3 Lecture 3

3.1 The Kuhn Tucker Conditions for Optimization with Equality Constraints

We will now use the concept of the tangent plane, along with Farkas' Lemma to derive first order conditions for the constrained problem

Lemma 1 *Let x^* be a regular point of $h(x) = 0$ and a local max of f in X. Then the system*

$$\begin{aligned} h'(x^*)d &= 0 \\ \langle \nabla f(x^*), d \rangle &> 0 \end{aligned}$$

has no solution.

Proof. *Assume by contradiction that d solves the above system. Then there exists an $x : (a, b) \to X$ such that $x^* = x(t^*)$ and $\dot{x}(t^*) = d$. Then*

$$\begin{aligned} f(x(t)) &= f(x(t^*)) + \frac{\partial}{\partial t} f(x(t))|_{t=t^*}(t - t^*) + o(t - t^*) \\ &= f(x^*) + \langle \nabla f(x^*), \dot{x}(t^*) \rangle (t - t^*) + o(t - t^*) \\ &= f(x^*) + \langle \nabla f(x^*), d \rangle (t - t^*) + o(t - t^*) \\ &\Rightarrow \frac{f(x) - f(x^*)}{(t - t^*)} = \langle \nabla f(x^*), d \rangle + \frac{o(t - t^*)}{(t - t^*)} \end{aligned}$$

which is greater than zero for t close to t^*, contradicting the idea that x^* is a local maximizer

■

And now, as if by magic, we can use this result and Farkas' lemma to derive the Kuhn-Tucker FOC

Theorem 8 (First Order Necessary Conditions) *Let x^* be a regular point of $h(x) = 0$ and a local maximizer of f in X. Then there exists $\lambda \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{M} \lambda_i \nabla h_i(x^*) = 0$$

Proof. *By the previous lemma, we know that the system of equations*

$$\begin{aligned} d^T h'(x^*) &= 0 \\ \langle \nabla f(x^*), d \rangle &> 0 \end{aligned}$$

has no solution. This implies that the system

$$\begin{aligned} d^T[h'(x^*) \quad &: \quad -h'(x^*)] \leq 0 \\ \langle \nabla f(x^*), d \rangle \quad &> \quad 0 \end{aligned}$$

Does not have a solutions (as this would require both $d^T h'(x^*) \leq 0$ and $-d^T h'(x^*) \leq 0$, and so $d^T h'(x^*) = 0$). By Farkas' lemma, this implies that the system

$$\begin{aligned} [h'(x^*) \quad : \quad -h'(x^*)] \begin{pmatrix} \lambda^+ \\ \lambda^- \end{pmatrix} &= \nabla f(x^*) \\ \lambda^+ \quad &\geq \quad 0 \\ \lambda^- \quad &\geq \quad 0 \text{ has a solution} \end{aligned}$$

Let $\lambda = \lambda^- - \lambda^+$, and this gives the required result ■

You should convince yourself that regularity is important here. Consider maximizing $-x_2$ subject to $x_1^6 - x_2^3 = 0$

As with the unconstrained case, we also have second order necessary conditions. First order necessary conditions find critical points - maxima or minima. The second order conditions help us to narrow down the possible critical points

Theorem 9 (Second Order Necessary Conditions) *Let x^* be regular for $h(x) = 0$ and a local maximizer, then there exists $\lambda \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{M} \lambda_i \nabla h_i(x^*) = 0$$

Moreover, the matrix

$$\Psi(x^*, \lambda) = F(x^*) + \sum \lambda_i H_i(x^*)$$

is negative semi definite in $T = \ker(h'(x^*))$ (i.e. $h'(x^*).d = 0$ implies $d^T \Psi(x^*, \lambda) d \leq 0$)

Note that

$$F(x^*) = \left[\frac{\partial^2 f}{\partial x_j \partial x_k}(x^*) \right]$$

and

$$H_i(x^*) = \left[\frac{\partial^2 h_i}{\partial x_j \partial x_k}(x^*)\right]$$

Proof. *Let $x : (a, b) \to X$ with $x(t^*) = x^*$ and $d = \dot{x}(t^*)$. From 1 dimensional calculus we know that, as $f(x(t))$ has a local maximum at t^*, it must be the case that*

$$0 \geq \frac{\partial^2}{\partial t^2}\left(f(x(t))\right)|_{t=t^*}$$

but

$$\frac{\partial}{\partial t}\left(f(x(t))\right) = \langle \dot{x}(t).\nabla f(x(t))\rangle$$

and so

$$\frac{\partial^2}{\partial t^2}\left(f(x(t))\right) = \langle \ddot{x}(t).\nabla f(x(t))\rangle + \dot{x}(t)^T F(x(t))\dot{x}(t)^T$$

implying

$$d^T F(x^*)d + .\nabla f(x^*).\ddot{x}(t) \leq 0$$

We also know that $h(x(t)) = 0$, and so

$$\sum \lambda_i h_i(x_t)) = 0$$

This implies that

$$\frac{\partial^2}{\partial t^2}\left(\sum \lambda_i h_i(x_t))\right)|_{t=t^*} = 0$$

and so

$$\sum_{i=1}^{m} \lambda_i \left(d^T H_i(x^*)d + .\nabla h_i(x^*).\ddot{x}(t)\right) = 0$$

Adding these two together gives

$$\begin{aligned} & \sum_{i=1}^{m} \lambda_i d^T H_i(x^*)d + d^T F(x^*)d \\ +\ddot{x}(t)\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)\right) \quad \leq \quad & 0 \end{aligned}$$

and so (as $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$)

$$d^T \Psi(x^*, \lambda)d \leq 0$$

■

So what are these second order necessary conditions telling us? We know that in the case of unconstrained optimization, the second order conditions basically ask about whether the function is locally concave or convex. Here, we are comparing the curvature of the objective function to that of the constraint. To see this, consider the following two example:

Example 4 *Consider the following two problems:*

1. $f(x) = (x_1 + \frac{1}{2})^2 + x_2^2 \; h(x) = (x_1 + 1)^2 + x_2^2 = 1$

2. $f(x) = (x_1 + 1)^2 + x_2^2 \; h(x) = (x_1 + \frac{1}{2})^2 + x_2^2 = \frac{1}{4}$

In both cases, $(0, 0)$ *satisfies the FONC. To see this, note that, for each system*

1. $\nabla f(x) \begin{pmatrix} 2(x_1 + \frac{1}{2}) \\ 2x_2 \end{pmatrix}$, $\nabla h(x) = \begin{pmatrix} 2(x_1 + 1) \\ 2x_2 \end{pmatrix}$ *and so* $\nabla f(0, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\nabla h(0, 0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, *and so* $\nabla f(0, 0) + \lambda \nabla h(0, 0) = 0$ *for* $\lambda_A = -\frac{1}{2}$

2. $\nabla f(x) \begin{pmatrix} 2(x_1 + 1) \\ 2x_2 \end{pmatrix}$, $\nabla h(x) = \begin{pmatrix} 2(x_1 + \frac{1}{2}) \\ 2x_2 \end{pmatrix}$ *and so* $\nabla f(0, 0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, $\nabla h(0, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, *and so* $\nabla f(0, 0) + \lambda \nabla h(0, 0) = 0$ *for* $\lambda_B = -2$

However, in the first case, we have found a local minimum and in the second case a local maximum. This is because, in the first case the objective function is 'more' curved than the constraint, while in the second case the constraint is more curved than the objective.

The SONC picks this up. Note that, for each problem

1. $\Psi((0, 0), \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, *or positive definite*

2. $\Psi((0, 0), \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - 2 \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$, *or negative definite everywhere*

As with the unconstrained case, we also have equivalent second order sufficient conditions:

Theorem 10 *Suppose x^* is feasible and $h(x^*)$ is regular. If there exists a $\lambda \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{M} \lambda_i \nabla h_i(x^*) = 0$$

Moreover, the matrix

$$\Psi(x^*, \lambda) = F(x^*) + \sum \lambda_i H_i(x^*)$$

is negative definite in $T = \ker(h'(x^*))$ (i.e. $h'(x^*).d = 0$ implies $d^T \Psi(x^*, \lambda) d < 0$), then x^* is a strict local maximizer.

follows