Caltech Division of the Humanities and Social Sciences

Constrained maxima and saddlepoints

KC Border

v. 2016.10.21::14.17

Definition 1 *Let $\varphi\colon X \times Y \to \boldsymbol{R}$. A point (x^*, y^*) in $X \times Y$ is a* ***saddlepoint of φ (over $X \times Y$)*** *if it satisfies*

$$\varphi(x, y^*) \leqslant \varphi(x^*, y^*) \leqslant \varphi(x^*, y) \quad \textit{for all} \quad x \in X,\ y \in Y.$$

That is, (x^*, y^*) is a saddlepoint of φ if x^* maximizes $\varphi(\cdot, y^*)$ over X and y^* minimizes $\varphi(x^*, \cdot)$ over Y. Saddlepoints of a function have the following nice interchangeability property.

Lemma 2 (Interchangeability of saddlepoints) *Let $\varphi\colon X \times Y \to \boldsymbol{R}$, and let (x_1, y_1) and (x_2, y_2) be saddlepoints of φ. Then (x_1, y_2) and (x_2, y_1) are also saddlepoints. Moreover*

$$\varphi(x_1, y_1) = \varphi(x_2, y_1) = \varphi(x_1, y_2) = \varphi(x_2, y_2).$$

Proof: We are given that

$$\varphi(x, y_1) \underset{(1a)}{\leqslant} \varphi(x_1, y_1) \underset{(1b)}{\leqslant} \varphi(x_1, y) \qquad x \in X,\ y \in Y, \tag{1}$$

and

$$\varphi(x, y_2) \underset{(2a)}{\leqslant} \varphi(x_2, y_2) \underset{(2b)}{\leqslant} \varphi(x_2, y) \qquad x \in X,\ y \in Y. \tag{2}$$

Evaluating (1a) at $x = x_2$ yields

$$\varphi(x_2, y_1) \leqslant \varphi(x_1, y_1) \tag{3}$$

and evaluating (2b) at $y = y_1$ yields

$$\varphi(x_2, y_2) \leqslant \varphi(x_2, y_1) \tag{4}$$

Combining these yields

$$\varphi(x, y_2) \underset{(2a)}{\leqslant} \varphi(x_2, y_2) \underset{(4)}{\leqslant} \varphi(x_2, y_1) \underset{(3)}{\leqslant} \varphi(x_1, y_1) \underset{(1b)}{\leqslant} \varphi(x_1, y) \qquad x \in X,\ y \in Y,$$

which implies that (x_2, y_1) is a saddlepoint. By symmetry, so is (x_1, y_2). The proof that φ assumes the same value at each of these four pairs follows from similar reasoning. ∎

Saddlepoints play an important role in the analysis of constrained maximum problems via Lagrangean functions.

Definition 3 *Given $f, g_1, \dots, g_m\colon C \to \boldsymbol{R}$, the associated* ***Lagrangean*** *$L\colon C \times \Lambda \to \boldsymbol{R}$ is defined by*

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x) = f(x) + \lambda \cdot g(x),$$

where Λ is an appropriate subset of $\boldsymbol{R}^{\mathrm{m}}$. (Usually $\Lambda = \boldsymbol{R}^{\mathrm{m}}$ or $\boldsymbol{R}^{\mathrm{m}}_{+}$.) The components of λ are called ***Lagrange multipliers****.*

The first result is that saddlepoints of Lagrangeans are constrained maxima. This result makes no restrictive assumptions on the domain or the functions.

Theorem 4 (Lagrangean saddlepoints are constrained maxima) *Let X be an arbitrary set, and let $f, g_1, \ldots, g_m \colon X \to \boldsymbol{R}$. Suppose that (x^*, λ^*) is a saddlepoint of the Lagrangean $L(x,\lambda) = f + \lambda \cdot g$ (over $X \times \boldsymbol{R}^{\mathrm{m}}_{+}$). That is,*

$$L(x,\lambda^*) \underset{(5a)}{\leqslant} L(x^*,\lambda^*) \underset{(5b)}{\leqslant} L(x^*,\lambda) \qquad x \in X,\ \lambda \geqq 0. \tag{5}$$

Then x^* maximizes f over X subject to the constraints $g_j(x) \geqslant 0$, $j = 1, \ldots, m$, and furthermore

$$\lambda^*_j g_j(x^*) = 0 \quad j = 1, \ldots, m. \tag{6}$$

Proof: Inequality (5b) implies $\lambda^* \cdot g(x^*) \leqslant \lambda \cdot g(x^*)$ for all $\lambda \geqq 0$. Therefore $g(x^*) \geqq 0$ (why?), so x^* satisfies the constraints. Setting $\lambda = 0$, we see that $\lambda^* \cdot g(x^*) \leqslant 0$. This combined with $\lambda \geqq 0$ and $g(x^*) \geqq 0$ implies $\lambda^* \cdot g(x^*) = 0$. Indeed it implies $\lambda^*_j g_j(x^*) = 0$ for $j = 1, \ldots, m$.

Now note that (5a) implies $f(x) + \lambda^* \cdot g(x) \leqslant f(x^*)$ for all x. Therefore, if x satisfies the constraints, $g(x) \geqq 0$, we have $f(x) \leqslant f(x^*)$, so x^* is a constrained maximizer. ∎

Condition (6) implies that if the multiplier λ^*_j is strictly positive, then the corresponding constraint is **binding**, $g_j(x^*) = 0$; and if a constraint is **slack**, $g_j(x^*) > 0$, then the corresponding multiplier satisfies $\lambda^*_j = 0$. These conditions are sometimes called the **complementary slackness** conditions.

The converse of Theorem 4 is not quite true, but almost. To state the correct result we now introduce the notion of a generalized Lagrangean.

Definition 5 *A **generalized Lagrangean** $L_\mu \colon C \times \Lambda \to \boldsymbol{R}$ is defined by*

$$L_\mu(x,\lambda) = \mu f(x) + \sum_{j=1}^{m} \lambda_j g_j(x),$$

where $\mu \geqslant 0$ and again Λ is an appropriate subset of $\boldsymbol{R}^{\mathrm{m}}$.

Note that each choice of μ generates a different generalized Lagrangean. However, as long as $\mu > 0$, a point (x, λ) is a saddlepoint of the Lagrangean if and only if it is a saddlepoint of the generalized Lagrangean. Thus the only case to worry about is $\mu = 0$.

The next results state that for concave functions satisfying a regularity condition, constrained maxima are saddlepoints of some generalized Lagrangean.

Theorem 6 (Concave constrained maxima are nearly Lagrangean saddlepoints) *Let $C \subset \boldsymbol{R}^{\mathrm{n}}$ be convex, and let $f, g_1, \ldots, g_m \colon C \to \boldsymbol{R}$ be concave. Suppose x^* maximizes f subject to the constraints $g_j(x) \geqslant 0$, $j = 1, \ldots, m$. Then there exist real numbers $\mu^*, \lambda^*_1, \ldots, \lambda^*_m \geqslant 0$, not all zero, such that (x^*, λ^*) is a saddlepoint of the generalized Lagrangean L_{μ^*}. That is,*

$$\mu^* f(x) + \sum_{j=1}^{m} \lambda^*_j g_j(x) \underset{(7a)}{\leqslant} \mu^* f(x^*) + \sum_{j=1}^{m} \lambda^*_j g_j(x^*) \underset{(7b)}{\leqslant} \mu^* f(x^*) + \sum_{j=1}^{m} \lambda_j g_j(x^*) \tag{7}$$

for all $x \in C$ and all $\lambda_1, \ldots, \lambda_m \geqslant 0$. Furthermore

$$\sum_{j=1}^{m} \lambda^*_j g_j(x^*) = 0. \tag{8}$$

Proof: Since x^* is a constrained maximizer there is no $x \in C$ satisfying $f(x) - f(x^*) > 0$ and $g(x) \geqq 0$. Therefore the Concave Alternative (see the notes on Separating Hyperplane Theorems) implies the existence of nonnegative $\mu^*, \lambda_1^*, \dots, \lambda_m^*$, not all zero, satisfying

$$\mu^* f(x) + \sum_{j=1}^{m} \lambda_j^* g_j(x) \leqslant \mu^* f(x^*) \quad \text{for every } x \in C.$$

Evaluating this at $x = x^*$ yields $\sum_{j=1}^{m} \lambda_j^* g_j(x^*) \leqslant 0$. But each term in this sum is the product of two nonnegative terms, so (8) holds. This in turn implies (7a). Given that $g_j(x^*) \geqslant 0$ for all j, (8) also implies (7b). ∎

Corollary 7 (When constrained maxima are true Lagrangean saddlepoints) *Under the hypotheses of Theorem 6 suppose in addition that* ***Slater's Condition****,*

$$\exists \bar{x} \in C \quad g(\bar{x}) \gg 0, \tag{$\boldsymbol{S}$}$$

is satisfied. Then $\mu^* > 0$, and may be taken equal to 1. Consequently $x^*, \lambda_1^*, \dots, \lambda_m^*$ is a saddlepoint of the Lagrangean for $x \in C$, $\lambda \geqq 0$. That is,

$$L(x, \lambda^*) \leqslant L(x^*, \lambda^*) \leqslant L(x^*, \lambda) \quad x \in C, \quad \lambda \geqq 0, \tag{9}$$

where $L(x, \lambda) = f(x) + \lambda \cdot g(x)$.

Proof: Suppose $\mu^* = 0$. Then evaluating (7) at $x = \bar{x}$ yields $\lambda^* \cdot g(\bar{x}) \leqslant 0$, but $g(\bar{x}) > 0$ implies $\lambda_j^* = 0$, $j = 1, \dots, m$. Thus $\mu = 0$ and $\lambda_j = 0$, $j = 1, \dots, m$, a contradiction. Therefore $\mu > 0$, and by dividing the Lagrangean by μ, we may take $\mu = 1$. The remainder is then just Theorem 6. ∎

Karlin [1, vol. 1, Theorem 7.1.1, p. 201] proposed the following alternative to Slater's Condition:

$$\forall \lambda > 0 \; \exists \bar{x}(\lambda) \in C \qquad \lambda \cdot g\big(\bar{x}(\lambda)\big) > 0,$$

which we may as well call **Karlin's condition.**

Theorem 8 *Let $C \subset \boldsymbol{R}^{\mathrm{n}}$ be convex, and let $g_1, \dots, g_m \colon C \to \boldsymbol{R}$ be concave. Then g satisfies Slater's Condition if and only it satisfies Karlin's Condition.*

Proof: Clearly Slater's Condition implies Karlin's. Now suppose g violates Slater's Condition. Then by the Concave Alternative Theorem, it must also violate Karlin's. ∎

The next example shows what can go wrong when Slater's Condition fails.

Example 9 In this example, due to Slater [2], $C = \boldsymbol{R}$, $f(x) = x$, and $g(x) = -(1-x)^2$. Note that Slater's Condition fails because $g \leqslant 0$. The constraint set $[g \geqslant 0]$ contains only 1. Therefore f attains a constrained maximum at 1. There is however no saddlepoint at all of the Lagrangean

$$L(x, \lambda) = x - \lambda(1-x)^2 = -\lambda + (1 + 2\lambda)x - \lambda x^2.$$

To see this, observe the first order condition for a maximum in x is $\frac{\partial L}{\partial x} = 0$, or $1 + 2\lambda - 2\lambda x = 0$, which implies $x > 1$ since $\lambda \geqslant 0$. But for $x > 1$, $\frac{\partial L}{\partial \lambda} = -(1-x)^2 < 0$, so no minimum with respect to λ exists. □

The role of Slater's Condition

In this section we present a geometric argument that illuminates the role of Slater's Condition in the Saddlepoint Theorem. Let us return to the underlying argument used in its proof. Define the function $h\colon C \to \boldsymbol{R}^{m+1}$ by

$$h(x) = \big(f(x) - f(x^*), g_1(x), \dots, g_m(x)\big)$$

and set

$$H = \{h(x) : x \in C\} \quad \text{and} \quad \hat{H} = \{y \in \boldsymbol{R}^{m+1} : \exists x \in C \quad y \leqq h(x)\}.$$

Then $\hat{H}$ is a convex set bounded by H. Figure 1 depicts the set $\hat{H}$ for Slater's example, where $f(x) - f(x^*)$ is plotted on the horizontal axis and $g(x)$ is plotted on the vertical axis. Now if

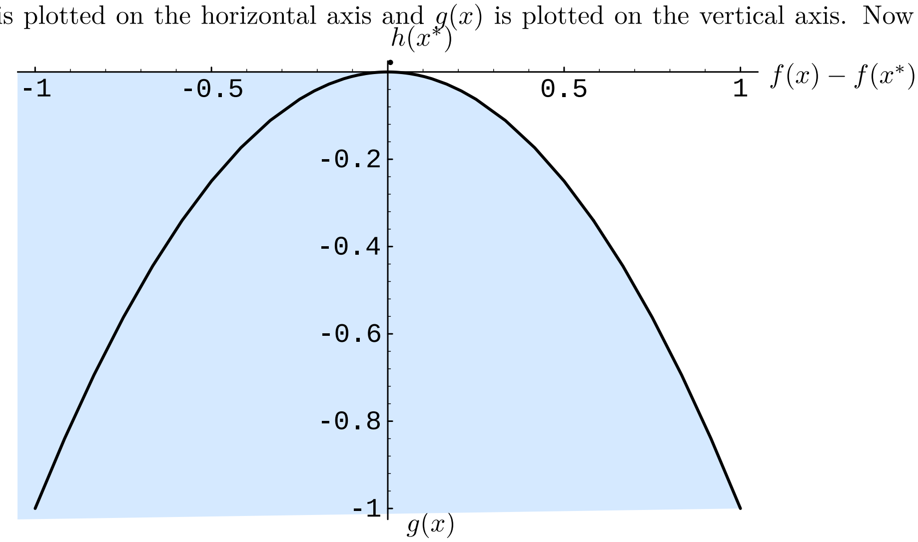

Figure 1. The set $\hat{H}$ for Slater's example.

x^* maximizes f over the convex set C subject to the constraints $g_j(x) \geqslant 0$, $j = 1, \dots, m$, then $h(x^*)$ has the largest horizontal coordinate among all the points in H whose vertical coordinates are nonnegative.

There is a semipositive $m{+}1$-vector $\hat{\lambda}^* = (\mu^*, \lambda_1^*, \dots, \lambda_m^*)$ obtained by separating the convex set $\hat{H}$ and $\boldsymbol{R}^{m+1}_{++}$. It has the property that

$$\hat{\lambda}^* \cdot h(x) \leqslant \hat{\lambda}^* h(x^*)$$

for all $x \in C$. That is, the vector $\hat{\lambda}^*$ defines a hyperplane through $h(x^*)$ such that the entire set $\hat{H}$ lies in one half-space. It is clear in the case of Slater's example that the hyperplane is a horizontal line, since it must be tangent to H at $h(x^*) = (0, 0)$. The fact that the hyperplane is horizontal means that μ^* (the multiplier on f) must be zero.

If there is a nonhorizontal hyperplane through $h(x^*)$, then μ^* is nonzero, so we can divide by it and obtain a full saddlepoint of the true Lagrangean. This is where Slater's condition comes in.

In the one dimensional, one constraint case, Slater's Condition reduces to the existence of $\bar{x}$ satisfying $g(\bar{x}) > 0$. This rules out having a horizontal supporting line through x^*. To see this, note that the first (horizontal) component of $h(x^*) = f(x^*) - f(x^*) = 0$. If $g(x^*) = 0$, then the horizontal line through $h(x^*)$ is simply the horizontal axis, which cannot be, since $h(\bar{x})$ lies above the axis. If $g(x^*) > 0$, then $\hat{H}$ includes every point below $h(x^*)$, so the only line separating $\hat{H}$ and $\boldsymbol{R}^2_{++}$ is vertical, not horizontal. See Figure 2.

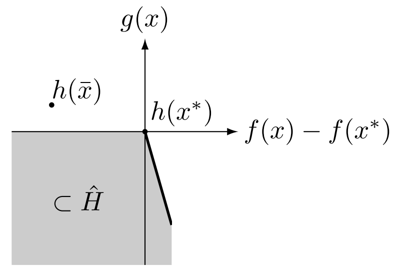

Figure 2. Slater's condition guarantees a nonhorizontal supporting line.

In Figure 2, the shaded area is included in $\hat{H}$. For instance, let $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x + 1$. Then the set $\hat{H}$ is just $\{y \in \boldsymbol{R}^2 : y \leqq (0, 1)\}$.

Later we shall see that if f and the g_js are linear, then Slater's Condition is not needed to guarantee a nonhorizontal supporting line. Intuitively, the reason for this is that for the linear programming problems considered, the set $\hat{H}$ is polyhedral, so even if $g(x^*) = 0$, there is still a nonhorizontal line separating $\hat{H}$ and $\boldsymbol{R}^{\mathrm{m}}_{++}$. The proof of this fact relies on our earlier results on linear inequalities. It is subtle because Slater's condition rules out a horizontal supporting line. In the linear case, there may be a horizontal supporting line, but if there is, there is also a nonhorizontal supporting line that yields a Lagrangean saddlepoint. As a case in point, consider $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x$. Then the set $\hat{H}$ is just $\{y \in \boldsymbol{R}^2 : y \leqq 0\}$, which is separated from $\boldsymbol{R}^2_{++}$ by every semipositive vector.

References

[1] S. Karlin. 1987. *Mathematical methods and theory in games, programming, and economics*. New York: Dover. Reprint of the 1959 two-volume edition published by Addison–Wesley.

[2] M. L. Slater. 1950. Lagrange multipliers revisited: A contribution to non-linear programming. Discussion Paper Math. 403, Cowles Commission. Reissued as Cowles Foundation Discussion Paper #80 in 1959. cowles.econ.yale.edu/P/cd/d00b/d0080.pdf