Optimization: Existence

Reading: Sundaram
Ch 3, Ch 4.1-4.5, (4.6 optional)
Ch 5.1-5.5 (5.2.3, 5.6 optional), Ch. 6.1-6.4 (6.1.2, 6.1.3 optional)

Weierstrass' Theorem
Let $S \subseteq \mathcal{R}^N$ be a compact set and $f : \mathcal{R}^N \to \mathcal{R}$ is a continuous function. There is a maximum and minimum of f in S. That is there exists $m_1, m_2 \in S$ such that $f(m_2) \leq f(x) \leq f(m_1)$ for all $x \in S$. These are also called **global maximum and global minimum**.

Proof: First we shall show that $f(S)$ is a compact set. Then we shall show that compact sets in $\mathcal{R}$ has maximum and minimum.
Step 1: Take any sequence $\{y_n\}_{n=1}^{\infty}$ in $f(S)$. $y_n \in f(S)$ implies there is $x_n \in S$ such that $y_n = f(x_n)$. Take the sequence $\{x_n\}_{n=1}^{\infty}$ in S. Since S is compact, $\{x_n\}_{n=1}^{\infty}$ has a subsequence $\{z_n\}_{n=1}^{\infty}$ which converges to a point z in S. Since f is continuous $\{f(z_n)\}_{n=1}^{\infty}$ converges to $f(z)$. Of course $\{f(z_n)\}_{n=1}^{\infty}$ is a subsequence of $\{y_n\}_{n=1}^{\infty}$ and $f(z) \in f(S)$ (because $z \in S$). Thus every sequence in $f(S)$ has convergent subsequence implying $f(S)$ is compact.

Step 2: $f(S)$ is compact in $\mathcal{R}$ implies $f(S)$ is closed and bounded. $f(S)$ is bounded implies $f(S)$ has a supremum (say M_1) and an infimum (say M_2). Since M_1 (similarly M_2) satisfies ϵ-test, it is a limit point of $f(S)$. Since $f(S)$ is closed, M_1 must be in $f(S)$. Thus there exists $m_1 \in S$ such that $f(m_1) = M_1 \geq f(x)$ for all $x \in S$.

Optimization: Existence

Examples

1. $f : \mathcal{R} \to \mathcal{R}$. $f(x) = x$. There is no maximum. Here f is continuous but domain is not compact because it is not bounded.
2. $f : (0, 1) \to \mathcal{R}$. $f(x) = x$. There is no maximum. f is continuous but once again domain is not compact because it is not closed.
3. $f : [0, 1] \to \mathcal{R}$. $f(x) = x$ for all $x \neq 1$ and $f(1) = 0$. There is no maximum. Here domain is compact but f is not continuous.

Weierstrass' Theorem gives 'sufficient condition' for existence of maximum and minimum. It is not necessary to have these properties for existence of optimum.

4. Take $f : [0, 1] \to \mathcal{R}$. $f(x) = x$ for all $x \neq 1/2$ and $f(1/2) = 0$. This has maximum and minimum.

Finding Interior Optima

Interior Optima

Let $f : S \to \mathcal{R}$, where $S \subseteq \mathcal{R}^N$. **We want to maximize (or minimize) f over S.** We shall focus on maximum. Note that minimum of f is the same as maximum of $-f$.

Instead of finding global maximum, we start from finding **interior local maxima** and use them to find global maximum.

$x^\star$ is called an **interior local maximum** if
(i) $x^\star$ is an interior point of S
(ii) There exists $\epsilon > 0$ such that $f(x^\star) \geq f(x)$ for all $x \in B_\epsilon(x^\star) \subseteq S$.
Note: At this moment we are excluding points that are not interior point of S.
Note: It is strict local maximum if the above inequality is strict.

First order necessary condition:
$x^\star$ is interior local maximum (or minimum) and f is differentiable at $x^\star \Rightarrow Df(x^\star) = \mathbf{0}$
Proof: Take $N = 1$. Take two sequence $\{y_n\}_{n=1}^\infty$ and $\{z_n\}_{k=1}^\infty$ such that $\lim y_n = \lim z_n = x^\star$; $y_n < x^\star$ and $z_n > x^\star$ for all n.

$$\frac{f(y_n) - f(x^\star)}{y_n - x^\star} \geq 0 \geq \frac{f(z_n) - f(x^\star)}{z_n - x^\star}$$

Taking limit, we get $f'(x^\star) \geq 0 \geq f'(x^\star)$. Thus $f'(x^\star) = 0$.
When $N > 1$, follow the same proof for every j. We get $\frac{\partial f}{\partial x_j} = 0$ for all $j = 1, 2, \ldots, N$.

Observation: This is not a sufficient condition. Example: $S = [-1, 1]$, $f(x) = x^3$. There is a critical point $x = 0$ but it is not an optimum.

Digression: Useful Results

Rolle's Theorem: $[a, b] \subset \mathcal{R}$. $f : [a, b] \to \mathcal{R}$ is a continuous function and has derivative at each point of (a, b). If $f(a) = f(b)$ then there is $c \in (a, b)$ such that $f'(c) = 0$
Proof: $[a, b]$ is a compact set and f is a continuous function. Therefore f has a global maximum and minimum in $[a, b]$. If the maximum or minimum is an interior point, then we have obtained a $c \in (a, b)$ such that $f'(c) = 0$. Otherwise both maximum and minimum are obtained at a and b, which implies that f is a constant function in $[a, b]$. Hence for all $c \in (a, b)$, $f'(c) = 0$.

Intermediate Value Theorem: $[a, b] \subset \mathcal{R}$. $f : [a, b] \to \mathcal{R}$ is a continuous function and has derivative at each point of (a, b). Then there is $c \in (a, b)$ such that $\frac{f(b)-f(a)}{b-a} = f'(c)$
Proof: Define $h(x) = f(x)(b - a) - x[f(b) - f(a)]$. Note that h is continuous on $[a, b]$ and has derivative at each point of (a, b). Moreover $h(a) = h(b)$. Using Rolle's Theorem, there is a $c \in (a, b)$ such that $h'(c) = f'(c)(b - a) - [f(b) - f(a)] = 0$

Result: $[a, b] \subset \mathcal{R}$. $f : [a, b] \to \mathcal{R}$ is a continuous function and has derivative at each point of (a, b).
(i) If $f'(x) > 0$ for all $x \in (a, b)$ then f is strictly increasing on $[a, b]$.
(ii) If $f'(x) < 0$ for all $x \in (a, b)$ then f is strictly decreasing on $[a, b]$.
(iii) If $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on $[a, b]$.
Proof: We just proof part (i). The rest are similar. Take any two points $x_1 < x_2$ in the interval $[a, b]$. By Intermediate Value Theorem, there exists $c \in (x_1, x_2)$ such that $f(x_2) - f(x_1) = (x_2 - x_1)f'(c)$. Right hand side is strictly positive because $x_2 > x_1$ and $f'(c) > 0$. Hence $f(x_2) > f(x_1)$

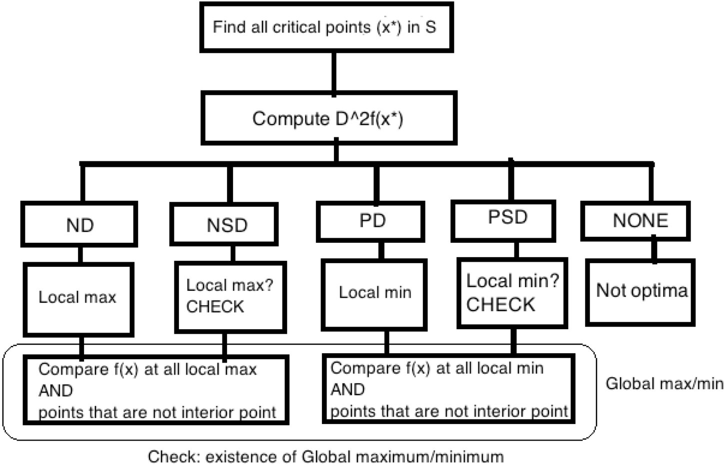
Second Order Conditions

Second order conditions: $f \in \mathcal{C}^2$
1. (necessary) $x^\star$ is interior local maximum (resp. minimum) $\Rightarrow z^T \left[D^2 f\right] z \leq 0$ (resp. $z^T \left[D^2 f\right] z \geq 0$) for all $z \in \mathcal{R}^N$ and $z \neq \mathbf{0}$,
2. (sufficient) $x^\star$ is a critical point and $z^T \left[D^2 f\right] z < 0$ (resp. $z^T \left[D^2 f\right] z > 0$) for all $z \in \mathcal{R}^N$ such that $z \neq \mathbf{0} \Rightarrow x^\star$ is a strict local maximum (resp. strict local minimum)

Proof: $(N = 1)$: First we show part 2. Take an interior point $x^\star$ such that $f'(x^\star) = 0$ and $f''(x^\star) < 0$ (since it is a scalar). Since $f \in \mathcal{C}^2$, f'' is a continuous function. Therefore we can find a neighbourhood around $x^\star$ such that for all $x \in B_r(x^\star)$, $f''(x) < 0$. Hence f' is strictly decreasing in $B_r(x^\star)$. Since $f'(x^\star) = 0$, we must have (i) $f'(z) > 0$ for $z < x^\star$ and $z \in B_r(x^\star)$, (ii) $f'(z) < 0$ for $z > x^\star$ and $z \in B_r(x^\star)$.
$(i) \Rightarrow f$ is strictly increasing when $z < x^\star$ and $z \in B_r(x^\star) \Rightarrow f(x^\star) > f(z)$ for all $z < x^\star$ and $z \in B_r(x^\star)$
$(ii) \Rightarrow f$ is strictly decreasing when $z > x^\star$ and $z \in B_r(x^\star) \Rightarrow f(x^\star) > f(z)$ for all $z > x^\star$ and $z \in B_r(x^\star)$
Combining two cases, we obtain, $f(x^\star) > f(z)$ for all $z \in B_r(x^\star)$. Hence $x^\star$ is a strict local maximum. Part 1 follows immediately from part 2.

$(N > 1)$: We shall only prove part 1. Pick any $z \in \mathcal{R}^N$, $z \neq 0$. Define $g(t) = f(x^\star + tz)$. Note that g is composition of two functions f (which is from $\mathcal{R}^N \to \mathcal{R}$) and ϕ, defined as , $\phi(t) = x^\star + tz$. Note further that $g(0) = f(x^\star) > g(t)$ for all t such that $(x^\star + tz)$ belongs to the neighbourhood that has $x^\star$ as the interior local maximum. Since g is a function from $\mathcal{R} \to \mathcal{R}$, we conclude that $g''(0) \leq 0$. By applying chain rule twice, we get $g''(0) = z^T \left[D^2 f(x^\star)\right] z$. Thus $z^T \left[D^2 f(x^\star)\right] z \leq 0$.

Pathway



Find all critical points (x*) in S
Compute D^2f(x*)
ND
NSD
PD
PSD
NONE
Local max
Local max? CHECK
Local min
Local min? CHECK
Not optima
Compare f(x) at all local max AND points that are not interior point
Compare f(x) at all local min AND points that are not interior point
Global max/min
Check: existence of Global maximum/minimum

Caution

Example: Caution 1: Be aware of necessary conditions.
$S = \mathcal{R}^2$, $f(x_1, x_2) = x_1^3 - 3x_1x_2^2$. Only critical point is $x^\star = (0, 0)$ and $f(0, 0) = 0$

$$D^2 f(x^\star) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

which is both PSD and NSD. However $x^\star$ is neither local min nor local max. For $\epsilon > 0$, $f(\epsilon, 0) > 0$ and $f(-\epsilon, 0) < 0$.

Example: Caution 2: Local maximum does not imply that it is global maximum. Global maximum and minimum may not exist.
$S = \mathcal{R}$, $f(x) = 2x^3 - 3x^2$. Critical points are $x^\star = 0, 1$. Second order conditions imply 0 is a strict local maximum and 1 is strict local minimum.
Our algorithm will say that global maximum is obtained at 0 and global minimum is obtained at 1.
However there is no global maximum or minimum. $f(x)$ strictly increases for $x > 1$ and $x < 0$. Of course a local maximum is obtained at 0 and a local minimum is obtained at 1.

Optimization with Equality Constraints

Suppose that the objective function is $f : \mathcal{R}^N \to \mathcal{R}$ and there are K $(< N)$ constraint functions, $g^1 : \mathcal{R}^N \to \mathcal{R}$, $g^2 : \mathcal{R}^N \to \mathcal{R}, \ldots, g^K : \mathcal{R}^N \to \mathcal{R}$. **We want to maximize (or minimize)** $f(x)$ **over all** $x \in \mathcal{R}^N$ **such that** $g^1(x) = 0, g^2(x) = 0, \ldots, g^K(x) = 0$.

Once again, we start from **local maxima** $x^\star$ that satisfy K constraints. That is
(i) $g^1(x^\star) = 0, g^2(x^\star) = 0, \ldots, g^K(x^\star) = 0$
(ii) There exists an open set U such that $f(x^\star) \geq f(x)$ for all $x \in U$ and $g^1(x) = 0, g^2(x) = 0, \ldots, g^K(x) = 0$.
Note: $K < N$ makes the domain non trivial.

First order necessary condition (Theorem of Lagrange):
Suppose that $f, g^1, \ldots, g^K$ are $\mathcal{C}^1$ functions.
$x^\star$ is local maximum (or minimum) of f $\Rightarrow$ There exists $\lambda_1^\star, \lambda_2^\star, \ldots, \lambda_K^\star \in \mathcal{R}^K$ such that

$$Df(x^\star) + \sum_{j=1}^{K} \lambda_j^\star Dg^j(x^\star) = \mathbf{0}$$

However this hold only when $rank(Dg(x^\star)) = K$. This condition is called *constraint qualification*.

$$Dg = \begin{bmatrix} \frac{\partial g^1}{\partial x_1} & \frac{\partial g^1}{\partial x_2} & \cdots & \frac{\partial g^1}{\partial x_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g^K}{\partial x_1} & \frac{\partial g^K}{\partial x_2} & \cdots & \frac{\partial g^K}{\partial x_N} \end{bmatrix}$$

Second Order Condition

Second order conditions: $f \in \mathcal{C}^2$
$1 - a)$ (necessary) $x^\star$ is interior local maximum $\Rightarrow$
$z^T\left[D^2 f(x^\star) + \sum_{j=1}^K \lambda_j^\star D^2 g^j(x^\star)\right]z \leq 0$ for all $z \in \mathcal{R}^N$ such that $z \neq \mathbf{0}$ and $Dg(x^\star)z = 0$.
$1 - b)$ (necessary) $x^\star$ is interior local minimum $\Rightarrow$
$z^T\left[D^2 f(x^\star) + \sum_{j=1}^K \lambda_j^\star D^2 g^j(x^\star)\right]z \geq 0$ for all $z \in \mathcal{R}^N$ such that $z \neq \mathbf{0}$ and $Dg(x^\star)z = 0$.
$2 - a)$ (sufficient) $x^\star$ is a critical point and $z^T\left[D^2 f(x^\star) + \sum_{j=1}^K \lambda_j^\star D^2 g^j(x^\star)\right]z < 0$ for all $z \in \mathcal{R}^N$, such that $z \neq \mathbf{0}$ and $Dg(x^\star)z = 0 \Rightarrow x^\star$ is a strict local maximum.
$2 - b)$ (sufficient) $x^\star$ is a critical point and $z^T\left[D^2 f(x^\star) + \sum_{j=1}^K \lambda_j^\star D^2 g^j(x^\star)\right]z > 0$ for all $z \in \mathcal{R}^N$, such that $z \neq \mathbf{0}$ and $Dg(x^\star)z = 0 \Rightarrow x^\star$ is a strict local minimum.

Pathway and Caution

Finding global maximum:
Step 1: Find all local maxima using First order condition and the constraints.
Step 2: Compare $f(x^\star)$ at all points obtained in Step 1.
Note 1: There is no need for second order condition.
Note 2: Usually constraint qualification hold at all points in the domain. If it does not there is no way to check whether CQ holds.

Caution 1: Be aware of necessary conditions.
Suppose $f(x_1, x_2) = x_1^3 + x_2^3$ and $g^1(x_1, x_2) = x_1 - x_2$. Note that CQ is met at all points.
First order condition:

$$
\begin{aligned}
3x_1^2 + \lambda_1 &= 0 \qquad (1) \\
3x_2^2 - \lambda_1 &= 0 \qquad (2) \\
x_1 - x_2 &= 0 \qquad (3)
\end{aligned}
$$

Only solution is $x_1 = x_2 = \lambda = 0$. Even second order conditions can not rule out $x_1 = x_2 = \lambda = 0$ as local maximum and minimum. However $(0, 0)$ is neither local maximum nor local minimum. For instance, take $f(\epsilon, \epsilon) = 2\epsilon^3 > 0 = f(0, 0)$, when $\epsilon > 0$.
This also suggests that there is no global maximum and global minimum, which is indeed true. $f(\epsilon, \epsilon) = 2\epsilon^3$ goes to ∞ as $\epsilon \to \infty$

Pathway and Caution

Caution 2: Algorithm may fail because CQ is not satisfied.
Suppose $f(x_1, x_2) = -x_2$ and $g^1(x_1, x_2) = x_2^3 - x_1^2$. This has unique global maximum at $(0, 0)$. Note that CQ fails at $(0, 0)$.
First order condition does not have any solution.

$$-2\lambda_1 x_1 = 0 \quad (4)$$

$$-1 + 3\lambda_1 x_2^2 = 0 \quad (5)$$

$$x_1^2 - x_2^3 = 0 \quad (6)$$

Caution 3: Global maximum and minimum may not exist.
Suppose $f(x_1, x_2) = \frac{1}{3}x_1^3 - \frac{3}{2}x_2^2 + 2x_1$ and $g^1(x_1, x_2) = x_1 - x_2$. First order condition

$$x_1^2 + 2 + \lambda_1 = 0 \quad (7)$$

$$-3x_2 - \lambda_1 = 0 \quad (8)$$

$$x_1 - x_2 = 0 \quad (9)$$

This system of equations have two solutions: $x_1 = x_2 = 1, \lambda_1 = -3$ and $x_1 = x_2 = 2, \lambda_1 = -6$. $f(1, 1) = \frac{5}{6}$ and $f(2, 2) = \frac{2}{3}$. However, global maximum and minimum does not exist. $f(\epsilon, \epsilon)$ is an increasing function when $\epsilon > 2$ and $\epsilon < 1$. Of course $(1, 1)$ is a local maximum and $(2, 2)$ is a local minimum.

Proof of FOC

Sketch of a proof: We prove a special case $N = 2$ and $K = 1$. That is we are maximizing $f(x_1, x_2)$ subject to $g(x_1, x_2) = 0$. We know that f and g are $\mathcal{C}^1$ functions. CQ implies $(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}) \neq (0,0)$ at optimum $(x_1^\star, x_2^\star)$. Assume that $\frac{\partial g}{\partial x_2}(x_1^\star, x_2^\star) \neq 0$.
By **Implicit function theorem**, $\exists h \in \mathcal{C}^1$ such that $x_2 = h(x_1)$ in an open set containing $(x_1^\star, x_2^\star)$. Thus

(A) $g(x_1, h(x_1)) = 0$. By chain rule, $\frac{\partial g}{\partial x_1} + \frac{\partial g}{\partial x_2} h'(x_1) = 0 \Rightarrow h'(x_1^\star) = -\frac{\frac{\partial g}{\partial x_1}(x_1^\star, x_2^\star)}{\frac{\partial g}{\partial x_2}(x_1^\star, x_2^\star)}$

(B) $f(x_1, h(x_1))$ is maximized at $x_1^\star$. By chain rule and unconstrained FOC,
$\frac{\partial f}{\partial x_1}(x_1^\star, x_2^\star) + \frac{\partial f}{\partial x_2}(x_1^\star, x_2^\star) h'(x_1^\star) = 0$
To show that, we can find $\lambda^\star \in \mathcal{R}$ such that

$$\frac{\partial f}{\partial x_1}(x_1^\star, x_2^\star) + \lambda^\star \frac{\partial g}{\partial x_1}(x_1^\star, x_2^\star) = 0 \quad (10)$$

$$\frac{\partial f}{\partial x_2}(x_1^\star, x_2^\star) + \lambda^\star \frac{\partial g}{\partial x_2}(x_1^\star, x_2^\star) = 0 \quad (11)$$

Equation (11) gives $\lambda^\star = -\frac{\frac{\partial f}{\partial x_2}(x_1^\star, x_2^\star)}{\frac{\partial g}{\partial x_2}(x_1^\star, x_2^\star)}$. Remains to show Equation (10) holds for $\lambda^\star$.

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1^\star, x_2^\star) + \lambda^\star \frac{\partial g}{\partial x_1}(x_1^\star, x_2^\star) &= \frac{\partial f}{\partial x_1}(x_1^\star, x_2^\star) + \frac{\partial f}{\partial x_2}(x_1^\star, x_2^\star) h'(x_1^\star) && \text{(by A)} \\ &= 0 && \text{(by B)} \end{aligned}$$

Implicit Function Theorem

Take a $\mathcal{C}^1$ function $\phi : D \to \mathcal{R}$ where $D \subseteq \mathcal{R}^N$ is an open set.

Suppose that $\phi(x_1^\star, x_2^\star, \cdots, x_N^\star) = 0$ and $\frac{\partial \phi}{\partial x_1}(x_1^\star, x_2^\star, \cdots, x_N^\star) \neq 0$.

Then we can find a neighbourhood $V_\delta(x_2^\star, \cdots, x_N^\star) \subseteq \mathcal{R}^{(N-1)}$ and an open set $W \subset \mathcal{R}^1$ containing $x_1^\star$ such that

(i) there exists a $\mathcal{C}^1$ function $\psi : V_\delta(x_2^\star, \cdots, x_N^\star) \to W$ and $x_1^\star = \psi(x_2^\star, \cdots, x_N^\star)$

(ii) moreover, $\frac{\partial \psi}{\partial x_k}(x_2^\star, \cdots, x_N^\star) = -\frac{\frac{\partial \phi}{\partial x_k}(x_1^\star, x_2^\star, \cdots, x_N^\star)}{\frac{\partial \phi}{\partial x_1}(x_1^\star, x_2^\star, \cdots, x_N^\star)}$ for all $k = 2, \cdots, N$

Optimization with Inequality Constraints

Suppose that the objective function is $f : \mathcal{R}^N \to \mathcal{R}$ and there are K constraint functions, $h^1 : \mathcal{R}^N \to \mathcal{R}$, $h^2 : \mathcal{R}^N \to \mathcal{R}, \ldots,$ $h^K : \mathcal{R}^N \to \mathcal{R}$. **We want to maximize (or minimize)** $f(x)$ **over all** $x \in \mathcal{R}^N$ **such that** $h^1(x) \geq 0, h^2(x) \geq 0, \ldots, h^K(x) \geq 0$.

Once again, we start from **local maxima** $x^\star$ that satisfy K constraints. That is
(i) $h^1(x^\star) \geq 0, h^2(x^\star) \geq 0, \ldots, h^K(x^\star) \geq 0$
(ii) There exists an open set U such that $f(x^\star) \geq f(x)$ for all $x \in U$ and $h^1(x) \geq 0, h^2(x) \geq 0, \ldots, h^K(x) \geq 0$

First order necessary condition (Theorem of Kuhn and Tucker):
Suppose that $f, h^1, \ldots, h^K$ are $\mathcal{C}^1$ functions.
$x^\star$ is local maximum of $f \Rightarrow$ There exists $\lambda_1^\star, \lambda_2^\star, \ldots, \lambda_K^\star \in \mathcal{R}^K$ such that
(KT1) $\lambda_j^\star \geq 0$ and $\lambda_j^\star h^j(x^\star) = 0$ for all $j = 1, 2, \ldots, K$
(KT2) $Df(x^\star) + \sum_{j=1}^K \lambda_j^\star Dh^j(x^\star) = \mathbf{0}$
under appropriate CQ condition (which we shall ignore).
In case of minimum replace f by $-f$ to obtain
(KT1) $\lambda_j^\star \geq 0$ and $\lambda_j^\star h^j(x^\star) = 0$ for all $j = 1, 2, \ldots, K$
(KT2) $Df(x^\star) - \sum_{j=1}^K \lambda_j^\star Dh^j(x^\star) = \mathbf{0}$

Finding global maximum:
Step 1: Find all local maxima using First order condition.
Step 2: Compare $f(x^\star)$ at all points obtained in Step 1.
Note: The usual cautions apply.

Envelope Theorem

Envelope Theorem

Objective function $f : \mathcal{R}^{(N+1)} \rightarrow \mathcal{R}$ is a function of variables $(x_1, x_2, \ldots x_N) \in \mathcal{R}^n$ and also a parameter $\alpha \in \mathcal{R}$. The parameter is held constant when f is maximized over $\mathcal{R}^N$ subject to $h(x_1, x_2, \ldots x_N, \alpha) \geq 0$.

Suppose that for every α, there is a unique interior maximizer, so we can say that $x^\star(\alpha)$ is the function that represents this relationship between parameter and maximizer. Suppose $f \in \mathcal{C}^1$ and $x^\star(\alpha)$ is differentiable. $V(\alpha) = f(x^\star(\alpha), \alpha)$ is the maximum value of f given α and is called a **value function**.

Change of $V(\alpha)$ with respect to α is given by
$V'(\alpha) = \frac{\partial \mathcal{L}}{\partial \alpha}(x^\star(\alpha), \lambda^\star(\alpha), \alpha)$, where $\mathcal{L} = f(x, \alpha) + \lambda h(x, \alpha)$

Proof (with a bit of hand-waving): $V(\alpha) = f(x^\star(\alpha), \alpha)$.

By chain rule, $V'(\alpha) = \frac{\partial f}{\partial \alpha}(x^\star(\alpha), \alpha) + \sum_{j=1}^{N} \left[\frac{\partial f}{\partial x_j}(x^\star(\alpha), \alpha) \cdot \frac{dx_j^\star}{d\alpha}(x^\star(\alpha), \alpha) \right]$

By (KT2) $\frac{\partial f}{\partial x_j}(x^\star(\alpha), \alpha) = -\lambda^\star \frac{\partial h}{\partial x_j}(x^\star(\alpha), \alpha)$ for all $j = 1, 2 \ldots n$.

Hence $V'(\alpha) = \frac{\partial f}{\partial \alpha}(x^\star(\alpha), \alpha) - \lambda^\star \sum_{j=1}^{N} \left[\frac{\partial h}{\partial x_j}(x^\star(\alpha), \alpha) \cdot \frac{dx_j^\star}{d\alpha}(x^\star(\alpha), \alpha) \right]$

By (KT1) either $\lambda^\star = 0$ or $h(x^\star(\alpha), \alpha) = 0$

Case 1: If $\lambda^\star = 0$ then $V'(\alpha) = \frac{\partial f}{\partial \alpha}(x^\star(\alpha), \alpha) = \frac{\partial \mathcal{L}}{\partial \alpha}(x^\star(\alpha), \lambda^\star(\alpha), \alpha)$

Case 2: Define $h(x^\star(\alpha), \alpha) = H(\alpha)$. Thus $H(\alpha) = 0$. Differentiating w.r.t α, we get,

$H'(\alpha) = \frac{\partial h}{\partial \alpha}(x^\star(\alpha), \alpha) + \sum_{j=1}^{N} \left[\frac{\partial h}{\partial x_j}(x^\star(\alpha), \alpha) \cdot \frac{dx_j^\star}{d\alpha}(x^\star(\alpha), \alpha) \right] = 0$

Therefore $V'(\alpha) = \frac{\partial f}{\partial \alpha}(x^\star(\alpha), \alpha) + \lambda^\star \frac{\partial h}{\partial \alpha}(x^\star(\alpha), \alpha) = \frac{\partial \mathcal{L}}{\partial \alpha}(x^\star(\alpha), \lambda^\star(\alpha), \alpha)$