Financial Optimization
ISE 347/447

Lecture 10

Dr. Ted Ralphs

Reading for This Lecture

- C&T Chapter 5

A Quick Tour of Nonlinear Programming

- Recall the fundamental theorem of asset pricing from Lecture 6.
- Stripping away the context, this theorem really just says something like:

  Theorem 1. *Let $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^m$. Then exactly one of the following systems has a solution:*

  I. $Ad \geq 0$ and $c^\top d < 0$.
  II. $A^\top y = c$ and $y \geq 0$.

- In this form, the theorem is know as the Farkas Theorem and can be seen as the basis for much of optimization theory.
- It is easy to see the relationship to LP duality and this is another way of deriving strong duality for LP.

Example: Arbitrage

We consider a market consisting of a risk-free asset S^0 and 2 risky assets S^1 and S^2. The current prices of the assets are

$$\begin{aligned} S_0^0 &= 1 \\ S_0^1 &= 7 \\ S_0^2 &= 5 \end{aligned}$$

The risk-free rate of return is 10%. At time 1, the state of the market is represented by three scenarios represented by events Ω_1, Ω_2, and Ω_3. The price distribution of assets 1 and 2 are

$$S_1^1(\omega) = \begin{cases} 10 & \text{if } \omega \in \Omega_1 \\ 1 & \text{if } \omega \in \Omega_2 \\ 20 & \text{if } \omega \in \Omega_3 \end{cases}$$

$$S_1^2(\omega) = \begin{cases} 5 & \text{if } \omega \in \Omega_1 \\ 12 & \text{if } \omega \in \Omega_2 \\ 1 & \text{if } \omega \in \Omega_3 \end{cases}$$

Example: Arbitrage

```
>>> A = np.matrix([[1.1, 10, 5], [1.1, 1, 12], [1.1, 20, 1]])
>>> print np.around(np.linalg.inv(A), 3)
[[ 6.39  -2.406 -3.075]
 [-0.324  0.118  0.206]
 [-0.559  0.294  0.265]]
>>> c = np.matrix([1, 7, 5])
>>> print c*np.linalg.inv(A)
[[ 1.3315508  -0.11229947 -0.31016043]]
```

How can we determine by inspection whether there is arbitrage?

Generalizing the Concepts

- The approach we took to LP duality theory earlier can be generalized to the nonlinear setting, but things are not as clean.
- For convex objective functions and feasible regions, the results generalize with little change.
- For nonconvex functions, we can get an analog of weak duality, but strong duality no longer holds.
- We know look briefly at the development of this theory and the core concepts of nonlinear programming.

The Primal Problem

- Given functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^l$, we consider the optimization problem (P), which we will call the *primal problem*:

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in X \end{aligned}$$

- As before, $X \subseteq \mathbb{R}^n$ is a set that implicitly enforces additional constraints.
- Note that we have explicitly separated the equality and inequality constraints.

The Lagrangian Function

- Recall the Lagrangian function we used to derive LP duality.
- We can use the same trick to obtain a dual of (P).
- We define the *Lagrangian function* as follows:

$$\Phi(x, u, v) \equiv f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x)$$

- As before, we are penalizing the violation of the constraints and turning an constrained optimization problem into an unconstrained one.
- The penalties are called *Lagrangian multipliers* or *dual prices*.

The Dual Problem

- As before, it is easy to show that for any $\hat{v} \in \mathbb{R}^l$ and $\hat{u} \in \mathbb{R}^m_+$

$$\inf_{x \in X} \Phi(x, \hat{u}, \hat{v})$$

  is a lower bound on the optimal solution value to (P).

- We can then formulate the following *dual problem* (D):

$$\begin{aligned} &\max \Theta(u, v) \\ &\text{s.t. } \quad u \geq 0 \end{aligned}$$

  where $\Theta(u, v) = \inf_{x \in X} \Phi(x, u, v)$.

- This problem finds the best lower bound over all choices of dual prices.

- Note that Θ is a concave function.

Weak Duality

Theorem 2. *Let x be a feasible solution to the primal problem (P) and let (u, v) be a solution to the dual problem (D). Then $f(x) \geq \Theta(u, v)$.*

Let $S = \{x \in X,\ g(x) \leq 0,\ h(x) = 0\}$

Corollary 1. $\inf\{f(x) \mid x \in S\} \geq \sup\{\Theta(u, v), u \geq 0\}$.

Corollary 2. *If $f(x^*) = \Theta(u^*, v^*)$ for some $x^* \in S, u \geq 0$, then x^* solves (P) and (u^*, v^*) solves (D).*

Corollary 3. *If $\sup\{\Theta(u, v), u \geq 0\} = \infty$, then the primal problem has no feasible solution.*

Theorem of the Alternative

Theorem 3. *Let X be a given nonempty convex set in R^n, $\alpha : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ given convex functions, and $h : \mathbb{R}^n \to \mathbb{R}^l$ a given affine function, i.e., $h(x) = Ax - b$ for $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, x \in \mathbb{R}^n$. Then if I has no solution, II has a solution. The converse holds if $u_0 > 0$.*

$$I. \;\; \alpha(x) < 0, g(x) \leq 0, h(x) = 0, \exists x \in X.$$

$$II. \;\; u_0\alpha(x) + u^\top g(x) + v^\top h(x) \geq 0, \forall x \in X, u \geq 0, (u, v) \neq 0.$$

- This generalizes the Farkas Theorem.
- Note, however, that there is a qualification needed for the converse to hold.

Strong Duality

Theorem 4. *Let X be a given nonempty convex set in $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be given convex functions, and $h : \mathbb{R}^n \to \mathbb{R}^l$ be given affine functions. If there exists $x' \in X$ such that $g(x') < 0$, $h(x') = 0$, and $0 \in int\ h(X)$, then*

$$\inf_{x \in S} f(x) = \sup_{u \geq 0} \Theta(u, v)$$

If the $\inf_{x \in S} f(x)$ is finite, then $\sup_{u \geq 0} \Theta(u, v)\}$ is achieved at (u^*, v^*) with $u^* \geq 0$. If the $\inf$ is achieved at x^*, then ${u^*}^\top g(x^*) = 0$.

Minimizing a Convex Function

Theorem 5. *Let S be a nonempty convex set on $\mathbb{R}^n$ and let $f : S \to \mathbb{R}$ be convex on S. Suppose that x^* is a local optimal solution to $min_{x \in S} f(x)$.*

- *Then x^* is a global optimal solution.*
- *If either x^* is a strict local optimum or f is strictly convex, then x^* is the unique global optimal solution.*

Necessary and Sufficient Conditions

Theorem 6. *Let S be a nonempty convex set on $\mathbb{R}^n$ and let $f : S \to \mathbb{R}$ be convex on S. The point $x^* \in S$ is an optimal solution to the problem $min_{x \in S}\ f(x)$ if and only if f has a subgradient ξ such that $\xi^\top (x - x^*) \geq 0\ \ \forall x \in S$.*

- This implies that if S is open, then x^* is an optimal solution if and only if there is a zero subgradient of f at x^*.

First-order Necessary Conditions (Unconstrained)

Theorem 7. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be differentiable at* x^**. If there is a vector* d *such that* $\nabla f(x^*)^\top d < 0$*, then there exists a* $\delta > 0$ *such that* $f(x^* + \lambda d) < f(x^*)$ *for each* $\lambda \in (0, \delta)$*.*

Corollary 4. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be differentiable at* x^**. If* x^* *is a local minimum, then* $\nabla f(x^*) = 0$*.*

- The direction d is called a *descent direction* or an *improving direction*.

Second-order Necessary Conditions (Unconstrained)

Theorem 8. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at x^*. If x^* is a local minimum, then $\nabla f(x^*) = 0$ and $H(x^*)$ is positive semi-definite.*

Sufficient Conditions (Unconstrained)

Theorem 9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at x^*. If $\nabla f(x^*) = 0$ and $H(x^*)$ is positive definite, then x^* is a local minimum.*

Theorem 10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let x^* be a point at which f is differentiable. Then x^* is a global minimum if and only if $\nabla f(x^*) = 0$.*

Feasible and Improving Directions

Definition 1. *Let S be a nonempty set in $\mathbb{R}^n$ and let $x^* \in cl(S)$. The* cone of feasible directions *of S at x^* is given by*

$$D = \{d : d \neq 0 \text{ and } x^* + \lambda d \in S, \forall \lambda \in (0, \delta), \exists \delta > 0\}$$

Definition 2. *Let S be a nonempty set in $\mathbb{R}^n$ and let $x^* \in cl(S)$. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the* cone of improving directions *of f at x^* is given by*

$$F = \{d : f(x^* + \lambda d) < f(x^*), \forall \lambda \in (0, \delta), \exists \delta > 0\}$$

Characterizing Set F

Define $F_0 = \{d : \nabla f(x^*)^\top d < 0\}$ and $F_0^{'} = \{d : d \neq 0, \nabla f(x^*)^\top d \leq 0\}$. Then $F_0 \subseteq F \subseteq F_0^{'}$.

Theorem 11. *Let S be a nonempty set in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given. Consider the constrained optimization problem*

$$\min f(x)$$
$$\textit{s.t. } x \in S$$

If x^* is a local optimal solution and f is differentiable at x^*, then $F_0 \cap D$ is empty. Conversely,...

Characterizing Set D

- Consider the feasible region $S = \{x \in X \mid g_i(x) \leq 0, i \in [1, m]\}$ where X is a nonempty open set in $\mathbb{R}^n$ and $g_i : \mathbb{R}^n \to \mathbb{R}, i \in [1, m]$.
- Given a feasible $x^* \in \mathbb{R}^n$, set $I = \{i \mid g_i(x^*) = 0\}$.
- Assume that g_i is differentiable at x^* for $i \in I$ and g_i is continuous at x^* for $i \notin I$ and define $G_0 = \{d \mid \bigtriangledown g_i(x^*)^\top d < 0 \ \forall i \in I\}$ and $G_0^{'} = \{d \mid d \neq 0, \bigtriangledown g_i(x^*)^\top d \leq 0 \ \forall i \in I\}$. Then $G_0 \subseteq D \subseteq G_0^{'}$.

More Optimality Conditions

Theorem 12. *Let X be a nonempty open set in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ be given. Consider the constrained optimization problem*

$$\begin{aligned} &\min f(x) \\ &\textit{s.t.}\ \ g_i(x) \leq 0 \\ &\qquad\ x \in X \end{aligned}$$

If x^* is a local optimal solution, then $F_0 \cap G_0$ is empty. Conversely, ...

Fritz-John Necessary Conditions

Theorem 13. *Consider the feasible region* $S = \{x \in X \mid g(x) \leq 0\}$ *where* X *is a nonempty open set in* $\mathbb{R}^n$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$*. Given a feasible* $x^* \in S$*, set* $I = \{i \mid g_i(x^*) = 0\}$*. Assume that* f *and* g_i *are differentiable at* x^* *for* $i \in I$ *and* g_i *is continuous at* x^* *for* $i \notin I$*. If* x^* *is a local minimum, then there exists* $u_0 \in \mathbb{R}$*,* $u \in \mathbb{R}^m$ *such that*

$$u_0 \bigtriangledown f(x^*) + \sum_{i=1}^{m} u_i \bigtriangledown g_i(x^*) = 0$$

$$u_i g_i(x^*) = 0 \qquad \forall i \in [1, m]$$

$$(u_0, u) \geq (0, 0)$$

$$(u_0, u) \neq (0, 0)$$

Terminology

- The u_i's are the *Lagrange multipliers* or *dual prices*.
- The requirement that $x^* \in S$ is called the *primal feasibility* (PF) condition.
- The requirement that $u_0 \bigtriangledown f(x^*) + \sum u_i \bigtriangledown g_i(x^*) = 0$ is called the *dual feasibility* (DF) condition.
- Note that dual feasibility is equivalent to $\bigtriangledown L(x^*) = 0$, where

$$L(x) \equiv \Phi(x, u^*) \equiv f(x) + \sum_{i \in I} u_i^* g_i(x)$$

  is called the *restricted Lagrangian function*.
- The requirement that $u_i g_i(x^*) = 0 \ \forall i \in [1, m]$ is called the *complementary slackness* (CS) condition.
- *FJ points* are those satisfying PF, DF and CS.
- Notice the similarity to LP optimality conditions.

Fritz-John Sufficient Conditions

- Note that a point is an FJ point if and only if $F_0 \cap G_0$ is empty.
- Notation and setup as for necessary conditions.

  Theorem 14. *If x^* is an FJ point and there exists $N_\epsilon(x^*), \epsilon > 0$ such that f is convex and $g_i, i \in I$ are strictly convex over $N_\epsilon(x^*) \cap S$, where S is the relaxed feasible region without the nonbinding constraints, then x^* is a local minimum.*

- There are also other possible sufficient conditions.

Remarks on the FJ conditions

- These conditions hold trivially in many cases.
- In particular, if $G_0 = \emptyset$, they will hold, regardless of the objective function (take $u_0 = 0$).
- Even for LP, there are non-optimal FJ points.
- We want to force $u_0 > 0$ in order to take the objective function into account.
- We do this by using a *constraint qualification*.

KKT Necessary Conditions

- If we require that the gradients of the binding constraints be linearly independent, this is equivalent to requiring that $u_0 > 0$.
- In this case, we can drop the condition that $u \neq 0$ and we get x^* locally optimal $\Rightarrow$ there exists $u \in \mathbb{R}^m$ such that

$$\begin{aligned} \bigtriangledown f(x^*) + \sum u_i \bigtriangledown g_i(x^*) &= 0 \\ u_i g_i(x^*) &= 0 \, \forall i \in [1, m] \\ u &\geq 0 \end{aligned}$$

Remarks on the KKT conditions

- Again, we have PF, DF and CS conditions. These make up the *KKT conditions*.
- x^* is a *KKT point* if the KKT conditions are satisfied at x^*.
- For a linear program, the KKT conditions are simply the standard optimality conditions for LP.
- Using previous notation, note that x^* is a KKT point if and only $F_0 \cap G_0'$ is empty
- Furthermore, x^* is a KKT point if and only if x^* is the solution to the first-order LP approximation to the NLP

  $$\min\{f(x^*) + \nabla f(x^*)^\top (x - x^*) \mid g_i(x^*) + \nabla g_i(x^*)^\top (x - x^*) \leq 0, i \in [1, m]\}$$

KKT Sufficient Conditions (1^{st} Order)

- We have the same setup as before.

Theorem 15. *Let x^* be a KKT point and let $I = \{i : g_i(x^*) = 0\}$. If f is convex at x^* and if $g_i, i \in I$ are differentiable and convex at x^*, x^* is then a global optimal solution.*

- There are other possible sufficient conditions.

KKT Sufficient Conditions (2^{nd} Order)

- **Theorem 16.** *Consider the problem to minimize* $f : \mathbb{R}^n \to \mathbb{R}$ *over* $S = \{x \in X : g(x) \leq 0\}$ *where* X *is a nonempty open set in* $\mathbb{R}^n$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$*. Assume* f *and all constraint functions are twice differentiable.*

  Suppose x^* *is a KKT point with restricted Lagrangian function* L*.*

  - *If* $\nabla^2 L(x)$ *is positive semi-definite* $\forall x \in S$*, then* x^* *is a global minimum.*
  - *If* $\nabla^2 L(x)$ *is positive semi-definite in a neighborhood of* x^**, then* x^* *is a local minimum.*
  - *If* $\nabla^2 L(x^*)$ *is positive definite, then* x^* *is a strict local minimum.*

Optimality Conditions for Convex Programs

- The KKT conditions are sufficient for *convex programs*:
  - f is convex
  - $g_1, ..., g_m$ is convex
  - $h_1, ..., h_l$ is linear
- The KKT conditions are necessary and sufficient for convex programs with all linear constraints.