4 Lecture 4

4.1 Constrained Optimization with Inequality Constraints

We are now going to move on to a discussion of Inequality constraints. Our canonical problem now looks as

Problem 11 (Constrained Optimization with Inequality 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*

$$\begin{aligned} h(x) &= 0 \\ g(x) &\leq 0 \end{aligned}$$

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

Our canonical problem therefore has m equality constraints and p inequality constraints.

When dealing with equality constraints, we solved this problem by characterizing the tangent plane. This was useful because it characterized all the feasible directions that one could move in from a particular point, and so we knew that for a local optimum it must be impossible to improve the objective function while moving in the direction of the tangent plane. Here it turns out we need to generalize the concept of the tangent plane. To see this, consider the constraint set in $\mathbb{R}^2$

$$\begin{aligned} g_1 &= x_1 - x_2 \\ g_2 &= -x_1 - x_2 \end{aligned}$$

and the point $(0, 0)$. These are two continuously differentiable functions, and yet, at $(0, 0)$, the tangent plane is given by $(0, 0)$, as the only continuously differentiable curve that goes through $(0, 0)$

is constant. We could characterize the tangent plane as the set $\{d | \langle \nabla g_i(0,0).d \rangle = 0,\ i = 1,2)\}$, and as

$$\nabla g_1(0,0) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and

$$\nabla g_2(0,0) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

This would give us back the tangent plane $(0,0)$. Thus, for this point, it true that for *any* objective function, the system

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

has no solution. So while it is true that any local maxima would have this property, it is also true for other points that are neither local maxima nor local minima of the problem. For example, consider the case of $f(x) = x_1$. We can find a pair of LeGrange multipliers λ_1, λ_2 such that

$$\begin{aligned} \frac{\partial f(x)}{\partial x_1} + \lambda_1 \frac{\partial g_1(x)}{\partial x_1} + \lambda_2 \frac{\partial g_2(x)}{\partial x_1} &= 0 \\ \frac{\partial f(x)}{\partial x_2} + \lambda_1 \frac{\partial g_1(x)}{\partial x_2} + \lambda_2 \frac{\partial g_2(x)}{\partial x_2} &= 0 \end{aligned}$$

at $(0,0)$, as

$$\begin{aligned} \frac{\partial f(x)}{\partial x_1} + \lambda_1 \frac{\partial g_1(x)}{\partial x_1} + \lambda_2 \frac{\partial g_2(x)}{\partial x_1} &= 1 + \lambda_1 - \lambda_2 \\ \frac{\partial f(x)}{\partial x_2} + \lambda_1 \frac{\partial g_1(x)}{\partial x_2} + \lambda_2 \frac{\partial g_2(x)}{\partial x_2} &= -\lambda_1 - \lambda_2 \end{aligned}$$

so, setting $\lambda_1 = -\frac{1}{2}$ and $\lambda_1 = \frac{1}{2}$ will do the trick.

The problem here is that the tangent plane is no longer a good description of the feasible directions that one can go from x^* (in the above case, one could go in any direction in the upwards cone from $(0,0)$). We therefore need to define the concept of a *tangent cone*

Definition 5 *A Tangent cone at x^* is*

$$C = \{\dot{x}(t^*) | x : (a,b) \to \mathbb{R}^n,\ x(t^*) = x^* \text{ and } x\left([t^*, b) \subset X\right)\}$$

In other words, a tangent cone is the derivative of all smooth parametric curves for which at least half the curve is in X (question - what is the tangent cone in the above example?)

So what we want to do is characterize the set C in terms of the gradients of the constraint functions. In order to do so, we are going to need some 'constraint qualifications', akin to the regularity conditions we used in the quality constraints case. The one we are going to use is in fact an extension of the concept of regularity. It should be noted that this condition is actually stronger than we need. There are other, weaker regularity conditions that do the job - for example the Mangasarian-Fromwitz CQ (wikipedia actually has a nice list of constraint qualifications).

We now need to define regularity at a point x^* in the case for inequality constraints. Basically, we still want the gradients of our constraints to be linearly independent, but we only care about constraints that bind - i.e. ones for which $g_i(x^*) = 0$. Thus, we define regularity as follows

Definition 6 *Let $x^* \in X$ for some constrained problem with slackness variables. Define $y_j^* = (-g_j(x^*))^2$. We say that x^* is regular if*

$$\left\{ \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} | i = 1, .., n \right\} \cup \left\{ \begin{pmatrix} \nabla g_j(x^*) \\ 2y_j^* e_j \end{pmatrix} | j = 1, .., p \right\}$$

is linearly independent, where e_j is a p length zero vector with a zero everywhere except in the jth position

Put another way, if we define

$$J(x^*) = \{j | g_j(x^*) = 0\}$$

Then the regularity condition demands that

$$\{(\nabla h_i(x^*)) | i = 1, .., n\} \cup \{(\nabla g_j(x^*)) | j \in J\}$$

is linearly independent.

We will state, but not prove, that under regularity, we can nicely characterize the tangent cone using the gradients of the constraints

Theorem 12 *if x^* is regular, then*

$$C(x^*) = \bar{K} = \left\{ d \in \mathbb{R}^n | h'(x^*).d = 0 \text{ and } \nabla g_j(x^*) d \leq 0, \; j \in J(x^*) \right\}$$

Proof. *Exercise* ■

So, in the previous example, the tangent cone is characterizes as $\beta \begin{bmatrix} \alpha \\ 1 \end{bmatrix}$ for $\alpha \in [-1,,1]$.

As before this allows us to determine that a particular series of equations does not have a solution

Lemma 2 *Suppose x^* is a local max and satisfies regularity. Then the system of equations*

$$\begin{aligned} h'(x^*).d &= 0 \\ g'_J(x^*).d &\leq 0 \\ f'(x^*).d &> 0 \end{aligned}$$

Has no solution

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

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

so if $\langle \nabla f(x^*), d \rangle > 0$, there exists an ε such that, for $t \in (t^*, t^* + \varepsilon)$ such that $f(x) - f(x^*) > 0$

■

Now we are in a position to use Farkas' lemma to prove the KKT conditions for the case with inequality constraints

Theorem 13 *Assume x^* is a local maximizer and satisfies regularity. Then there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^*)$$

$$\begin{aligned} \mu &\leq 0 \\ \textit{and } \mu_j . g_j(x^*) &= 0 \textit{ all } j \in (1..p) \end{aligned}$$

(we call these the KKT conditions)

Proof. *Again, this can be proved using Farkas' Lemma. We know from the previous lemma that if x^* is a local maxima, then the system*

$$\begin{aligned} h'(x^*).d &= 0 \\ g'_J(x^*).d &\leq 0 \\ f'(x^*).d &> 0 \end{aligned}$$

does not have a solution. This can be rewritten as can be written as

$$\begin{aligned} A^T d &\leq 0 \\ b^T d &> 0 \end{aligned}$$

where

$$\begin{aligned} b &= \nabla f\left(x^*\right) \\ A &= \left[\begin{array}{ccc} \nabla h\left(x^*\right) & -\nabla h\left(x^*\right) & \nabla g'_J\left(x^*\right) \end{array} \right] \end{aligned}$$

Using Farkas lemma, this tells us that

$$\begin{aligned} Ax &= b \\ x &\geq 0 \end{aligned}$$

has a solution, and so, letting $x = \left(\lambda^+, \lambda^-, -\mu\right)$, this gives

$$\left[\begin{array}{ccc} \nabla h\left(x^*\right) & -\nabla h\left(x^*\right) & \nabla g'_J\left(x^*\right) \end{array} \right] \left[\begin{array}{c} \lambda^+ \\ \lambda^- \\ -\mu \end{array} \right] = \nabla f\left(x^*\right)$$

and

$$\left[\begin{array}{c} \lambda^+ \\ \lambda^- \\ -\mu \end{array} \right] \geq 0$$

Setting $\mu_j = 0$ for all $j \notin J$, this gives us

$$\nabla f\left(x^*\right) + \sum_{i=1}^{m} \nabla h_i\left(x^*\right)\left(\lambda^+ - \lambda^-\right) + \sum_{j=1}^{m} \nabla g_j\left(x^*\right)\mu_j = 0$$

And, as $\mu \leq 0$, and $\mu_j = 0$ for all $j \notin J$ (and so for j such that $g_j(x^*) < 0$ we have the KKT conditions ■

In class we will draw a picture that will give some intuition to this result.

As you might expect, there are related second order necessary and sufficient conditions for the problem with inequality constraints. In order to do so, we need to define the concept of a second order test vector

Definition 7 *A vector $d \in \mathbb{R}^n$ is a second-order test vector at x^* if*

1. $g'_{J(x^*)}(x^*).d \leq 0$
2. $h'(x^*).d = 0$

Thus, a second order test vector is basically a feasible direction from x^*. Second order conditions are based on the idea that a particular matrix is negative semi-definite if we consider only test vectors.

Theorem 14 *Let x^* be a local maximizer, and regular. Then there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that (x^*, λ, μ) satisfy the KKT conditions, and*

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

for all second order test vectors d, where $\Psi(x^*, \lambda, \mu)$ is the Hessian matrix of

$$L(x^*, \lambda, \mu) = f(x^*) + \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{m} \mu_j g_j(x^*)$$

Similarly we can derive second order sufficient conditions.

Theorem 15 *Let $x^* \in \mathbb{R}^n$ such that there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ that satisfy the KKT conditions, and such that*

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

for all second order test vectors d, where $\Psi(x^*, \lambda, \mu)$ *is the Hessian matrix of*

$$L(x^*, \lambda, \mu) = f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i h_i\left(x^*\right) + \sum_{j=1}^{m} \mu_j g_j\left(x^*\right)$$

then x^* *is a local maximizer*