Envelope Theorems

Steven A. Matthews*

September 5, 2016

Theorem 1 below applies to optimization problems in which the parameter being studied does not affect the constraint set. Theorem 2 applies when the constraint set does depend on the parameter. The assumptions required for Theorem 2 are stronger and its conclusions weaker – it is basically the standard textbook Envelope Theorem.

1 Envelope Theorem 1

Theorem 1 applies to the following program,

$$(\text{P-}p) \quad V(p) := \max_{x \in X} f(x,p),$$

where f is a real-valued objective function, $x \in \mathbb{R}^n$ is the choice variable, and p is a parameter residing in some set $P \subseteq \mathbb{R}^m$. The only assumption on the constraint set X is that it is nonempty. The function V is called the *value function* of the program.[1]

Theorem 1 establishes conditions under which the upper envelope of a set of convex functions is (i) convex and (ii) has a slope at any point equal to the slope of the supporting function at that point.

Theorem 1

(i) *If P is a convex set and $f(x,\cdot)$ is a convex function of p for each $x \in X$, then V is a convex function.*

(ii) *For any p in the interior of P, and any x^* that solves* (P-p)*, suppose V and $f(x^*,\cdot)$ are differentiable at p. Then these two derivatives are equal:*[2]

$$\nabla V(p) = \nabla_p f(x^*, p).$$

The following subsections discuss, prove, and apply the theorem.

---

*Section 4 of this teaching document has been influenced by the notes of Kim Border on these topics, as cited in the References. I thank him for making his amazing notes available on his web site.

[1] Until Section 3 we assume (P-p) has a solution for each $p \in P$. This is for convenience; if we replace "max" by "sup" in the definition of V, the results of Sections 1 and 2 still hold.

[2] The usual proofs of this "derivative property" unnecessarily assume f is differentiable in x, $x^*(a)$ is a function and differentiable, and $x^*(a)$ is in the interior of X.

1.1 Discussion

Consider part (ii) first. Suppose for simplicity that $p \subseteq \mathbb{R}$, that (P-p) has a unique solution, $x^*(p)$, for each p, and that $x^*(\cdot)$ is a C^1 function (as is implied by the IFT under additional assumptions). Then,

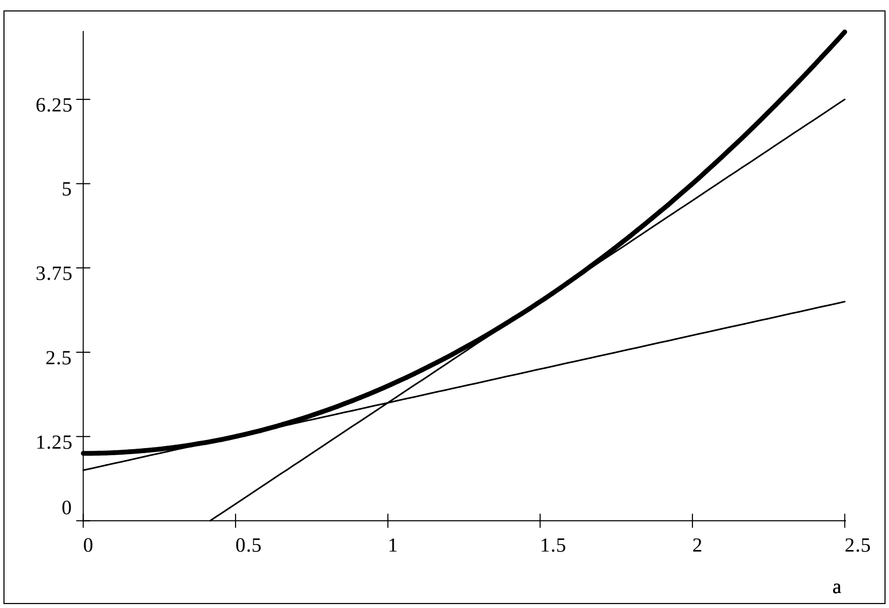
$$V(p) = f(x^*(p), p).$$

Changes in p affect the value function in two ways, directly through the second argument of f, and indirectly through $x^*(\cdot)$ in the first argument. The direct effect is of "first order" (recall Taylor expansions). The indirect effect is of second order: the effect of a change in p on x^* is of first order, but the effect of a change in x^* on f is second order because $x^*(p)$ maximizes $f(\cdot, p)$. Thus, to a first-order approximation, the effect on the value of a change in p is equal to its direct effect. These words are reflected in the following standard derivation, which uses the FOC, $f_x(x^*(p), p) = 0$:

$$\begin{aligned} V'(p) &= f_x(x^*(p), p) \cdot x^{*\prime}(p) + f_p(x^*(p), p) \\ &= f_p(x^*(p), p). \end{aligned}$$

Now turn to part (i) of Theorem 1. Part (i) is about a "dual" picture in which the parameter variable, p, is on the horizontal axis. In this picture, the definition of V is that it (or rather, its graph) is the upper envelope of the collection $\{f(x, \cdot)\}_{x \in X}$ of functions of p. As such, this upper envelope must be convex if each of the functions $f(x, \cdot)$ is convex (even linear).

As an example, suppose $f(x,p) = 2px - x^2 + 1$. Then $x^*(p) = p$ and $V(p) = p^2 + 1$. The graph of V and two of its supporting functions, $f(.5, p) = p + .75$ and $f(1.5, p) = 3p - 1.25$, are shown in the following figure:



Consider $p = .5$. The maximization problem, when $p = .5$, can be viewed as moving up the vertical line at $p = .5$ through the collection of straight lines corresponding to the functions

$f(x,p) = 2px - x^2 + 1$ for various fixed values of x until one gets to the highest such line. This highest line is the graph of $f(.5,p) = p + .75$, and its height at $p = .5$ gives us the value $V(.5) = 1.25$. The graphs of V and $f(.5,\cdot)$ touch at this point, i.e. at $(.5, 1.25)$. To a first order approximation, the two functions are the same in a neighborhood of $p = .5$. In other words, $V'(p) = f_p(.5,p) = 1$ at $p = .5$. If p were to increase from $p = .5$, but x were held fixed at $x^*(.5) = .5$, the value would increase as measured on the straight line $f(.5,p) = p + .75$. But if x is optimally adjusted as p increases, then the higher value, $V(p)$, is achieved. This is why the graph of V curves up and away from each of its supporting lines.

1.2 Proof of Theorem 1

(i) Assume P is a convex set and f is convex in p. Let $p^0, p^1 \in P$, $\lambda \in [0,1]$, and $p = \lambda p^1 + (1-\lambda)p^0$. Since P is convex, $p \in P$, and $f(x,p)$ is well-defined for all $x \in X$. Let x, x^0, and x^1 be solutions when the parameter is p, p^0, and p^1, respectively. Then,

$$
\begin{aligned}
V(p) &= f(x,p) && \text{(by the def of } x\text{)} \\
&\leq \lambda f(x,p^1) + (1-\lambda) f(x,p^0) && \text{(as } f(x,\cdot) \text{ is convex)} \\
&\leq \lambda f(x^1,p^1) + (1-\lambda) f(x^0,p^0) && \text{(by the defs of } x^1 \text{ and } x^0,\ \text{using } \lambda \in [0,1] \\
& && \text{and the independence of } X \text{ from } p\text{)} \\
&= \lambda V(p^1) + (1-\lambda) V(p^0).
\end{aligned}
$$

This proves V is convex.

(ii) Let x^* solve (P-p). Define a function $h : P \to \mathbb{R}$ by

$$h(\hat{p}) := V(\hat{p}) - f(x^*, \hat{p}).$$

Since X does not depend on p, $h(\hat{p}) \geq 0$ for all $p \in P$, and $h(p) = 0$. Hence, p minimizes h on P. By assumption, the derivative of h exists at p, and p is in the interior of P. The first-order condition for this minimization is thus $\nabla h(p) = 0$, or rather,

$$\nabla V(p) - \nabla_p f(x^*, p) = 0. \blacksquare$$

1.3 Applications of Theorem 1

Theorem 1 is used for doing comparative statics on optimization problems when the constraint set does not depend on the parameter in question. The following is a partial list of its comparative statics applications.

1. The profit function, $\pi(p) := \max_{y \in Y} p \cdot y$. Theorem 1(i) implies π is convex, and Theorem 1(ii) implies the derivative property for profit functions (Hotelling's lemma):

$$\frac{\partial \pi(p)}{\partial p_k} = y_k(p). \tag{1}$$

   Convexity and (1), together with the assumption that π is C^2, imply the laws of supply, demand, and reciprocity.

2. The expenditure function,

$$e(p, \bar{u}) := \min_{x \in \mathbb{R}^L_+ \geq 0} p \cdot x \text{ such that } u(x) \geq \bar{u}.$$

Theorem 1(i) implies e is concave in p. Theorem 1(ii) implies the derivative property (Shepard's Lemma) for expenditure functions:

$$\frac{\partial e(p, \bar{u})}{\partial p_i} = h_i(p, \bar{u}). \tag{2}$$

Since e is concave in p, if it is also C^2 in p, then (2) implies that the Hicksian demand functions satisfy the laws of demand and reciprocity.

3. The cost function,

$$c(w, q) := \min_{z \geq 0} w \cdot z \text{ such that } f(z) \geq q.$$

Theorem 1(i) implies e is concave in w. Theorem 1(ii) implies the derivative property (Shepard's Lemma) for cost functions:

$$\frac{\partial c(w, q)}{\partial w_i} = z_i(w, q). \tag{3}$$

Since e is concave in w, if it is alsoc C^2 in w, then (3) implies that the conditional factor demand functions satisfy the laws of demand and reciprocity.

4. In statistical decision theory, Theorem 1(i) implies that the Bayes risk function, $\rho^*(p) := \min_{d \in D} \mathbb{E}_p \rho(d, \omega)$, is concave in the probability distribution p over the state of nature ω.

2 Envelope Theorem 2

Theorem 2 states that when the constraint set does depend upon the parameter, in the form of an inequality constraint, then the derivative of the value function is equal to the partial derivative of the corresponding Lagrangian. This envelope theorem is less useful than the first one because it does not give us curvature properties; it corresponds only to part (ii) of Theorem 1. Theorem 2 gives us our interpretation of Lagrange multipliers as measuring the first order effect on the value of the problem of relaxing the constraint, its "shadow price." It is also Theorem 2 that is used to prove Roy's identity,

$$x_i(p, w) \frac{\partial v(p, w)}{\partial w} = -\frac{\partial v(p, w)}{\partial p_i}.$$

Theorem 2 concerns programs of the following form:

$$(\mathrm{P}'\text{-}p) \quad V(p) = \max_{x \in X} f(x, p) \text{ such that } g(x, p) \geq 0,$$

where f and g are real-valued C^2 functions on some open domain in $\mathbb{R}^n \times \mathbb{R}^m$. (The generalization to multiple constraints is straightforward.)

Theorem 2 *Suppose that for some $p \in \mathbb{R}^m$, a neighborhood N of p and a function $x^* : N \to \mathbb{R}^n$ exist such that for all $\hat{p} \in N$, the point $x^*(\hat{p})$ solves* (P′-$\hat{p}$)*. For each $\hat{p} \in N$, let $\lambda^*(\hat{p})$ be the corresponding Kuhn-Tucker multiplier. Suppose $x^*(\cdot)$ is differentiable at p, and that constraint qualification holds at $(x^*(p), \hat{p})$ for all $\hat{p} \in N$.*[3] *Lastly, suppose that if $g(x^*(p), p) = 0$, then $g(x^*(\hat{p}), \hat{p}) = 0$ for all $\hat{p} \in N$. Then V is differentiable at p, and its derivative at p is equal to the partial derivative of the Langrangian:*

$$\nabla V(p) = \nabla_p f(x^*(p), p) + \lambda^*(p) \nabla_p g(x^*(p), p). \tag{4}$$

Proof. For $\hat{p} \in N$ we have $V(\hat{p}) = f(x^*(\hat{p}), \hat{p})$, and so the derivative at $\hat{p} = p$ is

$$\nabla V = \nabla_x f \cdot x^{*\prime} + \nabla_p f,$$

where we have dropped the arguments of the functions for the sake of clarity. The FOC at p is

$$\nabla_x f + \lambda^* \nabla_x g = \mathbf{0}.$$

Hence,

$$\nabla V = -\lambda^* \nabla_x g \cdot x^{*\prime} + \nabla_p f. \tag{5}$$

Case 1.[4] $g(x^*(p), p) > 0$ (non-binding constraint). In this case, complementry slackness implies $\lambda^* = 0$, and so (5) immediately yields the desired result,

$$\nabla V = \nabla_p f = \nabla_p f + \lambda^* \nabla_p g.$$

Case 2. $g(x^*(p), p) = 0$ (binding constraint). Then, by assumption, $g(x^*(\hat{p}), \hat{p}) = 0$ for all $\hat{p} \in N$. Hence, the derivative of $g(x^*(\hat{p}), \hat{p})$ at $\hat{p} = p$ is zero:

$$\nabla_x g \cdot x^{*\prime} + \nabla_p g = \mathbf{0}.$$

Substituting this into (5) yields the desired result:

$$\begin{aligned} \nabla V &= -\lambda^* \nabla_x g \cdot x^{*\prime} + \nabla_p f \\ &= \lambda^* \nabla_p g + \nabla_p f. \ \blacksquare \end{aligned}$$

3 Differentiability of the Value Function

(This section is both more advanced and sketchy. Definitely optional! Although some of the early arguments are easy and possibly illuminating.)

An important assumption of Theorem 1(ii) is that the value function is differentiable. How can we know it is differentiable? We try to avoid just assuming it is, since it is an endogenous function (unlike the exogenous f). It turns out that often, the value function is differentiable precisely when the program has a unique solution.

[3] For example, $g_x(x(\hat{p}), \hat{p}) \neq 0$ if $g(x(\hat{p}), \hat{p}) = 0$.

[4] This case could also be proved directly by applying Theorem 1. Just take X and A to be small neighborhoods of $x^*(a)$ and a, respectively, such that $g(x, \hat{a}) > 0$ for all $(x, \hat{a}) \in X \times A$.

It is easy to see why uniqueness of the solution may be necessary for V to be differentiable. For, suppose that for some $\bar{p} \in P$, the program (P-$\bar{p}$) has two solutions, x^* and x^{**}. Then by Theorem 1(ii), if V and f are differentiable in p, we have two envelope equalities:

$$\nabla_p f(x^*, \bar{p}) = \nabla V(\bar{p}) = \nabla_p f(x^{**}, \bar{p}). \tag{6}$$

Thus, $x^* = x^{**}$ if $\nabla_p f(\cdot, \bar{p})$ is a one-to-one function. For example, in the important case of $f(x,p) = p \cdot x$, we have $\nabla_p f(x, \bar{p}) = x$, and so (6) immediately yields $x^* = x^{**}$.

A deeper analysis is required to show that the existence of a unique solution to (P-p) implies V is differentiable at p. We only sketch the argument here, for the important special case of $f(x,p) = p \cdot x$. We also modify (P-p) so that it handles cases in which a maximum does not exist, and call the resulting value function μ_X instead of V : For any nonempty $X \subseteq \mathbb{R}^n$ and $p \in \mathbb{R}^n$, define

$$\mu_X(p) := \sup_{x \in X} p \cdot x,$$

where "sup" is short for "supremum. (The function μ_X is called the *support function* of the set X.) Reviewing the proof of Theorem 1(i), note that it is easily modified to apply to (P-p) modified by replacing "max" by "sup," so that μ_X is indeed a convex function.[5] It is also true, and obvious, that μ_X is homogeneous of degree 1.

Theorem 3 *Let $X \subseteq \mathbb{R}^n$ be closed, and also bounded or convex.*[6] *Then for any $\bar{p} \in \mathbb{R}^n$, the function μ_X is differentiable at $\bar{p}$ if and only if there exists one and only one $\bar{x} \in X$ such that $\bar{p} \cdot \bar{x} = \mu_X(\bar{p})$.*

The "only if" direction of this theorem is proved above, using Theorem 1. The "if" direction is harder and deeper. A large part of it is the following result from convex analysis, which requires a few definitions results to state.

The *subgradient* of a convex function $h : \mathbb{R}^n \to \mathbb{R}$ at a point $p \in \mathbb{R}$ is a vector $x \in \mathbb{R}^n$ for which

$$h(q) \geq h(p) + x \cdot (q - p) \text{ for all } q \in \mathbb{R}^n.$$

A convex function has a subgradient at any p (in the relative interior of its domain, which we'll not bother talking about here). The set of all subgradients of h at p is denoted as $\partial h(p)$, and called the *subdifferential* of h at p. If h has only one subgradient x, then h is differentiable at p, and x is its derivative there: $\nabla h(p) = x$.

Lastly, for any $X \subseteq \mathbb{R}^n$, the *closed convex hull* of X, denoted as $\overline{\text{co}}X$, is the intersection of all closed convex sets that contain X, and can be shown to equal the intersection of all closed half spaces that contain X. It can also be shown that $\overline{\text{co}}X$ is equal to the closure of the convex hull of X, co$X$, which is the intersection of all convex sets that contain X. In turn, using Carathéodory's Theorem, co$X$ can be shown to equal the set of all $(n+1)$ convex combinations of points in X.

[5] However, $\mu_X(p) = \infty$ for any p at which $p \cdot x$ does not attain a maximum on X.

[6] In addition to defining the support function as a "sup" rather than the non-standard "inf" (an unimportant difference), this theorem differs from MWG's Proposition 3.F.1 by including the assumption that X is convex or bounded. MWG's proposition is not correct without the addition of one of these assumptions – see the section below.

The required result from convex analysis is that every maximizer of $p \cdot x$ on X is a subgradient of μ_X at p, and every subgradient of μ_X at p is a maximizer of $p \cdot x$ on $\overline{\text{co}}X$:

Theorem 4 *For any* $X \subseteq \mathbb{R}^n$ *and* $p \in \mathbb{R}^n$,

$$\arg\max_{x \in X} p \cdot x \subseteq \partial \mu_X(p) \subseteq \arg\max_{x \in \overline{co}X} p \cdot x. \tag{7}$$

Proof. The first inclusion follows from the definitions of subgradient and μ_X. The proof of the second inclusion first shows that any subgradient of μ_X at p is in $\overline{\text{co}}X$, using the facts that $\overline{\text{co}}X$ is a closed convex set, the Separating Hyperplane Theorem, and the homogeneity of μ_X. This result is then used to show that any subgradient of μ_X at p maximizes $p \cdot x$ on $\overline{\text{co}}X$. The details are left to the reader (or my future self). ■

Theorem 3 follows directly from Theorem 4 in the case X is closed and convex, as then $\overline{\text{co}}X = X$. Thus, in this case the left and right sides of (7) are the same, and so

$$\arg\max_{x \in X} p \cdot x = \partial \mu_X(p).$$

This equality implies that μ_X is differentiable at p, i.e., $\partial \mu_X(p)$ is a singleton set, if and only if $p \cdot x$ has a unique maximizer on X.

It takes more work to show that Theorem 3 implies Theorem 4 in the case X is compact but not necessarily convex. I leave this to the reader as well (or my future self). (Use Carathéodory's Theorem and limits of subsequences of convex combinations of points in X to show that if $p \cdot x$ has a unique maximizer on X, then that point is also the unique maximizer of $p \cdot x$ on $\overline{\text{co}}X$.)

There is also a connection between the differentiability of the value function and uniqueness of the solution to a constrained maximization program, useful for weakening the assumptions of Theorem 2. This I leave to a a future revision of these notes, if and when it occurs.

4 Counterexample to MWG's Proposition 3.F.1

The following is a counterexample to MWG's Proposition 3.F.1, i.e., to Theorem 3 above if the "bounded or convex" assumption is deleted. Think of it's X (which in MWG's notation would be K) as the production set of a competitive firm using input $x_1 \leq 0$ to produce output $x_2 \geq 0$:

$$f(x_1) := \begin{cases} 2x_1^2 & \text{for } x_1 \in [-1, 0] \\ 1 - x_1 + \dfrac{1 + x_1}{x_1^2} & \text{for } x_1 \leq -1 \end{cases}$$

$$X := \{x \in \mathbb{R}^2 : x_1 \leq 0,\ x_2 \leq f(x_1)\}.$$

This X is closed, but neither convex nor bounded. (Draw the figure.)

Let $\bar{p} = (1, 1)$ (be the price vector). Then

$$\mu_X(\bar{p}) = \max_{x \in X}(1, 1) \cdot x = \max_{x_1 \leq 0} f(x_1) + x_1.$$

This program is easily seen to have a unique solution, $\bar{x} = (-1, 2)$, yielding (profit) $\mu_X(\bar{p}) = 1$.

For $p_1 \in (1, 2)$ and $p_2 = 1$, the unique solution is still $\bar{x} = (-1, 2)$ (verify by looking at the figure), yielding profit $\mu_X(\bar{p}) = 2 - p_1$.

However, if the input costs less than 1, there is no upper bound to the profit the firm can obtain: for $p_1 < 1$ and $p_2 = 1$,

$$\begin{aligned}
\lim_{x_1 \to -\infty} [f(x_1) + p_1 x_1] &= \lim_{x_1 \to -\infty} \left[1 + \frac{1 + x_1}{x_1^2} + (p_1 - 1)\, x_1\right] \\
&= 1 + \lim_{x_1 \to -\infty} (p_1 - 1)\, x_1 \\
&= \infty.
\end{aligned}$$

We thus have

$$\mu_X(p_1, 1) = \begin{cases} \infty & \text{for } p_1 < 1 \\ 2 - p_1 & \text{for } p_1 \in [1, 2] \end{cases}.$$

This is clearly not a differentiable function of p_1 at $p_1 = 1$, and so μ_X is not differentiable at $\bar{p} = (1, 1)$. Having a unique solution does not imply the value function is differentiable.

(A mathematically more elegant counterexample can be constructed, but ....)

5 References

The standard reference on convex analysis is

Rockafellar, R. T. (1970), *Convex Analysis*, Princeton University Press.

An amazing source of excellent readable downloadable notes on mathematical economics has been made available by Kim Border at Caltech:

http://people.hss.caltech.edu/~kcb/Notes.shtml

I've used a bit here from the following notes on Kim's site:

"Envelope theorem": http://www.hss.caltech.edu/~kcb/Notes/Envelope.pdf

"Supergradients and subgradients": http://www.hss.caltech.edu/~kcb/Notes/Supergrad.pdf

(Note to me: In "Envelope theorem," Kim has a link to his "on line notes", a long document of miscellaneous results. Of relevance is section 5.10, "Value functions and envelope theorems" concerning the differentiability of of a Lagrangian (when the solution is a saddlepoint) of possible use for the possible future revision of these notes regarding Theorem 2.)