Econ 205 - Supplementary Slides on Optimization

Joel Sobel

September 10, 2010

I didn't lecture on this

I won't test you on this.

Saddle Points

To begin the study, we need a notion of a direction that is not ruled out by constraints.

Definition

The vector v enters S from x^* if there exists $\varepsilon > 0$ such that $x^* + tv \in S$ for all $t \in [0, \varepsilon)$.

If $x^* \in S$ and v enters S from x^*, then it is possible to start at x^*, go in the direction v, and stay in the feasible set. It follows that if x^* is a local maximum, then $D_v f(x^*) \geq 0$.
To fully describe the set of first-order conditions for a constrained optimization problem, we need to figure out a nice way to describe which vectors enter S at x and try to get a clean way of summarizing the resulting inequalities.

Definition

S is described by linear inequalities if there is a matrix $\mathbf{A}$ such that $S = \{x : Ax \geq b\}$.

- A general constraint set will be written in the form $S = \{x : g_i(x) \geq b_i, i \in I\}$.
- If the constraint set is described by linear inequalities, then we can take all of the g_i to be linear functions (that is, $g_i(x) = a_i \cdot x$, where a_i is the ith row of the matrix $\mathbf{A}$).
- A set described by inequalities includes an equality constrained set as a special case (to impose the constraint $g(x) = b$ substitute two inequalities: $g(x) \geq b$ and $-g(x) \geq -b$).
- Converse is false: Equality constrained problems are typically easier than inequality constrained problems.

Inward Normals

Definition

An inward pointing normal to the boundary of the set $S = \{x : g_i(x) \geq b_i, i = 1, \ldots, I\}$ at x^* is a direction of the form $\nabla g_i(x^*)$ where $g_i(x^*) = b_i$.

- When $g_i(x^*) = b_i$ the ith constraint is *binding*.
- If you move from x^* in the "wrong" direction, then the constraint will no longer be satisfied.
- If you move "into" S, then the constraint will be satisfied.
- When S is described by linear inequalities, then the inward directions are simply the a_i associated with binding constraints.

Theorem

Suppose that S is described by linear inequalities. v enters S from x^* if and only if v makes a nonnegative inner product with all of the inward pointing normals to the boundary of the feasible set at x^*.

Proof

1. Let $x^* \in S$ and let $a_k \cdot x^* = 0$ for $k \in J$ (J may be empty).
2. Let v enter S from x^*.
3. There exists $\varepsilon > 0$ such that $a_k \cdot (x^* + tv) \geq 0$ for $k \in I$ and $t \in [0, \varepsilon)$.
4. So if $i \in J$, then $a_k \cdot x^* = b_k$ and hence

$$a_k \cdot v \geq 0 \text{ for all } k \in J.$$

Converse

1. Assume $a_k \cdot v \geq 0$ for all $k \in J$.
2. For $k \notin K$, it must be that $a_k \cdot x^* > b_k$.
3. Therefore $a_k \, (x^* + tv) > b_k$ for sufficiently small t.
4. For $k \in J$ we have $a_k \cdot x^* = b_k$.
5. Hence $a_k \cdot (x^* + tv) \geq b_k$.
6. So v enters S at x^*.

Optimality

Theorem

If x^* maximizes $f(x)$ subject to $x \in S$, then $\nabla f(x^*) \cdot v \leq 0$ for all v entering S from v.

Proof.

There exists $\varepsilon > 0$ such that $x^* + tv \in S$ for $t \in [0, \varepsilon)$ and therefore $f(x^*) \geq f(x^* + tv)$ for $t \in [0, \varepsilon)$. Hence $D_v f(x^*) \geq 0$, which is equivalent to $\nabla f(x^*) \cdot v \leq 0$. □

Refinement

Theorem

Exactly one of the following two things holds:

$$\text{There exists } \lambda \geq 0 \text{ such that } \mathbf{A}^t \lambda = w.$$

or

$$\text{There exists } x \text{ such that } \mathbf{A}x \geq 0 \text{ and } w \cdot x < 0.$$

- The first condition in the theorem says that w is in the convex set generated by the rows of the matrix $\mathbf{A}$.
- So theorem says that if w fails to be in a certain convex set, then it can be separated from the set.
- In application $w = -\nabla f(x^*)$.
- λ in the second condition is the normal of the separating hyperplane.
- The second condition states that it is possible to find a direction that that makes an angle of less than ninety degrees with all of the rows of $\mathbf{A}$ and greater than ninety degrees with w.
- Second condition is essentially ruled out by the theorem, hence the first condition must hold.

Theorem

Suppose x^* *solves:* $\max f(x)$ *subject to* $x \in S$ *when* S *is defined by linear inequalities and* f *is differentiable. Then there exists* λ_k^* *such that*

$$\nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla (g_i(x^*) - b_i) = 0 \tag{1}$$

and

$$\sum_{i \in I} \lambda_i^* (g_i(x^*) - b_i) = 0. \tag{2}$$

1. Theorem gives us multipliers for an inequality constrained optimization problem provided that the constraint set is linear.
2. The only place where we used linearity of the constraint set was in the characterization of inward pointing normals.
3. We need an additional condition to get this result in general.

Pathological Example

$\max f(x_1, x_2) = x_1$ subject to $x_1, x_2 \geq 0$ and $(1 - x_1)^3 - x_2 \geq 0$.

- Solution: $(x_1, x_2) = (1, 0)$.
- Second and third constraints force $x_1 \leq 1$. Since $(1, 0)$ is feasible, it must be optimal.]
- The theorem would say:

$$1 + \lambda_1 + \lambda_3 \left(-3(1 - x^*)^2 \right) = 0, \tag{3}$$

  which has no non-negative solution when $x^* = 1$.
- Alternatively Equation (2) implies that $\lambda_3 = 0$ when $x^* < 1$, which again means that equation (3 cannot have a solution.
- Conclude that the conditions in Theorem 7 need not hold if S is not described by linear inequalities.

Diagnosis

1. The constraints that define S give inward pointing normals at $(1,0)$ as $(0,1)$ (from the binding constraint that $x_2 \geq 0$) and $(0,-1)$ (from the binding constraint that $(1-x_1)^3 - x_2 \geq 0$).
2. $v = (1,0)$ makes a zero inner product with both of these normals, but does not satisfy the definition of an entering direction.
3. The theorem places "too many" constraints on the derivatives of f so is not a useful characterization.
4. Need new definition of inward normal and then an assumption that rules out awkward cases.

Revised Definition of Inward Normal

Definition

The vector v enters S from x^* if there is a $\varepsilon > 0$ and a curve $\sigma : [0, \varepsilon) \longrightarrow S$ such that $\sigma(0) = x^*$, and $\sigma'(0) = v$.

- $\sigma'(0)$ denotes the right-hand derivative of σ at 0.
- Paths are not necessarily straight lines.
- Difference doesn't matter for linear constraint sets.
- It is immediate that $\nabla f(x^*) \cdot v \geq 0$ for all v entering S from x^* (according to Definition 8 as in Theorem 4).

Definition

If $S = \{x : g_i(x) \geq 0 \text{ for all } i \in I\}, g_i$ are differentiable and $x^* \in S$, then the *inward normals at* x^* are the vectors $\nabla g_i(x^*)$ for i such that $g_i(x^*) = 0$.

Constraint Qualification

Definition

Let $S = \{x : g_i(x) \geq 0 \text{ for all } i \in I\}$, g_i are differentiable, and $x^* \in S$. S satisfies the *constraint qualification* at x^* if $n_i \cdot v \geq 0$ for all inward normals n_i at x^* implies that v enters S from x^*.

- The example does not satisfy the constraint qualification.
- CQ will hold if the set of inward normals is linearly independent.
- When linear independence holds, it is possible to find a normal direction that strictly enters: $n_i \cdot w = 1$ for all i and with this you can construct a curve with the property that $\nabla g_i(x^*) \cdot \sigma'(0) > 0$ for all i such that $g_i(x^*)$.
- The linear independence condition will not hold if some of the constraints were derived from equality constraints (that is, one constraint is of the form $g(x) \geq 0$ and another constraint is of the form $-g(x) \geq 0$).
- Write equality constraints separately and impose CQ on binding constraints.

Kuhn-Tucker Again

Theorem

Suppose x^* *solves:* $\max f(x)$ *subject to* $x \in S$ *when* $S = \{x : g_i(x) \geq 0 \text{ for all } i \in I\}$ *and* f *and* g *are differentiable. If* x^* *satisfies the constraint qualification at* x^**, then there exists* λ_k^* *such that*

$$\nabla f(x^*) + \sum_{i \in I} \lambda_i \nabla g_i(x^*) - b_i) = 0 \tag{4}$$

and

$$\sum_{i \in I} \lambda_i \left(g_i(x^*) - b_i\right) = 0. \tag{5}$$

Look familiar?

Saddle Point Theorems

We have shown that if a "regularity" condition holds then

$$x^* \text{ solves } \max f(x) \tag{6}$$

subject to

$$g_i(x) \geq \tilde{b}_i, i \in I \tag{7}$$

then equations (1) and (2) must hold.
Note that

$$\nabla g_i(x^*) = \nabla (g_i(x^*) - b_i)$$

so equation (7) can be written without the b_i.
In particular, this result holds when the constraints are linear (which guarantees that the regularity condition will hold).

Linear Programming

Basic Problem:

$$\max c \cdot x \text{ subject to } Ax \leq b, x \geq 0. \tag{8}$$

- $f(x) = c \cdot x$.
- For each $i = 1, \ldots, m, g_i(x) = -a_i \cdot x$, where a_i is a row of A and for $i = m+1, \ldots, m+n, g_i(x) = x_i$.
- To put in general form, set the $\tilde{b}$ in the theorem equal to $(-b_1, \ldots, -b_m, 0, \ldots, 0)$. $\tilde{b}$ has $n + m$ components.

Observation

1. $\nabla f(x) \equiv c$ when $f(x) \equiv c \cdot x$ and similarly for the constraints.
2. If x^* solves (8) there exists $y^*, z^* \geq 0$ such that

$$c - y^* A + z^* = 0 \tag{9}$$

   and

$$y^* \cdot (b - Ax^*) + z^* \cdot x^* = 0. \tag{10}$$

3. λ divides into two parts:
   3.1 y^* contains the components of λ corresponding to $i \leq m$ – the constraints summarized in the matrix A
   3.2 z^* are the components of λ corresponding to $i > m$; these are the non-negativity constraints.

Massaging the Problem

$$s(x : y, z) = c \cdot x + y \cdot (b - Ax) + v \cdot x. \tag{11}$$

We have the following theorem.

Theorem

If x^* solves (8), then there exists y^* and $z^* \geq 0$ such that

1. $s(x^*; y^*, z^*) \geq s(x; y^*, z^*)$ *for all* x.
2. $s(x^*; y^*, z^*) \leq s(x^*; y, z)$ *for all* $y, z \geq 0$.
3. y^* *solves* $\min b \cdot y$ *subject to* $A^t y \geq c$ *and* $y \geq 0$.
4. $b \cdot y^* = c \cdot x^*$.

.

It follows from equation (9) that

$$s(x : y^*, z^*) = \left(c - A^t y * + z^*\right) x + b \cdot y^* = b \cdot y^* \tag{12}$$

for all x. Therefore the first claim of the theorem holds as an equation. Furthermore, equation (9) implies that

$$y^* \cdot (b - Ax^*) + x^* \cdot z^*. \tag{13}$$

Hence $s(x^*; y^*, z^*)$ $b\cdot^*$. Equation 13) proves that

$$c \cdot x^* = -(x^*; y^*, z^*) = b \cdot y^*. \tag{14}$$

This implies the fourth claim of the theorem.

To prove the second line, we must show that

$$s(x^*; y, z) = c \cdot x^* + y^* \cdot (b - Ax^*) + x \cdot z \geq s(x^*; y^*, z^*) = c \cdot x^* \text{ for all } y, z. \tag{15}$$

However, x^* and $b \geq Ax^*$ by the constraints in (8). Therefore, if $y, z \geq 0$, then $y\,(b - Ax^*) + z \cdot x$ and so the second claim in the theorem follows.

To prove the third claim, first note that

$$c - A^t y^* + z^* = 0 \text{ and } z^* \geq 0$$

implies that $c \leq A^t y^*$. Therefore y^* satisfies $A^t y^* \geq c$ and $y^* \geq 0$. If y also satisfies $A^t y \geq c$ and $y \geq 0$, then by (2) and

$$\begin{aligned} s(x^*; y, z^*) &= c \cdot x^* + y\,(b - Ax^*) + z^* \cdot x^* && (16)\\ &= \left(c - A^t y\right) x^* + b \cdot y + z^* \cdot x^* && (17)\\ &= \left(c - A^t y\right) x^* + b \cdot y && (18)\\ &= b \cdot y, && (19) \end{aligned}$$

1. Definition of $s(\cdot)$
2. Algebraic manipulation,
3. Equations (10), $y^* \geq 0$, and $b \geq Ax^*$ imply that $z^* \cdot x^* = 0$, and
4. the inequality follows because $x^* \geq 0$ and $c \leq A^t y$.

It follows that

$$b \cdot y \geq s(x^*; y, z^*) \geq s(x^*; y^*, z^*) = b \cdot y^*$$

so y^* solves the minimization problem (3).

If x^* solves the *Primal*:

$$\max c \cdot x \text{ subject to } \; Ax \leq b, x \geq 0,$$

then y^* solves the *Dual*:

$$\min b \cdot y \text{ subject to } A^t y \geq c, y \geq 0.$$

- Dual is equivalent to

$$\max -b \cdot y \text{ subject to } -A^t y \leq -c, y \geq 0$$

- This problem has the same general form as the Primal (with c replaced by $-b$ and A replaced by $-A^t$ and b replaced by $-c$).
- The Dual of the Dual is the Primal.

Fundamental Duality Theorem

Theorem

Associated with any Primal Linear Programming Problem

$$\max c \cdot x \ \textit{subject to} \ \ Ax \leq b, x \geq 0,$$

there is a Dual Linear Programming Problem:

$$\min b \cdot y \ \textit{subject to} \ A^t y \geq c, y \geq 0.$$

The Dual of the Dual is the Primal. If the Primal has a solution x^**, then the Dual has a solution* y^* *and* $b \cdot y^* = c \cdot x^*$*. Moreover, if the Dual has a solution, then the Primal has a solution and the problems have the same value.*

Comments

1. When you solve a constrained optimization problem you are simultaneously solving another optimization problem for a vector of multipliers.
2. There is one y_i for every constraint in the primal (in particular, there need not be the same number of variables in the primal as variables in the dual).
3. The *values* of the two problems are comparable: $b \cdot y^* = c \cdot x^*$.
4. Easy: if x satisfies the constraints of the Primal then
   $c \cdot x \leq c \cdot x^*$
   if y satisfies the constraints of the dual then $b \cdot y \geq b \cdot y^*$.
5. So feasible (P) values $\leq$ feasible (D) values.

Complementary Slackness

- $y^* (b - Ax^*) = 0$ and $(c - A^t y^*) x^* = 0$.
- If you know that $y_i > 0$, then you know that the associated constraint in binding (holds as an equation).
- and three other related kinds of implication.
- If "price" of input i is positive, you exhaust your supply.
- If you have excess supply of input i, its price is zero.
- If you produce positive amounts of product j ($x^* > 0$), then the value of ingredients equals its price.
- If the price of ingredients is greater than price, you don't produce.
- Equality of values means value of output is equal to value of inputs.

Primal has no solution

- Primal unbounded is possible (then dual not feasible).
- Primal infeasible is possible (then dual not feasible or unbounded).

To summarize:

1. If the primal has a solution, then so does the dual, and the solutions have the same value.
2. If the primal has no solution because it is unbounded, then the dual is not feasible.
3. If the primal is not feasible, then either the dual is not feasible or the dual is unbounded.
4. The three statements about remain true if you interchange the words "primal" and "dual."

If x^* solves the Primal, then there exist y^* and z^* such that

$$s(x^*; y, z) \geq s(x^*; y^*, z^*) \geq s(x; y^*, z^*). \tag{20}$$

Expression (20) states that the function s is maximized with respect to x at $x = x^*$, $(y, z) = (y^*, z^*)$ and minimized with respect to (y, z) at $(x^*, y^*, z^*), (y, z) \geq 0$. This makes the point (x^*, y^*, z^*) a *saddle point* of the function s.

Converse

Given the problem

$$\max f(x) \text{ subject to } g_i(x) \geq b_i, i \in I$$

consider the function

$$s(x, \lambda) = f(x) + \lambda \cdot (g(x) - b) \tag{21}$$

where $g(x) = (g_1(x), \ldots, g_I(x)), b = (b_1, \ldots, b_I) .(x^*, \lambda^*)$ is a *saddle point* if $\lambda^* \geq 0$ and s is maximized with respect to x and minimized with respect to λ at (x^*, λ^*). That is,

$$s(x, \lambda^*) \leq s(x^*, \lambda^*) \leq s(x^*, \lambda) \tag{22}$$

for all $\lambda \geq 0$.

Theorem

If (x^*, λ^*) is a saddle point of s, then x^* solves $\max f(x)$ *subject to* $g_i(x) \geq b_i$.

Since $s(x^*, \lambda^*) \leq s(x^*, \lambda)$ for all $\lambda \geq 0$,

$$\lambda^* \cdot (g(x^*) - b) \leq \lambda \cdot (g(x^*) - b) \text{ for all } \lambda \geq 0. \qquad (23)$$

Therefore (setting $\lambda = 0$ in inequality (23)) it must be that $\lambda^* \cdot (g(x^*) - b) \leq 0$.

Also: $g_i(x^*) - b_i \geq 0$ for all i.

- Suppose that there exists a k such that $g_j(x^*) - b_j < 0$ and let $\lambda_i = 0$ for $i \neq k$ and $\lambda_k = M$.
- It follows that

$$\lambda^* \cdot (g(x^*) - b) \leq \lambda \cdot (g(x^*) - b) = M\,(g_k(x^*) - b_k)\,.$$

- Impossible, since RHS is small as M grows.

- $\lambda^* \geq 0, g_i(x^*) - b_i \geq 0$ for all i, and $\lambda^* \cdot (g(x^*) - b) \leq 0$ implies $\lambda^* \cdot (g(x^*) - b) = 0$.
- So $s(x^*, \lambda^*) \leq s(x^*, \lambda)$ implies that x^* is feasible ($g(x^*) \geq b$) and that λ^*, x^* satisfies

$$\lambda^* \cdot (g(x^*) - b) = 0. \tag{24}$$

- To show: $g(x) \geq b$ implies that $f(x^*) \geq f(x)$.

- $s(x^*, \lambda^*) \geq s(x, \lambda^*)$
- $$s(x^* \lambda^*) = f(x^*) + \lambda^* \cdot (g(x^*) - b) = f(x^*).$$
- $$s(x, \lambda^*) = f(x) + \lambda^* \cdot (g(x) - b)$$
- if $g(x) - b \geq 0$, $\lambda^* \geq 0$ then $s(x, \lambda^*) \geq f(x)$.
- Hence

$$f(x^*) = s(x^*, \lambda^*) \geq s(x, \lambda^*) \leq f(x) \text{ whenever } g(x) \geq b.$$

The saddle point property guarantees optimality.

Converse

When can we say that if x^* solves $\max f(x)$ subject to $g(x) \geq b$, that there exists $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point?

Theorem

If f and g_i are concave and x^* solves $\max f(x)$ subject to $g(x) \geq b$ and a "constraint qualification" holds, then there exists $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point of $s(x, \lambda) = f(x) + \lambda \cdot (g(x) - b)$.

Comments

1. The "constraint qualification" plays the role that "regularity" played in the "inward normal" discussion.
2. This result is often called the Kuhn-Tucker Theorem. λ are called Kuhn-Tucker multipliers.
3. It is possible to prove essentially the same theorem with weaker assumptions (quasi concavity of g).
4. The beauty of this result is that it transforms a constrained problem into an unconstrained one. Finding x^* to maximize $s(x, \lambda^*)$ for λ^* fixed is relatively easy.
5. If f and g_i are differentiable, then the first-order conditions for maximizing s are the familiar ones:
$$\nabla f(x^*) + \lambda \cdot \nabla g(x^*) = 0.$$
6. The FOC that characterization the minimization problem are precisely the complementary slackness conditions. Since $s(x, \lambda)$ is concave in x when f and g are concave, it will turn out that
$$\nabla f(x^*) + \lambda \cdot \nabla g(x^*) = 0, \lambda^* \cdot (g(x^*) - b) = 0, g(x^*) \geq 0$$

We assumed concavity and we need to find "multipliers."
It looks like we need a separation theorem.
Let

$$K = \{(y,z): \text{ there exists } x \text{ such that } \; g(x) \geq y \text{ and } f(x) \geq z\}$$

1. K is convex since g_i and f are concave
2. $y \in \mathbb{R}^m$ and $z \in \mathbb{R}$.
3. If x^* solves the maximization problem, then $(b, f(x^*)) \in K$.
4. In fact, a boundary point. [Otherwise there exists $\varepsilon > 0$ and x such that $g_i(x) \geq b_i + \varepsilon > b_i$ for all i and $f(x) \geq f(x^*) + \varepsilon$. This contradicts the optimality of x^*.]

Hence there exists $(\lambda, \mu) \neq 0, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}$ such that

$$\lambda \cdot b + \mu f(x^*) \geq \lambda \cdot y \geq \mu z$$

for all $(y,z) \in K$.

1. $(\lambda, \mu) \geq 0$.
   If some component of λ, say $\lambda_j < 0$, you get a contradiction taking points $(y, z) \in K$ where, for example, $z = f(x^*), y_i = b_i, i \neq j, y_j = -M$. If M is large enough, this choice contradicts

$$\lambda \cdot b + \mu f(x^*) \geq \lambda \cdot y + \mu z.$$

   If $\mu < 0$ you obtain a contradiction because $(b, z) \in K$ for all arbitrarily small (large negative) z.
2. So $(\lambda, \mu) \geq 0, (\lambda, \mu) \neq 0$ and
3.

$$\lambda \cdot b + \mu f(x^*) \geq \lambda \cdot y + \mu z$$

   for all $(y, z) \in K$.
4. Also:

$$\lambda \cdot (g(x^*) - b) = 0$$

   T because $(g(x^*), f(x^*)) \in K$.
5. Hence

$$\lambda \cdot b + \mu f(x^*) \geq \lambda \cdot g(x^*) + \mu f(x^*).$$

6. and $0 \geq \lambda \left(g(x^*) - b \right)$.

Now **suppose** that $\mu > 0$. Put $\lambda^* = \lambda/\mu$. We have

$$f(x^*) + \lambda^* \cdot b \geq f(x) + \lambda^* \cdot g(x) \text{ for all } x \tag{25}$$

(since $(g(x), f(x)) \in K$) and $\lambda^* \cdot (g(x^*) - b) = 0$.
Now we can confirm the saddle-point property:

$$\begin{aligned} s(x^*, \lambda) \geq s(x^*, \lambda^*) \text{ for } \lambda \geq 0 \equiv \\ f(x^*) + \lambda \cdot (g(x^*) - b) \geq f(x^*) + \lambda^* \cdot (g(x^*) - b) \equiv \\ \lambda \cdot (g(x^*) - b) \geq 0 \text{ for } \lambda \geq 0. \end{aligned}$$

where we use $\lambda^* \cdot (g(x^*) - b) = 0$ and the last line holds because $g(x^*) \geq b$.

$$\begin{aligned} s(x^*, \lambda^*) \geq s(x, \lambda^*) \equiv \\ f(x^*) + \lambda^* \cdot (g(x^*) - b) \geq f(x) + \lambda^* \cdot (g(x) - b) \equiv \\ f(x^*) + \lambda^* \cdot b \ \geq f(x) + \lambda^* \cdot g(x) \end{aligned}$$

since $\lambda^* \cdot (g(x^*) - b) = 0$.

Constraint Qualification

When is $\mu > 0$? We have

$$\lambda \neq 0 \text{ such that } \lambda \cdot b \geq \lambda \cdot y \text{ for all } y \text{ such that } (y, z) \in K \text{ for some } z. \tag{26}$$

One condition that guarantees this cannot happen is:

$$\text{there exist } x^* \text{ such that } g_i(x^*) > b_i \text{ for all } i.$$

This condition is not consistent with (26) because we cannot have $\lambda \cdot b \geq \lambda \cdot g(x^*)$ if $\lambda \geq 0$, $\lambda \neq 0$, and $g_i(x^*) > b_i$ for all i.
The constraint qualification is fairly weak. It just requires that the constraint set has a non-empty interior.