IEOR 151 – Lecture 10
Optimization Review II

1 Optimality Conditions Example

Consider the following optimization problem

$$\begin{aligned}
\min\ & x_1 + x_2 \\
\text{s.t. } & x \in \mathbb{R}^2 \\
& -x_1 - 1 \leq 0 \\
& -x_2 - 1 \leq 0 \\
& 2x_1 + x_2 = 0 \\
& 4x_1 + 2x_2 = 0,
\end{aligned}$$

and use $\mathcal{X}$ to denote the feasible set defined in the normal way. If the point $x^* \in \mathcal{X}$ is a local minimizer, then the Fritz John conditions are that there exists $\lambda_0, \lambda_1, \lambda_2$ and $\mu_1, 2$ such that

$$\begin{aligned}
& \lambda_0 - \lambda_1 + 2\mu_1 + 4\mu_2 = 0 \\
& \lambda_0 - \lambda_2 + 1\mu_1 + 2\mu_2 = 0 \\
& \lambda_0, \lambda_1, \lambda_2 \geq 0 \\
& \lambda_1(-x_1 - 1) = 0 \\
& \lambda_2(-x_2 - 1) = 0 \\
& (\lambda_0, \lambda_1, \lambda_2, \mu_1, \mu_2) \neq 0.
\end{aligned}$$

The first question is: Which points x^* (not necessarily minimizers) satisfy the Fritz John conditions. Note that the values $(\lambda_0, \lambda_1, \lambda_2, \mu_1, \mu_2, x_1^*, x_2^*) = (0, 0, 0, -2, 1, 0.25, -0.5)$ satisfy the Fritz John conditions. In fact, by choosing $\mu_1 = -2$ and $\mu_2 = 1$ we can have an infinite number of points that satisfy the Fritz John conditions. This is discouraging because we would ideally like the number of points satisfying the optimality conditions to be small. Note that in these cases $\lambda_0 = 0$, meaning that the objective is not influencing the point.

The next question is: Does constraint qualification hold for this problem? This would be the first explanation for the degeneracy of the Fritz John conditions. Computing the gradients of the equality constraints gives

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \end{pmatrix}.$$

The two gradients are clearly *not* linearly independent. As a result, LICQ fails at every point in the feasible set.

Given the failure of the LICQ, how can we rewrite the optimization problem to ensure LICQ holds at every point? Removing the constraint $4x_1 + 2x_2 = 0$ fixes the problem. This is because if the $-x_1 - 1 \leq 0$ constraint is active (meaning $x_1 = -1$), then the equality constraint forces $x_2 = 2$. On the other hand, if the $-x_2 - 1 \leq 0$ constraint is active (meaning $x_2 = -1$), then the equality constraint forces $x_1 = 0.5$. The point is that only one inequality constraint can be active at a time, and so the gradient of an active inequality constraint (either $\begin{pmatrix}-1 & 0\end{pmatrix}'$ or $\begin{pmatrix}0 & -1\end{pmatrix}'$) cannot be linearly dependent with the gradient of the remaining equality constraint $\begin{pmatrix}2 & 1\end{pmatrix}'$. Thus, removing the constraint $4x_1 + 2x_2 = 0$ ensures that LICQ holds at every feasible point.

Since we not have LICQ of the modified optimization problem, we can now compute the KKT conditions. For a local minimizer $x^* \in \mathcal{X}$ there exists λ_1, λ_2 and μ_1 such that

$$\begin{aligned}
&1 - \lambda_1 + 2\mu_1 = 0 \\
&1 - \lambda_2 + 1\mu_1 = 0 \\
&\lambda_1, \lambda_2 \geq 0 \\
&\lambda_1(-x_1 - 1) = 0 \\
&\lambda_2(-x_2 - 1) = 0
\end{aligned}$$

So the natural question to ask is which points x^* satisfy the KKT conditions? Because of the complimentarily conditions, there are basically two interesting cases. Either (a) $x_1 = -1$ or (b) $x_2 = -1$; this is as expected because the optimal solution for a linear program lies on the boundary of the feasible set. In case (a), the equality constraint forces $x_2 = 2$; so $\lambda_2 = 0$ by the second complimentarily condition, and so the second stationarity condition forces $\mu_1 = -1$, which means that $\lambda_1 = -1$ by the first stationarity condition. But this is negative and not allowed. Hence, case (a) does not solve the KKT conditions. In case (b), the equality constraint forces $x_1 = 0.5$; so $\lambda_1 = 0$ by the first complimentarily condition, and so the stationarity condition forces $\mu_1 = -0.5$, which means that $\lambda_2 = 0.5$.

Since only one point satisfies the KKT conditions, it must be the global minimizer. The global minimizer is $x_1^* = 0.5$ and $x_2^* = -1$, and the minimum value is -0.5.

2 Parametric Optimization Problems

Many problems can be cast as a parametric optimization problem. These are optimization problems in which there is a free parameter $\theta \in \mathbb{R}^p$, and the optimization to be solved is different depending on the value of θ. Notationally, we write

$$\begin{aligned}
V(\theta) = \min\ & f(x;\theta) \\
\text{s.t. }\ & x \in C(\theta).
\end{aligned}$$

These types of problems are common in game theory, control engineering, and mathematical economics. One important question is whether the minimizers $x^*(\theta)$ and value function $V(\theta)$ are continuous with respect to the parameter θ.

2.1 Berge Maximum Theorem

The Berge Maximum Theorem provides an interesting set of conditions for when the minimizers and value function are appropriately continuous. The reason for stating appropriately continuous is that the function $C(\theta)$ is not a function in the normal sense; it is a multivalued function, and so continuity has to be redefined. The intuition of the theorem is that if the objective is continuous in the usual way, and the constraint set is continuous in an appropriate way, then the value function is continuous and the minimizer is continuous in an appropriate way. More formally:

If $f(x;\theta)$ is jointly continuous in x, θ and $C(\theta)$ is compact-valued, lower hemicontinuous, and upper hemicontinuous, then the value function $V(\theta)$ is continuous and the minimizer $x^*(\theta)$ is nonempty, compact-valued, and upper hemicontinuous.

Below we will define hemicontinuity; however, it is worth mentioning two important examples of constraint sets for which $C(\theta)$ is both upper and lower hemicontinuous. The first example is when the constraint set is independent of θ. The second example is when the constraint set is linear in both x and θ.

2.2 Upper Hemicontinuity

We say that $C(\cdot) : \mathbb{R}^p \to \mathbb{R}^n$ is upper hemicontinuous at $\theta \in \mathbb{R}^p$ if $\forall \theta_k \in \mathbb{R}^p$, $\forall x_k \in C(\theta_k)$, and $\forall x \in \mathbb{R}^n$; we have

$$\lim_{k\to\infty} \theta_k = \theta, \lim_{k\to\infty} x_k \to x \Rightarrow x \in C(\theta).$$

The intuition is that every sequence $x_k \in C(\theta_k)$ converges to a point $x \in C(\theta)$.

2.3 Lower Hemicontinuity

We say that $C(\cdot) : \mathbb{R}^p \to \mathbb{R}^n$ is lower hemicontinuous at $\theta \in \mathbb{R}^p$ if $\forall \theta_k \in \mathbb{R}^p$ such that $\theta_k \to \theta$, $\forall x \in C(\theta)$, there exists a subsequence θ_{k_j} and $x_k \in C(\theta_{k_j})$ such that $x_k \to x$. The intuition is that for every point $x \in C(\theta)$ there is a sequence $x_k \in C(\theta_k)$ that converges to the point.