IEOR 151 – Lecture 10
Nonlinear Programming

1 First-Order Optimality Conditions

We will consider the following optimization problem (P):

$$\min \ f(x)$$
$$\text{s.t. } x \in \mathbb{R}^n$$
$$g_i(x) \le 0, \forall i = 1, \ldots, m$$
$$h_i(x) = 0, \forall i = 1, \ldots, k$$

where $f(x), g_i(x), h_i(x)$ are continuously differentiable functions. We call the function $f(x)$ the objective, and we define the feasible set $\mathcal{X}$ as

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid g_i(x) \le 0 \ \forall i = 1, \ldots, m \wedge h_i(x) = 0 \ \forall i = 1, \ldots, k\}.$$

Note that this formulation also incorporates maximization problems such as $\max\{f(x) : x \in \mathcal{X}\}$ through rewriting the problem as $\min\{-f(x) : x \in \mathcal{X}\}$.

Let $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n : \|x - y\| \le r\}$ be a ball centered at point x with radius r. A point x^* is a local minimizer for (P) if there exists $\rho > 0$ such that $f(x^*) \le f(x)$ for all $x \in \mathcal{X} \cap \mathcal{B}(x^*, \rho)$. A point x^* is a global minimizer for (P) if $f(x^*) \le f(x)$ for all $x \in \mathcal{X}$. Note that a global minimizer is also a local minimizer, but the opposite may not be true.

1.1 UNCONSTRAINED OPTIMIZATION

When (P) does not have any constraints, we know from calculus (specifically Fermat's theorem) that the global minimum must occur at points where either (i) the slope is zero $f'(x) = 0$, (ii) at $x = -\infty$, or (iii) at $x = \infty$. More rigorously, the theorem states that if $f'(x) \ne 0$ for $x \in \mathbb{R}$, then this x is not a local minimum. This result is useful because it gives one possible approach to solve (P) in the case where there are no constraints: We can find all points where the slope is zero, evaluate the function in the limits as x tends to $x = -\infty$, or $x = \infty$, and then select the minimum amongst these points. One natural question to ask is how can we extend this approach to the general case (P).

1.2 Fritz John Conditions

If $x^* \in \mathcal{X}$ is a local minimizer for (P), then there exists λ_i for $i = 0, \ldots, m$ and μ_i for $i = 1, \ldots, k$ such that the following equations, known as the Fritz John conditions, hold:

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla h_i(x^*) = 0$$
$$\lambda_i \geq 0, \forall i = 0, \ldots, m$$
$$\lambda_i g_i(x^*) = 0, \forall i = 1, \ldots, m$$
$$(\lambda_0, \ldots, \lambda_m, \mu_1, \ldots, \mu_k) \neq 0.$$

We say that the j-th inequality constraint is active if $g_j(x^*) = 0$, and inactive if $g_j(x^*) < 0$. Let $\mathcal{I}(x^*) = \{j : g_j(x^*) < 0\}$ be indices of the inactive constraints. The complimentary slackness condition $\lambda_i g_i(x^*) = 0$ means that $\lambda_j = 0$ must be zero for any inactive constraints $j \in \mathcal{I}$. Similarly, we denote the indices of the active constraints as $\mathcal{J}(x^*) = \{1, \ldots, m\} \backslash \mathcal{I}(x^*)$.

The Fritz John conditions are *necessary* (but not sufficient) for optimality. There are a few important points to note.

1. It is possible for the Fritz John conditions to hold at some point x^* that is not a minimizer. For instance, if $x^* \in \mathcal{X}$ is a point such that $\nabla h_i(x^*)$ for all $i = 1, \ldots, k$ are linearly dependent, then $x^* \in \mathcal{X}$ satisfies the Fritz John conditions regardless of whether it is a minimizer.

2. If $\lambda_0 = 0$ then the minimizer x^* is independent of the objective $f(x)$. Additionally, it must be that the $\nabla g_j(x^*)$ for all $j \in \mathcal{J}(x^*)$ and $\nabla h_i(x^*)$ for all $i = 1, \ldots, k$ must be linearly dependent. (Recall that a finite set of vectors v_i are linearly independent if the only coefficients a_i that solve $\sum_i a_i v_i = 0$ are $a_i = 0$ for all i.) This is not a positive situation because it means that the optimization problem (P) may not be modeling what we are interested in.

These points are important enough that they require further elaboration. Basically, if the constraints are not well-behaved, then either a computational algorithm will have trouble with finding a minimizer or the computed value will not depend upon the objective. What would be more useful is a necessary condition for local optimality, but we will need to ensure that the constraints are well-behaved.

1.3 Constraint Qualification

If we want optimality conditions like the Fritz John conditions to actually be indicative of optimality, we require the constraints to be well-behaved. There are a number of mathematical conditions that ensure this. The simplest is arguably the Linear Independence Constraint Qualification (LICQ). The LICQ holds at a point x if $\nabla g_j(x)$ for all $j \in \mathcal{J}(x)$ and $\nabla h_i(x)$ for all $i = 1, \ldots, k$ are linearly independent.

Under LICQ at x^*, we have that $\lambda_0 \neq 0$ in the Fritz John conditions for a local minimizer $x^* \in \mathcal{X}$. This means that a local optimizer will depend upon the objective. Also, we cannot have a situation in which an arbitrary point with LICQ x^* satisfies the Fritz John conditions.

If the objective $f(x)$ is convex, the inequality constraints $g_i(x) \leq 0$ are convex, and $h_i(x)$ are affine functions, then Slater's condition is another situation that implies the constraints are well-behaved. Slater's condition is that there exists a point x such that $g_i(x) < 0$ for all $i = 1, \ldots, m$ and $h_i(x) = 0$ for all $i = 1, \ldots, k$. The intuition is that the feasible set $\mathcal{X}$ is convex and has an interior.

1.4 Karush-Kuhn-Tucker Conditions

If LICQ holds at a point $x^* \in \mathcal{X}$, then the Karush-Kuhn-Tucker (KKT) conditions are necessary for *local* optimality of x^*. The KKT conditions are the Fritz John conditions with $\lambda_0 = 1$ and with the positivity constraints $(\lambda_0, \ldots, \lambda_m, \mu_1, \ldots, \mu_k) \neq 0$ removed. Note that satisfaction of the KKT conditions at a point $x^* \in \mathcal{X}$ is also necessary for *global* optimality of the point.

2 Example

Consider the following optimization problem

$$\min x$$
$$\text{s.t. } x \in \mathbb{R}^n$$
$$- x^3 \leq 0.$$

If $x^* \in \{x \in \mathbb{R} : -x^3 \leq 0\}$ is the global minimizer of this optimization problem, then the Fritz John conditions are that there exists λ_0, λ_1 such that

$$\lambda_0 - 3\lambda_1 x^{*2} = 0$$
$$\lambda_0, \lambda_1 \geq 0$$
$$- \lambda_1 x^{*3} = 0$$
$$(\lambda_0, \lambda_1) \neq 0.$$

We can ask the reverse question, which is: What are the values of $(\lambda_0, \lambda_1, x)$ for which the Fritz John conditions hold? To answer this, we begin with the complimentary slackness condition: $-\lambda_1 x^3 = 0$. This means that either (a) $\lambda_1 = 0$ and $x^{\cdot} - x^3 \leq 0$, or (b) $\lambda_1 > 0$ and $x = 0$. For case (a), the stationarity condition: $\lambda_0 - 3\lambda_1 x^2 = 0$ implies that $\lambda_0 = 0$; however, this violates $(\lambda_0, \lambda_1) \neq 0$ and hence cannot be a solution. For case (b), the stationarity condition: $\lambda_0 - 3\lambda_1 x^2 = 0$ implies that $\lambda_0 = 0$. Summarizing, the only set of solutions are

$$\lambda_0 = 0$$
$$\lambda_1 : \lambda_1 > 0$$
$$x = 0.$$

Because $\lambda_0 = 0$ is the only possible value that satisfies the Fritz-John conditions, this means that the global minimizer does not satisfy the KKT conditions. The reason is that constraint qualifications fail: The gradient of the constraint $-x^3$ is 0 at $x = 0$. In this example, we can fix this issue by rewriting our constraints to ensure that LICQ holds. Specifically, we can reformulate the optimization as

$$\min x$$
$$\text{s.t. } x \in \mathbb{R}^n$$
$$-x \leq 0.$$

The KKT conditions are that the global minimum x^* satisfies

$$1 - \lambda_1 = 0$$
$$\lambda_1 \geq 0$$
$$-\lambda_1 x^* = 0$$

Also, all possible combinations of (λ_1, x) for which the KKT conditions hold can be computed. From the stationarity condition: $1 - \lambda_1 = 0$, we have that $\lambda_1 = 1$. Combining this with the complimentary slackness condition: $-\lambda_1 x = 0$ gives that $x = 0$. This is the only possible solution of the KKT conditions, and so the local minimizer must also be the global minimizer. Summarizing, the global minimizer is $x^* = 0$.