology on C. Also include intervals $[0, t^*)$ and $(\bar{t}, T]$ with the same properties; these intervals again are open in the relative topology on C. These open intervals form an open cover of C. Since C is compact, there exists a finite subcover. We let $t(A)$ contain one point t in $Disc(x)^c$ from each subinterval in the finite subcover; we also put $(x(t), t)$ into A. Let the set A be ordered according to the time points; i.e., $(z_1, t_1) \leq (z_2, t_2)$ if $t_1 \leq t_2$. So far, A contains points $(x(t), t)$ for $t \in Disc(x)^c$, including the boundary points t_i' and t_i'' of C. We have completed the definition of $t(A)$, which includes $Disc(x, \epsilon/2)$. If $\{t_i\}$ is the ordered set of points in $t(A)$, then the construction above implies that $|t_{i+1} - t_i| < \eta$ for all i (where η has been chosen so that $\bar{w}_s(x_n, x, \eta) < \epsilon/4$).

We now construct the set A_n. By Theorem 11.4.1, condition (11.8) implies that $x_n(t) \to x(t)$ for each $t \in Disc(x)^c$. For each $t \in t(A) - Disc(x, \epsilon/2)$, let $t \in t(A_n)$ and $(x_n(t), t) \in A_n$. Since each such t belongs to $Disc(x)^c$, there is $n_1 \geq n_0$ such that $\|x_n(t) - x(t)\| < \epsilon/4$ for all $t \in t(A) - Disc(x, \epsilon/2)$ and for all $n \geq n_1$. Hence we have established $d^*(A, A_n) < \epsilon/4$ for $n \geq n_1$ over C (outside the neighborhoods of $Disc(x, \epsilon/2)$). We complete the definition of A_n by adding finitely many points (z, t) for t in the open interval (t_i', t_i'') where t_i' and t_i'' are the adjacent points in $t(A)$ to $t_i \in Disc(x, \epsilon/2)$. We also do this for the interval $(t_{m+1}', T]$ if $T \in Disc(x, \epsilon/2)$. We do this for all $t_i \in Disc(x, \epsilon/2)$ so that overall $\hat{d}(A_n, \Gamma_{x_n}) < \epsilon/2$. This is always possible by Lemma 6.3.1. We next complete the definition of A by including a point (z, t_i) for each point (z, t) in A_n with $t \in (t_i', t_i'')$. This ensures that A_n and A have the same cardinality. Since $d(A_n, \Gamma_{x_n}) \leq \epsilon/2$, $\bar{w}_s(x_n, x, \eta) < \epsilon/4$,

$$\|x_n(t_i') - x(t_i')\| < \epsilon/4, \|x_n(t_i'') - x(t_i'')\| < \epsilon/4 \ ,$$

$$\|x(t_i') - x(t_i-)\| < \epsilon/4 \quad \text{and} \quad \|x(t_i'') - x(t_i)\| < \epsilon/4$$

for $n \geq n$, we can choose points in A so that $d^*(A_n, A) \leq \epsilon/2$ for $n \geq n_1$ and $\hat{d}(A, \Gamma_x) \leq \epsilon$, which completes the proof.

(vi)$\to$(i). Suppose that ϵ is given and the sets A and A_n in (vi) have points (z_i, t_i) and $(z_{n,i}, t_{n,i})$, $0 \leq i \leq m$, where $t_0 = 0$ and $t_m = T$. Construct arbitrary parametric representations of (u, r) of x and (u_n, r_n) of x_n such that

$$r(i/m) = t_i, \quad u(i/m) = z_i$$

and

$$r_n(i/m) = t_{n,i}, \quad u_n(i/m) = z_{n,i} \ .$$

Since $d^*(A_n, A) \leq \epsilon$,

$$\max_{0 \leq i \leq m} \{|r(i/m) - r_n(i/m)| \vee \|u(i/m) - u_n(i/m)\|\} < \epsilon \ .$$

Since $\hat{d}(A, \Gamma_x) < \epsilon$ and $\hat{d}(A_n, \Gamma_{x_n}) < \epsilon$ too, by the triangle inequality,

$$\|r - r_n\| \vee \|u_n - u\| < 3\epsilon \ .$$

(iv)$\leftrightarrow$(vii). Suppose that $0 \leq t \leq T$. If x is constant in the intervals $(0 \vee (t - 2\delta), t)$ and $[t, (t + 2\delta) \wedge T)$, then

$$[x(0 \vee (t' - \delta)), x((t' + \delta) \wedge T)] = [x(t-), x(t)]$$

for all t' with $0 \vee (t-\delta) < t' < (t+\delta) \wedge T$. Consequently, in that situation

$$\begin{aligned} &\sup_{0\vee(t-\delta)<t'<(t+\delta)\wedge T} \{\|x_n(t') - [x(0\vee(t'-\delta)), x((t'+\delta)\wedge T)]\|\} \\ &= \sup_{0\vee(t-\delta)<t'<(t+\delta)\wedge T} \{\|x_n(t') - [x(t-), x(t)]\|\} \ . \end{aligned} \tag{11.13}$$

Thus if x is piecewise constant with the distance between successive discontinuities at least δ, then $\bar{w}_s^*(x_n, x, \delta/2) = \bar{w}_s(x_n, x, \delta/2)$. Hence, for ϵ given suppose that we can choose η to make $\bar{w}_s(x_n, x, \eta) < \epsilon/3$. Then approximate x by $x_c \in D_c$ such that $\|x - x_c\| < \epsilon/3$. For that x_c, let α be the minimum distance between successive discontinuities. Then, for $\delta < \eta \wedge (\alpha/2)$,

$$\begin{aligned} \bar{w}_s^*(x_n, x, \delta) &\le \bar{w}_s^*(x_n, x_c, \delta) + \epsilon/3 \\ &\le \bar{w}_s(x_n, x_c, \delta) + \epsilon/3 \\ &\le \bar{w}_s(x_n, x, \eta) + 2\epsilon/3 \le \epsilon \ . \end{aligned} \tag{11.14}$$

Alternatively, for ϵ given, suppose that we can choose η to make $\bar{w}_s^*(x_n, x, \eta) < \epsilon/3$. Following the same reasoning,

$$\begin{aligned} \bar{w}_s(x_n, x, \delta) &\le \bar{w}_s(x_n, x_c, \delta) + \epsilon/3 \\ &\le \bar{w}_s^*(x_n, x_c, \delta) + \epsilon/3 \\ &\le \bar{w}_s^*(x_n, x, \eta) + 2\epsilon/3 \le \epsilon \ . \end{aligned} \tag{11.15}$$

Hence (iv) is equivalent to (vii).
(ii)$\rightarrow$(viii). By Theorem 11.4.1 and (ii)$\leftrightarrow$(v), (ii) implies that $x_n(t) \rightarrow x(t)$ for each $t \in Disc(x)^c$. It remains to show that (ii)$\rightarrow$(11.10). For $\epsilon > 0$ given, first pick a piecewise-constant x_c such that $\|x - x_c\| \le \epsilon/4$, which is possible by Lemma 6.3.1. Let γ be the $\mathbb{R}^k$-valued function with $\gamma^i(t) = 1$, $0 \le t \le T$, $1 \le i \le k$. Then $x_c - (\epsilon/4)\gamma \le x \le x_\epsilon + (\epsilon/4)\gamma$, i.e.,

$$x_c(t) - \epsilon/4 \le x(t) \le x_c(t) + \epsilon/4 \quad \text{for} \quad 0 \le t \le T \ .$$

Let the (α, β)-neighborhood of $x \in D$ be

$$N_{\alpha,\beta}(x) \equiv \{[[x(t) - \alpha\gamma, x(t) + \alpha\gamma]] \times [0 \vee (t-\beta), (t+\beta) \wedge T] : 0 \le t \le T\} \ . \tag{11.16}$$

Thus, $x \in N_{\epsilon/4,0}(x_c)$ and $x_c \in N_{\epsilon/4,0}(x)$. Now let α be the minimum distance between successive discontinuities in x_c, or to 0 or T for the leftmost and rightmost discontinuity points. Given (ii), choose n_0 so that $m_s(x_n, x) = \eta_n < \eta < (\epsilon \wedge \alpha)/4$ for $n \ge n_0$. Then $x_n \in N_{\eta+\epsilon/4,\eta}(x_c)$. Suppose that $\{t_i : 1 \le i \le m-1\}$ is the set of discontinuities of x_c, with $t_0 = 0$ and

$t_m = T$. By the construction above, the open intervals $(t_i - \eta, t_i + \eta)$ are disjoint, $1 \le i \le m-1$. Now let $\delta = 2\eta$. Hence, if $t' \in (t_i - \eta, t_i + \eta)$ for $t_i \in Disc(x_c)$, then

$$t_{i-1}+\eta < t'-\delta < t'-\delta/2 < t_i-\eta < t_i+\eta < t'+\delta/2 < t'+\delta < t_{i+1}-\eta \quad (11.17)$$

for all i, $1 \le i \le m-1$. On the other hand, if $t' \in [t_{i-1} + \eta, t_i - \eta] = B_{i,n}$, then necessarily either $(t' + \delta/2, t' + \delta)$ intersects $B_{i,n}$ or $(t' - \delta, t' - \delta/2)$ intersects $B_{i,n}$. Thus, for $n \ge n_0$ and each $t' \in [0, T]$, there exists $t_1 \in [0 \vee (t' - \delta), 0 \vee (t' - \delta) + \delta/2)$ and $t_2 \in (T \wedge (t' + \delta) - \delta/2, T \wedge (t' + \delta)]$ such that

$$\|x_n(t') - [x_n(t_1), x_n(t_2)]\| \le 2((\epsilon/4) + \eta) < \epsilon \; ; \quad (11.18)$$

i.e., $\bar{w}_s^*(x_n, \delta) < \epsilon$.
(viii)$\to$(v). For x, t and ϵ given, choose η so that $0 < t - \eta < t < t + \eta < T$, $v(x, [t - \eta, t)) < \epsilon/4$ and $v(x, [t, t + \eta]) < \epsilon/4$. Now choose $\delta < \eta$ and $t' \in (t - \delta/2, t + \delta/2)$. For δ and t' given, find t_1, t_2 in $Disc(x)^c$ such that $t_1 < t < t_2$, $t' - \delta < t_1 < t' - \delta/2$ and $t' + \delta/2 < t_2 < t' + \delta$. Then choose n_0 so that $\|x_n(t_i) - x(t_i)\| < \epsilon/4$ for $i - 1, 2$ and $n \ge n_0$. Apply (viii) to choose $n_1 \ge n_0$ so that $\bar{w}_s^*(x_n, \delta) \le \epsilon/2$. Then

$$\begin{aligned} \|x_n(t') - [x(t-), x(t)]\| &\le \|x_n(t') - [x(t_1), x(t_2)]\| + \epsilon/4 \\ &\le \|x_n(t') - [x_n(t_1), x_n(t_2)]\| + \epsilon/2 \\ &\le \bar{w}_s^*(x_n, \delta) + \epsilon/2 \le \epsilon \quad \text{for} \quad n \ge n_1 \end{aligned} \quad (11.19)$$

Since t' is arbitrary in $(t - \delta/2, t + \delta/2)$,

$$\bar{w}_s(x_n, x, t, \delta/2) \le \epsilon \quad \text{for} \quad n \ge n_1 \; ,$$

which implies (v). ■

Remark 6.11.1. The equivalence (iii)$\leftrightarrow$(vii)$\leftrightarrow$(viii) was established by Skorohod (1956). ■

Remark 6.11.2. There is no analog to characterization (v) involving $\bar{w}_s^*(x_n, x, t, \delta)$ in (11.4) instead of $\bar{w}_s(x_n, x, t, \delta)$. For $t \in Disc(x)^c$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \; \bar{w}_s^*(x_n, x, t, \delta) = 0$$

implies pointwise convergence $x_n(t) \to x(t)$, but not the local uniform convergence in Theorem 6.4.1. ■

6.11.3. WM_2 Convergence

Corresponding characterizations of WM_2 convergence follow from Theorem 6.11.1 because the WM_2 topology is the same as the product topology, by Theorem 6.10.2. Let

$$\bar{w}_w(x_1, x_2, \delta) \equiv \sup_{0 \le t \le T} \bar{w}_w(x_1, x_2, t, \delta) \tag{11.20}$$

for $\bar{w}_w(x_1, x_2, t, \delta)$ in (4.7).

Theorem 6.11.2. (characterizations of WM_2 convergence) *The following are equivalent characterizations of $x_n \to x$ as $n \to \infty$ in (D, WM_2):*

(i) $d_{w,2}(x_n, x) \to 0$ as $n \to \infty$ for $d_{w,2}$ in (11.2)*; i.e., for any $\epsilon > 0$ and all n sufficiently large, there exist $(u, r) \in \Pi_{w,2}(x)$ and $(u_n, r_n) \in \Pi_{w,2}(x_n)$ such that $\|u_n - u\| \vee \|r_n - r\| < \epsilon$.*

(ii) $m_p(x_n, x) \to 0$ as $n \to \infty$ for the metric m_p in (10.5).

(iii) Given $\bar{w}_w(x_1, x_2, \delta)$ defined in (11.20),

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ \bar{w}_w(x_n, x, \delta) = 0 \ .$$

(iv) For each t, $0 \le t \le T$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ \bar{w}_w(x_n, x, t, \delta) = 0 \ .$$

(v) For all $\epsilon > 0$ and all sufficiently large n, there exist finite ordered subsets A of G_x and A_n of Γ_{x_n}, of common cardinality m as in (3.9) *with $(z_1, t_1) \le (z_2, t_2)$ if $t_1 \le t_2$, such that $\hat{d}(A, G_x) < \epsilon$, $\hat{d}(A_n, \Gamma_{x_n}) < \epsilon$ and $d^*(A, A_n) < \epsilon$ for all $n \ge n_0$, for $\hat{d}$ in* (5.8) *and d^* in* (5.6).

Proof. (i)$\to$(ii). Clearly, $d_{w,2}(x_n, x) \to 0$ as $n \to \infty$ implies that $d_{s,2}(x_n^i, x^i) \to 0$ as $n \to \infty$ for each i. By Theorem 6.11.1, that implies $m_s(x_n^i, x^i) \to 0$ as $n \to \infty$ for each i, which implies (ii).
(ii)$\leftrightarrow$(iii). By Theorem 6.11.1, (ii) is equivalent to

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ \bar{w}_s(x_n^i, x^i, \delta) = 0 \tag{11.21}$$

for each i, but that is equivalent to (iii) because

$$\max_{1 \le i \le k} \bar{w}_s(x_n^i, x^i, \delta) = \bar{w}_w(x_n, x, \delta) \ . \tag{11.22}$$

(iii)$\leftrightarrow$(iv). By Theorem 6.11.1, (iii) is equivalent to

$$\lim_{\delta\downarrow 0} \overline{\lim_{n\to\infty}} \ \bar{w}_s(x_n^i, x^i, t, \delta) = 0 \tag{11.23}$$

for each i, but that is equivalent to (iv) because

$$\max_{1\le i\le k} \bar{w}_s(x_n^i, x^i, t, \delta) = \bar{w}_w(x_n, x, t, \delta) \ . \tag{11.24}$$

(iii)$\rightarrow$(v). Follow the proof of (iv)$\rightarrow$(vi) in Theorem 6.11.1. Use (**??**) to find an η and an n_0 such that $\bar{w}_w(x_n, x, \eta) < \epsilon/4$ for all $n \ge n_0$. Define $t(A)$ as before, first by including $Disc(x, \epsilon/2)$ and then by adding points from $Disc(x)^c$ in the complement of the union of the intervals about the points in $Disc(x, \epsilon/2)$. Let A be defined for $t \in t(A) - Disc(x, \epsilon/2)$ just as before. Let A_n be defined just as before. We complete the definition of A by including a point (z_i, t_i) for each point (z, t) in A_n with $t \in (t_i', t_i'')$. This ensures that A and A_n have the same cardinality. Since $d(A_n, \Gamma_{x_n}) \le \epsilon/2$, $\bar{w}_w(x_n, x, \eta) < \epsilon/4$, $\|x_n(t_i') - x(t_i')\| < \epsilon/4$, $\|x_n(t_i'') - x(t_i'')\| < \epsilon/4$, $\|x(t_i') - x(t_i'-)\| < \epsilon/4$ and $\|x(t_i'') - x(t_i)\| < \epsilon/4$ for $n \ge n_1$, we can choose these points to add to A so that $d^*(A_n, A) \le \epsilon/2$ for $n \ge n_1$ and $\hat{d}(A, G_x) \le \epsilon$. (Unlike in the proof of Theorem 6.11.1, here we cannot conclude that $\hat{d}(A, \Gamma_x) \le \epsilon$.)
(v)$\rightarrow$(i). Paralleling the proof of (v)$\rightarrow$(i) in Theorem 11.5.2, suppose that the conditions of (v) hold and A, A_n and ϵ are given. Let (u, r) and (u_n, r_n) be parametric representations of x and x_n such that

$$\begin{aligned} u(i/m) &= z_i, \ r(i/m) = t_i \quad \text{for} \quad (z_i, t_i) \in A \\ u_n(i/m) &= z_{n,i}, \ r_n(i/m) = t_{n,i} \quad \text{for} \quad (z_{n,i}, t_{n,i}) \in A_n \ . \end{aligned}$$

For any $s \in [0, 1]$ there is i such that $s_i \le s \le s_{i+1}$ and

$$\begin{aligned} &\|u_n(s) - u(s)\| \vee \|r_n(s) - r(s)\| \le \|(u_n(s), r_n(s)) - u_n(s_i), r_n(s_i)\| \\ &+\|(u_n(s_i), r_n(s_i)) - (u(s_i), r(s_i))\| + \|(u(s_i), r(s_i)) - (u(s), r(s))\| \\ &\le \hat{d}(A_n, G_{x_n}) + d^*(A_n, A) + \hat{d}(A, G_x) \le 3\epsilon \ . \quad \blacksquare \end{aligned}$$

Theorem 6.11.2 and Section 6.4 show that all forms of M convergence imply uniform convergence to continuous limit functions.

Corollary 6.11.1. (from WM_2 convergence to uniform convergence) *Suppose that $m_p(x_n, x) \to 0$ as $n \to \infty$.*
(i) If $t \in Disc(x)^c$, then

$$\lim_{\delta\downarrow 0} \overline{\lim_{n\to\infty}} \ v(x_n, x, t, \delta) = 0 \ .$$

(ii) If $x \in C$, then $\lim_{n\to\infty} \|x_n - x\| = 0$.

Proof. For (i) combine Theorems 6.4.1 and 6.11.2. For (ii) add Lemma 6.4.2. ▪

Convergence in WM_2 has the advantage that jumps in the converging functions must be inherited by the limit function.

Corollary 6.11.2. (inheritance of jumps) *If $x_n \to x$ in (D, WM_2), $t_n \to t$ in $[0,T]$ and $x_n^i(t_n) - x_n^i(t_n-) \geq c > 0$ for all n, then $x^i(t) - x^i(t-) \geq c$.*

Proof. Apply Theorem 6.11.2 (iv). ▪

Let $J(x)$ be the maximum magnitude (absolute value) of the jumps of the function x in D. We apply Corollary 8.5.1 to show that J is upper semicontinuous.

Corollary 6.11.3. (upper semicontinuity of J) *If $x_n \to x$ in (D, M_2), then*

$$\overline{\lim_{n\to\infty}}\ J(x_n) \leq J(x)\ .$$

Proof. Suppose that $x_n \to x$ in (D, WM_2) and there exists a subsequence $\{x_{n_k}\}$ such that $J(x_{n_k}) \to c$. Then there exist further subsubsequences $\{x_{n_{k_j}}\}$ and $\{t_{n_{k_j}}\}$, and a coordinate i, such that $t_{n_{k_j}} \to t$ for some $t \in [0,T]$ and $|x^i_{n_{k_j}}(t_{n_{k_j}}) - x^i_{n_{k_j}}(t_{n_{k_j}}-)| \to c$. Then Corollary 8.5.1 implies that $|x^i(t) - x^i(t-)| \geq c$. ▪

6.11.4. Additional Properties of M_2

We conclude this section by discussing additional properties of the M_2 topologies. First we note that there are direct M_2 analogs of the M_1 results in Theorems 6.6.1, 6.7.1, 6.7.2 and 6.7.3.

Theorem 6.11.3. (extending SM_2 convergence to product spaces) *Suppose that $m_s(x_n, x) \to 0$ in $D([0,T], \mathbb{R}^k)$ and $m_s(y_n, y) \to 0$ in $D([0,T], \mathbb{R}^l)$ as $n \to \infty$. If*

$$Disc(x) \cap Disc(y) = \phi\,,$$

then

$$m_s((x_n, y_n), (x, y)) \to 0 \quad in \quad D([0,T], \mathbb{R}^{k+l}) \quad as \quad n \to \infty\,.$$

Proof. We use characterization (v) in Theorem 6.11.1. Using the discontinuity condition, it is easy to show that (11.9) holds for $[(x_n, y_n), (x, y)]$ when it holds separately for $[x_n, x]$ and $[y_n, y]$, because i.e., at most one of the segments $[(x(t-), x(t)]$ and $[y(t-), y(t)]$ contains more than a single point. ∎

Corollary 6.11.4. (from WM_2 convergence to SM_2 convergence when the limit is in D_1) *If $m_p(x_n, x) \to 0$ as $n \to \infty$ and $x \in D_1$, then $m_s(x_n, x) \to 0$ as $n \to \infty$.*

Theorem 6.11.4. (Lipschitz property of linear functions of the coordinate functions) *For any x_1, $x_2 \in D([0,T], \mathbb{R}^k)$ and $\eta \in \mathbb{R}^k$,*

$$m(\eta x_1, \eta x_2) \leq (\|\eta\| \vee 1) m_s(x_1, x_2) \ .$$

Proof. For (**??**), the key property is that

$$\Gamma_{\eta x} = \{(\eta z, t) : (z, t) \in \Gamma_x\} \ .$$

It suffices to show that for all $\epsilon > 0$ and $(z_1', t_1) \in \Gamma_{\eta x_1}$ there exists $(z_2', t_2) \in \Gamma_{\eta x_2}$ such that

$$|z_1' - z_2'| \vee |t_1 - t_2| \leq (\|\eta\| \vee 1) m_s(x_1, x_2) + \epsilon \ .$$

However, for $(z_1', t_1) \in \Gamma_{\eta x_1}$, there exists $(z_1, t_1) \in \Gamma_{x_1}$ such that $\eta z_1 = z_1'$. Then choose $(z_2, t_2) \in \Gamma_{x_2}$ such that

$$\|z_1 - z_2\| \vee |t_1 - t_2| \leq m_s(x_1, x_2) + \epsilon$$

Let $(z_2', t_2) = (\eta z_2, t_2)$. Then

$$|z_1' - z_2'| = |\eta z_1 - \eta z_2| \leq \|\eta\| \|z_1 - z_2\| \ . \quad ∎$$

We have an analog of Corollary 6.7.1 for the M_2 topology.

Corollary 6.11.5. (SM_2-continuity of addition) *If $m_s(x_n, x) \to 0$ and $m_s(y_n, y) \to 0$ in $D([0,T], \mathbb{R}^k)$ and*

$$Disc(x) \cap Disc(y) = \phi \, ,$$

then

$$m_s(x_n + y_n, x + y) \to 0 \quad in \quad D([0,T], \mathbb{R}^k) \, .$$

Proof. First apply Theorem 6.11.3 to get $m_s((x_n, y_n), (x, y)) \to 0$ in $D([0,T], \mathbb{R}^{k+l})$. Then apply Theorem 6.11.4. ∎

Theorem 6.11.5. (characterization of SM_2 convergence by convergence of all linear functions of the coordinates) *There is convergence $x_n \to x$ in $D([0,T], \mathbb{R}^k)$ as $n \to \infty$ in the SM_2 topology if and only if $\eta x_n \to \eta x$ in $D([0,T], \mathbb{R}^1)$ as $n \to \infty$ in the M_2 topology for all $\eta \in \mathbb{R}^k$.*

Proof. One direction is covered by Theorem 6.11.4. Suppose that $x_n \not\to x$ as $n \to \infty$ in SM_2. Then apply part (v) of Theorem 6.11.1 to deduce that $\eta x_n \not\to \eta x$ as $\eta \to \infty$ for some η. Note that $\|a\| > 0$ for $a \in \mathbb{R}^k$ if and only if $|\eta a| > 0$ in $\mathbb{R}$ for some $\eta \in \mathbb{R}^k$. Also, $\|a - A\| > 0$ for $A \subseteq R^k$ if and only if $|\eta a - \eta A| > 0$ in $\mathbb{R}$ for some $\eta \in \mathbb{R}^k$, where $\eta A = \{\eta b : b \in A\}$. ∎

Just as with the M_1 topology, we can get convergence of sums under more general conditions than in Corollary 6.11.5. It suffices to have the jumps of x^i and y^i have common sign for all i. We can express this property by the condition (7.2).

Theorem 6.11.6. (continuity of addition at limits with jumps of common sign) *If $x_n \to x$ and $y_n \to y$ in $D([0,T], \mathbb{R}^k, SM_2)$ and if condition* (7.2) *holds, then*

$$x_n + y_n \to x + y \quad in \quad D([0,T], \mathbb{R}^k, SM_2) \ .$$

Proof. Apply the characterization of SM_2 convergence in Theorem 6.11.1 (v). At points t in $Disc(x)^c \cap Disc(y)^c$, use the local uniform convergence in Lemma 12.5.1 of the book and Corollary 6.11.1 here. For other t not in $Disc(x) \cap Disc(y)$, use Theorem 6.11.3. For $t \in Disc(x) \cap Disc(y)$, exploit condition (7.2) to deduce that, for all $\epsilon > 0$, there exists δ and n_0 such that

$$\bar{w}_s(x_n + y_n, x + y, t, \delta) \leq w_s(x_n, x, t, \delta) + w_s(y_n, y, t, \delta) + \epsilon \qquad (11.25)$$

for all $n \geq n_0$. ∎

We now apply Theorem 6.11.5 to extend a characterization of convergence due to Skorohod (1956) to $\mathbb{R}^k$-valued functions. For each $x \in D([0,T], \mathbb{R}^1)$ and $0 \leq t_1 < t_2 \leq T$, let

$$M_{t_1,t_2}(x) \equiv \sup_{t_1 \leq t \leq t_2} x(t) \ . \qquad (11.26)$$

The proof exploits the SM_2 analog of Corollary 6.9.1.

In preparation for the next result, we state a basic lemma about preservation of convergence under restriction maps. For $x \in D([0,T], \mathbb{R}^k)$ and

$0 \le t_1 < t_2 \le T^*$, let $r_{t_1,t_2} : D([0,T],\mathbb{R}^k) \to D([t_1,t_2],\mathbb{R}^k)$ be the restriction map, defined by $r_{t_1,t_2}(x)(s) = x(s)$, $t_1 \le s \le t_2$. We omit the proof.

Lemma 6.11.1. (continuity of restriction maps) *If $x_n \to x$ as $n \to \infty$ in $D([0,T],\mathbb{R}^k)$ with one of the SM_1, WM_1, SM_2 and WM_2 topologies and if $t_1, t_2 \in Disc(x)^c$, then*

$$r_{t_1,t_2}(x_n) \to r_{t_1,t_2}(x) \quad as \quad n \to \infty \quad in \quad D([t_1,t_2],\mathbb{R}^k)$$

with the same topology.

Theorem 6.11.7. (characterization of SM_2 convergence in terms of convergence of local extrema) *There is convergence $m_s(x_n,x) \to 0$ as $n \to \infty$ in $D([0,T],\mathbb{R}^k)$ if and only if*

$$M_{t_1,t_2}(\eta x_n) \to M_{t_1,t_2}(\eta x) \quad as \quad n \to \infty \tag{11.27}$$

for all $\eta \in \mathbb{R}^k$ and all points t_1, $t_2 \in \{T\} \cup Disc(x)^c$ with $t_1 < t_2$.

Proof. By Theorem 6.11.5, it suffices to consider the case of real-valued functions. By considering $\eta = \pm 1$ in (11.27), we get both the minimum and the maximum over $[t_1,t_2]$. it is easy to see that (11.27) for $\eta = \pm 1$ implies characterization (v) in Theorem 6.11.1: For x, t and ϵ given, choose γ so that $v(x,[t-\gamma,t)) < \epsilon/2$, $v(x,[t,t+\gamma]) < \epsilon/2$ and $0 < t-\gamma < t+\gamma < T$. Then finding n_0 such that $|M_{t_1,t_2}(\eta x_n) - M_{t_1,t_2}(\eta x)| < \epsilon/2$ for $n \ge n_0$, $\eta = \pm 1$ and

$$t - \gamma < t_1 < t - \delta < t < t + \delta < t_2 < t + \gamma$$

implies that $\bar{w}_s(x_n,x,t,\delta) < \epsilon$ for $n \ge n_0$. On the other hand, if $x_n \to x$ in $D([0,T],\mathbb{R}^1,M_2)$, then the restrictions converge in $D([t_1,t_2],\mathbb{R}^1,M_2)$ for all $t_1,t_2 \in Disc(x)^c$ by Lemma 6.11.1. If $m_s(x_n,x) < \epsilon$ in $D([t_1,t_2],\mathbb{R}^1,M_2)$, then clearly $|M_{t_1,t_2}(x_n) - M_{t_1,t_2}(x)| < \epsilon$ and $|M_{t_1,t_2}(-x_n) - M_{t_1,t_2}(-x)| < \epsilon$, so characterization (ii) of Theorem 6.11.1 implies (11.27). ∎

We can apply the characterization of M_2 convergence in Theorem 6.11.7 to show the preservation of convergence under bounding functions in the M_2 topology. See Corollary 12.11.6 in the book.

6.12. Compactness

We have nothing to add in this final section.

Chapter 7

Useful Functions

7.1. Introduction

This chapter contains proofs omitted from Chapter 13 of the book, with the same title. As before, the theorems to be proved are restated here. The section and theorem numbers parallel Chapter 13 in the book, so that the proofs should be easy to find.

We consider four basic functions introduced in Section 3.5 of the book: composition, supremum, reflection and inverse. Another basic function is addition, but it has already been treated in Sections 12.6, 12.7 and 12.11 of the book. Our treatment of useful functions follows Whitt (1980), but the emphasis there was on the J_1 topology, even though the M_1 topology was used in places. In contrast, here the emphasis is on the M_1 and M_2 topologies.

Here is how this chapter is organized: We start in Section 7.2 by considering the composition map, which plays an important role in establishing FCLTs involving a random time change. We consider composition without centering in Section 7.2; then we consider composition with centering in Section 7.3.

In Section 7.4 we study the supremum function, both with and without centering. In Section 7.5 we apply the supremum results to treat the (one-sided one-dimensional) reflection map, which arises in queueing applications.

We start studying the inverse function in Section 7.6. We study the inverse map without centering in Section 7.6 and with centering in Section 7.7. In Section 7.8 we apply the results for inverse functions to obtain corresponding results for closely related counting functions.

In Section 7.9 we apply the previously established convergence-preservation results for the composition and inverse maps to establish stochastic-process limits for renewal-reward stochastic processes. When the times between the renewals in the renewal counting process have a heavy-tailed distribution, we need the M_1 topology.

In Chapter 3 of the Internet Supplement we discuss pointwise convergence and its preservation under mappings. The prservation of pointwise convergence focuses on relations for inidvidual sample paths, as in the queueing book by El-Taha and Stidham (1999). There we see that a function-space setting is not required for all convergence preservation.

7.2. Composition

This section is devoted to the composition function, mapping (x, y) into $x \circ y$, where

$$(x \circ y)(t) \equiv x(y(t)) \quad \text{for all} \quad t \ .$$

The composition map is useful to treat random sums and, more generally, processes modified by a random time change; e.g., see Section 13.9 of the book on renewal-reward processes.

Henceforth in this chapter, unless stipulated otherwise, when $D \equiv D^k$, so that the range of functions is $\mathbb{R}^k$, we let D be endowed with the strong version of the J_1, M_1 or M_2 topology, and simply write J_1, M_1 or M_2. It will be evident that most results also hold with the corresponding weaker product topology.

7.2.1. Preliminary Results

To ensure that $x \circ y \in D$, we will assume that y is also nondecreasing. We begin by defining subsets of $D \equiv D^k \equiv D([0, \infty), \mathbb{R}^k)$ that we will consider. Let D_0 be the subset of all $x \in D$ with $x^i(0) \geq 0$ for all i. Let $D_\uparrow$ and $D_{\uparrow\uparrow}$ be the subsets of functions in D_0 that are nondecreasing and strictly increasing in each coordinate. Let D_m be the subset of functions x in D_0 for which the coordinate functions x^i are monotone (either increasing or decreasing) for each i. Let C_0, $C_\uparrow$, $C_{\uparrow\uparrow}$ and C_m be the corresponding subsets of C; i.e., $C_0 \equiv C \cap D_0$, $C_\uparrow \equiv C \cap D_\uparrow$, $C_{\uparrow\uparrow} = C \cap D_{\uparrow\uparrow}$, and $C_m = C \cap D_m$.

It is important that all of these subsets are measurable subsets of D with the Borel σ-fields associated with the nonuniform Skorohod topologies, which all coincide with the Kolmogorov σ-field generated by the projection maps; see Theorems 11.5.2 and 11.5.3 in the book.

Returning to the composition map, we state the condition for $x \circ y \in D$ as a lemma.

Lemma 7.2.1. (criterion for $x \circ y$ to be in D) *For each* $x \in D([0,\infty),\mathbb{R}^k)$ *and* $y \in D_{\uparrow}([0,\infty),\mathbb{R}_+)$, $x \circ y \in D([0,\infty),\mathbb{R}^k)$.

A basic result, from pp. 145, 232 of Billingsley (1968), is the following. The continuity part involves the topology of uniform convergence on compact intervals.

Theorem 7.2.1. (continuity of composition at continuous limits) *The composition map from* $D^k \times D^1_{\uparrow}$ *to* D^k *is measurable and continuous at* $(x,y) \in C^k \times C^1_{\uparrow}$.

Our goal now is to obtain additional positive continuity results under extra conditions. We use the following elementary lemma.

Lemma 7.2.2. *If* $y(t) \in Disc(x)$ *and* y *is strictly increasing and continuous at* t, *then* $t \in Disc(x \circ y)$.

The following is the J_1 result.

Theorem 7.2.2. (J_1-continuity of composition) *The composition map from* $D^k \times D^1_{\uparrow}$ *to* D^k *taking* (x,y) *into* $(x \circ y)$ *is continuous at* $(x,y) \in (C^k \times D^1_{\uparrow}) \cup (D^k \times C^1_{\uparrow\uparrow})$ *using the* J_1 *topology throughout.*

Proof. First suppose that $(x_n, y_n) \to (x,y)$ in $D^k \times D^1_{\uparrow}$ with $(x,y) \in C^k \times D_{\uparrow}$. Choose $t_1 \in Disc(y)^c$. Then $y_n \to y$ for the restrictions to $[0,t_1]$; i.e., there exist $\lambda_n \in \Lambda([0,t_1])$ such that $\|y_n - y \circ \lambda_n\|_{t_1} \vee \|\lambda_n - e\|_{t_1} \to 0$. Choose t_2 such that $y(t_1) \le t_2$ and $y_n(t_1) \le t_2$ for all $n \ge 1$. Since $x \in C^k$, $\|x_n - x\|_{t_2} \to 0$. By the triangle inequality,

$$\|x_n \circ y_n - x \circ y \circ \lambda_n\|_{t_1} \le \|x_n \circ y_n - x \circ y_n\|_{t_1} + \|x \circ y_n - x \circ y \circ \lambda_n\|_{t_1} . \quad (2.1)$$

The first term on the right in (2.1) converges to 0 because $\|x_n - x\|_{t_2} \to 0$ and the range of y_n is contained in $[0,t_2]$. The second term on the right in (2.1) converges to 0 because x is uniformly continuous over $[0,t_2]$ and $\|y_n - y \circ \lambda_n\|_{t_1} \to 0$.

Next suppose that $(x_n, y_n) \to (x,y)$ in $D^k \times D^1_{\uparrow}$ with $(x,y) \in D \times C_{\uparrow\uparrow}$. By Lemma 7.2.2 below, $y(t) \in Disc(x)^c$ for each $t \in Disc(x \circ y)^c$. However, for each $t' \in Disc(x)^c$, we have local uniform convergence of x_n to x, i.e.,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \, v(x_n, x, t', \delta) = 0 \; ; \quad (2.2)$$

see Section 12.4 in the book. Since $y_n(t) \to y(t)$ as $n \to \infty$, as a consequence of (2.2), we have $(x_n \circ y_n)(t) \to (x \circ y)(t)$ for each $t \in Disc(x \circ y)^c$. Now we show that the closure of the sequence $\{x_n \circ y_n : n \geq 1\}$ is compact in the J_1 topology. Since $(x_n \circ y_n)(t) \to (x \circ y)(t)$ for t in a countable dense subset, all limits of convergent subsequences must coincide with $x \circ y$. Since all convergent subsequences have the same limit, compactness implies that the sequence itself must converge; i.e., $x_n \circ y_n \to x \circ y$ (J_1). Hence it suffices to show that the closure of $\{x_n \circ y_n\}$ is compact, for which we apply Theorem 14.4 of Billingsley (1968). For an arbitrary t_1, choose $t_2 > y(t_1)$ with $t_2 \in Disc(x)^c$. Then, for all sufficiently large n, $y_n(t_1) < t_2$ and $x_n \to x$ for the restrictions in $D([0, t_2], \mathbb{R}^k)$. It is easy to see that condition (14.49) and (14.50) in Billingsley (1968) hold. First, (14.49) holds because

$$\sup_{\substack{0 \leq s \leq t_1 \\ n \geq 1}} \|x_n \circ y_n(s)\| \leq \sup_{\substack{0 \leq s \leq t_2 \\ n \geq 1}} \|x_n\| < \infty \,, \tag{2.3}$$

since $x_n \to x$ in $D([0, t_2], \mathbb{R}^k, J_1)$. Next (14.50) holds because the oscillation functions for $x_n \circ y_n$ over $[0, t_1]$ be bounded above by the oscillation functions of x_n over $[0, t_2]$; e.g., since $y \in C_{\uparrow\uparrow}$ and $\|y_n - y\|_{t_1} \to 0$, for any δ_2 there exists n_0 and δ_1 such that $w''_{x_n \circ y_n}(\delta_1) \leq w''_{x_n}(\delta_2)$ for all $n \geq n_0$. ∎

7.2.2. M-Topology Results

We have a different result for the M topologies.

Theorem 7.2.3. (M-continuity of composition) *If $(x_n, y_n) \to (x, y)$ in $D^k \times D^1_{\uparrow}$ and $(x, y) \in (D^k \times C^1_{\uparrow\uparrow}) \cup (C^k_m \times D^1_{\uparrow})$, then $x_n \circ y_n \to x \circ y$ in D^k, where the topology throughout is M_1 or M_2.*

In most applications we have $(x, y) \in D^k \times C^1_{\uparrow\uparrow}$, as is illustrated by the next section. That part of the M conditions is the same as for J_1. The mode of convergence in Theorem 7.2.3 for $y_n \to y$ does not matter, because on $D^1_{\uparrow}$, convergence in the M_1 and M_2 topologies coincides with pointwise convergence on a dense subset of $[0, \infty)$, including 0; see Corollary **??**.

It is easy to see that composition cannot in general yield convergence in a stronger topology, because $x \circ y = x$ and $x_n \circ y_n = x_n$, $n \geq 1$, when $y_n = y = e$, where $e(t) = t$, $t \geq 0$. Unlike for the J_1 topology, the composition map is in general *not* continuous at $(x, y) \in C \times D^1_{\uparrow}$ in the M topologies.

We actually prove a more general continuity result, which covers Theorem 7.2.3 as a special case.

Theorem 7.2.4. (more general M-continuity of composition) *Suppose that* $(x_n, y_n) \to (x, y)$ *in* $D^k \times D^1_\uparrow$. *If (i)* y *is continuous and strictly increasing at* t *whenever* $y(t) \in Disc(x)$ *and (ii)* x *is monotone on* $[y(t-), y(t)]$ *and* $y(t-), y(t) \notin Disc(x)$ *whenever* $t \in Disc(y)$, *then* $x_n \circ y_n \to x \circ y$ *in* D^k, *where the topology throughout is* M_1 *or* M_2.

Theorem 7.2.3 follows easily from Theorem 7.2.4: First, on $D^k \times C^1_\uparrow$, y is continuous, so only condition (i) need be considered; it is satisfied because y is continuous and strictly increasing everywhere. Second on $C^k_m \times D^1_\uparrow$, x is continuous so only condition (ii) need be considered; it is satisfied because x is monotone everywhere. Hence it suffices to prove Theorem 7.2.4, which is done in Section 1.8 of the Internet Supplement. The general idea in our proof of Theorem 7.2.4 is to work with the characterization of convergence using oscillation functions evaluated at single arguments, exploiting Theorems 6.5.1 (v), 6.5.2 (iv), 6.11.1 (v) and 6.11.2 (iv).

We obtain a stronger result (M_1 convergence of $x_n \circ y_n$ given only M_2 convergence of x_n) if we do not need to invoke condition (i) in Theorem 7.2.4. A sufficient condition is for x to be continuous.

Theorem 7.2.5. (obtaining SM_1 convergence from WM_2 convergence) *If the conditions of Theorem* 7.2.4 *hold with* $y(t) \notin Disc(x)$ *for all* t, *then* $x_n \circ y_n \to x \circ y$ *in* (D^k, SM_1) *even if* $x_n \to x$ *only in* (D^k, WM_2).

Proof. Apply Lemmas 7.2.4, 7.2.5 and 7.2.8 below. ∎

We prove Theorem 7.2.4 by identifying four different cases, with each either having $t \in Disc(x \circ y)$ or not.

Proof of Theorem 7.2.4. We will establish the appropriate characterization of convergence $x_n \circ y_n \to x \circ y$ at each t separately, using Theorems 12.5.1 (v), 12.5.2 (iv), 12.11.1 (v) and 12.11.2 (iv) in the book.

There are four cases to consider:

(i) $t \notin Disc(y)$ and $y(t) \notin Disc(x)$, so that $t \notin Disc(x \circ y)$;

(ii) $t \in Disc(y)$, $x(u) = x(y(t-)) = x(y(t))$ for all $u \in [y(t-), y(t)]$ and $y(t-)$, $y(t) \notin Disc(x)$, under which $t \notin Disc(x \circ y)$;

(iii) $t \in Disc(y)$, $x(y(t-)) \neq x(y(t))$, x is monotone on $[y(t-), y(t)]$ and $y(t-), y(t) \notin Disc(x)$, under which $t \in Disc(x \circ y)$;

(iv) $y(t) \in Disc(x)$ and y is continuous and strictly increasing at t so that $t \in Disc(x \circ y)$.

In case (ii) we have $t \notin Disc(x \circ y)$ even though $t \in Disc(y)$. The regularity conditions in case (ii) follow from condition (ii); since $x(y(t-)) = x(y(t))$, monotonicity reduces to a constant value over the subinterval. Case (iii) differs from case (ii) by having $x(y(t-)) \neq x(y(t))$, which makes $t \notin Disc(x \circ y)$. The regularity conditions in case (iii) again follow from condition (ii). The regularity conditions in case (iv) when $y(t) \in Disc(x)$ follow from condition (i). We use Lemma 7.2.2 in case (iv). In each case we know whether or not $t \in Disc(x \circ y)$. The four cases are covered by subsequent lemmas as follows: Case (i) by Lemmas 7.2.3–7.2.4; case (ii) by Lemma 7.2.5; case (iii) by Lemmas 7.2.6–7.2.8; and case (iv) by Lemma 7.2.10. ■

We now establish several lemmas in order to complete the proof of Theorem 7.2.4. Throughout, we assume that (x_n, y_n), $n \geq 1$, and (x, y) are elements of $D^k \times D_\uparrow^1$. Refer to Section 12.4 of the book for the oscillation functions.

Lemma 7.2.3. *If $v(y_n, y, t, \delta_1) \leq \delta_2$ in $D_\uparrow^1$, then*

$$u(x_n \circ y_n, x \circ y, t, \delta_1) \leq v(x_n, x, y(t), \delta_2) + \bar{v}(x \circ y, t, \delta_1)$$

for v in (12.4.2), u in (12.4.1) and $\bar{v}$ in (12.4.3), all in Section 12.4 of the book.

Proof. By the condition, $|y_n(t_1) - y(t)| \leq \delta_2$ provided that $0 \vee (t - \delta_1) < t_1 < (t + \delta_1) \wedge T$. Hence, for t_1 in that range,

$$\begin{aligned} \|(x_n \circ y_n)(t_1) - (x \circ y)(t_1)\| &\leq \|x_n(y_n(t_1) - x(y(t))\| \\ &\quad + \|x(y(t)) - x(y(t_1))\| \\ &\leq v(x_n, x, y(t), \delta_2) + \bar{v}(x \circ y, t, \delta_1) \ . \end{aligned}$$ ■

Lemma 7.2.4. *If $t \notin Disc(y)$, $y(t) \notin Disc(x)$,*

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ v(y_n, y, t, \delta) = 0 \tag{2.4}$$

and

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ v(x_n, x, y(t), \delta) = 0 \ , \tag{2.5}$$

then

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ v(x_n \circ y_n, x \circ y, t, \delta) = 0 \ . \tag{2.6}$$

Proof. Since $t \notin Disc(y)$ and $y(t) \notin Disc(x)$, $t \notin Disc(x \circ y)$ and $\bar{v}(x \circ y, t, \delta_1) \to 0$ as $\delta_1 \to 0$. We apply Lemma 7.2.3: For $\epsilon > 0$ given, choose δ_2 and n_1 so that

$$v(x_n, x, y(t), \delta_2) < \epsilon/2 \quad \text{for} \quad n \geq n_1 \ .$$

Then choose δ_1 and $n_2 \geq n_1$ so that $\bar{v}(x \circ y, t, \delta_1) < \epsilon/2$ and

$$v(y_n, y, t, \delta_1) \leq \delta_2 \quad \text{for} \quad n \geq n_2 \ .$$

By Lemma 7.2.3,

$$u(x_n \circ y_n, x \circ y, t, \delta_1) \leq \epsilon \quad \text{for} \quad n \geq n_2 \ .$$

Since ϵ was arbitrary, we have shown that

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ u(x_n \circ y_n, x \circ y, t, \delta) = 0 \ ,$$

which is equivalent to (2.6) by Theorem 12.4.1 in the book. ■

Recall the m_p is the product metric inducing the WM_2 topology.

Lemma 7.2.5. *Suppose that $t \in Disc(y)$ but $y(t) \notin Disc(x)$, $y(t-) \notin Disc(x)$ and $x(y(t)) = x(y(t-))$ so that $t \notin Disc(x \circ y)$, i.e., case (ii) in Theorem 7.2.4. If $m_p(y_n, y) \to 0$ and $m_p(x_n, x) \to 0$ as $n \to \infty$, and $x(u) = x(y(t))$ for all $u \in [y(t-), y(t)]$, then* (2.6) *holds.*

Proof. Since $u \notin Disc(x)$ for all $u \in [y(t-), y(t)]$ and $m_p(x_n, x) \to 0$ as $n \to \infty$, for $\epsilon > 0$ given, we can choose δ_1 and n_0 so that

$$\sup_{0 \vee (y(t-) - \delta_1) \leq u \leq (y(t) + \delta_1) \wedge T} \{\|x_n(u) - x(u)\| \leq \epsilon \tag{2.7}$$

for all $n \geq n_0$ by Lemma 12.4.2 in the book. Since $x(u) = x(y(t))$ for $y(t-) \leq u \leq y(t)$ and x is continuous at $y(t-)$ and $y(t)$, from (2.7) we can obtain δ_2 such that

$$\sup_{0 \vee (y(t-) - \delta_2) \leq u \leq (y(t) + \delta_2) \wedge T} \{\|x_n(u) - x(y(t))\|\} \leq 2\epsilon \tag{2.8}$$

for $n \geq n_0$. By right continuity and the existence of left limits, we can choose $t_1 < t < t_2$ such that

$$y(t_1) < y(t-) < y(t_1) + \delta_2/2 \ , \tag{2.9}$$

$$y(t) < y(t_2) < y(t) + \delta_2/2 , \tag{2.10}$$

$$\|(x \circ y)(t_j) - (x \circ y)(t)\| < \epsilon , \tag{2.11}$$

and $t_j \notin Disc(y)$ for $j = 1, 2$. Applying (2.4), we can choose $\delta_3 > 0$ and $n_1 \geq n_0$ so that

$$v(y_n, y, t_j, \delta_3) < \delta_2/2 \tag{2.12}$$

for all $n \geq n_1$ and $j = 1, 2$. Combining (2.9)–(2.12), and using the monotonicity of y_n and y, we have for $0 \vee (t - \delta_3) \leq t', t'' \leq (t + \delta_3) \wedge T$, $\|y_n(t') - \{y(t-), y(t)\}\| < \delta_2$. Thus, by (2.8),

$$\|x_n \circ y_n(t') - x \circ y(t'')\| \leq \|x_n \circ y_n(t') - x \circ y(t)\| + \|x \circ y(t) - x \circ y(t'')\| \leq 3\epsilon .$$

Since ϵ was arbitrary, we have established (2.6). ∎

We now turn to case (iii). We first show how we can exploit the monotonicity condition.

Lemma 7.2.6. (characterization of M_2 convergence at a monotone limit) *Suppose that x is monotone on $[a, b]$. Then $x_n \to x$ in $D([a, b], \mathbb{R}^k, WM_2)$ if and only if $x_n \to x$ pointwise on a dense subset of $[a, b]$ and*

$$\lim_{n\to\infty} w^*(x_n, [a, b]) = 0 , \tag{2.13}$$

where

$$w^*(x, [a, b]) \equiv \sup_{a \leq t_1 \leq t_2 \leq t_3 \leq b} \{\|x(t_2) - [x(t_1), x(t_3)]\|\} . \tag{2.14}$$

These imply that $x_n \to x$ as $n \to \infty$ in SM_1 as well.

Proof. Clearly $w_s(x, \delta) \leq w^*(x, [a, b])$ on $D([a, b], \mathbb{R}^k)$ for all $\delta > 0$, where

$$w_s(x, \delta) \equiv \sup_{a \leq t \leq b} w_s(x, t, \delta)$$

for $w_s(x, t, \delta)$ in equation (12.4.4) of the book, so that (2.13) plus the pointwise convergence implies that $x_n \to x$ as $n \to \infty$ in SM_1, by the basic characterization of SM_1 convergence, which in turn implies convergence in WM_2. To go the other way, suppose that $w^*(x_n, [a, b]) \not\to 0$ as $n \to \infty$. Then there exist $\epsilon > 0$ and subsequences $\{n_k\}$, $\{t_{n_k,j}\}$ for $j = 1, 2$ and 3 such that $n_k \to \infty$ and

$$\|x_{n_k}(t_{n_k,2}) - [x_{n_k}(t_{n_k,1}), x_{n_k}(t_{n_k,3})]\| > \epsilon \tag{2.15}$$

for all n_k. There are thus further subsequences $\{n'_k\}$, $\{t'_{n_k,2}\}$ for $j = 1, 2,$ and 3 so that $t'_{n_k,j} \to t_j$ as $n'_k \to \infty$ for each j, where $t_1 \le t_2 \le t_3$. Assuming that $x_n \to x$ as $n \to \infty$ in WM_2, we have $x_{n'_k}(t'_{n_k,j}) \to [[x(t_j-), x(t_j)]]$ as $n'_k \to \infty$, by the characterization of WM_2 convergence. This, with (2.15) and the monotonicity of x, implies that

$$\max_{1\le i\le k}\{\|x^i(t_2-) - [x^i(t_1-), x^i(t_3)]\|, \quad \|x^i(t_2) - [x^i(t_1-), x^i(t_3)]\| > 0\,,$$

which is impossible because x^i is monotone for each i. Hence, (2.13) must hold when $x_n \to x$ as $n \to \infty$ in WM_2. ∎

We will also apply the following elementary lemma, for which we omit the proof. We use the oscillation functions w_s in (12.4.4) and $\bar{v}$ in (12.4.3) of the book.

Lemma 7.2.7. *If*

$$y(t-) - \delta_2 \le y_n(t_1) \le y_n(t_2) \le y(t) + \delta_2$$

whenever $0 < t - \delta_1 \le t_1 \le t_2 \le t + \delta_1$, *then*

$$w_s(x_n \circ y_n, t, \delta_1) \le \bar{v}(x_n, y(t), \delta_2) + \bar{v}(x_n, y(t-), \delta_2) + w^*(x_n, [y(t-), y(t)])$$

for w^* *in (2.14).*

We apply Lemmas 7.2.6 and 7.2.7 to establish the following.

Lemma 7.2.8. *In case (iii), with* $t \in Disc(y)$, $y(t-)$, $y(t) \notin Disc(x)$ *and* x *monotone on* $[y(t-), y(t)]$, *if* $(x_n, y_n) \to (x, y)$ *in* $D^k(WM_2) \times D^1_{\uparrow}(WM_2)$, *then*

$$\lim_{\delta\downarrow 0}\ \overline{\lim_{n\to\infty}}\ w_s(x_n \circ y_n, t, \delta) = 0\ .$$

Proof. For any $\delta_2 > 0$ given, we can find δ_1 so that

$$y(t-) - \delta_2/2 \le y(t_1) \le y(t_2) \le y(t) + \delta_2/2$$

for $0 \vee (t - \delta_1) \le t_1 \le t_2 \le t + \delta_1$. By choosing continuity points of y, we can choose $n_2 \ge n_1$ so that

$$y(t-) - \delta_2 \le y_n(t_1) \le y_n(t_2) \le y(t) + \delta_2$$

for all $n \geq n_2$. Hence we can apply Lemmas 7.2.6 and 7.2.7. By Lemma 7.2.6, $w^*(x_n, [y(t-), y(t)]) \to 0$ as $n \to \infty$. Since $x_n \to x$ and $y(t-)$, $y(t) \notin Disc(x)$,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \; \bar{v}(x_n, y(t), \delta) = 0$$

and

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \; \bar{v}(x_n, y(t-), \delta) = 0 \; .$$

An application of Lemma 7.2.7 completes the proof. ∎

We now turn to case (iv). We first establish a preliminary result of independent interest, but which we do not directly need.

Lemma 7.2.9. *Suppose that $m_p(x_n, x) \to 0$ in D and $m(y_n, y) \to 0$ in $D^1_\uparrow$, but that $y(t) \in Disc(x)$. If y is strictly increasing and continuous in a neighborhood of t, then $(x_n \circ y_n)(t'_n) \to (x \circ y)(t')$ for all t' in a dense subset of neighborhood of t and all sequences $\{t'_n\}$ with $t'_n \to t'$.*

Proof. In the neighborhood of $y(t)$, there are at most countably many discontinuities of x. Since y is strictly increasing and continuous in a neighborhood of t, y is invertible there. Hence, for suitably small δ_2 and all but countably many t' in $(t - \delta_2, t + \delta_2)$, we simultaneously have y continuous at t' and x continuous at $y(t')$. At all such t', we have $y_n(t'_n) \to y(t')$ and $x_n(y_n(t')) \to x(y(t'))$ whenever $t'_n \to t'$, because m_p-convergence implies local uniform convergence at continuity points, by virtue of Theorem 12.4.1 in the book.

Corollary 7.2.1. *If y is strictly increasing and continuous whenever $y(t) \in Disc(x)$ and $(x_n, y_n) \to (x, y)$ in $D^1_\uparrow(M_1) \times D^1_\uparrow(M_1)$, then $x_n \circ y_n \to x \circ y$ in $D^1_\uparrow(M_1)$.*

Proof. By Lemma 7.2.6, M_1 convergence on $D^1_\uparrow$ coincides with pointwise convergence on a dense subset. Apply Lemma 7.2.9. ∎

Lemma 7.2.10. *If $m(y_n, y) \to 0$ in $D^1_\uparrow$, where y is continuous and strictly increasing at t, then for any $\delta > 0$, we can find $\delta_1 > 0$ such that, for all n sufficiently large,*

$$\begin{aligned} w_s(x_n \circ y_n, t, \delta_1) &\leq w_s(x_n, y(t), \delta) \; , \\ w_w(x_n \circ y_n, t, \delta_1) &\leq w_w(x_n, y(t), \delta) \; , \end{aligned}$$

$$\begin{aligned}\bar{w}_s(x_n \circ y_n, x \circ y, t, \delta_1) &\leq \bar{w}_s(x_n, x, y(t), \delta) , \\ \bar{w}_s(x_n \circ y_n, x \circ y, t, \delta_1) &\leq \bar{w}_w(x_n, x, y(t), \delta) .\end{aligned}$$

Proof. Since y is continuous at t, we can find $t_1 < t < t_2$ such that y is continuous at t_1 and t_2 and $|y(t) - y(t_j)| < \delta/2$ for $j = 1, 2$. Since $y_n \to y$ we can find n_0 such that $|y_n(t_j) - y(t_j)| < \delta/2$ for $n \geq n_0$ and $j = 1, 2$. By the triangle inequality, $|y_n(t_j) - y(t)| < \delta$ for $n \geq n_0$ and $j = 1, 2$. Let $\delta_1 = \min\{|t-t_1|, |t-t_2|\}$. Since y_n and y are nondecreasing, $|y_n(t')-y(t)| < \delta$ whenever $|t' - t| < \delta_1$. Hence

$$w_s(x_n \circ y_n, t, \delta_1) \leq w_s(x_n, y(t), \delta)$$

and

$$w_w(x_n \circ y_n, t, \delta_1) \leq w_w(x_n, y(t), \delta) .$$

Moreover, since y is continuous and strictly increasing, $x(y(t)-) = x(y(t-))$. Hence

$$\bar{w}_s(x_n \circ y_n, x \circ y, t, \delta_1) \leq \bar{w}_s(x_n, x, y(t), \delta)$$

and

$$\bar{w}_w(x_n \circ y_n, x \circ y, t, \delta_1) \leq \bar{w}_w(x_n, x, y(t), \delta) . \quad \blacksquare$$

7.3. Composition with Centering

This section considers the composition map with centering. Nothing was omitted from the book here.

7.4. Supremum

In this section we consider the supremum function, mapping $D \equiv D([0, T], \mathbb{R})$ into itself according to

$$x^{\uparrow}(t) = \sup_{0 \leq s \leq t} x(s), \quad 0 \leq t \leq T . \tag{4.1}$$

7.4.1. The Supremum without Centering

The following elementary result is stated without proof.

Theorem 7.4.1. (Lipschitz property of the supremum function) *For any* $x_1, x_2 \in D([0,T],\mathbb{R})$,

$$\begin{aligned} d_{J_1}(x_1^{\uparrow}, x_2^{\uparrow}) &\leq d_{J_1}(x_1, x_2) , \\ d_{M_1}(x_1^{\uparrow}, x_2^{\uparrow}) &\leq d_{M_1}(x_1, x_2) , \\ d_{M_2}(x_1^{\uparrow}, x_2^{\uparrow}) &\leq d_{M_2}(x_1, x_2) . \end{aligned}$$

The conclusion in Theorem 7.4.1 can be recast in terms of pointwise convergence: Since $x^{\uparrow}$ is nondecreasing, convergence $x_n^{\uparrow} \to x^{\uparrow}$ in the M topologies is equivalent to pointwise convergence at continuity points of $x^{\uparrow}$, because on $D_{\uparrow}$ the M_1 and M_2 topologies coincide with pointwise convergence on a dense subset of $\mathbb{R}_+$ including 0; see Corollary 12.5.1 in the book. Thus the M topologies have not contributed much so far. We obtain more useful convergence-preservation results for the supremum map with the M topologies when we combine supremum with centering. As before, let e be the identity map, i.e., $e(t) = t$, $0 \leq t \leq T$.

7.4.2. The Supremum with Centering

The following is the main result stated as Theorem 13.4.2 in the book. Our object here is to prove it.

Theorem 7.4.2. (convergence preservation with the supremum function and centering) *Suppose that* $c_n(x_n - e) \to y$ *as* $n \to \infty$ *in* $D([0,T],\mathbb{R})$ *with one of the topologies* J_1, M_1 *or* M_2, *where* $c_n \to \infty$.

(a) If the topology is M_1 *or* M_2, *then* $c_n(x_n^{\uparrow} - e) \to y$ *in the same topology.*

(b) If the topology is J_1, *then* $c_n(x_n^{\uparrow} - e) \to y$ *if and only if* y *has no negative jumps.*

Before proving Theorem 7.4.2, we establish some preliminary lemmas. We first give an alternative expression for the result, in the form of a continuous mapping theorem. Let $y_n \equiv c_n(x_n - e)$. Then $s_n(y_n) = c_n(x_n^{\uparrow} - e)$, where

$$s_n(y) \equiv (y + c_n e)^{\uparrow} - c_n e \quad \text{for} \quad y \in D \ . \tag{4.2}$$

Thus the conclusion of Theorem 7.4.2 can be expressed as $s_n(y_n) \to s(y) \equiv y$ when $y_n \to y$, with the appropriate topology.

Note that, for $x \in D$ and s_n in (4.2), $s_n(x)$ cannot have any negative jumps. For any $x \in D$, we can characterize $s_n(x)$ as the majorant which decreases by at most slope c_n at any time; i.e.,

$$s_n(x) = \inf\{y \in D : y \geq x,\ y(t_2) - y(t_1) \geq -c_n(t_2 - t_1)\}, \tag{4.3}$$

where we allow $0 \leq t_1 < t_2 \leq T$.

Lemma 7.4.1. *For any $x \in D$, $s_n(x)$ defined by (4.2) satisfies (4.3).*

Proof. First note that $s_n(x) \geq x$. Next note that

$$\begin{aligned} s_n(x)(t_2) - s_n(x)(t_1) &= (x + c_n e)^{\uparrow}(t_2) - (x + c_n e)^{\uparrow}(t_1) - c_n(t_2 - t_1) \\ &\geq -c_n(t_2 - t_1) . \end{aligned}$$

Finally, suppose that $y \geq x$ and $y(t_2) - y(t_1) \geq -c_n(t_2 - t_1)$ for all $0 \leq t_1 < t_2 < T$. Then $s_n(y) = y$. Since $y \geq x$, $s_n(y) \geq s_n(x)$. Hence $y \geq s_n(x)$. ∎

We can also bound $s_n(x)$ above for sufficiently large n by another majorant. Let the *left-local-majorant* of $x \in ([0, T], \mathbb{R})$ be

$$s_l^{\epsilon}(x)(t) = \sup_{0 \vee (t-\epsilon) \leq s \leq t} x(s), \quad 0 \leq t \leq T . \tag{4.4}$$

It is obvious that $x \leq s_l^{\epsilon}(x)$ for all x and $\epsilon > 0$. Moreover $s_l^{\epsilon}(x)(t)$ is nonincreasing as $\epsilon \downarrow 0$. We now show that $s_l^{\epsilon}(x) \to x$ in (D, M_2) as $\epsilon \downarrow 0$.

Lemma 7.4.2. *For any $x \in D$ and $\epsilon > 0$, there exists $\delta > 0$ such that*

$$d_{M_2}(x, s_l^{\delta}(x)) \leq \epsilon . \tag{4.5}$$

Proof. First, for x and ϵ given, apply Theorem 12.2.2 in the book to choose $x_c \in D_c$ such that $\|x - x_c\| < \epsilon/3$. For x_c, it is evident that there exists δ with $0 < \delta < \epsilon/3$ such that

$$d_{M_2}(s_l^{\delta}(x_c), x_c) < \delta < \epsilon/3 \quad \text{and} \quad \|s_l^{\delta}(x_c) - s_l^{\delta}(x)\| < \epsilon/3 .$$

Hence,

$$d_{M_2}(x, s_l^{\delta}(x)) \leq \|x - x_c\| + d_{M_2}(x_c, s_l^{\delta}(x_c)) + \|s_l^{\delta}(x_c) - s_l^{\delta}(x)\| < \epsilon \quad \blacksquare \tag{4.6}$$

We now show that $s_n(x) \to x$ as $n \to \infty$ in the M_2 topology, uniformly over a large class of functions x.

Lemma 7.4.3. *Let s_n be as in (4.2), where $c_n \to \infty$. For any M and $\epsilon > 0$, there is an n_0 such that*

$$d_{M_2}(s_n(x), x) < \epsilon, \quad n \geq n_0 , \tag{4.7}$$

for all x with $\|x\| \leq M$.

Proof. Let ϵ, M and x be given with $\|x\| \le M$. Apply Lemma 7.4.2 to find δ such that $m(s_l^\delta(x), x) < \delta < \epsilon$. Choose n_0 so that $c_n\delta > 2M$ for $n \ge n_0$. Then, for $n \ge n_0$,

$$x(s) + c_n s - c_n t \le x(t) \tag{4.8}$$

for all s, $0 \le s \le t - \delta$, $0 \le t \le T$, because under those conditions

$$x(s) + c_n s - c_n t \le M - c_n \delta \le -M \le x(t)\,. \tag{4.9}$$

Hence, for $n \ge n_0$,

$$x \le s_n(x) \le s_l^\delta(x)\,, \tag{4.10}$$

so that, by Lemma 7.4.2, $s_n(x)$ is contained in an M_2 ϵ-neighborhood of x; i.e., (4.7) holds. ∎

Next, for the J_1 results we need the following.

Lemma 7.4.4. *If $x \in D([0,T],\mathbb{R})$ and x has no negative jumps, then for any $\epsilon > 0$ there is a $\delta > 0$ such that*

$$v^-(x,\delta) \equiv \sup_{\substack{0\vee(t-\delta)\le t'\le t \\ 0\le t\le T}} \{x(t') - x(t)\} < \epsilon\,. \tag{4.11}$$

Proof. Under the condition, for any $\epsilon > 0$ and all $t \in (0,T]$, there is a $\delta(t)$ such that $0 < t - \delta(t) < t$ and

$$x(t') \le x(t) + \epsilon \quad \text{for all} \quad t' \in (t - \delta(t), t)\,. \tag{4.12}$$

By the right continuity of x at 0, there is a $\delta(0)$ such that $\|x(t') - x(0)\| < \epsilon$ for $0 \le t' \le \delta(0)$. The intervals $[0, \delta(0))$, $(t - \delta(t), t)$, $0 < t \le T$, form an open cover of the compact set $[0,T]$. Hence there is a finite subcover. Let the subcover be chosen (modified) so that each t is in at most two subintervals. Let δ be the minimum length of the overlapping intervals, i.e.,

$$\delta = \min_i\{|t_i + \delta(t_{i+1}) - t_{i+1}|\} \wedge \delta(0)\,. \tag{4.13}$$

Then, if t is any point in $[0,T]$, it either belongs to the subinterval $[0, \delta(0))$ or it is at least δ away from the left endpoint of one of its subintervals. Hence property (4.11) holds for δ in (4.13). ∎

Proof of Theorem 7.4.2. (a) We will show that $s_n(x_n) \to x$ whenever $x_n \to x$, for s_n in (4.2). First consider the M_2 topology. Let M be a constant so that $\|x\| \le M/2$. Since $d_{M_2}(x_n, x) \to 0$, there is an n_0 such that $\|x_n\| \le M$ for all $n \ge n_0$. By the condition and Lemma 7.4.3, for any $\epsilon > 0$ there is an $n_1 \ge n_0$ such that $d_{M_2}(x_n, x) < \epsilon/2$ and $d_{M_2}(s_n(x_n), x_n) < \epsilon/2$ for $n \ge n_1$. Hence, by the triangle inequality, for $n \ge n_1$,

$$d_{M_2}(s_n(x_n), x) \le d_{M_2}(s_n(x_n), x_n) + d_{M_2}(x_n, x) < \epsilon \ .$$

Next consider the M_1 topology. Since M_1 convergence implies M_2 convergence, we have $d_{M_2}(s_n(x_n), x) \to 0$ by the proof above. It thus suffices to strengthen convergence from M_2 to M_1. In particular, we can apply part (v) of Theorem 12.5.1 in the book. By Theorem 12.4.1 in the book, the M_2 convergence implies the local uniform convergence at continuity points in condition (12.5.4) in the book, so it only remains to establish the oscillation function limit at discontinuity points in condition (12.5.5) in the book; i.e.,

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ w_s(s_n(x_n), t, \delta) = 0 \ . \tag{4.14}$$

We show that if (4.14) fails, then necessarily we cannot have

$$\lim_{\delta \downarrow 0} \overline{\lim_{n\to\infty}} \ w_s(x_n, t, \delta) = 0 \ , \tag{4.15}$$

so that $x_n \not\to x$ (M_1), which is a contradiction. If (4.14) fails, then there must exist $\epsilon > 0$, $\delta_k \downarrow 0$ and $n_k \uparrow \infty$ such that

$$w_s(s_{n_k}(x_{n_k}), t, \delta_k) > \epsilon \quad \text{for all} \quad k \ . \tag{4.16}$$

Let $y_{n_k} = s_{n_k}(x_{n_k})$. Given (4.16), there are two cases: In the first case, there exist $t_{n_k,1}$, $t_{n_k,2}$ and $t_{n_k,3}$ such that

$$0 \vee (t - \delta_k) \le t_{n_k,1} < t_{n_k,2} < t_{n_k,3} \le (t + \delta_k) \wedge T \ , \tag{4.17}$$

$y_{n_k}(t_{n_k,2}) > y_{n_k}(t_{n_k,1}) + \epsilon$ and $y_{n_k}(t_{n_k,2}) > y_{n_k}(t_{n_k,3}) + \epsilon$. However, $y_{n_k}(t_{n_k,2}) > y_{n_k}(t_{n_k,1}) + \epsilon$ implies that there must exist $t'_{n_k,2}$ with $t_{n_k,1} < t'_{n_k,2} \le t_{n_k,2}$ and $x_{n_k}(t'_{n_k,2}) \ge y_{n_k}(t_{n_k,2})$. Since $y_{n_k}(t_{n_k,1}) \ge x_{n_k}(t_{n_k,1})$ and $y_{n_k}(t_{n_k,3}) \ge x_{n_k}(t_{n_k,3})$, we then must have $w_s(x_{n_k}, t, \delta_k) > \epsilon$, which contradicts (4.15).

In the second case, there exist $t_{n_k,1}$, $t_{n_k,2}$ and $t_{n_k,3}$ such that (4.17) holds, $y_{n_k}(t_{n_k,2}) < y_{n_k}(t_{n_k,1}) - \epsilon$ and $y_{n_k}(t_{n_k,2}) < y_{n_k}(t_{n_k,3}) - \epsilon$. By the last inequality, there must exist $t'_{n_k,3}$ with $t_{n_k,2} < t'_{n_k,3} \le t_{n_k,3}$ such that

$x_{n_k}(t'_{n_k,3}) \ge y_{n_k}(t_{n_k,3}) - \epsilon$. Since $x_n \le y_n$, $x_{n_k}(t_{n_k,2}) \le y_{n_k}(t_{n_k,2})$. Finally, since $\{x_{n_k}\}$ is uniformly bounded, there is δ'_k where $\delta'_k \downarrow 0$ as $k \to \infty$, and $t'_{n_k,1}$ with $0 \vee (t - (\delta_k + \delta'_k)) \le t'_{n_k,1} \le t_{n_k,1}$ with $x_{n_k}(t'_{n_k,1}) \ge y_{n_k}(t_{n_k,1})$. Hence, we must have

$$w_s(x_{n_k}, t, \delta_k + \delta'_k) > \epsilon \quad \text{for all} \quad k\ . \tag{4.18}$$

Since $\delta_k + \delta'_k \downarrow 0$ as $k \to \infty$, (4.18) again contradicts (4.15) and thus $x_n \to x$ (M_1). Thus, $d_{M_1}(s_n(x_n), x) \to 0$ as claimed.

(b) We now turn to the J_1 result. Given $c_n(x_n - e) \to y$ (J_1), there exists $\lambda_n \in \Lambda$ such that $\|c_n(x_n - e) - y \circ \lambda_n\| \to 0$ as $n \to \infty$. We want to show that $\|c_n(x_n^{\uparrow} - e) - y \circ \lambda_n\| \to 0$. Since $x_n^{\uparrow} \ge x_n$, it suffices to show, for any $\epsilon > 0$, that there is n_1 such that

$$c_n x_n(s') - c_n s \le y(\lambda_n(s)) + \epsilon \quad \text{for} \quad 0 \le s' \le s \le T \tag{4.19}$$

for $n \ge n_1$. Choose n_0 such that $\|c_n(x_n - e) - y \circ \lambda_n\| < \epsilon/2$ for $n \ge n_0$. From (4.19), we see that it suffices to show that there is $n_1 \ge n_0$ such that

$$y(\lambda_n(s')) \le y(\lambda_n(s)) + c_n(s - s') + \epsilon/2 \quad \text{for} \quad 0 \le s' \le s \le T\,. \tag{4.20}$$

Since y has no negative jumps, we can apply Lemma 7.4.4 to conclude that there is a δ such that $v^-(y,\delta) < \epsilon/2$ for $v^-(y,\delta)$ in (4.11). Then choose $n_1 \ge n_0$ such that $\|\lambda_n - e\| < \delta$ and $c_n\delta \ge \|y\|$ for $n \ge n_1$, and we obtain (4.20). Finally, recall that the maximum negative jump function is continuous, e.g., see p. 301 of Jacod and Shiryaev (1987); i.e.,

$$J_-(x) \equiv \sup_{0<t\le 1} \{x(t-) - x(t)\}\ . \tag{4.21}$$

Clearly, $J_-(c_n(x_n^{\uparrow} - e)) = 0$, so that if $c_n(x_n^{\uparrow} - e) \to y$ (J_1), then y must have no negative jumps. ∎

We now obtain joint convergence in the stronger topologies on $D([0,T],\mathbb{R}^2)$ under the condition that the limit function have no negative jumps.

Theorem 7.4.3. (criterion for joint convergence) *Suppose that $c_n(x_n - e) \to y$ as $n \to \infty$ in $D([0,T],\mathbb{R})$ with one of the J_1, M_1 or M_2 topologies, where $c_n \to \infty$. If, in addition, y has no negative jumps, then*

$$c_n(x_n - e, x_n^{\uparrow} - e) \to (y,y) \quad \text{as} \quad n \to \infty \tag{4.22}$$

in $D([0,T],\mathbb{R}^2)$ with the strong version of the same topology, i.e., with SJ_1, SM_1 or SM_2.

Proof. For the SM_1 and SM_2 topologies, we will work with parametric representations, using the parametric representation $((u,u),r)$ for (y,y). Given that $(c_n(x_n - e) \to y$, there exist parametric representations $(u_n, r_n) \in \Pi_s(c_n(x_n - e))$ and $(u,r) \in \Pi(y)$ such that $\|u_n - u\| \vee \|r_n - r\| \to 0$ as $n \to \infty$. We construct the desired parametric representations from these. Note that $(c_n^{-1}u_n + r_n, r_n) \in \Pi(x_n)$ and $(u_n', r_n) \in \Pi(c_n(x_n^{\uparrow} - e))$ for

$$u_n' = c_n((c_n^{-1}u_n + r_n)^{\uparrow} - r_n) = (u_n + c_n r_n)^{\uparrow} - c_n r_n\,. \tag{4.23}$$

Note that $x_n^{\uparrow}$ has the jumps up of x_n, while $x_n^{\uparrow}$ is continuous when x_n has a jump down. Thus $((u_n, u_n'), r_n) \in \Pi_s(y_n, y_n')$ for $y_n \equiv c_n(x_n - e)$ and $y_n' \equiv c_n(x_n^{\uparrow} - e)$. Of course $((u,u),r) \in \Pi_s((y,y))$. Thus it remains to show that

$$\|(u_n, u_n') - (u,u)\| \vee \|r_n - r\| \to 0 \quad \text{as} \quad n \to \infty\,. \tag{4.24}$$

Given that $\|u_n - u\| \vee \|r_n - r\| \to 0$, it suffices to show that $\|u_n' - u\| \to 0$. Clearly, $u_n' \geq u_n$ for all n, so that it suffices to show that, for all $\epsilon > 0$, there exist n_1 such that $u_n'(s) < u(s) + \epsilon$ for all $n \geq n_1$ and $s \in [0,1]$. Equivalently, by (4.23), it suffices to show that

$$u_n(s') + c_n(r_n(s') - r_n(s)) < u(s) + \epsilon, \quad 0 \leq s' \leq s \leq 1\ , \tag{4.25}$$

for all $n \geq n_1$. However, if we assume that the limit y has no negative jumps, then Lemma 7.4.4 implies that there is a $\delta > 0$ such that

$$u(s') \leq u(s) + \epsilon/2 \tag{4.26}$$

for all s, s' with $0 \leq s' \leq s \leq 1$ and $r(s) - r(s') < \delta$. Choose n_0 so that

$$\|u_n - u\| \vee \|r_n - r\| \leq (\delta \wedge \epsilon)/4 \quad \text{for} \quad n \geq n_0\,.$$

Choose $n_1 \geq n_0$ so that

$$c_n\delta/4 \geq 2\|x\| \quad \text{for} \quad n \geq n_1\ . \tag{4.27}$$

There are two cases: (i) $r_n(s) - r_n(s') \leq \delta/4$ and (ii) $r_n(s) - r_n(s') > \delta/4$. In case (i), $r(s) - r(s') < \delta$, so that by (4.26)

$$u_n(s') + c_n(r_n(s') - r_n(s)) \leq u_n(s') \leq u(s') + \epsilon/4 \leq u(s) + \epsilon\,, \tag{4.28}$$

so that (4.25) holds. In case (ii), by (4.27),

$$\begin{aligned} u_n(s') + c_n(r_n(s') - r_n(s)) &\leq u(s') + \epsilon/2 - c_n\delta/4 \\ &\leq u(s) + 2\|u\| - c_n\delta/4 + \epsilon/2 \\ &\leq u(s) + \epsilon\,, \end{aligned} \tag{4.29}$$

so that again (4.25) holds. Turning to J_1, we note that the result already follows from the proof of Theorem 7.4.2 (ii) because the same homeomorphims $\lambda_n \in \Lambda$ were used for both $c_n(x_n - e) \to y$ and $c_n(x_n^{\uparrow} - e) \to y$. ∎

Corollary 7.4.1. *Under the conditions of Theorem 7.4.3,*

$$\|c_n(x_n^{\uparrow} - x_n)\| \to 0 \quad as \quad n \to \infty .$$

Proof. Apply subtraction to get

$$c_n(x_n - x_n^{\uparrow}) = c_n(x_n - e) - c_n(x_n^{\uparrow} - e) \to x - x(M_2) .$$

Since the limit is continuous, the convergence holds in the uniform topology. ∎

We next give an elementary result about the supremum function when the centering is in the other direction, so that x_n must be rapidly decreasing. Convergence $x_n^{\uparrow}(t) \to x(0)$ as $n \to \infty$ is to be expected, but that conclusion can not be drawn if the M_2 convergence in the condition is replaced by pointwise convergence.

Theorem 7.4.4. (convergence preservation with the supremum function when the centering is in the other direction) *Suppose that $c_n \to \infty$ and $x_n + c_n e \to y$ in $D([0,T], \mathbb{R}, M_2)$. Then*

$$\|x_n^{\uparrow} - z(y)\| \to 0 \quad as \quad n \to \infty ,$$

where $z(y)(t) \equiv y(0)$, $0 \le t \le T$.

Proof. The assumed M_2 convergence implies local uniform convergence at the origin: For any $\epsilon > 0$, there is a δ and an n_0 such that

$$\sup_{0 \le t \le \delta} |x_n(t) + c_n t - y(0)| \le v(x_n, y, 0, \delta) < \epsilon$$

for $n \ge n_0$, where $v(x_1, x_2, t, \delta)$ is the modulus of continuity in (4.2) in Section 6.4. Hence, $x_n(t) \le y(0) + \epsilon$ for all t, $0 \le t \le \delta$, and $n \ge n_0$. Use the conditions to find $n_1 \ge n_0$ such that $\|x_n + c_n e\| \le \|y\| + \epsilon$ and $c_n \delta > 2\|y\|$ for $n \ge n_1$. Then, for $t > \delta$ and $n \ge n_1$,

$$x_n(t) = -c_n\delta + x_n(t) + c_n\delta \le -c_n\delta + \|x_n + c_n e\| \le -c_n\delta + \|y\| + \epsilon \le y(0) + \epsilon .$$

Hence, $x_n^{\uparrow}(t) \le y(0) + \epsilon$ for all t, $0 \le t \le T$, and $n \ge n_1$. On the other hand, for all t, $x_n^{\uparrow}(t) \ge x_n(0) \to y(0)$ as $n \to \infty$. ∎

7.5. One-Dimensional Reflection

Closely related to the supremum function is the one-dimensional (one-sided) reflection mapping, which we have used to construct queueing processes. Indeed, the reflection mapping can be defined in terms of the supremum mapping as

$$\phi(x) \equiv x + (-x \vee 0)^{\uparrow} ;$$

i.e.,

$$\phi(x)(t) = x(t) - (\inf\{x(s) : 0 \leq s \leq t\} \wedge 0) \ , \quad 0 \leq t \leq T \ , \qquad (5.1)$$

as in equation (2.5) in Section 5.2 of the book.

The Lipschitz property for the supremum function with the uniform topology in Lemma **??** immediately implies a corresponding result for the reflection map ϕ in (5.1).

Unfortunately, however, the Lipschitz property for the reflection map ϕ with the uniform topology does not even imply continuity in all the Skorohod topologies. In particular, ϕ is not continuous in the M_2 topology.

We do obtain positive results with the J_1 and M_1 topologies. As before, let d_{J_1} and d_{M_1} be the metrics in equations 3.2 and 3.4 in Section 3.3 of the book. For the J_1 result, we use the following elementary lemma.

Lemma 7.5.1. *For any $x \in D$ and $\lambda \in \Lambda$,*

$$\phi(x) \circ \lambda = \phi(x \circ \lambda) \ .$$

For the M_1 result, we use the following lemma. A fundamental difficulty for treating the more general multidimensional reflection map is that Lemma 7.5.2 below does not extend to the multidimensional reflection map; see Chapter 8.

Lemma 7.5.2. (preservation of parametric representations under reflections) *For any $x \in D$, if $(u, r) \in \Pi(x)$, then $(\phi(u), r) \in \Pi(\phi(x))$.*

Proof. In book. ■

Theorem 7.5.1. (Lipschitz property with the J_1 and M_1 metrics) *For any $x_1, x_2 \in D([0,T], \mathbb{R})$,*

$$d_{J_1}(\phi(x_1), \phi(x_2)) \leq 2 d_{J_1}(x_1, x_2)$$

and

$$d_{M_1}(\phi(x_1), \phi(x_2)) \leq 2 d_{M_1}(x_1, x_2)) \ ,$$

where ϕ is the reflection map in (5.1).

Proof. In book. ■

Theorem 7.5.1 covers the standard heavy-traffic regime for one single-server queue when $\rho = 1$, where ρ is the traffic intensity. The next result covers the other cases: $\rho < 1$ and $\rho > 1$. We use the following elementary lemma in the easy case of the uniform metric.

Lemma 7.5.3. *Let d be the metric for the U, J_1, M_1 or M_2 topology. Let $x \vee a : D \to D$ be defined by*

$$(x \vee a)(t) \equiv x(t) \vee a, \quad 0 \le t \le T. \tag{5.2}$$

Then, for any $x_1, x_2 \in D$,

$$d(x \vee a(x_1), x \vee a(x_2)) \le d(x_1, x_2) .$$

Theorem 7.5.2. (convergence preservation with centering) *Suppose that $x_n - c_n e \to y$ in $D([0,T], \mathbb{R})$ with the U, J_1, M_1 or M_2 topology.*

(a) If $c_n \to +\infty$, then

$$\phi(x_n) - c_n e \to y + \gamma(y) \quad as \quad n \to \infty \quad in \quad D$$

with the same topology, where

$$\gamma(y)(t) \equiv (-y(0)) \vee 0 = -(y(0) \wedge 0), \quad 0 \le t \le T .$$

(b) If $c_n \to -\infty$, $y(0) \le 0$ and y has no positive jumps, then

$$\|\phi(x_n) - 0e\| \to 0 \quad as \quad n \to \infty \quad in \quad D ,$$

where $e(t) = t$, $0 \le t \le T$.

Proof. (a) Note that

$$\phi(x_n) - c_n e = x_n - c_n e + (-x_n \vee 0)^{\uparrow} ,$$

where $(-x_n \vee 0)^{\uparrow} = (-x_n)^{\uparrow} \vee 0$. By assumption, $x_n - c_n e \to y$. By Theorem 7.4.4,

$$\|(-x_n)^{\uparrow} - z(-y)\| \to 0 \quad \text{as} \quad n \to \infty$$

where $z(-y)(t) = -y(0)$, $0 \le t \le T$. By Lemma 7.5.3,

$$\|(-x_n)^{\uparrow} \vee 0 - z(-y) \vee 0\| \to 0 \quad \text{as} \quad n \to \infty .$$

We obtain the desired convergence by adding, using the fact that the second term has a continuous limit.

(b) Apply the argument of Theorem 7.4.4 to show that, for all $\epsilon > 0$, there exists n_1 such that $-x_n(t) > -y(0) - \epsilon$ for all t, $0 \le t \le T$, and all $n \ge n_1$. Since $y(0) \le 0$, $-x_n(t) > -\epsilon$ for all t, $0 \le t \le T$, and all $n \ge n_1$. Thus,

$$(-x_n + c_n e, (-x_n) \vee 0 + c_n e) \to (-y, -y)$$

in $D([0,T], \mathbb{R}^2)$ with the appropriate strong topology. Then, by Theorem 7.4.3,

$$(-x_n + c_n e, (-x_n) \vee 0 + c_n e, (-x_n \vee 0)^{\uparrow} + c_n e) \to (-y, -y, -y) \tag{5.3}$$

in $D([0,T], \mathbb{R}^3)$ with the appropriate strong topology. Then, by applying subtraction to the first and third terms in (5.3), we get

$$\begin{aligned} \phi(x_n) &\equiv x_n + (-x_n \vee 0)^{\uparrow} \\ &= [(-x_n \vee 0)^{\uparrow} + c_n e] - [-x_n + c_n e] \\ &\to -y + y = 0e \end{aligned} \tag{5.4}$$

as $n \to \infty$. ■

7.6. Inverse

We now consider the inverse map.. It is convenient to consider the inverse map on the subset D_u of x in $D \equiv D([0,\infty), \mathbb{R})$ that are unbounded above and satisfy $x(0) \ge 0$. For $x \in D_u$, let the inverse of x be

$$x^{-1}(t) = \inf\{s \ge 0 : x(s) > t\}, \quad t \ge 0 . \tag{6.1}$$

As before, let D_0 be the subset of x in D with $x(0) \ge 0$, and let $D_{\uparrow}$ and $D_{\uparrow\uparrow}$ be the subsets of nondecreasing and strictly increasing functions in D_0. Let $D_{u\uparrow} \equiv D_u \cap D_{\uparrow}$ and $D_{u\uparrow\uparrow} \equiv D_u \cap D_{\uparrow\uparrow}$. Clearly,

$$D_{\uparrow\uparrow} \subseteq D_{\uparrow} \subseteq D_u \subseteq D_0 .$$

7.6.1. The M_1 Topology

Even for the M_1 topology, there are complications at the left endpoint of the domain $[0, \infty)$.

Example 7.6.1. *Complications at the left endpoint of the domain.* To see that the inverse map from $(D_\uparrow, U)$ to $(D_\uparrow, M_1)$ is in general not continuous, let $x(t) = 0$, $0 \le t < 1$, and $x(t) = t$, $t \ge 1$; Let $x_n = t/n$, $0 \le t < 1$ and $x_n(t) = t$, $t \ge 1$. Then $\|x_n - x\|_\infty = n^{-1} \to 0$, but $x_n^{-1}(0) = 0 \not\to 1 = x^{-1}(0)$, so that $x_n^{-1} \not\to x^{-1}$ (M_1). ■

To avoid the problem in Example 7.6.1, we can require that $x^{-1}(0) = 0$. To develop an equivalent condition, let $D_\epsilon^\uparrow$ be the subset of functions x in D_u such that $x(t) = 0$ for $0 \le t \le \epsilon$.

Then let

$$D_u^* \equiv \cap_{n=1}^\infty (D_{u,n^{-1}})^c \ . \tag{6.2}$$

Lemma 7.6.1. (measurability of D_u^*) *With the J_1, M_1 or M_2 topology, D_u^* in* (6.2) *is a G_δ subset of D_u and*

$$D_u^* = \{x \in D_u : x^{-1}(0) = 0\} \ . \tag{6.3}$$

Let $D_{u\uparrow}^* \equiv D_\uparrow \cap D_u^*$. A key property of $D_{u,\uparrow}^*$, not shared by $D_{u,\uparrow}$ because of the complication at the origin, is that parametric representation (u, r) for x directly serve as parametric representations for x^{-1} when we switch the roles of the components u and r.

Lemma 7.6.2. (switching the roles of u and r) *For $x \in D_{u,\uparrow}^*$, the graph Γ_x serves as the graph of $\Gamma_{x^{-1}}$ with the axes switched. Thus, $(u, r) \in \Pi(x)$ if and only if $(r, u) \in \Pi(x^{-1})$, where $\Pi(x)$ is the set of M_1 parametric representations.*

Corollary 7.6.1. (continuity on (D_u^*, M_1)) *The inverse map from (D_u^*, M_1) to $(D_{u,\uparrow}, M_1)$ is continuous.*

Proof. First apply Theorem 7.4.1 for the supremum. Then apply Lemma 7.6.2. ■

We now generalize Corollary 7.6.1 by only requiring that the limit be in D_u^*.

Theorem 7.6.1. (measurability and continuity at limits in D_u^*) *The inverse map in* (6.1) *from (D_u, M_2) to $(D_{u,\uparrow}, M_1)$ is measurable and continuous at $x \in D_u^*$, i.e., for which $x^{-1}(0) = 0$.*

Proof. First, recalling that the Borel σ-field on D coincides with the Kolmogorov σ-field generated by the projections, measurability follows from Lemma **??**; it suffices to show that $\{x : x^{-1}(t) \le a\}$ is measurable. However,

$$\begin{aligned} \{x : x^{-1}(t) \le a\} &= \cap_{j=1}^{\infty} \cap_{k=1}^{\infty} \{x : x^{-1}((t + j^{-1})-) \le a + k^{-1}\} \\ &= \cap_{j=1}^{\infty} \cap_{k=1}^{\infty} \{x : x^{\leftarrow}((t + j^{-1})) \le a + k^{-1}\} \\ &= \cap_{j=1}^{\infty} \cap_{k=1}^{\infty} \{x : x(a + k^{-1}) \ge t + j^{-1}\} , \end{aligned} \tag{6.4}$$

which is measurable. Next we turn to continuity. For any $x \in D_u$, $x^{-1} = (x^{\uparrow})^{-1}$, so it suffices to start from $x_n^{\uparrow} \to x^{\uparrow}$. By Theorem 7.4.1, the assumed convergence $x_n \to x$ in (D_u, M_2) implies that $x_n^{\uparrow} \to x^{\uparrow}$ in $(D_{\uparrow}, M_2)$. However, the M_1 and M_2 topologies coincide in $D_{\uparrow}$. So $x_n^{\uparrow} \to x^{\uparrow}$ in $(D_{\uparrow}, M_1)$. Since $x \in D_u^*$, $x^{\uparrow} \in D_{u,\uparrow}^*$. However, we need not have $x_n^{\uparrow} \in D_{u,\uparrow}^*$. We could directly apply Lemma 7.6.2 if $x_n^{\uparrow} \in D_{\uparrow}^*$ for all sufficiently large n. Hence suppose that is not the case. Then there exists a subsequence $\{x_{n_k}^{\uparrow}\}$ with $x_{n_k}^{\uparrow} \notin D_{u,\uparrow}^*$ for all n_k. Necessarily, then, $x_{n_k}^{\uparrow}(0) = 0$ for all n_k. Since $x_n^{\uparrow} \to x^{\uparrow}$, we can conclude that $x^{\uparrow}(0) = 0$. Since $x^{\uparrow}$ is right continuous and $x^{\uparrow} \in D_{u,\uparrow}^*$, for any $\epsilon > 0$, there exists δ, $0 < \delta < \epsilon/2$, such that $\delta \in Disc(x^{\uparrow})^c$ and $0 < x^{\uparrow}(\delta) < \epsilon/2$. Let n_0 then be such that $|x_n^{\uparrow}(0) - x^{\uparrow}(0)| < \epsilon/2$ and $|x_n^{\uparrow}(\delta) - x^{\uparrow}(\delta)| < \epsilon/2$ for all $n \ge n_0$. Hence, for $n \ge n_0$, we can define an approximation to $x_{n_k}^{\uparrow}$ which belongs to $D_{u,\uparrow}^*$. In particular, let $x_{n_k}^*(0) = x_{n_k}^{\uparrow}(0) = 0$ and let $x_{n_k}^*(t) = x_{n_k}^{\uparrow}(t)$ for all $t \ge \delta$ and let $x_{n_k}^*$ be defined by linear interpolation in $[0, \delta]$. Then $x_{n_k}^* \in D_{u,\uparrow}^*$, $\|x_{n_k}^* - x_{n_k}^{\uparrow}\| < \epsilon$ and $\|(x_{n_k}^*)^{-1} - x_{n_k}^{-1}\| < \epsilon$ for all $n_k \ge n_0$. For $n \ge n_0$ such that $x_n^{\uparrow} \in D_{u,\uparrow}^*$, let $x_n^* = x_n^{\uparrow}$. Since ϵ was arbitrary, we can choose x_n^* such that $x_n^* \to x^{\uparrow}$ (M_1), $\|x_n^* - x_n^{\uparrow}\| \to 0$ and $\|(x_n^*)^{-1} - x_n^{-1}\| \to 0$ as $n \to \infty$. By Lemma 7.6.2, $(x_n^*)^{-1} \to x^{-1}$ (M_1). Since $\|(x_n^*)^{-1} - x_n^{-1}\| \to 0$, $x_n^{-1} \to x^{-1}$ (M_1) as well. ∎

Corollary 7.6.2. . (continuity at strictly increasing functions) *The inverse map from (D_u, M_2) to $(D_{u,\uparrow}, U)$ is continuous at $x \in D_{u,\uparrow\uparrow}$.*

Proof. First, $D_{u,\uparrow\uparrow} \subseteq D_{u,\uparrow}^*$, so that we can apply Theorem 7.6.1 to get $x_n^{-1} \to x^{-1}$ in $(D_{u,\uparrow}, M_1)$. However, by Lemma **??**, $x^{-1} \in C$ when $x \in D_{u,\uparrow\uparrow}$. Hence the M_1 convergence $x_n^{-1} \to x^{-1}$ actually holds in the stronger topology of uniform convergence over compact subsets. ∎

7.6.2. The M_1' Topology

For cases in which the condition $x^{-1}(0) = 0$ in Theorem 7.6.1 is not satisfied, we can modify the M_1 and M_2 topologies to obtain convergence, following Puhalskii and Whitt (1997). With these new weaker topologies, which we call M_1' and M_2', we do not require that $x_n(0) \to x(0)$ when $x_n \to x$. We construct the new topologies by extending the graph of each function x by appending the segment $[0, x(0)] \equiv \{\alpha 0 + (1-\alpha)x(0) : 0 \le \alpha \le 1\}$. Let the new graph of $x \in D$ be

$$\begin{aligned}\Gamma_x' \quad = \quad & \{(z,t) \in \mathbb{R}^k \times [0,\infty) : z = \alpha x(t) + (1-\alpha)x(t-) \\ & \text{for } 0 \le \alpha \le 1 \text{ and } t \ge 0\}\,, \end{aligned} \tag{6.5}$$

where $x(0-) \equiv 0$. Let $\Pi'(x)$ and $\Pi_2'(x)$ be the sets of all M_1 and M_2 parametric representations of Γ_x', defined just as before. We say that $x_n \to x$ in (D, M_i') if there exist parametric representations $(u_n, r_n) \in \Pi'(x_n)$ and $(u, r) \in \Pi'(x)$, where Π' is the set of M_1' and M_2' parametric representations, such that

$$\|u_n - u\|_t \vee \|r_n - r\|_t \to 0 \quad \text{as} \quad n \to \infty \quad \text{for each} \quad t > 0\,. \tag{6.6}$$

With the M_1' topologies, we obtain a cleaner statement than Lemma 7.6.2.

Lemma 7.6.3. (graphs of the inverse with the M_1' topology) *For $x \in D_{u,\uparrow}$, the graph Γ_x' serves as the graph $\Gamma_{x^{-1}}'$ with the axes switched, so that $(u, r) \in \Pi'(x)$ $(\Pi_2'(x))$ if and only if $(r, u) \in \Pi'(x^{-1})$ $(\Pi_2'(x^{-1}))$.*

Thus we get an alternative to Theorem 7.6.1.

Theorem 7.6.2. (continuity in the M_1' topology) *The inverse map in (6.1) from (D_u, M_2') to $(D_{u,\uparrow}, M_1')$ is continuous.*

Proof. By the M_2' analog of Theorem 7.4.1, if $x_n \to x$ in (D_u, M_2'), then $x_n^{\uparrow} \to x^{\uparrow}$ in $(D_{u,\uparrow}, M_2')$. Since the M_2' topology coincides with the M_i' topology on $D_{\uparrow}$, we get $x_n^{\uparrow} \to x^{\uparrow}$ in $(D_{u,\uparrow}, M_1')$. By Lemma 7.6.3, we get $(x_n^{\uparrow})^{-1} \to (x^{\uparrow})^{-1}$ in $(D_{u,\uparrow}, M_1')$. That gives the desired result because $(x^{\uparrow})^{-1} = x^{-1}$ for all $x \in D_u$. ∎

An alternative approach to the difficulty at the origin besides M_i' topology on $D_u([0,\infty), \mathbb{R})$ is the ordinary M_i topology on $D_u((0,\infty), \mathbb{R})$. The difficulty at the origin goes away if we ignore it entirely, which we can do by making the function domain $(0, \infty)$ for the image of the inverse functions.

In particular, Theorem 7.6.2 implies the following corollary.

Corollary 7.6.3. (continuity when the origin is removed from the domain) *The inverse map in* (6.1) *from* $D_u([0,\infty), M_2)$ *to* $D_{u,\uparrow}((0,\infty), M_1)$ *is continuous.*

Proof. Since the M_2' topology is weaker than M_2, if $x_n \to x$ in $D_u([0,\infty), M_2)$, then $x_n \to x$ in $D_u([0,\infty), M_2')$. Apply Theorem 7.6.2 to get $x_n^{-1} \to x^{-1}$ in $D_{u,\uparrow}([0,\infty), M_1')$. That implies $x_n^{-1} \to x^{-1}$ for the restrictions in $D_\uparrow([t_1,t_2], M_1)$ for all $t_1, t_2 \in Disc(x^{-1})^c$, which in turn implies that $x_n^{-1} \to x^{-1}$ in $D_{u,\uparrow}((0,\infty), M_1)$. ■

However, in general we cannot work with the inverse on $D_u((0,\infty), \mathbb{R})$. We can obtain positive results if all the functions are required to be monotone. The following result is elementary.

Theorem 7.6.3. (equivalent characterizations of convergence for monotone functions) *For* x_n, $n \geq 1$, $x \in D_{u,\uparrow}([0,\infty), \mathbb{R})$, *the following are equivalent:*

$$x_n \to x \quad in \quad D_{u,\uparrow}((0,\infty), \mathbb{R}, M_1)\ ; \tag{6.7}$$

$$x_n \to x \quad in \quad D_{u,\uparrow}([0,\infty), \mathbb{R}, M_1')\ ; \tag{6.8}$$

$$x_n(t) \to x(t) \quad for\ all\ t\ in\ a\ dense\ subset\ of\ (0,\infty)\ ; \tag{6.9}$$

$$x_n^{-1} \to x^{-1} \quad in \quad D((0,\infty), \mathbb{R}, M_1)\ ; \tag{6.10}$$

$$x_n^{-1} \to x^{-1} \quad in \quad D([0,\infty), \mathbb{R}, M_1')\ ; \tag{6.11}$$

$$x_n^{-1}(t) \to x^{-1}(t) \quad for\ all\ t\ in\ a\ dense\ subset\ of\ (0,\infty). \tag{6.12}$$

Proof. Theorem 7.6.2 implies the equivalence of (6.8) and (6.11). Clearly, (6.8)→(6.7)→(6.9), so that (6.11)→(6.10)→(6.12). It thus suffices to show that (6.9)→(6.8). For any $\epsilon > 0$, we can find t and n_0 such that $0 < t < \epsilon$, $t \in Disc(x)$ and $|x_n(t) - x(t)| < \epsilon$ for $n \geq n_0$. Let $n_1 \geq n_0$ be such that $d_{M_2'}(x_n, x) < \epsilon$ for the restrictions to $[t, t']$ for any $t' > t$ with $t' \in Disc(x)^c$. Since x_n and x are nondecreasing and nonnegative, the bounds $d_{M_2'}(x_n, x) < \epsilon$ over $[t, t']$ and $|x_n(t) - x(t)| < \epsilon$ imply that $d_{M_2'}(x_n, x) < \epsilon$ for the restrictions over$[0, t']$. Since ϵ and t' were arbitrary, $x_n \to x$ in $D_\uparrow([0,\infty), \mathbb{R}, M_2')$, but the M_2' and M_1' topologies are equivalent on $D_\uparrow$. ■

In general, convergence in $D([0,\infty), \mathbb{R}, M_1')$ provides stronger control of the behavior at the origin than convergence in $D((0,\infty), \mathbb{R}, M_1)$. Nothing more is omitted from Section 13.6 of the book.

7.7. Inverse with Centering

We continue considering the inverse map, but now with centering. We start by considering linear centering. In particular, we consider when a limit for $c_n(x_n - e)$ implies a limit for $c_n(x_n^{-1} - e)$ when $x_n \in D_u \equiv D_u([0,\infty),\mathbb{R})$ and $c_n \to \infty$. By considering the behavior at one t, it is natural to anticipate that we should have $c_n(x_n^{-1} - e) \to -y$ when $c_n(x_n - e) \to y$. A first step for the M topologies is to apply Theorem 7.4.2, which yields limits for $c_n(x_n^{\uparrow} - e)$. Thus for the M topologies, it suffices to assume that $x_n \in D_{\uparrow}$.

Now we state the main limit theorem for inverse functions with centering.

Theorem 7.7.1. *Suppose that $c_n(x_n - e) \to y$ as $n \to \infty$ in $D([0,\infty),\mathbb{R})$ with one of the topologies M_2, M_1 or J_1, where $x_n \in D_u$, $c_n \to \infty$ and $y(0) = 0$.*

(a) If the topology is M_2 or M_1, then $c_n(x_n^{-1} - e) \to -y$ as $n \to \infty$ with the same topology.

(b) If the topology is J_1 and if y has no positive jumps, then $c_n(x_n^{-1} - e) \to -y$ as $n \to \infty$.

Proof. (a) The proof is easy for the M_i topologies when $x_n \in D_u^*$ for all sufficiently large n. First, given $c_n(x_n - e) \to y$ (M_i), we can apply Theorem 7.4.2 (a) to conclude that $c_n(x_n^{\uparrow} - e) \to y$ (M_i). Hence we can assume that $x_n \in D_{\uparrow}^*$. Thus there exist parametric representations $(u_n, r_n) \in \Pi(c_n(x_n - e))$ and $(u, r) \in \Pi(x)$ of the appropriate type such that $\|u_n - u\|_t \vee \|r_n - r\|_t \to 0$ as $n \to \infty$ for all $t > 0$. Then $(u_n', r_n) \in \Pi(x_n)$ for $u_n' = c_n^{-1}u_n + r_n$. Since $x_n \in D_{\uparrow}^*$ for n sufficiently large, $(r_n, u_n') \in \Pi(x_n^{-1})$ and $(c_n(r_n - u_n'), u_n') \in \Pi(c_n(x_n^{-1} - e))$ for sufficiently large n. However,

$$c_n(r_n - u_n') = -u_n \tag{7.1}$$

and

$$\|u_n' - r\|_t \to 0 \quad \text{as } n \to \infty \text{ for all } t > 0\,, \tag{7.2}$$

so that $c_n(x_n^{-1} - e) \to -y$ (M_i) as $n \to \infty$. However, in general we need not have $x_n \in D_u^*$ for all sufficiently large n. So, suppose that we do not. We then only have $x_n \in D_u$ for all n. As before, we can apply Theorem 7.4.2 to show that it suffices to assume that $x_n \in D_{\uparrow}$ for all n. We now show that we can approximate $x_n \in D_{\uparrow}$ by $x_n^* \in D_{\uparrow}^*$ for all n sufficiently large, so that

$$c_n\|x_n - x_n^*\| \to 0 \quad \text{and} \quad c_n\|x_n^{-1} - (x_n^*)^{-1}\| \to 0 \quad \text{as} \quad n \to \infty\,. \tag{7.3}$$

The limits in (7.3) plus the triangle inequality imply that

$$d(c_n(x_n^* - e), y) \leq d(c_n(x_n - e), y) + c_n\|x_n - x_n^*\| \to 0 \quad \text{as} \quad n \to \infty \quad (7.4)$$

and

$$\begin{aligned} &d(c_n(x_n^{-1} - e), -y) \\ &\quad \leq \|c_n(x_n^{-1} - (x_n^*)^{-1}\| + d(c_n((x_n^*)^{-1} - e), -y) \to 0 \end{aligned} \quad (7.5)$$

as $n \to \infty$, where d is the M_i metric. Thus, the remaining problem is to construct $x_n^* \in D_\uparrow^*$ satisfying (7.3). Since $y(0) = 0$ and $y \in D$, for all $\epsilon > 0$, there exists δ_1 such that $v(y, 0, \delta_1) < \epsilon/2$. Since $c_n(x_n - e) \to y$ (M_2), there exists n_0 and δ_2 such that $v(c_n(x_n - e), y, 0, \delta/2) < \epsilon/2$ for all $n \geq n_0$. Thus

$$t - c_n^{-1}\epsilon < x_n(t) \leq t + c_n^{-1}\epsilon \quad (7.6)$$

for all $n \geq n_0$ and t with $0 \leq t \leq \delta \equiv \delta_1 \wedge \delta_2$. By Lemma **??**,

$$t + c_n^{-1}\epsilon > x_n^{-1}(t-) \geq t - c_n^{-1}\epsilon \quad (7.7)$$

for all $n \geq n_0$ and t with $0 \leq t \leq \delta - c_n^{-1}\epsilon$. Now choose $n_1 \geq n_0$ so that $c_n^{-1}\epsilon < \delta/4$ for all $n \geq n_1$. Then, by (7.6), for $n \geq n_1$,

$$0 < x_n(\delta/4) < \delta/2 \quad (7.8)$$

and (7.7) holds for $0 \leq t \leq 3\delta/4$. Hence, if $n \geq n_1$ and $x_n \notin D_\uparrow^*$, we can construct $x_n^* \in D_\uparrow^*$ by letting $x_n^*(0) = x_n(0) = 0$, $x_n^*(t) = x_n(t)$, $t \geq \delta/4$, and letting x_n^* be defined by linear interpolation for t in $[0, \delta/4]$. By (7.8), $x_n^* \in D_\uparrow^*$. Since x_n^* is defined by linear interpolation over $[0, \delta/4]$, for $n \geq n_1$,

$$\|c_n(x_n^* - e)\|_{\delta/4} = \max\{c_n(x_n - e)(0), c_n(x_n - e)(\delta/4)\} \leq \epsilon\ , \quad (7.9)$$

so that

$$\|c_n(x_n^* - x_n)\| \leq \|c_n(x_n - e)\|_{\delta/4} + \|c_n(x_n^* - e)\|_{\delta/4} \leq 2\epsilon\,. \quad (7.10)$$

Similarly, $(x_n^*)^{-1}(t) = x_n^{-1}(t)$ for $t \leq x_n(\delta/4) < \delta/2$ and $n \geq n_1$, so that by (7.7)

$$\|c_n((x_n^*)^{-1} - x_n^{-1})\| \leq \|c_n(x_n^{-1} - e)\|_{\delta/2} + \|c_n((x_n^*)^{-1} - e)\|_{\delta/2} \leq 2\epsilon\,. \quad (7.11)$$

Since ϵ was arbitrary, (7.10) and (7.11) imply (7.3), as required.

(b) Since $c_n(x_n - e) \to y$ (J_1) and $c_n \to \infty$, $\|x_n - e\|_t \to 0$ as $n \to \infty$ for all $t > 0$. By Corollary 7.6.2, $\|x_n^{-1} - e\|_t \to 0$ as $n \to \infty$ for each

$t > 0$. By Theorem 7.2.2, we can apply the composition map to obtain $c_n(x_n \circ x_n^{-1} - x_n^{-1}) \to y \; (J_1)$. Hence it suffices to show that $c_n\|x_n \circ x_n^{-1} - e\|_t \to 0$ as $n \to \infty$ for all $t > 0$. However, by Corollary **??**,

$$\begin{aligned} c_n\|x_n \circ x_n^{-1} - e\|_t &\leq c_n J_{x_n^{-1}(t)}(x_n) \\ &= J_{x_n^{-1}(t)}(c_n(x_n - e)), \end{aligned} \tag{7.12}$$

where $J_t(x)$ is the maximum jump of x over $[0, t]$, treating $x(0-)$ as 0. Since $c_n(x_n - e) \to y$, $y(0) = 0$ and y has no positive jumps, $J_t(c_n(x_n - e)) \to 0$ as $n \to \infty$ for all $t > 0$, which implies the desired conclusion. ■

Nothing else is omitted from Section 13.7 of the book.

7.8. Counting Functions

Inverse functions or first-passage-time functions are closely related to counting functions. A counting function is defined in terms of a sequence $\{s_n : n \geq 0\}$ of nondecreasing nonnegative real numbers with $s_0 = 0$. We can think of s_n as the partial sum

$$s_n \equiv x_1 + \cdots + x_n, \quad n \geq 1, \tag{8.1}$$

by simply writing $x_i \equiv s_i - s_{i-1}$, $i \geq 1$. The associated *counting function* $\{c(t) : t \geq 0\}$ is defined by

$$c(t) \equiv \max\{k \geq 0 : s_k \leq t\}, \quad t \geq 0. \tag{8.2}$$

To have $c(t)$ finite for all $t > 0$, we assume that $s_n \to \infty$ as $n \to \infty$. We can reconstruct the sequence $\{s_n\}$ from $\{c(t) : t \geq 0\}$ by

$$s_n = \inf\{t \geq 0 : c(t) \geq n\}, \quad n \geq 0. \tag{8.3}$$

The sequence $\{s_n\}$ and the associated function $\{c(t) : t \geq 0\}$ can serve as sample paths for a stochastic point process on the nonnegative real line. Then there are (countably) infinitely many points with the n^{th} point being located at s_n. The summands x_n are then the intervals between successive points. The most familiar case is when the sequence $\{x_n : n \geq 1\}$ constitutes the possible values from a sequence $\{X_n : n \geq 1\}$ of i.i.d. random variables with values in $\mathbb{R}_+$. Then the counting function $\{c(t) : t \geq 0\}$ constitutes a possible sample path of an associated renewal counting process $\{C(t) : t \geq 0\}$; see Section 7.3 of the book.

Paralleling Lemma 13.6.3 in the book, we have the following basic inverse relation for counting functions.

Lemma 7.8.1. *For any nonnegative integer n and nonnegative real number t,*

$$s_n \le t \quad \textit{if and only if} \quad c(t) \ge n\,. \tag{8.4}$$

We can put counting functions in the setting of inverse functions on $D_\uparrow$ by letting

$$y(t) \equiv s_{\lfloor t \rfloor}, t \ge 0\,. \tag{8.5}$$

To have $y \in D_\uparrow$, we use the assumption that $s_n \to \infty$ as $n \to \infty$. if all the summands are strictly positive then

$$y^{-1}(t) = c(t) + 1, \quad t \ge 0\,, \tag{8.6}$$

where y^{-1} is the image of the inverse map in (6.1) applied to y in (8.5). With (8.6), limits for counting functions can be obtained by applying results in the previous two sections.

The connection to the inverse map can also be made when the summands x_i are only nonnegative. To do so, we observe that the counting function c is a time-transformation of y^{-1}. both are right-continuous, but $c(t) < y^{-1}(t)$. In particular, c and y can be expressed in terms of each other.

Lemma 7.8.2. (relation between counting functions and inverse functions) *For y in* (8.5) *and c in* (8.2),

$$\begin{aligned} c(t) &= y^{-1}(y(y^{-1}(t)-)-), \quad t \ge 0\,, & (8.7)\\ c(t) &= y^{-1}(t-) \quad \textit{for all} \quad t \in Disc(c) = Disc(y^{-1})\,, & (8.8)\\ y^{-1}(t) &= c(c^{-1}(c(t))), \quad t \ge 0\,. & (8.9) \end{aligned}$$

The three functions y, y^{-1} and c are depicted for a typical initial segment of a sequence $\{s_n : n \ge 0\}$ in Figure 13.1 of the book. We can apply (8.7)–(8.9) in Lemma 7.8.1 to show that limits for scaled counting functions with centering, are equivalent to limits for scaled inverse functions. We use the fact that the M topologies are not altered by changing to the left limits, because the graph is unchanged. We first consider the case of no centering; afterwards we consider the case of centering. When there is no centering, the M_1 and M_2 topologies coincide and reduce to pointwise convergence on a dense subset of $\mathbb{R}_+$ including 0.

Consider a sequence of counting functions $\{\{c_n(t) : t \ge 0\} : n \ge 1\}$ with associated processes

$$y_n^{-1}(t) \equiv c_n(c_n^{-1}(c_n(t))), \quad t \ge 0\,, \tag{8.10}$$

$y_n = (y_n^{-1})^{-1}$. Form scaled functions by setting

$$\mathbf{c}_n(t) = n^{-1}c_n(a_n t) \quad \text{and} \quad \mathbf{y}_n(t) = a_n^{-1}y_n(nt), \quad t \geq 0\,, \tag{8.11}$$

where a_n are positive real numbers with $a_n \to \infty$. Note that

$$\mathbf{c}_n^{-1}(t) = a_n^{-1}c_n^{-1}(nt) \quad \text{and} \quad \mathbf{y}_n^{-1}(t) = n^{-1}y_n(a_n t), \quad t \geq 0\,. \tag{8.12}$$

Theorem 7.8.1. (asymptotic equivalence of limits for scaled processes) *Suppose that* $\mathbf{y}_n \in D_{u,\uparrow}$, $n \geq 1$, *for* $\mathbf{y}_n$ *in* (8.11). *Then any one of the limits* $\mathbf{y}_n \to y$, $\mathbf{y}_n^{-1} \to y^{-1}$, $\mathbf{c}_n \to y^{-1}$ *or* $\mathbf{c}_n^{-1} \to y^{-1}$ *in* $D_{\uparrow}([0,\infty),\mathbb{R})$ *with the* M_2 $(= M_1)$ *topology, for* y_n^{-1}, $\mathbf{c}_n$ *and* $\mathbf{c}_n^{-1}$ *in* (8.11) *and* (8.12), *implies the others.*

Proof. The equivalence between $\mathbf{y}_n \to y$ and $\mathbf{y}_n^{-1} \to y^{-1}$, and between $\mathbf{c}_n \to y^{-1}$ and $\mathbf{c}_n^{-1} \to y$ follow from Theorem 7.6.1. We can relate the limits $\mathbf{c}_n \to y^{-1}$ and $\mathbf{y}_n \to y$ by applying (8.6), after modifying the summands $x_{n,i}$ in the sequences $\{s_{n,k} : k \geq 0\}$ to make them strictly positive. We can show that the limits are unaltered by adding suitably small positive values to the summands. Given $\epsilon > 0$ and $\{x_n : n \geq 1\}$, let

$$x'_n = x_n + \epsilon 2^{-n}, n \geq 1\,, \tag{8.13}$$

and let $x'_n = x'_1 + \cdots + x'_n$, $n \geq 1$, and $c'(t) = \max\{k \geq 0 : s'_n \leq t\}$, $t \geq 0$. Then

$$s_n \leq s'_n \leq s_n + \epsilon, \quad n \geq 0\,, \tag{8.14}$$

and

$$c((t-\epsilon) \vee 0) \leq c'(t) \leq c(t), \quad t \geq 0\,. \tag{8.15}$$

The actual limits we want to consider involve a sequence of sequences $\{\{s_{n,k} : k \geq 0\}, n \geq 1\}$ with $s_{n,0} = 0$ for each n. Let $\{\{c_n(t) : t \geq 0\}\}$ be the associated sequence of counting functions. Let $x'_{n,k}$ $s'_{n,k}$, $n'_n(t)$, $\mathbf{s}'_n$ and $\mathbf{n}'_n$ be associated quantities defined by the modification in (8.7), i.e., by letting

$$x'_{n,k} \equiv x_{n,k} + \epsilon_n 2^{-k}, \quad k \geq 1\,. \tag{8.16}$$

Given that scaled processes are formed as in (8.11) and (8.12). It is elementary that

$$\|\mathbf{y}_n - \mathbf{y}'_n\|_\infty \leq \epsilon_n / a_n \to 0 \tag{8.17}$$

so that, for appropriate choice of ϵ_n, e.g., $\epsilon_n = \epsilon$, $\epsilon_n/a_n \to 0$. The bound in (8.15) enables us to conclude that $\mathbf{c}_n \to c$ (M_2) if and only if $\mathbf{c}'_n \to c$ (M_2) by applying Corollary 12.11.6 in the book. Hence it suffices to assume that the sequences $\{s_{n,k} : k \geq 0\}$ are strictly increasing, which implies that (8.6) holds. Then, after scaling as in (8.11) and (8.12),

$$\|\mathbf{y}_n^{-1} - \mathbf{c}_n\|_\infty \leq 1/n \to 0\,,$$

which completes the proof. ∎

We now apply the results for inverse maps with centering in Section 7.7 to obtain limits for counting functions with centering. Consider a sequence of counting functions $\{\{c_n(t) : t \geq 0\} : n \geq 1\}$ associated with a sequence of nondecreasing sequences of nonnegative numbers $\{\{s_{n,k} : k \geq 0\} : n \geq 1\}$ defined as in (8.2). Let the scaled functions $\mathbf{c}_n$, $\mathbf{y}_n$, $\mathbf{c}_n^{-1}$ and $\mathbf{y}_n^{-1}$ be defined as in (8.10)–(8.12).

Theorem 7.8.2. (asymptotic equivalence of counting and inverse functions with centering) *Suppose that* $\mathbf{y}_n \in D_\uparrow$, $n \geq 1$, $b_n \to \infty$ *and* $y(0) = 0$. *Then any one of the limits* $b_n(\mathbf{y}_n - e) \to y$, $b_n(\mathbf{c}_n - e) \to -y$, $b_n(\mathbf{y}_n^{-1} - e) \to -y$ *or* $b_n(\mathbf{c}_n^{-1} - e) \to y$ *in* $D([0,\infty),\mathbb{R})$ *with the* M_1 *or* M_2 *topology, for* $\mathbf{y}_n$, $\mathbf{c}_n$, *and* $\mathbf{y}_n^{-1}$ *and* $\mathbf{c}_n^{-1}$ *in* (8.11) *and* (8.12), *implies the others with the same topology.*

Proof. The equivalence between $b_n(\mathbf{y}_n - e) \to y$ and $b_n(\mathbf{y}_n^{-1} - e) \to -y$ is contained in Theorem 7.7.1. Similarly, the equivalence between $b_n(\mathbf{c}_n - e) \to -y$ and $b_n(\mathbf{c}_n^{-1} - e) \to y$ is contained in Theorem 7.7.1. Let the topology be fixed at either M_1 or M_2. Given $b_n(\mathbf{y}_n^{-1} - e) \to -y$, we have $\|\mathbf{y}_n^{-1} - e\|_t \to 0$ and $\|\mathbf{y}_n - e\|_t \to 0$ as $n \to \infty$ for each $t > 0$. For any $x \in D$, let $\hat{x}$ denote the associated left-limit function; i.e., $\hat{x}(t) = x(t-)$. Then $\mathbf{c}_n = \hat{\mathbf{y}}_n^{-1} \circ \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1}$. Given $b_n(\mathbf{y}_n^{-1} - e) \to -y$, we have $\hat{\mathbf{y}}_n^{-1} \to e$, $\hat{\mathbf{y}}_n \to e$, $b_n(\hat{\mathbf{y}}_n^{-1} - e) \to -y$ and $b_n(\hat{\mathbf{y}}_n - e) \to y$, because the graphs are unchanged. Now we can apply the composition map to get $b_n(\hat{\mathbf{y}}_n^{-1} \circ \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1} - \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1}) \to -y$ and $b_n(\hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1} - \mathbf{y}_n^{-1}) \to y$. Hence, by Proposition **??**, for each $t \in Disc(y)^c$, we have

$$\begin{aligned} b_n(\mathbf{c}_n - e)(t) &= b_n(\hat{\mathbf{y}}_n^{-1} \circ \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1} - e)(t) \\ &= b_n(\hat{\mathbf{y}}_n^{-1} \circ \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1} - \hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1})(t) \\ &\quad + b_n(\hat{\mathbf{y}}_n \circ \mathbf{y}_n^{-1} - \mathbf{y}_n^{-1})(t) + b_n(\mathbf{y}_n^{-1} - e)(t) \\ &\quad \to -y(t) + y(t) - y(t) = -y(t)\,. \end{aligned} \tag{8.18}$$

Now we apply Theorems 6.5.1 (iv) and 6.11.1 (iv). Let $w(x,\delta)$ be the M_i oscillation function over the interval $[0,t]$. By (8.8), the oscillations of $b_n(\mathbf{c}_n - e)$ coincide with the oscillations of $b_n(\mathbf{y}_n^{-1} - e)$ at discontinuity points of $\mathbf{c}_n$ and $\mathbf{y}_n^{-1}$. Moreover, in between such discontinuity points, they have identical maximum oscillations. Hence, for any interval $[0,t]$ with $t \in Disc(y)^c$,

$$w(b_n(\mathbf{c}_n - e), \delta) < w(b_n(\mathbf{y}_n^{-1} - e), \delta) \ . \tag{8.19}$$

Since $b_n(\mathbf{y}_n^{-1} - e) \to -y$ by assumption,

$$\lim_{\delta \downarrow 0} \ \overline{\lim_{n\to\infty}} \ w(b_n(\mathbf{y}_n^{-1} - e), \delta) = 0 \tag{8.20}$$

Consequently,

$$\lim_{\delta \downarrow 0} \ \overline{\lim_{n\to\infty}} \ w(b_n(\mathbf{c}_n - e), \delta) = 0 \ . \tag{8.21}$$

Hence, we can conclude that $b_n(\mathbf{c}_n - e) \to -y$.

To go the other way, suppose that $b_n(\mathbf{c}_n - e) \to -y$. Applying Theorem 7.7.1, we have $b_n(\mathbf{c}_n^{-1} - e) \to y$, $\mathbf{c}_n \to e$ and $\mathbf{c}_n^{-1} \to e$. Then, paralleling (8.18), we can apply (8.9) to obtain

$$\begin{aligned} b_n(\mathbf{y}_n^{-1} - e)(t) &= b_n(\mathbf{c}_n \circ \mathbf{c}_n^{-1} \circ \mathbf{c}_n - e)(t) \\ &= b_n(\mathbf{c}_n \circ \mathbf{c}_n^{-1} \circ \mathbf{c}_n - \mathbf{c}_n^{-1} \circ \mathbf{c}_n)(t) \\ &\quad + b_n(\mathbf{c}_n^{-1} \circ \mathbf{c}_n - \mathbf{c}_n)(t) + b_n(\mathbf{c}_n - e)(t) \\ &\quad \to -y(t) + y(t) - y(t) = -y(t) \end{aligned} \tag{8.22}$$

for each $t \in Disc(y)^c$. Now let $w(x,\delta,t)$ denote the M_i oscillation function over the interval $[0,t]$ as a function of the right endpoint t. Then, paralleling (8.19), by (8.8), for all $t_1 \in Disc(y)^c$, there exists $t_2 > t_1$ with $t_2 \in Disc(y)^c$ such that

$$w(b_n(\mathbf{y}_n^{-1} - e), \delta, t_1) < w(b_n(\mathbf{c}_n - e), \delta, t_2) \tag{8.23}$$

for all n sufficiently large. Hence we can use the previous oscillation argument to conclude that $b_n(\mathbf{y}_n^{-1} - e) \to -y$. ∎

7.9. Renewal-Reward Processes

Nothing was omitted from Section 13.9 in the book.

Chapter 8

Queueing Networks

8.1. Introduction

This chapter contains proofs omitted from Chapter 14 of the book, with the same title. Section 8.9 also contains supplementary material on the existence of a limiting stationary version for a general reflected process. With the exception of Section 8.9, the section and theorem numbering here parallels Chapter 14, so that the proofs should be easy to find.

Here is how this chapter is organized: We start in Section 8.2 by carefully defining the multidimensional reflection map and establishing its basic properties. Since the definition (Definition 8.2.1) is somewhat abstract, a key property is having the reflection map be well defined; i.e., we show that there exists a unique function satisfying the definition (Theorem 8.2.1). We also provide multiple characterizations of the reflection map, one alternative being as the unique fixed point of an appropriate operator (Theorem 8.2.2), while another is a basic complementarity property (Theorem 8.2.3).

A second key property of the multidimensional reflection map is Lipschitz continuity in the uniform norm on $D([0,T],\mathbb{R}^k)$ (Theorem 8.2.5). We also establish continuity of the multidimensional reflection map as a function of the reflection matrix, again in the uniform topology (Theorems 8.2.8 and 14.2.9 in the book). It is easy to see that the Lipschitz property is inherited when the metric on the domain and range is changed to d_{J_1} (Theorem 8.2.7). However, a corresonding direct extension for the SM_1 metric d_s does not hold. Much of the rest of the chapter is devoted to obtaining positive results for the M_1 topologies.

Section 8.3 provides yet another characterization of the multidimensional reflection map via an associated instantaneous reflection map on $\mathbb{R}^k$.

Sections 8.4 and 8.5 are devoted to obtaining the M_1 continuity results.

In Section 8.4 we establish properties of reflection of parametric representations. We are able to extend Lipschitz and continuity results from the uniform norm to the M_1 metrics when we can show that the reflection of a parametric representation can serve as the parametric representation of the reflected function. The results are somewhat complicated, because this property holds only under certain conditions.

In Sections 8.6 and 8.7, respectively, we apply the previous results to obtain heavy-traffic stochastic-process limits for stochastic fluid networks and conventional queueing networks. In the queueing networks we allow service interruptions. When there are heavy-tailed distributions or rare long service interruptions, the M_1 topologies play a critical role.

In Section 8.8 we consider the two-sided regulator and other reflection maps. The two-sided regulator is used to obtain heavy-traffic limits for single queues with finite waiting space, as considered in Section 2.3 and Chapter 5 of the book. With the scaling, the size of the waiting room is allowed to grow in the limit as the traffic intensity increases, but at a rate such that the limit process involves a two-sided regulator (reflection map) instead of the customary one-sided one. Like the one-sided reflection map, the two-sided regulator is continuous on (D^1, M_1). Moreover, the content portion of the two-sided regulator is Lipschitz, but the two regulator portions (corresponding to the two barriers) are only continuous; they are not Lipschitz.

We also give general conditions for other reflection maps to have M_1 continuity and Lipschitz properties. For these, we require that the limit function to be reflected belong to D_1, the subset of functions with discontinuities in only one coordinate at a time.

In Section 8.9 we show that reflected stochastic processes have proper limiting stationary distributions and proper limiting stationary versions (stochastic-process limits for the entire time-shifted processes) under very general conditions. Our main result, Theorem 8.9.1, establishes such limits for stationary ergodic net-input stochastic processes satisfying a natural drift condition (9.7). It is noteworthy that a proper limit can exist even if there is positive drift in some (but not all) coordinates. Theorem 8.9.1 is limited by having a special initial condition: starting out empty. Much of the rest of Section 8.9 is devoted to obtaining corresponding results for other initial conditions. Theorem 8.9.6 establishes convergence for all proper initial contents when the net input process is also a Lévy process with mutually independent coordinate processes. Theorem 8.9.6 covers limit processes obtained in the heavy-traffic limits for the stochastic fluid networks in Section 14.6 of the book.

8.2. The Multidimensional Reflection Map

We start by giving basic definitions and establishing alternative characterizations. Then we establish continuity and Lipschitz properties.

8.2.1. Definition and Characterization

Let $\mathcal{Q}$ be the set of all reflection matrices, i.e., the set of all column-stochastic matrices Q (with $Q^t_{i,j} \geq 0$ and $\sum_{j=1}^k Q^t_{i,j} \leq 1$) such that $Q^n \to 0$ as $n \to \infty$, where Q^n is the n^{th} power of Q.

Definition 8.2.1. (reflection map) *For any* $x \in D^k \equiv D([0,T], \mathbb{R}^k)$ *and any reflection matrix* $Q \in \mathcal{Q}$, *let the* feasible regulator *set be*

$$\Psi(x) \equiv \{w \in D^k_\uparrow : x + (I - Q)w \geq 0\} \tag{2.1}$$

and let the reflection map *be* $R \equiv (\psi, \phi) : D^k \to D^{2k}$ *with* regulator component

$$y \equiv \psi(x) \equiv \inf \Psi(x) \equiv \inf\{w : w \in \Psi(x)\} \ , \tag{2.2}$$

i.e.,

$$y^i(t) \equiv \inf\{w^i(t) \in \mathbb{R} : w \in \Psi(x)\} \quad \text{for all} \quad i \quad \text{and} \quad t \ , \tag{2.3}$$

and content component

$$z \equiv \phi(x) \equiv x + (I - Q)y \ . \tag{2.4}$$

It remains to show that the reflection map is well defined by Definition 8.2.1; i.e., we need to know that the feasible regulator set $\Psi(x)$ is nonempty and that its infimum y (which necessarily is well defined and unique for nonempty $\Psi(x)$) is itself an element of $\Psi(x)$, so that $z \in D^k$ and $z \geq 0$.

To show that $\Psi(x)$ in (2.1) is nonempty, we exploit the well known fact that the matrix $I - Q$ has nonnegative inverse.

Lemma 8.2.1. (nonnegative inverse of reflection matrix) *For all* $Q \in \mathcal{Q}$, $I - Q$ *is nonsingular with nonnegative inverse*

$$(I - Q)^{-1} = \sum_{n=0}^{\infty} Q^n \ ,$$

where $Q^0 = I$.

Proof. Note that

$$(I - Q)(I + Q + \cdots + Q^{n-1}) = I - Q^n \ . \tag{2.5}$$

Since $Q^n \to 0$ as $n \to \infty$, $I - Q^n \to I$ as $n \to \infty$, where I has determinant 1. Hence, for all sufficiently large n, the left and right sides of (2.5) have nonzero determinant. Since the determinant of the product of two matrices is the product of the determinants, the determinant of $I - Q$ must be nonzero, so that $I - Q$ must be nonsingular. Now multiply both sides of (2.5) by this inverse, which we have shown exists, to obtain

$$I + Q + \cdots + Q^{n-1} = (I - Q)^{-1}(I - Q^n) \ .$$

Since the right side tends to the proper limit $(I - Q)^{-1}$ as $n \to \infty$, so does the left. ∎

The key to showing that the infimum belongs to the feasibility set is a basic result about semicontinuous functions. Recall that a real-valued function x on $[0, T]$ is *upper semicontinuous* at a point t in its domain if

$$\limsup_{t_n \to t} x(t_n) \leq x(t)$$

for any sequence $\{t_n\}$ with $t_n \in [0, T]$ and $t_n \to t$ as $n \to \infty$. The function x is upper semicontinuous if it is upper semicontinuous at all arguments t in its domain.

Lemma 8.2.2. (preservation of upper semicontinuity) *Suppose that $\{x_s : s \in S\}$ is a set of upper semicontinuous real-valued function on a subinterval of $\mathbb{R}$. Then the infimum $\underline{x} \equiv \inf\{x_s : s \in S\}$ is also upper semicontinuous.*

Proof. For any t and $\epsilon > 0$ given, we need to find δ such that $\underline{x}(t') \leq \underline{x}(t) + \epsilon$ whenever $|t' - t| < \delta$. Since $\underline{x}$ is the infimum, for any t and ϵ, we can find $x \in \{x_s : s \in S\}$ such that $x(t) \leq \underline{x}(t) + \epsilon/2$. Since x is upper semicontinuous, there exists δ such that $x(t') \leq x(t) + \epsilon/2$ for all t' with $|t - t'| < \delta$. As a consequence,

$$\underline{x}(t') \leq x(t') \leq x(t) + \epsilon/2 \leq \underline{x}(t) + \epsilon$$

whenever $|t - t'| < \delta$. ∎

Recall that $x^{\uparrow} \equiv \sup_{0 \leq s \leq t} x(s)$, $t \geq 0$, for $x \in D^1$. For $x \equiv (x^1, \ldots, x^k) \in D^k$, let $x^{\uparrow} \equiv ((x^1)^{\uparrow}, \ldots, (x^k)^{\uparrow})$.

Theorem 8.2.1. (existence of the reflection map) *For any* $x \in D^k$ *and* $Q \in \mathcal{Q}$,

$$(I-Q)^{-1}[(-x)^{\uparrow} \vee 0] \in \Psi(x) \ , \tag{2.6}$$

so that $\Psi(x) \neq \phi$,

$$y \equiv \psi(x) \in \Psi(x) \subseteq D_{\uparrow}^k \tag{2.7}$$

for y *in (2.2) and*

$$z \equiv \phi(x) = x + (I-Q)y \geq 0 \ . \tag{2.8}$$

Proof. The proof is in the book. ∎

We now characterize the regulator function $y \equiv \psi(x)$ as the unique fixed point of a mapping $\pi \equiv \pi_{x,Q} : D_{\uparrow}^k \rightarrow D_{\uparrow}^k$, defined by

$$\pi(w) = (Qw - x)^{\uparrow} \vee 0 \tag{2.9}$$

for $w \in D_{\uparrow}^k$. For this purpose, we use two elementary lemmas.

Lemma 8.2.3. (feasible regulator set characterization) *The feasible regulator set* $\Psi(x)$ *in* (2.1) *can be characterized by*

$$\Psi(x) = \{w \in D_{\uparrow}^k : w \geq \pi(w)\}$$

for π *in* (2.9).

Proof. The proof is in the book. ∎

Lemma 8.2.4. (closed subset of D) *With the uniform topology on* D, *The feasible regulator set* $\Psi(x)$ *is a closed subset of* $D_{\uparrow}^k$, *while* $D_{\uparrow}^k$ *is a closed subset of* D.

Theorem 8.2.2. (fixed-point characterization) *For each* $Q \in \mathcal{Q}$, *the regulator map* $y \equiv \psi(x) \equiv \psi_Q(x) : D^k \rightarrow D_{\uparrow}^k$ *can be characterized as the unique fixed point of the map* $\pi \equiv \pi_{x,Q} : D_{\uparrow}^k \rightarrow D_{\uparrow}^k$ *defined in* (2.9).

Proof. The proof is in the book. ∎

Theorem 8.2.3. (complementarity characterization) *A function* y *in the feasible regulator set* $\Psi(x)$ *in* (2.1) *is the infimum* $\psi(x)$ *in* (2.2) *if and only if the pair* (y, z) *for* $z \equiv x + (I-Q)y$ *satisfies the complementarity property*

$$\int_0^\infty z^i dy^i = 0, \quad 1 \leq i \leq k \ . \tag{2.10}$$

Proof. The proof is in the book. ∎

8.2.2. Continuity and Lipschitz Properties

We now establish continuity and Lipschitz properties of the reflection map as a function of the function x and the reflection matrix Q. We use the *matrix norm*, defined for any $k \times k$ real matrix A by

$$\|A\| \equiv \max_j \sum_{i=1}^{k} |A_{i,j}| \ . \tag{2.11}$$

We use the maximum column sum in (2.11) because we intend to work with the column-substochastic matrices in $\mathcal{Q}$. Note that

$$\|A_1 A_2\| \leq \|A_1\| \cdot \|A_2\|$$

for any two $k \times k$ real matrices A_1 and A_2. Also, using the sum (or l_1) norm

$$\|u\| \equiv \sum_{i=1}^{k} |u^i| \tag{2.12}$$

on $\mathbb{R}^k$, we have

$$\|Au\| \leq \|A\| \cdot \|u\| \tag{2.13}$$

for each $k \times k$ real matrix A and $u \in \mathbb{R}^k$. Indeed, we can also define the matrix norm by

$$\|A\| \equiv \max\{\|Au\| : u \in \mathbb{R}^n, \|u\| = 1\} \ , \tag{2.14}$$

using the sum norm in (2.12) in both places on the right. Then (2.11) becomes a consequence. Consistent with (2.12), we let

$$\|x\| \equiv \sup_{0 \leq t \leq T} \|x(t)\| \equiv \sup_{0 \leq t \leq T} \sum_{i=1}^{k} \|x^i(t)\| \tag{2.15}$$

for $x \in D([0,T], \mathbb{R}^k)$. Combining (2.13) and (2.15), we have

$$\|Ax\| \leq \|A\| \cdot \|x\| \tag{2.16}$$

for each $k \times k$ real matrix A and $x \in D([0,T], \mathbb{R}^k)$.

We use the following basic lemma.

Lemma 8.2.5. (reflection matrix norms) *For any $k \times k$ matrix $Q \in \mathcal{Q}$,*

$$\|Q\| \leq 1, \quad \|Q^k\| = \gamma < 1 \tag{2.17}$$

and

$$\|(I - Q)^{-1}\| \leq \frac{k}{1 - \gamma} \ . \tag{2.18}$$

Proof. The first relation in (2.17) is immediate. Since $Q^n \to 0$ for all Q in $\mathcal{Q}$, the Markov chain associated with Q^t is transient. Since the Markov chain has k states,

$$\sum_{j=1}^{k} (Q_{i,j}^t)^k < 1 \ , \tag{2.19}$$

which, with (2.11), implies the second relation in (2.17). Probabilistically, if the probability of eventually exiting the state space $\{1, \ldots, k\}$ of the Markov chain is 1, then the probability of immediately exiting the state space from some state must be positive. Then the probability of reaching that state or the exterior (leaving the state space) in one step must be positive from some other state. Proceeding on by induction, the state space must be exhausted after k steps, so that (2.19) holds. Finally,

$$\left\| \sum_{n=0}^{\infty} Q^n \right\| \le \sum_{n=0}^{\infty} \|Q^n\| \le \sum_{n=0}^{k-1} \|Q^n\| + \gamma \sum_{n=0}^{\infty} \|Q^n\|$$

so that (2.18) holds. ■

We now show that $\pi \equiv \pi_{x,Q}$ in (2.9) is a k-stage contraction map on $D_{\uparrow}^k$. Recall that for $x \in D$, $|x|$ denotes the function $\{|x(t)| : t \ge 0\}$ in D, where $|x(t)| = (|x^1(t)|, \ldots, |x^k(t)|) \in \mathbb{R}^k$. Thus, for $x \in D$, $|x|^{\uparrow} = (|x^1|^{\uparrow}, \ldots, |x^k|^{\uparrow})$, where $|x^i|^{\uparrow}(t) = \sup_{0 \le s \le t} |x^i(s)|$, $0 \le t \le T$.

Lemma 8.2.6. (π is a k-stage contraction) *For any $Q \in \mathcal{Q}$ and $w_1, w_2 \in D_{\uparrow}^k$,*

$$|\pi^n(w_1) - \pi^n(w_2)|^{\uparrow} \le |Q^n(|w_1 - w_2|^{\uparrow})| \quad \textit{for} \quad n \ge 1 \ , \tag{2.20}$$

so that

$$\|\pi^n(w_1) - \pi^n(w_2)\| \le \|Q^n\| \cdot \|w_1 - w_2\| \le \|w_1 - w_2\| \tag{2.21}$$

for $n \ge 1$ and

$$\|\pi^n(w_1) - \pi^n(w_2)\| \le \gamma \|w_1 - w_2\| \quad \textit{for} \quad n \ge k \ ,$$

where

$$\|Q^k\| \equiv \gamma < 1 \ .$$

Hence

$$\|\pi^n(w) - \psi(x)\| \to 0 \quad \textit{as} \quad n \to \infty \ .$$

Proof. The proof is in the book. ∎

We now establish inequalities that imply that the reflection map is a Lipschitz continuous map on $(D, \|\cdot\|)$. We will use the stronger inequalities themselves in Section 8.9.

Theorem 8.2.4. (one-sided bounds) *For any $Q \in \mathcal{Q}$ and $x_1, x_2 \in D$,*

$$-(I-Q)^{-1}\eta_1(x_1 - x_2) \le \psi(x_1) - \psi(x_2) \le (I-Q)^{-1}\eta_1(x_2 - x_1) \quad (2.22)$$

where $\eta_1(x) \equiv (\hat{\eta}_1(x^1), \ldots, \hat{\eta}_1(x^k))$ with $\hat{\eta}_1 : D^1 \to D^1$ defined by

$$\hat{\eta}_1(x^i) \equiv (x^i)^{\uparrow} \vee 0 \ .$$

Proof. The proof is in the book. ∎

As a direct consequence of Theorem 8.2.4, we obtain the desired Lipschitz property.

Theorem 8.2.5. (Lipschitz property with uniform norm) *For any $Q \in \mathcal{Q}$ and $x_1, x_2 \in D$,*

$$\begin{aligned} \|\psi(x_1) - \psi(x_2)\| &\le \|(I-Q)^{-1}\| \cdot \|x_1 - x_2\| \\ &\le \sum_{n=0}^{\infty} \|Q^n\| \cdot \|x_1 - x_2\| \\ &\le \frac{k}{1-\gamma}\|x_1 - x_2\| \ , \end{aligned} \quad (2.23)$$

where $\gamma \equiv \|Q^k\| < 1$, and

$$\begin{aligned} \|\phi(x_1) - \phi(x_2)\| &\le (1 + \|I-Q\| \cdot \|(I-Q)^{-1}\|)\|x_1 - x_2\| \\ &\le \left(1 + \frac{2k}{1-\gamma}\right)\|x_1 - x_2\| \ . \end{aligned} \quad (2.24)$$

Proof. The proof is in the book. ∎

We now summarize some elementary but important properties of the reflection map.

Theorem 8.2.6. (reflection map properties) *The reflection map satisfies the following properties:*

(i) adaptedness*: For any $x \in D$ and $t \in [0, T]$, $R(x)(t)$ depends upon x only via $\{x(s) : 0 \le s \le t\}$.*

(ii) monotonicity*: If $x_1 \le x_2$ in D, then $\psi(x_1) \ge \psi(x_2)$.*

(iii) rescaling*: For each* $x \in D([0,T],\mathbb{R}^k)$*,* $\eta \in \mathbb{R}^k$*,* $\beta > 0$ *and* γ *nondecreasing right-continuous function mapping* $[0,T_1]$ *into* $[0,T]$*,* $\eta + \beta(x \circ \gamma) \in D([0,T_1],\mathbb{R}^k)$ *and*

$$R(\eta + \beta(x \circ \gamma)) = \beta R(\beta^{-1}\eta + x) \circ \gamma \ .$$

(iv) shift*: For all* $x \in D$ *and* $0 < t_1 < t_2 < T$*,*

$$\psi(x)(t_2) = \psi(x)(t_1) + \psi(\phi(x)(t_1) + x(t_1 + \cdot) - x(t_1))(t_2 - t_1)$$

and

$$\phi(x)(t_2) = \phi(\phi(x)(t_1) + x(t_1 + \cdot) - x(t_1))(t_2 - t_1)$$

(v) continuity preservation*: If* $x \in C$*, then* $R(x) \in C$*.*

We can apply Theorems 8.2.5 and 8.2.6 (iii) to deduce that the reflection map inherits the Lipschitz property on (D, J_1) from (D, U). Unfortunately, we will have to work harder to obtain related results for the M_1 topologies.

Theorem 8.2.7. (Lipschitz property with d_{J_1}) *For any* $Q \in \mathcal{Q}$*, there exist constants* K_1 *and* K_2 *(the same as in Theorem 8.2.5) such that*

$$d_{J_1}(\psi(x_1), \psi(x_2)) \le K_1 d_{J_1}(x_1, x_2) \tag{2.25}$$

and

$$d_{J_1}(\phi(x_1), \phi(x_2)) \le K_2 d_{J_1}(x_1, x_2) \tag{2.26}$$

for all $x_1, x_2 \in D$*.*

Proof. The proof is in the book. ∎

We now want to consider the reflection map R as a function of the reflection matrix Q as well as the net input function x. We first consider the maps $\pi \equiv \pi^n_{x,Q}(0)$ in (2.9) and $\psi \equiv \psi_Q$ in (2.2) as functions of Q when Q is a strict contraction in the matrix norm (2.11), i.e., when $\|Q\| < 1$.

Theorem 8.2.8. (stability bounds for different reflection matrices) *Let* $Q_1, Q_2 \in \mathcal{Q}$ *with* $\|Q_1\| = \gamma_1 < 1$ *and* $\|Q_2\| = \gamma_2 < 1$*. For all* $n \ge 1$*,*

$$\|\pi^n_{x,Q_j}(0)\| \le (1 + \gamma_j + \cdots + \gamma_j^{n-1})\|x\| \tag{2.27}$$

and

$$\|\pi^n_{x,Q_1}(0) - \pi^n_{x,Q_2}(0)\| \le (1 + \gamma_2 + \cdots + \gamma_2^{n-1}) \frac{\|x\| \cdot \|Q_1 - Q_2\|}{1 - \gamma_1} \ , \tag{2.28}$$

so that

$$\|\psi_{Q_j}(x)\| \le \frac{\|x\|}{1-\gamma_j} \tag{2.29}$$

and

$$\|\psi_{Q_1}(x) - \psi_{Q_2}(x)\| \le \frac{\|x\| \cdot \|Q_1 - Q_2\|}{(1-\gamma_1)(1-\gamma_2)} . \tag{2.30}$$

Proof. First

$$\|\pi^1_{x,Q_j}(0)\| = \|(-x)^\uparrow \vee 0\| \le \|x\| .$$

Next, by induction,

$$\begin{aligned} \|\pi^{n+1}_{x,Q_j}(0)\| &= \|(Q_j \pi^n_{x,Q_j}(0) - x)^\uparrow \vee 0| \\ &\le \|Q_j\| \cdot \|\pi^n_{x,Q_j}(0)\| + \|x\| \\ &\le \gamma_j(1+\gamma_j+\cdots+\gamma_j^{n-1})\|x\| + \|x\| \\ &\le (1+\gamma_j+\cdots+\gamma_j^n)\|x\| . \end{aligned}$$

Similarly, by induction

$$\begin{aligned} \|\pi^{n+1}_{x,Q_1}(0) - \pi^{n+1}_{x,Q_2}(0)\| &\le \|Q_1\pi^n_{x,Q_1}(0) - Q_2\pi_{x,Q_2}(0)\| \\ &\le \|Q_1\pi^n_{x,Q_1}(0) - Q_2\pi^n_{x,Q_1}(0)\| + \|Q_2\pi^n_{x,Q_1}(0) - Q_2\pi^n_{x,Q_2}(0)\| \\ &\le \|Q_1 - Q_2\| \cdot \|x\|/(1-\gamma_1) + \|Q_2\| \cdot \|\pi^n_{x,Q_1}(0) - \pi^n_{x,Q_2}(0)\| \\ &\le (1+\gamma_2+\cdots+\gamma_2^n)\|Q_1 - Q_2\| \cdot \|x\|/(1-\gamma_1) . \end{aligned}$$

Finally, since $\|\pi^n_{x,Q}(0) - \psi_Q(x)\| \to 0$ as $n \to \infty$, the final two bounds (2.29) and (2.30) follow. ■

Nothin more is omitted from Section 14.2 of the book.

8.3. The Instantaneous Reflection Map

Nothing has been deleted from this section in the book.

8.4. Reflections of Parametric Representations

In order to establish continuity and stronger Lipschitz properties of the reflection map R on D with the M_1 topologies, we would like to have $(R(u), r)$ be a parametric representation of $R(x)$ when (u, r) is a parametric representation of x. That is not always true, but we now obtain positive results in that direction.

Theorem 8.4.1. (reflections of parametric representations) *Suppose that* $x \in D$, $(u,r) \in \Pi_s(x)$ *and* $r^{-1}(t) = [s_-(t), s_+(t)]$.

(a) If $t \in Disc(x)^c$, *then*

$$R(u)(s) = R(x)(t) \quad \textit{for} \quad s_-(t) \le s \le s_+(t) \ .$$

(b) If $t \in Disc(x)$, *then*

$$R(u)(s_-(t)) = R(x)(t-) \quad \textit{and} \quad R(u)(s_+(t)) = R(x)(t) \ .$$

(c) If $t \in Disc(x)$ *and* $x(t) \ge x(t-)$, *then*

$$\phi(u)(s) = \phi(x)(t-) + \left(\frac{u^j(s) - u^j(s_-(t))}{u^j(s_+(t)) - u^j(s_-(t))} \right) [x(t) - x(t-)]$$

for any j, $1 \le j \le k$, *and*

$$\psi(u)(s) = \psi(x)(t-) = \psi(x)(t) \quad \textit{for} \quad s_-(t) \le s \le s_+(t) \ ,$$

so that

$$R(u)(s) \in [R(x)(t-), R(x)(t)] \quad \textit{for} \quad s_-(t) \le s \le s_+(t) \ .$$

(d) If $t \in Disc(x)$ *and* $x(t) \le x(t-)$, *then* $\phi^i(u)$ *and* $\psi^i(u)$ *are monotone in* $[s_-(t), s_+(t)]$ *for each* i, *so that*

$$R(u)(s) \in [[R(x)(t-), R(x)(t)]] \quad \textit{for} \quad s_-(t) \le s \le s_+(t) \ .$$

We can draw the desired conclusion that $(R(u), r)$ is a parametric representation of $R(x)$ if we can apply parts (c) and (d) of Theorem 8.4.1 to all jumps. Recall that D_+ (D_s) is the subset of D for which condition (c) (condition (c) or (d)) holds at all discontinuity points of x. For $x \in D_s$, the direction of the inequality is allowed to depend upon t.

Theorem 8.4.2. (preservation of parametric representations under reflection) *Suppose that* $x \in D$ *and* $(u,r) \in \Pi_s(x)$.

(a) If $x \in D_+$, *then* $(R(u), r) \in \Pi_s(R(x))$.

(b) If $x \in D_s$, *then* $(R(u), r) \in \Pi_w(R(x))$.

We also have an analog of Theorems 8.4.1 and 8.4.2 for the case $x \in D_s$ and $(u,r) \in \Pi_w(x)$.

Theorem 8.4.3. (preservation of weak parametric representations) *If* $x \in D_s$ *and* $(u,r) \in \Pi_w(x)$, *then* $(R(u), r) \in \Pi_w(R(x))$.

As a basis for proving Theorem 8.4.1, we exploit piecewise-constant approximations.

Lemma 8.4.1. (left and right limits) *For any* $x \in D_c$, $(u, r) \in \Pi_s(x)$ *and* $r^{-1}(t) = [s_-(t), s_+(t)]$,

$$R(u)(s_-(t)) = R(x)(t-) \quad \textit{and} \quad R(u)(s_+(t)) = R(x)(t) \; . \tag{4.1}$$

In order to prove Lemma 8.4.1, we establish several other lemmas. First, the following property of the reflection map applied to a single jump at time t is an easy consequence of the definition of the reflection map. We consider the reflection map applied to the jump in two parts. Given the linear relationship in (2.4), it suffices to focus on only one of ψ or ϕ.

Lemma 8.4.2. (the case of a single jump) *For any* $b_1, b_2 \in \mathbb{R}^k$, $0 < \beta < 1$ *and* $0 < t \le T$,

$$\phi(b_1 + b_2 I_{[t,T]})(u) = \phi(\phi(b_1 + \beta b_2 I_{[t,T]})(t) + (1-\beta) b_2 I_{[t,T]})(u) \quad \textit{for} \quad t \le u \le T \; .$$

Lemma 8.4.3. (generalization) *For any* $b_1, b_2 \in \mathbb{R}^k$ *and right-continuous nondecreasing nonnegative real-valued function* α *on* $[0, T]$ *with* $\alpha(0) = 0$,

$$\phi(b_1 + \alpha b_2)(t) = \phi(b_1 + \alpha(t) b_2 I_{[0,T]})(t), \quad 0 \le t \le T \; . \tag{4.2}$$

Proof. Represent α as the uniform limit of nondecreasing nonnegative functions α_n in D_c. Then $\|\phi(b_1 + \alpha_n b_2) - \phi(b_1 + \alpha b_2)\| \to 0$ as $n \to \infty$ by the known continuity of ϕ in the uniform metric. Hence it suffices to assume that $\alpha \in D_c$. We then establish (4.2) by recursively considering the successive discontinuity points of α, using Lemma 8.4.2 and Theorem 8.2.6(iv). ∎

Proof of Lemma 8.4.1. Any $x \in D_c$ can be represented as

$$x = \sum_{j=0}^{m} b_j I_{[t_j, T]}$$

for $0 = t_0 < t_1 < \cdots < t_m \le T$ and $b_j \in \mathbb{R}^k$ for $0 \le j \le m$. Thus t_j is the j^{th} discontinuity point of x. Let $[s_-(t_j), s_+(t_j)] = r^{-1}(t_j)$ for each j. Since $(u, r) \in \Pi_s(x)$ instead of just $\Pi_w(x)$, u can be expressed as

$$u = \sum_{j=0}^{m} \alpha_j b_j \; ,$$

where $\alpha_0(s) = 1$ for all s and, for $j \geq 1$, $\alpha_j : [0,1] \to [0,1]$ is continuous and nondecreasing with $\alpha_j(s) = 0$, $s \leq s_-(t_j)$ and $\alpha_j(s) = 1$, $s \geq s_+(t_j)$. We can now consider successive intervals $[s_-(t_j), s_+(t_j)]$ recursively exploiting Lemma 8.4.3. First, for any s with $0 \leq s \leq s_-(t_1)$.

$$\phi(u)(s) = \phi(b_0 I_{[0,1]})(s) = \phi(x)(0) = \phi_0(x(0)) \ .$$

Now assume that (4.1) holds for all $j \leq m-1$ and consider $s \in [s_-(t_m), s_+(t_m)]$. By the induction hypothesis, Lemma 8.4.3 and Theorem 8.2.6(iv),

$$\begin{aligned} \phi(u)(s) &= \phi(\phi(x)(t_{m-1}) + \alpha_m b_m I_{[s_-(t_m),1]})(s) \\ &= \phi(\phi(x)(t_{m-1}) + \alpha_m(s) b_m I_{[s_-(t_m),1]})(s) \ , \end{aligned}$$

so that (4.1) holds for t_m. ■

Proof of Theorem 8.4.1. (a) Since $t \in Disc(x)^c$, $u(s) = x(t)$ for $s_-(t) \leq s \leq s_+(t)$. Given $x \in D$ with $t \in Disc(x)^c$, it is possible to choose $x_n \in D_c$ such that $t \in Disc(x_n)^c$ for all n and $\|x_n - x\| \to 0$, by a slight strengthening of Theorem 6.2.2 in Section 6.2. By characterization (i) of M_1 convergence in Theorem 6.1 in Section V.6, given $(u, r) \in \Pi_s(x)$, we can find $(u_n, r_n) \in \Pi_s(x_n)$ such that

$$\|u_n - u\| \vee \|r_n - r\| \to 0 \quad \text{as} \quad n \to \infty \ .$$

Since R is continuous in the uniform topology, $\|R(u_n) - R(u)\| \to 0$ and $\|R(x_n) - R(x)\| \to 0$ as $n \to \infty$. Let s_n be such that $r_n(s_n) = t$. Since $x_n \in D_c$ and $t \in Disc(x_n)^c$, $R(u_n)(s_n) = R(x_n)(t)$ by Lemma 8.4.1. Since $0 \leq s_n \leq 1$, $\{s_n\}$ has a convergent subsequence $\{s_{n_k}\}$. Let s' be the limit of that convergent subsequence. Since $r_{n_k}(s_{n_k}) = t$ for all n_k, we necessarily have $s' \in [s_-(t), s_+(t)]$. Since $\|R(u_n) - R(u)\| \to 0$, $R(x_{n_k})(t) = R(u_{n_k})(s_{n_k}) \to R(u)(s')$. Since we have already seen that $R(x_n)(t) \to R(x)(t)$, we must have $R(u)(s') = R(x)(t)$. Since $R(u)$ is constant on $[s_-(t), s_+(t)]$, we must have $R(u)(s) = R(x)(t)$ for all s with $s_-(t) \leq s \leq s_+(t)$.

(b) Since R maps D into D and C into C, $R(x)$ is right-continuous with left limits, while $R(u)$ is continuous. Given $t \in Disc(x)$, we can find $t_n \in Disc(x)^c$ with $t_n \uparrow t$. We can apply part (a) to obtain $R(u)(s_+(t_n)) = R(x)(t_n) \to R(x)(t-)$, but $s_+(t_n) \uparrow s_-(t)$, so that $R(u)(s_+(t_n)) \to R(u)(s_-(t))$. Hence, we have established the first claim: $R(u)(s_-(t)) = R(x)(t-)$. Similarly, we can find $t_n \in Disc(x)^c$ with $t_n \downarrow t$. Then we can apply part (a) again to obtain $R(u)(s_-(t_n)) = R(x)(t_n) \to R(x)(t)$. Since $s_-(t_n) \downarrow s_+(t)$, $R(u)(s_-(t_n)) \downarrow R(u)s_+(t))$. Hence $R(x)(t) = R(u)(s_+(t))$ as claimed.

(c) We can apply Lemma 14.3.4 (a) in the book. Since the increment $x(t) - x(t-)$ is nonnegative in each component,

$$z(t) = z(t-) + x(t) - x(t-)$$

and $y(t) = y(t-)$. Similarly,

$$\phi(u)(s) = \phi(u)(s_-(t)) + u(s) - u(s_-(t))$$

and $\psi(u)(s) = \psi(u)(s_-(t))$ for $s_-(t) \le s \le s_+(t)$.

(d) We apply Lemma 14.3.4 (b) in the book. Each coordinate $\phi^i(u)$ and $\psi^i(u)$ is monotone in s over $[s_-(t), s_+(t)]$, so that the desired conclusion holds.

Proof of Theorem 8.4.2. (a) We combine parts (a)–(c) of Theorem 8.4.1 to get $(R(u), r)(s) \in \Gamma_{R(x)}$ for all s. Since R maps C into C, $(R(u), r)$ is continuous. Also r is nondecreasing with $r(0) = 0$ and $r(1) = T$ because $(u, r) \in \Pi_s(x)$. Finally, $(R(u), r)$ maps $[0, 1]$ onto $\Gamma_{R(x)}$ and $(R(u), v)$ is nondecreasing with respect to the order on $\Gamma_{R(x)}$ because the increments of $R(u)$ coincide with the increments of u over each discontinuity in x because $x \in D_+$, and (u, r) has these properties.

(b) We incorporate part (d) of Theorem 8.4.1 to get $R(u)$ monotone over $[s_-(t), s_+(t)] = r^{-1}(t)$ for each $t \in Disc(x) = Disc(R(x))$. This allows us to conclude that $(R(u), r) \in \Pi_w(R(x))$. ∎

We now turn to the proof of Theorem 8.4.3. For the proof, we find it convenient to use a different class of approximating functions. Let D_l be the subset of all functions in D that (i) have only finitely many jumps and (ii) are continuous and piecewise linear in between jumps with only finitely many changes of slope. Let $D_{s,l} = D_s \cap D_l$.

Analogous to Theorem 6.2.2 in Section 6.2, we have the following result.

Lemma 8.4.4. (approximation of elements of D_s by elements of $D_{s,l}$) *For any $x \in D_s$, there exist $x_n \in D_{s,l}$ such that $\|x_n - x\| \to 0$ as $n \to \infty$.*

Proof. For $x \in D_s$ and $\epsilon > 0$ given, apply Theorem 6.2.2 in Section 6.2 to find $x_1 \in D_c$ (with only finitely many discontinuities) such that $\|x - x_1\| < \epsilon/4$. The function x_1 can have jumps of opposite sign, but the magnitude of the jumps in one of the two directions must be at most $\epsilon/2$. Form the desired function, say x_2, from x_1. Suppose that $\{t_1, \ldots, t_k\} = Disc(x_1)$. Suppose that x_1 has one or more negative jump at t_j, none of which has magnitude exceeding $\epsilon/2$. If x_1 has a negative jump at t_j in coordinate i for some i, then replace x_1^i over $[t_{j-1}, t_j)$ by the linear function connecting $x_1^i(t_{j-1})$ and $x_1^i(t_j)$. Similarly, if x_1 has one or more positive jumps at some t_j with all magnitudes less than $\epsilon/2$, then proceed as above. It is easy to see that $Disc(x_2) \subseteq Disc(x_1)$, $x_2 \in D_{s,l}$ and $\|x - x_2\| < \epsilon$. ■

We now show that limits of parametric representations are parametric representations when $\|x_n - x\| \to 0$.

Lemma 8.4.5. (limits of parametric representations) *If (i)* $\|x_n - x\| \to 0$ *as* $n \to \infty$, *(ii)* $(u_n, r_n) \in \Pi_z(x_n)$ *for each* n, *where* $z = s$ *or* w, *and (iii)* $\|u_n - u\| \vee \|r_n - r\| \to 0$ *as* $n \to \infty$ *where* u *and* r *are functions mapping* $[0,1]$ *into* $\mathbb{R}^k$ *and* $\mathbb{R}^1$, *respectively, then* $(u, r) \in \Pi_z(x)$ *for the same* z.

Proof. Since (u, r) is the uniform limit of the continuous functions (u_n, r_n), (u, r) is itself continuous. Since r is the limit of the nondecreasing functions r_n, r is itself nondecreasing. Since $r_n(0) = 0$ and $r_n(1) = T$ for all n, $r(0) = 0$ and $r(1) = T$. Since r is also nondecreasing and continuous, r maps $[0,1]$ onto $[0,T]$. Pick any s with $0 < s < 1$. Then $r(s) = t$ for some t, $0 \le t \le T$, and $r_n(s) = t_n \to t$ as $n \to \infty$. Suppose that $(u_n, r_n) \in \Pi_s(x_n)$ for all n. That means that

$$u_n(s) = \alpha_n(s) x_n(t_n) + (1 - \alpha_n(s)) x_n(t_n-)$$

for all n. Since $0 \le \alpha_n(s) \le 1$, there exists a convergent subsequence $\{\alpha_{n_k}(s)\}$ such that $\alpha_{n_k}(s) \to \alpha(s)$ as $n_k \to \infty$. At least one of the following three cases must prevail: (i) $t_{n_k} > t$ for infinitely many n_k, (ii) $t_{n_k} = t$ for infinitely many n_k and (iii) $t_{n_k} < t$ for infinitely many n_k. In case (i), we can choose a further subsequence $\{n_{k_j}\}$ so that $u_{n_{k_j}}(s) \to x(t)$; in case (ii), we can choose a further subsequence so that $u_{n_{k_j}}(s) \to \alpha(s)x(t) + [1 - \alpha(s)]x(t-)$; in case (iii) we can choose a further subsequence so that $u_{n_{k_j}}(s) \to x(t-)$. Since $u_n(s) \to u(s)$, the limit of the subsequence must be $u(s)$. Hence, $(u(s), r(s)) \in \Gamma_x$ for each s. Since (u, r) is continuous with $r(0) = 0$ and $r(1) = T$, (u, r) maps $[0,1]$ onto Γ_x. Since (u_n, r_n) is monotone as a function from $[0,1]$ to $(\Gamma_{x_n}, \le)$ and $\|u_n - u\| \vee \|r_n - r\| \to 0$, (u, r) is monotone from $[0,1]$ to $(\Gamma_x, \le)$. Hence, $(u, r) \in \Pi_s(x)$. Finally, suppose that $(u_n, r_n) \in \Pi_w(x_n)$ for all n. By the result above applied to the individual coordinates, $(u^i(s), r(s)) \in \Gamma_{x^i}$ and thus $(u^i, r) \in \Pi_s(x^i)$ for each i, which implies that $(u, r) \in \Pi_w(x)$. ■

Proof of Theorem 8.4.3. For $x \in D_s$, apply Lemma 8.4.4 to find $x_n \in D_{s,l}$ such that $\|x_n - x\| \to 0$. Suppose that $(u, r) \in \Pi_w(x)$. Then it is possible to find u_n such that $(u_n, r) \in \Pi_w(x_n)$ and $\|u_n - u\| \to 0$. To do so, let $u_n(s_-(t)) = x_n(t-)$ and $u_n(s_+(t)) = x_n(t)$, where $[s_-(t), s_+(t)] = r^{-1}(t)$ for each $t \in Disc(x)$. If $t \in Disc(x_n)^c$, let $u_n(s) = u_n(s_+(t))$ for $s_-(t) \leq s \leq s_+(t)$; if $t \in Disc(x_n)$, define u_n so that $\|u_n - u\| \to 0$. Given that $(u_n, r) \in \Pi_w(x_n)$, we can apply mathematical induction over the finitely many time points such that x_n has a jump or a change of slope to show that $(R(u_n), r) \in \Pi_w(R(x_n))$ for each n. We use Lemma 14.3.4 of the book critically at this point to treat the discontinuity points of x_n in $D_{s,l}$. The continuous linear pieces between discontinuities can be treated by applying the rescaling property in Theorem 8.2.6 (iii) with $\beta = 1$ and $\eta = 0$. Finally, we apply Lemma 8.4.5 to deduce that $(R(u), r) \in \Pi_w(R(x))$. For that, we use the fact that $\|R(x_n) - R(x)\| \to 0$ and $\|R(u_n) - R(u)\| \to 0$.

8.5. $\mathbf{M_1}$ Continuity Results

In this section we establish continuity and Lipschitz properties of the reflection map on $D \equiv D^k \equiv D([0, T], \mathbb{R}^k)$ with the M_1 topologies. Our first result establishes continuity of the reflection map R (for an arbitrary reflection matrix Q) as a map from (D, SM_1) to (D, L_1), where L_1 is the topology on D induced by the L_1 norm

$$\|x\|_{L_1} \equiv \int_0^T \|x(t)\| dt \ . \tag{5.1}$$

Under a further restriction, the map from (D, WM_1) to (D, WM_1) will be continuous.

Recall that D_s is the subset of functions in D without simultaneous jumps of opposite sign in the coordinate functions; i.e., $x \in D_s$ if, for all $t \in (0, T)$, either $x(t) - x(t-) \leq 0$ or $x(t) - x(t-) \geq 0$, with the sign allowed to depend upon t. The subset D_s is a closed subset of D in the J_1 topology and thus a measurable subset of D with the SM_1 and WM_1 topologies (since the Borel σ-fields coincide). The proofs of the main theorems here appear in Section 6.2 of the Internet Supplement.

Theorem 8.5.1. (continuity with the SM_1 topology on the domain) *Suppose that $x_n \to x$ in (D, SM_1).*

(a) Then

$$R(x_n)(t_n) \to R(x)(t) \quad in \quad \mathbb{R}^{2k} \tag{5.2}$$

for each $t \in Disc(x)^c$ and sequence $\{t_n : n \geq 1\}$ with $t_n \to t$,

$$\sup_{n \geq 1} \|R(x_n)\| < \infty \ , \tag{5.3}$$

$$R(x_n) \to R(x) \quad in \quad (D, L_1) \tag{5.4}$$

and

$$\psi(x_n) \to \psi(x) \quad in \quad (D, WM_1) \ . \tag{5.5}$$

(b) If in addition $x \in D_s$, then

$$\phi(x_n) \to \phi(x) \quad in \quad (D, WM_1) \ , \tag{5.6}$$

so that

$$R(x_n) \to R(x) \quad in \quad (D, WM_1) \ . \tag{5.7}$$

Proof. (a) We first prove (5.2). Since $x_n \to x$ in (D, SM_1), we can find parametric representations $(u, r) \in \Pi_s(x)$ and $(u_n, r_n) \in \Pi_s(x_n)$ for $n \geq 1$ such that

$$\|u_n - u\| \vee \|r_n - r\| \to 0 \ .$$

By Theorem 14.4.1 (a) in the book, $R(u)(s) = R(x)(t)$ for any $s \in [s_-(t), s_+(t)] \equiv r^{-1}(t)$, since $t \in Disc(x)^c$. Moreover, by Corollary 14.3.4 in the book, $t \in Disc(R(x))^c$. For any sequence $\{t_n : n \geq 1\}$ with $t_n \to t$, we can find another sequence $\{t'_n : n \geq 1\}$ such that $t'_n \to t$, $t'_n \in Disc(x_n)^c$ and $\|R(x_n)(t'_n) - R(x_n)(t_n)\| \to 0$ as $n \to \infty$. (Here we exploit the fact that $R(x_n) \in D$ for each n.) Consequently, $R(x_n)(t_n) \to R(x)(t)$ if and only if $R(x_n)(t'_n) \to R(x)(t)$. By Theorem 13.4.1 (a) again, $R(u_n)(s_n) = R(x)(t'_n)$ for any $s_n \in [s_-(t'_n), s_+(t'_n)] = r_n^{-1}(t'_n)$. Since $0 \leq s_n \leq 1$ for all n, any such sequence $\{s_n : n \geq 1\}$ has a convergent subsequence $\{s_{n_k} : k \geq 1\}$. Suppose that $s_{n_k} \to s'$ as $n_k \to \infty$. Since $t'_n \to t$ as $n \to \infty$ and $t'_{n_k} = r_{n_k}(s_{n_k}) \to r(s')$ as $n_k \to \infty$, we must have $s' \in [s_-(t), s_+(t)]$. Then, since $\|R(u_n) - R(u)\| \to 0$,

$$R(x_{n_k})(t'_{n_k}) = R(u_{n_k})(s_{n_k}) \to R(u)(s') = R(x)(t) \ .$$

Since every subsequence of $\{R(x_n)(t'_n) : n \geq 1\}$ must have a convergent subsequence with the same limit, we must have $R(x_n)(t'_n) \to R(x)(t)$ as $n \to \infty$, which we have shown implies that $R(x_n)(t_n) \to R(x)(t)$ as $n \to \infty$, as claimed in (5.2). Next we establish (5.3). For any $x \in D$, $\|x\| \equiv \sup_{0 \leq t \leq T} \|x(t)\| < \infty$. Since $d_s(x_n, x) \to 0$, $\|x_n\| \to \|x\|$ as $n \to \infty$. Hence, it suffices to show that there is a constant K such that

$$\|R(x)\| \leq K\|x\| \quad \text{for all} \quad x \in D \ ,$$

but that follows from Theorem 13.2.5. We apply the bounded convergence theorem with (5.2) and (5.3) to establish (5.4). We now turn to (5.5). Since $\psi(x_n)$ and $\psi(x)$ are nondecreasing in each coordinate the pointwise convergence established in (5.2) actually implies WM_1 convergence in (5.5); see Corollary 12.5.1 in the book.

(b) First, we use the assumed convergence $x_n \to x$ in (D, SM_1) to pick $(u,r) \in \Pi_s(x)$ and $(u_n, r_n) \in \Pi_s(x_n)$, $n \geq 1$, with

$$\|u_n - u\| \vee \|r_n - r\| \to 0 \ .$$

Since R is continuous on (D, U), we also have $\|R(u_n) - R(u)\| \to 0$. By part (a), we know that there is local uniform convergence of $R(x_n)$ to $R(x)$ at each continuity point of $R(x)$. Thus, by Theorem 12.5.1 (v) in the book, to establish $R(x_n) \to R(x)$ in (D, WM_1), it suffices to show that

$$\lim_{\delta \downarrow 0} \overline{\lim_{n \to \infty}} \ w_s(R^i(x_n), t, \delta) = 0 \tag{5.8}$$

for each i, $1 \leq i \leq 2k$, and $t \in Disc(R(x))$, where

$$w_s(x, t, \delta) \equiv \sup\{\|x(t_2) - [x(t_1), x(t_3)]\| : (t_1, t_2, t_3) \in A(t, \delta)\} \tag{5.9}$$

for

$$A(t, \delta) \equiv \{(t_1, t_2, t_3) : (t - \delta) \vee 0 \leq t_1 < t_2 < t_3 \leq (t + \delta) \wedge T\} \ .$$

(Since we are considering the i^{th} coordinate function $R^i(x_n)$, the function x in (5.9) is real-valued here.) Suppose that (5.8) fails for some i and t. Then there exist $\epsilon > 0$ and subsequences $\{\delta_k\}$ and $\{n_k\}$ such that $\delta_k \downarrow 0$, $n_k \to \infty$ and

$$w_s(R^i(x_{n_k}), t, \delta_k) > \epsilon \quad \text{for all} \quad \delta_k \quad \text{and} \quad n_k \ .$$

That is, there exist time points t_{1,n_k}, t_{2,n_k} and t_{3,n_k} with

$$(t - \delta_k) \vee 0 \leq t_{1,n_k} < t_{2,n_k} < t_{3,n_k} \leq (t + \delta_k) \wedge T \tag{5.10}$$

and

$$\|R^i(x_{n_k})(t_{2,n_k}) - [R^i(x_{n_k}(t_{1,n_k}), R^i(x_{n_k}(t_{3,n_k})]\| > \epsilon \ . \tag{5.11}$$

Since the values $R^i(x_{n_k})(t)$ are contained in the values $R^i(u_{n_k})(s)$ where $(u_{n_k}, r_{n_k}) \in \Pi_s(x_{n_k})$, we can deduce that there are points s_{j,n_k} for $j = 1,2,3$ such that $0 \leq s_{1,n_k} < s_{2,n_k} < s_{3,n_k} \leq 1$, $r_{n_k}(s_{j,n_k}) = t_{j,n_k}$ for $j = 1,2,3$ and all n_k, and

$$\|R^i(u_{n_k})(s_{2,n_k}) - [R^i(u_{n_k})(s_{1,n_k}), R^i(u_{n_k})(s_{3,n_k})]\| > \epsilon \ . \tag{5.12}$$

By (5.10) and (5.12), there then exists a further subsequence $\{n'_k\}$ such that $t_{j,n'_k} \to t$ and $s_{j,n'_k} \to s_j$ as $n'_k \to \infty$ for $j = 1, 2, 3$, where $0 \le s_1 \le s_2 \le s_3 \le 1$, $r_{n'_k}(s_{j,n'_k}) \to r(s_j) = t$ and

$$\|R^i(u)(s_2) - [R^i(u)(s_1), R^i(u)(s_3)]\| \ge \epsilon > 0 \ . \tag{5.13}$$

However, by Theorem 14.4.2 in the book, $(R(u), r) \in \Pi_w(R(x))$ since $x \in D_s$, so that $(R^i(u), r) \in \Pi_s(R^i(x))$. Hence $(R^i(u), r) \in \Pi_s(R^i(x))$. Since $R^i(u)$ is monotone on $[s_-(t), s_+(t)]$, (5.13) cannot occur. Hence (5.8) must in fact hold and $R^i(x_n) \to R^i(x)$ in (D, M_1). Since that is true for all i, we must have $R(x_n) \to R(x)$ in (D, WM_1). ∎

Under the extra condition in part (b), the mode of convergence on the domain actually can be weakened. However, little positive can be said if only $x_n \to x$ in (D, WM_1) without $x \in D_s$; see Example 14.5.3 in the book.

Theorem 8.5.2. (continuity with the WM_1 topology on the domain) *If $x_n \to x$ in (D, WM_1) and $x \in D_s$, then* (5.7) *holds.*

The proof of Theorem 8.5.2 is more difficult. We now work towards its proof. By Theorem 8.4.3, R is Lipschitz on (D_s, WM_1), but x_n need not be in D_s. We show that we can approximate x_n by elements of D_s.

We first restate Corollary 12.11.2 in the book as a lemma. It states that Convergence in WM_2, which of course is implied by convergence in WM_1, has the advantage that jumps in the converging functions must be inherited by the limit function.

Lemma 8.5.1. (inheritance of jumps) *If $x_n \to x$ in (D, WM_2), $t_n \to t$ in $[0, T]$ and $x_n^i(t_n) - x_n^i(t_n-) \ge c > 0$ for all n, then $x^i(t) - x^i(t-) \ge c$.*

For $x \in D$ and $t \in Disc(x)$, let $\gamma(x, t)$ be the largest magnitude (absolute value) of the jumps in x at time t of opposite sign to the sign of the largest jump in x at time t. Let $\gamma(x)$ be the maximum of $\gamma(x, t)$ over all $t \in Disc(x)$. We apply Lemma 8.5.1 to establish the next result.

Lemma 8.5.2. *If $x_n \to x$ in (D, WM_1), then*

$$\overline{\lim_{n\to\infty}}\ \gamma(x_n) \le \gamma(x) \ .$$

We only use the following consequence of Lemma 8.5.2.

Lemma 8.5.3. *If $x_n \to x$ in (D, WM_1) and $x \in D_s$, then $\gamma(x_n) \to 0$.*

We also use a generalization of Lemma 8.4.4 above, which is established in the same way.

Lemma 8.5.4. *For any $x \in D$, there exist $x_n \in D_{s,l}$ such that $\|x_n - x\| \to \gamma(x)$ as $n \to \infty$.*

We combine Lemmas 8.5.2 and 8.5.4 to obtain the tool we need.

Lemma 8.5.5. *If $x_n \to x$ in (D, WM_1) and $x \in D_s$, then there exists $x'_n \in D_{s,l}$ for $n \geq 1$ such that $\|x'_n - x_n\| \to 0$.*

Proof of Theorem 8.5.2. Given $x_n \to x$ in (D, WM_1), apply Lemma 8.5.5 to find $x'_n \in D_{s,l}$ for $n \geq 1$ such that $\|x'_n - x_n\| \to 0$ as $n \to \infty$. Then, by the triangle inequality, Theorem 14.2.5 in the book and Lemma 8.5.3 above,

$$\begin{aligned} d_p(R(x_n), R(x)) &\leq d_p(R(x_n), R(x'_n)) + d_p(R(x'_n), R(x)) \\ &\leq \|R(x_n) - R(x'_n)\| + d_w(R(x'_n), R(x)) \\ &\leq K\|x_n - x'_n\| + K d_w(x'_n, x). \end{aligned}$$

Since

$$\begin{aligned} d_p(x'_n, x) &\leq d_p(x'_n, x_n) + d_p(x_n, x) \\ &\leq \|x'_n - x_n\| + d_p(x_n, x) \\ &\to 0 \;, \end{aligned}$$

$d_w(x'_n, x) \to 0$. Hence, $d_p(R(x_n), R(x)) \to 0$ as claimed. ∎

Example 12.3.1 in the book shows that convergence $x_n \to x$ can hold in (D, WM_1) but not in (D, SM_1) even when $x \in D_s$. Thus Theorems 8.5.1 (a) and 8.5.2 cover distinct cases. An important special case of both occurs when $x \in D_1$, where D_1 is the subset of x in D with discontinuities in only one coordinate at a time; i.e., $x \in D_1$ if $t \in Disc(x^i)$ for at most one i when $t \in Disc(x)$, with the coordinate i allowed to depend upon t. In Section 6.7 it is shown that WM_1 convergence $x_n \to x$ is equivalent to SM_1 convergence when $x \in D_1$.

Just as with D_s above, D_1 is a closed subset of (D, J_1) and thus a Borel measurable subset of (D, SM_1). Since $D_1 \subseteq D_s$, the following corollary to Theorem 8.5.2 is immediate.

Corollary 8.5.1. (common case for applications) *If $x_n \to x$ in (D, WM_1) and $x \in D_1$, then $R(x_n) \to R(x)$ in (D, WM_1).*

We can obtain stronger Lipschitz properties on special subsets. Let D_+ be the subset of x in D with only nonnegative jumps, i.e., for which $x^i(t) - x^i(t-) \geq 0$ for all i and t. As with D_s and D_1 above, D_+ is a closed subset of (D, J_1) and thus a measurable subset of (D, SM_1).

Theorem 8.5.3. (*Lipschitz properties*) *There is a constant K (the same as associated with the uniform norm from Theorem 8.2.5) such that*

$$d_s(R(x_1), R(x_2)) \leq K d_s(x_1, x_2) \tag{5.14}$$

for all $x_1, x_2 \in D_+$, and

$$d_p(R(x_1), R(x_2)) \leq d_w(R(x_1), R(x_2)) \leq K d_w(x_1, x_2) \leq K d_s(x_1, x_2) \tag{5.15}$$

for all $x_1, x_2 \in D_s$.

Proof. Given that $x \in D_+$, apply Theorem 14.4.2 (a) in the book to get $(R(u), r) \in \Pi_s(R(x))$ when $(u, r) \in \Pi_s(x)$. Then

$$\begin{aligned} d_s(R(x_1), R(x_2)) &\equiv \inf_{\substack{(u_i', r_i) \in \Pi_s(R(x_i)) \\ i=1,2}} \{\|u_1' - u_2'\| \vee \|r_1 - r_2\|\} \\ &\leq \inf_{\substack{(u_i, r_i) \in \Pi_s(x_i) \\ i=1,2}} \{\|\phi(u_1) - \phi(u_2)\| \vee \|r_1 - r_2\|\} \\ &\leq \inf_{\substack{(u_i, r_i) \in \Pi_s(x_i) \\ i=1,2}} \{K\|u_1 - u_2\| \vee \|r_1 - r_2\|\} \\ &\leq K d_s(x_1, x_2) \end{aligned}$$

because $K \geq 1$. The other results are obtained in essentially the same way. Apply Theorem 14.4.3 in the book to get $(R(u), r) \in \Pi_w(R(x))$ when $(u, r) \in \Pi_w(x)$ and $x \in D_+$. When $x \in D_s$, apply Theorem 13.4.2 (b) to get $(R(u), r) \in \Pi_w(R(x))$ when $(u, r) \in \Pi_s(x)$. ∎

We can actually do somewhat better than in Theorem 8.5.1 when the limit is in D_+.

Theorem 8.5.4. (strong continuity when the limits is in D_+) *If*

$$x_n \to x \quad in \quad (D, SM_1) , \tag{5.16}$$

where $x \in D_+$, then

$$R(x_n) \to R(x) \quad in \quad (D, SM_1) . \tag{5.17}$$

Proof. Suppose that $x_n \to x$ in (D, SM_1). By Theorem 8.5.1(a), we have $\psi(x_n) \to \psi(x)$ in (D, WM_1). Since $x \in D_+$, $\psi(x) \in C$, by Corollary 14.3.5 in the book. Hence the WM_1 convergence is equivalent to uniform convergence; i.e.,

$$\psi(x_n) \to \psi(x) \quad \text{in} \quad D([0,T], \mathbb{R}^k, U) \ .$$

We can then apply addition with equation (14.2.6) in the book to get

$$R(x_n) \to R(x) \quad \text{in} \quad D([0,T], \mathbb{R}^{2k}, SM_1) \ . \quad \blacksquare$$

Our final result shows how the reflection map behaves as a function of the reflection matrix Q, as well as x, with the M_1 topologies.

Theorem 8.5.5. (continuity as a function of (x, Q)) *Suppose that* $Q_n \to Q$ *in* $\mathcal{Q}$.

(a) If $x_n \to x$ *in* (D^k, WM_1) *and* $x \in D_s$*, then*

$$R_{Q_n}(x_n) \to R_Q(x) \quad \textit{in} \quad (D^{2k}, WM_1) \ . \tag{5.18}$$

(b) If $x_n \to x$ *in* (D^k, SM_1) *and* $x \in D_+$*, then*

$$R_{Q_n}(x_n) \to R_Q(x) \quad \textit{in} \quad (D^{2k}, SM_1) \ . \tag{5.19}$$

Proof. We only prove the first of the two results, since the two proofs are essentially the same. If $x_n \to x$ in (D, WM_1) with $x \in D_s$, then we can find $x'_n \in D_{s,l}$ for $n \ge 1$ such that $\|x_n - x'_n\| \to 0$ by Lemma 8.5.5. By Theorem 14.2.5 in the book,

$$\|R_{Q_n}(x_n) - R_{Q_n}(x'_n)\| \le K_n \|x_n - x'_n\| \to 0 \tag{5.20}$$

because $K_n \to K < \infty$. By Theorem 14.4.3 in the book, $(R_Q(u), r) \in \Pi_w(R(x))$ when $x \in D_s$. So, for any $\epsilon > 0$ given, let $(u, r) \in \Pi_w(x)$ and $(u_n, r_n) \in \Pi_w(x'_n)$ such that $\|u_n - u\| \vee \|r_n - r\| \le \epsilon$. Then $(R_Q(u), r) \in \Pi_w(R_Q(x))$, $(R_{Q_n}(u_n), r_n) \in \Pi_w(R_{Q_n}(x'_n))$ for $n \ge 1$ and

$$\|R_{Q_n}(u_n) - R_Q(u)\| < K(\epsilon + \|Q_n - Q\|) \tag{5.21}$$

by Theorem 14.2.9 and equation (14.2.35) in the book, so that

$$R_{Q_n}(x'_n) \to R_Q(x) \quad \text{in} \quad (D^{2k}, WM_1) \ . \tag{5.22}$$

Combining (5.20), (5.22) and the triangle inequality with the metric d_p, we obtain (5.18). $\blacksquare$

We can apply Section 6.9 to extend the continuity and Lipschitz results to the space $D([0,\infty), \mathbb{R}^k)$.

Theorem 8.5.6. (extension of continuity results to $D([0,\infty),\mathbb{R}^k)$) *The convergence-preservation results in Theorems 8.5.1, 8.5.2 and 8.5.4 and Corollary 8.5.1 extend to* $D([0,\infty),\mathbb{R}^k)$.

Proof. Suppose that $x_n \to x$ in $D([0,\infty),\mathbb{R}^k)$ with the appropriate topology and that $\{t_j : j \geq 1\}$ is a sequence of positive numbers with $t_j \in Disc(x)^c$ and $t_j \to \infty$ as $j \to \infty$. Then, $r_{t_j}(x_n) \to r_{t_j}(x)$ in $D([0,\infty),\mathbb{R}^k)$ with the same topology as $n \to \infty$ for each j, where r_t is the restriction map to $D([0,t],\mathbb{R}^k)$. Under the specified assumptions,

$$r_{t_j}(R(x_n)) = R_{t_j}(r_{t_j}(x_n)) \to R_{t_j}(r_{t_j}(x)) = r_{t_j}(R(x)) \tag{5.23}$$

in $D([0,t_j],\mathbb{R}^{2k})$ with the specified topology as $n \to \infty$ for each j, which implies that

$$R(x_n) \to R(x) \quad \text{in} \quad D([0,\infty),\mathbb{R}^{2k}) \tag{5.24}$$

with the same topology as in (5.23). ∎

Theorem 8.5.7. (extension of Lipschitz properties to $D([0,\infty),\mathbb{R}^k)$) *Let* $R : D([0,\infty),\mathbb{R}^k) \to D([0,\infty),\mathbb{R}^{2k})$ *be the reflection map with function domain* $[0,\infty)$ *defined by Definition* 8.2.1. *Let metrics associated with domain* $[0,\infty)$ *be defined in terms of restrictions by equation* (9.1) *in Section* 12.9 *of the book. Then the conclusions of Theorems* 8.2.5, 8.2.7 *and* 8.5.3 *also hold for domain* $[0,\infty)$.

Proof. Apply Theorem 12.9.4 in the book. ∎

8.6. Limits for Stochastic Fluid Networks

Nothing has been omitted from Section 14.6 of the book.

8.7. Queueing Networks with Service Interruptions

Nothing has been omitted from Section 14.7 of the book.

8.8. The Two-Sided Regulator

Nothing has been omitted from Section 14.8 of the book.

8.9. Existence of a Limiting Stationary Version

In this section, drawing on and extending Kella and Whitt (1996), we show that there exists a proper limiting stationary version of a reflected stochastic process under natural conditions. We establish existence and uniqueness of the limiting stochastic process, but we do not otherwise characterize the limiting marginal distribution on $\mathbb{R}^k$ or determine how to calculate it.

Our existence and uniqueness results with general initial conditions cover the case of the reflected Lévy process obtained as the heavy-traffic limit of the vector-valued buffer-content stochastic processes in a stochastic fluid network, as in Section 14.6 of the book, when the exogenous input processes at the different nodes are independent Lévy processes (i.e., processes with stationary independent increments) under a natural condition on the net input rates. We also obtain useful results about more general reflected processes without the independence conditions.

8.9.1. The Main Results

We are given a net-input stochastic process $\{X(t) : t \geq 0\}$ and the associated reflected content stochastic process

$$Z(t) \equiv \phi(X)(t) \equiv X(t) + (I - Q)Y(t), \quad t \geq 0 , \tag{9.1}$$

where $Y \equiv \psi(X)$ is the minimal nondecreasing nonnegative stochastic process such that $Z \geq 0$, as in Definition 8.2.1. We want to consider the limiting behavior as $t \to \infty$. We want to determine conditions under which

$$(Z_s(t_1), \ldots, Z_s(t_m)) \Rightarrow (Z_*(t_1), \ldots, Z_*(t_m)) \quad \text{in} \quad \mathbb{R}^{km} \quad \text{as} \quad s \to \infty \tag{9.2}$$

for all positive integers m and any m time points t_i with $0 \leq t_1 < \cdots < t_m$, where

$$Z_s(t) \equiv Z(s + t), \quad t \geq 0, \quad s \geq 0 , \tag{9.3}$$

and the limiting stochastic process $Z_* \equiv \{Z_*(t) : t \geq 0\}$ is a *stationary stochastic process*, i.e., where

$$(Z_*(t_1 + h), \ldots, Z_*(t_m + h)) \stackrel{\mathrm{d}}{=} (Z_*(t_1), \ldots, Z_*(t_m))$$

for all positive integers m, any m time points t_i with $0 \leq t_1 < \cdots < t_m$ and all $h > 0$. We also want the limit process to be proper, i.e., we want to have

$$P(Z_*(t) < \infty) = 1 \quad \text{for all} \quad t .$$

We then call the stochastic process Z_* the *limiting stationary version* of Z.

We first observe that convergence of the finite-dimensional distributions in (9.2) for processes Z_s defined as in (9.3) directly implies that the limit process Z_* is stationary.

Lemma 8.9.1. (stationarity from convergence) *If*

$$(Z(s+t_1),\ldots,Z(s+t_m)) \Rightarrow (Z_*(t_1),\ldots,Z_*(t_m)) \quad in \quad \mathbb{R}^{km} \tag{9.4}$$

as $s \to \infty$ for all positive integers m and all m time points t_i with $0 \le t_1 < \cdots < t_m$, then Z_* is a stationary process.

Proof. If (9.4) holds, then

$$(Z(s+t_1+h),\ldots,Z(s+t_m+h)) \Rightarrow (Z_*(t_1+u),\ldots,Z_*(t_m+u)) \tag{9.5}$$

as $s \to \infty$ for any u, $0 \le u \le h$, because we can let $s' = s+h-u$, $t_i' = t_i + u$, $1 \le i \le m$, and let $s' \to \infty$ with (9.4). Hence the distribution of the random vector on the right in (9.5) must be independent of u. ■

In order to obtain a unique limiting stationary version, we will assume that the net-input process X has *stationary increments*, i.e., the joint distribution of the random vector

$$(X(t_1+s) - X(u_1+s),\ldots,X(t_m+s) - X(u_m+s))$$

in $\mathbb{R}^{km}$ is independent of s for all positive integers m and all m-tuples of real numbers $(t_1,\ldots,t_m)$ and $(u_1,\ldots,u_m)$. We assume that X is defined on the whole real line $(-\infty,\infty)$. As a consequence,

$$X_s \equiv \{X(t+s) - X(s) : t \ge 0\} \tag{9.6}$$

has a distribution as a random element of D^k independent of s. We will also assume that X has *ergodic increment*, i.e., the increment $X(t+s) - X(s)$ have finite mean and

$$t^{-1}X(t) \to E[X(1) - X(0)] \quad \text{w.p.1} \quad \text{as} \quad t \to \infty\,.$$

Here is our main result: In addition to the assumptions above, it depends on the special initial condition $X(0) = 0$, which forces $Z(0) = Y(0) = 0$. The proof of the following result and several others are given at the end of the section.

Theorem 8.9.1. (existence of a limiting stationary version) *If X has stationary ergodic increments with $X(0) = 0$ and*

$$((I - Q)^{-1} E[X(1) - X(0)])^i < 0, \quad 1 \le i \le k \ , \tag{9.7}$$

then (9.2) *holds, i.e., the finite-dimensional distributions of Z_s in* (9.3) *converge as $s \to \infty$ to the finite-dimensional distributions of a proper stationary stochastic process Z_*.*

We now show the necessity of condition (9.7), leaving untouched the boundary case of equality. In particular, we show that a proper limit cannot exist if the strict inequality in (9.7) is reversed in any coordinate i. Indeed, then the i^{th} coordinate of the reflected process grows without bound.

Theorem 8.9.2. (necessity of the drift condition) *Suppose that*

$$t^{-1} X(t) \to x \quad in \quad \mathbb{R}^k \quad w.p.1 \quad as \quad t \to \infty \, . \tag{9.8}$$

If

$$(I - Q)^{-1} x \le 0 \ , \tag{9.9}$$

then

$$t^{-1} Z(t) \to 0 \quad as \quad t \to \infty \quad w.p.1 \tag{9.10}$$

for Z in (9.1). *On the other hand, if $((I - Q^{-1})x)^i > 0$ for some i, then*

$$\liminf_{t \to \infty} t^{-1} Z^i(t) > 0 \quad for\ that \quad i \ . \tag{9.11}$$

Proof. By Corollary 3.2.1 in the Internet Supplement, the SLLN in condition (9.8) implies the stronger FSLLN

$$\mathbf{X}_n \to x\mathbf{e} \quad \text{in} \quad D \quad \text{w.p.1}$$

for

$$\mathbf{X}_n(t) \equiv n^{-1} X(nt), \quad t \ge 0 \, .$$

By Theorem 8.2.5,

$$\phi(\mathbf{X}_n) \to \phi(x\mathbf{e}) \quad \text{in} \quad D \quad \text{w.p.1} \quad \text{as} \quad n \to \infty \, .$$

However, condition (9.9) implies that $\phi(x\mathbf{e}) = \mathbf{0}$. Hence, (9.10) is obtained by applying the projection map $\pi_1(x) = x(1)$. Finally, we obtain (9.11) from (9.8) after noting from (9.1) that $(I - Q)^{-1} Z \ge (I - Q)^{-1} X$. ∎

Theorem 8.9.1 does not cover all cases, because it requires the special initial condition $X(0) = 0$. However, we also obtain additional results with other initial conditions below. A difficulty occurs because in general the initial condition $X(0)$ and the remaining net-input process $\{X(t) - X(0) : t \geq 0\}$ are dependent. Hence, in general we cannot talk about the increments process as if it did not depend upon the initial condition. Nevertheless, we are able to obtain some positive results. We first establish a tightness result; see Section 11.6 of the book.

Theorem 8.9.3. (tightness under general initial conditions) *If X has stationary ergodic increments, and if condition* (9.7) *holds, then the family of random variables $\{Z(t) : t \geq 0\}$ is tight in $\mathbb{R}^k$.*

Since tightness in product spaces is equivalent to tightness of the components in each coordinate by Theorem 11.6.7 in the book, Theorem 8.9.3 implies the following.

Corollary 8.9.1. (tightness of the finite-dimensional distributions) *Under the conditions of Theorem* 8.9.3*, the family $\{Z_s(t_1), \ldots, Z_s(t_m) : s \geq 0\}$ is tight in $\mathbb{R}^{km}$ for every positive integer m and m time points $0 \leq t_1 < \cdots < t_m$.*

We can combine Prohorov's theorem (Theorem 11.6.1 in the book) with monotonicity to obtain the following result.

Corollary 8.9.2. (convergence of subsequences) *Under the conditions of Theorem* 8.9.3*, every subsequence $\{Z(t_k) : k \geq 1\}$ based on a sequence $\{t_k : k \geq 1\}$ of nonnegative numbers has a convergent subsequence $\{Z(t_k') : k \geq 1\}$. If $Z(t_k) \Rightarrow L$ in $\mathbb{R}^k$ as $t_k \to \infty$, then*

$$Z_*(0) \leq_{st} L \ , \tag{9.12}$$

where Z_* is the stationary process obtained in Theorem 8.9.1 *and*

$$P(L^i < \infty) = 1, \quad 1 \leq i \leq k \, .$$

If we can conclude that the process Z gets arbitrarily close to the origin, then we can replace tightness in Theorem 8.9.3 with convergence.

Theorem 8.9.4. (convergence if the origin is approached) *If, in addition to the assumptions of Theorem* 8.9.3*, for any $\epsilon > 0$ there exists random time T_ϵ with*

$$P(T_\epsilon < \infty) = 1 \tag{9.13}$$

such that

$$\|Z(T_\epsilon)\| < \epsilon \ , \tag{9.14}$$

then the finite-dimensional distributions of Z_s *in* (9.3) *converge as* $s \to \infty$ *to the finite-dimensional distributions of the limit process* Z_* *in Theorem* 8.9.1.

We can obtain a stronger conclusion if the origin is actually hit for all initial positions.

Theorem 8.9.5. (coupling if the origin is always hit) *If, in addition to the assumptions of Theorem* 8.9.3, *for each initial value* $X(0)$, *there exists a random time* T *with* $P(T < \infty) = 1$ *such that* $Z(T) = 0$, *then the process* $\{Z(t) : t \geq 0\}$ *couples with the stationary version in finite time, so that*

$$\lim_{s\to\infty} Ef(Z_s) = Ef(Z_*)$$

for all measurable real-valued functions f *on* D^k.

However in general $\{Z(t) : t \geq 0\}$ need never visit a neighborhood of the origin.

Example 8.9.1. *The process* Z *need not vist a neighborhood of the origin.* To see that it is possible to have $Z(t) \neq (0, \ldots, 0)$, and even $\|Z(t)\| > c > 0$ for some constant c, for all $t \geq 0$ under the conditions of Theorem 8.9.3, consider a two-dimensional case in which *either* $X^1(t+\epsilon) - X^1(t) > \delta\epsilon$ or $X^2(t+\epsilon) - X^2(t) > \delta\epsilon$ for all t, where ϵ and δ are small positive constants. For example, let

$$V^1(t) = \begin{cases} \delta, & 3k \leq t < 3k+2 \\ -1, & 3k+2 \leq t < 3k+3 \end{cases}$$

and

$$V^2(t) = \begin{cases} \delta, & 3k+1 \leq t < 3k+3 \\ -1, & 3k \leq t < 3k+1 \end{cases}$$

for all nonnegative integers k. Let U be uniformly distributed on $[0,3]$. Then $\{V(t) : t \geq 0\} \equiv \{(V^1(t+U), V^2(t+U)) : t \geq 0\}$ is a stationary process on the positive half line, so that $X(t) \equiv \int_0^t V(u)du$ is a net input process with stationary increments. It is easy to see that the content process associated with $Q = 0$ never hits the origin after time 0, and yet for $\delta < 1/2$

it has a proper steady-state distribution. Indeed, eventually $Z(t)$ follows the deterministic trajectory with $Z(3k-U)=(2\delta,0)$, $Z(3k+1-U)=(0,\delta)$ and $Z(3k+2-U)=(\delta,2\delta)$. This steady-state trajectory is reached for

$$t \geq 3\left(1+\frac{\max\{Z^1(0),Z^2(0)\}}{1-2\delta}\right) .$$

By an appropriate choice of units, the limiting trajectory falls outside any neighborhood of the origin.

We can also modify Example 8.9.1 to construct two stable content processes which differ only in their initial conditions but do *not* couple in finite time.

Example 8.9.2. *Failure to couple in finite time.* We modify Example 8.9.1 by letting $Q^t_{1,2}=P^t_{2,1}=\epsilon$ for $0<\epsilon<\delta$. The content process now approaches the deterministic trajectory with $Z(3k-U)=(2\delta-\epsilon+\epsilon',0)$, $Z(3k+1-U)=(0,\delta-\epsilon+\epsilon')$ and $Z(3k+2-U)=(\delta,2\delta-\epsilon+\epsilon')$, where $\epsilon'=(2\delta^2-\epsilon\delta)/(1+\delta)$. However, unlike Example 8.9.1, the content process typically does not reach this cycle in finite time. Suppose one of the two content processes starts above another, where they have the same net input process X. They move together until they hit a boundary. However, when the lower process is on a boundary and the other is not, the other coordinate of the two processes moves away from each other at rate ϵ. Hence the processes cannot couple on any boundary, although they do get closer in an appropriate metric as they hit the boundaries.

Since many of the limiting net-input processes X will be Lévy processes (i.e., will have stationary independent increments), we now add the independent increments property.

Theorem 8.9.6. (existence and uniqueness for Lévy net-input processes with independent coordinate processes) *Suppose that $X\equiv(X^1,\ldots,X^k)$ has mutually independent marginal processes X^i, $1\leq i\leq k$, each with stationary and independent increments, $X(0)$ is proper and condition* (9.7) *holds. Then the limit* (9.2) *holds and the limit has the same distribution as the limit $Z_*(0)$ associated with $X(0)=0$.*

As mentioned in the beginning of this section, Theorem 8.9.6 applies to the limit process in Section 8.6 when the scaled versions of the exogenous arrival process C converge to a Lévy process with mutually independent coordinate processes, because the only stochastic component in the net-input process X^i is C^i. However, in general, Theorem 8.9.6 does not apply to the heavy-traffic limits for the queueing network in Section 8.7. It does in the special case in which the coordinate limit process $\mathbf{X}^i$ depends only on the limit of the scaled process associated with the i^{th} coordinate arrival process.

It remains to establish more general conditions under which the assumptions of Theorems 8.9.4 and 8.9.5 are satisfied. It also remains to find useful expressions for the limiting distributions. Explicit expressions for the Laplace transforms of non-product-form two-dimensional stationary buffer-content distributions of stochastic fluid networks with Lévy exogenous input processes have been determined by Kella and Whitt (1992a) and Kella (1993).

8.9.2. Proofs

We now provide the missing proofs for the results above. We first establish some bounds and inequalities to be used in the proofs. Let $D_{\downarrow}^k$ be the subset of nonnegative nonincreasing functions in D^k. As before, let $D_{\uparrow}^k$ be the subset of nonnegative nondecreasing functions in D^k.

Theorem 8.9.7. (bounds and inequalities for the reflection map) *Assume that $x_1, x_2 \in D$ with $x_2 - x_1 \in D_{\uparrow}^k$, $x_3 = x_1 + (I - Q)\psi(x_2)$ and $w \geq 0$ in $\mathbb{R}^k$. Then*

(i) $\phi(x_2) \geq \phi(x_1)$,

(ii) $\psi(x_1) - \psi(x_2) \in D_{\uparrow}^k$,

(iii) $\psi(x_1) - \psi(x_2) \leq (I - Q)^{-1}(x_2 - x_1)$,

(iv) $\psi(x_3) = \psi(x_1) - \psi(x_2)$,

(v) $0 \leq (I - Q)^{-1}(\phi(x_2) - \phi(x_1)) \leq (I - Q)^{-1}(x_2 - x_1)$,

(vi) $0 \leq 1(\phi(x_2) - \phi(x_1)) \leq 1(x_2 - x_1)$,

(vii) $(I - Q)^{-1}(\phi(x_1 + w) - \phi(x_1)) \in D_{\downarrow}^k$,

(viii) $1(\phi(x_1 + w) - \phi(x_1)) \in D_{\downarrow}^k$.

Proof. Parts (i) and (ii) follow for $x_1, x_2 \in D_c$ by induction from Corollary 14.3.2 and Lemma 14.3.3 in the book. They then follow for $x_1, x_2 \in D$ by taking limits: Given $x_1, x_2 \in D$ with $x_2 - x_1 \in D_\uparrow$, it is possible to find $x_{1,n}$ and $x_{2,n} \in D_c$ with $x_{2,n} - x_{1,n} \in D_\uparrow$ for all n and $\|x_{j,n} - x_j\| \to 0$ as $n \to \infty$ for $j = 1, 2$. Part (iii) follows from Theorem 14.2.4 in the book because

$$\eta_1(x_2 - x_1) = x_2 - x_1 \quad \text{for} \quad x_2 - x_1 \in D_\uparrow \ .$$

Turning to (iv), note that

$$0 \leq \phi(x_3) = x_1 + (I - Q)(\psi(x_2) + \psi(x_3)) \tag{9.15}$$

and

$$0 \leq \phi(x_1) = x_3 + (I - Q)(\psi(x_1) - \psi(x_2)) \ . \tag{9.16}$$

From (9.15) and minimality of $\psi(x_1)$, it follows that $\psi(x_1) \leq \psi(x_2) + \psi(x_3)$ for any choice of x_1 and x_2. From (9.16) and minimality of $\psi(x_3)$, it follows that $\psi(x_3) \leq \psi(x_1) - \psi(x_2)$. Hence we must have $\psi(x_3) = \psi(x_1) + \psi(x_2)$ as claimed. Parts (v) –(viii) follow from the relations $(I - Q)^{-1}\phi(x) = (I - Q)^{-1}x + \psi(x)$ and $1(I - Q) \geq 0$, and Theorem 14.2.4 in the book. ∎

We now apply Theorem 8.9.7 to determine the shape of several mean values as a function of time.

Corollary 8.9.3. (concavity of mean values) *If X has stationary increments with $X(0) = 0$, then the functions $((I - Q)^{-1}E\phi(X)(t))^i$, $E\psi^i(X)(t)$ and $1E\phi(X)(t)$ are concave functions of t for each i.*

Proof. Apply parts (vii), (ii) and (viii) of Theorem 8.9.7, respectively. We will only prove the first result because the three proofs are essentially the same. It suffices to show that

$$((I - Q)^{-1}E[\phi(X)(t + s) - \phi(X)(s)])^i$$

is nonincreasing in s for all t, but that follows from Theorem 8.9.7(vii), because $\phi(X)(t + s)$ is distributed as the reflection of $X_s(t) \equiv X(s + t) - X(s)$ starting at $\phi(X)(s)$ evaluated at t, while $\phi(X)(s)$ is distributed as the reflection of X_s starting at 0 evaluated at t, since the law of X_s is independent of s. ∎

We say that a real-valued function f on $\mathbb{R}_+$ is *subadditive* if

$$f(t_1 + t_2) \leq f(t_1) + f(t_2)$$

for all $t_1, t_2 \in \mathbb{R}_+$. We say that an $\mathbb{R}^k$-valued stochastic process $\{X(t) : t \geq 0\}$ is *stochastically increasing and subadditive* (SIS) if

$$Ef(X(t_1 + t_2)) \leq Ef(X(t_1)) + Ef(X(t_2))$$

for all nondecreasing subadditive real-valued functions f on $\mathbb{R}^k$.

Corollary 8.9.4. (SIS property) *If X has stationary increments with $X(0) = 0$, then $(I - Q)^{-1}Z$ and $1Z$ are stochastically increasing and subadditive stochastic processes.*

Proof. Since the two results are proved similarly, we only prove the first. Let

$$\tilde{Z}_{s_1,s_2}(t) \equiv (I - Q)^{-1}Z(t)$$

with Z having initial value $Z(s_1)$ and net input $X_{s_2}(t) \equiv X(s_2 + t) - X(s_2)$, $t \geq 0$, where $0 \leq s_1 \leq s_2$. By Theorem 8.9.7(vii),

$$\tilde{Z}_{s,s}(t) - \tilde{Z}_{0,s}(t) \leq \tilde{Z}_{s,s}(0) - \tilde{Z}_{0,s}(0) = \tilde{Z}_{s,s}(0)$$

for all $s, t \geq 0$, or

$$\tilde{Z}_{0,0}(s + t) = \tilde{Z}_{s,s}(t) \leq \tilde{Z}_{0,s}(t) + \tilde{Z}_{s,s}(0) \ ,$$

so that, for any subadditive function f,

$$\begin{aligned} E[f(\tilde{Z}_{0,0}(s + t)) &\leq E[f(\tilde{Z}_{0,s}(t) + \tilde{Z}_{s,s}(0))] \\ &\leq E[f(\tilde{Z}_{0,s}(t))] + E[f(\tilde{Z}_{s,s}(0))] \\ &\leq E[f(Z_{0,0}(t))] + E[f(Z_{0,0}(s))] \ , \end{aligned}$$

with the last line holding because there is equality in distribution for the respective terms. ■

A key to establishing the important Theorems 14.8.1 and 14.8.6 in the book is the following stochastic increasing property, which we deduce from Theorem 8.9.7.

Theorem 8.9.8. (stochastic increasing starting empty) *If X has stationary increments and $X(0) = 0$, then the family of processes $\{Z_s : s \geq 0\}$ in (9.3) is stochastically increasing in s, i.e.,*

$$Ef(Z_{s_1}) \leq Ef(Z_{s_2})$$

for $0 \leq s_1 < s_2$ and all bounded measurable nondecreasing real-valued functions f on $D \equiv D([0, \infty), \mathbb{R}^k)$, using the componentwise order on D.

Proof. Let $\hat{Z}_s(t)$ $(Z_s(t))$ be the content with $Z(0) = 0$ $(Z(0) = Z(s))$ and input increments from X_s in equation (14.8.6) in the book. Then, for $0 \le s_1 < s_2$,

$$\hat{Z}_{s_2-s_1}(s_1+t) \le Z_{s_2-s_1}(s_1+t) \quad \text{for all} \quad t \ge 0 \quad \text{w.p.1}\,,$$

by Theorem 8.9.7 because $\hat{Z}_{s_2-s_1}(0) \equiv 0 \le Z_{s_2-s_1}(0) \equiv Z(s_2-s_1)$ and both processes have the common input increments from X_s. Hence,

$$Ef(\hat{Z}_{s_2-s_1}) \le Ef(Z_{s_2-s_1})$$

for all nondecreasing bounded measurable real-valued functions f on D, using the usual componentwise order. However, since $X_s \stackrel{\mathrm{d}}{=} X$,

$$\{\hat{Z}_{s_1-s_1}(s_1+t) : t \ge 0\} \stackrel{\mathrm{d}}{=} \{Z(s_1+t) : t \ge 0\} \equiv Z_{s_1}$$

and

$$\{Z_{s_2-s_1}(s_1+t) : t \ge 0\} = \{Z(s_2+t) : t \ge 0\} \equiv Z_{s_2}\,.$$

These last three relations combine to establish the desired conclusion. ∎

We use the following result to establish Theorem 14.8.3 in the book.

Theorem 8.9.9. (tightness solidarity) *Suppose that X has stationary increments. Then $\{Z(t) : t \ge 0\}$ is tight for all proper distributions of $X(0)$ if and only if it is tight for any one.*

Proof. Note that $\{Z(t) : t \ge 0\}$ is tight if and only if $\{(I-Q)^{-1}Z(t) : t \ge 0\}$ is tight. By Theorem 8.9.7, the processes $(I-Q)^{-1}Z(t)$ starting at $X(0)$ and 0, with common increments from X, differ by at most $(I-Q)^{-1}\|X(0)\|$. Hence they are tight or nontight together. Hence, the tightness of the process with one proper initial condition implies the tightness of the process starting at 0. Then the tightness of the process starting at 0 implies the tightness of any other process with another initial condition. ∎

The key to our tightness results, and thus also our convergence results, is our ability to bound the marginal processes Z^i associated with a k-dimensional reflected process $Z \equiv (Z^1, \ldots, Z^k)$ by related well-studied and well-understood one-dimensional reflections. For that purpose, we have the following bounds.

Theorem 8.9.10. (one-dimensional reflection bounds) *For any $x \in D^k$ and $Q \in \mathcal{Q}$,*

$$\psi_1((I-Q)^{-1}x) \le \psi(x) \le (I-Q)^{-1}\psi_1(x) \tag{9.17}$$

and

$$\phi_1((I-Q)^{-1}x) \le (I-Q)^{-1}\phi(x) \le (I-Q)^{-1}\phi_1(x) \ , \tag{9.18}$$

where $(\psi_1, \phi_1) : D^k \to D^{2k}$ *with*

$$(\psi_1(x)^i, \phi_1(x)^i) \equiv \hat{\psi}_1(x^i), \hat{\phi}_1(x^i)), \quad 1 \le i \le k \ ,$$

and $(\hat{\psi}_1, \hat{\phi}_1) : D \to D^2$ *being the one-dimensional reflection map, i.e.,*

$$\hat{\phi}_1(x^i) \equiv x^i + \hat{\psi}_1(x^i)) \tag{9.19}$$

and

$$\hat{\psi}_1(x^i) \equiv - \inf_{0 \le s \le t} \{x_i(s)^-\}, \quad t \ge 0 \, . \tag{9.20}$$

Proof. For the upper bounds, note that

$$\phi_1(x) = x + \psi_1(x) = x + (I-Q)(I-Q)^{-1}\psi_1(x) \ .$$

By the minimality of $\psi(x)$ in the definition of (ψ, ϕ),

$$\psi(x) \le (I-Q)^{-1}\psi_1(x) \ .$$

Therefore,

$$(I-Q)^{-1}\phi(x) = (I-Q)^{-1}x + \psi(x) \le (I-Q)^{-1}x + (I-Q)^{-1}\psi_1(x) = (I-Q)^{-1}\phi_1(x) \, .$$

Similarly, for the lower bound,

$$\phi_1((I-Q)^{-1}x) = (I-Q)^{-1}x + \psi_1((I-Q)^{-1}x) \tag{9.21}$$

and

$$(I-Q)^{-1}\phi(x) = (I-Q)^{-1}x + \psi(x) \, .$$

Since $(I-Q)^{-1}\phi(x) \ge 0$, we can apply the minimality of ψ_1 in (9.21) to deduce that

$$\psi_1((I-Q)^{-1}x) \le \psi(x)$$

and

$$\phi_1((I-Q)^{-1}x) \le (I-Q)^{-1}\phi(x) \ . \quad \blacksquare$$

In order to apply the one-dimensional reflection bounds in Theorem 8.9.10, we need to have a net input process X with negative drift in each coordinate. However, from (9.7), we only have X such that $(I-Q)^{-1}X$ has negative drift in each coordinate. We now show that, given X such that $(I-Q)^{-1}X$ has negative drift, we can bound $(I-Q)^{-1}\phi(X)$ above by $(I-Q)^{-1}\phi(X_y)$, where $X_y(t) \equiv X(t) - yt$, $t \ge 0$ and X_y has negative drift in each coordinate.

Theorem 8.9.11. (upper bound with negative drift) *Let X be a random element of D^k with stationary increments such that*

$$E[X(1) - X(0)] = x \quad \textit{and} \quad ((I-Q)^{-1}x)^i < 0, \quad 1 \le i \le k\,.$$

For any $y \in \mathbb{R}^k$ with $y^i > x^i$ and $((I-Q)^{-1}y)^i < 0$, $1 \le i \le k$ (there necessarily is one), let

$$X_y(t) \equiv X(t) - yt, \quad t \ge 0\,.$$

Then X_y has stationary increments (and ergodic increments if X does) with

$$E[X_y(1) - X_y(0)]^i = x^i - y^i < 0, \quad 1 \le i \le k\,,$$

and

$$(I-Q)^{-1}\phi(X) \le (I-Q)^{-1}\phi(X_y)\ . \tag{9.22}$$

Proof. Only the final conclusion (9.22) requires discussion. Let e be the identity map, i.e., $e(t) = t$, $t \ge 0$. Recall that

$$\begin{aligned}
\phi(X)(t) &\equiv X(t) + (I-Q)\psi(X)(t) \\
\phi(X_y)(t) &\equiv X(t) - yt + (I-Q)\psi(X_y)(t) \\
\phi(ye)(t) &\equiv yt + (I-Q)\psi(ye)(t), \quad t \ge 0\,.
\end{aligned}$$

First, since $(I-Q)^{-1}y \le 0$, it is easy to see that

$$\phi(ye)(t) = 0 \quad \text{and} \quad \psi(ye)(t) = -(I-Q)^{-1}yt\ .$$

Then

$$\phi(X_y)(t) \equiv \phi(X_y)(t) + \phi(ye)(t) = X(t) + (I-Q)(\psi(X_y)(t) + \psi(ye)(t))\,.$$

By the minimality of $\psi(X)$,

$$\psi(X) \le \psi(X_y) + \psi(ye)$$

and

$$\begin{aligned}
(I-Q)^{-1}\phi(X) &= (I-Q)^{-1}X + \psi(X) \\
&\le (I-Q)^{-1}X + \psi(X_y) + \psi(ye) = (I-Q)^{-1}\phi(X_y)\,. \quad \blacksquare
\end{aligned}$$

We now state the classical one-dimensional result, which depends on the fact that the reflected content $\phi(X)(t)$ has the same distribution as the supremum of the time-reversed net-input process for each t (but not for multiple t).

Theorem 8.9.12. (classical one-dimensional result) *If X is a real-valued stochastic process with stationary increments such that*

$$X_r(t) \equiv -X(-t) \to -\infty$$

as $t \to \infty$ and $X(0)$ is proper, then there exists a proper random variable L such that

$$\phi(X)(t) \Rightarrow L \quad in \quad \mathbb{R} \quad as \quad t \to \infty\,.$$

Proof. First assume that $X(0) = 0$. Given the time reversed process $X_r(t) \equiv -X(-t)$, $t \geq 0$, note that

$$\phi(X)(t) \stackrel{\mathrm{d}}{=} X_r^{\uparrow}(t) \quad \text{for each} \quad t \geq 0\,.$$

Since $X_r(t) \to -\infty$ as $t \to \infty$ and $X_r \in D$,

$$X_r^{\uparrow}(t) \to X_r^{\uparrow}(\infty) < \infty \quad \text{as } t \to \infty \quad \text{w.p.1}\,.$$

Hence the desired conclusion holds with the proper limit $L \stackrel{\mathrm{d}}{=} X_r^{\uparrow}(\infty)$. Now suppose that $X(0) \neq 0$. Since $X(t) \to -\infty$ w.p.1, the processes $Z(t)$ starting at 0 and $X(0)$, with common net input process X, couple w.p.1. Hence we can invoke Theorem 8.9.5. ■

We now provide the missing proofs of theorems earlier in this section.

Proof of Theorem 8.9.1. By Theorem 8.9.8, the family of processes Z_s in (9.3) are stochastically increasing in s. Consequently, the finite-dimensional distributions of Z_s are stochastically increasing in s. The cumulative distribution functions (cdf's) of $(Z_s(t_1), \ldots, Z_s(t))$ in $\mathbb{R}^{km}$ thus converge as $s \to \infty$ to a possibly improper cdf; e.g., see Chapter VIII of Feller (1971). It thus suffices to show that $\{Z^i(t) : t \geq 0\}$ is tight for each i, for which it suffices to show that $\{((I-Q)^{-1}Z(t))^i : t \geq 0\}$ is tight for each i. (The tightness implies that the limiting cdf is proper.) By Theorem 8.9.11, we can bound $(I-Q)^{-1}\phi(X)$ above by $(I-Q)^{-1}\phi(X_y)$, where $X_y(t) \equiv X(t) - yt$ for appropriate $y \in \mathbb{R}^k$ and

$$-\infty < E[X_y^i(1) - X_y^i(0)] < 0 \quad \text{for all} \quad i\,. \tag{9.23}$$

By (9.18) in Theorem 8.9.10, we can bound $(I-Q)^{-1}\phi(X_y)$ above by $(I-Q)^{-1}\phi_1(X_y)$, where ϕ_1 is the vector of one-dimensional reflection maps. Hence it suffices to show that $\{\hat{\phi}_1(X_y^i(t) : t \geq 0\}$ is tight for each i, where

$\hat{\phi}_1$ is the one-dimensional reflection map in (9.19). However, $\hat{\phi}_1(X_y^i)(t)$ converges to a proper limit by Theorem 8.9.12. The condition $-X(-t) \to -\infty$ in Theorem 8.9.12 holds for X_y by virtue of (9.23) and that fact that $\{-X(t)\}$ is a process with stationary ergodic increments (Stationarity and metric transitivity are invariant under time reversal, and ergodicity is equivalent to metric transitivity.) The assumptions imply that

$$-t^{-1}X_y^i(-t) \to E[X_y^i(1) - X_y^i(0)] \quad \text{as} \quad t \to \infty \quad \text{w.p.1}$$

for each i, which implies that $-X_y^i(-t) \to -\infty$ w.p.1 as $t \to \infty$ for each i. ∎

Proof of Theorem 8.9.3. By Theorem 8.9.1, we have convergence to a proper limit L for the process $\{Z_0(t) : t \geq 0\}$ starting from the origin. By the continuous mapping theorem,

$$(I-Q)^{-1}Z_0(t) \Rightarrow (I-Q)^{-1}L \quad \text{as} \quad t \to \infty\,.$$

If $X(0)$ is proper, then so is $X(0)^+ \equiv (X^1(0)^+, \ldots, X^k(0)^+)$. Then, from Theorem 8.9.7(i) and (v),

$$0 \leq (I-Q)^{-1}Z_{X(0)}(t) \leq (I-Q)^{-1}Z_{X(0)^+}(t) \leq (I-Q)^{-1}Z_0(t) + (I-Q)^{-1}X(0)^+\,,$$

where here $Z_w(t)$ denotes the process governed by X with initial position w. Hence

$$P(|((I-Q)^{-1}Z_{X(0)}(t))^i| > 2K) \leq P(|(I-Q)^{-1}Z_0(t))^i| > K) \\ + P(|(I-Q)^{-1}X(0)^+)^i| > K)\,,$$

so that the tightness holds by the results above. ∎

Proof of Corollary 8.9.2. We can combine Prohorov's theorem (Theorem 11.6.1 in the book) with monotonicity. By Theorem 8.9.7,

$$Z_0(t) \leq Z_{X(0)}(t) \quad \text{for all} \quad t\,. \tag{9.24}$$

Since $Z_0(t) \Rightarrow Z_*(0)$ by Theorem 8.9.1, (9.12) must hold. (Stochastic order on $\mathbb{R}^k$ is preserved under weak convergence.) ∎

Proof of Theorem 8.9.4. Since (9.24) holds and

$$(I-Q)^{-1}(Z_{X_0} - Z_0) \in D_{\downarrow}^k\,,$$

by Theorem 8.9.7(vii),

$$0 \leq (I-Q)^{-1}(Z_{X(0)}(t) - Z_0(t)) \leq (I-Q)^{-1}1\epsilon \quad \text{for all} \quad t \geq T_\epsilon\,.$$

Since $Z_0(t) \Rightarrow L$ as $t \to \infty$ by Theorem 8.9.1 and ϵ is arbitrary, we must have $Z_{X(0)}(t) \Rightarrow L$ too. ∎

Proof of Theorem 8.9.5. The processes starting at 0, $X(0)$ or $Z_*(0)$ can all be given a common net input process $X(t) - X(0)$, $t \geq 0$. Hence, they all must couple when the process starting at $Z(0) \vee Z^*(0)$ first hits the origin. ∎

In preparation for the proof of Theorem 8.9.6, we now establish a property of the limiting distribution in the one-dimensional case when X is a Lévy process.

Theorem 8.9.13. (mass near the origin) *If, in addition to the assumptions of Theorem* 8.9.12*, the one-dimensional net-input process X has independent increments, then*

$$P(L < \epsilon) > 0 \quad \textit{for all} \quad \epsilon > 0\,,$$

where L is the limiting random variable.

Proof. Consider the time reversed process X_r defined in Theorem 8.9.12. It suffices to show that $P(X_r^{\uparrow}(\infty) < \epsilon) > 0$. Suppose not. Then $P(X_r^{\uparrow}(\infty) \geq \epsilon) = 1$, which implies that $P(T_\epsilon < \infty) = 1$, where

$$T_\epsilon = \inf\{t > 0 : X_r(t) \geq \epsilon\}\ .$$

Using the regeneration property associated with the stationary independent increments, that in turn implies that

$$\limsup_{t \to \infty} X_r(t) = +\infty \quad \text{w.p.1}\ ,$$

which contradicts the limit $X_r(t) \to -\infty$ w.p.1. Hence we must have $P(X_r^{\uparrow}(\infty) < \epsilon) > 0$ for all $\epsilon > 0$ as claimed. ∎

Proof of Theorem 8.9.6. The conditions allow us to apply Theorem 8.9.4. Theorems 8.9.10 and 8.9.11 allow us to bound the process $(I - Q)^{-1}\phi(X)(t)$ above by $(I - Q)^{-1}\phi_1(X_y)(t)$, as in the proof of Theorem 8.9.1. However, $\phi_1(X_y)$ has mutually independent coordinate processes. Let L^i be the limit random variable for the one-dimensional process associated with $\phi_1(X_y)$ and coordinate i. Since, for any $\epsilon > 0$,

$$P(L^1 \leq \epsilon, \ldots, L^k \leq \epsilon) = \prod_{i=1}^{k} P(L^i \leq \epsilon) > 0$$

by the independence and Theorem 8.9.13 we must have $P(T_\epsilon < \infty) = 1$ for the random time T_ϵ in Theorem 8.9.4. ∎

As mentioned earlier, Theorem 8.9.6 applies to the limit process in Section 14.6 in the book when the scaled versions of the exogenous arrival process C converge to a Lévy process with mutually independent coordinate processes, because the only stochastic component in the net-input process X^i is C^i. However, in general, Theorem 8.9.6 does not apply to the heavy-traffic limits for the queueing network in Section 14.7 of the book. It does in the special case in which the coordinate limit process $\mathbf{X}^i$ depends only on the limit of the scaled process associated with the i^{th} coordinate arrival process.

Chapter 9

Nonlinear Centering and Derivatives

9.1. Introduction

In this chapter we continue to study the useful functions introduced in Section 3.5 of the book and investigated in Chapter 13 of the book. Now we consider supremum, reflection and inverse maps with nonlinear centering.

Following Mandelbaum and Massey (1995), we identify the limit of the properly scaled function as a derivative. We also show how the convergence-preservation results for the reflection map can be applied to establish heavy-traffic limits for nonstationary queues.

To explain the derivative representation, recall that our previous results on the preservation of convergence with linear centering started with the assumed convergence

$$c_n(x_n - e) \to y \quad \text{in} \quad D\,, \tag{1.1}$$

where $c_n \to \infty$ and e is the identity function, i.e., $e(t) = t$, $t \geq 0$. Given (1.1), we found conditions under which

$$c_n(\phi(x_n) - e) \to z \quad \text{in} \quad D \tag{1.2}$$

for various functions ϕ and we identified the limit z. We also obtained some extensions in which the linear centering function e in (1.1) is replaced by a nonlinear function x; i.e., instead of (1.1), we assumed that

$$c_n(x_n - x) \to y \quad \text{in} \quad D \quad \text{as} \quad n \to \infty\,, \tag{1.3}$$

where $c_n \to \infty$. In particular, see Theorems 13.3.2, 13.7.2 and 13.7.4 and Corollaries 13.4.1, 13.7.1 and 13.7.2 in the book. We now want to obtain some further results of this kind.

Given (1.3), we have as a consequence

$$x_n \to x \quad \text{in} \quad D\,. \tag{1.4}$$

Hence, for any continuous function ϕ, we have

$$\phi(x_n) \to \phi(x) \quad \text{in} \quad D\,. \tag{1.5}$$

Thus we want to find functions $z \in D$ and regularity conditions such that

$$c_n(\phi(x_n) - \phi(x)) \to z \quad \text{in} \quad D\,. \tag{1.6}$$

The previous results with centering by e were of this form, where $\phi(x) = x = e$. The M topologies play an important role, because the limit z in (1.6) may have discontinuities even when y, x and x_n are all continuous functions.

In a probability context, (1.6) is interesting because it corresponds to a FCLT refinement to a nonlinear FLLN. We may have scaled stochastic processes $\{X_n(t) : t \geq 0\}$ which obey a nonlinear FWLLN of the form

$$X_n \Rightarrow x \quad \text{in} \quad D\,, \tag{1.7}$$

where x is a nonlinear deterministic function, and a FCLT refinement of the form

$$c_n(X_n - x) \Rightarrow Y \quad \text{in} \quad D\,, \tag{1.8}$$

where $c_n \to \infty$. From the FWLLN (1.7) it follows directly that

$$\phi(X_n) \Rightarrow \phi(x) \quad \text{in} \quad D \tag{1.9}$$

for a continuous function ϕ. Our goal is to establish the FCLT refinement of (1.9), i.e.,

$$c_n(\phi(X_n) - \phi(x)) \Rightarrow Z \quad \text{in} \quad D\,. \tag{1.10}$$

As before, (1.10) follows from (1.8) when (1.6) follows from (1.3). Hence we focus on obtaining (1.6) from (1.3).

It is interesting that, under regularity conditions, z in (1.6) can be thought of as a derivative of the map ϕ, in particular, a directional derivative of ϕ in the direction y, evaluated at x. To see that, it is convenient to index the functions by ϵ in such a way that x_n becomes x_ϵ and c_n becomes ϵ^{-1}. (That is without loss of generality.) Then (1.3) is equivalent to

$$\epsilon^{-1}(x_\epsilon - x) \to y \quad \text{as} \quad \epsilon \downarrow 0\,. \tag{1.11}$$

Without being too precise, we can rewrite (1.11) as

$$x_\epsilon = x + \epsilon y + o(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0 \; . \tag{1.12}$$

Now, assuming that the function $\phi : D \to D$ satisfies

$$\phi(\tilde{x}_\epsilon + o(\epsilon)) - \phi(\tilde{x}_\epsilon) = o(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0 \tag{1.13}$$

for any $\tilde{x}_\epsilon$ with $\tilde{x}_\epsilon \to x$ in D as $\epsilon \downarrow 0$ (which is not automatic), we have

$$\phi(x_\epsilon) = \phi(x + \epsilon y) + o(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0 \tag{1.14}$$

and, given the ϵ-analog of (1.6),

$$\phi(x + \epsilon y) = \phi(x) + \epsilon z + o(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0 \; . \tag{1.15}$$

From (1.15), it is evident that z can be given the directional derivative interpretation. Moreover, (1.14) and (1.15) together imply that

$$\epsilon^{-1}(\phi(x_\epsilon) - \phi(x)) \to z \quad \text{as} \quad \epsilon \downarrow 0 \; . \tag{1.16}$$

Equivalently, (1.3), (1.13) and (1.16) imply the desired (1.6).

Here is how the present chapter is organized: In Section 2 we investigate when the convergence-preservation question (when (1.3) implies (1.6)) can be reduced to the derivative determination in (1.15). Unfortunately, we are not able to show that this can be done as generally as we would like. This step seems to be the weak link in our analysis in this chapter. Hopefully future research will provide further insights.

In Sections 9.3 – 9.5 we determine sufficient conditions for the derivatives of the supremum and reflection maps to exist and determine their form. As should be anticipated from Chapter 13 in the book, the reflection derivative can be expressed in terms of the supremum derivative. The M_1 topology plays an important role even if x and y in (1.3) are both continuous.

In Section 9.6 we apply the derivative calculation and convergence-preservation results for the reflection map to establish heavy-traffic limits for nonstationary queues. For example, these results cover the $\mathrm{M_t/M_t/1}$ queue with time-dependent arrival and service rates.

Finally, in Section 9.7 we consider the derivative of the inverse map.

9.2. Nonlinear Centering and Derivatives

In this section we investigate when the desired convergence-preservation (when (1.11) implies (1.16)) can be deduced by determining the derivative via (1.15). For any function $\phi : D \to D$, a general approach to establish the desired limit (1.16) for $\phi(x_\epsilon)$ is to exploit the triangle inequality:

$$\begin{aligned} d(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], z) \quad &\leq \quad d(\epsilon^{-1}[\phi(x + \epsilon y) - \phi(x)], z) \\ &\quad + d(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], \epsilon^{-1}[\phi(x + \epsilon y) - \phi(x)]) \end{aligned} \tag{2.1}$$

for an appropriate metric d. A limit for the first term in (2.1) as $\epsilon \downarrow 0$ identifies z as the derivative of ϕ in the direction y evaluated at x. In addition to establishing the existence of this derivative, we must also show that the second term in (2.1) converges to 0 as $\epsilon \downarrow 0$. Surprisingly, the second term presents difficulties. However, we are able to show that it is negligible under regularity conditions. The results are in a good form when $y \in C$, but not so good when only $y \in D$. (Recall that the limit z in (1.16) may be discontinuous even if $y \in C$, so the case $y \in C$ is interesting and important.)

We now obtain results about the second term in (2.1) for general functions $\phi : (D_1, d_1) \to (D_2, d_2)$, where $D_i \equiv D([0, t_i], \mathbb{R}^{k_i})$ for $i = 1, 2$.

Theorem 9.2.1. (reduction of convergence preservation to the derivative) *Suppose that $\phi : (D_1, d_1) \to (D_2, d_2)$, where the metrics d_i satisfy the properties:*

$$d_i(cx_1, cx_2) = cd_i(x_1, x_2) \quad \textit{for all} \quad c > 0, \; i = 1, 2 \; , \tag{2.2}$$

$$d_i(x_1 + x_3, x_2 + x_3) = d_i(x_1, x_2), \; i = 1, 2, \tag{2.3}$$

$$d_2(\phi(x_1), \phi(x_2)) \leq K d_1(x_1, x_2) \quad \textit{for some} \quad K > 0 \; , \tag{2.4}$$

for all x_1, x_2, $x_3 \in D_i$. Then

$$d_2(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], \epsilon^{-1}[\phi(x + \epsilon y) - \phi(x)]) \leq K d_1(\epsilon^{-1}(x_\epsilon - x), y) \; . \tag{2.5}$$

Proof. The conditions imply that

$$\begin{aligned} d_2(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], \epsilon^{-1}[\phi(x + \epsilon y) - \phi(x)]) &= \epsilon^{-1} d_2(\phi(x_\epsilon), \phi(x + \epsilon y)) \\ &\leq \epsilon^{-1} K d_1(x_\epsilon, x + \epsilon y) \\ &= K d_1(\epsilon^{-1}(x_\epsilon - x), y) \; . \quad \blacksquare \end{aligned}$$

Notice that the uniform metric satisfies conditions (2.2) and (2.3). The following application of Theorem 9.2.1 is elementary.

Theorem 9.2.2. (reduction for the supremum and reflection maps with the uniform metric) *If d_1 and d_2 in Theorem* 9.2.1 *above are the uniform* *metric on* $D([0,t],\mathbb{R})$ *and* ϕ *is the supremum function in equation* (13.4.1) *in the book or the reflection map in equation* (13.5.1) *in the book, then the conditions of Theorem* 9.2.1 *above are satisfied, so that conclusion* (2.5) *holds.*

Proof. It is evident that the uniform metric on D satisfies conditions (2.2) and (2.3). The supremum and reflection functions also satisfy (2.4) with respect to the uniform metric by Lemmas 13.4.1 and 13.5.1 in the book.

Example 9.2.1. *The need for the map* ϕ *to be Lipschitz.* To see the need for $\phi : D \to D$ being Lipschitz in Theorem 9.2.1, let $\phi(x)(t) = \sqrt{x(1)}$, $t \geq 0$. If $\|x_\epsilon - x\|_t \to 0$ for $t > 1$, then $\|\phi(x_\epsilon) - \phi(x)\|_t \to 0$, but ϕ is not Lipschitz. Suppose that $x(t) = 0$, $y(t) = 1$ and $x_\epsilon(t) = x(t) + \epsilon y(t) = \epsilon$, $t \geq 0$. Then $\|\epsilon^{-1}(x_\epsilon - x) - y\| = 0$ for all ϵ,

$$\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)](t) = \epsilon^{-1}[\sqrt{\epsilon} - 0] = \epsilon^{-1/2} \to \infty \quad \text{as} \quad \epsilon \downarrow 0 \ . \quad \blacksquare \qquad (2.6)$$

Unfortunately, for the nonuniform Skorohod metrics on D, which we will want to consider when $y \notin C$, we do not have properties (2.2) and (2.3) in Theorem 9.2.1.

Example 9.2.2. *Failure for nonuniform metrics.* Unlike with the uniform metric, we cannot conclude that $d(\epsilon^{-1}x_\epsilon, \epsilon^{-1}x + y) \to 0$ as $\epsilon \downarrow 0$ when $d(\epsilon^{-1}(x_\epsilon - x), y) \to 0$ as $\epsilon \downarrow 0$ if d is the J_1, M_1 or M_2 metric and y is not continuous. To see this, let $x(t) = tI_{[0,1)}(t) + (2-t)I_{[1,2]}(t)$, $y = I_{[0,1)} - I_{[1,2]}$ and $x_\epsilon = (x + \epsilon)I_{[0,1-\epsilon]} + (x - \epsilon)I_{[1-\epsilon,2]}$ in $D([0,2],\mathbb{R})$. Then $\epsilon^{-1}(x_\epsilon - x) = y \circ \lambda_\epsilon$, where $\lambda_\epsilon \in \Lambda$ with $\lambda_\epsilon(1) = 1 - \epsilon$, $\lambda_\epsilon(0) = 0$ and $\lambda_\epsilon(2) = 2$. Hence $d_{J_1}(\epsilon^{-1}(x_\epsilon - x), y) = \|\lambda_\epsilon - e\| = \epsilon \to 0$ as $\epsilon \downarrow 0$. However $\epsilon^{-1}x_\epsilon^{\uparrow}(2) = \epsilon^{-1}x_\epsilon^{\uparrow}(1-\epsilon) = \epsilon^{-1}$, while $(\epsilon^{-1}x + y)^{\uparrow}(2) = (\epsilon^{-1}x + y)(1-) = \epsilon^{-1} + 1$, so that $d_{M_2}(\epsilon^{-1}x_\epsilon, \epsilon^{-1}x + y) \geq 1$. $\blacksquare$

However, under regularity conditions, we can also establish results starting from J_1, M_1 and M_2 convergence. We state the following results for the strong SJ_1, SM_1 and SM_2 metrics on $D([0,t],\mathbb{R}^k)$. Corresponding results for the product metrics for Lemmas 9.2.1 and 9.2.2 below follow; just consider one coordinate at a time.

Recall that x is Lipschitz on $[0,t]$ if there is a constant K so that $|x(t_1) - x(t_2)| \leq K|t_1 - t_2|$ for $0 \leq t_1, t_2 \leq t$. This regularity condition is typically satisfied in applications, because x often satisfies an ordinary differential equation (ODE). If x is absolutely continuous with derivative $\dot{x}$, where $\dot{x} \in$

D, then for each $t > 0$, there exists K such that $|\dot{x}(s)| \le K$ for $0 \le s \le t$ and, for $0 \le t_1 < t_2 \le t$,

$$|x(t_2) - x(t_1)| \le \int_{t_1}^{t_2} |\dot{x}(s)| ds \le K|t_2 - t_1| \ , \tag{2.7}$$

so that x is Lipschitz.

Lemma 9.2.1. (subtracting a common Lipschitz function) *Suppose that x is Lipschitz in $[0,t]$ with Lipschitz constant K. If d_t is the SJ_1, SM_1 or SM_2 metric on $D([0,t], \mathbb{R}^k)$, then*

$$d_t(x_1 - x, x_2 - x) \le (1 + K) d_t(x_1, x_2) \ . \tag{2.8}$$

Proof. First consider J_1. For all $\epsilon > 0$, there exist $\eta(\epsilon) > 0$ and increasing homeomorphisms λ_ϵ of $[0,t]$ such that

$$\|x_1 - x_2 \circ \lambda_\epsilon\|_t \vee \|\lambda_\epsilon - e\|_t \le (1 + \eta(\epsilon)) d_t(x_1, x_2) \ .$$

It follows that

$$\begin{aligned} \|(x_1 - x) - (x_2 - x) \circ \lambda_\epsilon\|_t &\le \|x_1 - x_2 \circ \lambda_\epsilon\|_t + \|x - x \circ \lambda_\epsilon\|_t \\ &\le (1 + \eta(\epsilon)) d_t(x_1, x_2) + K\|\lambda_\epsilon - e\|_t \\ &\le (1 + \eta(\epsilon) + K[1 + \eta(\epsilon)]) d_t(x_1, x_2) \ . \end{aligned}$$

Since $\eta(\epsilon)$ can be made arbitrarily small, the proof for J_1 is complete. Now consider M_1. For all $\epsilon > 0$ and $t > 0$, there exist $\eta(\epsilon) > 0$ and parametric representations $(u_1, r_{1\epsilon})$ of x_1 and $(u_{2\epsilon}, r_{1\epsilon})$ of x_2 such that

$$\|u_{1\epsilon} - u_{2\epsilon}\| \vee \|r_{1\epsilon} - r_{2\epsilon}\| \le (1 + \eta(\epsilon)) d_t(x_1, x_2) \ .$$

Since x is continuous, $(x \circ r_{1\epsilon}, r_{1\epsilon})$ and $(x \circ r_{2\epsilon}, r_{2\epsilon})$ are parametric representations of x, $(u_{1\epsilon} - x \circ r_{1\epsilon}, r_{1\epsilon})$ and $(u_{2\epsilon} - x \circ r_{2\epsilon}, r_{2\epsilon})$ are parametric representations of $x_1 - x$ and $x_2 - x$, and

$$\begin{aligned} \|(u_{1\epsilon} - x \circ r_{1\epsilon}) - (u_{2\epsilon} - x \circ r_{2\epsilon})\| &\le \|u_{1\epsilon} - u_{2\epsilon}\| + \|x \circ r_{1\epsilon} - x \circ r_{2\epsilon}\| \\ &\le (1 + \eta(\epsilon)) d_t(x_1, x_2) + K\|r_{1\epsilon} - r_{2\epsilon}\| \\ &\le (1 + \eta(\epsilon) + K[1 + \eta(\epsilon)]) d_t(x_1, x_2) \ . \end{aligned}$$

Since $\eta(\epsilon)$ can be arbitrarily small, the proof for M_1 is complete. Now consider M_2. let $(z_1, t_1) \in \Gamma_{x_1}$. If $(z_2, t_2) \in \Gamma_{x_2}$ is such that $\|(z_1, t_1) - (z_2, t_2\| < \delta$, then $(z_1 - x(t_1), t_1) \in \Gamma_{x_1 - x}$, $(z_2 - x(t_2), t_2) \in \Gamma_{x_2 - x}$ and

$$\begin{aligned} \|(z_1 - x(t_1), t_1) \vee (z_2 - x(t_2), t_2)\| &\le \|(z_1, t_1) - (z_2, t_2)\| + \|x(t_1) - x(t_2)\| \\ &\le \delta + K\|t_1 - t_2\| \le (1 + K)\delta \ . \quad \blacksquare \end{aligned}$$

Next we generalize (2.2).

Lemma 9.2.2. (deterministic scaling) *Let d_t be the SJ_1, SM_1 or SM_2 metric on $D([0,t],\mathbb{R}^k)$. For any $c > 0$,*

$$d_{ct}(cx_1 \circ c^{-1}e, cx_2 \circ c^{-1}e) = cd_t(x_1, x_2) \tag{2.9}$$

or, equivalently,

$$d_t(cx_1, cx_2) = cd_{t/c}(x_1 \circ ce, x_2 \circ ce) \leq (c \vee 1)d_t(x_1, x_2)\ . \tag{2.10}$$

Proof. First, for SJ_1, note that $\lambda \in \Lambda_t$ if and only if $c\lambda \circ c^{-1}e \in \Lambda_{ct}$ for $c > 0$ and

$$\|c\lambda \circ c^{-1}e - e\|_{ct} = c\|\lambda - e\|_t\ .$$

Hence

$$\begin{aligned} &d_{ct}(cx_1 \circ c^{-1}e, cx_2 \circ c^{-1}e) \\ &\quad = \inf_{\lambda \in \Lambda_t} \{\|cx_1 \circ c^{-1}e - (cx_2 \circ c^{-1}e) \circ (c\lambda \circ c^{-1}e)\|_{ct} \vee \|c\lambda \circ c^{-1}e - e\|_{ct}\} \\ &\quad = \inf_{\lambda \in \Lambda_t} \{c\|x_1 - x_2 \circ \lambda\|_t \vee c\|\lambda - e\|_t\} \\ &\quad = cd_t(x_1, x_2)\ . \end{aligned}$$

Next, for SM_2, note that $c\Gamma_{x_i}$ is the graph of $cx_i \circ c^{-1}e$ over $[0, ct]$ if and only if Γ_{x_i} is the graph of x_i over $[0,t]$. Hence (2.9) holds. Finally, for SM_1, note that (cu_i, cr_i) is a parametric representation of $cx_i \circ c^{-1}e$ over $[0, ct]$ if and only if (u_i, r_i) is a parametric representation of x_i over $[0,t]$. Hence (2.9) holds. ∎

Our next result goes beyond Theorem 9.2.1 by allowing the map ϕ to be Lipschitz with respect to the SJ_1, SM_1 or SM_2 metrics, but not the uniform metric.

Theorem 9.2.3. (Lipschitz functions with respect to nonuniform metrics) *Suppose that $y \in D([0,t_1],\mathbb{R}^{k_1})$ and x, x_ϵ, $x + \epsilon y$ all belong to a subset A of $D([0,t_1],\mathbb{R}^{k_1})$ for sufficiently small $\epsilon > 0$. Suppose that $\phi : A \to D([0,t_2],\mathbb{R}^{k_2})$ is Lipschitz with respect to the metrics d_1 on A and d_2 on $D([0,t_2],\mathbb{R}^{k_2})$, i.e., there is a constant K such that*

$$d_2(\phi(x_1), \phi(x_2)) \leq K d_1(x_1, x_2) \tag{2.11}$$

for all $x_1, x_2 \in A$, where d_1 and d_2 are nonuniform Skorohod metrics (not necessarily the same). Suppose that x is Lipschitz on $[0,t_1]$ and $\phi(x)$ is Lipschitz on $[0,t_2]$. Then there is a constant K' such that

$$\begin{aligned} &d_2(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], \epsilon^{-1}[\phi(x + \epsilon y) - \phi(x)]) \\ &\qquad \leq K'\epsilon^{-1}d_1(x_\epsilon - x, \epsilon y) \\ &\qquad \leq K'\|\epsilon^{-1}(x_\epsilon - x) - y\|_{t_1}\ . \end{aligned} \tag{2.12}$$

Proof. By Lemmas 9.2.2 and 9.2.1 and the assumptions, for $\epsilon < 1$, there are constants K_1, K_2 and K_3 such that

$$\begin{aligned} d_2(\epsilon^{-1}[\phi(x_\epsilon) - \phi(x)], \epsilon^{-1}[\phi(x+\epsilon y) - \phi(x)]) \\ \le \epsilon^{-1} d_2(\phi(x_\epsilon) - \phi(x), \phi(x+\epsilon y) - \phi(x)) \\ \le K_1 \epsilon^{-1} d_2(\phi(x_\epsilon), \phi(x+\epsilon y)) \\ \le K_1 K_2 \epsilon^{-1} d_1(x_\epsilon, x+\epsilon y) \\ \le K_1 K_2 K_3 \epsilon^{-1} d_1(x_\epsilon - x, \epsilon y) \\ \le K_1 K_2 K_3 \epsilon^{-1} \|x_\epsilon - x - \epsilon y\|_{t_1} \\ \le K_1 K_2 K_3 \|\epsilon^{-1}(x_\epsilon - x) - y\|_{t_1} . \quad \blacksquare \end{aligned} \tag{2.13}$$

The final upper bound in Theorem 9.2.3 does not help with the supremum and reflection maps because the supremum and reflection maps already have the required Lipschitz properties with respect to the uniform metric, by Theorem 9.2.2. In order to apply Theorem 9.2.3 without having to resort to the cruder uniform metric bound, we need to have

$$d_1(x_\epsilon - x, \epsilon y) = o(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0 . \tag{2.14}$$

First, from this analysis, we see the need to be precise about what we mean about $o(\epsilon)$ terms in (1.12)–(1.15). Next, we observe that $d_1(\epsilon^{-1}[x_\epsilon - x], y) \to 0$ does not directly imply that $d_1(x_\epsilon - x, \epsilon y) = o(\epsilon)$ as $\epsilon \downarrow 0$, but that it is possible to have $d_1(x_\epsilon - x, \epsilon y) = o(\epsilon)$ as $\epsilon \downarrow 0$ without having $\|x_\epsilon - x - \epsilon y\|_{t_1} = o(\epsilon)$ as $\epsilon \downarrow 0$.

Example 9.2.3. *Condition* (2.14) *is weaker than the usual limit.* We would like to have $\epsilon^{-1} d_t(x_\epsilon - x, \epsilon y) \to 0$ as $\epsilon \downarrow 0$ whenever $d_t(\epsilon^{-1}(x_\epsilon - x), y) \to 0$ as $\epsilon \downarrow 0$, so that we could improve upon (2.12), but that implication is not valid. To see that, let $x = y = I_{[1,2]}$ in $D([0,2], \mathbb{R})$ and let $x_\epsilon = x + \epsilon I_{[1+\delta_\epsilon, 2]}$. Then $\epsilon^{-1}(x_\epsilon - x) = I_{[1+\delta_\epsilon, 2]}$ and $d_{J_1}(\epsilon^{-1}(x_\epsilon - x), y) = \delta_\epsilon$. On the other hand $\epsilon^{-1} d_{J_1}(x_\epsilon - x, \epsilon y) = \epsilon^{-1}(\epsilon \wedge \delta_\epsilon)$, which converges to 0 if and only if $\epsilon^{-1}\delta_\epsilon \to 0$ as $\epsilon \to 0$. Hence, we do not necessarily have $\epsilon^{-1} d_t(x_\epsilon - x, \epsilon y) \to 0$ as $\epsilon \downarrow 0$, given $d_t(\epsilon^{-1}(x_\epsilon - x, y) \to 0$, but we could have it, as is the case here when $\epsilon^{-1}\delta_\epsilon \to 0$ as $\epsilon \downarrow 0$. On the other hand, $\|\epsilon^{-1}(x_\epsilon - x) - y\| = 1$ for all $\epsilon > 0$. $\blacksquare$

Example 9.2.4. *A parametric family of examples.* Consider Example 9.2.2 modified by having

$$x_\epsilon = (x + \epsilon) I_{[0, 1-\epsilon^p)} + (x - \epsilon) I_{[1-\epsilon^p, 2]} \ . \tag{2.15}$$

Then $x_\epsilon - x = \epsilon y \circ \lambda_\epsilon$ for $\lambda_\epsilon(1) = 1 - \epsilon^p$, $\lambda_\epsilon(0) = 0$ and $\lambda_\epsilon(2) = 2$ with λ_ϵ defined by linear interpolation elsewhere. Thus

$$d_{J_1}(x_\epsilon - x, \epsilon y) = d_{J_1}(\epsilon^{-1}(x_\epsilon - x), y) = \|\lambda_\epsilon - e\| = \epsilon^p \ , \tag{2.16}$$

so that condition (2.14) holds if $p > 1$, but not if $0 < p \leq 1$.

9.3. Derivative of the Supremum Function

In this section we consider the derivative of the supremum function; i.e., we find conditions under which the limit (1.15) is valid and identify the limit z when $\phi : D \to D$ is the supremum function. The supremum function maps $x \in D \equiv D([0,T], \mathbb{R})$ into $x^\uparrow \in D$ for

$$x^\uparrow(t) \equiv \sup_{0 \leq s \leq t} x(s), \quad 0 \leq t \leq T \, . \tag{3.1}$$

In order to treat the derivatives, we will find it necessary to consider functions outside of D. Thus let D_{lim} be the set of functions with left and right limits everywhere, but without having to be either left continuous or right continuous at each discontinuity point. In general, we will only be able to conclude (in Theorem 9.3.2 below) that the derivative belongs to D_{lim}. In our definition of the derivative, we start by allowing one function to be in D_{lim}. For $x \in D$, $y \in D_{lim}$ and $\epsilon > 0$, let

$$z_\epsilon \equiv z_\epsilon(x,y) \equiv \epsilon^{-1}[(x + \epsilon y)^\uparrow - x^\uparrow] = (\epsilon^{-1}x + y)^\uparrow - \epsilon^{-1}x^\uparrow \, . \tag{3.2}$$

The derivative of the supremum function (in the direction y, evaluated at x) is the limit of z_ϵ as $\epsilon \downarrow 0$, if it exists. We will show that the limit does exist under regularity conditions and identify it. In this section we consider pointwise convergence for all t; in the next section we consider M_1 and M_2 convergence.

We start by stating two elementary lemmas; the second follows from the first.

Lemma 9.3.1. (the case of constant y) *If $y(s) = c$, $0 \leq s \leq t$, then $z_\epsilon(t) = c$ for all ϵ.*

For $z^\downarrow$ be the infimum function; i.e., $z^\downarrow = -(-z)^\uparrow$.

Lemma 9.3.2. (monotone bounds) *For all $\epsilon > 0$, $y^\downarrow \leq z_\epsilon \leq y^\uparrow$.*

Even though x is right-continuous, it can approach its supremum from the left $(x(s) = sI_{[0,t_0)}(s))$ or right $(x(s) = -sI_{[t_0,t_1)}(s))$. Let $\Phi_x^L(t)$ and $\Phi_x^R(t)$ be the subsets of time points in $[0,t]$ at which the left and right limits of x attain the supremum; i.e.,

$$\Phi_x^L(t) = \{s : 0 < s \le t, x(s-) = x^\uparrow(t)\} \tag{3.3}$$

and

$$\Phi_x^R(t) = \{s : 0 \le s \le t, x(s+) = x^\uparrow(t)\} \ . \tag{3.4}$$

Let $\Phi_x(t) = \Phi_x^L(t) \cup \Phi_x^R(t)$. When $x \in C$, $\Phi_x^L(t) = \Phi_x^R(t)$.

Example 9.3.1. *The possibility of empty sets.* It is possible for Φ_x^L or $\Phi_x^R(t)$ to be empty: Let $x(t) = tI_{[0,1)}(t)$, $t \ge 0$. Then, for $t \ge 1$, $\Phi_x^L(t) = \{1\}$, while $\Phi_x^R(t) = \phi$. However, $\Phi_x^L(t) \cup \Phi_x^R(t) \ne \phi$. ■

These subsets need not be closed, but they have the following partial closure property.

Lemma 9.3.3. (partial closure property) *For any $x \in D$ and $t \ge 0$, $\Phi_x^L(t)$ is closed from the left, while $\Phi_x^R(t)$ is closed from the right; i.e., if $s_n \uparrow s$ in $[0,t]$ and $s_n \in \Phi_x^L(t)$ for all n, then $s \in \Phi_x^L(t)$; if $s_n \downarrow s$ and $s_n \in \Phi_x^R(t)$ for all n, then $s \in \Phi_x^R(t)$. Moreover, if $s_n \uparrow s$ in $[0,t]$ and $s_n \in \Phi_x^R(t)$ for all n, then $s \in \Phi_x^L(t)$; if $s_n \downarrow s$ in $[0,t]$ and $s_n \in \Phi_x^L(t)$ for all n, then $s \in \Phi_x^R(t)$.*

Corollary 9.3.1. (compactness of $\Phi_x(t)$) *For each $t > 0$, $\Phi_x(t)$ is a compact subset of $[0,t]$.*

We next show that z_ϵ is monotone in ϵ.

Lemma 9.3.4. (monotonicity in ϵ) *For z_ϵ in (3.2), $z_\epsilon(t)$ decreases as ϵ decreases for each t.*

Proof. We want to show that

$$(\epsilon_2^{-1}x + y)^\uparrow - \epsilon_2^{-1}x^\uparrow < (\epsilon_1^{-1}x + y)^\uparrow - \epsilon_1^{-1}x^\uparrow$$

for $\epsilon_1 > \epsilon_2$ or, equivalently,

$$(\epsilon_2^{-1}x + y)^\uparrow - (\epsilon_1^{-1}x + y)^\uparrow < (\epsilon_2^{-1} - \epsilon_1^{-1})x^\uparrow \ . \tag{3.5}$$

However, (3.5) follows from the relation

$$x_1^\uparrow - x_2^\uparrow \le (x_1 - x_2)^\uparrow \ . \quad ■$$

We first establish pointwise convergence for z_ϵ in (3.2).

Theorem 9.3.1. (pointwise convergence) *For each* $x \in D, y \in D_{lim}$ *and* $t \geq 0$,

$$\lim_{\epsilon \downarrow 0} z_\epsilon(t) = z(t) \equiv \sup_{s \in \Phi_x^L(t)} y(s-) \vee \sup_{s \in \Phi_x^R(t)} \{y(s), y(s+)\} \ . \tag{3.6}$$

Proof. The convergence follows from the monotonicity established in Lemma 9.3.3. Lemma 9.3.2 above provides a lower bound, which implies that there is a proper limit for each t. For any $\delta > 0$, let $s_\epsilon(t)$ be a point in $[0, t]$ such that

$$(\epsilon^{-1}x + y)(s_\epsilon(t)) \geq (\epsilon^{-1}x + y)^{\uparrow}(t) - \delta \ . \tag{3.7}$$

(Since x and y need not be continuous, the supremum of $\epsilon^{-1}x + y$ need not be attained.) Then

$$\begin{aligned} y(s_\epsilon(t)) &\geq y(s_\epsilon(t)) + \epsilon^{-1}\{x[s_\epsilon(t)] - x^{\uparrow}(t)\} \\ &\geq y(s) + \epsilon^{-1}[x(s) - x^{\uparrow}(t)] - \delta \quad \text{for} \quad 0 \leq s \leq t \\ &\geq \begin{cases} y(s-) - \delta & \text{for} \quad s \in \Phi_x^L(t) \\ y(s) - \delta & \text{for} \quad s \in \Phi_x^R(t) \\ y(s+) - \delta & \text{for} \quad s \in \Phi_x^R(t) \ , \end{cases} \end{aligned} \tag{3.8}$$

implying that

$$\varliminf_{n \to \infty} y(s_\epsilon(t)) \geq z(t), \quad t \geq 0 \ . \tag{3.9}$$

We now verify that

$$\varlimsup_{n \to \infty} y(s_\epsilon(t)) \leq z(t), \quad t \geq 0 \ . \tag{3.10}$$

Start by choosing $\{s_\epsilon(t)\}$ such that $y(s_\epsilon(t)) \to \varlimsup_{n \to \infty} y(s_\epsilon(t))$ as $\epsilon \downarrow 0$. Since $s_\epsilon(t) \in [0, t]$, any subsequence from $\{s_\epsilon(t)\}$ has a convergent subsequence $\{s_{\epsilon'}(t)\}$ as $\epsilon' \downarrow 0$. (Let $\epsilon' \downarrow 0$ through countably many values.) So suppose that $s_{\epsilon'}(t) \to s_0(t)$ as $\epsilon' \downarrow 0$. Without loss of generality, by taking a further subsequence if necessary, we can assume that either $s_{\epsilon'}(t) \uparrow s_0(t)$ with $s_{\epsilon'}(t) < s_0(t)$ for all $\epsilon' > 0$ or $s_{\epsilon'}(t) \downarrow s_0(t)$ with $s_{\epsilon'}(t) \geq s_0(t)$ for all $\epsilon' > 0$. Suppose that $s_{\epsilon'}(t) \uparrow s_0(t)$. Then $y(s_{\epsilon'}(t)) \to y(s_0(t)-)$. We can deduce from (3.8) that there is a constant K such that, for all ϵ',

$$-K \leq \epsilon^{-1}[x(s_{\epsilon'}(t)) - x^{\uparrow}(t)] \leq 0 \ , \tag{3.11}$$

implying that $x(s_{\epsilon'}(t)) \to x^{\uparrow}(t)$ as $\epsilon' \to 0$, so that $x(s_0(t)-) = x^{\uparrow}(t)$ and $s_0(t) \in \Phi_x^L(t)$. By this argument,

$$\overline{\lim_{n\to\infty}}\ y(s_\epsilon(t)) \le \sup_{s\in\Phi_x^L(t)} y(s-)\ . \tag{3.12}$$

On the other hand, if $s_{\epsilon'}(t) \downarrow s_0(t)$, we can deduce by the same reasoning that

$$\overline{\lim_{n\to\infty}}\ y(s_\epsilon(t)) \le \sup_{s\in\Phi_x^R(t)} \{y(s), y(s+)\}\ . \tag{3.13}$$

Since one of (3.12) or (3.13) must hold, we have established (3.10). Finally, from the first and last lines of (3.8),

$$0 \ge \epsilon^{-1}\{x(s_\epsilon(t)) - x^{\uparrow}(t)\} \ge z(t) - y(s_\epsilon(t))\ . \tag{3.14}$$

Since $y(s_\epsilon(t)) \to z(t)$, $\epsilon^{-1}\{x(s_\epsilon(t)) - x^{\uparrow}(t)\} \to 0$ as $\epsilon \downarrow 0$, which implies that $z_\epsilon(t) \to z(t)$ as $\epsilon \downarrow 0$. ■

Corollary 9.3.2. (simplification under extra conditions) *Suppose that $x \in C$ and $y \in D_{lim}$. Then the limit z in* (3.6) *is*

$$z(t) = \sup_{s\in\Phi_x(t)} \{y(s-), y(s), y(s+)\}\,. \tag{3.15}$$

If, in addition, $y \in C$, then

$$z(t) = \sup_{s\in\Phi_x(t)} \{y(s)\}\ . \tag{3.16}$$

We now determine the structure of the limit function z in (3.6). Since $\Phi_x^L(t)$, $\Phi_x^R(t)$ and $\Phi_x(t)$ are subsets of $[0,t]$, we need a notion of convergence of sets. For subsets A_n and A of $\mathbb{R}$, we say that $A_n \to A$ if (i) for all $a_n \in A_n$, $n \ge 1$, $\{a_n\}$ has a convergent subsequence and the limits of all convergent subsequences belong to A, and (ii) for all $a \in A$, there exists $a_n \in A_n$, $n \ge 1$, such that $a_n \to a$ as $n \to \infty$. In our set limits involving $\Phi_x^L(t)$ and $\Phi_x^R(t)$, only three special cases arise: (i) A_n is independent of n for all sufficiently large n, (ii) the sequence $\{A_n\}$ is eventually monotone, i.e., either $A_n \subseteq A_{n+1}$ for all sufficiently large n or $A_n \supset A_{n+1}$ for all sufficiently large n, and (iii) $A = \{a\}$, i.e., the limit set contains a single point.

When we consider $\Phi_x(t) \equiv \Phi_x^L(t) \cup \Phi_x^R(t)$, we have compact subsets of $[0,t]$. Then the notion of set convergence above is induced by the Hausdorff metric on the space $\mathcal{C} \equiv \mathcal{C}([0,\infty))$ of compact subsets of $[0,\infty)$, defined in (2.8) in Chapter V.

However, even if x and $x^\uparrow$ are continuous in t, $\Phi_x(t)$ is in general *not* continuous in t. Moreover, at some time points, $\Phi_x(t)$ is neither left-continuous nor right-continuous.

Example 9.3.2. *Lack of continuity from left or right in $\Phi_x(t)$.* Suppose that $x(t) = (1-t)I_{[0,1)}(t) + (t-1)I_{[1,\infty)}(t)$. Then $\Phi_x(t) = \{0\}$, $0 \le t < 2$, $\Phi_x(2) = \{0,1\}$ and $\Phi_x(t-1) = \{t\}$, $t > 2$, so that Φ_x is neither left-continuous nor right-continuous at $t = 2$. However, $\Phi_x(2)$ is the union of the left and right limits $\Phi_x(2-)$ and $\Phi_x(2+)$. ∎

Example 9.3.3. *Neither left-continuous everywhere nor right-continuous everywhere.* We can extend Example 9.3.2 to show that the limit z need not be either a left-continuous function or a right-continuous function, even if x and y are both continuous. Let

$$x(t) = (1-t)I_{[0,1)}(t) + (t-1)I_{[1,3)}(t) + (5-t)I_{[3,4)}(t) + (t-3)I_{[4,\infty)}(t) \quad (3.17)$$

and

$$y(t) = -tI_{[0,2.5]} + 6(t-2.5)I_{[2.5,\infty)}(t)\,. \quad (3.18)$$

Then

$$\begin{aligned} \Phi_x(t) &= \{0\}, \quad 0 \le t < 2, \quad \Phi_x(2) = \{0,2\} \\ \Phi_x(t) &= \{t\}, \quad 2 < t \le 3, \quad \Phi_x(t) = \{3\}, \quad 3 \le t < 5\,, \\ \Phi_x(5) &= \{3,5\}, \quad \Phi_x(t) = \{t\}, \quad t > 5\,, \\ z(2) &= 0 \quad \text{and} \quad z(5) = 15\,. \end{aligned} \quad (3.19)$$

Then z is discontinuous at $t = 2$ and $t = 5$, with z being left-continuous at 2 and right-continuous at 5. Hence z is neither left-continuous everywhere nor right-continuous everywhere. On the positive side, z is either left-continuous or right-continuous at each t and z is upper semicontinuous everywhere. ∎

Example 9.3.4. *Neither left-continuous nor right-continuous at one t.* We now show that the limit z in (3.8) need not be either left-continuous or right-continuous at a single argument t when $x \in C$ and $y \in D$ but $y \notin C$. We construct y and x so that y and Φ_x have only one common discontinuity. Let

$$y(t) = tI_{[0,1)}(t) + I_{[1,2)}(t), \quad t \ge 0\,, \quad (3.20)$$

and

$$x(t) = -tI_{[0,1)}(t) + (t-2)I_{[1,\infty)}(t), \quad t \ge 0\,, \quad (3.21)$$

so that

$$\Phi_x(t) = \{0\}, 0 \le t < 2, \Phi_x(2) = \{0, 2\} \quad \text{and} \quad \Phi_x(t) = t, \quad t > 2 \ . \qquad (3.22)$$

Hence y and Φ_x are continuous everywhere except $t = 2$. Moreover,

$$z(2) = \sup_{s \in \{0,2\}} \{y(s)\} \vee \sup_{s \in \{2\}} \{y(s-)\} = 0 \vee 1 = 1 \ , \qquad (3.23)$$

while $z(t) = 0$ for all other t. Hence the left and right limits coincide at $t = 2$ but do not equal $z(2)$, so that $z \notin D$. It is easy to see that $z_\epsilon(2) = 1$ and

$$z_\epsilon(t) = 0, \quad 0 \le t \le 2 - \epsilon \quad \text{and} \quad t \ge 2 + \epsilon \ ,$$

with z_ϵ defined by linear interpolation elsewhere. Hence, z_ϵ has slope ϵ^{-1} on $[2 - \epsilon, 2]$, slope $-\epsilon^{-1}$ on $[2, 2 + \epsilon]$ and is 0 elsewhere. Consistent with Theorem 9.3.1, z_ϵ converges pointwise to z. We will want to impose regularity conditions to prevent such pathological behavior. As an alternative, we could conclude that z_ϵ converges to a limit in one of the larger spaces E or F in Chapter X. ■

We now introduce a regularity condition under which the limit z in (3.6) has left and right limits everywhere and is either left continuous or right continuous everywhere (without necessarily being right continuous everywhere). Let $D_{l,r}$ denote this space. We first define some subsets of $[0, \infty)$. (We could alternatively restrict attention to a subinterval $[0, T]$.) For any $x \in D$, let $Rinc(x)$ and $Linc(x)$ be the set of right-increase and left-increase points of x, let $Lconst(x)$ be the set of left-constant points of x, and let $Amax(x)$ be the argmax set of x, i.e., the set of arguments at which x equals its supremum, i.e.,

$$Rinc(x) \equiv \{t \ge 0 : x(t) < x(t + \epsilon) \quad \text{for all sufficiently small} \quad \epsilon > 0\} \qquad (3.24)$$

$$Linc(x) \equiv \{t \ge 0 : x(t - \epsilon) < x(t) \quad \text{for all sufficiently small} \quad \epsilon > 0\} \qquad (3.25)$$

$$Lconst(x) \equiv \{t \ge 0 : x(t - \epsilon) = x(t) \quad \text{for all sufficiently small} \quad \epsilon > 0\} \qquad (3.26)$$

$$Amax(x) \equiv \{t \ge 0 : t \in \Phi_x^R(t)\} \ . \qquad (3.27)$$

We will look at these sets for the functions x and $x^\uparrow$. Of course, $x^\uparrow$ is nondecreasing and right-continuous. Let $Disc(x)$ be the set of discontinuity points of x.

Theorem 9.3.2. (regularity properties of the limit z) *Suppose that x, $y \in D$. Then $z \in D_{lim}$, where z is the limit in* (3.6). *At all t not in the set*

$$Bad(x) \equiv Rinc(x^\uparrow) \cap Lconst(x^\uparrow) \cap Disc(x)^c \cap Linc(x) \cap Amax(x) \ , \qquad (3.28)$$

z is either left-continuous or right-continuous. For $t \in Bad(x)$, $z(t+) = y(t)$, $z(t-)$ is independent of $\{y(t-), y(t)\}$ and $z(t) = \max\{z(t-), y(t-), y(t)\}$, so that z is left-continuous at t if $z(t-) \geq y(t-) \vee y(t)$, right-continuous at t if $y(t) \geq y(t-) \vee z(t-)$, and neither left-continuous nor right-continuous if $y(t-) > y(t) \vee z(t-)$. If

$$y(t-) \leq z(t-) \vee y(t) \tag{3.29}$$

for all $t \in Bad(x)$, for which a sufficient condition is

$$Disc(y) \cap Bad(x) = \phi \ , \tag{3.30}$$

then z is either left-continuous or right-continuous at all t, so that $z \in D_{l,r}$.

Corollary 9.3.3. (regularity for continuous y) *If $x \in D$ and $y \in C$, then $z \in D_{l,r}$.*

Remark 9.3.1. *Sufficient condition for having more than one point in the set.* Let $|\Phi_x(t)|$ be the cardinality of the set $\Phi_x(t)$. Note that $|\Phi_x(t)| \geq 2$ when $t \in Lconst(x^{\uparrow}) \cap Amax(x)$, i.e.,

$$Lconst(x^{\uparrow}) \cap Amax(x) \subseteq \{t : |\Phi_x(t)| \geq 2\} \ , \tag{3.31}$$

so that $t \in Bad(x)$ when $|\Phi_x(t)| \geq 2$ and $x(t-\epsilon) < x(t-) = x(t) = x^{\uparrow}(t) < x^{\uparrow}(t+\epsilon)$ for all suitably small $\epsilon > 0$. ■

Remark 9.3.2. *The set $Bad(x)$ is at most countably infinite.* From (3.28), it follows that $Bad(x) \subseteq Disc(\Phi_x)$, where $\Phi_x \in D([0,\infty), (\mathcal{C}, h))$. Therefore, $Bad(x)$ is a countable set. ■

Corollary 9.3.4. (regularity properties of the limit Z when Y is a stochastic process) *Suppose that $\{Y(t) : t \geq 0\}$ is a stochastic process with sample paths in D. If $x \in D$ and if $P(t \in Disc(Y)) = 0$ for each $t > 0$, then $P(Z \in D_{l,r}) = 1$, where Z is the limiting stochastic process defined by applying* (3.6) *to Y.*

Proof. In Remark 9.3.2 it was noted that the set $Bad(x)$ in (3.28) is countable. Consequently,

$$P(Z \in D_{l,r}) = P(Disc(Y) \cap Bad(x) = \phi) = 1 \ . \quad ■ \tag{3.32}$$

Theorem 9.3.2 is proved by examining all relevant cases. We identify appropriate cases and results for those cases in the following theorem.

Theorem 9.3.3. (identification of relevant cases) *The following is a set of exhaustive and mutually exclusive cases and subcases when $x, y \in D$:*

1. $t \notin Amax(x)$, *i.e.*, $t \notin \Phi_x^R(t)$: *z is right-continuous with a left limit at t.*

2. $\Phi_x^R(t) = \Phi_x^L(t) = \{t\} : z(t) = y(t-) \vee y(t)$, *$z$ is either right-continuous or left-continuous at t.*

3. $\Phi_x^R(t) = \{t\}$, $\Phi_x^L(t) = \phi$: *z is right-continuous with a left limit at t.*

4. $t \in \Phi_x^R(t) \subseteq \Phi_x(t) \neq \{t\}$, *so that cases 1–3 do not hold;*

   (a) $t \notin Rinc(x^{\uparrow})$, *i.e.*, $\Phi_x(t) \subseteq \Phi_x(u)$ *for some* $u > t$: *z is right-continuous with a left limit at t.*

   (b) *Condition* (a) *does not hold and* $t \in Lconst(x^{\uparrow}) \cap Linc(x)^c$, *i.e., t is not isolated in* $\Phi_x(t)$: $z(t-) \geq y(t-)$ *and* $z(t+) = y(t)$, *so that z is left (right) continuous at t if* $z(t-) \geq (\leq)\ y(t)$.

   (c) *Condition* (a) *does not hold, t is isolated in* $\Phi_x(t)$ *and* $t \in Disc(x)$: $z(t+) = y(t)$ *and* $z(t) = \max\{z(t-), y(t)\}$, *so that z is left (right) continuous if* $z(t-) \geq (\leq)\ y(t)$. *(In this case $z(t)$ does not depend upon $y(t-)$.)*

   (d) *Condition* (a) *does not hold, t is isolated in* $\Phi_x(t)$ *and* $t \notin Disc(x)$, *i.e.,* $t \in Bad(x)$ *in* (3.28): $z(t+) = y(t)$, $z(t-)$ *is independent of* $\{y(t-), y(t)\}$ *and* $z(t) = \max\{z(t-), y(t-), y(t)\}$. *Hence z is neither left-continuous nor right-continuous at t if and only if* $y(t-) > z(t-) \vee y(t)$.

Proof. We prove Theorem 9.3.3 by examining all relevant subcases. We provide a further characterization below, but do not give all details. For this purpose, let

$$\Psi_x^L(t) = \{s : 0 < s \leq t, x(s-) = x^{\uparrow}(t-)\}\ , \tag{3.33}$$

$$\Psi_x^R(t) = \{s : 0 \leq s \leq t, x(s) = x^{\uparrow}(t-)\} \tag{3.34}$$

and $\Psi_x(t) = \Psi_x^L(t) \cup \Psi_x^R(t)$.

Case 1: In this case, $x(t) < x^{\uparrow}(t)$ and $x^{\uparrow}(t-) = x^{\uparrow}(t)$. Since x and $x^{\uparrow}$ are right-continuous, Φ_x^L and Φ_x^R are constant in $[t, t+\epsilon)$ for all suitably small $\epsilon > 0$, so that z is necessarily right-continuous. We identify three subcases:

(i) If $t \notin \Psi_x^L(t)$, then $x(t-) < x^{\uparrow}(t-)$, so that Φ_x^L and Φ_x^R are constant in $(t-\epsilon, t+\epsilon)$ for all suitably small $\epsilon > 0$, so that z is constant in the same subinterval. (ii) If $t \in \Psi_x^L(t) = \Phi_x^L(t)$ and $\Phi_x(t) \neq \{t\}$, then x jumps down at time t, so that $x(t-) = x^{\uparrow}(t-) = x^{\uparrow}(t) > x(t)$. Since $\Phi_x(t) \neq \{t\}$, $x^{\uparrow}$ must be constant in $(t-\epsilon, t]$ for all suitably small $\epsilon > 0$ and there must exist $s < t$ such that $x(s) = x^{\uparrow}(t)$ or $x(s-) = x^{\uparrow}(t)$. Hence for $s < t' < t$, $\Phi_x^L(t')$ and $\Phi_x^R(t')$ increase as t' increases. Since $t \notin \Phi_x^R(t)$, $\Phi_x^R(t') \uparrow \Phi_x^R(t)$ as $t' \uparrow t$, so that Φ_x^R is continuous at t. Since $\Phi_x^L(t')$ increases as t' increases, $\Phi_x^L(t')$ has a limit as $t' \uparrow t$, but this limit set may be separated from $t \in \Phi_x^L(t)$. Hence, in general z is right-continuous with a left limit at t, with $z(t)$ depending upon $y(t-)$ but not $y(t)$. In this case z is continuous at t if and only if $z(t-) \geq y(t-)$. (iii) If $t \in \Psi_x^L(t) = \Phi_x^L(t)$ and $\Phi_x(t) = \{t\}$, then again x jumps down at time t, $x(t-) = x^{\uparrow}(t-)$. Since $x^{\uparrow}$ is increasing from the left at t, there exists a sequence $\{t_n\}$ with $t_n \uparrow t$ as $n \to \infty$ such that $x(t_n\pm) = x^{\uparrow}(t_n)$ and $\Phi_x(t_n) = \{t_n\}$. Moreover, for any s with $t_n < s < t$, necessarily $\Phi_x(s) \subseteq [t_n, s]$. Hence, $\Phi_x(s) \to \Phi_x(t)$ as $s \uparrow t$. This implies that z is continuous at t with $z(t) = y(t-)$. We remark that the case $t \in \Psi_x^L(t)$ but $t \notin \Phi_x^L(t)$ cannot occur because it requires $x(t-) = x^{\uparrow}(t-) < x^{\uparrow}(t)$, which implies that x make a jump up to a new maximum at time t, i.e., $t \in \Phi_x^R(t)$, which contradicts our original assumption.

Case 2: $\Phi_x^R(t) = \Phi_x^L(t) = \{t\}$.

In this case $x(t-) = x(t) = x^{\uparrow}(t)$, so that x is continuous at t. Since $\Phi_x(t) = \{t\}$, $\Phi_x(u) \subseteq [t, u]$ for all $u > t$. Hence $\Phi_x(u) \to \Phi_x(t) = \{t\}$ as $u \downarrow t$, so that Φ_x is right-continuous and z has a limit from the right with $z(t+) = y(t)$. In this case $x^{\uparrow}$ is increasing at t, and $\Phi_x(s) \to \Phi_x^L(t)$ as $s \uparrow t$, so that Φ_x is continuous at t and z has the left limit $z(t-) = y(t-)$. Since $z(t) = y(t) \vee y(t-)$, z is either left-continuous or right-continuous at t; z is continuous at time t if and only if y is.

Case 3: $\Phi_x^R(t) = \{t\}$ and $\Phi_x^L(t) = \phi$.

In this case $x(t-) \neq x(t) = x^{\uparrow}(t)$, so that x is discontinuous at t. As in case 2 above, $\Phi_x(s) \to \Phi_x(t) = \{t\}$ as $s \downarrow t$, so that Φ_x is right-continuous at t and z has the right limit $z(t+) = y(t)$. Since $z(t) = y(t)$, z is right-continuous in this case. We identify three subcases: (i) If $t \notin \Psi_x^L(t)$, then $x(t-) < x^{\uparrow}(t-) < x^{\uparrow}(t)$, so that x jumps up to a new maximum at time t and Φ_x^L and Φ_x^R are constant in $(t-\epsilon, t)$ for all suitably small ϵ. Hence Φ_x^L, Φ_x^R and z have limits from the left, but may be discontinuous at t. (ii) If $\Psi_x(t) = \{t\}$, then $x(t-) = x^{\uparrow}(t-) < x^{\uparrow}(t)$. As in (ii), x jumps up to a new maximum at t. Since $\Phi_x(t) = \{t\}$, $x^\uparrow$ is increasing from the left at t. Hence, there exists a sequence $\{t_n\}$ with $t_n \uparrow t$ as $n \to \infty$ such that $x(t_n\pm) = x^\uparrow(t_n) \uparrow x^\uparrow(t-)$ and $\Phi_x(t_n) = \{t_n\}$. Hence $\Phi_x(s) \subseteq [t_n, s]$ for all s with $t_n < s < t$. Hence, $\Phi_x(s) \to \Psi_x^L(t) = \{t\}$ as $s \uparrow t$, so that Φ_x and z have limits from the left at t, with $z(t-) = y(t-)$. (iii) Suppose that $\Phi_x^L(t) = \phi$ and $t \in \Psi_x(t) \neq \{t\}$. This is similar to case (ii). Since $\Psi_x(t) \neq \{t\}$, $x^\uparrow$ is constant in $[t-\epsilon, t)$ for all suitably small ϵ. Thus, over $(t-\epsilon, t)$, $\Phi_x^L(s)$ and $\Phi_x^R(s)$ increase to $\Psi_x^L(t)$ and $\Psi_x^R(t)$ as $s \uparrow t$. Hence, z has a left limit at t. In general, z need not be continuous at t.

Case 4(a): In this case $x^\uparrow(t) = x^\uparrow(u)$ for some $u > t$. Hence $\Phi_x^L(u) \downarrow \Phi_x^L(t)$ and $\Phi_x^R(u) \downarrow \Phi_x^R(t)$ as $u \downarrow t$ so that z is right-continuous at t. If t is not isolated in $\Phi_x(t)$, as in Case 4(b), then there exists $t_n \uparrow t$ with $x(t_n-) = x^\uparrow(t)$ or $x(t_n) = x^\uparrow(t)$, so that $x^\uparrow$ is constant in $[t-\epsilon, t]$ for all suitably small ϵ. Moreover, $\Phi_x^L(s) \uparrow \Phi_x^L(t)$ and $\Phi_x^R(s) \uparrow \Phi_x^R(t)$ as $s \uparrow t$. Hence z has a left limit $z(t-) \geq y(t-)$. Moreover, Φ_x^L and Φ_x^R are continuous at t. If $y(t-) \leq z(t-) < y(t)$, then y is right-continuous but not continuous. On the other hand, if $z(t-) \geq y(t)$, then z is continuous at t. If instead t is isolated in $\Phi_x(t)$, as in Case 4(c), then $\Phi_x^L(s)$ and $\Phi_x^R(s)$ are constant in $(t-\epsilon, t)$ for all suitably small ϵ, but $\Phi_x^R(t) = \Phi_x^R(t-) \cup \{t\}$. Hence, Φ_x^L and Φ_x^R have limits from the left at t. Thus z has a limit from the left at t, which does not depend on $y(t-)$. If $z(t-) < y(t)$, then z is discontinuous at t; otherwise it is continuous.

Case 4(b): As in case 4(a), z has a left limit at t. If Case 4(a) does not hold, then $x^\uparrow(t) < x^\uparrow(t+\epsilon)$ for all sufficiently small ϵ. In this case, $\Phi_x(s) \to \{t\}$ as $s \downarrow t$, so that Φ_x and z have limits from the right with $z(t+) = y(t)$. However, since $\Phi_x(t) \neq \{t\}$ by assumption, Φ_x is not right-continuous. In this case z is left (right) continuous if $z(t-) \geq (\leq)\ y(t)$.

Case 4(c): In this case

$$t \in OK(x) \equiv Rinc(x^\uparrow) \cap Lconst(x^\uparrow) \cap Disc(x) \cap Linc(x) \cap Amax(x) \ . \quad (3.35)$$

Note that $OK(x)$ in (3.35) differs from $Bad(x)$ in (3.28) only by having $x(t-) < x(t)$. As noted for case 4(a) and 4(b), z has left limit $z(t-)$ and right limit $z(t+) = y(t)$ at t, with $z(t) = z(t-) \vee y(t)$. However, since $x(t-) < x(t) = x^\uparrow(t)$, $t \notin \Phi_x^L(t)$, so that $z(t)$ does not depend upon $y(t-)$. Hence z is either left-continuous or right-continuous at t.

Case 4(d): In this case $t \in Bad(x)$. Since $x(t-) = x(t) = x^{\uparrow}(t)$, $t \in \Phi_x^L(t)$ and $z(t) \geq y(t-)$. As in Case 4(c), z has left and right limits at t with $z(t+) = y(t)$ and $z(t) = \max\{z(t-), y(t-), y(t)\}$. ∎

Theorem 9.3.2 concluded that $z \in D_{lim}$ when $x, y \in D$. By the same reasoning, examining the cases in Theorem 9.3.3, we can obtain the same conclusion when $y \in D_{lim}$.

Theorem 9.3.4. (extension when $y \in D_{lim}$) *Suppose that $x \in D$ and $y \in D_{lim}$. Then $z \in D_{lim}$. At all t not in the set*

$$Bad(x, y) = [Bad_1(x) \cap Disc(y)] \cup Bad_2(y) \ , \qquad (3.36)$$

where

$$Bad_1(x) \equiv Rinc(x^{\uparrow}) \cap Lconst(x^{\uparrow}) \cap Linc(x) \cap Amax(x) \qquad (3.37)$$

and

$$Bad_2(y) \equiv \{t \in [0, T] : y(t) > y(t-), y(t+)\} \ , \qquad (3.38)$$

z is either left-continuous or right-continuous. At $t \in Bad(x) \cap Disc(x)$, $z(t+) = y(t+)$, $z(t-)$ is independent of $y(t-)$ and $z(t) = z(t-) \vee y(t) \vee y(t+)$, so that z is left-continuous if $z(t-) \geq y(t) \vee y(t+)$, right-continuous if $y(t+) \geq y(t) \vee z(t-)$ and neither right-continuous nor left-continuous if $y(t) > z(t-) \vee y(t+)$. At $t \in Bad(x) \cap Disc(x)^c$, $z(t+) = y(t+)$, $z(t-)$ is independent of $y(t-)$ and $z(t) = z(t-) \vee y(t-) \vee y(t) \vee y(t+)$, so that z is left-continuous if $z(t-) \geq y(t-) \vee y(t) \vee y(t+)$, right-continuous if $y(t+) \geq z(t-) \vee y(t-) \vee y(t)$ and neither left-continuous nor right-continuous if $y(t-) \vee y(t) > z(t-) \vee y(t+)$.

We get extra regularity conditions if we assume that $x \in C$. Recall that z is upper semicontinuous at t if $\lim_{s \to t} z(s) \leq z(t)$; z is upper semicontinuous if it is upper semicontinuous at all t. Let D_{usc} be the subset of upper semicontinuous functions in D_{lim}.

Theorem 9.3.5. (upper-semicontinuity when $x \in C$) *Suppose that $x \in C$ and $y \in D_{lim}$, then $z \in D_{usc}$. Then $\Phi_x^L(t) = \Phi_x^R(t) = \Phi_x(t)$ for all $t > 0$ and*

$$z(t) = \sup_{s \in \Phi_x(t)} \{y(s-) \vee y(s) \vee y(s+)\} \ . \qquad (3.39)$$

Proof. Since $x \in C$, the only relevant cases in Theorem 9.3.3 are: $1(i)$, 2 and 4. Formula (3.39) follows directly from formula (3.6). The upper semicontinuity follows from by considering the cases in Theorem 9.3.3. ∎

Remark 9.3.3. *The need for x to be continuous.* Without assuming that $x \in C$, we need not have z be upper semicontinuous. In Case 3 of Theorem 9.3.3, we can have $z(t-) > z(t) = z(t+) = y(t+)$. ∎

From the point of view of applications, the two most common cases are

$$\begin{aligned} &\text{(i)} \quad x \in C \quad \text{and} \quad y \in C \\ &\text{(ii)} \quad x \in C \quad \text{and} \quad y \in D\,. \end{aligned} \tag{3.40}$$

We thus summarize the situation in these two important cases.

First, with case (i) in (3.40) when both $x \in C$ and $y \in C$, we can apply Corollary 9.3.3 and Theorem 9.3.5 above to conclude that $z \in D_{l,r} \cap D_{usc}$, but Example 9.3.3 shows that we need not have $z \in D$. Indeed, we will always have $z \in D_{l,r} \cap D_{usc}$ instead. For $x \in D$, we have $x \in D_{usc}$ only if $x(t) \geq x(t-)$ for all t. So it is important to have the space $D_{l,r} \cap D_{usc}$.

Second, with Case (ii) in (3.40) when $x \in C$ but only $y \in D$, Theorem 9.3.2 shows that $z \in D_{lim}$, but Example 9.3.4 shows that we need not have $z \in D_{l,r}$ in general. However, under condition (3.29), which is implied by condition (3.30), Theorem 9.3.2 implies that we do have $z \in D_{l,r}$. Moreover Theorem 9.3.5 shows that $z \in D_{usc}$. So, in Case (ii) we should also have $z \in D_{l,r} \cap D_{usc}$, but we need to impose condition (3.30).

Because we assumed only that $y \in D_{lim}$ in Theorem 9.3.1, we can consider z playing the role of y. For example, we could start by considering $z_\epsilon(x_1, y)$ in (3.2) for some $x_1 \in D$ and obtain $z_1 = z(x_,y)$ as $\epsilon \downarrow 0$. Then we could consider $z_\epsilon(x_2, z_1)$ in (3.2) for another $x_2 \in D$ and obtain $z_2 = z(x_2, z_1)$ as $\epsilon \downarrow 0$.

9.4. Extending Pointwise Convergence to M_1 Convergence

We now want to extend the pointwise convergence of z_ϵ to z as $\epsilon \downarrow 0$ in Theorem 9.3.1 to M_1 convergence. We first observe that monotone pointwise convergence of continuous functions in D does not by itself imply M_1 convergence.

Example 9.4.1. *Monotone pointwise convergence of continuous functions does not imply M_1 convergence.* To see that monotone pointwise convergence of continuous functions does not imply M_1 convergence in $D([0,2],\mathbb{R})$, let

$$x_{2^{-n}}(0) = x_{2^{-n}}(1-2^{-n}) = x_{2^{-n}}(1-2^{-(n+1)}) = 0$$

$$x_{2^{-n}}(1-3(2^{-(n+2)})) = x_{2^{-n}}(1-2^{-(n+1)}+2^{-(2n+1)}) = x_{2^{-n}}(2) = 1$$

for $n \geq 1$, with $x_{2^{-n}}$ defined by linear interpolation elsewhere. Clearly $x_{2^{-n}}$ is continuous for each n. Let $x_\epsilon = x_{2^{-n}}$ for $2^{-n} \geq \epsilon > 2^{-(n+1)}$, $n \geq 1$. It is easy to see that $x_{2^{-n}}(t) \geq x_{2^{-(n+1)}}(t) \downarrow x(t)$ as $n \to \infty$ for each $t \geq 0$, so that $x_\epsilon(t) \downarrow x(t)$ as $\epsilon \downarrow 0$ for each $t \geq 0$. Moreover $x_\epsilon \to x$ in D as $\epsilon \downarrow 0$ with the M_2 topology, but not in the M_1 topology, because, for any $\delta > 0$, x_ϵ crosses the strip $(1/3, 2/3)$ for t in $[1-\delta, 1+\delta]$ three times for all sufficiently small ϵ, whereas x crosses it only once; see Theorem 12.5.1 (v) in the book. ■

In general (without continuity conditions) monotone pointwise convergence does not imply even M_2 convergence.

Example 9.4.2. *Monotone pointwise convergence without continuity does not imply M_2 convergence.* To see that M_2 convergence does not follow from monotone pointwise convergence in or $D_{l,r}$ when neither the limit nor the converging functions need be continuous, let $x = I_{[1,2]}$ and $x_n = 2I_{[1-n^{-1},1)} + I_{[1,2]}$, $n \geq 1$. ■

However, we can obtain a positive result when the converging functions are continuous (without relying on the special structure associated with the supremum).

Theorem 9.4.1. (M_2 convergence from monotone pointwise convergence of continuous functions) *If $x \in D_{l,r}$, $x_\epsilon \in C$ for all ϵ and $x_\epsilon(t) \downarrow x(t)$ as $\epsilon \downarrow 0$ for all $t \geq 0$, then $x_\epsilon \to x$ in $(D_{l,r}, M_2)$ as $\epsilon \downarrow 0$.*

We can combine Theorems 9.3.1 and 9.4.1 above to obtain the following corollary.

Corollary 9.4.1. (M_2 convergence of the supremum derivative) *In the setting of Theorem 9.3.1, if x and y are both continuous, then $z_\epsilon \to z$ in $(D_{l,r}, M_2)$ as $\epsilon \downarrow 0$.*

However, by exploiting the special structure of the supremum function, we will actually establish the stronger M_1 convergence under weaker conditions. To prove Theorem 9.4.1, we exploit approximations by piecewise-constant functions see Section 12.2 in the book.

Proof of Theorem 9.4.1. Since the pointwise convergence is monotone, $x_\epsilon(t) \ge x(t)$ for all t and ϵ. For any u and $\delta > 0$, let $\tilde{x}$ be a piecewise-constant function in D with $\|x - \tilde{x}\|_u < \delta$. Then $x(t) \le \tilde{x}(t) + \delta$ for $0 \le t \le u$. Let $\hat{x}$ be the upper boundary (containing only vertical and horizontal pieces) of the δ neighborhood of the completed graph $\Gamma_{\tilde{x}+\delta}$ of $\tilde{x} + \delta$ for the time set $[0, t]$, using the Hausdorff metric, as depicted in Figure 9.1. Note that $\hat{x}(s) \ge x(s)$ for $0 \le s \le t$ and $h_t(\Gamma_x, \Gamma_{\hat{x}}) \le 3\delta$, where h_t is the Hausdorff metric applied to the graphs with time set $[0, t]$. It thus suffices to show that $x_\epsilon(s) \le \hat{x}(s)$ for all s, $0 \le s \le t$, for all sufficiently small ϵ.

Consequently, it suffices to show that $x_\epsilon(s) \vee \hat{x}(s)$ converges uniformly to $\hat{x}(s)$ for $0 \le s \le t$ as $\epsilon \downarrow 0$. However, $\hat{x}$ has only finitely many discontinuities. Since $x_\epsilon \vee \hat{x}$ is continuous and nonincreasing in ϵ, we can apply Dini's theorem to get uniform convergence in any compact subset of $[0, t]$ excluding arbitrarily small open neighborhoods of each of the finitely many discontinuities. To treat the discontinuities, we need to carefully treat the neighborhood to the left (right) of a jump up (down). On the other side, the limit function constrains $x_\epsilon(s) \vee \hat{x}(s)$ as $\epsilon \downarrow 0$. Now suppose that t is one of the finitely many discontinuities of $\hat{x}$. Then there is $\epsilon_0(t)$ such that $|x_\epsilon(t) - x(t)| < \delta/2$ for all $\epsilon < \epsilon_0(t)$ by the pointwise convergence. Let ϵ_0 be the minimum of the finitely many $\epsilon_0(t)$. For any $\epsilon \le \epsilon_0$ given, the continuity of x_{ϵ_0} implies that, for each discontinuity point t, there is an $\eta(t) \equiv \eta(t, \epsilon) > 0$ such that $|x_\epsilon(t) - x_\epsilon(s)| < \delta/2$ for all s with $|s - t| < \eta(t)$. Thus, $|x_\epsilon(s) - x(t)| < \delta$ for $|s - t| < \eta(t)$. On the critical side of each discontinuity, the monotonicity implies that

$$x_{\epsilon'}(s) \le x_\epsilon(s) \le x_\epsilon(t) + \delta/2$$

for all $\epsilon' \le \epsilon$. Let the open neighborhood about t be $(t - \eta(t)/4, t + \eta(t)/4)$. Outside the finite union of those open intervals, we have the uniform convergence; inside those intervals we have established that $x_\epsilon(s) \vee \hat{x}(s) < \hat{x}(s) + \delta$. Hence Γ_{x_ϵ} is contained in the 4δ-neighborhood of Γ_x for suitably small ϵ, which implies the M_2 convergence. ■

We will want to approximate $y \in D$ by $y \in D_c$. For this purpose, it is important to understand how z_ϵ and z in (3.2) and (3.6) depend upon y.

Lemma 9.4.1. (uniform Lipschitz property of z_ϵ as a function of y) *For any $\epsilon > 0$, $t > 0$, $x \in D$ and y_1, $y_2 \in D$,*

$$\|z_\epsilon(x, y_1) - z_\epsilon(x, y_2)\|_t \le \|y_1 - y_2\|_t \tag{4.1}$$

and

$$\|z(x, y_1) - z(x, y_2)\|_t \le \|y_1 - y_2\|_t \ . \tag{4.2}$$

Proof. Property (4.2) follows immediately from (3.6). For (4.1), note that

$$\begin{aligned}\|z_\epsilon(x,y_1) - z_\epsilon(x,y_2)\|_t &= \epsilon^{-1}\|(x+\epsilon y_1)^\uparrow - (x+\epsilon y_2)^\uparrow\|_t \\ &\leq \epsilon^{-1}\|(x+\epsilon y_1) - (x+\epsilon y_2)\|_t \\ &= \|y_1 - y_2\|_t \ . \ \blacksquare\end{aligned}$$

We also employ the following elementary, but useful, lemma.

Lemma 9.4.2. ($z \in D_c$ when $y \in D_c$) *Suppose that x, $y \in D$. If, in addition, $y \in D_c$ and*

$$Disc(y) \cap Bad(x) = \phi \tag{4.3}$$

for $Bad(x)$ in (3.28)*, then $z \in D_c$. If y has k discontinuity points in $(0,t)$, then z has at most k discontinuity points in $[0,t]$.*

Proof. We use Theorem 9.3.2 to show that $z \in D$. Since $y \in D_c$, for any given interval $[0,t]$, there are time points $t_0 = 0 < t_1 < \cdots < t_k = t$ such that y is constant on $[t_{j-1}, t_j)$ and $[t_{k-1}, t]$ for $1 \leq j \leq k$. Note that $z(t) = y(0)$ for $t \in [0, t_1)$. From (3.6), it is obvious that z can only assume one of the k values $y(t_{j-1})$, $1 \leq j \leq k$. The function z may change to $y(t_{j-1})$ in the interval $[t_{j-1}, t_j)$, but it can only do so once. Transitions from $z(t_{j-1}-) < y(t_{j-2})$ to $y(t_{j-2}) < y(t_{j-1})$ to $y(t_{j-1})$ at t_{j-1} are ruled out by condition (4.3). ∎

Theorem 9.4.2. (M_1 convergence of the supremum derivative) *Suppose that $x, y \in D$ and* (4.3) *holds for $Bad(x)$ in* (3.28)*. Then*

$$z_\epsilon \to z \quad in \quad (D_{lr}, M_1) \quad as \quad \epsilon \downarrow 0$$

for z_ϵ in (3.2) *and z in* (3.6)*.*

Proof. Lemmas 9.4.1 and 9.4.2 imply that it suffices to consider $y \in D_c$ in order to establish the M_1 convergence. By Theorem 9.3.2 and Example 9.3.2, the discontinuity condition (4.3) is necessary and sufficient to have $z \in D$. Under condition (4.3), it is possible to choose the piecewise-constant approximation to y so that it too satisfies (4.3). So, henceforth, assume that $y \in D_c$ and satisfies (4.3). By Lemma 9.4.2, $z \in D_c$ as well. Now, by applying mathematical induction over the successive discontinuities of z, it is not difficult to show that, for all sufficiently small $\epsilon > 0$, $z_\epsilon(t) = z(t)$ for all t outside a union of open neighborhoods of the discontinuities of z. (We strongly exploit D_c at this step.) For given discontinuities of y and z, by making ϵ suitably small, these neighborhoods can be chosen to be disjoint with the property that z_ϵ is monotone on each interval. The monotonicity together with the pointwise convergence established in Theorem 9.3.1 implies the local characterization of M_1 convergence in Theorem 12.5.1 in the book. ■

Example 9.4.3. *The need for M_1 convergence.* It is possible to have $z_\epsilon = z$ at a discontinuity point of z: For $x(t) = 0$, $t \geq 0$, $z_\epsilon(t) = z(t) = y^{\uparrow}(t)$ for all $t \geq 0$. Then z_ϵ and z have the discontinuities of $y^{\uparrow}$. A typical case requiring the M_1 convergence is $y = I_{[1,\infty)}$ and $x(t) = -tI_{[0,1)}(t) + (t-2)I_{[1,\infty)}(t)$. Then

$$z_\epsilon(t) = \epsilon^{-1}(2 - t + \epsilon)I_{[2-\epsilon,2)}(t) + I_{[2,\infty)}(t) \to z(t) = I_{[2,\infty)}(t) \quad \text{in} \quad (D, M_1) \ .$$

Finally, we can combine Theorems 9.2.3, 9.4.2 and the triangle inequality (2.1) to obtain a preservation-of-convergence result for the supremum function.

Theorem 9.4.3. (convergence preservation for the supremum map with nonlinear centering) *For $\epsilon > 0$, let x_ϵ, $y \in D$ and let x be a Lipschitz function in C. If*

$$d_{M_1}(x_\epsilon - x, \epsilon y) = o(\epsilon) \quad as \quad \epsilon \downarrow 0 \ , \tag{4.4}$$

for which a sufficient condition is

$$\|\|\epsilon^{-1}(x_\epsilon - x) - y\|_t \to 0 \quad as \quad \epsilon \downarrow 0 \quad for\ all \quad t > 0 \ , \tag{4.5}$$

and if (4.3) holds for $Bad(x)$ in (3.28), then

$$\epsilon^{-1}(x_\epsilon^{\uparrow} - x^{\uparrow}) \to z \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0 \tag{4.6}$$

for z in (3.6).

Corollary 9.4.2. (convergence preservation starting with the standard initial limit (4.5)) *For $\epsilon > 0$, let $x_\epsilon \in D$ and $x, y \in C$ with x being Lipschitz. If (4.5) holds, then (4.6) holds for z in* (3.39) *and $z \in D_{usc} \cap D_{l,r}$.*

9.5. Derivative of the Reflection Map

Now we consider the reflection map $\phi : D \to D$ defined by

$$\phi(x) \equiv x + (-x \vee 0)^{\uparrow} \ ; \tag{5.1}$$

see Section 13.4 in the book.

Results for the reflection map ϕ in (5.1) above follow from the results for the supremum map in Sections 9.3 and 9.4 above, because

$$\phi_\epsilon(x,y) \equiv \epsilon^{-1}[\phi(x+\epsilon y) - \phi(x)] = y + m_\epsilon(-x,-y) \;, \tag{5.2}$$

where

$$m_\epsilon(x,y) = \epsilon^{-1}[((x+\epsilon y)^\uparrow \vee 0 - (x^\uparrow \vee 0)] \;. \tag{5.3}$$

Note that $m_\epsilon(x,y)$ in (5.3) differs from $z_\epsilon(x,y)$ in (3.2) only by the extra maximum with respect to 0. In most applications, we will have $x(0) = y(0) = 0$, in which case the extra maximum $\vee 0$ is superfluous; then $m_\epsilon(x,y) = z_\epsilon(x,y)$. Thus, in this common case we can immediately apply the results in Section 9.3 to obtain corresponding results for the reflection map.

Theorem 9.5.1. (derivative of the reflection map in the common case) *Suppose that $x \in D$, $y \in D$ and $x(0) = y(0) = 0$. Then, for each $t > 0$,*

$$\lim_{\epsilon\downarrow 0} \phi_\epsilon(x,y)(t) = \dot{\phi}(t) \;, \tag{5.4}$$

where

$$\begin{aligned} \dot{\phi}(t) &\equiv \dot{\phi}(x,y)(t) \\ &\equiv y(t) - \left(\inf_{s\in\Phi^L_{-x}(t)} \{y(s-)\} \wedge \inf_{s\in\Phi^R_{-x}(t)} \{y(s)\} \right) \end{aligned} \tag{5.5}$$

and $\dot{\phi} \in D_{lim}$. If, in addition,

$$Disc(y) \cap Bad(-x) = \phi \;, \tag{5.6}$$

then $\dot{\phi} \in D_{l,r}$ and

$$\phi_\epsilon(x,y) \to \dot{\phi}(x,y) \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0\,. \tag{5.7}$$

If, in addition, $x \in C$, then $\dot{\phi} \in D_{usc}$. If, in addition, x is Lipschitz and $y \in C$, then there is convergence preservation: If

$$\|\epsilon^{-1}(x_\epsilon - x) - y\|_t \to 0 \quad as \quad \epsilon \downarrow 0 \quad for\ all \quad t \tag{5.8}$$

then

$$\epsilon^{-1}(\phi(x_\epsilon) - \phi(x)) \to \dot{\phi}(x,y) \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0\,. \tag{5.9}$$

for

$$\dot{\phi}(t) = y(t) - \inf_{s\in\Phi_{-x}(t)} \{y(s)\}\,. \tag{5.10}$$

where

$$\Phi_{-x}(t) = \{s : 0 \le s \le t, x(s) = x^\downarrow(t)\}, \quad t \ge 0\,. \tag{5.11}$$

Proof. The pointwise limit in (5.4) follows from Theorem 9.3.1, noting that $-(-y)^{\uparrow} = y^{\downarrow}$. The fact that $\dot{\phi} \in D_{lim}$ follows from Theorem 9.3.2. The stronger conclusion that $\dot{\phi} \in D_{l,r}$ under condition (5.6) also follows from Theorem 9.3.2, exploiting condition (3.30). The M_1 convergence in (5.7) follows from Theorem 9.4.2. Finally, the convergence preservation ((5.8) implies (5.9)) follows from Corollary 9.4.3. ∎

We now return to the general case. For that purpose, let

$$t_l \equiv t_l(x) \equiv \inf\{t > 0 : x^{\uparrow}(t) = 0\} \tag{5.12}$$

and

$$t_u = t_u(x) \equiv \sup\{t > 0 : x^{\uparrow}(t) = 0\} \ , \tag{5.13}$$

with $t_l = t_u = \infty$ if $x^{\uparrow}(t) < 0$ for all t. In many applications we will have $x(0) = 0$; then $t_l = 0$ and $t_u = \infty$. It is easy to see that for any t, $0 \le t < t_l$, $m_\epsilon(x,y)(t) = 0$ for all sufficiently small positive ϵ. Similarly, for any t, $t_u < t < \infty$, $m_\epsilon(x,y)(t) = z_\epsilon(x,y)(t)$ for all sufficiently small positive ϵ. We need to examine the interval $(t_l - \epsilon, t_u + \epsilon)$ more carefully. To do so, we exploit the following analog of Lemma 9.4.1, which is proved in the same way.

Lemma 9.5.1. (uniform Lipschitz property for m_ϵ as a function of y) *For any $\epsilon > 0$, $t > 0$, $x \in D$ and $y_1, y_2 \in D$,*

$$\|m_\epsilon(x,y_1) - m_\epsilon(x,y_2)\|_t \le \|y_1 - y_2\|_t \ .$$

Our analog of Theorems 9.3.1, 9.3.2, 9.3.5 and 9.4.2 for m_ϵ is the following.

Theorem 9.5.2. (the derivative in the general case) *Suppose that $x, y \in D$. For each $t \ge 0$, $m_\epsilon(x,y)(t)$ is decreasing in ϵ and*

$$\lim_{\epsilon \downarrow 0} m_\epsilon(x,y)(t) = m(x,y)(t) \equiv \begin{cases} 0, & t < t_l \\ y(t-) \vee y(t) \vee 0, & t = t_l \\ z(t) \vee 0, & t_l < t < t_u \\ z(t-) \vee 0 \vee y(t), & t = t_u \\ z(t), & t > t_u \end{cases} \tag{5.14}$$

for m_ϵ in (5.3), t_l in (5.12), t_u in (5.13) and $z(t)$ in (3.6). The limit $m(x,y)$ in (5.14) has limits from the left and right at all t. If $x \in C$, then z is given by (3.39) and z and m are upper semicontinuous. At all t not in the set

$$B(x) \equiv \{t_l\} \cup (Bad(x) \cap (t_l, \infty)) \tag{5.15}$$

for $Bad(x)$ in (3.28), *m is either left-continuous or right-continuous. At $t = t_l$, m is left-continuous if $y(t-) \vee y(t) \le 0$, m is right-continuous if $y(t) \ge y(t-) \vee 0$, and neither left-continuous nor right-continuous if $y(t-) > y(t) \vee 0$. If*

$$(i)\ y(t-) \le z(t-) \vee y(t) \vee 0 \quad for \quad t \in B(x) \cap [t_l, t_u] \tag{5.16}$$

and

$$(ii)\ y(t-) \le z(t-) \vee y(t) \quad for \quad t \in B(x) \cap (t_u, \infty)\ , \tag{5.17}$$

for which a sufficient condition is

$$Disc(y) \cap B(x) = \phi\ , \tag{5.18}$$

then m is either left-continuous or right-continuous at all t, so that $m \in D_{l,r}$. Then

$$m_\epsilon(x,y) \to m(x,y) \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0\ .$$

Proof. First, for any $\delta > 0$ and $T > 0$, $m_\epsilon(x,y)(t) = 0$ in $[0, (0 \vee (t_l - \delta)) \wedge T]$ and $m_\epsilon(x,y)(t) = z_\epsilon(x,y)(t)$ in $[(t_u + \delta) \wedge T, T]$ for all sufficiently small positive ϵ. We apply Theorems 9.3.1, 9.3.2 and 9.4.2 to treat the subinterval $[(t_u + \delta) \wedge T, T]$. Hence it suffices to focus on the subinterval $(t_l - \delta, t_u + \delta)$. By Lemmas 9.4.1 and 9.5.1, it suffices to assume that $y \in D_c$. The argument then is as for Theorems 9.3.1, 9.3.2, 9.3.5 and 9.4.2. ■

Corollary 9.5.1. (convergence) *If $x, y \in D$, then*

$$\phi_\epsilon(x,y)(t) \downarrow y(t) + m(-x,-y)(t) \quad as \quad \epsilon \downarrow 0$$

for ϕ_ϵ in (5.2), *each $t \ge 0$ and m in* (5.14). *If in addition* (5.18) *holds, then*

$$\phi_\epsilon(x,y) \to y + m(-x,-y) \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0\ .$$

Finally, paralleling Theorem 9.4.3 for the supremum function, we can combine Theorems 9.2.3, 9.5.2 and the triangle inequality in (2.1) to obtain a preservation-of-convergence result for the reflection map.

Theorem 9.5.3. (M_1 convergence for the reflection derivative) *For $\epsilon > 0$, let $x_\epsilon, y \in D$ and let x be a Lipschitz function in C. If condition (4.4) holds, for which a sufficient condition is (4.5), and if* (5.18) *holds, then*

$$\epsilon^{-1}(\phi(x_\epsilon) - \phi(x)) \to y + m(-x,-y) \quad in \quad (D_{l,r}, M_1) \quad as \quad \epsilon \downarrow 0 \tag{5.19}$$

for m in (5.14).

Corollary 9.5.2. *For $\epsilon > 0$, let $x_\epsilon \in D$ and $x, y \in C$ with x being Lipschitz. If* (4.5) *holds, then* (5.19) *holds for m in* (5.14), *where $m \in D_{usc} \cap D_{l,r}$.*

9.6. Heavy-Traffic Limits for Nonstationary Queues

In this section we apply the convergence-preservation results in the last section to establish heavy-traffic limits for nonstationary queues. We assume that the queue-length process can be represented directly as the reflection map applied to a net-input process, which is the difference of two nondecreasing processes admitting nonstationary rates.

As background, note that the queue-length process $\{Q(t) : t \geq 0\}$ in the M/M/1 queue starting empty with arrival rate λ and service rate μ has such a representation. In particular, for the M/M/1 queue,

$$Q(t) = \phi(X)(t), \quad t \geq 0 , \tag{6.1}$$

where X is the net-input process, satisfying

$$X(t) = X^{+}(\Lambda^{+}(t)) - X^{-}(\Lambda^{-}(t)) , \tag{6.2}$$

with X^{+} and X^{-} being rate-1 Poisson processes and

$$\Lambda^{+}(t) = \lambda t \quad \text{and} \quad \Lambda^{-}(t) = \mu t, \quad t \geq 0 . \tag{6.3}$$

Then $X^{+} \circ \Lambda^{+}$ is a rate-λ Poisson process.

Similarly, for the $M_t/M_t/1$ queue with (integrable) time-dependent arrival-rate function $\lambda(t)$ and service-rate function $\mu(t)$, (6.1) and (6.2) remain valid with Λ^{+} and Λ^{-} redefined as

$$\Lambda^{+}(t) = \int_0^t \lambda(s) ds \quad \text{and} \quad \Lambda^{-}(t) = \int_0^t \mu(s) ds . \tag{6.4}$$

It is easy to see that there are many generalizations. First, we obtain the queue-length process in an MMPP/MMPP/1 queue with independent Markov modulated Poisson process (MPPP) arrival and service processes if Λ^{+} and Λ^{-} are independent stationary versions of finite-state continuous-time Markov chains. (We then assume that X^{+}, X^{-}, Λ^{+} and Λ^{-} are mutually independent. We obtain the queue-length process in a more general $\mathrm{MMPP_t}/\mathrm{MMPP_t}/1$ queue with independent time-dependent MMPP arrival and service processes if Λ^{+} and Λ^{-} are independent time-dependent finite-state CTMCs, governed by time-dependent transition functions.

We construct associated fluid queue models by letting X^{+} and X^{-} be other Lévy processes instead of Poisson processes. Without loss of generality, these again can be rate-1 processes. For nodes in a communication network with fixed bandwidth, it is natural to let $X^{-}(t) = t$, $t \geq 0$, but generalizations are possible.

We now establish limits for a sequence of models indexed by n. For each n, we have the four-tuple of stochastic processes $(X_n^+, X_n^-, \Lambda_n^+, \Lambda_n^-)$ with sample paths in D^4. We then form the associated scaled stochastic processes by letting

$$\begin{aligned}
\mathbf{X}_n^+(t) &\equiv c_n^{-1}[X_n^+(nt) - nx^+(t)] \\
\mathbf{X}_n^-(t) &\equiv c_n^{-1}[X_n^-(nt) - nx^-(t)] \\
\mathbf{\Lambda}_n^+(t) &\equiv c_n^{-1}[\Lambda_n^+(t) - ny^+(t)] \\
\mathbf{\Lambda}_n^-(t) &\equiv c_n^{-1}[\Lambda_n^-(t) - ny^-(t)] \\
\mathbf{X}_n(t) &\equiv c_n^{-1}[X_n^+(\Lambda_n^+(t)) - X_n^-(\Lambda_n^-(t)) - nx^+(y^+(t)) - x^-(y^-(t))] \\
\hat{\mathbf{X}}_n(t) &\equiv n^{-1}[X_n^+(\Lambda_n^+(t)) - X_n^-(\Lambda_n^-(t))].\ t \ge 0\,, \qquad (6.5)
\end{aligned}$$

We think of the centering terms x^+, x^-, y^+ and y^- as deterministic functions, but that is not necessary.

The following limit for the net-input process is a direct consequence of Theorem 13.3.2 in the book.

Theorem 9.6.1. (FLLN and FCLT for the net-input process) *Suppose that*

$$(\mathbf{X}_n^+, \mathbf{X}_n^-, \mathbf{\Lambda}_n^+, \mathbf{\Lambda}_n^-) \Rightarrow (\mathbf{U}^+, \mathbf{U}^-, \mathbf{V}^+, \mathbf{V}^-) \quad in \quad (D^4, WM_1) \qquad (6.6)$$

for the processes in (6.5), *where* x^+ *and* x^- *have continuous derivatives* $\dot{x}^+$ *and* $\dot{x}^-$, y^+ *and* y^- *are continuous nonnegative and strictly increasing,* $c_n \to \infty$, $n/c_n \to \infty$ *and*

$$\begin{aligned}
&Disc(\mathbf{U}^+ \circ y^+) \cap Disc(\mathbf{V}^+) = \phi \\
&Disc(\mathbf{U}^- \circ y^-) \cap Disc(\mathbf{V}^-) = \phi \\
&Disc(\mathbf{U}^+ \circ y^+ + (\dot{x}^+ \circ y^+)\mathbf{V}^+) \cap \\
&\quad Disc(\mathbf{U}^- \circ y^- + (\dot{x}^- \circ y^-)\mathbf{V}^-) = \phi\,. \qquad (6.7)
\end{aligned}$$

Then

$$\hat{\mathbf{X}}_n \Rightarrow x \quad in \quad (D, M_1) \qquad (6.8)$$

and

$$\mathbf{X}_n \Rightarrow \mathbf{X} \quad in \quad (D, M_1)\ , \qquad (6.9)$$

for $\hat{\mathbf{X}}_n$ *and* $\mathbf{X}_n$ *in* (6.5), *where*

$$x \equiv x^+ \circ y^+ - x^- \circ y^- \qquad (6.10)$$

and

$$\mathbf{X} \equiv \mathbf{U}^+ \circ y^+ + (\dot{x}^+ \circ y^+)\mathbf{V}^+ - \mathbf{U}^- \circ y^- - (\dot{x}^- \circ y^-)\mathbf{V}^-\,. \qquad (6.11)$$

Proof. As usual, start by applying the Skorohod representation theorem to replace the convergence in distribution in (6.6) by convergence w.p.1 for special versions, without introducing new notation for the special versions. Then apply Theorem 13.3.2 in the book, after rewriting $\mathbf{X}_n^+$ as

$$\mathbf{X}_n^+(t) \equiv (n/c_n)[n^{-1}X_n^+(nt) - x^+(t)], \quad t \geq 0\,, \tag{6.12}$$

and similarly for the other functions. That yields

$$\begin{aligned} &c_n^{-1}(X_n^+ \circ \Lambda_n^+ - nx^+ \circ y^+, X_n^- \circ \Lambda_n^- - nx^- \circ y^-) \\ &\quad \Rightarrow (U^+ \circ y^+ + (\dot{x}^+ \circ y^+)V^+, U^- \circ y^- + (\dot{x}^- \circ y^-)V^-) \end{aligned} \tag{6.13}$$

in (D^2, WM_1). Multiply by c_n/n in (6.13) to get

$$n^{-1}(X_n^+ \circ \Lambda_n^+, X_n^- \circ \Lambda_n^-) \Rightarrow (x^+ \circ y^+, x^- \circ y^-) \quad \text{in} \quad (D^2, WM_1) \tag{6.14}$$

Finally, given the last condition in (6.7), we can apply addition to go from (6.13) and (6.14) to (6.9) and (6.8). ■

We now apply Theorem 9.5.1 to obtain a corresponding result for the queue-length processes. Let

$$\mathbf{Q}_n(t) \equiv c_n^{-1}(Q_n(nt) - nq(t)), \quad t \geq 0\,. \tag{6.15}$$

and

$$\hat{\mathbf{Q}}_n(t) \equiv n^{-1}Q_n(nt), \quad t \geq 0\,. \tag{6.16}$$

Theorem 9.6.2. (FLLN and FCLT for the queue-length process) *If, in addition to the assumptions of Theorem 9.6.1, y^+ and y^- are Lipschitz continuous, $x(0) = 0$, $P(\mathbf{X}(0) = 0) = 1$ and*

$$P((\mathbf{U}^+, \mathbf{U}^-, \mathbf{V}^+, \mathbf{V}^-) \in C^4) = 1\,, \tag{6.17}$$

then

$$\hat{\mathbf{Q}}_n \Rightarrow q \quad in \quad (D, M_1) \tag{6.18}$$

and

$$\mathbf{Q}_n \Rightarrow \mathbf{Q} \quad in \quad (D_{l,r}, M_1) \tag{6.19}$$

for $\hat{\mathbf{Q}}_n$ in (6.16) and $\mathbf{Q}_n$ in (6.15), where

$$q = \phi(x) \tag{6.20}$$

for x in (6.10) and

$$\mathbf{Q} = \mathbf{X} + z(-x, -\mathbf{X}) \tag{6.21}$$

for x in (6.10), $\mathbf{X}$ in (6.11) and z in (3.16). The limit process $\mathbf{Q}$ then has upper semicontinuous sample paths.

Example 9.6.1. *The* $\mathrm{M_t/M_t/1}$ *queue.* Now let us examine the special case of the $\mathrm{M_t/M_t/1}$ queue in more detail. For the $\mathrm{M_t/M_t/1}$ queue, $c_n = \sqrt{n}$, $x^+ = x^- = e$ and U^+, U^- are independent Brownian motions. It is natural to have

$$\Lambda_n^{\pm}(t) = \int_0^t \lambda_n^{\pm}(s)ds \quad \text{and} \quad y^{\pm}(t) = \int_0^t \lambda^{\pm}(s)ds \tag{6.22}$$

where $\lambda_n^{\pm}$ and $\lambda^{\pm}$ are deterministic functions. We can then have

$$n^{-1/2}(\lambda_n^{\pm}(t) - n\lambda^{\pm}(t)) \to \gamma^{\pm}(t) \quad \text{as} \quad n \to \infty \tag{6.23}$$

uniformly in $[0, T]$, where γ^+ and γ^- are deterministic, which implies that

$$\mathbf{\Lambda}_n^{\pm}(t) \to \int_0^t \gamma^{\pm}(s)ds \equiv \mathbf{V}^{\pm} \; . \tag{6.24}$$

Thus the assumptions of Theorems 9.6.1 and 9.6.2 are satisfied and

$$x(t) = \int_0^t [\lambda^+(s) - \lambda^-(s)]ds, \quad t \geq 0\,, \tag{6.25}$$

while

$$\begin{aligned} \mathbf{X}(t) \;\; &= \;\; \mathbf{U}^+ \left(\int_0^t \lambda^+(s)ds \right) \\ &\quad - \mathbf{U}^- \left(\int_0^t \lambda^-(s)ds \right) + \int_0^t [\gamma^+(s) - \gamma^-(s)]ds \end{aligned} \tag{6.26}$$

where $\mathbf{U}^+$ and $\mathbf{U}^-$ are independent standard Brownian motions and the rest involves continuous deterministic functions. It is easy to see that X is equal in distribution (on D) to

$$U\left(\int_0^t [\lambda^+(s) + \lambda^-(s)]ds \right) + \int_0^t [\gamma^+(s) - \gamma^-(s)]ds, \quad t \geq 0\,, \tag{6.27}$$

where U is a standard Brownian motion.

The FWLLN limits x and q can be regarded as the net-input and buffer-content processes, respectively, in a fluid-queue model with time-dependent deterministic input rate $\lambda^+(t)$ and time-dependent deterministic potential output rate $\lambda^-(t)$. Then

$$-(-x)^{\downarrow} = - \min_{0 \leq s \leq t} \left\{ \int_0^s [\lambda^-(r) - \lambda^+(r)]dr \right\} \tag{6.28}$$

represents the cumulative potential output that is lost (i.e., does not occur during the interval $[0,t]$ because of insufficient input. Then

$$\Phi_{-x}(t) = \{s : 0 \le s \le t, q(s) = 0, -(-x)^{\downarrow}(s) = -(-x)^{\downarrow}(t)\} \qquad (6.29)$$

i.e., $\Phi_{-x}(t)$ is the set of times s at which the buffer is empty $(q(s) = 0)$ and there is no potential output los over $[s,t]$.

An important special case is when λ_n^+ and λ_n^- in (6.22) are independent of n. Then $\gamma^+(t) = \gamma^-(t) = 0$ for all $t \ge 0$ and the deterministic function $\int_0^t [\gamma^+(s) - \gamma^-(s)]ds$ in (6.27) is identically 0. Then the limit for the queue-length process has one of three forms over subintervals: time-scaled standard Brownian motion (BM), time-scaled canonical reflected Brownian motion (RBM) and the zero function. There can be discontinuities in the sample path when the set function $\Phi_{-x}(t)$ is discontinuous in t. We display possible sample paths of (λ^+, λ^-), $(-x, (-x)^{\uparrow})$, $\Phi_{-x}(t)$, q and $\mathbf{Q}$ when λ^- is the constant function in Figure 9.2 below. We identify nine intervals associated with nine time points $t_0 \equiv 0 < t_1 < \cdots < t_8$.

In this example, the fluid rates start out ordered by $\lambda^+(t) < \lambda^-(t)$. Thus $-x(t) \equiv \int_0^t [\lambda^-(s) - \lambda^+(s)]ds$ is initially increasing, which implies that $\Phi_{-x}(t) = \{t\}$. Thus $Q(t) = q(t) = 0$ for these t. At time t_1, the ordering switches to $\lambda^+(t) > \lambda^-(t)$. Thus after t_1, $-x$ is decreasing, so that $\Phi_{-x}(t) = \{t_1\}$. At time t_2, the ordering switches back to $\lambda^+(t) < \lambda^-(t)$, but $-x(t)$ does not reach $(-x)^{\uparrow}(t) = (-x)(t_1)$ and $q(t)$ does not return to 0 until $t = t_3$. In the interval (t_1, t_3), q is positive and Q is time-scaled BM.

At time t_3, there is a discontinuity in the set-valued function Φ_{-x} and a corresponding jump in the stochastic process Q. In the interval (t_3, t_4), $-x$ is still increasing and $\Phi_{-x}(t) = \{t\}$, so that $q(t) = Q(t) = 0$, just as in $[0, t_1)$. In the interval (t_4, t_5), $\lambda^+(t) = \lambda^-$, so that $-x$ is constant and $\Phi_{-x}(t) = [t_4, t]$, $t_4 \le t \le t_5$. In the interval (t_4, t_5), Q evolves as RBM. At t_5, λ^+ increases, so that $-x$ decreases and $\Phi_{-x}(t) = \Phi_{-x}(t_5) = [t_4, t_5]$ for $t_5 \le t < t_7$. At t_6, λ^+ starts to decrease again and at t_7 $q(t) = 0$ for the first time. Hence, Q evolves as BM in the interval (t_5, t_7).

At t_7, there is a second discontinuity in Φ_{-x} and a corresponding jump in Q. In the subsequent interval $[t_7, t_8]$, $\lambda^+(t) = \lambda^-$, so that $-x$ remains constant. Then $\Phi_{-x}(t) = [t_4, t_5] \cup [t_7, t]$ for $t_7 \le t < t_8$. During the interval $[t_7, t_8]$, $q(t) = 0$ and Q evolves as RBM. At t_8, λ^+ starts to decrease and thereafter remains below λ^-. Hence, Φ_{-x} has another discontinuity at t_8. After t_8, $\Phi_{-x}(t) = \{t\}$ and $q(t) = Q(t) = 0$.

We conclude this section by relating the three possible kinds of heavy-traffic limits for the case of the $\mathrm{M_t/M_t}/1$ queue with fixed arrival and service rate functions $\lambda^+(t)$ and $\mu^-(t)$ to the values of a *time-dependent traffic intensity*, defined by

$$\rho^*(t) \equiv \sup_{0 \le s \le t} \left\{ \int_0^t \lambda^+(r)dr / \int_s^t \lambda^-(r)dr \right\}, \quad t \ge 0 . \tag{6.30}$$

Notice that the buffer-content deterministic fluid limit q satisfies

$$\begin{aligned} q(t) &= x(t) - \inf_{0 \le s - t} x(s) \\ &= \sup_{0 \le s \le t} \{x(t) - x(s)\} \\ &= \sup_{0 \le s \le t} \left\{ \int_s^t [\lambda^+(r) - \lambda^-(r)]dr \right\} , \end{aligned} \tag{6.31}$$

so that $q(t) > 0$ if and only if $\rho^*(t) > 1$.

Moreover, we can have $q(t) = 0$ but $P(Q(t) = 0) = 0$ for all t in an interval (a, b) if and only if $\rho^*(t) = 1$ in (a, b). First, we must have $\rho^* \le 1$ since $q(t) = 0$. However, in this region we must also have

$$\int_s^t [\lambda^+(r) - \lambda^-(r)]dr = 0 \tag{6.32}$$

for some s suitably chose to t. For that s,

$$\int_s^t \lambda^+(r)dr / \int_s^t \lambda^-(r)dr = 1 \tag{6.33}$$

which implies that $\rho^*(t) \ge 1$. Since both $\rho^*(t) \le 1$ and $\rho^*(t) \ge 1$, we must have $\rho^*(t) = 1$.

We thus say that the queue is *overloaded*, *critically loaded* or *underloaded* in an open interval (a, b) if $\rho^*(t) > 1$, $\rho^*(t) = 1$ or $\rho^*(t) < 1$ throughout the interval (a, b). In Figure 9.2 above, in the intervals $(0, t_1)$, (t_1, t_3), (t_3, t_4), (t_4, t_5), (t_5, t_7), (t_7, t_8) and (t_8, T), we have successively $\rho^*(t) < 1$, > 1, < 1, $= 1$, > 1, $= 1$ and < 1.

9.7. Derivative of the Inverse Map

In this section we obtain convergence-preservation results for the inverse map

$$x^{-1}(t) \equiv \inf\{s \ge 0 : x(s) > t\}, \quad t \ge 0 , \tag{7.1}$$

defined on the subset D_u of functions unbounded above in $D \equiv D([0,\infty),\mathbb{R})$, as in Section 13.6 of the book. As in previous sections here, we approach convergence preservation through a derivative representation.

To determine the derivative of the inverse map, we introduce yet another topology on D. Recall that we introduced the M_1' topology on $D([0,t],\mathbb{R})$ by appending a segment to the graphs, i.e., by letting

$$\Gamma_x' = \Gamma_x \cup \{(\alpha x(0), 0) : 0 \le \alpha \le 1\} \ , \tag{7.2}$$

where Γ_x is the graph of x, i.e.,

$$\begin{aligned}\Gamma_x \equiv \{&(z,s) \in \mathbb{R} \times [0,t] : \\ &z = \alpha x(s-) + (1-\alpha)x(s) \quad \text{for some } \alpha,\ 0 \le \alpha \le 1\}\,.\end{aligned} \tag{7.3}$$

We now construct a similar M_1'' topology on $D([0,t),\mathbb{R})$ by also appending the vertical line at t to the graph, i.e., by setting

$$\Gamma_x'' = \Gamma_x' \cup (\mathbb{R} \times \{t\}) \tag{7.4}$$

for Γ_x' in (7.2). Note that the function value at the right endpoint t plays no role in the M_1'' topology.

As done before for the graph Γ_x in (7.3), we define a lexicographic *order relation* on the graph Γ_x'' by saying that $(z_1,s_1) \le (z_2,s_2)$ if either (i) $s_1 < s_2$ or (ii) $s_1 = s_2$ and $|x(s_1-) - z_1| \le |x(s_1-) - z_2|$. The definition makes the relation $\le$ a total order on the graph Γ_x''. A parametric representation of the graph Γ_x'' or the function x is a continuous nondecreasing function (u,r) mapping $[0,1]$ *into* the graph Γ_x'' such that $r(0) = 0$, $u(0) = 0$ and $r(1) = t$. We allow the parametric representation of Γ_x'' to cover only part of the vertical line at t. If $r(s) < t$ for all $s < 1$, then the parametric representation (u,r) covers only the single point $(x(t-),t)$. If $r(s) = t$ for $a \le s \le 1$, then (u,r) covers a compact subinterval of either $\{(z,t) : z \ge x(t-)\}$ or $\{(z,t) : z \le x(t-)\}$. (Since (u,r) maps $[0,1]$ into Γ_x'', we must have $(u(1),r(1)) \in \Gamma_x''$, which implies that $|u(1)| < \infty$.) Let $\Pi''(x)$ be the set of all parametric representations of Γ_x''.

A metric d_t'' inducing the M_1'' topology on $D([0,t),\mathbb{R})$ is defined by letting

$$d_t''(x_1,x_2) = \inf_{\substack{(u_i,r_i)\in\Pi''(x_i)\\ i=1,2}} \{\|u_1 - u_2\|_1 \vee \|r_1 - r_2\|_1\} \ . \tag{7.5}$$

We have the following lemma linking the M_1' and M_1'' topologies with bounded function domains.

Lemma 9.7.1. *Let $x, x_n \in D([0,\infty),\mathbb{R})$. If $x_n \to x$ as $n \to \infty$ for the restrictions in $D([0,t_2),\mathbb{R},M_1'')$ for $0 < t_2 < \infty$, then $x_n \to x$ as $n \to \infty$ for the restrictions in $D([0,t_1],\mathbb{R},M_1')$ for each $t_1 \in Disc(x)^c$ with $0 < t_1 < t_2$.*

As before, we say that $x_n \to x$ in $D([0,\infty),R)$ with any of the topologies M_1, M_1' or M_1'' if $x_n \to x$ for the restrictions in $D([0,t],\mathbb{R})$ ($D([0,t),\mathbb{R})$ for M_2'') with the same topology for all t in a sequence $\{t_k\}$ with $t_k \to \infty$ as $k \to \infty$. (The boundary points t_k can be taken from $Disc(x)^c$.) We obtain the following result from Lemma 9.7.1.

Lemma 9.7.2. *The M_1' and M_1'' topologies coincide on $D([0,\infty),\mathbb{R})$.*

We can combine Lemma 9.7.2 here and Theorem 13.6.3 in the book to obtain the following connection between M_1'' and M_1.

Lemma 9.7.3. *If*

$$x_n \to x \quad in \quad D([0,\infty),\mathbb{R},M_1'') \ ,$$

where $x(0) = 0$, then

$$x_n \to x \quad in \quad D([0,\infty),\mathbb{R},M_1) \ .$$

A metric d'' inducing the M_1'' topology on $D([0,\infty),\mathbb{R})$ is defined by letting

$$d''(x_1,x_2) = \int_0^\infty e^{-t}[1 \wedge d_t''(x_1,x_2)]dt \ , \tag{7.6}$$

where $d_t''(x_1,x_2)$ is understood to be the d_t'' metric applied to the restrictions of x_1 and x_2 to $[0,t)$. There is convergence $d''(x_n,x) \to 0$ if and only if there exist parametric representations (u,r) of x and (u_n,r_n) of x_n, $n \geq 1$ with domains $[0,\infty)$, such that $\|u_n - u\|_t \vee \|r_n - r\|_t \to 0$ as $n \to \infty$ for each t.

To apply the approach in Section 9.2, we need the inverse map to be Lipschitz. The Lipschitz property is valid on an appropriate subset of D with an appropriate choice of metrics. Recall that D_u is the subset of functions x in $D \equiv D([0,\infty),\mathbb{R})$ that are unbounded above and have $x(0) \geq 0$. For positive t_1, t_2, let $D_u(t_1,t_2)$ be the subset of x in D_u with $x^{\uparrow}(t_1) \geq t_2$. Clearly $D_u(t_1,t_2)$ is a closed subset of D_u. Moreover,

$$D_u = \cap_{m=1}^{\infty} \cup_{k=1}^{\infty} D(k,m) \ . \tag{7.7}$$

We now show that the inverse map from $D_u(t_1,t_2) \subseteq D_u([0,t_1],\mathbb{R},M_1)$ to $D([0,t_2),\mathbb{R},M_1'')$ is Lipschitz.

Lemma 9.7.4. *For $t > 0$, let d''_t be the M''_1 metric on $D([0,t),\mathbb{R})$ and let d_t be the M_1 metric on $D([0,t],\mathbb{R})$. If x_1, $x_2 \in D_u(t_1,t_2)$, then*

$$d''_{t_2}(x_1^{-1},x_2^{-1}) \le d_{t_1}(x_1^{\uparrow}\wedge t_2, x_2^{\uparrow}\wedge t_2) \le d_{t_1}(x_1^{\uparrow},x_2^{\uparrow}) \le d_{t_1}(x_1,x_2)\ . \quad (7.8)$$

where $(x_i^{\uparrow}\wedge t_2)(s) = x_i^{\uparrow}(s)\wedge t_2$, $0 \le s \le t_1$.

Proof. For $x_i \in D_u(t_1,t_2)$, let (u_i,r_i) be an arbitrary M_1 parametric representation of $x_i^{\uparrow}\wedge t_2$ over $[0,t_1]$. Then (r_i,u_i) is an M''_1 parametric representation of x_i^{-1} over $[0,t_2)$ with the special property that $u_i(1) = t_1$. Hence

$$d''_{t_2}(x_1^{-1},x_2^{-1}) \le d_{t_1}(x_1^{\uparrow}\wedge t_2, x_2^{\uparrow}\wedge t_2)\ . \quad (7.9)$$

It is not difficult to see that

$$d_{t_1}(x_1^{\uparrow}\wedge t_1, x_2^{\uparrow}\wedge t_2) \le d_{t_1}(x_1^{\uparrow},x_2^{\uparrow}) \le d_{t_1}(x_1,x_2)\ .$$

Hence the proof is complete. ■

Lemmas 9.7.2 and 9.7.4 imply that the inverse map from $D_u([0,\infty),\mathbb{R},M_1)$ to $D_u([0,\infty),\mathbb{R},M'_1)$ is continuous, which is weaker than Theorem 13.6.2 in the book. We now want to establish an analog of Theorem 9.2.3. For that purpose, we need both x and x^{-1} to be Lipschitz on $[0,t]$ for all $t > 0$. The following lemma provides natural conditions.

Lemma 9.7.5. (conditions for both x and x^{-1} to be Lipschitz) *If $x \in D_u$ is absolutely continuous, i.e., $x(t) = \int_0^t \dot{x}(s)ds$ for $t > 0$, with $\dot{x} \in D$ and with $l(t) \le \dot{x}(t) \le u(t)$ for all $t \ge 0$ where $0 < l^{\downarrow}(t) < u^{\uparrow}(t) < \infty$ for all t, then*

$$x^{-1}(t) = \int_0^t [1/\dot{x}(x^{-1}(s))]ds \quad \textit{for all} \quad t > 0 \quad (7.10)$$

and x and x^{-1} are both Lipschitz on $[0,t]$ for all $t > 0$, with

$$\dot{x}^{-1}(t) \equiv \frac{d}{dt}(x^{-1})(t) = 1/\dot{x}(x^{-1}(t))\ . \quad (7.11)$$

Proof. Clearly x is strictly increasing and continuous, so that x is a homeomorphism of $[0,\infty)$ and $x \circ x^{-1} = e$, where $\circ$ is the composition map. Thus

$$x(x^{-1}(t)) = \int_0^{x^{-1}(t)} \dot{x}(s)ds = t, \quad t \ge 0\ ,$$

which implies that

$$x^{-1}(t) = \int_0^t [1/\dot{x}(x^{-1}(s))]ds, \quad t \geq 0 .$$

The Lipschitz properties hold because

$$|x(t_2) - x(t_1)| = \int_{t_1}^{t_2} \dot{x}(s)ds \leq u^{\uparrow}(t_2)|t_2 - t_1|$$

and

$$|x^{-1}(t_2) - x^{-1}(t_1)| = \int_{t_1}^{t_2} [1/\dot{x}(x^{-1}(s))]ds \leq |t_2 - t_1|/l^{\downarrow}(t_2) . \quad \blacksquare$$

We now want to establish an analog of Theorem 9.2.3. Since the M_1'' analog of Lemma 9.2.2 is evident, we only establish the M_1'' analog of Lemma 9.2.1.

Lemma 9.7.6. (reduction of convergence to the derivative with the M_1'' topology) *Suppose that x is Lipschitz on $[0,t]$ with Lipschitz constant K. Let d_t'' be the M_1'' metric on $D([0,t),\mathbb{R})$. Then*

$$d_t''(x_1 - x, x_2 - x) \leq (1 + K)d_t''(x_1, x_2) .$$

Proof. For all $\epsilon > 0$, there exists $\eta(\epsilon) > 0$ and parametric representations $(u_{i,\epsilon}, r_{i,\epsilon}) \in \Pi_t''(x_i)$ such that

$$\|u_{1,\epsilon} - u_{2,\epsilon}\| \vee \|r_{1,\epsilon} - r_{2,\epsilon}\| \leq (1 + \eta(\epsilon))d_t''(x_1, x_2) . \tag{7.12}$$

We now want natural modifications of the parametric representations of x_i to serve as parametric representations of x and $x_i - x$. To obtain such parametric representations for x, we need to allow for the line segment joining $(x(0), 0)$ to $(0, 0)$. Hence we first modify the parametric representations of x_i. Let $(u'_{i,\epsilon}, r'_{i,\epsilon}) \in \Pi_t''(x_i)$ be scaled versions of the parametric representations $(u_{i,\epsilon}, r_{i,\epsilon})$ on $[\delta, 1]$ with $(u'_{i,\epsilon}(s), r'_{i,\epsilon}(s)) = (0,0)$, $0 \leq s \leq \delta$, i.e.,

$$(u'_{i,\epsilon}(\delta + s), r'_{i,\epsilon}(\delta + s)) = (u_{i,\epsilon}((1-\delta)^{-1}s), r_{i,\epsilon}((1-\delta)^{-1}s)), \quad 0 \leq s \leq 1 - \delta . \tag{7.13}$$

Then

$$\|u'_{1,\epsilon} - u'_{2,\epsilon}\| \vee \|r'_{1,\epsilon} - r'_{2,\epsilon}\| = \|u_{1,\epsilon} - u_{2,\epsilon}\| \vee \|r_{1,\epsilon} - r_{2,\epsilon}\| . \tag{7.14}$$

Since $x \in C$, $(u''_{i,\epsilon}, r'_{i,\epsilon}) \in \Pi''(x)$ for $i = 1, 2$, if

$$u''_{i,\epsilon}(s) = \begin{cases} x \circ r'_{i,\epsilon}, & \delta \le s \le 1 \\ 0, & s = 0 \end{cases}$$

with $u''_{i,\epsilon}$ defined by linear interpolation on $(0, \delta)$. Then $(u'_{i,\epsilon} - u''_{i,\epsilon}, r'_{i,\epsilon}) \in \Pi''(x_i - x)$ and

$$\begin{aligned} d''_t(x_1 - x, x_2 - x) &\le \|(u'_{1,\epsilon} - u''_{1,\epsilon}) - (u'_{2,\epsilon} - u''_{2,\epsilon})\| \vee \|r'_{1,\epsilon} - r'_{2,\epsilon}\| \\ &\le (\|u'_{1,\epsilon} - u'_{2,\epsilon}\| + \|x \circ r_{1,\epsilon} - x \circ r_{2,\epsilon}\|) \vee \|r'_{1,\epsilon} - r'_{2,\epsilon}\| \\ &\le (1 + K)(1 + \eta(\epsilon)) d''_t(x_1, x_2) \,. \end{aligned}$$

Since $\eta(\epsilon) \to 0$ as $\epsilon \to 0$, the proof is complete. ∎

We now obtain the M''_1-analog of Theorem 9.2.3. By Lemma 9.7.2, the M'_1 and M''_1 topologies agree on $D([0, \infty), \mathbb{R})$.

Theorem 9.7.1. *Suppose that* $x, x_\epsilon \in D_u([0, \infty), \mathbb{R})$ *and that* x *satisfies the condition of Lemma 9.7.5. If* $d_t(x_\epsilon - x, \epsilon y) = o(\epsilon)$ *as* $\epsilon \to 0$ *for* t *in a sequence* $\{t_k\}$ *with* $t_k \to \infty$ *as* $k \to \infty$*, for which a sufficient condition is* $\|\epsilon^{-1}(x_\epsilon - x) - y\|_t \to 0$ *as* $\epsilon \downarrow 0$ *for all* $t > 0$*, then*

$$d'(\epsilon^{-1}[x_\epsilon^{-1} - x^{-1}], \epsilon^{-1}[(x + \epsilon y)^{-1} - x^{-1}]) \to 0 \quad as \quad \epsilon \downarrow 0 \,, \tag{7.15}$$

where d' *is the* M'_1 *metric on* $D([0, \infty, \mathbb{R})$.

Proof. For any $t_2 > 0$, choose t_1 such that $d_{t_1}(x_\epsilon - x, \epsilon y) = o(\epsilon)$ as $\epsilon \downarrow 0$ and $x^{-1}(t_2) < t_1$. The assumptions imply that $\|x_\epsilon - x\|_{t_1} \to 0$ and $\|\epsilon y\|_{t_1} \to 0$ as $\epsilon \downarrow 0$. Hence, for all sufficiently small ϵ, x_ϵ, $x + \epsilon y \in D_u(t_1, t_2)$. On $D_u(t_1, t_2)$, we can apply Lemmas 9.7.4 and 9.7.6, and the M''_1 analog of Lemma 9.2.2 to conclude for $\epsilon \le 1$ that there are constants K_1 and K_2 such that

$$\begin{aligned} & d''_{t_2}(\epsilon^{-1}[x_\epsilon^{-1} - x^{-1}], \epsilon^{-1}[(x + \epsilon y)^{-1} - x^{-1}]) \\ & \qquad \le \epsilon^{-1} d''_{t_2}(x_\epsilon^{-1} - x^{-1}, (x + \epsilon y)^{-1} - x^{-1}) \\ \le & \qquad K_1 \epsilon^{-1} d''_{t_2}(x_\epsilon^{-1}, (x + \epsilon y)^{-1}) \\ \le & \qquad K_1 \epsilon^{-1} d_{t_1}(x_\epsilon, x + \epsilon y) \\ \le & \qquad K_1 K_2 \epsilon^{-1} d_{t_1}(x_\epsilon - x, \epsilon y) \\ \le & \qquad K_1 K_2 \|(\epsilon^{-1}(x_\epsilon - x) - y\|_{t_1} \,. \end{aligned} \tag{7.16}$$

This argument applies for arbitrarily large t_2 provided that we increase t_1 appropriately. ∎

We now focus on the derivative of the inverse map. Let

$$z_\epsilon \equiv z_\epsilon(x,y) \equiv \epsilon^{-1}[(x+\epsilon y)^{-1} - x^{-1}] \; . \tag{7.17}$$

We first observe that z_ϵ in (7.17) is monotone decreasing in y.

Lemma 9.7.7. *For any $x \in D_u$ and $y \in D$, if $y_1(t) \le y_2(t)$ for all t, then $z_\epsilon(x,y_1)(t) \ge z_\epsilon(x,y_2)(t)$ for all ϵ and t, where z_ϵ is defined in* (7.17).

We now show that it suffices to consider piecewise-constant functions y, because under regularity conditions, $z_\epsilon(x,y)$ as a function of y is Lipschitz. Hence, for x and y given, we can replace y by $y_c \in D_c$.

Lemma 9.7.8. *Suppose that $x \in D_u$, $\dot{x} \in D$, $y_1 \in D$, $t_1 = x^{-1}(t_2)+1$, $0 < a \le \|\dot{x}\|_{t_1} < \infty$ and $\|y_1\|_{t_1} \le K$. If $\|y_1 - y_2\|_{t_1} < 1$, then*

$$\|z_\epsilon(x,y_1) - z_\epsilon(x,y_2)\|_{t_2} \le (2/a)\|y_1 - y_2\|_{t_1}$$

provided that $\epsilon \le a/[K+1]$.

Proof. By the monotonicity established in Lemma 9.7.7,

$$z_\epsilon(x,y_1-\delta) \ge z_\epsilon(x,y_1), z_\epsilon(x,y_2) \ge z_\epsilon(x,y_1+\delta)$$

on $[0,t]$ provided that $\|y_1 - y_2\|_{t_1} \le \delta$ for a suitably large t_1. For the given t_1 and $\delta \le 1$,

$$\begin{aligned}(x+\epsilon y_i)^{-1}(t) &\le (x+\epsilon(y_1-\delta))^{-1}(t)\\ &\le (x-\epsilon(K+\delta))^{-1}(t)\\ &\le x^{-1}(t) + \frac{\epsilon(K+\delta)}{a} \le t_1\end{aligned}$$

provided that $\delta \le 1$ and $\epsilon \le a/(K+1)$. Hence, if $\|\dot{x}\|_{t_1} \ge a$ and $\|y_1\|_{t_1} \le K$ for that t_1, the inverses are all contained in $[0,t_1]$. Then, for $\|y_1 - y_2\|_{t_1} \le \delta \le 1$,

$$\begin{aligned}\|z_\epsilon(x,y_1) - z_\epsilon(x,y_2)\|_{t_2} &\le \|z_\epsilon(x,y_1+\delta) - z_\epsilon(x,y_1-\delta\|_{t_2}\\ &= \epsilon^{-1}\|(x+\epsilon(y_1-\delta))^{-1} - (x+\epsilon(y_1+\delta))^{-1}\|_{t_2}\\ &\le x^{-1}(t_2) + 2\delta/a \; . \quad \blacksquare\end{aligned}$$

We now establish pointwise convergence. For this purpose, let

$$Pos(x) = \{t \ge 0 : x(t) > 0\} \; . \tag{7.18}$$

We obtain the following result by examining the indicated cases.

Theorem 9.7.2. *If $y \in D$ and $x \in D_u$ satisfies the condition of Lemma 9.7.5, then*

$$z_\epsilon(t) \equiv \epsilon^{-1}[(x+\epsilon y)^{-1}(t) - x^{-1}(t)] \to z(t) \quad in \quad \mathbb{R} \quad as \quad \epsilon \downarrow 0$$

for each t, where

$$(i) \quad z(t) = \frac{-y(x^{-1}(t)-)}{\dot{x}(x^{-1}(t)-)} < 0 \tag{7.19}$$

if $y(x^{-1}(t)-) > 0$;

$$(ii) \quad z(t) = \frac{-y(x^{-1}(t))}{\dot{x}(x^{-1}(t))} > 0 \tag{7.20}$$

if $y(x^{-1}(t)-) < 0$ and $y(x^{-1}(t)) < 0$ or if $y(x^{-1}(t)-) = 0$, $\sup\{Pos(y \circ x^{-1}) \cap [0,t)\} < t$ and $y(x^{-1}(t)) < 0$;

$$(iii) \quad z(t) = 0 \tag{7.21}$$

otherwise: if one of: (a) $y(x^{-1}(t)-) = 0$ and $\sup\{Pos(y \circ x) \cap [0,t)\} = t$, (b) $y(x^{-1}(t)-) < 0$ and $y(x^{-1}(t)) = 0$, (c) $y(x^{-1}(t)-) = 0$, $\sup\{Pos(y \circ x) \cap [0,t)\} < t$ and $y(x^{-1}(t)) = 0$, or (d) $y(x^{-1}(t)-) < 0 < y(x^{-1}(t))$.

Consequently, z is either left-continuous or right-continuous at t unless $y(x(t)-) < 0 < y(x(t))$, in which case $z(t-) > z(t) > z(t+)$.

Proof. It is elementary that $z_\epsilon(t)$ converges pointwise to $z(t)$ for $z(t)$ in (7.20) when both $\dot{x}$ and y are continuous at $x^{-1}(t)$, so that z is continuous at t. For the other cases, we apply Lemma 9.7.8 to approximate y by a piecewise-constant function. We then exploit Lemma 9.7.7 and the fact that $\dot{x}$ and y are elements of D. We obtain the conclusions by examining the different cases. ∎

Remark 9.7.1. In order to have the pointwise convergence in Theorem 9.7.2, at a single t, it suffices to have the conditions on x and y hold only in a neighborhood of $x^{-1}(t)$. Then x need not be absolutely continuous or strictly increasing everywhere.

Remark 9.7.2. We have difficulty at some t if x is only an increasing homeomorphism of $[0,\infty)$. Then we can have $\dot{x}(x^{-1}(t)) = 0$ and $\dot{x}^{-1}(t)) = \infty$ for some t, so that $z_\epsilon(t) \to \infty$ as $\epsilon \downarrow 0$.

We now want to establish M_1' convergence in D. However, first we note that the limit z does not necessarily belong to D, because it may be neither left-continuous nor right-continuous at discontinuity points.

Example 9.7.1. *We need not have $z \in D$.* To see that we need not have $z \in D$, even if $\dot{x} \in C$, let $x = e$ and let $y = -I_{[0,1)} + I_{[1,\infty)}$. Then

$$z_\epsilon(t) = I_{[0,1-\epsilon)}(t) + \epsilon^{-1}(1-t)I_{[1-\epsilon,1+\epsilon)}(t) - I_{[1+\epsilon,\infty)}(t) \tag{7.22}$$

and

$$z = I_{[0,1)} - I_{(1,\infty)} \tag{7.23}$$

so that $z(1) = 0$, but $z(1-) = 1$ and $z(1+) = -1$. However, $z(1)$ is in between $z(1-)$ and $z(1+)$. ∎

Since $z(t)$ lies between $z(t-)$ and $z(t+)$ for all t, the space D^* of such functions with the M_1 and M', topologies is equivalent to D because functions in D and D^* have the same graphs.

Theorem 9.7.3. (conditions for convergence to the right-continuous version) *If $y \in D$ and $x \in D_u$ satisfies the condition of Lemma 9.7.5 with $\dot{x} \in D$, then*

$$z_\epsilon \to z_+ \quad in \quad (D, M_1') \quad as \quad \epsilon \downarrow 0$$

for z_ϵ in (7.17) *and z_+ the right-continuous version of z, i.e. $z_+(t) = z(t+)$, $t \geq 0$ and z in* (7.19). *If $z_+(0) = 0$, the convergence is in M_1.*

Proof. First, for x and y given, with $\dot{x}$ satisfying the conditions of Lemma 9.7.5, the conditions of Lemma 9.7.8 are satisfied. Since $\dot{x} \in D$ and $y \in D$, $z \in D^*$ for z defined in (7.19). Start by replacing z by its right-continuous version, which has the same graph. Invoking Lemma 9.4.1, for any $t > 0$, let $\tilde{z} \in D_c$ be such that $\|z - \tilde{z}\|_t \leq \delta_1$. Suppose that $x^{-1}(t_1)$ and $x^{-1}(t_2)$ are two successive discontinuity points of y (where $t_1, t_2 < t$), regarded as an element of D_c. Suppose that $y(s) = c > 0$ in $[x^{-1}(t_1), x^{-1}(t_2))$. Then, for any $\delta_2 > 0$, $z_\epsilon(s) \uparrow z(s)$ in $(t_1 + \delta_2, t_2 - \delta_2)$. Since z_ϵ and $\tilde{z} + \delta_1$ are both continuous in $(t_1 + \delta_2, t_2 - \delta_2)$, we can apply Dini's theorem to conclude that $z_\epsilon(s) \wedge \tilde{z}(s) - \delta_1$ converges uniformly to $\tilde{z}(s) - \delta_1$ in $(t_1 + \delta_2, t_2 - \delta_2)$. Similarly, if $y(s) = c < 0$ in $[t_1, t_2)$, then we can conclude that $z_\epsilon(s) \vee (\tilde{z}(s) + \delta_1)$ converges uniformly to $\tilde{z}(s) + \delta_1$ in $(t_1 + \delta_2, t_2 - \delta_2)$. It thus suffices to establish local M_1 convergence at each of the isolated discontinuity points of $\tilde{z}$; see Theorem 6.5.1. However, z_ϵ is monotone in a neighborhood of each of these discontinuity points for all sufficiently small ϵ. Together with the pointwise convergence at all continuity points established in Theorem 9.7.2, this implies the required local M_1 convergence. To get the strengthened convergence to M_1, apply Theorem 13.6.3 in the book. ∎

The derivative result in Theorem 9.7.3 holds for arbitrary $y \in D$. By applying Theorem 9.7.1, we obtain a corresponding preservation result, but only under the extra condition of uniform convergence of $\epsilon^{-1}(x_\epsilon - x)$ to y as $\epsilon \downarrow 0$, which holds if $y \in C$.

Below let U be the topology on $D([0,\infty),\mathbb{R})$ of uniform convergence over compact subsets.

Corollary 9.7.1. *Suppose that x_ϵ, $x \in E$. Under the conditions of Theorem 9.7.3, if $\|\epsilon^{-1}(x_\epsilon - x) - y\|_t \to 0$ as $\epsilon \downarrow 0$ for all $t > 0$, then*

$$\epsilon^{-1}(x_\epsilon^{-1} - x^{-1}) \to z_+ \quad in \quad (D, M_1') \quad as \quad \epsilon \downarrow 0$$

for z_+ as in Theorem 9.7.3.

9.8. Chapter Notes

As indicated at the outset, this chapter is largely based on Mandelbaum and Massey (1995). They formulate convergence preservation in terms of the directional derivative. We focus on the second term of the triangle inequality in (2.1). Thus The results in Section 9.2 here are new. It would be nice if the upper bound $K\epsilon^{-1}d_1(x_\epsilon - x, y)$ in Theorem 9.2.3 could be replaced by $Kd_1(\epsilon^{-1}(x_\epsilon - x), y)$ under reasonable regularity conditions. The existing bound in terms of $K\epsilon^{-1}d_1(x_\epsilon - x, y)$ may be suitable for applying strong approximations. It thus also would be nice to develop such strong approximations to apply with Theorem 9.2.3 here.

Section 9.3 on the derivative of the supremum function is also based on Mandelbaum and Massey (1995). We provide extensions allowing the functions x and y appearing in $z_\epsilon(x, y)$ in (3.2) to be discontinuous. We also do not require that the limit z have only finitely many discontinuities in each finite interval. The arguments are quite a bit more complicated as a result. Some simplification is achieved here by exploiting approximations by piecewise-constant functions. In particular, for establishing M_1 convergence, Lemma 9.4.2 is key.

Given the intimate connection between the reflection and supremum maps, most of the work on the derivative of the reflection map in Section 9.5 is done in Sections 9.3 and 9.4. The application of Sections 9.3 – 9.5 in Section 9.6 to obtain heavy-traffic limits for nonstationary queues also follows Mandelbaum and Massey (1995). They focused on the $\mathrm{M_t/M_t/1}$ queue with fixed arrival-rate and service-rate functions $\lambda^+(t)$ and $\lambda^-(t)$, drawing on the strong approximation for Poisson processes. We show how the results can be generalized by applying convergence-preservation results for the composition function with nonlinear centering in Chapter 13 of the book.

Section 9.7 on the derivative of the inverse function is new. The M_1'' topology extends the M_1' topology introduced in Puhalskii and Whitt (1997).

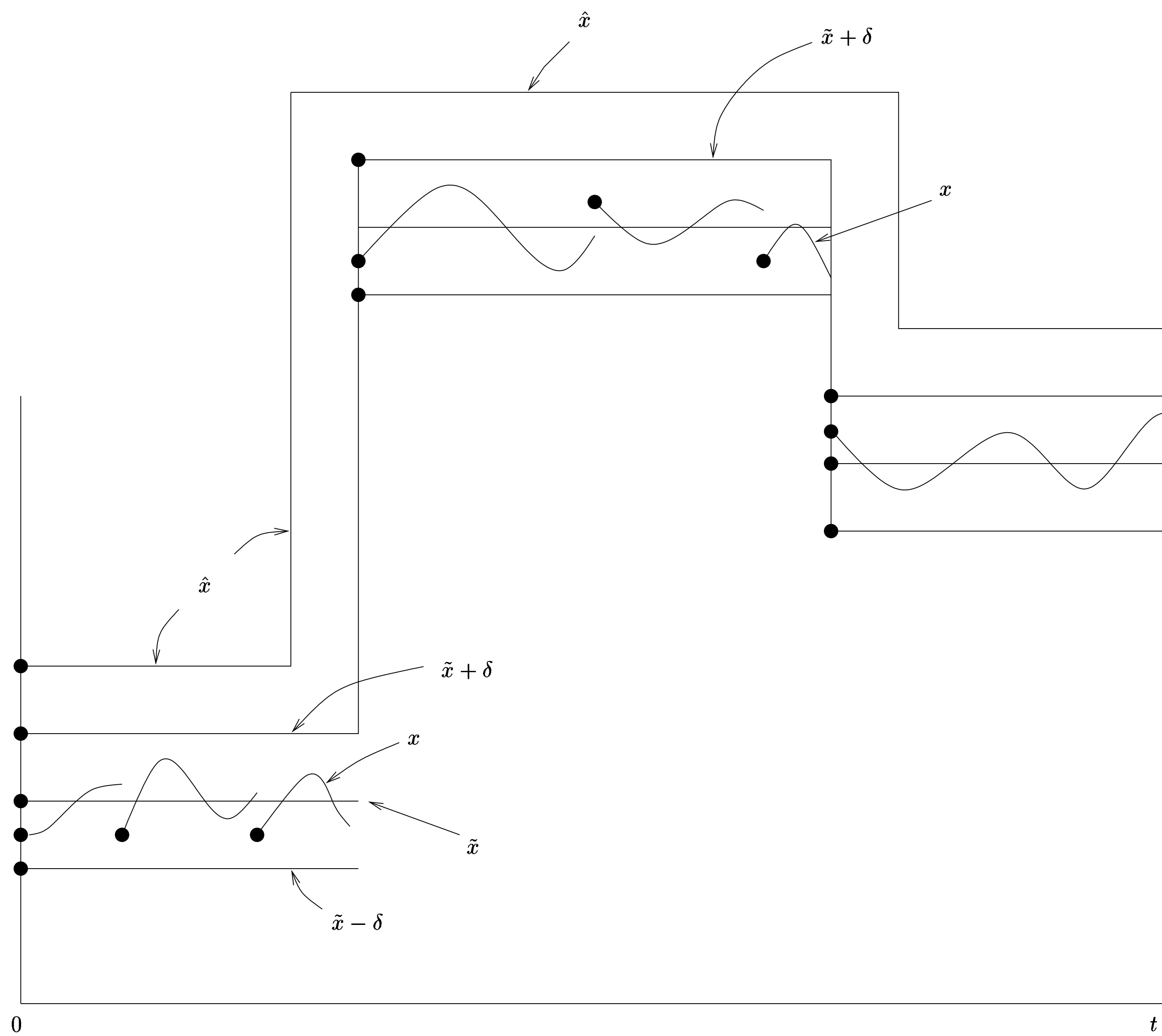

Figure 9.1: A possible function x, piecewise-constant approximation $\tilde{x}$, upper bound $\tilde{x}+\delta$ and upper boundary $\hat{x}$ of the δ-neighborhood of the graph $\Gamma_{\tilde{x}+\delta}$ used in the proof of Theorem 9.4.1.

Figure 9.2: Graphs of the time-dependent arrival-rate and service-rate functions $(\lambda^+(t), \lambda^-(t))$ with λ^- constant, the functions $(-x, (-x)^\uparrow)$, the set-valued function Φ_{-x} and the limits q and Q for a typical realization of the $\mathrm{M_t/M_t/1}$ queue.

Chapter 10

Errors and Comments After Publication

This chapter contains a list of errors in the book found since publication. It also contains comments on related work.

p. 80 While $\|x_n - x\| = 1$ in Figure 3.1, $d_{J_1}(x_n, x) = 1/2$, not 1, as claimed just below Figure 3.1. [Anton Kleywegt, Georgia Tech]

p. 100 third paragraph: A real-valued random variable (no s). [Aubin Whitley, University of California at San Diego]

p. 114-115 format $\frac{F^c(x)}{G^c(x)}$ should be used consistently, in (5.22) as well as (5.27).

p. 124 Formula (6.12) is incorrect. It should be

$$C_2 = \frac{c^2 \Gamma(1-\gamma)\Gamma(2\gamma - 1)}{\Gamma(\gamma)(3-2\gamma)(1-\gamma)} \ .$$

Lemma 4.6.1 applies; (6.12) was calculated incorrectly given Lemma 4.6.1 and formula (6.9). In Example 4.6.1 this makes the coefficient three times smaller when $\gamma = 0.75$. [Michael Roginsky, Berkeley]

p. 126 The direct simulation method described here tends to be inefficient. Michael Roginsky at Berkeley has been investigating alternative procedures, including the Choleski decomposition, which directly constructs normal random variables with the prescribed correlations. A method based on wavelet decompositions is described in Chapter 2 of Park and Willinger (2000). [Michael Roginsky, Berkeley]

p. 297 A finite-waiting room version of Theorem 9.3.4 is established in

W. Whitt, Heavy-traffic limits for loss proportions in single-server queues, 2002.

This paper can be downloaded from the "Recent Papers" section on my homepage.

p. 358 Formulas (4.15) and (4.19) are correct for the $M/M/m$ model, where $\xi = \beta$, but they are incorrect for the $GI/M/m$ model. These errors appear in Halfin and Whitt (1981) and were perpetuated in Whitt (1993) and here. Formula (4.15) should be

$$p \equiv p(\beta, z) = \alpha(\beta/\sqrt{z}) \ ,$$

where

$$\alpha(\beta) \equiv p(\beta, 1) = [1 + \beta\Phi(\beta)/\phi(\beta)]^{-1}$$

and

$$z = (1 + c_a^2)/2 \ .$$

Formula (4.19) should be

$$P(Z \le x | Z \le 0) = \Phi((x + \beta)/\sqrt{z})/\Phi(\beta/\sqrt{z})$$

for β in (4.14) and z above.

See the papers:

W. Whitt, Heavy-traffic limits for the $G/H_2^*/n/m$ queue, 2002.

W. Whitt, A diffusion approximation for the $G/GI/n/m$ queue, 2002.

These papers can be downloaded from the "Recent Papers" section on my homepage.

p. 360 Condition (4.23) should read

$$(m-\gamma)/\sqrt{\gamma} \to \beta \quad \text{for} \quad -\infty < \beta < \infty$$

[Garud Iyengar, Columbia]

Chapter 11

Bibliography

Bibliography

[1] Abate, J., Choudhury, G. L. and Whitt, W. (1999) An introduction to numerical transform inversion and its application to probability models. *Computational Probability*, W. Grassman (ed.), Kluwer, Boston, 257–323.

[2] Abate, J. and Whitt, W. (1992a) The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems* 10, 5–88.

[3] Abate, J. and Whitt, W. (1992b) Numerical inversion of probability generating functions. *Operations Res. Letters* 12, 245–251.

[4] Abate, J. and Whitt, W. (1995) Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Computing* 7, 36–43.

[5] Abate, J. and Whitt, W. (1996) An operational calculus for probability distributions via Laplace transforms. *Adv. Appl. Prob.* 28, 75–113.

[6] Albert, A. (1966) Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* 37, 1602–1630.

[7] Aldous, D. (1978) Stopping times and tightness. *Ann. Prob.* 6, 335–340.

[8] Anscombe, F. J. (1952) Large sample theory for sequential estimation. *Proc. Cambrdige Philos. Soc.* 48, 600–607.

[9] Anscombe, F. J. (1953) Sequential estimation. *J. Roy. Statist. Soc. Ser. B* 15, 1–21.

[10] Asmussen, S. (1987) *Applied Probability and Queues*, Wiley, New York.

[11] Baccelli, F. and Brémaud, P. (1994) *Elements of Queueing Theory*, Springer, New York.

[12] Beran, J. (1994) *Statistics for Long-Memory Processes*, Chapman and Hall, New York.

[13] Bertoin, J. (1996) *Lévy Processes*, Cambridge University Press, Cambridge, UK.

[14] Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley, New York.

[15] Billingsley, P. (1999) *Convergence of Probability Measures*, second edition, Wiley, New York.

[16] Bingham, N. H. (1975) Fluctuation theory in continuous time. *Adv. Appl. Prob.* 7, 705–766.

[17] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989) *Regular Variation*, Cambridge University Press, Cambridge, UK.

[18] Bondesson, L. (1992) *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, Springer, New York.

[19] Box, G. E. P. and Jenkins, G. M. (1970) *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco.

[20] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994) *Time Series Analysis, Forecasting and Control*, third edition, Prentice Hall, NJ.

[21] Bratley, P. , Fox, B. L. and Schrage, L. E. (1987) *A Guide to Simulation*, second edition, Springer, New York.

[22] Chen, H. and Mandelbaum, A. (1994) Heirarchical modeling of stochastic networks, part II: strong approximations. *Stochastic Modeling and Analysis of Manufacturing Systems*, D. D. Yao, ed., Springer-Verlag, New York, 107–131.

[23] Chen, H. and Shen, X. (2000) Strong approximations for multiclass feedforward queueing networks. *Ann. Appl. Prob.* 10, 828–876.

[24] Chen, H. and Yao, D. D. (2001) *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization*, Springer, New York.

[25] Choudhury, G. L., Lucantoni, D. M. and Whitt, W. (1994) Multidimensional transform inversion with application to the transient $M/G/1$ queue. *Ann. Appl. Prob.* 4, 719–740.

[26] Chow, Y. S. and Robbins, H. (1965) On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36, 457–462.

[27] Crane, M. A. and Lemoine, A. J. (1977) *An Introduction to the Regenerative Method for Simulation Analysis. Lecture Notes in Control and Inform. Sci.*, Springer, New York.

[28] Csörgő, M. and Horvath, L. (1993) *Weighted Approximations in Probability and Statistics*, Wiley, New York.

[29] Csörgő, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics*, Academic Press, New York.

[30] Damerdji, H. (1994) Strong consistency of the variance estimator in steady-state simulation output analysis. *Math. Oper. Res.* 19, 494–512.

[31] Damerdji, H. (1995) Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Res.* 43, 282–291.

[32] Dudley, R. M. (1968) Distances of probability measures and random variables. *Ann. Math. Statist.* 39, 1563–1572.

[33] Dunford, N. and Schwartz, J. T. (1958) *Linear Operators. Part I: General Theory.* Interscience, New York.

[34] Einmahl, U. (1989) Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Anal.* 28, 20–68.

[35] El-Taha, M. and Stidham, S., Jr. (1999) *Sample-Path Analysis of Queueing Systems*, Kluwer, Boston.

[36] Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes, Characterization and Convergence*, Wiley, New York.

[37] Feller, W. (1968) An Introduction to Probability Theory and its Applications, vol. I, third edition, Wiley, New York.

[38] Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, vol. II, second edition, Wiley, New York.

[39] Fishman, G. S. (1977) Achieving specific accuracy in simulation output analysis. *Comm. ACM* 20, 310–315.

[40] Fox, B. L. and Glynn, P. W. (1989) Replication schemes for limiting expectations. *Prob. Eng. Inf. Sci.* 3, 299–318.

[41] Gleser, L. J. (1965) On the asymptotic theory of fixed-size confidence bounds for linear regression parameters. *Ann. Math. Statist.* 36, 463–467. [Correction note: 37 (1966) 1053–1055].

[42] Glynn, P. W. (1994) Poisson's equation for the recurrent M/G/1 queue. *Adv. Appl. Prob.* 26, 1044–1062.

[43] Glynn, P. W. and Heidelberger, P. (1989) Jackknifing under a budget constraint. Unpublished manuscript, Dept. Operations Research, Stanford Univ.

[44] Glynn, P. W. and Iglehart, D. L. (1988) A new class of strongly consistent variance estimators for a steady-state simulation. *Stochastic Process. Appl.* 16, 71–80.

[45] Glynn, P. W. and Meyn, S. (1996) A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* 24, 916–931.

[46] Glynn, P. W. and Whitt, W. (1987) Sufficient conditions for functional-limit-theorem versions of $L = \lambda W$. *Queueing Systems* 1, 279–287.

[47] Glynn, P. W. and Whitt, W. (1988) Ordinary CLT and WLLN versions of $L = \lambda W$. *Math. Oper. Res.* 13, 674–692.

[48] Glynn, P. W. and Whitt, W. (1989) Indirect estimation via $L = \lambda W$. *Operations Res.* 37, 82–103.

[49] Glynn, P. W. and Whitt, W. (1991a) A new view of the heavy-traffic limit theorem for the infinite-server queue. *Adv. Appl. Prob.* 23, 188–209.

[50] Glynn, P. W. and Whitt, W. (1991b) Departures from many queues in series. *Ann. Appl. Prob.* 1, 546–572.

[51] Glynn, P. W. and Whitt, W. (1992a) The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Prob.* 2, 180–198.

[52] Glynn, P. W. and Whitt, W. (1992b). The asymptotic efficiency of simulation estimators. *Operations Res.* 40, 505–520.

[53] Glynn, P. W. and Whitt, W. (1993) Limit theorems for cumulative processes. *Stoch. Proc. Appl.* 47, 299–314.

[54] Glynn, P. W. and Whitt, W. (2000) N&S conditions for the Markov chain CLT. AT&T Labs, Florham Park, NJ.

[55] Gnedenko, B. V. and Kolmogorov, A. N. (1968) *Limit Distributions for Sums of Independent Random Variables*, revised edition, Addison Wesley, Reading, MA.

[56] Govindarajulu, Z. (1987) *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*, 2nd ed. American Sciences Press, Columbus, Ohio.

[57] Granger, C. W. J. and Joyeux, R. (1980) An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.* 1, 15–30.

[58] Gut, A. (1988) *Stopped Random Walks*, Springer-Verlag, New York.

[59] Halmos, P. R. (1956) *Measure Theory* Van Nostrand, Princeton, NJ.

[60] Heidelberger, P. and Welch, P. D. (1981a) A spectral method for confidence interval generation and run length control in simulations. *Comm. ACM* 24, 233–245.

[61] Heidelberger, P. and Welch, P. D. (1981b) Adapative spectral methods for simulation output analysis. *IBM J. Res. Develop.* 25, 860–876.

[62] Heidelberger, P. and Welch, P. D. (1983) Simulation run length control in the presence of an initial transient. *Oper. Res.* 31, 1109–1144.

[63] Hosking, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

[64] Jacod, J. and Shiryaev, A. N. (1987) *Limit Theorems for Stochastic Processes*, Springer-Verlag, New York.

[65] Kella, O. (1993) Parallel and tandem fluid networks with dependent Lévy inputs. *Ann. Appl. Prob.* 3, 682–695.

[66] Kella, O. (1996) Stability and nonproduct form of stochastic fluid networks with Lévy inputs. *Ann. Appl. Prob.* 6, 186–199.

[67] Kella, O. and Whitt, W. (1992a) A tandem fluid network with Lévy input. In *Queues and Related Models*, I. Basawa and N. V. Bhat (eds.), Oxford University Press, Oxford, 112–128.

[68] Kella, O. and Whitt, W. (1992b) Useful martingales for stochastic storage processes with Lévy input, *J. Appl. Prob.* 29, 396–403.

[69] Kella, O. and Whitt, W. (1996) Stability and structural properties of stochastic storage networks. *J. Appl. Prob.* 33, 1169–1180.

[70] Kemeny, J. G. and Snell, J. L. (1960) *Finite Markov Chains*, Van Nostrand, Princeton.

[71] Kemeny, J. G. and Snell, J. L. (1961) Finite continuous time Markov chains. *Theor. Probability Appl.* 6, 101–105.

[72] Kennedy, D. P. (1973) Limit theorems for finite dams. *Stoch. Proc. Appl.* 1, 269–278.

[73] Kogan, Ya., Liptser, R. Sh. and Smorodinskii, A. V. (1986) Gaussian diffusion approximations of Markovian closed models of computer communication networks. *Problemy Peredachi Informatsii* 22, 49–65.

[74] Komlós, J., Major, P. and Tusnády, G. (1975) An approximation of partial sums of independent R.V.'s and the sample DF, I. *Zeitschrift für Wahrscheinlichkeitstheorie verw. Gebiete* 32, 111–131.

[75] Komlós, J., Major, P. and Tusnády, G. (1976) An approximation of partial sums of independent R.V.'s and the sample DF, II. *Zeitschrift für Wahrscheinlichkeitstheorie verw. Gebiete* 34, 33–58.

[76] Kurtz, T. G. (1996) Limit theorems for workload input models. In *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary and I. Ziedins (eds.) Oxford University Press, Oxford, U.K., 119–139.

[77] Kushner, H. J. (2001) *Heavy Traffic Analysis of Controlled and Uncontrolled Queueing and Communication Networks*, Springer, New York.

[78] Lavenberg, S. S. and Sauer, C. H. (1977) Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop.* 21, 545–558.

[79] Law, A. M. and Carson, J. S. (1979) A sequential procedure for determining the length of a steady-state simulation. *Oper Res.* 27, 1011–1025.

[80] Law, A. M. and Kelton, W. D. (1982) Confidence intervals for steady-state simulation. II. A survey of sequential procedures. *Management Sci.* 28, 550–562.

[81] Law, A. M., Kelton, W. D. and Koenig, L. W. (1981) Relative width sequential confidence intervals for the mean. *Comm. Statist. B — Simulation Comput.* 10, 29–39.

[82] Lindvall, T. (1973) Weak convergence in the function space $D[0, \infty)$. *J. Appl. Prob.* 10, 109–121.

[83] Lindvall, T. (1992) *Lectures on the Coupling Method*, Wiley, New York.

[84] Lucantoni, D. M. (1993) The BMAP/G/1 queue: a tutorial. *Models and Techniques for Performance Evaluation of Computer and Communications Systems*, J. Donatiello and R. Nelson (eds.), Springer, New York, 330–358.

[85] Maigret, N. (1978) Théorème de limite centrale fonctionnel pour une chaine de Markov récurrente au sens de Harris et positive. *Ann. Inst. H. Poincaré Sect. B (N. S.)* 14, 425–440.

[86] Mandelbaum, A. and Massey, W. A. (1995) Strong approximations for time-dependent queues. *Math. Oper. Res.* 20, 33–64.

[87] Mandelbaum, A., Massey, W. A. and Reiman, M. I. (1998) Strong approximations for Markovian service networks. *Queueing Systems* 30, 149–201.

[88] Mandelbaum, A., Massey, W. A., Reiman, M. I. and Stolyar, A. (1999) Waiting time asymptotics for time varying multiserver queues with abandonments and retrials. Proceedings of 37^{th} Allerton Conference on Communication, Control and Computing, September 1999, 1095–1104.

[89] Massey, W. A. and Whitt, W. (1994) Unstable asymptotics for nonstationary queues. *Math. Oper. Res.* 19, 267–291.

[90] McLeish, D. L. (1976) Functional and random central limit theorems for the Robbins-Monro process. *J. Appl. Probab.* 13, 148–154.

[91] Miller, R. G. (1964) A trustworthy jackknife. *Ann. Math. Statist.* 35, 1594–1605.

[92] Miller, R. G. (1974) The jackknife — a review. *Biometrika* 61, 1–15.

[93] Nadas, A. (1969) An extension of a theorem of Chow and Robbins on sequential confidence intervals for the mean. *Ann. Math. Statist.* 40, 667–671.

[94] Neuts, M. F. (1989) *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York.

[95] Neveu, J. (1965) *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco.

[96] Park, K. and Willinger, W. (2000) *Self-Similar Network Traffic and Performance Evaluation*, Wiley, New York.

[97] Parthasarathy, K. R. (1967) *Probability Measures on Metric Spaces*, Academic, New York.

[98] Philipp, W. and Stout, W. (1975) *Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables*, Mem. Amer. Math. Soc. **161**, Providence, RI.

[99] Prabhu, N. U. (1998) *Stochastic Storage Processes*, second ed. Springer, New York.

[100] Prohorov, Yu. V. (1956) Convergence of random processes and limit theorems in probability. *Theor. Probability Appl.* 1, 157–214.

[101] Puhalskii, A. A. (1994) On the invariance principle for the first passage time. *Math. Oper. Res.* 19, 946–954.

[102] Puhalskii, A. A. and Whitt, W. (1997) Functional large deviation principles for first-passage-time processes. *Ann. Appl. Prob.* 7, 362–381.

[103] Resnick, S. and Stărică, C. (1997) Smoothing the Hill estimator. *Adv. Appl. Prob.* 29, 271–293.

[104] Rogers, L. C. G. (2000) Evaluating first-passage probabilities for spectrally one-sided Lévy processes. J. Appl. Prob. 37, 1173–1180.

[105] Ruppert, D. (1982) Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Probab.* 10, 178–187.

[106] Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable NonGaussian Random Processes*, Chapman-Hall, New York.

[107] Siegmund, D. (1985) *Sequential Analysis.* Springer, New York.

[108] Simmons, G. F. (1963) *Topology and Modern Analysis*, McGraw-Hill, New York.

[109] Skorohod, A. V. (1956) Limit theorems for stochastic processes. *Theor. Probability Appl.* 1, 261–290. (also republished in *Skorohod's Ideas in Probability Theory*, V. Korolyuk, N. Portenko and H. Syta (eds.), institute of Mathematics of the National Acafdemy of Sciences of the Ukraine, Kyiv, Ukraine, 23–52.)

[110] Skorohod, A. V. (1957) Limit theorems for stochastic processes with independent increments. *Theor. Probability Appl.* 2, 138–171.

[111] Skorohod, A. V. (1961) *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, MA.

[112] Srivastava, M. S. (1967) On fixed-width confidence bounds for regression parameters and the mean vector. *J. Roy. Statist. Soc. Ser. B* 29, 132–140.

[113] Starr, N. (1966) The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* 37, 36–50.

[114] Strassen, V. (1965) The existence of probability measures with given marginals. *Ann. Math. Statist.* 36, 423–439.

[115] Takács, L. (1967) *Combinatorial Methods in the Theory of Stochastic Processes*, Wiley, New York.

[116] Thorin, O. (1977a) On the infinite divisibility of the Pareto distributions. *Scand. Act. J.* 60, 31–40.

[117] Thorin, O. (1977b) On the infinite divisibility of the lognormal distribution. *Scand. Act. J.* 60, 121–148.

[118] Venter, J. H. (1967) An extension of the Robbins-Monro procedure. *Ann. Math. Statist.* 38, 181–190.

[119] Wetherill, G. B. and Glazebrook, K. D. (1986) *Sequential Methods in Statistics*, 3rd ed., Chapman and Hall, London.

[120] Whitt, W. (1974) Preservation of rates of convergence under mappings. *Zeitschrift für Wahrscheinlichskeitstheorie verw. Gebiete* 29, 39–44.

[121] Whitt, W. (1980) Some useful functions for functional limit theorems. *Math. Oper. Res.* 5, 67–85.

[122] Whitt, W. (1991) A review of $L = \lambda W$ and extensions. *Queueing Systems* 9, 235–268. (Correction Note on $L = \lambda W$. *Queueing Systems* 12, 431–432.)

[123] Whitt, W. (1992a) Asymptotic formulas for Markov processes with applications to simulation. *Operations Res.* 40, 279–291.

[124] Whitt, W. (1992b) $H = \lambda G$ and the Palm transformation. *Adv. Appl. Prob.* 24, 755–758.

[125] Whitt, W. (2001) The reflection map with discontinuities. *Math. Oper. Res.* 26, 447–484.

[126] Wichura, M. J. (1970) On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* 41, 284–291.

Index