Generalized Envelope Theorems*

Olivier Morand† Kevin Reffett‡ Suchismita Tarafdar§

First draft: August, 2012
This draft: September 2013

Abstract

We develop new envelope theorems for a broad class of parameterized nonsmooth optimization problems typical of economic applications where nonconvexities play a key role. We provide sufficient conditions for the value function of a nonconvex and nonsmooth Lipschitz program to be locally Lipschitz. We obtain bounds for upper and lower Dini derivatives of this value function, as well as sufficient conditions for the directional differentiability and/or differentiability of the value function.

Several applications of these results are presented, illustrating how the use of generalized envelopes is a natural extension of standard envelope theorems to nonsmooth and nonconvex cases.

1 INTRODUCTION

The use of envelope theorems to characterize optimal solutions of constrained optimization problems is a powerful tool of microeconomic and macroeconomic theory.

A classical envelope (theorem) for a parameterized optimization problem is basically an equality between the derivative of the value function along a direction and the derivative of the objective evaluated at the optimum along that same direction, ignoring "indirect" changes in the objective due to changes in the optimum. In constrained programs one often appeals to a duality argument to read the envelope of a primal program using an appropriate dual program, and in general smooth convex program, this envelope is a standard $C^1$ derivative (see, for instance, Rincon-Zapatero and Santos [33]). Unfortunately, in dynamic programming with discrete (or integer) choices, in problems in mechanism design with Lipschitz primitive data, in constrained lattice programming problems, or in "bi-level" programming problems (Stackelberg games) as well as in many other economic problems, this result does not apply.

*We thank Rabah Amir, Hector Chade, Cuong LeVan, Len Mirman, Ed Prescott, Juan-Pablo Rincon-Zapatero, Manuel Santos, and Carlo Strub for helpful conversations. The usual caveats apply.
†Department of Economics, University of Connecticut
‡Department of Economics, WP Carey School of Business, Arizona State University
§Department of Economics, Shiv Nadar University
Evidently, in nonconvex non-smooth but Lipschitz programs (constrained or unconstrained), we cannot expect the envelope to be a simple derivative as many technical difficulties arise simultaneously. Some constraints may occasionally be active, the Lagrange multiplier may not be unique, and obtaining directional derivatives can be a problem since traditional derivatives (or gradients) become generalized gradients (which are not necessarily singleton) when working with Lipschitz functions. Even Lagrange multiplier rules or Karash-Kuhn-Tucker conditions cannot generally be stated in a simple form.

The goal of this paper is to produce generalized envelope theorems for a wide class of "Lipschitz programs". For these programs, in which concavity/convexity but also standard $C^1$ differentiability assumptions are relaxed, we give sufficient conditions for generalized envelope theorems to exist, and show that classical $C^1$ envelopes are a special case of our more general findings. At the core of our analysis is a fundamental result showing the existence of bounds for the Dini derivatives of the value function of a Lipschitz program. This key result also help establish that the value function is locally Lipschitz, providing a non-smooth constraint qualification is met. Our results also help to characterize and develop methods for computing solutions of dynamic programming problems with discrete/integer decisions, binding constraints, and noninterior solutions, as well as establishing Lipschitz bounds for the value function of a dynamic program.\footnote{And thus help construct error bounds for dynamic programs that are only (locally) Lipschitz via discretization methods.}

It is important to bear in mind that even something as simple as the continuity of the value function can be difficult to establish in Lipschitz programs with general constraint systems. Berge’s theorem of the maximum has, since it requires the constraint correspondence $s \mapsto D(s)$ to be continuous for Berge’s theorem of the maximum to have a chance to deliver the desired continuity of the value function. This very often fails in fairly simple models, but we present an alternative to Berge’s theorem of the maximum that uses a local condition to guarantee the continuity of the value function.

The paper is laid out as follows. In the next section, we describe our benchmark class of Lipschitz programs and the type of nonsmooth constraint qualification used throughout the paper. In Section three we construct Dini bounds for the value function and give various envelope theorems. In Section four we prove that the value function is Lipschitz and give its Clarke gradient. The final Section presents a variety of applications, including examples in growth theory, industrial organization, macroeconomics, and constrained lattice programming.

## 2 LIPSCHITZ PROGRAMS AND CONSTRAINT QUALIFICATIONS

We consider Lipschitz programs of the form:

$$\max_{a \in D(s)} f(a, s)$$

(1)

where the choice set $A$ and the state space $S$ are open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, $f : A \times S \to \mathbb{R}$ is the objective function, and $D : S \rightrightarrows A$ is the feasible correspondence.
defined as:

\[ D(s) = \{ a | g_i(a, s) \leq 0, \quad i = 1, \ldots, p \text{ and } h_j(a, s) = 0, \quad j = 1, \ldots, q \} \]

where \( g_i : A \times S \to \mathbb{R}, \ i = 1, \ldots, p \) and \( h_j : A \times S \to \mathbb{R}, \ j = 1, \ldots, n \). Objectives and constraints are only locally Lipschitz in \((a, s)\), in contrast to "smooth" programs with \(C^1\) (or, once continuously differentiable) objective and constraints.

The function \( V : S \to \mathbb{R}, \ V(s) = \max_{a \in D(s)} f(a, s) \) is called the value function, and the optimal solution correspondence is \( A^* : S \Rightarrow A \) defined as:

\[ A^*(s) = \arg \max_{a \in D(s)} f(a, s) \]

The classical Lagrangian\(^2\) associated with this program is:

\[ L(a, s, \lambda, \mu) = f(a, s) - \lambda g(a, s) - \mu h(a, s) \]

where \( \lambda \) and \( \mu \) are row vectors in \( \mathbb{R}^p \) and \( \mathbb{R}^q \) respectively.

### 2.1 Constraint Qualifications

Given \( s \in S \), a point \( a \in D(s) \) is a Karush-Kuhn-Tucker (KKT) point of Program (1) if there exists a vector \((\lambda, \mu)\) of multipliers with \( \lambda \geq 0 \) such that:

\[ 0 \in \partial_a(f - \sum_{i=1}^{p} \lambda_i g_i(a, s) - \sum_{j=1}^{q} \mu_j h_j(a, s)) \]

and \( \lambda_i g_i(a, s) = 0 \) for all \( i = 1, \ldots, p \). The necessarily closed and convex (but perhaps empty) set of vectors \((\lambda, \mu)\) satisfying the above "multiplier rule" at \((a, s)\) will be denoted \( K(a, s) \) (or simply \( K \)).

To guarantee that \( K \) is non-empty and bounded (thus compact) requires additional assumptions in the form of constraint qualifications. In smooth programs the most important CQ is the Mangasarian-Fromovitz Constraint Qualification.

**Definition 1** The Mangasarian-Fromovitz Constraint Qualification (MFCQ) is satisfied at \( a^*(s) \in A^*(s) \) if there exists \( y \in \mathbb{R}^n \) such that:

\[ \nabla_a g_i(a^*(s), s) \cdot y < 0, \quad i \in I(a^*(s), s), \]

\[ \nabla_a h_j(a^*(s), s) \cdot y = 0, \quad j = 1, \ldots, q \]

where \( I(a^*(s), s) \) is the set identifying the active inequality constraints (those for which \( g_i(a^*(s), s) = 0 \)), and the matrix \( \nabla_a h(a^*(s), s) \) has full rank.

In smooth programs, Gauvin [14] (Lemma 1) proves that MFCQ at \( a^*(s) \) is equivalent to the compactness of \( K(a^*(s), s) \). Kyparisis [24] sharpens this result under the following slightly less general condition which simply treats active inequality constraints for which the multiplier is strictly positive as equality constraints.

\(^2\)If the choice set \( A \) is closed, then the abstract constraint \( a \in A \) induces an additional term in the Lagrangian (see, for instance, Clarke [8], Chapter 6).
**Definition 2** The Strict Mangasarian-Fromovitz Constraint Qualification (SMFCQ) is satisfied at $a^*(s) \in A^*(s)$ if there exists $y \in \mathbb{R}^n$ such that:

$$\nabla_a g_i(a^*(s), s) \cdot y < 0, \quad i \in I_s(a^*(s), s)$$

$$\nabla_a g_i(a^*(s), s) \cdot y = 0, \quad i \in I_b(a^*(s), s)$$

$$\nabla_a h_j(a^*(s), s) \cdot y = 0, \quad j = 1, \ldots, q$$

where $I_s(a^*(s), s) = \{i \in I(a^*(s), s), \lambda_i = 0\}$, and $I_b(a^*(s), s) = \{i \in I, \lambda_i > 0\}$, and the vectors $\nabla_a g_i(a^*(s), s), i \in I_b(a^*(s), s), \nabla_a h_j(a^*(s), s), j = 1, \ldots, q$ are linearly independent.

In smooth programs, Kyparsis [24] (Proposition 1.1) (see also Bonnans and Shapiro [6], Remark 4.49, page 298) shows that SMFCQ is both necessary and sufficient for $K$ to be a singleton.

No classical gradients generically exist in Lipschitz programs, so we rely on a generalization of the MFCQ (referred to as "Generalized MFCQ", or GMFCQ) introduced by Hiriart-Urruty [20] and stated in terms of Clarke’s generalized gradients.³ Denoting by $\overline{g}(a^*(s), s)$ the vector of binding inequality constraints at $a^*(s)$, so that $\overline{g} : A \times S \rightarrow \mathbb{R}^\overline{p}$ (where $\overline{p} = \text{Card}(I(a^*(s), s)))$, GMFCQ can be stated as follows:

**Definition 3** The Generalized Mangasarian-Fromovitz Constraint Qualification (GMFCQ) is satisfied at $a^*(s) \in A^*(s)$ if there exists $y \in \mathbb{R}^n$ such that:

$$\forall (\gamma_a, v_a) \in \partial_a(\overline{g}, h)(a^*(s), s), \quad \gamma_a \cdot y < 0, \quad \text{and} \quad v_a \cdot y = 0$$

and $\partial_a h(a^*(s), s)$ is of maximal rank.

Hiriart-Urruty [20] (Theorem 4.2) proves that GMFCQ implies the non-emptiness of the set of KKT multipliers (and thus the absence of "abnormal" multipliers, in Clarke’s terminology).

We note that:

$$\partial_a \overline{g}(a^*(s), s) \subset \prod_{i \in I(a^*(s), s)} \partial_a g_i(a^*(s), s)$$

so this version of GMFCQ is slightly more general than that of Auslender [4] (Theorem 2.1).

### 2.2 Additional Assumptions

The local Lipschitzness of all the functions in Program (1) is of course not sufficient for the value function $V$ to be locally Lipschitz (or even continuous). Clearly, additional assumptions are needed to establish any kind of differentiability property of $V$ at a point $s$. We adopt the general hypothesis made in Clarke [8] (Hypothesis 6.5.1, page 241), and show that it is satisfied in many important cases.

³See Appendix for the Math
Theorem 6 is not continuous at \( f \). The correspondence \( D \) of \( V \) and the optimal correspondence is a powerful result analogous to Berge’s maximum theorem: The value function is continuous, necessary \( A^*(s') \cap \Lambda \neq \emptyset \).

Although Clarke’s Hypothesis is not expressed in terms of primitive data of the problem, we present three basic conditions under which it is satisfied. The first is a simple condition on the objective and is mentioned by Clarke. The second is the inf-compactness condition used in Bonnans and Shapiro [6] to derive stability results for differentiable programs. The third is the uniform compactness condition used for instance in Gauvin and Dubeau [15].

Proposition 5 Clarke’s Hypothesis is satisfied if: (a) for every \( r \in \mathbb{R} \), the set \( \{(a', s') \in A \times S, f(a', s') \geq r \} \) is compact, or (b) there exists \( r \in \mathbb{R} \) and a compact set \( \Omega \subset A \) such that for every \( s' \) in a neighborhood of \( s \) the set \( \{a' \in D(s'), f(a', s') \geq r \} \) is a nonempty set contained in \( \Omega \), or (c) there exists a neighborhood \( S' \) of \( s \) such that \( cl[\cup_{s' \in S'} D(s)] \) is compact.

Proof. (a) See Clarke [8] page 241. (b). By inf-compactness at \( s \), there exists \( r \) such that if \( s' \in \varepsilon_0 B(s) \) then \( \{a' \in D(s'), f(a', s') \geq r \} \) is nonempty and contained in \( \Omega \). Thus, letting \( \Lambda = \Omega, \forall s' \in \varepsilon_0 B(s), A^*(s') \subset \Omega \) and \( A^*(s') \) is nonempty, and Clarke’s Hypothesis is satisfied. (c). Uniform compactness of \( D \) near \( s \) implies the existence of a neighborhood \( S' \) of \( s \) such that \( cl[\cup_{s' \in S'} D(s)] \) is compact. As a result:

\[
\exists \varepsilon_0 > 0, s' \in \varepsilon_0 B(s) \implies D(s') \text{ is compact}
\]

since \( D(s') \) is a closed subset of the compact \( cl[\cup_{s' \in S'} D(s)] \). This implies that \( A^*(s') \) is nonempty (the objective \( f \) is continuous on the compact \( D(s') \)), and Clarke’s Hypothesis is satisfied by letting \( \Lambda = cl[\cup_{s' \in S'} D(s)] \).

When combined with GMFCQ, any of the conditions of Proposition 5 implies a very powerful result analogous to Berge’s maximum theorem: The value function is continuous, and the optimal correspondence \( A^* \) is upper hemicontinuous. This upper hemicontinuity is a very important property, since it implies that as \( s_n \) converges to \( s \), the maxima of \( f(.,s_n) \) get arbitrarily close to some of the maxima of \( f(.,s) \). Note also that the continuity of \( V \) cannot come from a direct application of Berge’s maximum theorem since the feasible correspondence \( D \) is not necessarily continuous, even though all constraints are continuous.

The correspondence \( D \) defined as:

\[
D(s) = \{(x, y), x + y \leq s \text{ and } (s - 11)(10 - x) \leq 0\}
\]

is not continuous at \( s = 11 \).

Theorem 6 If GMFCQ holds at any \( a^*(s) \in A^*(s) \) and any of the conditions of Proposition 5, \( V \) is Lipschitz near \( s \) and \( A^* \) is upper hemicontinuous at \( s \). Moreover:

\[
\partial V(s) \subset co \left\{ \bigcup_{a^*(s) \in A^*(s)} \bigcup_{(\lambda, \mu) \in K^*(a^*_n(s), s)} \partial_s (f - \lambda g - \mu h)(a^*(s), s) \right\}
\]
Proof. The Lipschitz property of $V$ and the formula for the generalized gradient follow directly from Clarke [8] Corollary 1 and rely on the combination of Clarke’s Hypothesis and GMFCQ. The upper semicontinuity of $A^*$ is established separately for the three cases.

(a) Given any $s_n \to s$ and any sequence $\{a_n\}$ such that $a_n \in A^*(s_n) \subset D(s_n)$, necessarily $V(s_n) = f(a_n, s_n) \to V(s)$. Given $\varepsilon' > 0$ there exists $N$ such that $\forall n \geq N$, $f(a_n, s_n) \geq V(s) - \varepsilon'$. By the growth condition the sequence $\{a_n, s_n\}_{n \geq N}$ thus belongs to a compact set, and therefore has a convergent subsequence to $(a, s)$. By continuity of $f$, $f(a_n, s_n) \to f(a, s)$, hence $f(a, s) = V(s)$, and by closeness of $D$ at $a$, $a \in A^*(s)$, which establishes the upper hemicontinuity of $A^*$ at $s$.

(b) Under the inf-compactness condition, there exists $r$ such that the set $A_s = \{a' \in D(s') : f(a', s') \geq r\}$ is nonempty for all $s' \in \delta B(s)$ and is included in a compact set $\Omega$. Thus there exists $N$ such that $\forall n \geq N$, $a_n \in A^*(s_n) \subset A_s \subset \Omega$ and the sequence $\{(a_n, s_n)\}_{n \geq N}$ therefore has a convergent subsequence to $(a, s)$. As in (a), the continuity of $f$ and closeness of $D$ imply the desired result.

(c) Since $V$ is continuous at $s$, the map $L : s \to \{a, f(a, s) - V(s) \geq 0\}$ is closed at $s$. Under uniform compactness, since $A^*(s) = L(s) \cap D(s)$, the correspondence $A^* : s \to A^*(s)$ is the intersection of the closed mapping $L$ with the upper hemicontinuous (since closed and uniformly compact) mapping $D$. Consider then $s_n \to s$ and any $a_n \in A^*(s_n) = L(s_n) \cap D(s_n)$. Since $D$ is upper hemicontinuous at $s$, there exists a subsequence of $a_n$ converging to some $a \in D(s)$. Since $L$ is closed at $s$, the limit $a$ of the subsequence of $a_n$ necessarily belong to $L(s)$. Thus, $a \in A^*(s) = L(s) \cap D(s)$, which proves that $A^*$ is upper hemicontinuous at $s$. ■

3 STABILITY BOUNDS FOR THE VALUE FUNCTION

In this section we present sufficient conditions for the existence of an upper (resp. lower) bound for the upper (resp. lower) Dini derivatives of the value function. This important result is the basis for our generalized envelope theorems.

3.1 A Central Result

We use an important result on the existence of bounds for the Dini derivatives of the value function for general Lipschitz programs obtained as a consequence of Clarke [8] Corollary 4 (page 243) (see also Tarafdar [37] for alternative proof independent of Clarke’s results).

Theorem 7 If GMFCQ holds at any $a^*(s) \in A^*(s)$ and under Clarke’s Hypothesis, for any $x \in \mathbb{R}^m$:

$$D^+ V(s; x) \leq \max_{a^*(s) \in A^*(s)} \left( \sup_{\lambda \in K(a^*(s), s)} \left( \max_{\theta \in \partial_h (f - \lambda g - \mu h)(a^*(s), s)} \theta \cdot x \right) \right)$$

and:

$$\max_{a^*(s) \in A^*(s)} \inf_{\lambda \in K(a^*(s), s)} \left( \min_{\theta \in \partial_h (f - \lambda g - \mu h)(a^*(s), s)} \theta \cdot x \right) \leq D_- V(s; x)$$

6
Proof. We judiciously rewrite our maximization problem so as to apply results from Clarke [8], omitting the equality constraints $h$ to simplify notations. As in Clarke [8], we work with a minimization problem, and explain how the results follow for a maximization problem. Consider:

$$V(s) = \min f(a, s) \text{ s.t. } g(a, s) \leq 0$$

Denote $y = (a, a')$ and rewrite the problem as:

$$V(s) = \min f(y) \text{ s.t. } g(y) \leq 0 \text{ and } -k(y) + s = 0$$

in which $k(y) = (0, 1) \cdot y$ (so the equality condition is simply $a' - s = 0$). Form the Lagrangian:

$$L(y, r, \theta, s) = f(y) + \lambda g(y) + \theta [-k(y) + s]$$

By Theorem 6.1.1 in Clarke [8], there exists $\lambda \geq 0$, and $\theta$ such that $\lambda g(y) = 0$ and:

$$0 \in \partial_y L(y, \lambda, \theta, s)$$

Recall that for a Lipschitz function $f$, at the point $(a, s)$:

$$\partial_y f((a, s)) \subset \partial_a f((a, s)) \times \partial_a' f((a, s))$$

Given that $\partial_y h(x) = (0, 1)$ the multiplier rule implies that there exists $(\sigma + \lambda \gamma_s, \sigma_s + \lambda \gamma_s) \in \partial_a (f(a, s) + \lambda g(a, s)) \times \partial_{a'} (f(a, s) + \lambda \partial_a g(a, s))$ such that, for all $(u, v)$:

$$0 = (\sigma + \lambda \gamma_a) \cdot u + (\sigma_{a'} + \lambda \gamma_{a'}) \cdot v - \theta \cdot v$$

Thus:

$$\sigma + \lambda \gamma_a = 0$$

or, equivalently

$$0 \in \partial_a (f(a, s) + \lambda (a, s))$$

and:

$$\theta = \sigma_{a'} + \lambda \gamma_{a'}$$

The assumptions of Corollary 4 in Clarke [8] are satisfied (as noted before, GMFCQ implies that $M^0(\sum) = \{0\}$), so:

$$V^+(s; x) \leq \sup_{(a, s) \in \sum} \inf_{(\lambda, \theta) \in M^1(a, s)} \theta \cdot x = \inf_{a \in A(s)} \sup_{\lambda \in K(a, s)} \sup_{\theta \in \partial_n (f(a, s) + \lambda g(a, s))} \theta \cdot x$$

and similarly for $V_+(s; x)$.

The value function $W$ associated with the maximization problem:

$$W(s) = \max f(a, s) \text{ s.t. } g(a, s) \leq 0$$
satisfies $W(s) = -V(s)$ where:

$$V(s) = \min -f(a, s) \text{ s.t. } g(a, s) \leq 0$$

Since:

$$W^+(s; x) = \limsup_{t \downarrow 0} \frac{W(s + tx) - W(s)}{t} = -\liminf_{t \downarrow 0} \frac{V(s + tx) - V(s)}{t} = -V^+(s; x)$$

we have:

$$W^+(s; x) \leq -\inf_{a \in A^+(s)} \inf_{\lambda \in K(a, s)} \inf_{\theta \in \partial_d \left( -f + \lambda g \right)(a, s)} \theta \cdot x = \sup_{a \in A^+(s)} \sup_{\lambda \in K(a, s)} \sup_{\theta \in \partial_d \left( -f + \lambda g \right)(a, s)} \theta \cdot x$$

and, similarly:

$$W^+(s; x) \geq -\inf_{a \in A^+(s)} \sup_{\lambda \in K(a, s)} \sup_{\theta \in \partial_d \left( f - \lambda g \right)(a, s)} \theta \cdot x = \sup_{a \in A^+(s)} \inf_{\lambda \in K(a, s)} \inf_{\theta \in \partial_d \left( f - \lambda g \right)(a, s)} \theta \cdot x$$

which proves the result. Both the generalized gradient and the optimal solution correspondence $A^*(s)$ are compact valued, so inf and sup become min and max. ■

### 3.2 Consequences

Characterizing additional properties of the value function requires more from the primitive data than just the local Lipschitz structure. First, we note that with continuously differentiable functions Clarke’s generalized gradient becomes a singleton. Assuming then that SMFCQ holds implies that the multiplier is unique. Consequently, the upper and lower Dini derivative of $V$ coincide, and the value function is directionally differentiable.

**Corollary 8** Under Clarke’s Hypothesis, if the SMFCQ holds at every optimal solution $a^*(s) \in A^*(s)$, and the primitive data is $C^1$ in $s$, then the value function is directionally differentiable with:

$$V'(s; x) = \max_{a^*(s) \in A(s)} L_2(a^*(s), s, \mu) \cdot x$$

An alternative to SMFCQ is to assume enough concavity to "squeeze" the lower and upper bounds to obtain a directional envelope, as done in Milgrom and Segal [26] Corollary 5. We derive such a result from Theorem 7 for a less restrictive environment.

**Corollary 9** Under Clarke’s hypothesis, if the primitive data is $C^1$ in $s$, $f$, and $-g$ concave, $h$ affine in $a$, the derivatives $f_s$, $-g_s$, $h_s$ are upper semicontinuous in $a$, and GMFCQ holds at all $a^*(s)$ in $A^*(s)$, then for any direction $x \in \mathbb{R}^m$:

$$V''(s; x) = \max_{a^*(s) \in A^*(s)} \min_{(\lambda, \mu) \in K(a^*(s), s)} L_3(a^*(s), s, \mu) \cdot x$$
Proof. By Theorem 7:

\[
\max_{a^*(s) \in K(a^*(s), s)} \min_{\lambda \in K(a^*(s), s)} \lambda L_2(a^*(s), s, \lambda, \mu) \leq D_+ V(s; x)
\]

Imposing additional conditions on the primitive data helps tighten the upper bound as follows. Choose a sequence \( \{t_n\} \) converging to 0 such that:

\[
\limsup_{t \to 0^+} \frac{V(s + tx) - V(s)}{t} = \lim_{n \to \infty} \frac{V(s + t_n x) - V(s)}{t_n}
\]

Consider a sequence \( \{a^*(s + t_n x)\} \) where \( a^*(s + t_n x) \in A^*(s + t_n x) \). By Clarke’s Hypothesis (and since \( V \) is continuous at \( s \)), for \( n \) large enough all \( a^*(s + t_n x) \) belong to a compact set, so without loss of generality we may assume that \( a^*(s + t_n x) \) converges to some \( a^* \). By closeness of \( D \), necessarily in \( a^* \in D(s) \), and by continuity of \( V, V(s) = f(a^*, s) \). Hence \( a^* \in A^*(s) \) is a global maxima. By strong duality, the Lagrangian has a global saddle point at \( (a^*, s, \lambda, \mu) \) where \( (\lambda, \mu) \in K(a^*, s) \). Thus, for any \( (\lambda, \mu) \in K(a^*, s) \):

\[
\lim_{n \to \infty} \frac{V(s + t_n x) - V(s)}{t_n} = \lim_{n \to \infty} \frac{L(a^*(s + t_n x), s + t_n x, \lambda_n, \mu_n) - L(a^*, s, \lambda, \mu)}{t_n}
\]

where \( (\lambda_n, \mu_n) \in K(a^*(s + t_n x), s + t_n x) \). Consequently:

\[
\lim_{n \to \infty} \frac{V(s + t_n x) - V(s)}{t_n} \leq \lim_{t_n \to 0^+} \frac{L(a^*(s + t_n x), s + t_n x, \lambda, \mu) - L(a^*, s, \lambda, \mu)}{t_n} \leq \lim_{t_n \to 0^+} \frac{L(a^*(s + t_n x), s + t_n x, \lambda, \mu) - L(a^*(s + t_n x), s, \lambda, \mu)}{t_n} = L_2(a^*(s + t_n x), s, \lambda, \mu) \cdot x
\]

The first and the second inequality follows from the fact that \( (a^*(s + t_n x), s + t_n x, \lambda_n, \mu_n) \) and \( (a^*, s, \lambda, \mu) \) are global saddle points of \( L \) for any \( s + t_n x \) and \( s \) respectively.

The above inequality is true for all \( (\lambda, \mu) \in K(a^*, s) \), thus necessarily:

\[
\limsup_{t \to 0^+} \frac{V(s + tx) - V(s)}{t} \leq \min_{(\lambda, \mu) \in K(a^*(s), s)} L_2(a^*(s + t_n x), s, \lambda, \mu) \cdot x
\]

Finally, since \( a^*_n(s) \to a^* \in A^*(s) \) and \( L_2(.) \) is upper semicontinuous in its first argument, the above inequality implies:

\[
D_+ V(s; x) \leq \max_{a^*(s) \in A^*(s)} \min_{(\lambda, \mu) \in K(a^*(s), s)} L_2(a^*(s), s, \lambda, \mu) \cdot x
\]
Under SMFCQ, by a direct application of Corollary 9: the gradient of the value function. This is precisely because the value function remains concave. Any optimal solution along with its corresponding multiplier can be used to calculate the lower bound for every optimal solution the value function becomes once continuously differentiable.

\[ V'(s) = \max_{a^*(s) \in A^*(s)} L_2(a^*(s), s, \lambda, \mu) \cdot x \]

**Proof.** Under SMFCQ, by a direct application of Corollary 9:

\[ V'(s, x) = \max_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x \]

Evidently, for all directions \( x \):

\[
-V'(s, -x) = - \max_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot (-x) \\
= \min_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x \\
\leq \max_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x \\
= V'(s, x)
\]

The concavity of \( f \), and \(-g\) is transmitted to \( V \) which in turn implies that:

\[-V'(s, -x) \geq V'(s, x),\]

Thus:

\[-V'(s, -x) = V'(s, x)\]

and therefore

\[ \min_{\xi \in \partial V} \{\xi \cdot x\} = \max_{\xi \in \partial V} \{\xi \cdot x\} \]

for all \( x \in \mathbb{R}^n \). Thus, the Clarke gradient is singleton implying \( V \) is strictly differentiable. By Rockafellar and Wets [34] (Corollary 9.19) \( V \) is \( C^1 \). Also, note

\[-V'(s, -x) = V'(s, x) \]

\[ \Leftrightarrow \min_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x = \max_{a^*(s) \in A^*(s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x \]

for all \( x \in \mathbb{R}^n \). Thus the gradient of \( V \) is given by

\[ V'(s) = F_s(a^*(s), s) - \lambda g_s(a^*(s), s) \]

for any \( a^*(s) \in A^*(s) \) and \( \lambda \in K(a^*(s), s) \). ■

Corollary 10 does not require the set of optimal solutions to be singleton. However, any optimal solution along with its corresponding multiplier can be used to calculate the gradient of the value function. This is precisely because the value function remains concave.
4 APPLICATIONS AND EXTENSIONS

4.1 Lipschitz Dynamic Programming

We extend the results of Laraki and Sudderth [25] and Hinderer [19] on the preservation of Lipschitz continuity in recursive dynamic programs by weakening the global Lipschitz conditions on the primitive data to local Lipschitzness.

Consider a $N$ period/stage Lipschitz dynamic program, to which we associate the recursive (primal) formulation:

$$V_n(s) = \max_a \{ F(a, s) + \beta V_{n+1}(a) \}$$

subject to $g(a, s) \leq 0$, $n = 1, ..., N - 1$, $a \in A \subset \mathbb{R}^n$ and $s \in S \subset \mathbb{R}^m$, and $V_N = 0$, and the corresponding (dual) programs:

$$L_{N-j}(a, s) = F(a, s) + \beta V_{N-j+1}(a) - \lambda g(a, s)$$

with $j = 1, ..., N - 1$. We make the following assumptions:

Assumption 4.1: $F$ and $g$ are locally Lipschitz, and admit differential extension on the boundary of $A \times S$, and the feasible correspondence is uniformly compact for all $s \in S$.

Although $V_n$ is not necessarily concave or/and $C^1$, the nonsmooth constraint qualification GMFCQ allows us to characterize the local Lipschitz structure of the value function of the primal program using the Lagrangian dual formulation as follows:

**Proposition 11** Under assumption 4.1, if GMFCQ is satisfied for every optimal solution $a^*(s) \in A^*(s)$, then each $V_n$ is locally Lipschitz with Clarke gradient:

$$\partial V_n(s) \subset co \left\{ \bigcup_{a^*_n(s) \in A^*_n(s)} \bigcup_{\lambda \in K^*(a^*_n(s), s)} \partial_s (F - \lambda g)(a^*_n(s), s) \right\}$$

**Proof.** Follows directly from Theorem 6. □

Since each $V_i$ is locally Lipschitz, the generalized (period $n$) multiplier rule is:

$$0 \in \partial_a (F(a, s) + \beta V_{n+1}(a) - \lambda g(a, s))(a^*(s))$$

Note that if $a^*(s)$ is interior this condition becomes:

$$0 \in \partial_a (F(a, s) + \beta V_{n+1}(a))(a^*(s))$$

which, whenever $F$ is $C^1$ in $a$, further simplifies to:

$$-D_a F(a^*(s), s) \in \partial_a (\beta V_n(a^*(s)))$$

Taking $N \to \infty$, the repeated application of Proposition 11 above generates a sequence $\{V_n\}$ of locally Lipschitz functions converging uniformly (via a standard application of the
contraction mapping theorem) to the unique continuous function $V^*$ solving Bellman’s equation. Unfortunately, uniform limits of sequences of locally Lipschitz functions are not necessarily locally Lipschitz.\footnote{By the Weistrass Approximation Theorem any continuous functions on $[0,1]$ (including non Lipschitz continuous functions) may be uniformly approximated by polynomials (which are Lipschitz).} Nevertheless, there are well-known conditions on this problem yielding a locally Lipschitz limit. One can, for instance, impose sufficient structure on primitives to make sure that the optimal solutions are "uniformly interior".\footnote{E.g., see the discussion in Montrucchio [30]. Sufficient conditions in economic models for such properties typically involve multivariate Inada conditions (e.g., as in Askri and LeVan [3]), or strong concavity conditions such as in Montrucchio [29].} Alternatively, concavity requirement may be sufficient to guarantee that $V^*$ is locally Lipschitz as discussed next.

### 4.1.1 Concave Dynamic Programming

Consider the infinite horizon version of the previous dynamic program (as in Bonnisseau and LeVan [7] or Rincon-Zapatero and Santos [33]), for which we make the following assumptions, typically satisfied by a large class of dynamic programs.

Assumption 4.1.1: $F$ and $g$ are jointly locally Lipschitz and concave in $(a, s)$, $C^1$ in $s$, and admit differential extensions on the boundary of $A \times S$. In addition, the derivatives $F_s$ and $g_s$ are upper semicontinuous in $a$, and the feasible correspondence is uniformly compact for all $s \in S$.

Although we do not impose smoothness in the choice variable $a$, concavity, along with the right constraint qualification, will guarantee the existence of directionally differentiable envelope for the primal dynamic program using a recursive Lagrangian dual formulation, as given by the next corollary.

**Proposition 12** Under assumption 4.1.1 and if GMFCQ is satisfied for every optimal solution $a^*(s) \in A^*(s)$, then $V$ is concave and locally Lipschitz, and:

$$V'(s, x) = \max_{a^*(s) \in A^*(s)} \min_{\lambda \in K(a^*(s), s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x$$

If the stronger assumption of SMFCQ is made for $C^1$ primitives, then $V$ is concave and $C^1$ on the interior of $S$, with derivative given by:

$$V'(s) = (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x$$

for any $a^*(s) \in A^*(s)$.

**Proof.** A recursive application of Bellman’s operator generates a sequence $\{V_n\}$ of concave functions converging uniformly to a concave function $V$. Concavity of $V$ implies almost everywhere differentiability. At points of nondifferentiability $V$ is at least directionally differentiability (a bit more, since concave functions are lower Clarke regular), and thus locally Lipschitz. Recalling that $V$ solves the Lipschitz program:

$$V(s) = \max \{F(a, s) + \beta V(a)\}$$
subject to $g(a, s) \leq 0$, under GMFCQ a direct application of Corollary 9 establishes that:

$$V'(s, x) = \max_{a^*(s) \in A^*(s)} \min_{\lambda \in K(a^*(s), s)} (F_s(a^*(s), s) - \lambda g_s(a^*(s), s)) \cdot x$$

Next, recalling that SMFCQ for $C^1$ programs implies uniqueness of the multipliers, Corollary 10 establishes that:

$$V_0(s) = F_s(a^*(s), s) - \lambda g_s(a^*(s), s)$$

for any $a^*(s) \in A^*(s)$. ■

Our result on $C^1$ envelopes generalizes Benveniste and Scheinkman [5] by allowing the inequality constraints to be active at the optimal solution. We also generalize Rincon-Zapatero and Santos [33] by considering a less restrictive constraint qualification, namely the SMFCQ.

### 4.1.2 Differentiability of the Pareto Frontier

Consider the model in Kocherlakota and Koeppel (see also Rincon-Zapatero and Santos [33]) of an exchange economy in which two infinitely lived agents receive a stochastic endowment in each period, which they mutually share under limited commitment. As in Koeppel, the endowment for agent $i = 1, 2$ in period $t$ is $(\omega^1_s, \omega^2_s)$ which is determined by the realization of $\theta_t$, $\theta = \{\theta_1, \theta_2, \ldots, \}$ is a sequence of iid random variables, each having finite support $\Theta = \{1, 2, \ldots, S\}$. The probability that $\theta_t$ equals $s$ is denoted by $\pi_s$ for all $s$ in $\Theta$. We will assume the following, in which we relax Koeppel’s assumptions of strict monotonicity, strict concavity and $C^2$ utility function.

**Assumption 4.1.2:** The utility function $u: \mathbb{R}_+ \rightarrow \mathbb{R}$ is increasing, concave and $C^1$, with differential extensions on the boundaries. $\lim_{c \rightarrow 0^+} u'(c) = \infty$, $0 < \beta < 1$, and for each $U_0$, the feasible set is uniformly compact.

We characterize the incentive feasible allocations (see Koeppel for details).

$$V(U_0) = \max_{\{c_s, u_s\}_{s=1}^S} \sum_{s=1}^S \pi_s [u(\omega_s - c_s) + \beta V(U_s)]$$

subject to

$$U_0 - \sum_{s=1}^S \pi_s [u(c_s) + \beta U_s] \leq 0$$

$$u(\omega^1_s) + \beta U_{aut} - u(c_s) - \beta U_s \leq 0$$

$$u(\omega_s - \omega^1_s) + \beta U_{aut} - u(\omega_s - c_s) - \beta V(U_s) \leq 0$$

$$U_s \in [U_{aut}, U_{max}]$$

The set of optimal solutions is denoted $Y^*(U_0)$, and a typical element of this set is $\{c^*_s, U^*_s\}_{s=1}^S$; the KKT multiplier vector takes the form $(\lambda_1, \{\lambda_{2s}\}, \{\lambda_{3s}\}, \{\lambda_{4s}\}, \{\lambda_{5s}\}) \in K(\{c^*_s, U^*_s\})$. Given assumption 4.1.2, and the joint concavity of the objective and the constraints in $(\{c_s, u_s\}, U_0)$, the hypothesis of Proposition 12 are met. Thus, by Proposition 12, if GMFCQ
is satisfied for every optimal solution \((c^*(U_0), U^*(U_0)) \in Y^*(U_0)\), then \(V\) is concave with directional derivatives given by,

\[
V'(U_0; x) = \max_{(c^*, U^*) \in Y^*(U_0)} \min_{\lambda \in K(c^*, U^*)} \{-\lambda_1 \cdot x\}
\]

Further, if SMFCQ is satisfied for every optimal solution \((c^*(U_0), U^*(U_0)) \in Y^*(U_0)\), then \(V\) is concave and \(C^1\) with derivative given by:

\[
\nabla V(U_0) = -\lambda_1
\]

for any \((\lambda_1, \{\lambda_{2s}\}, \{\lambda_{3s}\}, \{\lambda_{4s}\}, \{\lambda_{5s}\}) \in K(\{c^*_s, U^*_s\})\).

### 4.2 Optimization problems with discrete choice variables

Consider a Lipschitz program in which one of the decision variable (here \(a_1\)) is constrained to take only one of \(r\) possible values, that is:

\[
\max_{a \in D(s)} f(a, s)
\]  

where \(D(s) = \{g(a, s) \leq 0, \text{ and } a_1 = b_j \ j = 1, \ldots, r\}\). We simply rewrite the \(r\) equality constraints \(a_1 = b_j\) as the \(C^1\) equality constraint \(h(a, s) = \prod_{j=1}^{r} (a_1 - b_j) = 0\). At any optimum \(a^*(s), a^*_1(s)\) must equal some \(b_k\), so that \(\nabla_a h(a^*(s), s) = (\prod_{j \neq k} (b_k - b_j), 0, \ldots, 0) \neq 0\).

Associated with this standard maximization problem is the Lagrangian:

\[
L(a, \lambda, \mu; s) = f(a, s) - \lambda g(a, s) - \mu h(a, s)
\]

and \(a^*(s) = (b_k, a^*_2(s), \ldots, a^*_n(s)) \in A^*(s)\) is a KKT point if there exists a vector \(\lambda \geq 0\) of multipliers and \(\mu \in \mathbb{R}\) such that:

\[
\mu \nabla_a h(a^*(s)) \in \partial_a (f - \sum_{i \in I(a^*(s), s)} \lambda_i g_i)(a^*(s), s))
\]

where \(I(a^*(s), s)\) is the set of active inequality constraints. In this problem, GMFCQ is defined as follows.

**Definition 13** A feasible point \(a \in D(s)\) satisfies the GMFCQ if there exists a \(\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n) \in \mathbb{R}^n\) such that:

\[
\exists \tilde{y} \in \mathbb{R}^n, \forall \gamma \in \partial_a \overline{g}(a, s), \gamma \cdot \tilde{y} < 0, \text{ and } h_1(a, s)\tilde{y} = 0.
\]

Note that necessarily \(\tilde{y}_1 = 0\) for \(\tilde{y}\) to meet this condition. By a straightforward application of Theorem 7: 

14
Proposition 14 If $D(s)$ is nonempty-valued and uniformly compact near $s$, and if the GM-FCQ holds for every $a^*(s) \in A^*(s)$, then for any direction $x \in \mathbb{R}^m$:

$$\lim \inf_{t \to 0^+} \frac{V(s + tx) - V(s)}{t} \geq \inf_{(\lambda, \mu) \in K(a^*(s), s)} \min_{\theta \in \partial_x (f - \lambda g - \mu h)(a^*(s), s)} \theta \cdot x$$

and:

$$\lim \sup_{t \to 0^+} \frac{V(s + tx) - V(s)}{t} \leq \sup_{(\lambda, \mu) \in K(a^*(s), s)} \max_{\theta \in \partial_x (f - \lambda g - \mu h)(a^*(s), s)} \theta \cdot x$$

$V$ is locally Lipschitz with Clarke gradient:

$$\partial V(s) \subset \text{co} \left\{ \bigcup_{a^*(s) \in \Lambda^*(s)} \bigcup_{(\lambda, \mu) \in K(a^*(s), s)} \partial_x (f - \lambda g - \mu h)(a^*(s), s) \right\}$$

Such result directly applies to the Consider a finite horizon ($N$ periods) labor-leisure choice problem in which labor takes only the binary values \{0, 1\}. Thus we formulate a $N$ period problem as,

$$V_n(k_n) = \max_{c_n, k_{n+1}, l_n} \left\{ u(c_n, 1 - l_n) + \beta V_{n+1}(k_{n+1}) \right\}$$

subject to

$$c_n + k_{n+1} - f(k_n, l_n) \leq 0$$

$$-c_n \leq 0$$

$$-k_{n+1} \leq 0$$

$$l_n(1 - l_n) = 0$$

for all $n \leq N$, and $(k_{N+1}, V_{N+1}) = (0, 0)$. We also assume the following:

Assumption 4.2: Functions $u : \mathbb{R}_+^n \times \mathbb{R}_+ \to \mathbb{R}$ and $f : \mathbb{R}_+^n \times \mathbb{R}_+ \to \mathbb{R}_+$ are locally Lipschitz with differential extensions on the boundaries, and strictly increasing in both arguments. The feasible set $D_n : \mathbb{R}_+^n \Rightarrow \mathbb{R}_+^n \times \mathbb{R}_+ \times \mathbb{R}_+$ is nonempty-valued and uniformly compact near for all $l_t$.

Since $u$ is strictly increasing the first inequality constraint always holds with an equality, and the associated Lagrangian is:

$$L_n(k_n) = u(f(k_n, l_n) - k_{n+1}, 1 - l_n) + \beta V_{n+1}(k_{n+1}) + \lambda_1^1(f(k_n, l_n) - k_{n+1})$$

$$+ \lambda_2^2 k_{n+1} + \mu_n(l_n(1 - l_n))$$

Under assumption and if the GMFCQ holds for every optimal solution $y_n^*(k_n) \in V_n^*(k_n)$, a direct consequence of Proposition 14, is that for any direction $x \in \mathbb{R}^n$:

$$\lim \inf_{t \to 0^+} \frac{V_n(k_n + tx) - V_n(k_n)}{t} \geq \inf_{(\lambda_n, \mu_n) \in K^*(y_n^*(k_n), k_n)} \min_{\theta \in \partial_{k_n} L_n(y_n^*(k_n), k_n)} \theta \cdot x$$

and:

$$\lim \sup_{t \to 0^+} \frac{V_n(k_n + tx) - V_n(k_n)}{t} \leq \sup_{(\lambda_n, \mu_n) \in K^*(y_n^*(k_n), k_n)} \max_{\theta \in \partial_{k_n} L_n(y_n^*(k_n), k_n)} \theta \cdot x$$

and $V_n$ is locally Lipschitz for all $n = 1, \ldots, N$. 

15
4.3 Computing Markov equilibrium in growth models with nonsmooth nonconvex technologies

Recent work on optimal growth models with nonsmooth and nonconvex technologies have largely set aside the issue of existence of decentralized recursive and/or sequential equilibrium (see, for instance, Kamihigashi and Roy [22][23]). We propose to establish the existence of decentralized Markov equilibrium in these models by combining the envelope results in this paper with the methodology of Mirman, Morand, and Reffett [28].

We consider a class of model with a continuum of identical infinitely-lived households/firms and where time periods are indexed by $t = \{0, 1, 2, \ldots\}$. A household enters each period with an individual stock of capital $k$ and an endowment of one unit of time (supplied inelastically to firms). For simplicity, uncertainty is modeled as a sequence of iid shocks $z \in \mathbb{Z}$ with density function $\chi$, where $\mathbb{Z}$ is assumed to be countable. We put standard restrictions on the primitive data to guarantee that aggregate per-capita capital stock $K$ and individual capital $k$ remain in a compact set $K \subset \mathbb{R}_+$. The aggregate state of an economy is given by the vector $S = (K, z) \in K \times \mathbb{Z} \subset \mathbb{R}_+^2$, and the individual state by $p = (k, S) \in K \times K \times \mathbb{Z}$.

As common in this literature (e.g., Coleman [9], Greenwood and Huffman [16]), and consistent with recent work (e.g., [36] [22] [23]) we use a "reduced-form" production function $F(x, n, K, N, z)$ with $F$ increasing, locally Lipschitz, lower Clarke regular jointly in $(x, n, K, N, z)$. We assume that $F$ is CRS in private inputs $(x, n)$ for each $(K, N, z)$, where $x$ is the firms capital input, $n$ is the firms labor input, and that Inada conditions hold when $n = N = 1$, $K > 0$, all $z$. We state explicitly the class of economies under consideration in the following assumption. When convenient we use the notation $f(k, K, z) = F(k, 1, K, 1, z)$.

Assumption 4.3: Let functions $u$ and $F$ satisfy the following conditions:

(i) $u : K \mapsto \mathbb{R}$ is isotone, concave, locally Lipschitz in the interior of $K$, and satisfies $u(0) = 0$ and $\lim_{c \to 0} u'(c) \to \infty$ at points of differentiability;

(iii) $F : K \times N \times K \times N \times \mathbb{Z} \to \mathbb{R}$ is isotone, locally Lipschitz, and exhibits constant returns to scale in $(k, n)$. For all $z \in \mathbb{Z}$ there exists $k(z) > 0$ such that $F(k(z), 1, k(z), 1, z) = k(z)$ and $F(k, 1, k, 1, z) < k$ for all $k > k(z)$.

(iv) Almost everywhere $u'(c(k(K, z)))f_1(k, K, z)$ is increasing in $K$ for any function $c$ increasing in $k$.

Given factor prices $(r, w)$ and aggregate inputs $(K, N, z) > 0$, the profit function is:

$$\pi(r, w) = F(x, n, K, N, z) - rx - wn$$

In equilibrium $N = n = 1$, and the profit maximizing conditions are simply:

$$r \in \partial_1 F(K, 1, K, 1, z)$$
$$w \in \partial_2 F(K, 1, K, 1, z)$$

---

6 So we can set aside the measurability issue. See Mirman, Morand, and Reffett [28] for more general shocks processes in compact finite dimensional intervals.

7 As $u$ and $f$ are locally Lipschitz, this derivative exists almost everywhere, and the condition is thus an almost everywhere version of the complementarity condition between private and social returns developed Mirman, Morand, and Reffett [28].
and almost everywhere:
\[ r = F_1(K, 1, K, 1, z); w = F_2(K, 1, K, 1, z) \]

Household lifetime preferences are given by:
\[ E_0 \left\{ \sum_{i=0}^{\infty} \beta^i u(c_i) \right\} \]  \hspace{1cm} (4)

The household begins any period in state \((k, K, z) \in K \times K \times Z\), takes as given a law of motion \(h\) for the per-capita capital stock, and chooses investment \(y\).

We consider the set \(B\) of laws of motion defined as:
\[ B = \{ h : K \times Z \to K, 0 \leq h(k, z) \leq f(k, k, z) \} \]
and endowed with the topology of pointwise convergence. Under the pointwise partial order \(B\) is a lattice, and we introduce the sublattice \(USC \subset B\) of functions \(h\) that are isotone in \((k, z)\) and upper semicontinuous in \(k\), that is: i.e.,
\[ USC = \{ h \in B, \text{usc in } k \} \]
The set \(USC\) is a complete lattice under the pointwise partial order (see for instance Davey and Priestley [12], 2.31), with top and bottom elements \((f,0)\) continuous in \(k\).

The value function \(v^*\) associated with the household’s maximization problem satisfies the standard functional equation:
\[ v^*(k, K, z; h) = \sup_{y \in \Gamma(k, K, z)} \left\{ u(f(k, K, z) - y) + \beta \int v^*(y, h(K, z), z'; h) \chi(dz') \right\}, \]
where \(\Gamma(k, K, z) = \{ y \in K, 0 \leq y \leq f(k, K, z) \}\) is the non-empty, continuous, compact and convex valued feasible correspondence. A standard argument shows that \(v^*(., h) : P \to \mathbb{R}\) is bounded, isotone and continuous in \(k\) for each \((K, z)\). We denote:
\[ Y^*(k, K, z; h) = \arg \sup_{y \in \Gamma(k, K, z)} \left\{ u(f(k, K, z) - y) + \beta \int v(y, h(K, z), z'; h) \chi(dz') \right\} \]
the set of optimal investment solutions.

To compute a recursive equilibrium for this economy, we propose to modify the lattice programming argument of Mirman, Morand, and Reffett [28]. The main difficulty consists in obtaining sufficient conditions for the preservation of the supermodularity of the value function under the sup operation in the dynamic program above. The traditional result of Topkis (e.g. [38], Theorem 2.7.6) does not apply, so our approach is based on the generalized envelope theorems in the previous section.

Let \(\forall^*\) and \(\land^*\) denote the greatest and least selection of the optimal solution correspondence \(Y^*\), and \(\geq_a\) Veinott’s strong set ordering. Then, we have the following result:

---

Using least selectors, we can also verify existence of lower semicontinuous recursive equilibrium in the space \(LSC(S)\), which is the space \(USC(S)\) only assuming each element is lsc, not usc. See Mirman, Morand, and Reffett [28] for discussion.
Lemma 15 Given any \( h \) in USC, \( Y^* \) is isotone in \( h \) in the strong set order for any given \((k, K, z)\). The maximal selection \( \forall Y^* \) is isotone in \((k, K, z; h)\). The operator \( A : h \rightarrow Ah \) defined as \( Ah(k, z) = \forall Y^*(k, k, z; h) \) is order continuous on USC.9

Proof. We prove below that \( v^* \) is supermodular in \((k, K)\) by obtaining \( v^* \) as the pointwise limit of a sequence of functions (recursively constructed with Bellman’s operator) that are supermodular. The supermodularity of \( v^* \) is the critical building block in our argument, and it has two important consequence. First, the term \( Ev^*(y, h(K, z), z') \) in Bellman’s equation above has increasing differences with respect to \( v \) and it has two important consequence. First, the term \( Ev^*(y, h(K, z), z') \) in Bellman’s equation above has increasing differences with respect to \( (y, (k, K, z)) \) which implies again by the same Theorems in Topkis\[38\] (Theorems 2.8.1 and 2.8.3) the optimal correspondence \( Y^* \) is strong set order ascending in \( h \) and both \( \forall Y^* \) and \( \wedge Y^* \) are isoton in \( h \). Second, given that \( h \) is increasing in its arguments (and given our assumptions on \( u \) and \( F \), the right hand side in Bellman’s equation above has increasing differences with respect to \( (y, (k, K, z)) \) which implies again by the same Theorems in Topkis\[38\] that \( Y^* \) is strong set order ascending in \((k, K, z)\), so that both both \( \forall Y^* \) and \( \wedge Y^* \) are isoton in \((k, K, z)\). The order continuity of the maximal selection follows from Mirman, Morand, and Reffett (\[28\], Theorem 6).

We turn now to the proof of supermodularity of \( v^* \) in \((k, K)\). Fixing \( h \), and suppressing it from the notation, and given \((K, z)\), \( K > 0 \), \( z \in Z \), consider Bellman’s operator \( T \) constructed as:

\[
v_n(k, K, z) = T^n v_0(k, K, z) = T v_{n-1}(k, K, z) = \max_{0 \leq y_0 \leq f(k, K, z)} \{ u(f(k, K, z) - y_0) + \beta \int Z v_{n-1}(y_n, h(K, z), z' \chi(dz')) \}
\]

Given \( v_0 = 0 \), \( T^n v_0 \) is locally Lipschitz in \( k \) by a recursive application of Theorem 6. By Theorem 7 and noting that Inada conditions imply that \( \lambda_1 = \lambda_2 = 0 \):

\[
\max_{y^*_n(p) \in Y^*_n(p)} \left( \min_{\theta \in \partial_k(u(f(p) - y^*_n(p)))} \theta \cdot x \right) \leq D^+ v_n(p; x)
\]

and:

\[
D^+ v_n(p; x) \leq \max_{y^*_n(p) \in Y^*_n(p)} \left( \max_{\theta \in \partial_k(u(f(p) - y^*_n(p)))} \theta \cdot x \right)
\]

The local Lipschitz property of \( v_n \) in \((k)\) imply that it is almost everywhere differentiable, and at points of differentiability the above bounds necessarily coincide so that:

\[
\min_{\theta \in \partial_k(u(f(p) - y^*_n(p)))} \theta \cdot x = \max_{\theta \in \partial_k(u(f(p) - y^*_n(p)))} \theta \cdot x
\]

However, \( u \circ f \) is locally Lipschitz and also almost everywhere differentiable, so that almost everywhere:

\[
\min_{\theta \in \partial_k(u(f(p) - y^*_n(p)))} \theta \cdot x = \left( u'(f(p) - y^*_n(p)) f_1(p) \right) \cdot x
\]

---

9For a partially ordered set \((C, \leq)\), an operator \( g : C \rightarrow C \) is order continuous if for any countable subchain \( C_n \subset C \), \( \sup_n \vee g(C_n) = \sup_n g(\vee C_n) \) (dually, if \( \inf_n \wedge g(C_n) = \inf_n g(\wedge C_n) \)). Note, an order continuous operator is necessarily isoton.
As a result, almost everywhere:

\[
\partial_k v_n(p) = D_k v_n(p) = \max_{y_n(p) \in V_n(p)} \left( u'(f(p) - y_n^*(p)) \right) f_1(p)
\]

A standard argument shows that under the Inada condition, consumption \( c_n^*(p) = f(p) - y_n^*(p) \) is bounded away from 0 on any \([\theta, k_{\text{max}}]\) where \( \theta > 0 \), which implies that the derivative \( D_k v_n(\cdot, K, z) \) is bounded on \([\theta, k_{\text{max}}]\). As a result, for each \( K > 0 \) the function \( v_n(\cdot, K, z) \) is continuous, locally Lipschitz, differentiable a.e. with a bounded derivative where it exists; it is therefore necessarily globally Lipschitz on \([\theta, k_{\text{max}}]\).

We next prove that \( v_n \) is absolutely continuous in its first argument on \([0, k_{\text{max}}]\) for each \((K, z), K > 0\). Absolute continuity means that, given any \((K, z)\) with \( K > 0 \) and any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that for any finite collection of non-overlapping intervals \( \{ (x_i, x'_i) \}_{i=1}^n \) satisfying:

\[
\sum_{i=1}^N |x'_i - x_i| < \delta
\]

it is the case that:

\[
\sum_{i=1}^N |v_n(x'_i, K, z) - v_n(x_i, K, z)| < \varepsilon.
\] (5)

Our proof of absolute continuity consists in dividing the collection of non-overlapping intervals into two disjoint subcollections. The first subcollection is near 0, and the continuity of \( v_n \) to the right of 0 implies that the corresponding part of the sum can be made arbitrarily small. The remainder of the sum in (5) (corresponding to the second subcollection) can also be made arbitrarily small by appealing to the Lipschitz property of \( v_n \) on any \([\theta, k_{\text{max}}]\), \( \theta > 0 \). The details are as follows. As \( v_n(\cdot, K, z) \) is continuous, for any \( \varepsilon/2 \), there exists \( k_{\text{max}} > k_0 > 0 \) such that \( v_n(k_0, K, z) - v_n(0, K, z) < \varepsilon/2 \). Set \( \theta = k_0 \) and without loss of generality, there exists \( P \geq 1 \) such that \((x_i, x'_i)\) belongs to the interval \([0, \theta]\) for \( i = 1, \ldots, P \) and \((x_i, x'_i)\) belongs to \([\theta, k_{\text{max}}]\) for \( j = P + 1, \ldots, N \). On \([\theta, k_{\text{max}}]\), \( v_n(\cdot, K, z) \) is globally Lipschitz so there exists \( M_n(K, z) \) such that:

\[
\sum_{i=j+1}^n |v_n(x'_i, K, z) - v_n(x_i, K, z)| \leq M_n(K, z) \sum_{j=1}^n |x'_j - x_j| < M_n(K, z) \delta
\]

On \([0, \theta]\), \( v_n \) is increasing and continuous in its first argument so that:

\[
\sum_{i=1}^j |v_n(x'_i, K, z) - v_n(x_i, K, z)| \leq v_n(\theta, K, z) - v_n(0, K, z) \leq v_n(k_0, K, z) - v_n(0, K, z) < \varepsilon/2
\]

Together these two inequalities imply that:

\[
\sum_{i=1}^n |v_n(x'_i, K, z) - v_n(x_i, K, z)| < M_n(K, z) \delta + \varepsilon/2
\]

19
and therefore $\delta = \varepsilon/(2M_n(K, z))$ establishes the desired result.

Having established each $v_n(., K, z)$ is absolutely continuous in $k$, by Theorem 5.14 in Royden [35], we have:

$$v_n(k, K, z) = \int_0^k D_1 v_n(x, K, z)dx.$$ 

where:

$$D_1 v_n(k, K, z) = u'(f(k, K, z) - \wedge Y_n^*(k, K, z))f_1(k, K, z)$$

Being able to express $v_n$ as the integral of its derivative will permit to use the monotonicity of $D_1 v_n$ in $k$ to prove the supermodularity of $v_n$ whenever $h$ is increasing in $k$ by induction. $v_0 = 0$ is trivially supermodular, so assume that $v_{n-1}$ is supermodular in $(k, K)$ and consider again the nth iteration of the Bellman operator:

$$v_n(k, K, z) = \max_{0 \leq k' \leq f(k, K, z)} \{u(f(k, K, z) - k') + \beta E v_{n-1}(k', h(K, z), z')\}.$$ 

Given iid shocks and $h(K, z)$ increasing in $K$, $\beta E v_{n-1}(k', h(K, z), z')$ has increasing differences in $(k'; (K, z))$. Also, $u(f(k, K, z) - k')$ has increasing differences in $(k'; (K, z))$ (as $u'' \leq 0$ and $F_k \geq 0$ and $F_K \geq 0$) and the objective on the right hand side of Bellman's equation has increasing differences with respect to $(k'; (k, K, z))$, while the choice correspondence $[0, f(k, K, z)]$ is strong set order ascending. By Topkis (Theorems 2.8.1 and 2.8.3), $Y_n^*$ is strong set order ascending, and $\vee Y_n^*$ and $\wedge Y_n^*$ are both increasing selections in $(k, K, z)$. Recall that almost everywhere:

$$D_1 v_n(k, K, z) = u'(f(k, K, z) - \wedge Y_n^*)f_1(k, K, z)$$

Under the curvature condition in Assumption 4.3 implies, if consumption $(f - y_n^*)(k, K, z)$ is increasing in $K$, $D_1 v_n$ is increasing in $K$. If $(f - y_n^*)(k, K, z)$ decreases when $K$ rises, then the supermodularity of $f$ in $(k, K)$ together with the concavity of $u$ imply that $D_1 v_n$ increases.

Thus, for $k' > k$ and $K' > K$, each $z$:

$$v_n(k', K', z) - v_n(k, K', z) = \int_k^{k'} D_1 v_n(x, K', z)dx$$

$$\geq$$

$$\int_k^{k'} D_1 v_n(x, K, z)dx = v_n(k', K, z) - v_n(k, K, z)$$

\[\text{[10]}

Because the shocks are iid, the value function being supermodular in $(k, K, z)$ is not needed to get monotone controls in $z$. 

20
i.e., \( v_n \) is a supermodular function in \((k,K)\), each \( z \). By induction, the supermodularity property is true for all \( n \), and the sequence of functions \( \{v_n\}_{n=0}^\infty \) is a collection of supermodular functions in \((k,K)\). Its pointwise limit, which is precisely \( v^* \) must therefore share that same property.

We now prepared to prove the existence of Markovian equilibrium. For any \( h \in \text{USC} \), recall the construction of the operator \( A \) as \( Ah(k,z) = \vee Y^*(k,k,z;h) \), and that a Markov Equilibrium Decision Process is precisely a fixed point of \( A \). The previous lemma established that \( A \) maps \( \text{USC} \) into itself, and that \( A \) is order continuous (and thus isotone). Since \( \text{USB} \) is a complete lattice, by Tarski’s fixed point theorem applied to \( A \), there exists a nonempty complete lattice of MEDP The fact the greatest fixed point can be computed by successive approximate follows from the Tarski-Kantorovich theorem (e.g., Dugundji and Granas ([11], Theorem 4.2), noting the order continuity of \( A \); and the fact that \( A \) maps the greatest element of \( \text{USB} \) (the function \( h_0(k,z) = f(k,k,z) \)) down.

**Theorem 16** Under Assumption 4.3, the set of MEDPs is a nonempty, complete lattice in \( \text{USC} \). The sequence \( A^n h_0 \) where \( h_0(k,z) = f(k,k,z) \) converges pointwise to the greatest MEDP.

As a final note, we emphasize that it is easy to obtain equilibrium comparative statics with respect to some of the deep parameters of this economy, even given the nonsmooth primitive data (in particular, equilibrium comparative statics relative to the space of production functions \( F \) and changes in the discount rate \( \beta \), for example). For example, if we order \( F \) using the "gradient monotonicity": such that \( f' \geq g \) when \( f'(k,K,z) - f(k,K,z) \) is increasing in \( k \), with \( f' - f = 0 \), when \( k = 0 \), then for fixed \((K,z)\), two production functions are ordered if (i) their "levels" in \( k \) are ordered, for all \((K,z)\), and (ii) their gradient fields at points of differentiability (or private marginal products) are ordered in \( k \), for all \((K,z)\). At points of nondifferentiability, integral forms of increasing differences as in the proof of the theorem above can be used to check increasing differences between \((y;f)\) for each \( \beta \). Then, as the operator \( \vee Y^*(p;h,f) \) is jointly isotone in \((h,f)\), each \( p \), it is easy to show we have monotone equilibrium monotone comparative statics for the set of recursive equilibria in \( \text{USC} (S) \) using related arguments to those in Mirman, Morand, and Reffett [28]; we just check a new version of the increasing differences using our nonsmooth envelope theorem in the previous section.

5 **APPENDIX A: Mathematical Tools**

5.1 Derivatives and subgradients

5.1.1 Differentiability

A function \( f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m \) is said to be:

(i). Locally Lipschitz at \( x \) with modulus \( k \), \( 0 \leq k < \infty \), if there exists \( \delta > 0 \) such that for all \( x',x'' \) in \( B(x,\delta) \):

\[
|f(x'') - f(x')| \leq k |x'' - x'|
\]
If $k$ can be chosen independently of $x$ on an open subset of $\Omega$, $f$ is said to be globally Lipschitz on that subset. By Rademacher’s theorem, if $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz at all point of an open set $\Theta \subset \Omega$, then it is almost everywhere differentiable on $\Theta$.

(ii). Directionally differentiable (or Gateaux differentiable) at $x_0$ if the directional derivative:

$$f'(x_0; d) = \lim_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t},$$

exists for all $d$.

(iii). Differentiable at $x_0 \in X$ if it is directionally differentiable and if $f'(x_0; d) = \nabla f(x_0) \cdot d$ (note, for instance that the function $x \to |x|$ is Gateaux differentiable but not differentiable), continuously differentiable at $x_0$ if the function $\nabla f(.) : \Omega \to \mathbb{R}^{n \times m}$ is continuous at $x_0$, and strictly differentiable at $x_0$ if there exists a continuous linear (in $d$) function $Df(x_0)$ such that:

$$\lim_{z \to x_0, t \to 0^+} \frac{f(z + td) - f(z)}{t} = Df(x_0)(d)$$

A strictly differentiable function is obviously differentiable (the converse is not true) but not necessarily continuously differentiable.

### 5.1.2 Differentiability of Lipschitz functions

Directional derivatives of locally Lipschitz functions do not necessarily exist. However, if $f$ is Lipschitz at $x$, the following derivatives always exist (i.e., are finite quantities):

(i). The upper and lower (right hand) Dini derivative, respectively defined as the function:

$$D^+ f(x; d) = \limsup_{t \to 0^+} \frac{f(x + td) - f(x)}{t} \quad \text{and:}$$

$$D_+ f(x; d) = \liminf_{t \to 0^+} \frac{f(x + td) - f(x)}{t}$$

(ii). The Clarke upper and lower derivatives at $x$ respectively defined as the functions:

$$f^+(x; d) = \limsup_{y \to x, t \to 0^+} \frac{f(y + td) - f(y)}{t} \quad \text{and:}$$

$$f^- (x; d) = \liminf_{y \to x, t \to 0^+} \frac{f(y + td) - f(y)}{t}$$

If $f$ is locally Lipschitz and directionally differentiable at $x$, then:

$$f^-(x; d) \leq f'(x; d) \leq f^+(x; d)$$

A function $f$ is said to be upper (resp. lower) Clarke regular at $x$ if it is directionally differentiable at $x$ and if its Clarke upper (resp. lower) derivative coincide with its directional derivative, i.e., if $f^+(x; d) = f'(x; d)$ (resp. $f^-(x; d) = f'(x; d)$).
5.1.3 Clarke gradient

The Clarke gradient of a Lipschitz function $f$ at $x$ is the nonempty compact convex set:

$$\partial f(x) = \overline{\text{co}} \{ \lim \nabla f(x_i) : x_i \rightarrow x, x_i \notin \Theta, x_i \notin \Omega_f \}$$

where $\overline{\text{co}}$ denotes the convex hull\(^{11}\), $\Theta$ is any set of Lebesgue measure zero in the domain, and $\Omega_f$ is a set of points at which $f$ fails to be differentiable. Clarke [8] (Proposition 2.1.5) shows that $x \mapsto \partial f(.)$ is an upper hemicontinuous correspondence.

For a convex (concave) function $f$, the Clarke gradient at $x$ coincides with the set of $p \in M_{m \times n}$ satisfying:

$$p \cdot d \leq (\geq) f(x_0 + d) - f(x_0)$$

for all directions $d \in \mathbb{R}^n$. Clarke [8] (Proposition 2.1.2) also shows that:

$$f^o(x; d) = \max_{\zeta \in \partial f(x)} \{ \zeta, d \}$$

which implies that $f^o(x; d)$ is a convex function of $d$ (equivalently, $f^{-o}(x; d)$ is concave in $d$). This also implies that:

$$-f^o(x; -d) = f^{-o}(x; d) \leq f^o(x; d) = -f^{-o}(x; -d)$$

and, equivalently, that:

$$f^{-o}(x; d) \leq -f^{-o}(x; -d) = f^o(x; d)$$

Finally, we note that for a Lipschitz function at $x$, strict differentiability at $x$ is equivalent to the equality of the Clarke derivatives at $x$ so that:

$$f^{-o}(x; d) = \lim_{y \rightarrow x, t \rightarrow 0^+} \frac{f(y + td) - f(y)}{t} = f^o(x; d).$$

The mean value theorem asserts that if $f$ is locally Lipschitz on $\Omega$, then for all $x, y \in \Omega$, $\exists \gamma \in \text{co} \{ \cup_{z \in [x,y]} \partial f(z) \}$ such that $f(y) - f(x) = \gamma \cdot (y - x)$ where $[x,y]$ is the line segment joining $x$ and $y$ (Clarke [8] Theorem 2.3.7).

5.2 Properties of Correspondences

We work in metric spaces, so we can state topological properties of correspondences exclusively in terms of sequences.

\(^{11}\)In the formula, either $\text{co}$ or $\overline{\text{co}}$ will do since we work with finite dimensional spaces.
Definition 17  Given $A \subset R^n$ and $S \subset R^m$, a non-empty valued correspondence $D : S \rightarrow A$ is:

(i) lower hemicontinuous at $s$ if for every $a \in D(s)$ and every sequence $s_n \rightarrow s$ there exists a sequence $\{a_n\}$ such that $a_n \rightarrow a$ and $a_n \in D(s_n)$.

(ii) upper hemicontinuous at $s$ if for every sequence $s_n \rightarrow s$ and every sequence $\{a_n\}$ such that $a_n \in D(s_n)$ there exists a convergent subsequence of $\{a_n\}$ whose limit point $a$ is in $D(s)$.

(iii) closed at $s$ if $s_n \rightarrow s$, $a_n \in D(s_n)$ and $a_n \rightarrow a$ implies that $a \in D(s)$ (in particular, this implies that $D(s)$ is a closed set).

(iv) open at $s$ if for any sequence $s_n \rightarrow s$ and any $a \in D(s)$, there exists a sequence $\{a_n\}$ and a number $N$ such that $a_n \rightarrow a$ and $a_n \in D(s_n)$ for all $n \geq N$.

Note that $D(s) = \{a \in A, g_i(a, s) \leq 0, i = 1, ..., p\}$, in which the $g_i$ are locally Lipschitz (and thus continuous), is necessarily closed at $s$. The same property holds true in the presence of locally Lipschitz equality constraints.

Another property of correspondences which is critical in our analysis is that of uniform compactness.

Definition 18  A non-empty valued correspondence $D$ is said to be uniformly compact near $s$ if there exists a neighborhood $S'$ of $s$ such that $\text{cl} \left[ \bigcup_{s' \in S'} D(s) \right]$ is compact.

We note the result in Hogan [21] that if $D$ is uniformly compact near $s$, then $D$ is closed at $s$ if and only if $D(s)$ is a compact set and $D$ is upper hemicontinuous at $s$. When $D$ is defined by a system of continuous equality and inequality constraints, uniform compactness near $s$ thus implies compactness and upperhemicontinuity at $s$. In fact, for any $s'$ sufficiently close to $s$, since $D(s')$ is a closed subset of $\text{cl} \left[ \bigcup_{s' \in S'} D(s) \right]$ it is therefore compact.

Finally, we will need the following property of hemicontinuous correspondences (and thus of Clarke gradients).

Proposition 19  If $D$ is an upper hemicontinuous correspondence, then for every compact neighborhood $K$ of $x$, the set:

$$\bigcup_{z \in K} D(z)$$

is compact.

Proof. Consider a sequence $\{y_n\}$ in $\bigcup_{z \in K} D(z)$ so that $y_n \in D(z_n)$ for some $z_n$ in $K$. The sequence $\{z_n\}$ is the compact $K$, so there exists a subsequence of $\{z_{\varphi(n)}\}$ of $\{z_n\}$ converging to some $z' \in K$. By upper hemicontinuity of $D$ at $z'$, there exists a subsequence of $\{y_{\varphi(n)}\}$ converging to some $y \in D(z')$. This proves that the initial sequence $\{y_n\}$ has a convergent subsequence, and therefore that the set $\bigcup_{x \in K} D(x)$ is compact. ■
References


