2 Lecture 2

2.1 Defining Optimization with Equality Constraints

So far we have been concentrating on an arbitrary set X. Because of this, we could of course incorporate constrains directly into the set. This, however, turns out to be a bad way of dealing with constraints. If you look at the above theorems, they all require finding out what a 'feasible direction' is for any $x \in X$. If we are not in the interior of x, this is a pain. Much better to deal with constraints explicitly, in the form of a constrained optimization problem. We will start off dealing with equality constraints, then move on to inequality constraints.

Problem 5 (Constrained Optimization with Equality Constraints) *Let $f : \mathbb{R}^n \to \mathbb{R}$. A constrained optimization problem is the following:*

- *Find $x \in \mathbb{R}^n$*
- *In order to maximize $f(x)$*
- *Subject to*

$$\begin{aligned} h_1(x) &= 0 \\ h_2(x) &= 0 \\ &\vdots \\ h_m(x) &= 0 \end{aligned}$$

or

$$h(x) = 0$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$

In general, we will assume that h_i are $\mathcal{C}^1$.

In other words, this is an 'unconstrained' optimization problem where

$$X = \{x \in \mathbb{R}^n | h(x) = 0\}$$

2.2 The Tangent Plane

In order to take on this problem, we are going to need to define two new objects:

Definition 2 *A parametric curve on X is the image of a continuous function $x : (a, b) \to X$. The curve is differentiable if*

$$\dot{x} = \frac{d}{dt} x(t)$$

exists for all $t \in (a, b)$.

The curve $x : (a, b) \to X$ passes through x^* if there exists a $t^* \in (a, b)$ such that $x(t^*) = x^*$

Definition 3 *The tangent plane to X at x^* is*

$$T = \{\dot{x}(t^*) | x : (a, b) \to X, x(t^*) = x^*\}$$

In other words, the tangent plane is the derivative of all parameterized curves that go through x^*. This may not seem like a very intuitive definition, but it is getting at a very intuitive concept. Let's first think of an example of X in $\mathbb{R}^2$

Example 1 *Let $X \subset \mathbb{R}^2$ be defined by the constraint $x_2 - x_1^2 = 0$. In general, a parametric curve that lives in this space is defined by*

$$f(t) = \begin{pmatrix} g(t) \\ g(t)^2 \end{pmatrix}$$

Where $g(t)$ is some real valued function. For any value of t, $f(t)$ lies inside X. What is the derivative of this function with respect to t?

$$f'(t) = \begin{pmatrix} g'(t) \\ 2g'(t)g(t) \end{pmatrix}$$

Now let's consider a point $x^* = \begin{pmatrix} x^* \\ x^{*2} \end{pmatrix}$ that lives in X. let t^* be the value of t such that $f(t^*) = x^*$. Thus, according to the definition we have above the tangent plane to X at x^* must contain the points $\begin{pmatrix} g'(t^*) \\ 2g'(t^*)g(t^*) \end{pmatrix}$ for any differentiable function g. So for example, if $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, the

the tangent plane contains all vectors $\begin{pmatrix} g'(t^*) \\ 0 \end{pmatrix}$. *If* $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, *then the tangent plane is all vector of the form* $\begin{pmatrix} & 1 \\ g'(t^*) & \\ & 2 \end{pmatrix}$. *Note that we can find a differentiable function such that* $g'(t^*)$ *is any real number (think of linear functions) But what IS this? Well, it is just the equation of hyperplane that is tangent to* X *at* x^**, displaced by* x^**. In other words, the hyperplane that is tangent to* x^* *is defined by* $T + x^*$

The idea of a tangent place just extends this notion to $\mathbb{R}^n$

How do we characterize the tangent plane of X? If the constraints are well behaved, then we have a nice characterization. By 'well behaved', we mean that all the constraints matter.

Definition 4 *A point* x^* *satisfying* $h(x^*) = 0$ *is regular if*

$$\{\nabla h_1(x^*), ..., \nabla h_m(x^*)\}$$

is linearly independents.

If this is the case, then the tangent space can be characterized as the set of vectors that are orthogonal to derivatives of the constraints

Theorem 6 *If* x^* *is regular, then the tangent space to* X *at* x^* *is*

$$T = \left\{d \in \mathbb{R}^n | h'(x^*).d = 0\right\} = \ker(h'(x^*))$$

where

$$\begin{aligned} h'(x^*) &= \left[\frac{\partial h_i}{\partial h_j}(x^*)\right] \in \mathbb{R}^{m \times n} \\ &= \begin{bmatrix} \nabla h_1(x^*)^T \\ \vdots \\ \nabla h_m(x^*)^T \end{bmatrix} \end{aligned}$$

Before we prove this theorem, we need to remember what the implicit function theorem states

Theorem 7 *Let $\theta : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be such that $\theta(x^*, y^*) = 0$ for some $x^* \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^m$. Then if $\theta_x(x^*, y^*)$ is of full rank (and therefore invertible), then there exists a function*

$$\begin{aligned} \zeta \quad &: \quad B(y^*, \varepsilon_1) \to B(x^*, \varepsilon_2) \\ \textit{such that } \theta(\zeta(y), y) \quad &= \quad 0 \ \forall \ y \in B(y^*, \varepsilon_2) \end{aligned}$$

We are not going to prove the theorem, but to get an idea of why full rank is important, consider the following example

Example 2 *Let $\theta : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be defined as*

$$\theta(x_1, x_2, y) = \begin{Bmatrix} ax_1 + bx_2 + y \\ cx_1 + dx_2 + y \end{Bmatrix}$$

Clearly, $\theta(0,0,0) = 0$. Now try to construct an implicit function: i.e a mapping $\zeta : B(0, \varepsilon_1) \to B(0, \varepsilon_2)$ such that

$$\theta(\zeta(y), y) = 0$$

We can figure out what this function must look like by solving

$$\begin{aligned} ax_1 + bx_2 + y = 0 \\ cx_1 + dx_2 + y = 0 \end{aligned}$$

Rearranging tells us that

$$x_2 = \frac{a\left(\frac{c}{a} - 1\right)}{(da - bc)} y$$

which is well defined unless $da = bc$. But notice that the matrix

$$\theta_x(x^*, y^*) = \begin{matrix} a & b \\ c & d \end{matrix}$$

Which will have full rank unless $\frac{a}{c} = \frac{b}{d}$, i.e $da = bc$

We now move on to prove the representation of the tangent space

Proof (Representation of the Tangent Space). We first need to show that

$$T \subset \ker(h'(x^*))$$

Let $x : (a, b) \to X$ be a tangent curve that goes through x^*. Then $h(x(t)) = 0 \; \forall \; t \in (a, b)$. Thus, if x is differentiable then,by chain rule

$$h'(x(t^*))\dot{x}(t^*) = 0 = h'(x^*)\dot{x}(t^*)$$

thus, for every $\dot{x}(t^*)$ such that $x : (a, b) \to X, x(t^*) = x^*$, $h'(x^*)\dot{x}(t^*) = 0$

Next we need to show that

$$\ker(h'(x^*)) \subset T$$

In other words, we need to show that, for any d such that $h'(x^*).d = 0$, there exists some parametric curve x that passes through x^* such that $\dot{x}(t^*) = d$

Consider the following system of equations in t and u

$$h(x^* + td + h'(x^*)^T u) = 0$$

This system of equations has a solution, as $t = t^* = 0, \; \text{u} = u^* = 0$ is one solution, The fact that $h_u(t^*, u^*) = h'(x^*)h'(x^*)^T$ is of full rank follows from the fact that x^* is regular (check). Thus, by the implicit function theorem there exists $(-\varepsilon, \varepsilon)$ and $u : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ such that

$$h(x^* + td + h'(x^*)^T u(t)) = 0$$

for all $t \in (-\varepsilon, \varepsilon)$ and $u(0) = 0$

Our candidate parametric curve is therefore

$$x(t) = x^* + td + h'(x^*)^T u(t)$$

Thus,

$$\begin{aligned} 0 &= \frac{\partial(h(x(t))}{\partial(t)}|_{t=0} = h'(x^*)d + h'(x^*)h'(x^*)^T \dot{u}(0) \\ &= h'(x^*)h'(x^*)^T \dot{u}(0) \end{aligned}$$

which implies $\dot{u}(0) = 0$, and $\dot{x}(0) = d$ ■

Note that, if the point is not regular, then the above relationship may not hold, as the following example demonstrates

Example 3 *Let* $X \subset \mathbb{R}^n$ *be defined by* $h(x) = x_2^3 - x_1^6$*, so* $h'(x) = 3x_2^2 - 6x_1^5$*. At* $x^* = (0, 0)$, $\ker(h'(x^*)) = \{d \in \mathbb{R}^n | h'(x^*).d = 0\} = \mathbb{R}^2$*, which is not equal to the tangent plane at* x^**.*

This should give a hint that regularity is not a property of sets, but of how you define the constraints