A Generalized Approach to Envelope Theorems

Olivier Morand* Kevin Reffett† Suchismita Tarafdar‡

August 2012
Preliminary and Incomplete

Abstract

We develop a generalized approach to envelope theorems that applies across a broad class of parameterized nonlinear optimization problems that arise typically in economic applications. In particular, we provide sufficient conditions under which the value function for a nonconvex, and/or nonsmooth program is locally Lipschitz and/or Clarke differentiable. We then apply our results to Lipschitz dynamic programming problems with and without concavity assumptions, to discrete labor choice models, and provide some examples from the redistributive taxes, and Research and Development literature. Also, our companion paper, Morand, Reffett and Tarafdar [23] extends the results of this paper to provide sufficient conditions for the existence of directionally differentiable and C^1 envelopes.

1 INTRODUCTION

Constrained maximization problems are an essential building block of both microeconomic and macroeconomic theory, and powerful results exists that characterize the solutions to these problems and the properties of the objective at the maximum (i.e., the value function) in the form of envelope theorems. Such problems typically translate into finding the values of a decision variable a that maximize an objective $f(a, s)$ subject to some inequality constraints $g(a, s) \leq 0$ (and possibly also to the equality constraints $h(a, s) = 0$), and traditional results are been derived under the assumption of smoothness (i.e., once continuously differentiable, or C^1 constraints and objectives) and concavity. Important recent work has focused on loosening these assumptions, as for instance in Rincon-Zapateros and Santos (2009), Milgrom and Segal [21], Clausen and Straub ([9]), and Marimon and Werner [20]. The goal of this paper is to provide a baseline result for Lipschitz programs without any concavity or differentiability assumption.

There are many situations in which the traditional results of convex analysis do not apply. The objective and/or the inequality constraints may neither be smooth nor concave, in which case the standard optimality conditions of the theory of Lagrange multipliers (i.e.,

*Department of Economics, University of Connecticut
†Department of Economics, WP Carey School of Business, Arizona State University
‡Planning Unit, Indian Statistical Institute (ISI), Delhi

the Karush-Kuhn-Tucker condition or the Fritz-Jones condition) clearly do not exist, and standards results on the differential properties of the value function (i.e., envelope theorems) are also not available. The lack of concavity is a problem, since multiple maximizers appear, and it is compounded with the lack of smoothness which trigger the loss of uniqueness of Lagrange multipliers.

One must also bear in mind that even something as simple as the continuity of the value function may be hard to establish, since it often requires the mapping $s \rightrightarrows D(s) = \{a, g(a,s) \leq 0 \text{ and } h(a,s) = 0\}$ associated with the feasible set must be a continuous correspondence for Berge's theorem of the maximum to deliver the desired continuity. Unfortunately, it very often fails to be so, even in fairly simple models

Finally, we note that our results are applicable to programs with discrete choices and therefore helpful in computing solutions to these programs, since this involves discreticizing the state space. We leave the full details of this application to a future paper on Lipschitz dynamic programming.

To handle these issues, we provide a fundamental result showing the existence of bounds for the Dini derivatives of the value function of Lipschitz programs, and prove the Lipschitz property of the value function for such programs. This fundamental result is key to developing envelope theorems for non-smooth non-concave programs that extend the existing standard results of Morand, Reffett and Tarafdar [23]. In our proof, we also present an alternative to Berge's theorem of the maximum that uses a local condition to guarantee the continuity of the value function.

2 LIPSCHITZ PROGRAMS AND CONSTRAINT QUALIFICATIONS

We consider Lipschitz programs of the form:

$$\max_{a \in D(s)} f(a,s) \tag{1}$$

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

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

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

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

is the optimal solution correspondence. In Lipschitz programs objective and constraints are only locally Lipschitz in (a,s), in contrast to "smooth programs" in which objective and constraints are typically assumed to be continuously differentiable. In this paper unless otherwise mentioned the objective and the inequality constraints are locally Lipschitz, but the equality constraints are smooth(or, C^1). We'll maintain the following assumption through out the paper:

Assumption: The sets A and S are each convex in $\mathbb{R}^n$ and $\mathbb{R}^m$ *respectively, and the feasible correspondence $D : S \rightrightarrows A$ is nonempty, continuous and compact-valued. Further, the objective $f(a,s)$ and the constraints $g(a,s)$* and $h(a,s)$ *both admit differential extensions over the boundary of $A \times S$.*

2.1 Optimality conditions

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 (λ, μ) of multipliers with $\lambda \geq 0$ such that:

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

The closed and convex (but perhaps empty) set of vectors (λ, μ) satisfying the above multiplier rule will be denoted $K(a,s)$ (or simply K) . To guarantee the non-emptiness and boundedness (and thus compactness) of K at a local optimum requires assumptions in the form of constraint qualifications (the so-called "optimality conditions"). In programs with smooth primitive data, constraints qualifications are restrictions involving the (classical) gradients of the constraints. For instance, it is well known that a point $a^*(s) \in A^*(s)$ satisfying the Mangasarian Fromovitz constraint qualification stated immediately below is a KKT point and is such that K is compact (see for instance Gauvin [12]).

Definition 1 *The point $a^*(s) \in A^*(s)$ satisfies the Mangasarian-Fromovitz Constraint Qualification (MFCQ) if, there exist $\widetilde{y} \in \mathbb{R}^n$ such that*

$$\begin{aligned} \nabla_a g_i(a^*(s),s) \cdot \widetilde{y} &< 0,\ i \in I(a^*(s),s), \\ \nabla_a h_j(a^*(s),s) \cdot \overline{y} &= 0\ j = 1,...,q \end{aligned}$$

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

Kyparisis [19] has showed that a slightly less general condition, the Strict Mangasarian-Fromovitz Constraint Qualification (SMFCQ), which simply treats active inequality constraints for which the multiplier is strictly positive as equality constraints, is necessary and sufficient for the uniqueness of the multiplier vector in smooth programs.

Definition 2 *The SMFCQ is satisfied at $a^*(s) \in A^*(s)$ if, there exist $\widetilde{y} \in \mathbb{R}^n$ such that*

$$\begin{aligned} &\nabla_a g_i(a^*(s),s) \cdot \widetilde{y} < 0,\ i \in I_s(a^*(s),s) \\ &\nabla_a g_i(a^*(s),s) \cdot \widetilde{y} = 0,\ i \in I_b(a,s) \\ &\nabla_a h_j(a^*(s),s) \cdot \overline{y} = 0\ j = 1,...,q \end{aligned}$$

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 $\nabla_a g_i(a^*(s),s)$, $i \in I_b(a^*(s),s)$, $\nabla_a h_j(a^*(s),s)$, $j = 1,...,q$ are linearly independent.

With Lipschitz constraints no classical gradients generically exist, but the following generalization of the MFCQ for Lipschitz programs (which we denote GMFCQ) has been shown by Hiriart-Urruty [17] to be sufficient for the non-emptiness of K. This non-smooth version of the MFCQ uses generalized gradients (or "subdifferentials") as developed by Clarke (see Appendix A for mathematical tools and definitions). To simplify the notations, we denote by $\overline{g}(a^*(s), s)$ the vector of binding inequality constraints at point $a^*(s) \in A^*(s)$, so that $\overline{g} : A \times S \to \mathbb{R}^{\overline{p}}$ (where $\overline{p} = Card(I(a^*(s), s)) \leq p$).

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

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

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

For simplicity of exposition, the results and proofs in the core of the paper are for Lipschitz programs without equality constraints, i.e. with the feasible correspondence taking the form:

$$D(s) = \{a \in A,\ g_i(a, s) \leq 0, \quad i = 1, ..., p\},$$

in which $g_i : A \times S \to \mathbb{R}$, $i = 1, ..., p$. As shown in Appendix C of the paper, all results extend to the case of programs with smooth equality constraints. With no equality constraints, the GMFCQ is simpler and reduces to the following.

Definition 4 *The Generalized Mangasarian-Fromovitz Constraint Qualification (GMFCQ) is satisfied at $a^*(s) \in A^*(s)$ if there exist $\widetilde{y} \in \mathbb{R}^n$ such that*

$$\exists \widetilde{y} \in \mathbb{R}^n,\ \forall \gamma_a \in \partial_a \overline{g}(a^*(s), s),\ \gamma_a \cdot \widetilde{y} < 0. \tag{2}$$

Note that $\overline{g}$ is the vector of active inequality constraints and that:

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

so this version of the GMFCQ is slightly more general than that of Auslender [4].

2.2 Consequences of the GMFCQ

An envelope theorem for a parameterized program is basically a result equating the derivative of the value function along a particular direction in the state space to 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. When deriving such a result, technical difficulties arise when some inequality constraints are active (i.e. corner solutions), especially when Lagrange multipliers are not unique, and when working with Lipschitz functions, since traditional derivatives, or gradients (singletons) give place to generalized gradients (not necessarily singleton)[1]. In face of these difficulties, and absent sufficient smoothness, convexity

[1] For ease of notations, "generalized gradients" will be denoted as simply "gradients" in the sequel.

and regularity conditions, a simple envelope for parametrized Lipschitz programs seems out of reach. However, the GMFCQ and the uniform compactness assumptions help us get closer to our goal.

First, the upper hemicontinuity of Clarke's generalized gradients implies that the set of multipliers is closed. In addition, GMFCQ implies that the set of multiplier is nonempty and bounded as demonstrated for instance in Nguyen, Strodiot and Mifflin, [25]. It is therefore a non-empty compact set.[2]

Second, under GMFCQ we characterize the lower bound of the gradient of the Lagrangian (with respect to the state s) in the direction x at an optimum in terms of the saddle value of a function $\mathcal{S}^L$. The lower bound of the gradient of the Lagrangian is precisely the lower Clarke derivative of the Lagrangian in the direction x, and this result is essential to finding bounds for the Dini derivative of the value function. In the remaining section we consider a Lipschitz program with only inequality constraints.

Theorem 5 *Suppose GMFCQ holds at the optimal solution $a^*(s)$. Then, for any $x \in \mathbb{R}^m$:*

$$\mathcal{S}^L(y,\lambda) = \min_{(\varsigma_a,\gamma_a)\in\partial_a(f,\overline{g})(a^*(s),s)} (\varsigma_a - \lambda^T\gamma_a)\cdot y + \min_{(\varsigma_s,\gamma_s)\in\partial_s(f,\overline{g})(a^*(s),s)} (\varsigma_s - \lambda\lambda^T\gamma_s)\cdot x$$

has a saddle point, and:

$$-\infty < \inf_{\lambda\geq 0}\sup_y \mathcal{S}^L(y,\lambda) = \sup_y \inf_\lambda \mathcal{S}^L(y,\lambda) = \inf_{\lambda\in K} L_s^{-o}(a^*(s),s;\lambda;x) < +\infty$$

where:

$$L_s^{-o}(a^*(s),s;\lambda;x) = \min_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta\cdot x$$

Proof. See Appendix B. Note that $L_s^{-o}(a^*(s),s;\lambda;x)$ is the lower Clarke derivative of the Lagrangian at $(a^*(s),s)$. ■

Third, we show that the GMFCQ implies that for any perturbation of s in the direction x, there exists of a "feasible direction" $\overline{y}$ such that the gradient of the objective at the optimal and in the particular direction $(\overline{y},x)$ is arbitrarily close to the lower bound (over all KKT vectors) of the gradient of the Lagrangian in the direction x. This is precisely the two inequalities in the following lemma.

Lemma 6 *Assume that GMFCQ holds at $a^*(s)\in A^*(s)$. Then for any direction of perturbation $x\in\mathbb{R}^m$, and any $\varepsilon>0$, there exists a vector $\overline{y}(x,\varepsilon)$ such that for all $(\varsigma,\gamma)\in\partial(f,\overline{g})(a^*(s),s)$:*

$$\gamma\cdot(\overline{y}(x,\varepsilon),x) < 0$$

and:

$$\varsigma\cdot(\overline{y}(x,\varepsilon),x) > \inf_{\lambda\in K} L_s^o(a^*(s),s;\lambda;x) - \varepsilon$$

[2]Nguyen, Strodiot and Mifflin [25] proves that GMFCQ is actually equivalent to the non-emptiness and boundedness of the multiplier set.

Proof. From our previous result (5) the saddle point of S is such that:

$$\begin{aligned}\inf_{\lambda\geq 0}\sup_{y}\mathcal{S}^L(y,\lambda) &= \sup_{y}\inf_{\lambda\geq 0}\mathcal{S}^L(y,\lambda)\\ &= \inf_{\lambda} L_s^{-o}(a^*(s),s;\lambda;x) = \sup_{y\in G}\min_{(\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s)}(\varsigma_a\cdot y+\varsigma_s\cdot x)\end{aligned}$$

In this last expression the supremum may not be attained since the set G (defined in the proof of the Theorem 6 in Appendix B) is not necessarily bounded. However, for any given ε, there exists $y(x,\varepsilon)$ in G such that:

$$\min_{(\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s)}[\varsigma_a\cdot y(x,\epsilon)+\varsigma_s\cdot x]\geq \inf_{\lambda\in K} L_s^{-o}(a^*(s),s;\lambda;x)-\varepsilon/2$$

Next, define $\overline{y}=y(x,\varepsilon)+\delta\widetilde{y}$, where $\delta>0$ is arbitrarily small and $\widetilde{y}$ satisfies GMFCQ, so that:

$$\forall\gamma_a\in\partial_a\overline{g}(a,s),\ \delta\gamma_a\cdot\widetilde{y}<0$$

Thus, $\forall\left((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s)\right)\in\partial_a(f,\overline{g})\times\partial_s(f,\overline{g})$:

$$\gamma\cdot(\overline{y}(x,\varepsilon),x)=(\gamma_a\cdot y(x,\varepsilon)+\gamma_s\cdot x)+\delta\gamma_a\cdot\widetilde{y}<0$$

since $y(x,\varepsilon)$ was chosen in G (and thus $\varsigma_a\cdot y(x,\varepsilon)+\varsigma_s\cdot x<0$). We also have:

$$\min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))\in\partial_a(f,\overline{g})\times\partial_s(f,\overline{g})}[\varsigma_a\cdot\overline{y}+\varsigma_s\cdot x]\geq\inf_{\lambda\in K}L_s^{-o}(a^*(s),s;\lambda;x)-\varepsilon/2+\delta\min(\varsigma_a\cdot\widetilde{y})$$

and we choose δ small enough such that $\delta\min(\varsigma_a\cdot\widetilde{y})>-\varepsilon/2$. Recalling that $\partial(f,g)\subset\partial_a(f,\overline{g})\times\partial_s(f,\overline{g})$, we obtain that $\forall(\varsigma,\gamma)\in\partial(f,\overline{g})(a^*(s),s)$:

$$\forall(\varsigma,\gamma)\in\partial(f,\overline{g})(a^*(s),s),\ \varsigma\cdot(\overline{y},x)>\inf_{\lambda\in K}L_s^{-o}(a^*(s),s;\lambda;x)-\varepsilon$$

and:

$$\gamma\cdot(\overline{y}(x,\varepsilon),x)<0.$$

■

Finally, when combined with the assumption of uniform compactness of D, the GMFCQ yields a very powerful results similar to that of Berge's maximum theorem: the value function is continuous, and the optimal correspondence if upper hemicontinuous, as shown in the following lemma. It is important to note that the continuity of V cannot come from a direct application of Berge's maximum theorem, since the domain D is not necessarily continuous even though all constraints are continuous. Consider for instance the domain D in $\mathbb{R}^2$ defined as:

$$D(m)=\{(x,y),x+y\leq m\text{ and }(m-11)(10-x)\leq 0\}$$

which is not continuous at $m=11$.

Note also that the upper semicontinuity of A^* is a key 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 $f(.,s)$. It also implies that $A^*(s)$ is compact.

Lemma 7 *If GMFCQ holds, and if D is non-empty valued and uniformly compact in a neighborhood of s, then the value function V is continuous at s, and the optimal correspondence A^* is upper hemicontinuous at s.*

Proof. D is closed at s and uniformly compact near s, and therefore upper hemicontinuous at s, so by Berge V is upper semicontinuous at s. To prove lower semicontinuity of V consider a sequence $\{s_n\}$ converging to s and such that $\lim_{s' \to s} \inf V(s') = \lim_{s_n \to s} V(s_n)$. Let $\widetilde{s}_n = \frac{(s_n - s)}{\|s_n - s\|}$ and $t_n = \|s_n - s\|$ so that $s_n = s + t_n \widetilde{s}_n$. Since $\|\widetilde{s}_n\| = 1$, the sequence $\{\widetilde{s}_n\}$ has a convergent subsequence. Thus, without loss of generality, assume that $\lim_{n \to \infty} \widetilde{s}_n = \widetilde{s}$.

In the direction $\widetilde{s}$, there exists some $\widetilde{y}$ satisfying lemma 6 above. By the mean value theorem (See Appendix A), there exists $\gamma(t) \in co_{x \in T}\{\partial \overline{g}(x)\}$ where $T = [(a^*(s), s), (a^*(s) + t\overline{y}, s + t\widetilde{s})]$ such that:

$$\overline{g}(a^*(s) + t\overline{y}, s + t\widetilde{s}) - \overline{g}(a^*(s), s) = t\gamma(t) \cdot (\widetilde{y}, \widetilde{s})$$

and by upper hemicontinuity $\gamma(t)$ converges to some $\gamma \in \partial \overline{g}(a^*(s), s)$, and by construction $\gamma \cdot (\widetilde{y}, \widetilde{s}) < 0$. Thus it must be the case that $\overline{g}(a^*(s) + t\overline{y}, s + t\widetilde{s}) - \overline{g}(a^*(s), s) < 0$ for t small enough, or, equivalently:

$$\overline{g}(a^*(s) + t_n\overline{y}, s + t_n\widetilde{s}) = \overline{g}(a^*(s) + t_n\overline{y}, s_n) < 0$$

for n large enough. This inequality obviously holds for the inequality constraints that are not active at $(a^*(s), s)$ and thus for n large enough:

$$g(a^*(s) + t_n\overline{y}, s_n) < 0$$

This implies that $a^*(s) + t_n\overline{y}$ is in the feasible domain $D(s_n)$ so that $V(s_n) \geq f(a^*(s) + t_n\overline{y}, s_n)$. Thus:

$$\begin{aligned}
\liminf_{s' \to s} V(s') &= \lim_{s_n \to s} V(s_n) \\
&\geq \lim_{n \to \infty} f(a^*(s) + t_n\overline{y}, s_n) \\
&= f(a^*(s), s) \\
&= V(s)
\end{aligned}$$

which proves that V is lower semicontinuous.

Since V is continuous at s, the map $L : s \to \{a, f(a, s) - V(s) \geq 0\}$ is closed at s. 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 mapping D (recall that uniform compactness and closeness imply upper hemicontinuity), and is therefore upper hemicontinuous. Indeed consider $s_n \to s$ 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, $a \in A^*(s) = L(s) \cap A(s)$, which proves that A^* is upper hemicontinuous at s (and $A^*(s)$ is therefore a compact set). ■

3 DINI BOUNDS FOR THE VALUE FUNCTION

3.1 Main Result

In this section we establish lower and upper bounds for the Dini derivatives[3]:

$$\begin{aligned} D_+V(s) &= \lim_{t\longrightarrow 0^+}\inf \frac{V(s+tx)-V(s)}{t} \\ &\leq \\ D^+V(s) &= \lim_{t\rightarrow 0^+}\sup \frac{V(s+tx)-V(s)}{t} \end{aligned}$$

Constructing the lower bound for D_+V relies heavily on lemma 6 above. Briefly, by that Lemma there exists a feasible direction $\overline{y}$ such that a subgradient ς of the objective function at the optimal and in the direction $(\overline{y},x)$ (i) is arbitrarily close to the minimum subgradient (with respect to s) of the Lagrangian in the direction x, and (ii) is a lower bound for the rate of growth of the value function in the direction x. Combining these two properties gives the desired result. The upper bound for D^+V is obtained in a similar manner.

Theorem 8 *If D is nonempty valued and uniformly compact near s, and if the GMFCQ holds at every optimal solution $a^*(s)\in A^*(s)$, then for any direction of perturbation $x\in\mathbb{R}^m$, (i) for all $a^*(s)\in A^*(s)$*

$$-\infty < \inf_{\lambda\in K(a^*(s),s)}\left(\min_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta\cdot x\right) \leq D_+V(s;x)$$

and, (ii)

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

Proof. Consider any $a^*(s)$ in $A^*(s)$. Given ε and x, consider $\overline{y}=\overline{y}(x,\varepsilon)$ satisfying lemma 6. By the mean value theorem, there exists $(\varsigma(t),\gamma(t))\in co\{\cup_{x\in T}\partial(f,\overline{g})(x)\}$ where $T=[(a^*(s),s),(a^*(s)+t\overline{y},s+tx]$,

$$(f,\overline{g})(a^*(s)+t\overline{y},s+tx)-(f,\overline{g})(a^*(s),s)=t(\varsigma(t),\gamma(t))\cdot(\overline{y},x), \tag{3}$$

By upper hemicontinuity of the subdifferential, without loss of generality we may assume that $(\varsigma(t),\gamma(t))$ converges to some subgradient $(\varsigma,\gamma)\in\partial(f,\overline{g})(a^*(s),s)$ as $t\downarrow 0$. Since by construction $\overline{y}$ satisfies:

$$\gamma\cdot(\overline{y},x)<0$$

and

$$\varsigma\cdot(\overline{y},x) > \inf_{\lambda\in K} L_s^{-o}(a^*(s),s;\lambda;x)-\varepsilon$$

[3] Appendix C generalizes the bounds of the Dini derivatives by including C^1 equality constraints.

then for t small enough:

$$\gamma(t).(\overline{y}, x) < 0$$

and

$$\varsigma(t) \cdot (\overline{y}, x) > \inf_{\lambda \in K} L_s^{-o}(a^*(s), s; \lambda; x) - \varepsilon$$

Using (3), this implies that:

$$\overline{g}(a^*(s) + t\overline{y}, s + tx) - \overline{g}(a^*(s), s) = \gamma(t) \cdot (\overline{y}, x) < 0$$

(and thus $a^*(s) + t\overline{y} \in D(s + tx)$), and:

$$f(a^*(s) + t\overline{y}, s + tx) - f(a^*(s), s) = t\varsigma(t) \cdot (\overline{y}, x) \geq t \left(\inf_{\lambda \in K} L_s^{-o}(a^*(s), s; \lambda; x) - \varepsilon \right)$$

As $a^*(s) + t\overline{y} \in D(s + tx)$, necessarily $V(s + tx) \geq f(a^*(s) + t\overline{y}, s + tx)$, and therefore:

$$\begin{aligned} \frac{V(s+tx) - V(s)}{t} &\geq \frac{f(a^*(s) + t\overline{y}, s + tx) - f(a^*(s), s)}{t} \\ &\geq \inf_{\lambda \in K} L_s^{-o}(a^*(s), s; \lambda; x) - \varepsilon \end{aligned}$$

As ε may be chosen arbitrarily small we obtain that:

$$\lim_{t \longrightarrow 0^+} \inf \frac{V(s+tx) - V(s)}{t} \geq \inf_{\lambda \in K} L_s^{-o}(a^*(s), s; \lambda; x)$$

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

The upper bound is a little harder to obtain. First, choose a sequence $\{t_n\}$ converging to 0 such that, in the direction x:

$$D^+V(s; x) = \lim_{t \to 0^+} \sup \frac{V(s+tx) - V(s)}{t} = \lim_{n \to \infty} \frac{V(s + t_n x) - V(s)}{t_n}.$$

To each n corresponds some $a_n^*(s)$ in $A^*(s + t_n x)$ such that $V(s + t_n x) = f(a_n^*(s), s + t_n x)$. Given the upper hemicontinuity of the optimal correspondence A^* at s established in lemma 7, there exists a subsequence of the sequence $\{a_n^*(s)\}_{n \geq N}\}$ converging to some $a^*(s)$ in $A^*(s)$. Thus, without loss of generality we may assume that $\lim_{n \to \infty} a_n^*(s) = a^*(s) \in D(s)$. Exploiting the continuity of V at s established in that same lemma:

$$\lim_{n \to \infty} V(s + t_n x) = V(s)$$

that is:

$$\lim_{n \to \infty} f(a_n^*(s), s + t_n x) = f(a^*(s), s).$$

Next, let $\overline{y} = \overline{y}(-x, \varepsilon)$ satisfy lemma 6 above for the direction $-x$, and let $a_n(s) = a_n^*(s) + t_n\overline{y}$. By the mean value theorem, there exists some $(\varsigma(t_n), \gamma(t_n))$ in $\overline{co}\{\cup_{x\in T}\partial(f, \overline{g})(a^*(s), s)\}$ where $T = [(a_n(s), s), (a_n^*(s), s + t_n\overline{y})]$ such that:

$$(f, \overline{g})(a_n(s), s) - (f, \overline{g})(a_n^*(s), s + t_n x) = t_n(\varsigma(t_n), \gamma(t_n)) \cdot (\overline{y}, -x) \tag{4}$$

As $n \to \infty$, $t_n \to 0$ and $a_n^*(s) \to a^*(s)$, by upper hemicontinuity of the subdifferential, without loss of generality the sequence $\{(\varsigma(t_n), \gamma(t_n))\}$ converges to some $(\varsigma, \gamma) \in \partial(f, \overline{g})(a^*(s), s)$. By definition of $\overline{y}$:

$$\gamma \cdot (\overline{y}, -x) < 0$$

and:

$$\varsigma \cdot (y, -x) > \inf_{\lambda\in K} L_s^{-o}(a^*(s), s; \lambda; -x) - \varepsilon$$

or:

$$-\varsigma \cdot (y, -x) < \sup_{\lambda\in K} -L_s^{-o}(a^*(s), s; \lambda; -x) + \varepsilon$$

Recall that:

$$-L^{-o}(a^*(s), s; \lambda; -x) = -\min_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta\cdot(-x) = \max_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta\cdot x = L^o(a^*(s), s; \lambda; x)$$

so that:

$$-\varsigma \cdot (y, -x) < \sup_{\lambda\in K} \left(\max_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right) + \varepsilon$$

Then, for n large enough, it must also be the case that:

$$\gamma(t_n) \cdot (\overline{y}, -x) < 0 \tag{5}$$

and:

$$-\varsigma(t_n) \cdot (\overline{y}, -x) < \sup_{\lambda\in K} \left(\max_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right) + \varepsilon \tag{6}$$

Inequalities (4) and (5) imply $\overline{g}(a_n(s), s) < \overline{g}(a_n^*(s), s + t_n x)$, so that $a_n(s) \in D(s)$ since $\overline{g}(a_n^*(s), s + t_n x) \le 0$. Inequalities (4) and (6) imply:

$$f(a_n^*(s), s + t_n x) - f(a_n(s), s) < t_n \sup_{\lambda\in K} \left(\max_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right) + \varepsilon.$$

Given that $a_n(s) \in D(s)$ it must be that $V(s) \ge f(a_n(s), s)$ so that:

$$\begin{aligned}
\limsup_{t\to 0^+} \frac{V(s+tx) - V(s)}{t} &\le \lim_{n\to\infty} \frac{f(a_n^*(s), s + t_n x) - f(a_n(s), s)}{t_n} \\
&\le \sup_{\lambda\in K(a^*(s),s)} [L_s^o(a^*(s), s; \lambda; x)] + \varepsilon \\
&= \sup_{\lambda\in K(a^*(s),s)} \left(\max_{\theta\in\partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right) + \varepsilon
\end{aligned}$$

Since ε was chosen arbitrary small, this shows that for the direction x there exists $a^*(s)$ in $A^*(s)$ such that:

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

Thus, for any direction x,

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

As in the smooth case of , it may be shown that the supremum over all $a^*(s)$ in $A^*(s)$ is attained, so that max may legitimately be substituted for sup. Indeed, consider a sequence $\{a_n, \lambda_n, \theta_n\}$ where $\theta_n \in \partial_s(f - \lambda_n \overline{g})(a_n)$, $\lambda_n \in K(a_n)$, and $a_n \in A^*(s)$ and such that:

$$\lim_{n\to\infty} \theta_n \cdot x = \sup_{a^*(s)\in A^*(s)} \sup_{\lambda \in K(a^*(s))} \left(\max_{\theta \in \partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right)$$

Then, by the compactness of $A^*(s)$ (see Lemma 7), the upper semicontinuity of the subdifferential (see Appendix A), and the upper semicontinuity of the set of all multipliers $\cup_{a^*(s)\in A^*(s)} K(a^*(s), s)$ (see 12 in the next section), there exists a subsequence $\{a_m, \lambda_m, \theta_m\}$ such that $a_m \to a \in A^*(s)$, $\lambda_m \to \lambda' \in K(a)$ and $\theta_m \to \theta' \in \partial_s(f - \left(\lambda^T\right)' \overline{g})(a)$ such that:

$$\sup_{a^*(s)\in A^*(s)} \sup_{\lambda \in K(a^*(s))} \left(\max_{\theta \in \partial_s(f-\lambda^T g)(a^*(s),s)} \theta \cdot x \right) = \theta' \cdot x \leq \sup_{\lambda' \in K(a)} \left(\max_{\theta \in \partial_s(f-\left(\lambda^T\right)' \overline{g})(a)} \theta \cdot x \right)$$

which implies that the sup must be attained for some a in $A^*(s)$. ■

3.2 Consequences

Getting additional properties of the value function requires more from the primitive data than just local Lipschitzness. While we explore various consequences of Theorem 8 in a companion paper, we mention here two interesting results as corollaries. First, under the assumption of strictly differentiability of the primitive data, the upper and lower Clarke derivative of the Lagrangian coincide. Add to this the assumption that the SMFCQ holds at every $a^*(s)$ and the multiplier set become a singleton, as explained in Kyparisis [19] (in Kyparisis strict differentiability and SMFCQ is in fact sufficient to establish uniqueness of multipliers). The upper bound is attained for some a in $A^*(s)$, which then necessarily coincides with the lower bound, and the value function is thus directionally differentiable.

Corollary 9 *If D is nonempty valued and uniformly compact near s, if the SMFCQ holds at every optimal solution $a^*(s) \in A^*(s)$, and if the primitive data is C^1, then the value function is directionally differentiable, and there exists $a^*(s)$ in $A^*(s)$ such that:*

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

and it is upper Clarke regular.

Second, an important result in Milgrom and Segal [21] (Corollary 5) also follows directly from our Dini bounds. In it, rather than assuming SMFCQ to guarantee the uniqueness of the KKT multiplier, it is the concavity of the problem that permits to "squeeze" lower and upper bounds.

Corollary 10 *If D is nonempty valued and uniformly compact near s, if the primitive data is C^1 in s with $f_2(\cdot, s)$, $g_2(\cdot, s)$, $h_2(\cdot, s)$ continuous, f, $-g$ concave and h affine in a, and the MFCQ holds at all $a^*(s)$ in $A^*(s)$, then for any direction $x \in \mathbb{R}^m$, the directional envelope is given by:*

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

Proof. Lemma 8 provides the lower bounds of the Dini derivative, as it implies that:

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

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

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

Since $D(s)$ is uniformly compact near s, for n large, there exists $a^*(s+t_nx) \in D(s+t_nx)$ such that $V(s+t_nx) = f(a^*(s+t_nx), s+t_nx)$. Since the sequence $\{a^*(s+t_nx)\}$ is in a compact domain, without loss of generality we may assume that that $a^*(s+t_nx)$ converges to some $a^*(s)$ in $A^*(s) \subset D(s)$, and by continuity of V, $V(s) = f(a^*(s), s)$. As any $a^*(s) \in A^*(s)$ is a global maxima, appealing to strong duality, the Lagrangian has a global saddle point at $(a^*(s), s, \lambda, \mu)$ where $(\lambda, \mu) \in K(a^*(s), s)$. Thus, for any $(\lambda, \mu) \in K(a^*(s), s)$:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{V(s+t_nx) - V(s)}{t_n} \\
&= \\
&\lim_{n\to\infty} \frac{L(a^*(s+t_nx), s+t_nx, \lambda_n, \mu_n) - L(a^*(s), s, \lambda, \mu)}{t_n}
\end{aligned}$$

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

$$\begin{aligned}
&\lim_{n\to\infty} \frac{V(s+t_nx) - V(s)}{t_n} \\
&\leq \lim_{t_n\to 0^+} \frac{L(a^*(s+t_nx), s+t_nx, \lambda, \mu) - L(a^*(s), s, \lambda, \mu)}{t_n} \\
&\leq \lim_{t_n\to 0^+} \frac{L(a^*(s+t_nx), s+t_nx, \lambda, \mu) - L(a^*(s+t_nx), s, \lambda, \mu)}{t_n} \\
&= L_2(a^*(s+t_nx), s, \lambda, \mu) \cdot x
\end{aligned}$$

where the first inequality follows from:

$$L(a_n^*(s), s + t_n x, \lambda_n, \mu_n) < L(a_n^*(s), s + t_n x, \lambda, \mu)$$

since $(a_n^*(s + t_n x), s + t_n x, \lambda_n, \mu_n)$ is a saddle point, and the second inequality from:

$$L(a^*(s), s, \lambda, \mu) > L(a_n^*(s), s, \lambda, \mu)$$

since $(a^*(s), s, \lambda, \mu)$ is a saddle point.

Since this is true for all $(\lambda, \mu) \in K(a^*(s), s)$, necessarily:

$$\begin{aligned} &\limsup_{t \to 0^+} \frac{V(s + tx) - V(s)}{t} \\ &\leq \min_{(\lambda,\mu) \in K(a^*(s),s)} L_2(a_n^*(s), s, \lambda, \mu) \cdot x \end{aligned}$$

Finally, since $a_n^*(s) \to a^*(s) \in A^*(s)$ and $L_2(.)$ is continuous in its first argument the above inequality imply

$$\begin{aligned} &D^+V(s; x) \\ &\leq \max_{a^*(s) \in A^*(s)} \min_{(\lambda,\mu) \in K(a^*(s),s)} L_2(a^*(s), s, \lambda, \mu) \cdot x \end{aligned}$$

The lower bound for D_+V coincides with the upper bound for D^+V and the result thus follows. ■

4 LIPSCHITZ PROPERTY OF THE VALUE FUNCTION

In this section we show that when the upper and lower bounds of the Dini derivatives of the value function exist at s, they also exist in a neighborhood of s, i.e. V is locally Lipschitz near s[4]. Several technical issues need to be addressed. First, we show in Lemma 11 below that the GMFCQ is a local property, in the sense that when it holds at some $(a^*(s), s)$, then it necessarily also holds in some neighborhood of that point. This property is, of course, obvious in the case of LICQ with C^1 constraints, but less trivial under the GMFCQ assumption. Recall that the GMFCQ is:

(i) h is differentiable with respect to a, $\nabla_a h$ is jointly continuous in a neighborhood of $(a^*(s), s)$, and the matrix $\nabla_a h(a^*(s), s)$ has full rank;

(ii) there exists $\overline{y} \in \mathbb{R}^n$ such that:

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

Second, since the Dini bounds in Theorem 8 are obtained as extrema we guarantee in lemma 12 below that the sup and inf of these extrema exist in a neighborhood of s because of the upper semicontinuity of the joint set of KKT vectors and subgradients.

[4]In this section we will consider a general Lipschitz program with inequalities and C^1 equality constraints.

Lemma 11 *If GMFCQ holds at $(a^*(s), s)$ then it holds in a neighborhood of $(a^*(s), s)$.*

Proof. If $\nabla_a h$ is jointly continuous in a neighborhood of $(a^*(s), s)$, and the matrix $\nabla_a h(a^*(s), s)$ has full rank, then it obviously has full rank in some neighborhood of $(a^*(s), s)$ as well. Next since h is differentiable with respect to a, in a neighborhood of $(a^*(s), s)$:

$$\partial_a(\overline{g}, h)(a, s') = \partial_a \overline{g}(a, s') \times \{\nabla_a h(a, s')\}$$

Consider[5]:

$$y(a, s') = (I_n - \nabla_a h(a, s')^{PI} \nabla_a h(a, s')) \cdot \overline{y}$$

in which case,

$$\nabla_a h(a, s') y(a, s') = (\nabla_a h(a, s') - \nabla_a h(a, s')) \cdot \overline{y} = 0.$$

Next, since $\partial_a \overline{g}(a^*(s), s)$ is a closed set and GMFCQ holds at $(a^*(s), s)$, there exists $\varepsilon > 0$ such that:

$$\forall \gamma_a \in \partial_a \overline{g}(a^*(s), s),\ \gamma_a \cdot \overline{y} < -2\varepsilon.$$

Let $M = \sup\{\|\gamma_a\|, \gamma_a \in \partial_a \overline{g}(a^*(s), s)\}$. As $(a, s') \to (a^*(s), s)$, by construction $y(a, s') \to \overline{y}$ so there exists a neighborhood $A' \times S'$ of $(a^*(s), s)$ such that:

$$\|y(a, s') - \overline{y}\| < 2/M$$

Let $K = \sup\{\|y(a, s')\|, (a, s') \in A' \times S'\}$. Since $\partial_a \overline{g}$ is upper semicontinuous, there exists a neighborhood $A'' \times S''$ of $(a^*(s), s)$ such that:

$$\forall (a, s') \in A'' \times S'',\ \forall \gamma'_a \in \partial_a \overline{g}(a, s'),\ \exists \gamma_a \in \partial_a \overline{g}(a^*(s), s),\ \ \|\gamma_a - \gamma'_a\| \leq \varepsilon / K.$$

Thus, on this neighborhood, for every $\gamma'_a \in \partial_a \overline{g}(a, s')$:

$$\gamma'_a \cdot y(a, s') = (\gamma'_a - \gamma_a) \cdot y(a, s') + \gamma_a \cdot (y(a, s') - \overline{y}) + \gamma_a \cdot \overline{y} < \varepsilon + \varepsilon - 2\varepsilon = 0$$

which shows that $y(a, s')$ satisfies (ii) of the GMFCQ. ■

Lemma 12 *If the GMFCQ holds at $(a^*(s), s)$, then the correspondence $s \to \cup_{a^*(s) \in A^*(s)} K(a^*(s), s) \times \partial_s(f, \overline{g}, h)(a^*(s), s)$ is upper hemicontinuous at s.*

Proof. Consider a sequence $\{s_n\} \to s$, and a corresponding sequence given by $\{\alpha_n\} = \{(\lambda_n, \mu_n); (\varsigma_{s,n}, \gamma_{s,n}, \nabla_s h(a^*(s_n), s_n))\}$ each α_n in $K(a^*(s_n), s_n) \times \partial_s(f, \overline{g}, h)(a^*(s_n), s_n)$. By definition of the multipliers, for all n there exists $(\varsigma_{a,n}, \gamma_{a,n}) \in \partial_a(f, \overline{g})(a^*(s_n), s_n)$ such that:

$$\mu_n \cdot \nabla_a h(a^*(s_n), s_n) = \varsigma_{a,n} - \lambda_n \cdot \gamma_{a,n} \tag{7}$$

Since the subdifferential is an upper hemicontinuous correspondence, there must exists a subsequence of $\{(\varsigma_n, \gamma_n,)\}$ converging to some $(\varsigma, \gamma) \in \partial_s(f, \overline{g})(a^*(s), s)$. Without loss of

[5] M^{PI} is the pseudo-inverse of the full rank matrix M.

generality assume that $\{(\varsigma_n, \gamma_n,)\} \rightarrow (\varsigma, \gamma)$. Next, recall that A^* is an upper hemicontinuous correspondence by lemma 7 so that $a^*(s_n)$ converges to some element of $a^*(s) \in A^*(s)$, as $s_n \rightarrow s$. By the preceding lemma, the GMFCQ holds for $a^*(s_n)$ sufficiently close to s, i.e. for $n \geq N$. As in the proof of lemma 11 above, consider the sequence $\{\overline{y}_i\}$ defined as:

$$\overline{y}_i = \overline{y}(a^*(s_i), s_i) = (I_n - \nabla_a h(a^*(s_i), s_i)^{PI} \nabla_a h(a^*(s_i), s_i)) \cdot \overline{y}$$

so that $\lim_{i \rightarrow +\infty} \overline{y}_i = \overline{y}$, and $\overline{y}_i$ satisfies the GMFCQ at $(a^*(s_i), s_i)$ for i sufficiently large. That is:

$$\gamma_i \cdot \overline{y}_i < 0, \text{ and } \nabla_a h(a^*(s_i), s_i) \cdot \overline{y}_i = 0$$

and therefore, from (7):

$$\varsigma_{a,i} \cdot \overline{y}_i = \lambda_i \cdot \gamma_{a,i} \cdot \overline{y}_i \tag{8}$$

Since $(\varsigma_{a,i}, \gamma_{a,i}, \overline{y}_i) \rightarrow (\varsigma_a, \gamma_a, \overline{y})$, by 8 $\lambda_i \rightarrow \lambda$. Then by (7), $\mu_i \rightarrow \mu$ and $\mu \cdot \nabla_a h(a^*(s), s) = \varsigma_a - \lambda \cdot \gamma_a$, which proves that (λ, μ) is in $K(a^*(s), s)$. Having demonstrated the existence a subsequence of $\{\alpha_n\}$ converging to some $(\lambda, \mu, \varsigma_s, \gamma_s)$ in $K(a^*(s), s) \times \partial_s(f, \overline{g}, h)(a^*(s), s)$ for some $a^*(s) \in A^*(s)$ proves the desired upper hemicontinuity. ■

Theorem 13 *If D is uniformly compact near s, and if the GMFCQ holds at every $(a^*(s), s)$, then the value function V is locally Lipschitz near s.*

Proof. Since GMFCQ holds in a neighborhood of every $(a^*(s), s)$, and since by Lemma 7 above each maximum $a^*(s')$ is in a neighborhood of some $a^*(s) \in A^*(s)$ for s' sufficiently close to s, the Dini bounds established in the previous section also hold in a neighborhood of s. In particular, for all s' in a small compact neighborhood S' of s, recalling that $\partial_s(f - \lambda^T g - \mu^T h) \subset \partial_s f - \lambda \partial_s \overline{g} \mu - \partial_s \mu^T h$ we have:

$$\min_{(\varsigma,\gamma,\rho) \in \partial_s(f,\overline{g},h)(a^*(s'),s')} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x \leq \min_{\theta \in \partial_s(f - \lambda^T g - \mu h)(a^*(s'),s')} \theta \cdot x$$

and:

$$\max_{\partial_s(f - \lambda^T g - \mu h)(a^*(s'),s')} \theta \cdot x \leq \max_{(\varsigma,\gamma,\rho) \in \partial_s(f,\overline{g},h)(a^*(s'),s')} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x$$

Noting that at any $(a^*(s'), s')$ the set K is nonempty and bounded (see [25]), and using these inequalities in the Dini bounds of Theorem 26 gives:

$$\begin{aligned}
& \min_{a^*(s') \in A^*(s')} \min_{(\lambda,\mu) \in K(a^*(s'),s')} \min_{(\varsigma,\gamma,\rho) \in \partial_s(f,\overline{g},h)(a^*(s'),s')} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x \\
\leq\ & \inf_{(\lambda,\mu) \in K(a^*(s'),s')} \left(\min_{\theta \in \partial_s(f - \lambda^T g - \mu h)(a^*(s'),s')} \theta \cdot x \right) \\
\leq\ & D_+ V(s', x) \\
\leq\ & D^+ V(s', x) \\
\leq\ & \sup_{(\lambda,\mu) \in K(a^*(s'),s')} (\max_{\partial_s(f - \lambda^T g - \mu h)(a^*(s'),s')} \theta \cdot x) \\
\leq\ & \max_{a^*(s') \in A^*(s')} \max_{(\lambda,\mu) \in K(a^*(s'),s')} \max_{(\varsigma,\gamma,\rho) \in \partial_s(f,\overline{g},h)(a^*(s'),s')} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x
\end{aligned}$$

and therefore:

$$\begin{aligned}\min_{(\lambda,\mu,\varsigma,\gamma,\rho)\in\Pi_{S'}}(\varsigma-\lambda\gamma-\mu\rho)\cdot x &\leq D_+V(s';x)\\ &\leq D^+V(s';x)\leq \max_{(\lambda,\mu,\varsigma,\gamma,\rho)\in\Pi_{S'}}(\varsigma-\lambda\gamma-\mu\rho)\cdot x\end{aligned}$$

where $\Pi_{S'}$ is the set:

$$\Pi_{S'}=\bigcup_{s'\in S'}\left(\cup_{a^*(s)\in A^*(s)}K(a^*(s),s)\times\partial_s(f,\overline{g},h)(a^*(s),s)\right)$$

Note that the set $\Pi_{S'}$ is compact since $s\to F(s)=\cup_{a^*(s)\in A^*(s)}K(a^*(s),s)\times\partial_s(f,\overline{g},h)(a^*(s),s)$ is an upper hemicontinuous correspondence (as established in lemma 12 above), and so the min and max are attained on $\Pi_{S'}$. Consequently, there exists $\pi_1,\pi_2\in\Pi_{S'}$ such that:

$$\pi_1\cdot x\leq D_+V(s';x)\leq D^+V(s';x)\leq\pi_2.x$$

Consider any pair (s',s'') of points in the interior of S' and the direction $x=\frac{s''-s'}{\|s''-s'\|}$. The function: $t\in[0,\|s''-s'\|]\subset\mathbb{R}\to V(s'+tx)$ is locally Lipschitz on $[0,\|s''-s'\|]$ since:

$$\begin{aligned}\left|\lim_{t'\to t}\frac{V(s'+tx)-V(s'+t'x)}{t-t'}\right| &\leq \max(|D_+V(s'+tx,x)|,\left|D^+V(s'+tx,x)\right|)\\ &\leq \max(|\pi_1|,|\pi_2|)\,\|x\|=\max(|\pi_1|,|\pi_2|)\end{aligned}$$

and it is therefore differentiable almost everywhere, and also absolutely continuous on $[0,\|s''-s'\|]$ so that:

$$V(s'')-V(s^{'})=\int_0^{\|s''-s'\|}V'(s'+tx)dt$$

Therefore for s' and s'' in a neighborhood of s:

$$\left|V(s'')-V(s^{'})\right|\leq\max(|\pi_1|,|\pi_2|)\,\|s''-s'\|\,\|x\|=\max(|\pi_1|,|\pi_2|)\,\|s''-s'\|$$

which proves that V is locally Lipschitz in a neighborhood of s. ■

Corollary 14 *If D is uniformly compact near s and if the GMFCQ holds at every element of $A^*(s)$, then:*

$$\partial V(s)\subset co\left\{\bigcup_{a^*(s)\in A^*(s)}\ \bigcup_{(\lambda,\mu)\in K^*(a^*(s),s)}\partial_s(f-\lambda^Tg-\mu^Th)(a^*(s),s)\right\}$$

Proof. Since V is locally Lipschitz near s, its subdifferential at s is given by:

$$\partial V(s)=co\left\{\lim\nabla V(s_n):s_n\to s,s_n\in dom\nabla V\right\}$$

Consider the sequence $\{s_n\}$ converging to s and such that $\nabla V(s_n) \to \varphi$ (elements of $\partial V(s)$ are convex combinations of φ). For s_n sufficiently close to s, the Dini bounds hold and thus for n large enough (say $n \geq N$) and for any direction x :

$$\begin{aligned}\nabla V(s_n) \cdot x &= D^+ V(s_n; x) \\ &\leq \max_{a^*(s_n) \in A^*(s_n)} \max_{(\lambda,\mu) \in K(a^*(s_n), s_n)} \max_{(\varsigma,\gamma,\rho) \in \partial_s (f,\overline{g},h)(a^*(s_n),s_n)} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x \\ &\leq \max_{(\lambda,\mu,\varsigma,\gamma,\rho) \in \Pi_{s_n}} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x\end{aligned}$$

where Π_{s_n} is the compact set $\cup_{a^*(s_n) \in A^*(s_n)} K(a^*(s_n), s_n) \times \partial_s (f, \overline{g}, h)(a^*(s_n), s_n)$. For each s_n, $n \geq N$, the maximum is attained at some $(\lambda_n, \mu_n, \varsigma_n, \gamma_n, \rho_n) \in \Pi_{s_n}$. By lemma 12 the correspondence $s_n \to \Pi_{s_n}$ is upper hemicontinuous, and therefore the sequence of maxima $\{(\lambda_n, \mu_n, \varsigma_n, \gamma_n, \rho_n)\}_{n=N}^{\infty}$ has a subsequence converging to some $(\lambda, \mu, \varsigma, \gamma, \rho) \in \Pi_s = \cup_{a^*(s) \in A^*(s)} K(a^*(s), s) \times \partial_s (f, \overline{g}, h)(a^*(s), s)$. Taking limits along this subsequence in the above inequalities, we have that for any φ:

$$\varphi \cdot x \leq \max_{(\lambda,\mu,\varsigma,\gamma,\rho) \in \Pi_s} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x$$

Consequently:

$$\max_{\varphi \in \partial V(s)} \varphi \cdot x \leq \max_{(\lambda,\mu,\varsigma,\gamma,\rho) \in \Pi_s} (\varsigma - \lambda^T \gamma - \mu^T \rho) \cdot x = \max_{a^*(s) \in A^*(s)} \max_{(\lambda,\mu) \in K(a^*(s),s)} \max_{\theta_s \in \partial_s (f - \lambda^T g - \mu h)(a^*(s),s)} \theta_s \cdot x$$

Since $\partial V(s)$ is formed by the convex combinations of all the φ, necessarily:

$$\partial V(s) \subset \bigcup_{a^*(s) \in A^*(s)} \bigcup_{(\lambda,\mu) \in K^*(a^*(s),s)} \partial_s (f - \lambda^T g - \mu^T h)(a^*(s), s).$$

■

5 APPLICATIONS AND EXTENSIONS

5.1 Lipschitz Dynamic Programming

Consider the N periods Lipschitz program:

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

subject to $g(a,s) \leq 0$, for all $n = 1, 2, ..N$, $a \in A \subset \mathbb{R}^n$ and $s \in S \subset \mathbb{R}^m$, and $V_0 = 0$. The associated Lagrangian takes the form:

$$L(a,s) = F(a,s) + \beta V_n(a) - \lambda^T g(a,s)$$

so that

$$\partial_s L(a,s) = \partial_s (F(a,s) - \lambda^T g(a,s)) \subset \partial_s F(a,s) - \lambda^T \partial_s g(a,s)$$

with equality if either the objective F or the constraint g is strictly differentiable in s (or both). Clearly, V_n is not necessarily concave or/and C^1, but we will provide sufficient conditions under which it is locally Lipschitz under the following assumption:

Assumption 5.1: F, and g are locally Lipschitz, and the feasible correspondence is uniformly compact for all $s \in S$.

Proposition 15 *Under assumption 5.1 if GMFCQ is satisfied for every optimal solution $a^*(s) \in A^*(s)$, then each V_n, $n = 1, ..N$ is locally Lipschitz with Clarke gradient:*

$$\partial V_n(s) \subset co \left\{ \bigcup_{a_n^*(s) \in A_n^*(s)} \bigcup_{\lambda \in K^*(a_n^*(s),s)} \partial_s (f - \lambda^T g)(a_n^*(s), s) \right\}$$

Proof. Follow directly from Theorem 13 and Theorem 8. ■

Since each V_i is locally Lipschitz, the generalized multiplier rule for each period is given by:

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

and if $a^*(s)$ is interior, then:

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

which simplifies to:

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

whenever F is C^1 in a.

Taking $N \to \infty$, a consequence of Proposition 15 above is that the sequence of functions $\{V_n\}$ (which converges uniformly to the unique function V solving Bellman's equation through the contraction mapping theorem) is in fact a sequence of locally Lipschitz functions. Unfortunately, uniform limits of sequences of locally Lipschitz functions are not necessarily locally Lipschitz.[6] However, additional properties of V may be obtained by putting more restrictions on the primitive data, as shown in the next section.

5.1.1 Concave Dynamic Programming

Consider the infinite horizon version of the previous program for which we make the following assumptions, typically satisfied by a large class of dynamic programs:

Assumption 5.1.1: F, and g are jointly concave and C^1 in (a, s). The feasible correspondence is uniformly compact for all $s \in S$.

This assumption of concavity along with the right constraint qualification guarantees the existence of directionally differentiable envelope, as given by the next corollary.

Proposition 16 *Under assumption 5.1.1 (i) if GMFCQ is satisfied for every optimal solution $a^*(s) \in A^*(s)$, then V is concave and locally Lipschitz, with bounds of the directional derivatives given by,*

$$V'(s,x) = \max_{a^*(s) \in A^*(s)} \min_{\lambda \in K(a^*(s),s)} \left(F_s(a^*(s), s) - \lambda^T g_s(a^*(s), s) \right) \cdot x$$

and (ii) if SMFCQ is satisfied for every optimal solution $a_n^*(s) \in A_n^*(s)$, then V is concave and C^1 with derivative given by

$$V'(s) = \max_{a^*(s) \in A^*(s)} \left(F_s(a^*(s), s) - \lambda^T g_s(a^*(s), s) \right) \cdot x$$

[6] Indeed by the Weistrass Approximation Theorem, any continuous functions on $[0, 1]$ (including non Lipschitz continuous functions) may be uniformely approximated by polynomials (which are Lipschitz).

Proof. A recursive applications of Bellman's operator generate a sequence $\{V_n\}$, of concave functions converging uniformly to a concave function V. Concavity of V implies almost everywhere differentiability. At points of nondifferentiablity V is at least directionally differentiability 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 10 establishes that:

$$V'(s,x) = \max_{a^*(s)\in A^*(s)} \min_{\lambda \in K(a^*(s),s)} \big(F_s(a^*(s),s) - \lambda^T g_s(a^*(s),s)\big) \cdot x$$

Next, recalling that SMFCQ implies uniqueness of the multipliers (when primitives are C^1, see Corollary 9) so that:

$$V'(s,x) = \max_{a^*(s)\in A^*(s)} \big(F_s(a^*(s),s) - \lambda^T g_s(a^*(s),s)\big) \cdot x$$

Moreover, note,

$$\begin{aligned}
-V'(s,-x) &= -\max_{a^*(s)\in A^*(s)} \big(F_s(a^*(s),s) - \lambda^T g_s(a^*(s),s)\big) \cdot (-x) \\
&= \min_{a^*(s)\in A^*(s)} \big(F_s(a^*(s),s) - \lambda^T g_s(a^*(s),s)\big) \cdot x \\
&\leq \max_{a^*(s)\in A^*(s)} \big(F_s(a^*(s),s) - \lambda^T g_s(a^*(s),s)\big) \cdot x \\
&= V'(s,x)
\end{aligned}$$

for all directions $x \in \mathbb{R}^m$. The concavity of V implies that

$$-V'(s,-x) \geq V'(s,x),$$

and therefore, for all $x \in \mathbb{R}^m$,

$$-V'(s,-x) = V'(s,x)$$

Given the lower Clarke regularity of V, this precisely means that:

$$\begin{aligned}
&-\min_{z\in\partial V_{n+1}(s)}\{z\cdot(-x)\} = \min_{z\in\partial V_{n+1}(s)}\{z\cdot x\} \\
\Longleftrightarrow\quad &\max_{z\in\partial V_{n+1}(s)}\{z\cdot x\} = \min_{z\in\partial V_{n+1}(s)}\{z\cdot x\}
\end{aligned}$$

and thus, V is strictly differentiable. Strict differentiability imply C^1, by Rockafellar and Wets [28] Corollary 9.19. ■

In the following section we provide an example of concave Dynamic Programming where we apply the above proposition.

5.1.2 Application: Differentiability of the Pareto Frontier

Consider the model of Kocherlakota for an exchange economy in which two infinitely lived agents receive a stochastic endowment in each period t. Since endowment is stochastic, the two agents mutually share their endowment under limited commitment. The endowment for agent i in period t is (ω_s^1, ω_s^2) which is determined by the realization of θ_t. The sequence of independently and identically distributed random variable $\theta = \{\theta_1, \theta_2, ...\}$ have finite support $\Theta = \{1, 2, ..., S\}$. The probability that θ_t equals s is denoted by π_s for all s in Θ. We will assume the following:

Assumption 5.1.2: The utility function $u : \mathbb{R}_+ \to \mathbb{R}$ *is increasing, concave and* C^1*. Further* $\lim_{c \to 0^+} u'(c) = \infty$*.* $0 < \beta < 1$*. And for each* U_0*, the feasible set is uniformly compact.*

Note we have relaxed the assumptions in Koeppl of strictly concave and strictly increasing and C^2 utility function. We characterize the incentive feasible allocations (see Koeppl for details).

$$V(U_0) = \max_{c_s, u_s} \sum_{s=1}^{S} \pi_s \left[u(\overline{\omega}_s - c_s) + \beta V(U_s) \right] \tag{9}$$

subject to

$$\begin{aligned}
&U_0 - \sum_{s=1}^{S} \pi_s \left[u(c_s) + \beta U_s \right] \leq 0 \\
&u(\omega_s^1) + \beta U_{aut} - u(c_s) - \beta U_s \leq 0 \\
&u(\overline{\omega}_s - \omega_s^1) + \beta U_{aut} - u(\overline{\omega}_s - c_s) - \beta V(U_s) \leq 0 \\
&U_s \in [U_{aut}, U_{\max}]
\end{aligned}$$

Here, the set of optimal solution is denoted by $Y^*(s)$, and a typical element of this set is (c_s^*, U_s^*), and the KKT multipliers are denoted as $(\lambda_1, \lambda_{2s}, \lambda_{3s}, \lambda_{4s}, \lambda_{5s}) \in K(c_s^*, U_s^*)$. Note, from assumption *5.1.2*, and separability structure of programming problem the objective and the constraints are jointly concave in (c_s, u_s, U_0). Thus, the assumptions of the last section hold here. As a direct consequence of Proposition 16, under assumption *5.1.2* 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[7]

$$\nabla V(U_0) = \lambda_1$$

Moreover here we also show that at points of nondifferentiability, the limiting value function is directionally differentiable. Under assumption *5.1.2,* if GMFCQ is satisfied for every optimal solution $(c^*(U_0), U^*(U_0)) \in Y^*(U_0)$, then from Proposition 16, V is concave with directional derivatives given by,

$$V'(U_0; x) = \max_{(\lambda_1, \lambda_{2s}, \lambda_{3s}) \in (c_s^*, U_s^*)} \lambda_1 \cdot x$$

[7] This result is also discussed by Zapatero and Santos [26].

5.2 Optimization problems with discrete choice variables

Consider a Lipschitz program in which the decision variable a_1 may only take one of r possible values. That is consider the program:

$$\max_{a \in D(s)} f(a,s) \tag{10}$$

where $D(s) = \{g(a,s) \leq 0, \text{ and } a_1 = b_j \; j = 1,...,r\}$. We simply rewrite the r equality constraints $a_1 = b_j$ as one unique constraint $h(a,s) = \prod_{j=1}^{r}(a_1 - b_j) = 0$. Clearly $h : A \times S \to \mathbb{R}$ is C^1, with gradient given by $\nabla_a h(a,s) = h_1(a) = (\frac{\partial h}{\partial a_1}, 0, ..., 0)$. At any optimum $a^*(s) \in A^*(s)$, $a_1^*(s)$ must equal some b_k, so that $\nabla_a h(a^*(s), s) = (\prod_{j \neq k}(b_k - b_j), 0, ..., 0)$ is always distinct from 0.

The Lagrangian associated with this standard maximization problem is given by:

$$L(a, \lambda, \mu; s) = f(a,s) - \lambda^T \cdot g(a,s) - \mu h(a,s)$$

and $a^*(s) = (b_k, a_2^*(s), ..., a_n^*(s)) \in A^*(s)$ is a KKT point if there exists a vector $\lambda \geq 0$ of multipliers and $\mu \in \mathbb{R}$ such that:

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

where $I(a^*(s), s)$ is the set of identifying the $\overline{p}$ active inequality constraints (those for which $g_i(a^*(s), s) = 0$). In this problem, GMFCQ is defined as follows.

Definition 17 *A feasible point $a \in D(s)$ satisfies the Generalized Mangasarian-Fromovitz Constraint Qualifier (GMFCQ) if there exists a $\widetilde{y} \in \mathbb{R}^n$ such that,*

$$\exists \widetilde{y} \in \mathbb{R}^n, \; \forall \gamma_a \in \partial_a \overline{g}(a,s), \; \gamma \cdot \widetilde{y} < 0, \; \textit{and } h_1(a,s)\widetilde{y} = 0.$$

Note that necessarily $\widetilde{y}_1 = 0$ for $\widetilde{y}$ to meet this condition. It is then a straightforward application of Theorem 8 for Lipschitz programs with C^1 equality constraints (specifically, Theorem 26 in Appendix C) to claim the following:

Proposition 18 *If $D(s)$ is nonempty-valued and uniformly compact near s, and if the GM-FCQ holds for every optimal solution $a^*(s) \in A^*(s)$, then for any direction of perturbation $x \in \mathbb{R}^m$:*

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

and:

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

Also, V is locally Lipschitz and the Clarke gradient is given by:

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

5.2.1 Application: Labor Leisure Choice

Consider a finite horizon labor-leisure choice problem where labor takes only the binary values $\{0,1\}$. Thus we formulate a N period problem as,

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

subject to

$$\begin{aligned} c_n + k_{n+1} - f(k_n, l_n) &\leq 0 \\ -c_n &\leq 0 \\ -k_{n+1} &\leq 0 \\ l_n(1-l_n) &= 0 \end{aligned}$$

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

Assumption 5.2.1: The utility function $u : \mathbb{R}^n_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ and the production function $f : \mathbb{R}^n_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are locally Lipschitz and increasing in both arguments. The feasible set $D_n : \mathbb{R}^n_+ \rightrightarrows \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}_+$ is nonempty-valued and uniformly compact near for all k_t.

Since the utility function is increasing in the first argument for each period n, the first inequality always holds with an equality. Therefore we substitute c_n in the objective with the first constraint. The per period Lagrangean is given by;

$$\begin{aligned} L_n(k_n) &= u(f(k_n, l_n) - k_{n+1}, 1-l_n) + \beta V_{n+1}(k_{n+1}) + \lambda^1_n(f(k_n, l_n) - k_{n+1}) \\ &\quad + \lambda^2_n k_{n+1} + \mu_n(l_n(1-l_n)) \end{aligned}$$

and a typical element of the optimal choice correspondence is denoted by $y^*_N(k_N) \in Y^*_N(k_N)$. Now by recursively applying Proposition 18, we get the following result.

Under assumption 5.2.1, if the GMFCQ holds for every optimal solution $y^*_n(k_n) \in Y^{;*}_n(k_n)$, then for any direction of perturbation $x \in \mathbb{R}^n$:

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

and:

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

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

5.3 Some Examples

The economic literature is depleted with examples with multi-period optimization problems. Typically in these examples the value function of the final period (for finite period problems) enter the optimization problem of the previous period so on and so forth. The value function of period t may enter optimization problem of the Period $t-1$, either in the objective or the constraints or both. To make calculations easy what is done in practice is to assume enough structure such that the value function of each period is C^1, implying the optimization problem in each period is in the class of C^1 and/or concave functions. However, as we show in this paper concavity and C^1 differentiability is not necessary to assume, albeit it makes calculations easier but at the cost of strong assumptions. Next, we solve two multi period optimization problem where the objective need not be C^1, and thus we use generalized derivatives to derive the optimal solutions.

5.3.1 Redistributive Taxes

Consider an endowment economy with a benevolent social planner who collects taxes to redistribute it; and N agents each endowed with wealth W^i that chooses one unit of labor between leisure and labor supply to maximize their utility. We consider a sequential problem: In the first period the planner chooses an optimal tax levied an wage income to maximize his social welfare function. In the second period given the tax rate, the redistributed income, wages and prices of the final good each agent maximizes their utility $u^i : \mathbb{R}_+ \times [0,1] \to \mathbb{R}_+$. Agent i gets $\alpha^i \in [0,1]$, with $\sum_{i=1}^N \alpha^i = 1$, proportion of total tax collected. We assume the following:

Assumption 5.3.1: The utility function u^i for each agent i, is locally Lipschitz, and nondecreasing in all arguments.

The first period problem is:

$$U^i(\tau, W^i, w, \alpha^i) = \max_{l^i, c^i} u^i(c^i, 1 - l^i)$$

subject to

$$\begin{aligned} c^i - w(1-\tau)l^i - \alpha^i w\tau \sum_{i=1}^N l^i - W^i &\leq 0 \\ -l^i &\leq 0 \\ l^i - 1 &\leq 0 \end{aligned}$$

From, Theorem 13, under assumption 5.3.1 if the feasible set for each agent i, $D^i(\tau, W^i, w, \alpha^i)$ is uniformly compact and GMFCQ hold for all $(c^{i*}, l^{i*})(\tau, W^i, w, \alpha^i) \in Y^*(\tau, W^i, w, \alpha^i)$ then the indirect utility function for each agent i, U^i is locally Lipschitz in (τ, W^i, w, α^i).Thus, recalling Corollary 14, the Clarke derivative of U with respect to τ, α^i are given by

$$\begin{aligned} \partial_\tau U(\tau, G, w) &\subset co\{ \bigcup_{(c^{i*}, l^{i*}) \in Y^*(\tau, W^i, w, \alpha^i)} \bigcup_{(\lambda_1^1, \lambda_2^1, \lambda_2^3) \in K} \{\partial_c u(c^{i*}, 1 - l^{i*}) - \lambda_1^1 (w l^{i*} - \alpha^i w \sum_{i=1}^N l^{i*})\}\} \\ \partial_{\alpha^i} U(\tau, G, w) &\subset co\{ \bigcup_{(c^{i*}, l^{i*}) \in Y^*(\tau, W^i, w, \alpha^i)} \bigcup_{(\lambda_1^1, \lambda_2^1, \lambda_2^3) \in K} \{\partial_c u(c^{i*}, 1 - l^{i*}) - \lambda_1^1 (w\tau \sum_{i=1}^N l^{i*})\}\} \end{aligned}$$

where the KKT multipliers of the first period constraints $(\lambda_1^1, \lambda_2^1, \lambda_2^3) \in K(c^{i*}, l^{i*}(\tau, W^i, w, \alpha^i))$.

Now the first period problem of the social planner takes the form:

$$v(w) = \max_{\tau, \alpha^i} \sum_{i=1}^N \beta^i U^i(\tau, W^i, w, \alpha^i)$$

subject to

$$\begin{aligned}
-\tau &\leq 0 \\
-\alpha^i &\leq 0 \\
\sum_{i=1}^{N} \alpha^i &= 1
\end{aligned}$$

where $\sum_{i=1}^{N} \beta^i = 1$.

The objective of the first period is not necessarily C^1, but locally Lipschitz and the constraints are C^1, thus, the generalized first order conditions are given by

$$\begin{aligned}
0 &\in \partial_\tau \left(U(\tau, W^i, w, \alpha^i) + \lambda_1^2 \right) \\
0 &\in \partial_{\alpha^i} \left(U(\tau, W^i, w, \alpha^i) + \lambda_2^2 - \mu^2 \right)
\end{aligned}$$

where the KKT multipliers of the second period constraints $(\lambda_1^2, \lambda_2^2, \mu^2) \in K(\tau^*, \alpha^{i*}(w))$. Any optimal tax rate satisfies the above two inclusion relation.

5.3.2 Research and Development

We consider a two stage Monopolist problem: in the first stage an R and D investment level that might reduces cost in chosen, and in the second stage the firm enters the product market. A I amount of R and D investment reduces constant marginal cost from $\bar{c}$ to $\underline{c}$ with probability $z(I)$. The inverse demand function $q : \mathbb{R}_+ \to \mathbb{R}_+$, and the probability function $z : \mathbb{R}_+ \to [0, 1]$ are both locally Lipschitz.

In the second period the monopolist is faced with the following optimization problem:

$$\Pi^1(I) = \max_p \left\{ q(p_)(p - \underline{c}z(I) \right\}$$

subject to

$$\begin{aligned}
-q &\leq 0 \\
-p &\leq 0
\end{aligned}$$

Let the set of optimal solution be denoted by $P^*(I)$ with typical element $p^*(I)$, and the set of multipliers by $\lambda = (\lambda_1^1, \lambda_2^1) \in K(p^*(I), P^*(I))$. From Theorem 13, $\Pi^1(I)$, is locally Lipschitz with Clarke gradient

$$\begin{aligned}
\partial \Pi^1(I) &\subset co \left\{ \bigcup_{p^*(I) \in P^*(I)} \bigcup_{\lambda \in K(p^*(I), P^*(I))} \{ \partial_I q(p^*) \underline{c} z(I) \} \right\} \\
&= co \left\{ \bigcup_{p^*(I) \in P^*(I)} \{ \partial_I q(p^*) \underline{c} z(I) \} \right\}
\end{aligned}$$

Thus, in the first period the monopolist encounters a Lipschitz optimization problem with C^1 constraints given by:

$$\max_I \left\{ \Pi^1(I) - I \right\}$$

subject to

$$-I \leq 0$$

Consequently the optimal investment satisfies the first order inclusion condition given by

$$0 \in \partial \Pi^1(I) - 1 + \lambda^2.$$

6 APPENDIX A: Mathematical Tools

6.1 Derivatives and subgradients

6.1.1 Lipschitz functions

A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is *locally Lipschitz at x with modulus $k(x)$*, $0 \leq k < \infty$, if there exists $\delta > 0$ such that for all x', x'' in $B(x, \delta)$:

$$|f(x'') - f(x')| \leq k(x) \left| x'' - x' \right|.$$

When $k(x)$ may be chosen independently of x, f is said to be *Lipschitz.* By Rademacher's theorem, a Lipschitz function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is almost everywhere differentiable. Recall that the *directional derivative at x_0* is the function:

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

and, when this quantity exists for all d, we say that f is *directionally differentiable* (or Gateaux differentiable)[8] at x_0. f is *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) and *continuously differentiable at x_0* if the function $\nabla f(.) : \mathbf{\Omega} \to \mathbb{R}^{n \times m}$ is continuous at x_0. f is *strictly differentiable* at x if there exists a continuous linear (in d) function $Df(x)$ such that:

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

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

6.1.2 Dini and Clarke derivatives

Directional derivatives of locally Lipschitz functions do not necessarily exist. However, the *upper (right hand) Dini derivative* defined as the function*:*

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

[8] For locally Lipschitz functions in finite dimensional spaces, the notion of Gateaux differentiability and Frechet differentiablility coincide, and $f'(x_0; d)$ is homogenous of degree one in d.

and the *lower (right hand) Dini derivative* defined as the function:

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

of locally Lipschitz functions always exist (i.e. are finite quantities). Upper and lower Dini derivatrives coincide at points x where the directional derivative exists

The *Clarke upper and lower (directional) derivatives* at x_0 are, respectively, the functions:

$$\begin{aligned} d &\longmapsto f^o(x;d) = \lim\sup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td)-f(y)}{t} \\ d &\longmapsto f^{-o}(x;d) = \lim\inf_{\substack{y\to x\\ t\to 0^-}} \frac{f(y+td)-f(y)}{t} \end{aligned}$$

Clarke derivatives of Lipschitz functions always exist, and if f is locally Lipschitz and directionally differentiable at x,

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

6.1.3 Clarke gradient

The Clarke generalized gradient (sometimes called "subdifferential") of a Lipschitz function f at x is the nonempty compact convex set defined as:

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

where $\overline{co}$ denotes the convex hull[9], Θ is any set of Lebesgue measure zero in the domain, and Ω_f is a set of points at which f fails to be differentiable. Clarke (Proposition 2.1.5) shows that $x \rightrightarrows \partial f(.)$ is an upper hemicontinuous correspondence. Elements of the subdifferential are called subgradients, and for a convex function f, the subdifferential at x is the set of $p \in M_{m\times n}$ satisfying:

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

for all directions $d \in \mathbb{R}^n$. Clarke [6] (Proposition 1.4) 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 (since it is the supremum of a family of linear functions) (equivalently, $f^{-o}(x;d)$ is concave in d). This also implies that:

$$-f^o(x;-d) = -\max_{\zeta\in\partial f(x)}\{\zeta.d\} = \min_{\zeta\in\partial f(x)}\{\zeta.d\} = f^{-o}(x;d) \leq f^o(x;d)$$

[9]In the formula, either co or $\overline{co}$ will do since we work in finite dimensional spaces.

and, equivalently, that:

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

We also note that if the Clarke derivatives of a locally Lipschitz function coincide at x if then f is strictly differentiable at x since

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

is a concave and convex (thus linear) function of d, and this expression is by construction continuous in x. Clearly, the Clarke derivatives of a strictly differentiable function coincide.

A function f is said to be *upper* (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 $f^{o}(x;d) = f'(x;d)$. (resp. $f^{-o}(x;d) = f'(x;d)$). We note that the sum of two Clarke regular functions is itself a Clarke regular function (since $\partial(f+g) \subset \partial f + \partial g$ so $f' + g' \leq (f+g)^0 \leq f^0 + g^0 = f' + g'$).

6.2 Properties of Correspondences

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

Definition 19 *A non-empty valued correspondence $D : S \twoheadrightarrow A$ is:*

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

(ii) upper hemicontinuous at s if for every sequence $s_n \to s$ 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 $s_n \to s$, $a_n \in D(s_n)$ and $a_n \to a$ 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 $s_n \to s$ and any $a \in D(s)$, there exists a sequence $\{a_n\}$ and a number N such that $a_n \to a$ and $a_n \in D(s_n)$ for all $n \geq N$.

Note that if the feasible domain is $D(s) = \{a \in A,\ g_i(a,s) \leq 0,\ i = 1, ..., p\}$, in which the g_i are locally Lipschitz (hence jointly continuous), then D is closed at any $s \in S$. The same property holds true in the presence of locally Lipschitz equality constraints. Another property of correspondence which will be critical in our analysis is that of uniform compactness.

Definition 20 *A non-empty valued correspondence D is said to be uniformly compact near s if there exists a neighborhood S' of s such that $cl\,[\cup_{s' \in S'} D(s)]$ is compact.*

We note the result in Hogan [18] 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. In particular, closed and uniformly compact imply upper hemicontinuous. Finally, we will need the following property of hemicontinuous correspondences (a property that hence applies to subdifferentials).

Proposition 21 *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. ■

7 APPENDIX B

Proof of theorem 5

Proof. Notice that the maximization domains in the definition of $\mathbb{S}^L$ are generalized gradients, which we know are non-empty and compact. Given any x, define G as follows:

$$\begin{aligned} G &= \{y \in \mathbb{R}^n,\ \forall(\varsigma_a,\gamma_a) \in \partial_a(f,\overline{g})(a^*(s),s) \text{ and } \forall(\varsigma_s,\gamma_s) \in \partial_s(f,\overline{g})(a^*(s),s) \\ \gamma_a \cdot y + \gamma_s \cdot x &\leq 0\} \end{aligned}$$

By GMFCQ, G is non-empty, and both G and $K = K(a^*(s),s)$ are closed convex sets. Note that if $\lambda \in K$ then:

$$\forall y \in \mathbb{R}^n,\ \min_{(\varsigma_a,\gamma_a)} \left(\varsigma_a - \lambda^T\gamma_a\right)\cdot y = \min_{\theta\in\partial_a(f-\lambda^T g)(a^*(s),s)} \theta\cdot y \leq 0$$

(or else there exists some y such that $\min_{(\varsigma_a,\gamma_a)}(\varsigma_a - \lambda\gamma_a)\cdot y > 0$ which would contradict $0 \in \partial_a(f - \sum_{i=1}^p \lambda_i g_i)(a^*(s),s))$. Consequently:

$$\sup_y \mathbb{S}^L(y,\lambda) = \left\{ \begin{array}{c} \min_{(\varsigma_s,\gamma_s)}(\varsigma_s - \lambda^T\gamma_s)\cdot x \text{ if } \lambda \in K \\ +\infty \end{array} \right\}$$

which implies that:

$$\inf_{\lambda\geq 0}\sup_y \mathbb{S}^L(y,\lambda)) = \inf_{\lambda\in K}\sup_y \mathbb{S}^L(y,\lambda)) < +\infty \tag{B1}$$

Next, we also have:

$$\begin{aligned} \inf_{\lambda\geq 0} \mathbb{S}^L(y,\lambda) &= \inf_{\lambda\geq 0}\left(\min_{(\varsigma_a,\gamma_a)}(\varsigma_a - \lambda^T\gamma_a)\cdot y + \min_{(\varsigma_s,\gamma_s)}(\varsigma_s - \lambda^T\gamma_s)\cdot x\right) \\ &= \min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))}\left((\varsigma_a\cdot y + \varsigma_s\cdot x) + \inf_{\lambda\geq 0}(-\lambda^T(\gamma_a\cdot y + \gamma_s\cdot x))\right) \end{aligned}$$

If $y \notin G$, then $\gamma_a \cdot y + \gamma_s \cdot x > 0$ and thus:

$$\inf_{\lambda \geq 0} -\lambda^T \left(\gamma_a \cdot y + \gamma_s \cdot x\right) = -\infty$$

so that:

$$\inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) = \left\{ \begin{array}{l} \min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))}(\varsigma_a \cdot y + \varsigma_s \cdot x) \;\; (\text{if } y \in G) \\ -\infty \text{ otherwise} \end{array} \right\}$$

and therefore:

$$\sup_y \inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) > -\infty \tag{B2}$$

Naturally for all $\lambda \geq 0$:

$$\inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) \leq \sup_y \mathbb{S}^L(y, \lambda)$$

and, therefore (using the results B1 and B2 above):

$$-\infty < \sup_y \inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) \leq \inf_{\lambda \geq 0} \sup_y \mathbb{S}^L(y, \lambda)) = \inf_{\lambda \in K} \sup_y \mathbb{S}^L(y, \lambda) < +\infty$$

Our purpose is to show that this last weak inequality is in fact an equality, and this will prove that $\mathbb{S}^L$ has a saddle value.

From the inequalities above, we have:

$$\begin{aligned} -\infty < & \sup_y \left(\min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))} (\varsigma_a \cdot y + \varsigma_s \cdot x) \right) \\ \leq & \inf_{\lambda \in K} \left(\min_{(\varsigma_s,\gamma_s)} (\varsigma_s - \lambda^T \gamma_s) \cdot x \right) < +\infty \end{aligned}$$

Next, for any $((\varsigma_a', \gamma_a'), (\varsigma_s', \gamma_s')) \in \partial_a(f, \overline{g})(a^*(s), s) \times \partial_s(f, \overline{g})(a^*(s), s)$:

$$\begin{aligned} & \inf_{\lambda \geq 0} \sup_y \left[\begin{array}{c} \left(\varsigma_a' - \lambda^T \gamma_a'\right) \cdot y \\ + \left(\varsigma_s' - \lambda^T \gamma_s'\right) \cdot x \end{array} \right] \\ \geq & \min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))} \inf_{\lambda \geq 0} \sup_y \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \gamma_a\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \gamma_s\right) \cdot x \end{array} \right] \\ = & \inf_{\lambda \geq 0} \min_{(\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s)} \sup_y \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \gamma_a\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \gamma_s\right) \cdot x \end{array} \right] \\ \geq & \inf_{\lambda \geq 0} \left(\sup_y \mathbb{S}^L(y, \lambda) \right) \\ \geq & \sup_y \left(\inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) \right) \\ = & \sup_y \min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))} \inf_{\lambda \geq 0} \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \gamma_a\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \gamma_s\right) \cdot x \end{array} \right] \\ = & \min_{((\varsigma_a,\gamma_a),(\varsigma_s,\gamma_s))} \sup_y \inf_{\lambda \geq 0} \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \gamma_a\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \gamma_s\right) \cdot x \end{array} \right] \end{aligned} \tag{B3}$$

Note that the last equality in this sequence is due to the existence of a saddle point for the saddle function $\overline{s}$:

$$\overline{s}(\lambda; (\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s)) = \inf_{\lambda \geq 0} \begin{bmatrix} \left(\varsigma_a - \lambda^T \gamma_a\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \gamma_s\right) \cdot x \end{bmatrix}$$

Indeed, since $\overline{s}$ is convex in y, concave in $((\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s)) \in \partial_a(f, \overline{g}) \times \partial_s(f, \overline{g})$ (which is bounded), then $\overline{s}$ has a (local) saddle value. Therefore,

$$\min_{(\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s)} \sup_y \overline{s} = \sup_y \min_{(\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s)} \overline{s}$$

Denoting $(\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s)$ as one of the particular subgradients for which expression (B3) above is attained, we can summarize the above sequence of inequalities as follows:

$$\begin{aligned} & \inf_{\lambda \geq 0} \sup_y \begin{bmatrix} (\varsigma_a - \lambda \gamma_a) \cdot y \\ + (\varsigma_s - \lambda \gamma_s) \cdot x \end{bmatrix} \\ \geq \quad & \sup_y \inf_{\lambda \geq 0} \begin{bmatrix} (\varsigma_a - \lambda \gamma_a) \cdot y \\ + (\varsigma_s - \lambda \gamma_s) \cdot x \end{bmatrix} \end{aligned} \tag{B4}$$

Now, consider the following linear program:

$$\varsigma_s \cdot x - \max (\lambda \gamma_s) \cdot x \tag{P}$$

subject to:

$$\lambda \geq 0 \text{ and } \varsigma_a - \lambda \gamma_a = 0$$

The dual to (P) is simply:

$$\varsigma_s \cdot x - \min \varsigma_a \cdot y$$

subject to:

$$\gamma_a \cdot y + \gamma_s \cdot x \leq 0 \text{ and } y \text{ unrestricted}$$

Clearly, any $y \in G$ is feasible for the dual, and since B1 above established that:

$$+\infty > \inf_{\lambda \in K} \big(\min_{(\varsigma_s, \gamma_s)} (\varsigma_s - \lambda^T \gamma_s) \cdot x \big) = \inf_{\lambda \geq 0} \sup_y \mathbb{S}^L(y, \lambda)$$

program P is feasible and there is no duality gap. The zero duality gap condition is precisely (B4) with an equality. Referring to the sequence of inequality above, we see that this implies that $\inf_{\lambda \geq 0} \sup_y \mathbb{S}^L(y, \lambda)$ and $\sup_y \inf_{\lambda \geq 0} \mathbb{S}^L(y, \lambda)$ must coincide. That is, $\mathbb{S}^L$ has a saddle value, and we have:

$$\begin{array}{rcl} \sup\limits_y \inf\limits_{\lambda \geq 0} \mathbb{S}^L(y, \lambda) & = & \sup\limits_{y \in G} \left(\min\limits_{((\varsigma_a, \gamma_a), (\varsigma_s, \gamma_s))} (\varsigma_a \cdot y + \varsigma_s \cdot x) \right) \\ & = & \\ \inf\limits_{\lambda \in K} \sup\limits_y \mathbb{S}^L(y, \lambda) & = & \inf\limits_{\lambda \in K} \left(\min\limits_{(\varsigma_s, \gamma_s)} (\varsigma_s - \lambda^T \gamma_s) \cdot x \right) = \inf\limits_{\lambda \in K} L_s^o(a^*(s), s; \lambda; x) \end{array}$$

in which:

$$L_s^o(a^*(s), s; \lambda; x) = \min_{\theta \in \partial_s (f - \lambda^T g)(a^*(s), s)} \theta \cdot x$$

■

8 APPENDIX C. EQUALITY CONSTRAINTS

In this section we consider the Lipschitz optimization program:

$$\max_{a \in D(s)} f(a,s) \tag{11}$$

in which:

$$D(s) = \{a | g_i(a,s) \leq 0, i = 1, ..., p \text{ and } h_j(a,s) = 0, \;\; j = 1, ..., q\}.$$

In addition to being locally Lipschitz, we assume that the equality constraints are differentiable and modify the GMFCQ to:

(i) h is differentiable with respect to a, $\nabla_a h$ is jointly continuous in a neighborhood of $(a^*(s), s)$, and the matrix $\nabla_a h(a^*(s), s)$ has full rank;

(ii) there exists $\overline{y} \in \mathbb{R}^n$ such that:

$$\forall \gamma_a \in \partial_a \overline{g}(a^*(s), s), \;\; \gamma_a \cdot \overline{y} < 0, \text{ and } \nabla_a h(a^*(s), s) \cdot \overline{y} = 0 \tag{12}$$

A couple of issues arise when seeking to relax the smoothness assumption (with respect to a) on the equality constraints and to only assume Lipschitzness. First, part of the GMFCQ needs to be generalized as specified in definition 3 of Section 2 and this obviously imposes additional restrictions . Second, as shown below, proofs of existence of envelopes in programs with equality constraints make use of an implicit function theorem. Unfortunately, implicit function theorems for Lipschitz functions fail to provide enough characterization of the subdifferential of the implicit function, given the weak composition rule for Lipschitz functions.[10] Consequently, and in line with the optimization literature, we only consider only smooth constraints and use the classical implicit function theorem stated below without proof.[11]

Proposition 22 *(Classical Implicit Function Theorem). Let $h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ be C^1, and suppose that $h(u,v) = 0$ and that $\nabla_u h(u,v)$ has maximal rank. Then there exists open neighborhoods U of (u,v) and W of v, and a C^1 mapping $w : W \to \mathbb{R}^k$ such that:*

$$(x,y) \in U \text{ and } h(x,y) = 0 \Longleftrightarrow y \in W \text{ and } x = w(y)$$

A direct application of this theorem shows that there exists a neighborhood U of $(a^*(s), s)$ and W of (a_I^*, s) in which:

$$h(a,s) = 0 \text{ and } (a,s) \in U \Longleftrightarrow (a_I^*, s) \in W \text{ and } a = (a_D(a_I, s), a_I)$$

where a_D is C^1 in (a_I, s) and $a_I \in \mathbb{R}^{n-q}$. With this change of variable, we define a new reduced-form objective function as $\widetilde{f}(a_I, s) = f(a_D(a_I, s), a_I, s)$, a new reduced-form set of constraints $\tilde{g}(a_I, s)$ and $\tilde{h}(a_I, s)$, and the following reduced-form problem:

$$\max_{a_I \in D(s)} \widetilde{f}(a_I, s) \tag{13}$$

[10] Something stronger than Lipschitzness but weaker than smoothness (in a) would be needed, perhaps strong differentiability (in a) in a neighborhood of $(a^*(s), s)$ with the strong differential having maximal rank.

[11] The Lipschitz version of this theorem simply assumes that h is locally Lipschitz and that $\partial_u h(u,v)$ has maximal rank, but only states that the implicit function w is locally Lipschitz on W.

subject to

$$D(s) = \{a_I \in \mathbb{R}^{n-q}, \, \widetilde{g}_i(a_I, s) \leq 0, \, i = 1, ..., p\}$$

Note that program (13) is only defined in a neighborhood W of (a_I^*, s), and that for corresponding values $(a, s) \in F(s)$ and $(a_I, s) \in W$, the constraints and the objective functions of the reduced-form programs and of the original program take the same values. As a result, the two programs (13) and (11) have corresponding feasible points and maxima, the same value function at s $V(s) = \widetilde{V}(s)$, as well as the same binding inequality constraints at $(a^*(s), s)$ and (a_I^*, s). Having eliminated the equality constraints through a change of variables, we prove in the next lemma that the GMFCQ constraint qualification for the full program (11) is transmitted to the reduced-form program (13) and becomes the GMFCQ/R:

$$\exists \overline{y} \text{ such that } \forall \widetilde{\gamma} \in \partial_{a_I} \widetilde{\overline{g}}(a_I^*, s), \; \widetilde{\gamma} \cdot \overline{y} < 0.$$

Lemma 23 *If the GMFCQ holds for some point $(a^*(s), s)$ for (11), then the GMFCQ/R holds at (a_I^*, s) for (13).*

Proof. Let y satisfy the GMFCQ at (a^*, s) so that:

$$\nabla_{a_D} h(a^*, s) \cdot y_D + \nabla_{a_I} h(a^*, s) \cdot y_I = 0$$

or, equivalently:

$$y_D = -\nabla_{a_D}^{-1} h(a^*, s) \circ \nabla_{a_I} h(a^*, s) y_I = \nabla_{a_I} a_D(a_I^*, s) \cdot y_I^*,$$

Also, by definition of y :

$$\forall \gamma \in \partial_a \overline{g}(a^*, s), \; \gamma \cdot y < 0,$$

and therefore:

$$\gamma_D \circ \nabla_{a_I} a_D(a_I^*, s) \cdot y_I + \gamma_I \cdot y_I = \gamma_D \cdot y_D + \gamma_I \cdot y_I < 0$$

Clearly, the same inequality holds for any convex combination of the $\gamma_D \in \partial_{a_D} \overline{g}(a^*, s)$. Recalling the change of variables $\widetilde{\overline{g}}(a_I, s) = \overline{g}(a_D(a_I, s), a_I, s)$ and the chain rule for generalized gradients:

$$\partial_{a_I} \widetilde{\overline{g}}(a_I^*, s) \subset co\{\partial_{a_D} \overline{g}(a^*, s) \circ \nabla_{a_I} a_D(a_I^*, s)\} + \partial_{a_I} \overline{g}(a^*, s))$$

which implies, by the last inequality above, that:

$$\widetilde{\gamma} \cdot y_I < 0 \text{ for any } \widetilde{\gamma} \in \partial_{a_I} \widetilde{\overline{g}}(a_I^*, s)$$

This proves that y_I satisfies the GMFCQ/R at (a_I^*, s). ■

There is, however, one major difference between the two programs. While it is possible for the full program to have several optima at a given s, the reduced form program is constructed in the neighborhood of only one maximum, and can only mimic the behavior of f around that particular maximum. If the maximum is unique, that is not a problem. If there are several maxima, then the reduced form program cannot capture all the information about the behavior of f, so one can only assert that $V(s') \geq \widetilde{V}(s')$ in a neighborhood of s, of course with equality at s. However, we know that the reduced form value function is continuous at s, and we have the following results:

Lemma 24 *Suppose that D is nonempty valued and uniformly compact in a neighborhood of s. (i) If the GMFCQ holds at $a^*(s) \in A^*(s)$, then V is continuous at s, (ii) If the GMFCQ holds at every element of $A^*(s)$ then V and $\widetilde{V}$ coincide in every direction of perturbation x in some neighborhood of s, and (iii) A^* is upper hemicontinuous at s.*

Proof. (i) As noted before, D is closed at s, and being uniformly compact near s, it is necessarily upper hemicontinuous at s so V is (at least) upper semicontinuous at s. The GMFCQ is transmitted to the GMFCQ/R for the reduced form program so $\widetilde{V}$ is continuous at s (from Lemma 7 above). Moreover $V(s') \geq \widetilde{V}(s')$ in a neighborhood of s, and thus:

$$\liminf_{t\to 0} V(s+tx) \geq \liminf_{t\to 0} \widetilde{V}(s+tx) = \widetilde{V}(s) = V(s)$$

which proves that V is lower semicontinuous at s as well.

(ii). Consider the sequence $\{t_n\}$ converging to 0 so that $\lim_{t_n\to 0}\{s+t_n x\} = s$. Since D is uniformly compact near s, and since D is closed at s, then every $D(s')$ with s' in some neighborhood of s is compact. This implies that for n large enough (say $n \geq N$, or, equivalently, t_n small enough), $D(s+t_n x)$ is compact and therefore f attains its maximum at on $D(s+t_n x)$. Consider then the sequence $\{a^*(s+t_n x)\}_{n\geq N}$ such that $V(s+t_n x) = f(a^*(s+t_n x), s+t_n x)$, with $a^*(s+t_n x) \in D(s+t_n x)$. By uniform compactness of D near s, there exists a neighborhood $N(s)$ such that the closure of $\cup_{s'\in N(s)} D(s')$ is compact, so that the sequence $\{a^*(s+t_n x)\}_{n\geq N}$ has a convergent subsequence converging to some a^*. Without loss of generality we may assume that $\lim_{t_n\to 0} a^*(s+t_n x) = a^*$. By continuity of V and f, and since D is closed at s, $a^* \in D(s)$ and $V(s) = f(a^*, s)$ which implies that $a^* \in A^*(s)$.[12] Because a^* is in $A^*(s)$, lemma23 above applies and the GMFCQ is transmitted to the reduced form problem at the corresponding maximum a_I^* of $\widetilde{f}$, and $\widetilde{V}$ is continuous at s as well. Uniform compactness of D near s implies uniform compactness of $\widetilde{D}$ near s, so there exists N' such that for all $n \geq \max(N, N')$, $\widetilde{f}$ attains its maximum at on $\widetilde{D}(s+t_n x)$ and therefore:

$$\widetilde{V}(s+t_n x) = \widetilde{f}(a_I^*(s+t_n x), s+t_n x) \text{ with } a_I^*(s+t_n x) \in \widetilde{D}(s+t_n x).$$

By construction $f(a^*(s+t_n x), s+t_n x) = \widetilde{f}(a_I^*(s+t_n x), s+t_n x)$ and thus, for n sufficiently large (i.e., for t_n small enough):

$$V(s+t_n x) = f(a^*(s+t_n x), s+t_n x) = \widetilde{f}(a_I^*(s+t_n x), s+t_n x) = \widetilde{V}(s+t_n x).$$

■

The initial and reduced-form programs also coincide at another level: A correspondence exists between the KKT vectors of the reduced program and those of the full program, as shown in the following lemma.

Lemma 25 *If λ is a KKT vector for program (13) at a_I^*, then there exists some $\mu \in \mathbb{R}^q$ such that (λ, μ) is a KKT vector for program (11) at $a^* = (a_D(a_I^*, s), a_I^*)$.*

[12] Summarizing, we just proved that for every sequence $\{s_n\}$ that converges to s *in the direction* x and every corresponding sequence $\{a^*(s_n)\}$ with $a^*(s_n) \in A^*(s_n)$ for each n, there exists a convergent subsequence of $\{a^*(s_n)\}$ whose limit point a^* is in $A^*(s)$. This is very close to showing that $A^* : S \to A$ is upper hemicontinuous at s (which in fact we prove below).

Proof. Since $\widetilde{h}$ is identical to 0 on W, and recalling that $\widetilde{h}_j(a_I,t) = h(a_D(a_I,t),a_I,t)$, then for all $j = 1,..,q$:

$$0 = \nabla_{a_D} h(a,t) \circ \nabla_{a_I} a_D(a_I,t)] + \nabla_{a_I} h(a,t)$$

Since $\nabla_{a_D} h(a,t)$ has maximal rank at $(a^*(s),s)$, by continuity of $\nabla_a h$, it has maximal rank on a neighborhood of $(a^*(s),s)$ included in U and thus:

$$\nabla_{a_I} a_D(a_I,t) = \{-\nabla_{a_D}^{-1} h(a,t) \circ \nabla_{a_I} h(a,t)\},$$

The reduced form equality constraints $\widetilde{h}$ being identical to 0, for any $\mu \in R^q$ we have:

$$\partial_{a_I}(\widetilde{f} - \lambda^T \widetilde{g})(a_I^* s) = \partial_{a_I}(\widetilde{f} - \lambda^T \widetilde{g} - \mu^T \widetilde{h})(a_I^* s)$$

and by the chain rule:

$$\partial_{a_I}(\widetilde{f} - \lambda^T \widetilde{g})(a_I^* s) \subset \partial_a(\widetilde{f} - \lambda^T \widetilde{g} - \mu^T \widetilde{h})(a^* s) \circ \{(\varepsilon, I_{n-q})\}$$

where I_{n-q} is the $(n-q)\times(n-q)$ identity matrix, and $\varepsilon = \{-\nabla_{a_D}^{-1} h(a^*(s),s) \circ \nabla_{a_I} h(a^*(s),s)\}$ is the gradient of a_D at (a_I^*,s) obtained above. Thus if λ is a Kuhn-Tucker vector for program (13) at a_I^*, there exists $(\varsigma,\gamma) \in \partial_a(f,\overline{g})(a^*(s),s)$ such that:

$$0 = (\varsigma_D - \lambda^T \gamma_D - \mu^T \nabla_{a_D} h(a^*(s),s)) \circ \varepsilon + \varsigma_I - \lambda^T \gamma_I - \mu^T \nabla_{a_I} h(a^*(s),s)$$

for any choice of μ. By assumption $\nabla_{a_D} h(a^*(s),s)$ has maximal rank, so for the specific choice of μ equal to $-\nabla_{a_D}^{-1} h(a^*,s)(\varsigma_D + \lambda \gamma_D)$ we have:

$$\varsigma_D - \lambda^T \gamma_D - \mu^T \rho_D = 0$$

and thus:

$$\varsigma_I - \lambda^T \gamma_I - \mu^T \rho_I$$

This establishes that:

$$\varsigma - \lambda^T \gamma - \mu^T \rho = 0$$

which proves that (λ,μ) is a Kuhn-Tucker vector for program (11) at $(a^*(s),s)$. ■

Consider now the two Lagrangians $\widetilde{L}(a_I,s;\lambda) = (\widetilde{f} - \lambda^T \widetilde{g})(a_I,s)$ and $L(a,s;\lambda,\mu) = (f - \lambda^T g - \mu^T h)(a,s)$, and the set $H = \{(a,s) : h(a,s) = 0\}$. By construction, for a given λ and for any μ, $\widetilde{L}$ takes the same values in a neighborhood W of (a_I^*,s) as the restriction of L to $H \cap U$ near (a^*,s). By Lemma 23, if the constraint qualification holds at (a^*,s), it also holds at (a_I^*,s). This implies that L may be used to calculate the generalized gradient of $\widetilde{L}$ at the optimum, so that:

$$\begin{aligned}
\partial_s \widetilde{L}(a_I^*,s,\lambda) &= \overline{co}\{\theta : \exists\{(a_n,s_n)\} \in dom \nabla_s L \cap (H \cap U),\ (a_n,s_n) \to (a^*,s),\ \nabla_s L(a_n,s_n,\lambda,\mu) \to \theta\} \\
&\subset \partial_s L(a^*,s)
\end{aligned}$$

The inclusion is generally not an equality, since the elements of $\partial\widetilde{L}$ are limits of sequences of derivatives evaluated at (a_n, s_n) belonging to H, whereas no such restriction is imposed on the sequences of derivatives used in defining the elements of ∂L. This inclusion implies that for any direction of perturbation x:

$$\widetilde{L}_s^{-o}(a_I^*, s; \lambda; x) = \min_{\widetilde{\theta}\in\partial_s(\widetilde{f}-\lambda\widetilde{g})(a_I^*(s),s)} \widetilde{\theta}\cdot x \geq \min_{\theta\in\partial_s(f-\lambda^T g-\mu h)(a^*(s),s)} \theta\cdot x = L_s^{-o}(a^*, s; \lambda, \mu; x)$$

so that:

$$\widetilde{L}_s^{-o}(a_I^*, s; \lambda; x) \geq L_s^{-o}(a^*, s; \lambda, \mu; x)$$

Furthermore, by Lemma 25, if $\lambda \in \widetilde{K}(a_I^*(s), s)$ there exists some μ such that $(\lambda, \mu) \in K(a^*(s), s)$, and thus:

$$\inf_{\lambda\in\widetilde{K}(a_I^*(s),s)} \widetilde{L}_s^{-o}(a_I^*, s; \lambda; x) \geq \inf_{(\lambda,\mu)\in K(a^*(s),s)} L_s^{-o}(a^*, s; \lambda, \mu; x) \tag{14}$$

Similarly:

$$\sup_{\lambda\in\widetilde{K}(a_I^*(s),s)} \max_{\widetilde{\theta}\in\partial_s(\widetilde{f}-\lambda^T\widetilde{g})(a_I^*(s),s)} \widetilde{\theta}\cdot x \leq \sup_{(\lambda,\mu)\in K(a^*(s),s)} \max_{\theta\in\partial_s(f-\lambda^T g-\mu^T h)(a^*(s),s)} \theta\cdot x \tag{15}$$

With these results we are finally equipped to state a nonsmooth envelope theorem with equality and inequality constraints.

Theorem 26 *If D is nonempty-valued and uniformly compact near s, and if the GMFCQ holds for every optimal solution $a^*(s) \in A^*(s)$, then for any direction of perturbation $x \in \mathbb{R}^m$:*

$$\liminf_{t\to 0^+} \frac{V(s+tx) - V(s)}{t} \geq \inf_{(\lambda,\mu)\in K(a^*(s),s)} \{L_s^{-o}(a^*(s), s; \lambda, \mu; x)\},$$

and:

$$\limsup_{t\to 0^+} \frac{V(s+tx) - V(s)}{t} \leq \sup_{(\lambda,\mu)\in K(a^*(s),s)} \{L_s^{o}(a^*(s), s; \lambda, \mu; x)\},$$

where

$$L_s^{-o}(a^*(s), s; \lambda, \mu; x) = \min_{\theta\in\partial_s(f-\lambda^T g-\mu^T h)(a^*(s),s)} \theta\cdot x$$

so that:

$$L_s^{o}(a^*(s), s; \lambda, \mu; x) = \max_{\theta\in\partial_s(f-\lambda^T g-\mu^T h)(a^*(s),s)} \theta\cdot x$$

Proof. By construction $V(s') \geq \widetilde{V}(s')$ in a neighborhood of s and $V(s) = \widetilde{V}(s)$, so that:

$$\begin{aligned} \liminf_{t\to 0^+} \frac{V(s+tx)-V(s)}{t} &\geq \liminf_{t\to 0^+} \frac{\widetilde{V}(s+tx)-\widetilde{V}(s)}{t} \\ &\geq \inf_{\lambda\in\widetilde{K}(a_I^*)} \{\widetilde{L}_s^{-o}(a_I^*, s; \lambda; x)\} \\ &\geq \inf_{(\lambda,\mu)\in K(a^*(s),s)} \{L_s^{-o}(a^*, s; \lambda, \mu; x\} \end{aligned}$$

the last inequality being precisely (14). Also, as in the proof without equality constraints, choose a sequence $\{t_n\}$ converging to 0 such that:

$$\begin{aligned} \limsup_{t\to 0^+} \frac{V(s+tx)-V(s)}{t} &= \lim_{n\to\infty} \frac{V(s+t_nx)-V(s)}{t_n} \\ &= \lim_{n\to\infty} \frac{\widetilde{V}(s+t_nx)-\widetilde{V}(s)}{t_n} \\ &\leq \sup_{\lambda\in\widetilde{K}(a^*(s),s)} \quad \widetilde{L}_s^o(a_I^*, s; \lambda; x) \\ &\leq \sup_{(\lambda,\mu)\in K(a^*(s),s)} \quad L_s^o(a^*, s; \lambda; \mu; x) \end{aligned}$$

the last inequality being simply (15). ■

References

[1] R. Amir, L. Mirman, W. Perkins, One-sector nonclassical optimal growth: optimality conditions and comparative dynamics, Int. Econ. Rev. 32 (1991) 625-644.

[2] R. Amir, Sensitivity analysis of multisector optimal economic dynamics, J. Math. Econ. 25 (1996) 123-141.

[3] K. Askri, C. LeVan, Differentiability of the value function of nonclassical optimal growth models, Journal of Optimization Theory and Applications 97 (1998), 591-604.

[4] A. Auslender, Differentiable Stability in non convex and non differentiable Programming, Math. Programming Stud. 10 (1979) 29-41.

[5] Benveniste, L. and J. Scheinkman, On the differentiability of the value function in dynamic models of economics, Econometrica 47 (1979) 727-32.

[6] F. Clarke, Generalized gradients and applications, Transactions of the American Mathematical Society 205 (1975) 247-262.

[7] F. Clarke, Generalized gradients of Lipschitz functionals, Advances in Mathematics 40 (1981) 52-67.

[8] F. Clarke, Optimization and Nonsmooth Analysis. SIAM 1983.

[9] A. Clausen, C. Strub, Envelope Theorems for Non-Smooth and Non-Concave Optimization, Mimeo.

[10] J. Danksin, The Theory of Max-Min, Springer, 1967.

[11] G. Fontanie, Subdifferential stability in Lipschitz programming, MS, Operations Research and Systems Analysis Center, University of North Carolina1980.

[12] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming, Mathematical Programming 12 (1979) 136-138.

[13] J. Gauvin, F. Dubeau, Differential properties of the marginal function in mathematical programming, Math. Programming Stud. 19 (1982) 101-119.

[14] J. Gauvin, J.W. Tolle, Differential stability in nonlinear programming, SIAM Journal of Control and Optimization 15 (1977) 294-311.

[15] E. G. Gol'stein, Theory of Convex Programming, Translations of Mathematical Monographs, Providence: American Mathematical Society, 1972.

[16] J.B. Hiriart Urruty, On optimality conditions in nondifferentiable programming, Mathematical Programming 14 (1978) 73-86.

[17] J.B. Hiriart Urruty, Refinement of necessary optimality conditions in nondifferentiable programming, Applied Mathematical Optimization 5 (1979) 62-82.

[18] W. Hogan, Point-to-Set Maps in Mathematical Programming, SIAM Review 15 (1973) 591-603.

[19] J. Kyparisis, On the uniqueness of Kuhn-Tucker Multiplier in nonlinear programming, Mathematical Programming 32 (1985) 242-246.

[20] R. Marimon, J.Werner On the Envelope Theorem without Differentiability, Mimeo.

[21] P. Milgrom I. Segal, Envelope theorems for arbitrary choices, Econometrica 70 (2002) 583-601.

[22] L. Mirman, I. Zilcha, On optimal growth under uncertainty, Journal of Econ. Theory 11 (1975) 329-339.

[23] O. Morand, K. Reffett, S. Tarafdar, A Nonsmooth Approach to Envelope Theorems, MS, Arizona State University, 2011.

[24] O. Morand, K. Reffett, S. Tarafdar, Nonsmooth methods in Lipschitzian stochastic dynamic programming, MS, Arizona State University, 2011.

[25] V.H. Nguyen, J.J. Strodiot, R. Mifflin, On conditions to have bounded multipliers in locally Lipschitz programming, Mathematical Programming 18 (1980) 100-106.

[26] J. Rincon-Zapatero, M. Santos, Differentiability of the value function without interiority assumptions Journal of Econ. Theory 144 (2009), 1948-1964.

[27] R.T. Rockafellar, Convex Analysis, Princeton, 1970.

[28] R.T. Rockafellar, R J-B Wets, Variational Analysis Springer.

[29] S. Tarafdar, Optimization in economies with nonconvexities, Ph.D. Dissertation, Arizona State University, 2010.