Envelope Theorems for Non-Smooth and Non-Concave Optimization[*]

Andrew Clausen[†]
University of Pennsylvania

Carlo Strub[‡]
University of St. Gallen

First version: June 9, 2010
This version: January 23, 2012

Abstract

We study general dynamic programming problems with continuous and discrete choices and general constraints. The value functions may have kinks arising (1) at indifference points between discrete choices and (2) at constraint boundaries. Nevertheless, we establish a general envelope theorem: first-order conditions are necessary at interior optimal choices. We only assume differentiability of the utility function with respect to the continuous choices. The continuous choice may be from any Banach space and the discrete choice from any non-empty set.

1 Introduction

Optimization problems that involve both discrete and continuous choices are common in economics. Examples include the trade-off between consumption and savings alongside

---

[*] We are very grateful to Guy Arie, Martin Barbie, Christian Bayer, Aleksander Berentsen, Carlos Carillo-Tudela, Harold Cole, Russell Cooper, Thomas Davoine, Jesús Fernández-Villaverde, Marcus Hagedorn, Christian Hellwig, Leo Kaas, Andreas Kleiner, Dirk Krueger, Tzuo Hann Law, Yvan Lengwiler, Alexander Ludwig, George Mailath, Steven Matthews, Guido Menzio, Nirav Mehta, Leonard Mirman, Kurt Mitman, Georg Nöldeke, Nicola Pavoni, Andrew Postlewaite, José Rodrigues-Neto, Felipe Saffie, Florian Scheuer, Ilya Segal, Shouyong Shi, Lones Smith, Kyungchul Song, Xi Weng and Mark Wright for fruitful discussions.

[†] clausen@sas.upenn.edu

[‡] cs@carlostrub.ch

the discrete decision of whether to work, accept a job offer, declare bankruptcy, go to college, or enroll children in child care.[1] In addition, we show how non-smooth optimization problems, such as capital adjustment in the presence of fixed costs, may be recast as mixed continuous and discrete choice problems. In the absence of lotteries or other smoothing mechanisms, such problems create kinks in the value function where agents are indifferent between two discrete choices. A second type of kink arises when constraints become binding. As a result, the value function is non-differentiable, non-concave, and may even lack directional derivatives. Can first-order conditions be applied under such circumstances?

This paper provides two general *envelope theorems*. The first relates to static optimization problems. Figure 1a illustrates an example where an investor maximizes his profit by choosing the size of his investment, c, and the product, d_1 or d_2, to invest in. The investor takes the upper envelope over the two per-product profits f and maximizes it with respect to the continuous choice c. We assume $f(\cdot, d)$ is differentiable for each discrete choice $d \in \{d_1, d_2\}$. Observe that the upper envelope has only downward kinks but no upward kinks. Moreover, maxima may not occur at downward kinks. Therefore, our static envelope theorem concludes that interior maxima only occur at differentiable points. In other words, at an investment level where the investor is indifferent between the two products, he strictly prefers to increase the investment and choose product d_1, or decrease the investment and choose product d_2. Amir, Mirman and Perkins (1991, Lemma 3.4) and Milgrom and Segal (2002, Corollary 2) provide special cases of this theorem under the assumptions of supermodularity and equidifferentiability, respectively.

Our second envelope theorem applies this intuition to dynamic settings. When an agent makes both discrete and continuous choices subject to some constraint, the value function has (potentially infinitely many) kinks. In Figure 1b, c represents effort, and the two curves represent the payoffs from attending college or not. As before, discrete choices may lead to downward kinks. In addition, binding constraints may lead to upward (or downward) kinks. Nevertheless, we show that at interior optimal choices, the value function is differentiable; the agent never chooses a savings level where he is indifferent between college or not.[2] We show this is true for a very general notion of interior choice, relaxing the traditional notion due to Benveniste and Scheinkman (1979, Assumption 4). We say a choice is a *one-period interior choice* if the agent is able to increase or decrease his continuous choice today without changing any other choices, e.g. the agent can increase or decrease his savings today without changing his college or future savings decisions.

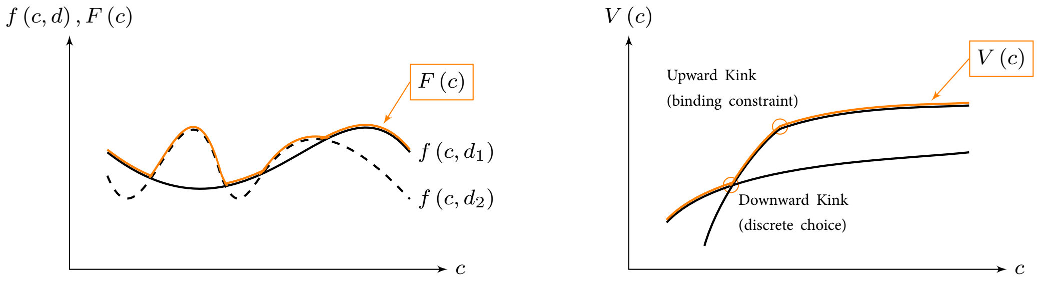
Previous envelope theorems in dynamic settings do not accommodate discrete choices, and impose additional assumptions. Mirman and Zilcha (1975, Lemma 1) and Benveniste

---

[1] Eckstein and Wolpin (1989), Rust (2008), or Aguirregabiria and Mira (2010) list many more examples.

[2] In complementary work, Rincón-Zapatero and Santos (2009) study the differentiability of value functions at boundary choices.

and Scheinkman (1979, Theorem 1) impose concavity assumptions, and Amir *et al.* (1991, Lemma 3.4) assume supermodularity.



(a) Static Problem (b) Constrained Dynamic Problem

Figure 1: Mixed continuous and discrete choice problems

The concept of directional derivatives is central to the proofs of previous envelope theorems. However, in our fully general setting, directional derivatives may not exist everywhere, so a new approach is required. We apply Fréchet sub- and superdifferentials, and their one-dimensional analogues which we call Dini sub- and superdifferentials.[3] They capture what we think of as upward and downward kinks.

This paper is organized as follows: Section 2 states our envelope theorems. All proofs go into Section 3 which contains additional general lemmata on kinks and upper envelopes. Section 4 illustrates the breadth of applications of our envelope theorems to non-smooth and non-concave dynamic programming problems. The proofs of the Banach space versions of our theorems are in the appendix. Nevertheless, we recommend reading them as they are much more elegant (but less intuitive) than the standard versions.

2 Theorems

An agent makes a *continuous choice* $c \in C$ and a *discrete choice* $d \in D$. Initially, we require the continuous choice set C to be a subset of $\mathbb{R}$; Appendix A generalizes all theorems to allow C to be a subset of any Banach space. We allow the discrete choice set D to be any non-empty set, e.g. a finite set such as $\{$full time work, part time work, not work$\}$, a continuous space such as $\mathbb{R}^2$, or an infinite dimensional space such as $C[0,1]$.

---

[3] The terminology "Dini sub- and superdifferential" does not appear to be widespread. On the other hand, "Fréchet sub- and superdifferential" is a standard generalization of the convex analysis notion of "subdifferential" to non-convex functions.

Definition 1. *We say F is the **upper envelope** of $\{f(\cdot, d)\}$ if $F(c) = \sup_{d \in D} f(c, d)$.*

Our static envelope theorem asserts that non-differentiable points are never optimal choices. An agent is never indifferent between two discrete choices after making an optimal continuous choice (unless the discrete choices are locally equivalent).

Theorem 1. *Suppose F is the upper envelope of a (possibly infinite) set of differentiable functions $\{f(\cdot, d)\}$. If $(\hat{c}, \hat{d})$ maximizes f, and $\hat{c} \in \operatorname{int}(C)$, then F is differentiable at $\hat{c}$ and satisfies the first-order condition $F'(\hat{c}) = f_c(\hat{c}, \hat{d}) = 0$.*

Our second result builds on Theorem 1 to study dynamic programming problems with continuous and discrete choices and states. In every period, the agent makes a continuous and discrete choice (c', d') based on the state variable (c, d) consisting of the previous period's choices. We denote the set of possible states by Ω. The agent may only make choices that satisfy the constraint

$$(c, c', d, d') \in \Gamma.$$

It will be convenient to write

$$\begin{aligned}
\Gamma(c, \cdot; d, \cdot) &= \{(c', d') : (c, c', d, d') \in \Gamma\} \\
\Gamma(c, \cdot; d, d') &= \{c' : (c, c', d, d') \in \Gamma\} \\
\Gamma(\cdot, c'; d, d') &= \{c : (c, c', d, d') \in \Gamma\}.
\end{aligned}$$

Let us assume that the agent has a feasible choice at every state, i.e. $\Gamma(c, \cdot, d, \cdot) \subseteq \Omega$ is non-empty for all $(c, d) \in \Omega$.

Problem 1. *Consider the following dynamic programming problem:*

$$V(c, d) = \sup_{(c', d') \in \Gamma(c, \cdot; d, \cdot)} u(c, c'; d, d') + \beta V(c', d'), \tag{1}$$

where the domain of V is Ω. We assume that $u(\cdot, c'; d, d')$ and $u(c, \cdot; d, d')$ are differentiable on $\operatorname{int}(\Gamma(\cdot, c'; d, d'))$ and $\operatorname{int}(\Gamma(c, \cdot; d, d'))$, respectively.[4]

There are two sources of non-differentiability in Problem 1. As before, the value function may have downward kinks where the agent is indifferent between two discrete choices. In addition, the value function may have upward kinks where the agent makes a boundary choice, but has zero marginal cost of changing to an interior choice nearby. As in Theorem 1, our approach is to focus on differentiability at optimal choices away from boundaries.

---

[4] Since we neither study nor require the existence of optimal policies or value functions, we do not impose conditions such as $\beta \in (0, 1)$. In particular, if the value function takes infinite values, then there are no maxima and the conditions for our theorems are violated.

The notion of interior choice is subtle in this context. For our purposes, a choice is an interior choice if the agent may increase or decrease his choice without changing any of his other (present and future) choices. This involves two conditions, which we illustrate with an agent who faces a savings choice in which consumption is uniquely determined by a budget constraint. First, the agent must be able to afford to save a bit more (perhaps by consuming a bit less today), and must be able to save a bit less so that no borrowing constraints bind. Second, if the agent decreases his savings a little bit today, then he must be able to maintain the same savings choice tomorrow (perhaps by consuming a bit less tomorrow). This leads to the following definition.

Definition 2. *The choice c' is a* ***one-period interior choice*** *with respect to (c, c'', d, d', d'') if*

(i) $c' \in \text{int}(\Gamma(c, \cdot; d, d'))$ *and*

(ii) $c' \in \text{int}(\Gamma(\cdot, c''; d', d''))$.

There is an important difference between Theorem 1 and Theorem 2. In the former, the optimal choice is the maximum of an upper envelope. In the latter, the optimal choice involves a trade-off between the cost of a choice today and its continuation value. We establish the value function is differentiable at optimal choices, rather than at the maximum of the value function (which typically does not exist).

Theorem 2. *Suppose $(\hat{c}', \hat{c}'', \hat{d}', \hat{d}'')$ are optimal choices following (c, d) in Problem 1. If $\hat{c}'$ is a one-period interior choice with respect to $(c, \hat{c}'', d, \hat{d}', \hat{d}'')$, then $V(\cdot, \hat{d}')$ is differentiable at $\hat{c}'$ and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}') = \beta\, V_c(\hat{c}', \hat{d}') = \beta\, u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}'').$$

The optimal one-period interior choice condition is weak in two respects. First, it only requires that the supremum is attained in these two periods. Second, it does not rule out any constraints binding on any future choices. In particular, the value function may have an upward kink at the future boundary choice $(\hat{c}'', \hat{d}'')$, but this does not matter. As long as $\hat{c}'$ satisfies the one-period interior choice condition, the value function is differentiable at $(\hat{c}', \hat{d}')$.

A natural extension of Problem 1 is the following version of a dynamic programming problem that incorporates stochastic shocks.

Problem 2. *Consider the following stochastic dynamic programming problem:*

$$V(c, d, \theta) = \sup_{(c', d') \in \Gamma(c, \cdot; d, \cdot; \theta)} u(c, c'; d, d'; \theta) + \beta \sum_{\theta' \in \Theta} \pi(\theta' | \theta)\; V(c', d', \theta'),$$

where the domain of V is $\Omega \times \Theta$. We assume that $u(\cdot, c'; d, d'; \theta)$ and $u(c, \cdot; d, d'; \theta)$ are differentiable on $\text{int}(\Gamma(\cdot, c'; d, d'; \theta))$ *and* $\text{int}(\Gamma(c, \cdot; d, d'; \theta))$*, respectively.*

The following theorem establishes, that the value function is differentiable at optimal choices. It is a stochastic version of Theorem 2.

Definition 3. *The choice c' is a **stochastic one-period interior choice** with respect to $(c, c''(\cdot), d, d', d''(\cdot))$ at θ if*

(i) $c' \in \text{int}(\Gamma(c, \cdot; d, d'; \theta))$ *and*

(ii) $c' \in \text{int}(\Gamma(\cdot, c''(\theta'); d', d''(\theta'); \theta'))$ *for all θ'.*

Theorem 3. *Suppose $(\hat{c}', \hat{d}')$ are optimal choices following (c, d, θ) in Problem 2, and $(\hat{c}''(\cdot), \hat{d}''(\cdot))$ are optimal policies for the following period's choices as a function of θ''. If $\hat{c}'$ is a one-period interior choice with respect to $(c, \hat{c}''(\cdot), d, \hat{d}', \hat{d}''(\cdot))$, then $V(\cdot, \hat{d}')$ is differentiable at $\hat{c}'$ and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}'; \theta) = \beta \sum_{\theta'} \pi(\theta'|\theta)\, V_c(\hat{c}', \hat{d}', \theta') = \beta \sum_{\theta'} \pi(\theta'|\theta)\, u_c(\hat{c}', \hat{c}''(\theta'); \hat{d}', \hat{d}''(\theta'); \theta').$$

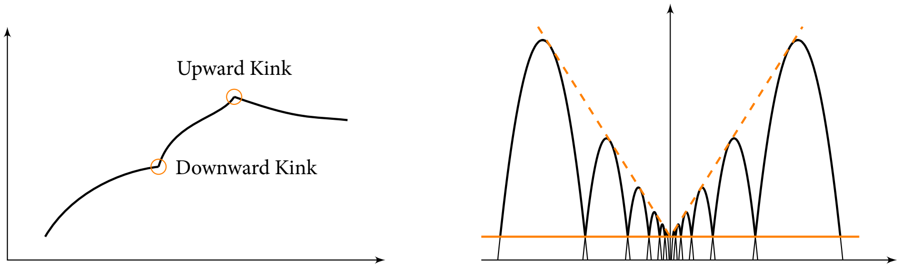
We omit the proof of this theorem, as it is a straightforward generalization of Theorem 2. The main difference is that there is a convex combination of value functions in the Bellman equation, rather than one single value function. This requires a simple generalization of Lemma 2 Part (iii) to finite sums. Generalizing to continuous random variables would require generalizing this lemma to integrals.

3 Proofs

3.1 Classification of Non-Differentiable Points

This section develops a classification of non-differentiable points of functions. We define upward and downward kinks in terms of (Dini) sub- and superderivatives. Then, we show that every non-differentiable point is either an upward or a downward kink and provide a lemma on important algebraic operations.

Intuitively, we would like to define an upward kink as a point where the slope approaching from the left is greater than the slope approaching from the right (see Figure 2a). Downward kinks would have the converse property.



(a) Upward and Downward Kinks (b) Bouncing Ball Function

Figure 2: Classifying non-differentiable points

However, we can not use directional derivatives because they may not exist. For instance, consider the bouncing ball function F depicted in Figure 2b, is the upper envelope of a countable set of parabolas $\{f(\cdot,d)\}_{d\in D}$, where

$$f\left(c,d\right)=-\frac{1}{|d|}\left(c-d\right)\left(c-\frac{d}{2}\right)\quad\text{and}\quad D=\left\{\frac{s}{2^n}:s\in\{-1,1\},n\in\mathbb{N}\right\}.$$

This function has directional derivatives everywhere except at $c=0$. In particular, the right directional derivative at $c=0$,

$$\lim_{\Delta c\to 0^+}\frac{F(\Delta c)-F(0)}{\Delta c}$$

does not exist because the slope oscillates between 0 and $(\sqrt{2}-1)^2$. We resolve this problem by taking limits inferior and superior of the slope, which always exist. According to our classification, $c=0$ is a downward kink but not an upward kink.[5]

Definition 4. *The (Dini)* ***sub- and superdifferentials*** *of f at $c\in\operatorname{int}(C)$ are*

$$\partial_D f\left(c\right)=\left\{m\in\mathbb{R}:\limsup_{\Delta c\to 0^-}\frac{f\left(c+\Delta c\right)-f\left(c\right)}{\Delta c}\le m\le\liminf_{\Delta c\to 0^+}\frac{f\left(c+\Delta c\right)-f\left(c\right)}{\Delta c}\right\}$$
$$\partial^D f\left(c\right)=\left\{m\in\mathbb{R}:\liminf_{\Delta c\to 0^-}\frac{f\left(c+\Delta c\right)-f\left(c\right)}{\Delta c}\ge m\ge\limsup_{\Delta c\to 0^+}\frac{f\left(c+\Delta c\right)-f\left(c\right)}{\Delta c}\right\}.$$

If $\partial_D f\left(c\right)$ is non-empty, then we say f is (Dini) subdifferentiable at c. Similarly, if $\partial^D f\left(c\right)$ is non-empty, then we say f is (Dini) superdifferentiable at c.

---

[5] In similar examples, there are points that are both upward and downward kinks. For example, the function $f(x)=x\sin\frac{1}{x}$ has an upward and downward kink at $x=0$.

Definition 5. *If f is not subdifferentiable at c, then we say it has an **upward kink** at c. Similarly, if f is not superdifferentiable at c, then we say it has a **downward kink** at c.*

The following lemma establishes that a non-differentiable point of a function can be classified as either an upward kink or a downward kink.

Lemma 1 (Differentiability). *A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at c if and only if f is both sub- and superdifferentiable at c. Moreover, if f is differentiable at c then $\{f'(c)\} = \partial_D f(c) = \partial^D f(c)$.*

Proof. The forward direction is straightforward. For the reverse direction, suppose that f is both sub- and superdifferentiable at c, so that $m_* \in \partial_D f(c)$ and $m^* \in \partial^D f(c)$. From the definitions,

$$\limsup_{\Delta c \to 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} \leq m_* \leq \liminf_{\Delta c \to 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c}$$
$$\liminf_{\Delta c \to 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} \geq m^* \geq \limsup_{\Delta c \to 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c}.$$

Since infima are weakly less than suprema, going clockwise, each expression is weakly less than the following one. Therefore, all of the expressions are equal. Thus, f is differentiable at c with $f'(c) = m^* = m_*$. ☐

The following Lemma provides some calculus properties of sub- and superdifferentials. Part (iii) provides a sufficient condition for the differentiability of a sum of functions, and plays an important role in the proof of Theorem 2.

Lemma 2 (Differential Calculus). *The following statements are true at any c (along with their superdifferentiable counterparts):*

(i) If g and h are subdifferentiable, then so is $g + h$.

(ii) g is subdifferentiable if and only if $-g$ is superdifferentiable.

(iii) If g and h are subdifferentiable and $g + h$ is superdifferentiable, then g, h, and $g + h$ are differentiable.

Proof. (i) This result follows from the subadditivity property of limits superior that allows us to write

$$\limsup_{c \to 0^-} [g(c) + h(c)] \leq \limsup_{c \to 0^-} \left[g(c) + \limsup_{c \to 0^-} h(c) \right] = \limsup_{c \to 0^-} g(c) + \limsup_{c \to 0^-} h(c),$$

and the analogous right limit inferior inequality.

(ii) Trivial.

(iii) From Part (i), $g + h$ is subdifferentiable, and hence differentiable by Lemma 1. From Part (ii), $-g$ is superdifferentiable, and Part (i) implies $h = (g + h) + (-g)$ is superdifferentiable. Therefore, Lemma 1 implies h is differentiable.

□

History: The notions of Dini (and Fréchet) sub- and superdifferentials generalize classical notions from convex analysis to non-convex functions. However, according to Kruger (2003), previous work in mathematics has not applied these concepts, because of "rather poor calculus" as $\partial_F(f+g)(x) \neq \partial_F f(x) + \partial_F g(x)$. Our approach appears to be novel: we simultaneously study sub- and superdifferentiability of functions to establish full differentiability.

The notions of Fréchet sub- and superdifferentials defined in Appendix A are standard, and appear in Schirotzek (2007, Chapter 9), although Fréchet superdifferentials only appear in a two-page section on Hamilton-Jacobi equations. The special case of Dini sub- and superdifferentials is non-standard. The pioneering papers that lead to these definitions are Clarke (1975, 1976), Penot (1974, 1978), and Bazaraa, Goode and Nashed (1974). Lemma A.1 – the Banach space version of Lemma 1 – appears without proof as Proposition 1.3 in Kruger (2003), but does not appear in Schirotzek. Parts (i) and (ii) of Lemma A.2 – the Banach space version of Lemma 2 – appear without proof in discussion on pages 172 and 181 of Schirotzek, respectively. We believe Part (iii) is novel.

3.2 Proof of Theorem 1

To establish Theorem 1, we prove that

(i) non-differentiable points are either upward or downward kinks or both (Lemma 1),

(ii) optimal choices may not occur at downward kinks (Lemma 3, Figure 3a), and

(iii) upper envelopes may not contain upward kinks (Lemma 4, Figure 3b).

Lemma 3. *If $\hat{c} \in \text{int}(C)$ is a maximum of $g : \mathbb{R} \to \mathbb{R}$, then g is superdifferentiable at $\hat{c}$ with $0 \in \partial^D g(\hat{c})$.*

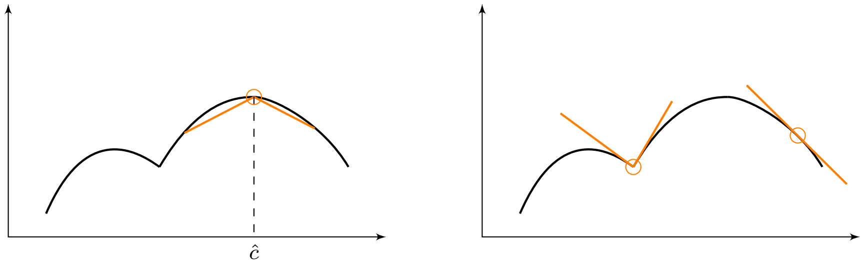
Proof. Since $\hat{c}$ is a maximum, the slope on the left is weakly positive, and the slope on the right is weakly negative. In other words, for any $\Delta c > 0$,

$$\frac{g\left(\hat{c} - \Delta c\right) - g\left(c\right)}{-\Delta c} \geq 0 \geq \frac{g\left(\hat{c} + \Delta c\right) - g\left(c\right)}{\Delta c}.$$

Taking limits gives

$$\liminf_{\Delta c \to 0^-} \frac{g\left(\hat{c} + \Delta c\right) - g\left(c\right)}{\Delta c} \geq 0 \geq \limsup_{\Delta c \to 0^+} \frac{g\left(\hat{c} + \Delta c\right) - g\left(c\right)}{\Delta c},$$

which establishes $0 \in \partial^D g(\hat{c})$. □



(a) Lemma 3 (b) Lemma 4

Figure 3: Illustration of Proof of Theorem 1

Lemma 4. *If F is the upper envelope of a (possibly infinite) set of differentiable functions $\{f\left(\cdot, d\right)\}$, and $c \in \text{int}(C)$, and $F(c) = f(c, \hat{d})$, then F is subdifferentiable at c with $f_c(c, \hat{d}) \in \partial_D F(c)$.*

Proof. Since $\hat{d}$ is the optimal choice at c,

$$f\left(c + \Delta c, \hat{d}\right) - f\left(c, \hat{d}\right) \leq F\left(c + \Delta c\right) - F\left(c\right).$$

Dividing by $\Delta c > 0$, and taking limits gives

$$f_c\left(c, \hat{d}\right) \leq \liminf_{\Delta c \to 0^+} \frac{F\left(c + \Delta c\right) - F\left(c\right)}{\Delta c}.$$

Similarly, dividing by $\Delta c < 0$ and taking limits gives

$$f_c\left(c, \hat{d}\right) \geq \limsup_{\Delta c \to 0^-} \frac{F\left(c + \Delta c\right) - F\left(c\right)}{\Delta c}.$$

Therefore, $f_c(c, \hat{d}) \in \partial_D F(c)$. □

We are ready now to prove Theorem 1 which is restated here.

Theorem 1. *Suppose* F *is the upper envelope of a (possibly infinite) set of differentiable functions* $\{f(\cdot,d)\}$. *If* $(\hat{c},\hat{d})$ *maximizes* f, *and* $\hat{c}\in\operatorname{int}(C)$, *then* F *is differentiable at* $\hat{c}$ *and satisfies the first-order condition* $F'(\hat{c})=f_c(\hat{c},\hat{d})=0$.

Proof. Lemmata 3 and 4 establish that F is super- and subdifferentiable at $\hat{c}$ with $0\in\partial^D F(\hat{c})$ and $f_c(\hat{c},\hat{d})\in\partial_D F(\hat{c})$. Applying Lemma 1, we conclude that F is differentiable at $\hat{c}$ with $F'(\hat{c})=0=f_c(\hat{c},\hat{d})$. ☐

History: We sketch the history of the proof steps (i)–(iii). Mirman and Zilcha (1975, Lemma 1) introduced (iii) in the context of a growth model. Instead of using (ii), they ensure that there are no downward kinks by assuming that the objective is concave. These results were generalized by Benveniste and Scheinkman (1979). Rockafellar (1970, Theorem 23.1) proves the existence of directional derivatives of concave functions, which he traces as far back as Stolz (1893, Satz 10, p. 35) who describes it as a standard result from geometry.

Amir *et al.* (1991, Lemmata 3.3 and 3.4) introduced the proof strategy of (i)–(iii), also in the context of a growth model. To ensure that directional derivatives exist in step (i), they impose a supermodularity assumption on the underlying function $f(\cdot,\cdot)$.

Milgrom and Segal (2002, Corollary 2) were the first to notice that this logic applies without any topological or monotonicity assumptions on the discrete choice set D. In step (i), they assumed that $\{f(\cdot,d)\}_{d\in D}$ is an equidifferentiable set of functions. Their Theorem 3 generalizes Clarke (1975, Theorem 2.1), which in turn generalizes Danskin (1966, Theorem 1).

3.3 Proof of Theorem 2

In this section, we prove Theorem 2 which is restated here:

Theorem 2. *Suppose* $(\hat{c}',\hat{c}'',\hat{d}',\hat{d}'')$ *are optimal choices following* (c,d) *in Problem 1. If* $\hat{c}'$ *is a one-period interior choice with respect to* $(c,\hat{c}'',d,\hat{d}',\hat{d}'')$, *then* $V(\cdot,\hat{d}')$ *is differentiable at* $\hat{c}'$ *and satisfies the first-order condition*

$$-u_{c'}(c,\hat{c}';d,\hat{d}')=\beta\,V_c(\hat{c}',\hat{d}')=\beta\,u_c(\hat{c}',\hat{c}'';\hat{d}',\hat{d}'').$$

The following lemma implies that $V(\cdot,\hat{d}')$ is subdifferentiable at the optimal choice $\hat{c}'$ when $\hat{c}'\in\operatorname{int}(\Gamma(\cdot,\hat{c}'';\hat{d}',\hat{d}''))$. In other words, upward kinks may only arise when a constraint binds on today's choice; upward kinks in the value function at future dates do not propagate backwards. Note that the lemma is written with different timing, and is applicable in a more general setting than the theorem.

Lemma 5. *Suppose* $(\hat{c}', \hat{d}')$ *are optimal choices given* (c, d) *in Problem 1. If* $c \in \text{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$*, then the value function* $V(\cdot, d)$ *is subdifferentiable at* c *with* $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$*.*

Proof. For all $c + \Delta c \in \Gamma(\cdot, \hat{c}'; d, \hat{d}')$, we have the inequality

$$\begin{aligned} & V(c+\Delta c, d) - V(c, d) \\ & \leq \left[u\left(c+\Delta c, \hat{c}'; d, \hat{d}\right) + \beta V\left(\hat{c}', \hat{d}'\right)\right] - \left[u\left(c, \hat{c}'; d, \hat{d}\right) + \beta V\left(\hat{c}', \hat{d}'\right)\right] \\ & = u\left(c+\Delta c, \hat{c}'; d, \hat{d}\right) - u\left(c, \hat{c}'; d, \hat{d}\right). \end{aligned}$$

Since $c \in \text{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$, this inequality holds for all Δc in an open neighborhood of 0. Dividing both sides by $\Delta c > 0$ and taking limits gives

$$\liminf_{\Delta c \to 0^+} \frac{V(c+\Delta c, d) - V(c, d)}{\Delta c} \leq u_c\left(c, \hat{c}'; d, \hat{d}'\right).$$

Similarly, dividing both sides by $\Delta c < 0$ and taking limits gives

$$\limsup_{\Delta c \to 0^-} \frac{V(c+\Delta c, d) - V(c, d)}{\Delta c} \geq u_c\left(c, \hat{c}'; d, \hat{d}'\right).$$

Therefore, $V(\cdot, d)$ is subdifferentiable at c with $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$. ☐

It remains to show that $V(\cdot, \hat{d}')$ is superdifferentiable at the optimal choice $\hat{c}'$. The Bellman equation in Problem 1 may be decomposed into the recursive equations

$$v(c'; c, d) = \sup_{d' \in \Gamma(c, c'; d, \cdot)} u(c, c'; d, d') + \beta V(c', d') \tag{2a}$$

$$\begin{aligned} V(c, d) = & \sup_{c' \in C} v(c'; c, d) \\ & \text{s.t. } c' \in \Gamma(c, \cdot; d, d') \text{ for some } d' \in D. \end{aligned} \tag{2b}$$

Our approach is to strip away the operations on the right side of (2a) until we arrive at V, showing that each expression is superdifferentiable with superderivative 0 at each step. Surprisingly, the subdifferentiability of $V(\cdot, \hat{d}')$ established above plays a key role.

Since Theorem 2 requires that the optimal choice $\hat{c}'$ lies in $\text{int}(\Gamma(c, \cdot; d, \hat{d}'))$, Lemma 3 implies that $v(\cdot; c, d)$ is superdifferentiable at $\hat{c}'$, and 0 is a superderivative. Therefore, both sides of (2a) are superdifferentiable in c' at $(c, \hat{c}', d)$.

The right side of (2a) can be written as

$$G(c') = \sup_{d' \in \Gamma(c, c'; d, \cdot)} g(c', d') \tag{3a}$$

$$g(c', d') = u(c, c'; d, d') + \beta V(c', d'). \tag{3b}$$

The following lemma establishes that $g(\cdot,\hat{d}')$ is also superdifferentiable at $\hat{c}'$ with a superderivative of 0.

Lemma 6. *Suppose F is the upper envelope of the set of functions $\{f(\cdot,d)\}$, and that $F(c^*) = f(c^*,d^*)$. If F is superdifferentiable at c^*, then $f(\cdot,d^*)$ is also superdifferentiable at c^* with $\partial^D f(c^*,d^*) \supseteq \partial^D F(c^*)$.*

Proof. Since F is superdifferentiable at c^*, there is some slope $m^* \in \partial^D F(c^*)$ with

$$\liminf_{\Delta c \to 0^-} \frac{F\left(c^* + \Delta c\right) - F\left(c^*\right)}{\Delta c} \geq m^* \geq \limsup_{\Delta c \to 0^+} \frac{F\left(c^* + \Delta c\right) - F\left(c^*\right)}{\Delta c}.$$

Since $F(c) \geq f(c,d^*)$, we know that

$$F\left(c^* + \Delta c\right) - F\left(c^*\right) \geq f\left(c^* + \Delta c, d^*\right) - f\left(c^*, d^*\right).$$

Dividing by $\Delta c > 0$ and taking limits, we find that

$$m^* \geq \limsup_{\Delta c \to 0^+} \frac{F\left(c^* + \Delta c\right) - F\left(c^*\right)}{\Delta c} \geq \limsup_{\Delta c \to 0^+} \frac{f\left(c^* + \Delta c, d^*\right) - f\left(c^*, d^*\right)}{\Delta c}.$$

Along with the analogous inequality on the left, this establishes $m^* \in \partial^D f(c^*,d^*)$. □

So far, we have established that $g(\cdot,\hat{d}')$ is superdifferentiable at $\hat{c}'$ and that each term in its sum is subdifferentiable. Therefore, Lemma 2 Part (iii) implies $g(\cdot,\hat{d}')$ and $V(\cdot,\hat{d}')$ are differentiable at $\hat{c}'$. We also established that 0 is a superderivative of $g(\cdot,\hat{d}')$ and $u_c(\hat{c}',\hat{c}'';\hat{d}',\hat{d}'')$ is a subderivative of $V(\cdot,\hat{d}')$, so these are in fact the derivatives. The equality of Theorem 2 follows, and this completes the proof.

History: Lemma 5 is a straightforward generalization of Lemma 4, whose history is discussed above. The early envelope theorems for dynamic programming problems (Mirman and Zilcha (1975, Lemma 1) and Benveniste and Scheinkman (1979)) imposed concavity assumptions to establish a form of superdifferentiability to complete the proof. In particular, Lemma 6 is reminiscent of Benveniste and Scheinkman (1979, Lemma 1). Amir *et al.* (1991, Lemma 3.4) do not have to address the possibility of cancellation of kinks, because the flow value is differentiable. Their proof has a similar structure to our Theorem 1.

4 Applications

To illustrate how broadly our theorems may be applied, we present two examples. The first is a classical dynamic programming problem with binary labor choice. The second is a capital adjustment problem with fixed costs. The general approach is to realize that there are many ways to decompose the agent's choices into continuous and discrete choices, and to select an appropriate decomposition to obtain first-order conditions.

Binary Labor Choice: Consider the following dynamic programming problem with consumption, savings, and a discrete labor choice:

$$\begin{aligned} W(a) = \max_{(c,a',\ell)\in\mathbb{R}^3} \; & u(c,\ell) + \beta\, W(a') \\ \text{s.t.} \quad & c \geq 0, a' \geq 0, \ell \in \{0,1\}, \\ & c + a' = R\,a + w\,\ell, \end{aligned} \tag{4}$$

where $u(\cdot,\ell)$ is differentiable and $\beta \in (0,1)$. To apply Theorem 2, we reformulate the problem as follows:

$$\begin{aligned} \tilde{W}(a,\ell) = \max_{(a',\ell')\in\Gamma(a,\cdot;\ell,\cdot)} \; & u(R\,a + w\,\ell - a', \ell) + \beta\,\tilde{W}(a',\ell'), \\ & \text{where } \Gamma = \{(a,a';\ell,\ell') \text{ s.t. } (a,\ell)\in\Omega,\ (a',\ell')\in\Omega,\ R\,a + w\,\ell - a' \geq 0\} \\ & \text{and } \Omega = [0,\infty)\times\{0,1\}. \end{aligned}$$

Suppose $\hat{\ell}$, $\hat{a}'$, $\hat{c}$ and $\hat{c}'$ are optimal choices given a.[6] The one-period interior choice condition of Theorem 2 is that (i) $\hat{c} > 0$ and $\hat{a}' > 0$, and (ii) $\hat{c}' > 0$. If this condition is met, then the theorem establishes that $\tilde{W}(\cdot,\hat{\ell}')$ is differentiable at $\hat{a}'$, and the first-order condition

$$u_c(\hat{c},\ell) = \beta\,\tilde{W}_{a'}\left(\hat{a}',\hat{\ell}'\right)$$

is satisfied. To summarize, W has two types of kinks: those arising at savings choices where the agent is indifferent between working or not, and those arising where constraints change from non-binding to binding. The first type of kink can not be an optimal choice, and is irrelevant. The second type of kink is only relevant when tomorrow's choice may become infeasible after a small change in today's choice. In particular, the first-order conditions are necessary even if $\hat{a}'' = 0$, because saving nothing tomorrow is feasible regardless of today's choices.

---

[6] c and $\hat{c}'$ are short hand for $\hat{c} = R\hat{a} + w\hat{\ell} - \hat{a}'$ and $\hat{c}' = R\hat{a}' + w\hat{\ell}' - \hat{a}''$.

Fixed Costs of Capital Adjustment: In many markets, there are fixed costs associated with adjusting capital stocks. For example, expanding office space involves searching for a new building, transporting furniture, and so on. The techniques explored here are also applicable to more general adjustment costs, as well as irreversible investment, and problems with bid-ask spreads.

Each period, the firm's production technology allows it to use k units of capital to produce $f(k)$ units of output, which trade at price p. Adjusting the capital stock k requires a fixed cost of c units of output. If the firm decides to pay this fixed cost, then it may buy or sell units of capital at a price of r. The firm discounts future profits at rate β, and has the following dynamic programming problem:

$$V(k) = \max \begin{cases} p\, f(k) + \beta\, V(k), \\ \max_{k' \geq 0} \quad p\,[f(k) - c] - r\,(k' - k) + \beta\, V(k'). \end{cases}$$

The value function V has downward kinks at points k at which the firm is indifferent between making an adjustment or not. We apply Theorem 2 to establish: if $\hat{k}' > 0$ and $\hat{k}'' > 0$ are optimal choices given k and $\hat{k}' \neq k$, then the value function V is differentiable at $\hat{k}'$ and satisfies the first-order condition

$$r = \beta\, V'(\hat{k}') = \beta\, p\, f'(\hat{k}').$$

To apply Theorem 2, we need to reformulate the problem for two reasons. First, the theorem requires full differentiability with respect to continuous choices. So, we express the adjustment cost as a discrete choice $\alpha \in \{0, 1\}$ of whether to make an adjustment, and impose a constraint that $k = k'$ when $\alpha = 0$. Second, this new constraint of $k = k'$ must be formulated to accommodate the one-period interior choice condition (Definition 2). This condition requires that when changing k' by a small amount, all other choices remain feasible. However, if the agent changes k' but does not make an adjustment tomorrow, then k'' must also change, violating the condition. We resolve this problem by reformulating the decision as an investment choice $i = k' - k$ rather than a capital stock choice k'. (Formally speaking, the capital stock becomes a discrete "choice" that is entirely constrained by the other choices.) We use a subscripted '$-$' to indicate choices made yesterday.

$$\begin{aligned} W(i_-; k_-) &= \max_{(i;k,\alpha) \in \Gamma(i_-,\cdot;k_-,\cdot,\cdot)} u(i_-, i; k_-, k, \alpha) + \beta\, W(i; k) \\ \text{where} \quad & u(i_-, i; k_-, k; \alpha) = p\,[f(k_- + i_-) - \alpha c] - ri, \\ & \Gamma = \{(i_-, 0; k_-, k_- + i_-, 0) : k_- \geq 0, k_- + i_- \geq 0\} \\ & \quad \cup \{(i_-, i; k_-, k_- + i_-; 1) : k_- \geq 0, k_- + i_- \geq 0, k_- + i_- + i \geq 0\}. \end{aligned}$$

Note that the reformulation is related by the equation $W(i_-; k_-) = V(i_- + k_-)$. The one-period interior choice condition in the reformulated problem is that $k_- + i_- + i > 0$ and $k_- + i_-i + i' > 0$ (i.e. $k' > 0$ and $k'' > 0$). Theorem 2 implies that if the firm chooses investment $\hat{i} \neq 0$ along with $(\hat{i}'; \hat{k}, \hat{k}', \hat{\alpha}')$ satisfying the one-period interior choice condition, then this choice satisfies the first-order condition,

$$-u_i(i_-, \hat{i}; k_-, \hat{k}; 1) = \beta\, W_i(\hat{i}; \hat{k}) = \beta\, u_{i_-}(\hat{i}, \hat{i}'; \hat{k}, \hat{k}'; \hat{\alpha}').$$

Theorem 3 would similarly apply if the firm faced productivity shocks.

Numerical Analysis: Our results may be useful for numerical analysis. Fella (2011) applies our theorems in his generalization of the endogenous grid method of Carroll (2006). He finds his method is substantially faster and more accurate than discretization methods.

A Banach Space Version

For many dynamic discrete choice problems in economics, it is natural to allow for more than one continuous choice (we already allowed for arbitrary "discrete" choice spaces above). We generalize our results to multidimensional spaces. Let $(X, \|\cdot\|)$ be a Banach space (for example, X could be $\mathbb{R}^n$). We denote

$$X^* = \{\phi : X \to \mathbb{R} \text{ such that } \phi \text{ is linear and continuous}\}$$

as its topological dual space. The standard notion of differentiability in Banach spaces is due to Fréchet.

Definition A.2. *A function $f : X \to \mathbb{R}$ is* ***Fréchet differentiable*** *at x if there is some $\phi^* \in X^*$ such that*

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} = 0.$$

ϕ^* is called the Fréchet derivative of f at x, and may be written as $f'(x)$ or $f_x(x)$.

Theorem A.1. *Suppose F is the upper envelope of a (possibly infinite) set of Fréchet differentiable functions $\{f(\cdot, d)\}$. If $(\hat{c}, \hat{d})$ maximizes f, then F is Fréchet differentiable at $\hat{c}$ with $F'(\hat{c}) = f_c(\hat{c}, \hat{d}) = 0$.*

The statement of the generalization of Theorem 2 is identical to the original, apart from the use of Fréchet derivatives.

Theorem A.2. *Suppose* $(\hat{c}', \hat{c}'', \hat{d}', \hat{d}'')$ *are optimal choices following* (c, d) *in Problem 1 (in which the utility functions are Fréchet differentiable in the analogous way). If* $\hat{c}'$ *is a one-period interior choice with respect to* $(c, \hat{c}'', d, \hat{d}', \hat{d}'')$*, then* $V(\cdot, \hat{d}')$ *is Fréchet differentiable at* $\hat{c}'$ *and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}') = \beta V_c(\hat{c}', \hat{d}') = \beta u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}''). \tag{5}$$

Notice that the following proofs are shorter than the standard proofs (because we do not have to deal with left and right limits), however this comes at the cost of a loss in economic intuition. We keep the order and numbering similar to the proofs in Section 3 but we omit all the surrounding text, discussing these results.

Definition A.3. *The* ***Fréchet subdifferential*** *of* $f : X \to \mathbb{R}$ *is*

$$\partial_F f(x) = \left\{ \phi^* \in X^* : \liminf_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \geq 0 \right\},$$

and f *is Fréchet subdifferentiable if* $\partial_F f(x)$ *is non-empty. Similarly, the* ***Fréchet super-differential*** *of* f *is*

$$\partial^F f(x) = \left\{ \phi^* \in X^* : \limsup_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \leq 0 \right\},$$

and f *is Fréchet superdifferentiable if* $\partial^F f(x)$ *is non-empty.*

For completeness, we prove the following standard result which generalizes Lemma 1.

Lemma A.1. *A function* $f : X \to \mathbb{R}$ *is Fréchet differentiable if and only if it is both Fréchet sub- and superdifferentiable.*

Proof. It is straightforward to show that differentiable functions are sub- and superdifferentiable. Conversely, suppose f is both Fréchet sub- and superdifferentiable, so that $\phi_* \in \partial_F f(x)$ and $\phi^* \in \partial^F f(x)$. Then from the definitions,

$$\liminf_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x) - \phi_* \Delta x}{\|\Delta x\|} \geq 0,$$
$$\limsup_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \leq 0.$$

The second inequality may be rewritten as

$$\liminf_{\Delta x \to 0} -\frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \geq 0.$$

From the superadditivity of limits inferior, we deduce

$$\liminf_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x) - \phi_* \Delta x}{\|\Delta x\|} - \frac{f(x+\Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \geq 0$$

$$\liminf_{\Delta x \to 0} [\phi_* - \phi^*] \frac{\Delta x}{\|\Delta x\|} \geq 0.$$

But this final equality is only satisfied when $\phi_* = \phi^*$. Therefore, the Fréchet sub- and superdifferentials coincide on a singleton, which must be the Fréchet derivative. □

Lemma A.2. *The following statements are true at any c:*

(i) If g and h are Fréchet subdifferentiable, then so is $g + h$.

(ii) g is Fréchet subdifferentiable if and only if $-g$ is Fréchet superdifferentiable.

(iii) If g and h are Fréchet subdifferentiable and $g + h$ is Fréchet superdifferentiable, then g, h, and $g + h$ are Fréchet differentiable.

Lemma A.3. *If $\hat{c} \in \text{int}(C)$ is a maximum of $f : C \to \mathbb{R}$, then f is superdifferentiable at $\hat{c}$ with $0 \in \partial^F f(\hat{c})$.*

Proof. Since $\hat{c}$ is a maximum, $f(\hat{c}+\Delta c) - f(\hat{c}) \leq 0$ for sufficiently small $\Delta c \in X$. Dividing by $\|\Delta c\|$ and taking limits gives

$$\limsup_{\Delta c \to 0} \frac{f(\hat{c} + \Delta c) - f(\hat{c})}{\|\Delta c\|} \leq 0.$$

Therefore $0 \in \partial^F f(\hat{c})$. □

Lemma A.4. *If F is the upper envelope of the set of Fréchet differentiable functions $\{f(\cdot, d)\}$, and $c^* \in \text{int}(C)$, and $F(c^*) = f(c^*, d^*)$, then F is subdifferentiable with $f_c(c^*, d^*) \in \partial_F F(c^*)$.*

Proof. If $F(c^*) = f(c^*, d^*)$, then we have

$$f(c^* + \Delta c, d^*) - f(c^*, d^*) \leq F(c^* + \Delta c) - F(c^*).$$

Subtracting $\phi \Delta c$, dividing by $\|\Delta c\|$, and taking limits on both sides gives

$$\liminf_{\Delta c \to 0} \frac{f(c^* + \Delta c, d^*) - f(c^*) - \phi\, \Delta c}{\|\Delta c\|} \leq \liminf_{\Delta c \to 0} \frac{F(c^* + \Delta c) - F(c^*) - \phi\, \Delta c}{\|\Delta c\|}.$$

After setting $\phi = f_c(c^*, d^*)$, the left side is zero. Therefore, the right side is non-negative, so $f_c(c^*, d^*) \in \partial_F F(c^*)$. □

Lemma A.5. *Suppose* $(\hat{c}', \hat{d}')$ *are optimal choices given* (c, d) *in Problem 1. If* $c \in \operatorname{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$, *then the value function* $V(\cdot, d)$ *is subdifferentiable at* c *with* $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$.

Proof. This proof is omitted; it is straightforward to adapt the proof of Lemma 5 using the technique in the proof of Lemma 4. □

Lemma A.6. *Suppose* F *is the upper envelope of the set of functions* $\{f(\cdot, d)\}$, *and that* $F(c^*) = f(c^*, d^*)$. *If* F *is Fréchet superdifferentiable at* c^*, *then* $f(\cdot, d^*)$ *is also Fréchet superdifferentiable at* c^*.

Proof. Since F is Fréchet superdifferentiable at c^*, there is some $\phi^* \in \partial^F F(c^*)$ with

$$\limsup_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x) - \phi^* \Delta x}{\|\Delta x\|} \leq 0.$$

Since $F(c) \geq f(c, d^*)$, we know that

$$F(c^* + \Delta c) - F(c^*) \geq f(c^* + \Delta c, d^*) - f(c^*, d^*).$$

Subtracting $\phi^* \Delta c$, dividing by $\|\Delta c\|$ and taking limits on both sides yields

$$\limsup_{\Delta x \to 0} \frac{F(c^* + \Delta c) - F(c^*) - \phi^* \Delta c}{\|\Delta c\|} \geq \limsup_{\Delta x \to 0} \frac{f(c^* + \Delta c, d^*) - f(c^*, d^*) - \phi^* \Delta c}{\|\Delta c\|}.$$

From the first inequality, the left side is less than 0, which establishes that $\phi^* \in \partial^F f(c^*, d^*)$. □

References

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, **156** (1), 38 – 67, structural Models of Optimization Behavior in Labor, Aging, and Health.

Amir, R., Mirman, L. J. and Perkins, W. R. (1991). One-sector nonclassical optimal growth: Optimality conditions and comparative dynamics. *International Economic Review*, **32** (3), 625–644.

Bazaraa, M. S., Goode, J. J. and Nashed, M. Z. (1974). On the cones of tangents with applications to mathematical programming. *Journal of Optimization Theory and Applications*, **13**, 389–426, 10.1007/BF00934938.

Benveniste, L. M. and Scheinkman, J. A. (1979). On the differentiability of the value function in dynamic models of economics. *Econometrica*, **47** (3), 727–732.

Carroll, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization proble. *Economics Letters*, **91** (3), 312–320.

Clarke, F. H. (1975). Generalized gradients and applications. *Transactions of The American Mathematical Society*, **205**, 247–262.

— (1976). A new approach to lagrange multipliers. *Mathematics of Operations Research*, **1** (2), 165–174.

Danskin, J. M. (1966). The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, **14** (4), 641–664.

Eckstein, Z. and Wolpin, K. I. (1989). The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, **24** (4), 562–598.

Fella, G. (2011). *A Generalized Endogenous Grid Method for Non-concave Problems*. Working Papers 677, Queen Mary, University of London, School of Economics and Finance.

Kruger, A. Y. (2003). On Fréchet subdifferentials. *Journal of Mathematical Sciences*, **116**, 3325–3358, 10.1023/A:1023673105317.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, **70** (2), 583–601.

Mirman, L. J. and Zilcha, I. (1975). On optimal growth under uncertainty. *Journal of Economic Theory*, **11** (3), 329–339.

Penot, J.-P. (1974). Sous-différentiels de fonctions numériques non convexes. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Série A*, **278**, 1553–1555.

— (1978). Calcul sous-différentiel et optimisation. *Journal of Functional Analysis*, **27** (2), 248–276.

Rincón-Zapatero, J. P. and Santos, M. S. (2009). Differentiability of the value function without interiority assumptions. *Journal of Economic Theory*, **144** (5), 1948–1964.

Rockafellar, R. T. (1970). *Convex Analysis*. Prince.

Rust, J. (2008). Dynamic programming. In S. N. Durlauf and L. E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Basingstoke: Palgrave Macmillan.

Schirotzek, W. (2007). *Nonsmooth Analysis*. Springer.

Stolz, O. (1893). *Grundzüge der Differential- und Integralrechnung*. Leipzig: B. G. Teubner.