About the "Envelope Theorem"

I have been taken to task by some colleagues because no where in the book do I give the general "Envelope Theorem." Applications of this "theorem" appear in several places, in Propositions 9.22, 10.6, and 11.1. But the general "theorem" is never discussed.

The scare quotes in the previous paragraph signal my excuse: As far as I can tell, there is no single, official result that can be called "The General Envelope Theorem," but instead a number of more or less general results that are given that title. In spirit, they are all of the following form: In the parametric optimization problem

$$\text{Maximize (or minimize) } F(x,\theta) \text{ over } x \in A(\theta),$$

if we write $f(\theta)$ as the maximized value, then under conditions $\ldots$, the function f is differentiable in θ and its derivative is $\ldots$ (involving the partial derivative of F in θ at the optimal value of x for θ.) As you might imagine from the three propositions listed, assumptions about the uniqueness of the solution to the problem at θ will be involved, and the precise connection between the derivative of f and the partial derivative of F will depend on whether and how the constraint set $A(\theta)$ changes with θ. Also, to deal with situations in which the solution to the problem is not unique for all θ but is unique for "most" of them, variations are given in "integrated" form.

Hence, there are a lot of variations.

If and when Volume II is done, I'll need to confront at least some of these variations, as the general result is employed in mechanism-design problems. But, in the meantime, in case you were feeling bereft of a general result of this type, here is a result that is a bit more complex than is the situation in Propositions 9.22 and 10.6 but still without confronting some of the difficulties in Proposition 11.1. (The objective function is pretty general, but the constraint set is constant.) You should be able to prove this on your own after you've consumed the proofs of the three propositions—it is good practice—but in case you need some help, I give the proof.

One variation on "The Envelope Theorem." *Consider the problem*

$$\textit{Maximize } F(x,\theta) \textit{ in } x, \textit{ for fixed } \theta \in \Theta, \textit{ subject to } x \in X.$$

Assume that $F : X \times \Theta \to R$, where $X \subseteq R^n$ is compact and $\Theta \subseteq R^m$ is open. Let $f(\theta)$ be the maximized value; that is,

$$f(\theta) := \mathit{sup}\,\{F(x,\theta); x \in X\}, \textit{ for each } \theta \in \Theta$$

Assume that F is jointly continuous in x and θ and that F is continuously differentiable in θ for each x; assume further that the gradient of F with respect to θ is jointly continuous in x and θ. Then if, for a given θ^0, there is a unique solution to the problem x^0, f is differentiable at θ^0, and

$$\left.\frac{\partial f}{\partial \theta_i}\right|_{(\theta^0)} = \left.\frac{\partial F}{\partial \theta_i}\right|_{(x^0,\theta^0)}.$$

Proof. The first step is to apply Berge's Theorem. This is a parametric optimization problem. Moreover, it is a very simple parametric optimization problem: The objective function is continuous, and the constraint set is compact and doesn't change with the parameter. Hence we know that Berge's Theorem applies; the optimal value function is continuous in θ, and the solution set correspondence is upper semi-continuous (and locally bounded, but since X is compact, that's superfluous).

Moreover, if at θ^0 there is a unique solution x^0, then if x^n is *any* solution at θ^n for a sequence $\{\theta^n\}$ that approaches θ^0, upper semi-continuity of the solution correpondence tells us that $x^n \to x^0$.

Now examine what we want to show: If $\{\theta^n\}$ is a sequence with limit θ^0, we wish to show that

$$\lim_{n\to\infty} \frac{f(\theta^n) - f(\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0)}{\|\theta^n - \theta^0\|} = 0,$$

where $\nabla_\theta F(x^0,\theta^0)$ is the gradient vector of F in θ evaluated at (x^0,θ^0).

Let x^n be any solution at θ^n. We know that

$$f(\theta^n) = F(x^n,\theta^n) \geq F(x^0,\theta^n) \quad \text{and} \quad f(\theta^0) = F(x^0,\theta^0) \geq F(x^n,\theta^0).$$

Therefore, for each n,

$$f(\theta^n) - f(\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0) = F(x^n,\theta^n) - F(x^0,\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0)$$

and therefore

$$\begin{aligned}
&F(x^n,\theta^n) - F(x^n,\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0) \geq \\
&f(\theta^n) - f(\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0) \geq \\
&F(x^0,\theta^n) - F(x^0,\theta^0) - \nabla_\theta F(x^0,\theta^0)\cdot(\theta^n - \theta^0).
\end{aligned}$$

So we have the desired result if we can show that the limit of the terms on the left- and right-hand side of the string of inequalities just given, when divided by $\|\theta^n - \theta^0\|$, are both zero.

Take the left-hand side first. By the exact form of Taylor's Theorem (which is just the mean-value theorem),

$$F(x^n, \theta^n) - F(x^n, \theta^0) = \nabla_\theta F(x^n, \hat{\theta}^n) \cdot (\theta^n - \theta^0),$$

where $\hat{\theta}^n$ is some convex combination of θ^n and θ^0. Hence,

$$F(x^n, \theta^n) - F(x^n, \theta^n) - \nabla_\theta F(x^0, \theta^0) \cdot (\theta^n - \theta^0) = \big[\nabla_\theta F(x^n, \hat{\theta}^n) - \nabla_\theta F(x^0, \theta^0)\big] \cdot (\theta^n - \theta^0).$$

Of course,

$$\big|\big[\nabla_\theta F(x^n, \hat{\theta}^n) - \nabla_\theta F(x^0, \theta^0)\big] \cdot (\theta^n - \theta^0)\big| \leq \big\|\nabla_\theta F(x^n, \hat{\theta}^n) - \nabla_\theta F(x^0, \theta^0)\big\| \times \|\theta^n - \theta^0\|.$$

Hence

$$\lim_{n\to\infty} \left| \frac{F(x^n, \theta^n) - F(x^n, \theta^0) - \nabla_\theta F(x^0, \theta^0) \cdot (\theta^n - \theta^0)}{\|\theta^n - \theta^0\|} \right| \leq$$

$$\lim_{n\to\infty} \left| \frac{\big\|\nabla_\theta F(x^n, \hat{\theta}^n) - \nabla_\theta F(x^0, \theta^0)\big\| \times \|\theta^n - \theta^0\|}{\|\theta^n - \theta^0\|} \right| = \lim_{n\to\infty} \big\|\nabla_\theta F(x^n, \hat{\theta}^n) - \nabla_\theta F(x^0, \theta^0)\big\|.$$

But the last limit is zero, since $(x^n, \hat{\theta}^n) \to (x^0, \theta^0)$, and the gradient of F in θ is (by assumption) jointly continuous in x and θ. The other side is handled in entirely similar fashion, completing the proof.