Optimization in Economies with Nonconvexities*

Suchismita Tarafdar†

(Job Market Paper)
Preliminary Draft: February 2010

Abstract

Nonconvex optimization is becoming the fashion to solve constrained optimization problems in economics. Classical Lagrangian does not necessarily represent a nonconvex optimization problem. In this paper, we give conditions under which the Classical Lagrangian serves as an exact penalization of a nonconvex programming. This has a simple interpretation and is easy to solve. We use this Classical Lagrangian to provide sufficient conditions under which value function is Clarke differentiable with differential bounds. The existence of Clarke envelopes has numerous potential examples in lattice programming, nonclassical growth theory and macroeconomics, Negishi methods, nonstationary dynamic lattice programming, and duopoly problems. Most importantly, the nonlinear duality theorem of this paper is used to provide generalized envelopes discussed in our companion paper.

1 Introduction

Nonconvex optimization problems arise naturally in many economic models. For example, starting in the 1950s with the work of Farrell [19], Rothenberg [58], Koopmans [36], Reiter [51] and others, emphasized the potential importance of nonconvexities in general equilibrium theory. More recently, nonconvexities have been used in many models in macroeconomics and growth theory to develop interesting new results. For example, in theoretical work,

*I especially thank Olivier Morand and Kevin Reffett for directing me through this entire project. I also thank Manjira Datta, Yiannis Vailakis, and Lukasz Woźny for helpful discussions during the writing of this paper. This is a preliminary draft of the paper. Errors may remain. Do not recirculate, rather write the author for the current version.

†Department of Economics, WP Carey School of Business, Arizona State University

for the case of one-sector growth, Dechert and Nishimura [15], Amir, Mirman, and Perkins [2], Hopenhayn and Prescott [27], Nishimura, and Rudnicki, and Stachurski [48], and Kamihigashi and Roy[30] [31] have studied the structure of "nonclassical" optimal growth models. In such models, optimal dynamics can be very different than their "classical" counterparts. In Amir [3], these ideas generalizes the nonclassical multisector models. For nonclassical stochastic growth models, Nishimura and Stachurski [47] develops a Foster-Lyapunov method to characterize conditions under which optimal dynamics in growth models with nonconvexities are stochastically stable. Recently, Mirman, Morand, and Reffett [43] study issues associated with recursive equilibrium in economics with nonconvexities, and argue the Lipschitzian structure of dynamic programs plays a critical role in developing nonsmooth characterizations of dynamic complementarities. In all of this work, to avoid issues associated with Lagrangian methods, artificially strong interiority conditions are imposed on optimal solutions. Similarly, in applied work in macroeconomics, nonconvexities in labor services have been proposed as a key ingredient in resolving important labor market puzzles in macroeconomic models (e.g., see Prescott, Rogerson, and Wallenius ([49]) and Rogerson and Wallenius ([55])). When studying the structure of asymptotic growth, Romer [56][57] emphasizes the importance of nonconvexities in production in growth models with endogenous technological innovations. Khan and Thomas [33] emphasize the importance of lumpy investment and nonconvexities in adjustment costs when explaining various investment puzzles at the plant and aggregate level in (s, S) models of dynamic equilibrium. Finally, in microeconomics and industrial organization models, there is a large literature on two-part marginal pricing equilibria emphasizing the role of nonconvexities and increasing returns in the theory of the firm and optimal pricing equilibrium (e.g., see Brown, Heller, and Starr [10]).

Despite this recent emphasis on nonconvexities in the economic literature, general results for optimization methods used to characterize the structure of optimal solutions to agent optimization problems in models with nonconvexities has been woefully neglected. What is typically done in the existing literature when Langrangian methods (or even unconstrained methods) are used is that the methods of convex programming are assumed to be, in some sense, applicable, and researchers proceed under this assumption that careful variations of the existing methods can be used to characterize optimal solutions. Unfortunately, it is well-known that for programming problems with nonconvexities, such an approach can be inappropriate. For example, unlike convex optimization problems, the mathematical foundations of the Kuhn-Tucker theory and duality for classical Lagrangians does not apply.[1] In convex problems, under suitable constraint qualifications (e.g., Slater conditions), constrained optimizations problems can be restated as unconstrained linear (or Classical) Lagrangian problems. Under standard differentiability conditions, it is straight forward to develop sharp necessary and sufficient first order conditions that can be used to characterize the set of optimal solutions. Furthermore, the classical Lagrangian satisfies zero duality gap. If the primitive data is smooth, and regularity conditions for boundary solutions are imposed (e.g., Rincón-Zapatero and Santos [52]), then classical (once-continuous differentiable or C^1) envelope theorems for the value function exist. In this case the Classical Lagrangian serves as an exact penalization representation of the original problem. In general, none of these statements are the case in optimization problems with nonconvexities.

More specifically, Classical Lagrangian methods do not necessarily achieve zero duality. First order conditions are necessary (not sufficient), even in smooth problems. In multistage nonconvex programs (i.e., nonconcave dynamic programming), objective functions are not even differentiable (rather, at best, locally Lipschitz). Conditions for classical C^1 envelope theorems cannot be easily produced. In textbook treatments of nonlinear programming that are not convex, to develop Lagrangian methods, authors typically impose very strong constraint qualifications on the problem. Further, and most importantly, they impose global smoothness of primitive data. This latter condition fails in many very simple examples of optimization problems with nonconvexities (e.g., dynamic programming problems and/or consumer choice problems). To deal with this situation in the mathematical programming literature, various other types of "augmented" Lagrangian methods have been developed in the literature, such as the quadratic penalty, or the absolute value penalty functions. (See Rockafellar and Wets ([54]) for a discussion). Unfortunately, the augmented Lagragians are difficult to solve in even very simple problems.

In this paper, we take a direct approach to the problem of characterizing optimal solutions in constrained optimization problems with nonconvexities. In particular, we develop sufficient conditions under which the classical Lagrangian can be applied, and serves as exact penalization of the constrained problem with nonconvexities. Further, we give conditions under which gen-

[1] By a "classical Lagrangian", we mean a Lagrangian that is linear in the Lagrange multipliers (as opposed, to "augment" Lagrangians as discussed in Rockafellar and Wets [54] or "Lagrangian-type" functions as discussed in Rubinov and Yang [59]).

eralized first order conditions can be delivered for the classical Lagrangian (i.e., generalized Kuhn-Tucker conditions), as well as provide simple formulas for generalized envelope theorems for the value function that can be constructed by min-max operation on the differential structure of the Lagrangian. These results allow the application of Classical Lagrangian methods to a broad class of optimization problems in models with nonconvexities, and, therefore, greatly simplifies the characterization of all optimal solutions for nonconvex problems. To achieve our strongest results (e.g., exact penalization, zero duality gap, and useful generalized Kuhn-Tucker conditions), our sufficient conditions involve Mangasarian-Fromowitz constraint qualification (MFCQ), which are very mild regularity conditions for most interesting economic models. In addition, under this very relaxed regularity condition, MFCQ, we provide conditions under which the value function has differential bounds. We apply the results of this paper to provide the most relaxed sufficient conditions in our companion paper Morand, Reffett and Tarafdar [45] under which (i) Clarke envelopes, (ii) directionally differentiable envelopes and finally (iii) once continuously differentiable envelopes exists. To get directional differntiable and continuosly differentiable envelopes we impose stronger conditions such as strict Mangasarian-Fromowitz constraint qualification (SMFCQ).

Our approach to nonlinear programming is very closely related to the work of Gauvin and Tolle [24], Auslender [5], and Gauvin and Dubeau [22]. These papers give conditions under which nonsmooth envelopes exist. Following the work of Fontanie [21], we improve upon this literature in two ways; (1) we relax the assumption of C^1 (i.e., smooth) objective functions, and (2) we apply a weaker constraint qualifications to obtain many of our results. Relaxing the assumption of continuous differentiability opens up many applications in dynamic programming, where the value function is not necessarily smooth even with all smooth primitive data, especially in the presence of nonconvexities. Relative to existing methods in the economics literature (e.g., Milgrom and Segal [41] and Rincon-Zapatero and Santos [52]), we not only weaken the constraint qualification needed for classical smooth envelope theorems, but we also unify many results within a broader approach of generalized envelope theorems for nonconvex Lipschitzian programming problems.

The paper is laid out as follows. In the next section, we introduce much of the mathematical terminology we need in the paper. Section 3 develops a nonlinear duality theory and gives the appropriate first order conditions. In Section 4, we provides numerous economic applications of our results.

2 Mathematical Preliminaries

We begin with a number of mathematical definitions that we shall use in this paper. See Rockafellar [53], Clarke [12], and Rockafellar and Wets [54] for further discussion.

2.1 Structural Properties of Functions

Let (X, ρ_X), (Y, ρ_Y), and (T, ρ_T) be metric spaces, and $f : X \to Y$ be a continuous function. The function f is *Lipschitz with module (*or *modulus)* k, $0 \leq k < \infty$, if for all $x, x' \in X$,

$$\rho_Y(f(x), f(x') \leq k\rho_X(x, x')$$

The function f is *locally Lipschitz near* $x \in X$ *of modulus* $k(x)$ if f is Lipschitz of module $k(x)$ on a neighborhood $N(x; e)$, $e > 0$. A function $f : X \times T \to Y$ is *uniformly Lipschitz in* t of modulus k if

$$\sup_{x \in X} \rho_Y(f(x, t), f(x, t')) \leq k\rho_T(t, t')$$

for all $t, t' \in T$. Finally, f is *uniformly locally Lipschitz* near $t \in T$ of modulus $k(t)$ if on a neighborhood $N(t, e)$,

$$\sup_{x \in X} |f(x, t') - f(x, t'') \leq k(t)\rho_T(t', t''), t', t'' \in N(t, e)$$

A particular type of locally Lipschitz function often used in economic optimization is a proper convex function. Let X be a convex set. A real valued function $f : X \to \mathbf{R}$ is *(strictly) convex* if for all $x, y \in X$, and all $\lambda \in (0, 1)$

$$f(\lambda x + (1 - \lambda)y) \leq (<)\lambda f(x) + (1 - \lambda)f(y)$$

The function $f(x)$ is strongly convex if $\exists$ a constant $\sigma > 0$ for all $x, y \in X$, and for all $\lambda \in (0, 1)$ such that

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y) + \frac{1}{2}\sigma\lambda(1 - \lambda)\left\|x - y\right\|^2$$

The function f is *essentially strongly concave (resp, strongly concave, strictly concave, concave) if* $-f$ is essentially strongly convex, (resp, strongly convex, strictly convex, convex). A *proper convex function* is a *convex function*

$f : X \to Y$, where Y is the *extended reals.* A proper convex function is locally Lipschitz for any open set in X.

In this paper, we will consider many different notions of differentiability for Lipschitz functions. Consider a Lipschitz continuous function $f : I \subset R^n \to \mathbf{R}^m$ of modulus k. At a point $x_0 \in I$, we first consider a number of different types of generalized smoothness of f in direction $x \in \mathbf{R}^n$ that will be used in some of our proofs:

Upper radical right Dini derivative is defined as:

$$D^+ f(x_0; d) = \lim\sup_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t}$$

Lower radical right dini derivative is defined as:

$$D_+ f(x_0; d) = \lim\inf_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t}$$

The *upper radical left dini derivative and lower right radical dini derivative* are defined similarly simply changing $t \to 0^+$ to $t \to 0^-$.

The *directional derivative at* $x_0 \in X$ in the direction $d \in \mathbf{R}^n$ is defined to be,

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

and *Clarke's upper and lower generalized directional derivative* at x_0 in the direction $d \in \mathbf{R}^n$ is,

$$\begin{aligned} f^o(x_0; d) &= \lim\sup_{\substack{y \to x_0 \\ t \to 0^+}} \frac{f(y + td) - f(y)}{t} \\ f^{-o}(x_0; d) &= \lim\inf_{\substack{y \to x_0 \\ t \to 0^+}} \frac{f(y + td) - f(y)}{t} \end{aligned}$$

It is important to remember that Clarke generalized derivatives of Lipschitz functions always exist, while directional derivatives of such functions need not. We say a function f is *Clarke regular* if its Clarke generalized directional derivative equals its directional derivative in all directions d.

A function f is *differentiable at* $x_0 \in X$ if the directional derivative exist in all direction and $f'(x_0; d) = \nabla_x f(x_0) \cdot d$. In this case, the derivative of is given by

$$\nabla_x f(x_0) = \lim_{h\to 0^-} \frac{f(x_0+h)-f(x_0)}{h} = \lim_{h\to 0^+} \frac{f(x_0+h)-f(x_0)}{h}$$

The function f has a *strict derivative* at x_0, denoted by $D_s f(x_0)$, when for all $d \in \mathbf{R}^n$.

$$\langle D_s f(x_0), d\rangle = \lim_{\substack{x\to x_0 \\ t\downarrow 0^-}} \frac{f(x+td)-f(x)}{t}$$

Finally, we say that f is *continuously differentiable* if $D_s f(x) : \mathbf{R}^n \to \mathbf{R}^{n\times m}$ is continuous at x_0. On a finite dimension domain, a strictly differentiable function is continuous differentiable.

Finally, recall that the subgradient of a convex function f, is the set of $p \in M_{m\times n}$ satisfying:

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

for direction $d \in \mathbf{R}^n$. The set of subgradients of a convex function is subdifferential. Dually we can define a subdifferential for any function, but for a non convex function this set may or may not exist.

Since Lipschitz functions may not necessarily have subgradients, we define Clarke's generalized gradient as

$$\partial f(x_0) = co\left\{\lim \nabla f(x_i) : x_i \to x_0, x_i \notin S, x_i \notin \Omega_f\right\}$$

where *co* denotes the convex hull, S 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.

2.2 Structural Properties of Correspondences

As we study the value functions and optimal solutions of collections of parameterized optimization problems for economic decision makers, it turns out that the topological properties of feasible correspondences prove very important in our work. Let X and Y be topological spaces, and $F : X \rightrightarrows Y$ a correspondence. A correspondence $F(x)$ is *upper semicontinuous (or u.s.c.)* at $x_0 \in X$ if for any two sequences $\{x_n\}$ and $\{y_n\}$ such that $x_n \to x_0$, $y_n \to y_0$ with $y_n \in F(x_n)$ and implies $y_0 \in F(x_0)$. If F is *upper* *semicontinuous* at x for all $x \in X$, then it is *upper semicontinuous.* $F(x)$ is *lower semicontinuous (or l.s.c)* at $x_0 \in X$ if for any two sequences $\{x_n\}$ and $\{y_n\}$ such that $x_n \to x_0$, with $y_n \in F(x_n)$ and $y_0 \in F(x_0)$ implies $y_n \to y$. If F is *lower semicontinuous* at x for all $x \in X$, then it is *lower semicontinuous.* [2] $F(x)$ is *continuous* at $x_0 \in X$ if it is both u.s.c and l.s.c. at $x_0 \in X$, and a *continuous correspondence* if it is continuous for all $x \in X$.

As with functions, we can characterize the metric properties of correspondences. Let (X, ρ_X) and (Y, ρ_Y) be metric spaces,the correspondence $F : X \rightrightarrows Y$, and 2^X the powersets of X. A useful metric for correspondences is the Hausdorff metric. Define Hausdorff distance between $F(x'')$ and $F(x')$ for all $x'', x' \in X$ by

$$\rho_y^H\left(F(x'), F(x'')\right) = Max\left[\sup_{y' \in F(x')} \rho_y(y', F(x'')), \sup_{y'' \in F(x'')} \rho_y(y'', F(x'))\right]$$

where $\rho_y(y', F(x'')) = \inf_{y'' \in F(x'')} \rho_y(y', y'')$ and $\rho_y(y'', F(x')) = \inf_{y' \in F(x')} \rho_y(y', y'')$

We say a correspondence $F(x)$ is *Lipschitz continuous of modulus k on X* if $\forall\ x',\ x'' \in X$,

$$\rho_y^H\left(F(x'), F(x'')\right) \le k\rho_x\left(x'', x'\right)$$

$F(x)$ is *locally Lipschitz continuous near $x \in X$ of modulus $k(x)$ on a neighborhood* $N(x, e)$ if it is Lipschitz continuous of modulus k on $N(x, e)$. In other words, for all x', $x'' \in N(x, e)$,

$$\rho_y^H\left(F(x'), F(x'')\right) \le k(x)\rho_x\left(x'', x'\right)$$

Finally, we say $F(x)$ is *uniformly compact near x* if there is a neighborhood $N(x)$ of x such that the closure of $\cup_{x' \in N(x)}\Gamma(x')$ is compact for all x'.

[2] See Berge ([9], p108) for discussion. Note also that to define u.s.c. and l.s.c. of a correspondence, we only need topological spaces (i.e., the topological spaces need not be metrizable).

3 Classical Lagrangian and the Non Linear Duality

We focus in this paper is on parameterized Lipschitzian optimization problems. Consider a collection of parameterized optimization problems describe as follows: let $a \in A$ be the space for the choice variables, $s \in S$ be the parameter space, $f : A \times S \to \mathbf{R}$ be an objective function, and $D : S \rightrightarrows A$ a correspondence that describes the feasible set of actions in each state $s \in S$. We consider a family of *Parameterized Lipschitzian Optimization Problem*:

$$V(s) = \max_{a \in D(s)} f(a, s) \tag{3.0.1}$$

where $D(s)$ is given by:

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

with the function $g_i(a, s)$ being inequality constraints, and $h_i(a, s)$ the equality constraints. The optimal solution correspondence in problem (3.0.1) is denoted by $A^* : S \rightrightarrows A$, and defined follows:

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

We make four assumptions that we maintain throughout the paper:

Assumption 1: The primitive data in (3.0.1) *satisfies the following conditions:*

(a) A is a sequentially compact topological space;

(b) (S, ρ_S) a metric space;

(c) the objective function $f : A \times S \to \mathbf{R}$ is continuous in (a, s);

(d) the feasible correspondence $D : S \rightrightarrows A$ is a non empty-valued continuous, compact-valued correspondence.

Assumption 2: (A, ρ_A) is a metric space.

Assumption 3: (i) The spaces (A, ρ_A) and (S, ρ_S) are each convex in R^n and R^m, respectively, and (ii) the constraints g_i, $i = 1, .., p$ and $h_j = 0$, $j = 1, .., q$ are jointly C^1.

Assumption 4: The objective functions and the constraints are defined on a set A' where *$A \subset A'$.*

The fourth assumption enables us to treat the noninterior solutions at par with the interior ones.

The value function $V(s)$, and the set of optimal solutions $A^*(s)$, are well-defined by Berge's Maximum Theorem under assumption 1. For the sake of completeness, we state the maximum theorem that is appropriate in our setting.

Proposition 1 *(Berge, [9], Maximum Theorem, p 116). Under Assumption 1, in Problem (3.0.1), (i) V is continuous on S, and (ii) A^* is upper hemicontinuous on S.*

In this section, we first develop a nonlinear duality theory for problems with inequality constraints.[3] Specifically, we prove a key theorem that characterizes the nature of this nonlinear duality present in particular parameterized Lipschitzian versions of the problem in (3.0.1). We show that the gradient of the standard Lagrangian obtains a zero-duality gap, and satisfies a local saddlepoint property under a weak constraint qualification. Then we built our main nonlinear duality result: the classical Lagrangian serves as an exact penalization function for a nonconvex optimization problem under some conditions. Before we get to the main theorem we introduce the min-max Lagrangian and the constraint qualifications used in the paper.

3.1 Min-Max Lagrangian

The Classical Lagrangian corresponding to (3.0.1) is defined as

$$\begin{aligned} L(a,\lambda,\mu;s) &= f(a,s)-\lambda^T g(a,s)-\mu^T h(a,s) \text{ if } a\in D(s) \qquad (3.1.1)\\ &= -\infty \text{ otherwise} \end{aligned}$$

where $D(s)$ is the feasible set given by the inequality and equality constraints. Further:

$$g(a,s)=[g_1(a,s),...,g_p(a,s)]^T;\ h(a,s)=[h_1(a,s),...,h_q(a,s)]^T$$

For any $s\in S$, $(\lambda,\mu)\in K(s)$, where $K(s)$ is nonempty and convex .From Rockafellar ([53], Lemma 36.1)

$$\sup_{a\in D(s)}\inf_{(\lambda,\mu)\in K} L(a,\lambda,\mu;s)\leq \inf_{(\lambda,\mu)\in K}\sup_{a\in D(s)} L(a,\lambda,\mu;s)$$

If the reverse inequality hold, then the Lagrangian has a min-max value or a saddle value and achieves a zero duality gap. In such a case there exist $(a^*(s),\lambda^*(s),\mu^*(s))$ such that The Lagrangian satisfies the following,

$$L(a,\lambda^*(s),\mu^*(s);s)\leq L(a^*(s),\lambda^*(s),\mu^*(s);s)\leq L(a^*(s),\lambda,\mu;s)$$

for all $a\in D(s)$ and for all $(\lambda,\mu)\in K(s)$. Unlike in nonconvex optimization programming case, in convex optimization programming the classical

[3] Morand, Reffett and Tarafdar [45] extend the results in this paper by introducing smooth equality constraints.

Lagrangian has a saddle value and achieves a zero duality gap. Since this is not true for nonconvex programming problems, we do not know if the classical Lagrangian is the right penalization function. There are various other penalization functions discussed in the literature (Rubinov and Yang [59]). In this paper we show under a not so strong constraint qualifier the classical Lagrangian achieves zero duality gap. This merits the discussion of constraint qualification, and defining the various forms of constraint qualifications that we shall use in this paper.

3.2 Constraint Qualifications

To obtain Lagrange multiplier rules and to characterize the local saddlepoint properties of $L(a, s; \lambda, \mu)$, we need to make restrictions on the functions g and h that define the feasible correspondence $D(s)$. In particular, we need constraint qualifications. Under Assumption 3, in (3.0.1), we will consider three types of constraint qualifications:[4].

(i) the Mangasarian-Fromowitz constraint qualification (MFCQ),

(ii) the strict Mangasarian-Fromowitz constraint qualification (SMFCQ), and

(iii) the linear independence constraint qualification (LICQ).

Among the three, the weakest form of constraint qualification is the MFCQ. We say a feasible point $a \in D(s)$ satisfies the *Mangasarian-Fromovitz Constraint Qualifier (MFCQ) if:*

(i) the following vectors are linearly independent

$$\nabla_a h_i(a, s),\ j = 1, ..., q$$

(ii) there exists a $\widetilde{y} \in \mathbf{R}^n$ such that,

$$\nabla_a g_i(a, s)\widetilde{y} < 0,\ i \in I;\ \nabla_a h_j(a, s)\widetilde{y} = 0, \quad j = 1, ..., q$$

where $I = \{i : g_i(a, s) = 0\}$.

We also consider two other constraint qualifications that are stronger than MFCQ. The first (and weaker) constraint qualification is the SMFCQ. We say a feasible point $a \in D(s)$ satisfies the *Strict Mangasarian-Fromovitz Constraint Qualifier (SMFCQ) if:*

(i) the following vectors are linearly independent

$$\nabla_a g_i(a, s),\ i \in I_b;\ \nabla_a h_i(a, s),\ j = 1, ..., q$$

[4] For the sake of comparison with Milgrom and Segal [41], for convex problems, for $s \in S$, we say a constraint system satisfies a *Slater condition* if there exists a point a$\in D(s)$ such that $h(a, s) = 0$ and $g_i(a, s) < 0$ for all constraints i that are active.

(ii) there exist $\overline{y} \in \mathbf{R}^n$ such that,

$$\nabla_a g_i(a,s)\overline{y} < 0,\ i \in I_s;\ \nabla_a g_i(a,s)\overline{y} = 0,\ i \in I_b$$
$$\nabla_a h_j(a,s)\overline{y} = 0, \quad j = 1, ..., q$$

where $I_b = \{i \in I : \lambda_i > 0\}$, $I_s = \{i \in I : \lambda_i = 0\}$ and $I = \{i : g_i(a,s) = 0\}$.

A third constraint qualification is the strongest we consider, and is the focus of the recent work of Rincon-Zapatero and Santos [52]. We say a feasible point $a \in D(s)$ satisfies the *Linear Independence Constraint Qualifier (LICQ) if,* the following vectors are linearly independent,

$$\nabla_a g_i(a,s),\ i \in I, \quad \nabla_a h_i(a,s),\ j = 1, ..., q$$

where $I = \{i : g_i(a,s) = 0\}$.

3.3 Lagrange Multiplier Rule

Now we discuss various first order conditions for nonconvex optimization problems with nonsmooth objectives. If the objective is not continuously differentiable the standard first order conditions do not apply. In this paper we will always assume the objective to be at least jointly locally Lipschitz. Thus, we provide usable first order conditions for locally Lipschitz and/or directionally differentiable objectives. Note, the first order conditions can be easily adapted for locally Lipschitz and/or directionally differentiable inequality constraints.

Proposition 2 *Under Assumptions 1-4, $f(a^*(s), s)$ attains a local maxima for a given s and $a^*(s)$ satisfies MFCQ, (i) if $f(a,s)$ is directional differentiable then for each s and each direction $x_a \in \mathbf{R}^n$*

$$f'_a(a^*(s), s; x_a) - \left(\lambda^T \nabla g_a(a^*(s), s) - \mu^T \nabla h_a(a^*(s), s)\right) \cdot x_a \leq 0 \quad (3.3.1)$$

(ii) if $f(a,s)$ locally Lipschitz then there exist $\varsigma_{a^*(s)}(a^*(s), s) \in \partial_a f(a^*(s), s)$ such that

$$\varsigma_{a^*(s)}(a^*(s), s) = \lambda^T \nabla g_a(a^*(s), s) + \mu^T \nabla h_a(a^*(s), s) \quad (3.3.2)$$

also (iii) if $f(a,s)$ locally Lipschitz then for each s and each direction $x_a \in \mathbf{R}^n$

$$f_a^{-o}(a^*(s), s; x_a) - \left(\lambda^T \nabla g_a(a^*(s), s) - \mu^T \nabla h_a(a^*(s), s)\right) \cdot x_a \leq 0 \quad (3.3.3)$$

Proof. Appendix ■

The relevant first order conditions can be similarly stated for minimization problems. These first order conditions will be useful for economic applications in section 4 where the objective of an optimization problem is not smooth (or C^1). This will be specially important for dynamic programming problems with some nonconvexities. Here, the Bellman equation will not necessarily be smooth due to presence of nonconvexities even if all primitive data is smooth. Consequently the importance of the first order conditions of proposition (2) are significant.

3.4 Non-Linear Duality

We will now state a result on Nonlinear Duality for problem (3.0.1) with the constraint set comprising of only inequalities. Morand, Reffett and Tarafdar [45] show that smooth equality constraints and easily be added to this result by standard implicit function theorem. From Auslander ([5]) we know under MFCQ a standard Lagrangian exist. The first order conditions of this Lagrangian is discussed above. Our next theorem show, Clarke's generalized directional derivative of the Lagrangian satisfies saddle point property in the direction of perturbation in the choice variable and the Kuhn-Tucker multipliers. This saddle function satisfies zero duality and can be used to study the structure of the value function in (3.0.1). Further the saddle value of this function is independent of any perturbation in the choice variable a. In other words, under the conditions of Theorem (3) for any direction of perturbation of the parameter s, the effect of the choice variable a is enveloped out. This observation is crucial in obtaining nonsmooth envelope theorems which is the focus of Morand, Reffett and Tarafdar [45].

Theorem 3 *At $s \in S$, suppose MFCQ holds at an optimal solution $a^*(s) \in A^*(s)$. Then, under Assumptions 1-4, for any direction of perturbation $x \in \mathbf{R}^m$, and for all $(\varsigma_a, \varsigma_s) \in \partial_a f(a^*(s), s) \times \partial_s f(a^*(s), s)$:*

$$\begin{aligned} \check{S}(y, \lambda) \quad &= \min_{\varsigma_a \in \partial_a f(a^*(s), s)} (\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s), s)) \cdot y \\ &\quad + \min_{\varsigma_s \in \partial_s f(a^*(s), s)} (\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s), s)) \cdot x \end{aligned}$$

is a saddle function with saddle value

$$L_s^{-o}(a^*(s), s, \lambda, \mu; x) = \min_{\varsigma_s \in \partial_s f(a^*(s), s)} \left[\left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s), s) \right) \cdot x \right]$$

Here $\overline{g}$ denotes the active inequality constraints.

Proof. Consider for any $x \in \mathbf{R}^m$:

$$\begin{aligned}\check{S}(y,\lambda) &= \min_{\varsigma_a \in \partial_a f(a^*(s),s)} (\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s)) \cdot y \\ &\quad + \min_{\varsigma_s \in \partial_s f(a^*(s),s)} (\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)) \cdot x\end{aligned}$$

To simplify notation, it shall be understood that that minimizations with respect to ς_i are taken over the elements of the generalized gradients $\partial_i f$ for $i = a, s$ which is a nonempty-compact-valued correspondence for each $s \in S$.

Given any direction x of perturbation, define $D'(a^*(s),s)$ and $D''(a^*(s),s)$, as:

$$\begin{aligned}D'(a^*(s),s) &= \{y \in \mathbb{R}^n : \ \nabla_a \overline{g}(a^*(s),s) \cdot y + \nabla_s \overline{g}(a^*(s),s) \cdot x < 0\} \\ D''(a^*(s),s) &= \{y \in \mathbb{R}^n : \ \nabla_a \overline{g}(a^*(s),s) \cdot y + \nabla_s \overline{g}(a^*(s),s) \cdot x = 0\end{aligned}$$

Let $K(a^*(s),s)$ be the set of Kuhn-Tucker multipliers for state s at the optima $a^*(s) \in A^*(s)$. By MFCQ, $D' \cup D''$ is non-empty. Further, clearly $D' \cup D''$ and K are closed convex sets. Note, that if $\lambda \in K$, it must be:

$$\forall y \in \mathbb{R}^n, \ \min (\varsigma_a - \lambda \nabla_a \overline{g}(a^*(s),s)) \cdot y \leq 0,$$

so,

$$\sup_{y \geq 0} \check{S}(y,\lambda) = \left\{ \begin{array}{c} \min \left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right) \cdot x \text{ if } \lambda \in K \\ +\infty \text{ otherwise} \end{array} \right\} \tag{3.4.1}$$

Also, if $y \notin D' \cup D''$, then $\nabla_a \overline{g}(a^*(s),s) \cdot y + \nabla_s \overline{g}(a^*(s),s) \cdot x \geq 0$; therefore,

$$\begin{aligned}&\inf_{\lambda \geq 0} - [\nabla_a \overline{g}(a^*(s),s) \cdot y + \nabla_s \overline{g}(a^*(s),s) \cdot x] \\ =\ & -\infty \text{ or } 0\end{aligned}$$

so

$$\inf_{\lambda \geq 0} \check{S}(y,\lambda) = \left\{ \begin{array}{l} \min(\varsigma_a \cdot y + \varsigma_s \cdot x) \ \ (\text{if } y \in D' \cup D'') \\ -\infty \text{ otherwise} \end{array} \right\} \tag{3.4.2}$$

Now, we can show that:

$$-\infty < \sup_y \inf_{\lambda \geq 0} \check{S}(y,\lambda) \leq \inf_{\lambda \geq 0} \sup_y \check{S}(y,\lambda) < +\infty \tag{3.4.3}$$

Next, we have the following sequences of inequalities:

$$
\begin{aligned}
& \min_{(\varsigma_a,\varsigma_s)} \sup_y \inf_{\lambda\geq 0} \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s)\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right) \cdot x \end{array} \right] \\
\leq \quad & \inf_{\lambda\geq 0} \sup_y \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s)\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right) \cdot x \end{array} \right]
\end{aligned} \tag{3.4.4}
$$

Denoting $(\varsigma_a, \varsigma_s)$ as one of the particular subgradients for which (3.4.4) is attained, we can summarize the above sequence of inequalities as follows:

$$
\begin{aligned}
& \sup_y \inf_{\lambda\geq 0} \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s)\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right) \cdot x \end{array} \right] \\
\leq \quad & \inf_{\lambda\geq 0} \sup_y \left[\begin{array}{c} \left(\varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s)\right) \cdot y \\ + \left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right) \cdot x \end{array} \right]
\end{aligned} \tag{3.4.5}
$$

Now, consider the following dual pair of linear programs: the primal is given as:

$$\varsigma_s \cdot x - \max \left(\lambda^T \nabla_s \overline{g}(a^*(s),s) - \mu^T \nabla_s h(a^*(s),s)\right) \cdot x$$

subject to

$$
\begin{aligned}
& \lambda \geq 0 \\
& \varsigma_a - \lambda^T \nabla_a \overline{g}(a^*(s),s) = 0
\end{aligned}
$$

while, the dual is given as:

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

subject to:

$$
\begin{aligned}
& \nabla_a \overline{g}(a^*(s),s) \cdot y + \nabla_s \overline{g}(a^*(s),s) \cdot x \leq 0 \\
& y \text{ unrestricted}
\end{aligned}
$$

Note, for any $y \in D' \cup D''$ feasible in the dual, by (3.4.3), the dual objective is bounded below since:

$$-\infty < \sup_{y\in x} (\min_{\varsigma_a,\varsigma_s} (\varsigma_a \cdot y + \varsigma_s \cdot x)) < +\infty$$

Therefore, there is also no duality gap; hence, (3.4.5) is an equality, so $\check{S}$ has a saddle value.

So we have:

$$
\begin{aligned}
& \sup_{y}[\min_{\varsigma_a,\varsigma_s}(\varsigma_a \cdot y + \varsigma_s \cdot x)] \\
= \;& \sup_{y} \inf_{\lambda\geq 0} \check{S}(y,\lambda) \\
= \;& \inf_{\lambda\geq 0} \sup_{y} \check{S}(y,\lambda) \\
= \;& \inf_{\lambda\geq 0} \min_{\varsigma_s} \left[\left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right)\cdot x\right] \\
= \;& \inf_{\lambda\geq 0} L_s^{-o}(a^*(s),s,\lambda;x)
\end{aligned}
$$

Summarizing:

$$
\sup_{y}[\min_{\varsigma_a,\varsigma_s}(\varsigma_a \cdot y + \varsigma_s \cdot x)] = \inf_{\lambda\geq 0} L_s^{-o}(a^*(s),s,\lambda;x)
$$

Thus $\check{S}(y,\lambda)$, is a saddle function with saddle value

$$
L_s^{-o}(a^*(s),s,\lambda,\mu;x) = \min_{\varsigma_s\in\partial_s f(a^*(s),s)} \left[\left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s),s)\right)\cdot x\right]
$$

■

Our main contribution is to show that the classical Lagrangian satisfies zero duality gap and achieves saddle value. To do this, first we need to show the value functions is locally Lipschitz, albeit under some conditions. Before we proceed we state the following corollary to Theorem (3). The corollary provides bounds on the Clarke gradient of the objective function and the gradient of the inequality constraints, which is critical for our result.

Corollary 4 *At $s \in S$, suppose MFCQ holds at an optimal solution $a^*(s) \in A^*(s)$. Then, under Assumptions 1-4 and no equality constraints, for any direction of perturbation $x \in \mathbf{R}^m$, and any $\epsilon > 0$, there exists a vector $\overline{y}(\epsilon,x)$ such that for all $(\varsigma_a,\varsigma_s) \in \partial_a f(a^*(s),s)\times\partial_s f(a^*(s),s)$ the saddle value satisfy*

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

and the active inequality constraints satisfy

$$
\nabla_a \overline{g}(a^*(s),s)\cdot\overline{y} + \nabla_s \overline{g}(a^*(s),s)\cdot x < 0
$$

Proof. From Theorem (3) $\check{S}(y,\lambda)$ is a saddle function with saddle value

$$
\inf_{\lambda\geq 0} \sup_{y} \check{S}(y,\lambda) = \inf_{\lambda\in K} L_s^{-o}(a^*(s),s;\lambda;x)
$$

Further by (3.4.2) of the previous theorem, we have

$$\sup_{y} \inf_{\lambda \geq 0} \check{S}(y, \lambda) = \sup_{y \in x} \min_{(\varsigma_a, \varsigma_s)} (\varsigma_a \cdot y + \varsigma_s \cdot x)$$

Therefore, appealing to the fact that $\check{S}$ has a saddle value, we have:

$$\inf_{\lambda \in K} L_s^{-o}(a^*(s), s; \lambda; x) = \sup_{y \in D'} \min_{(\varsigma_a, \varsigma_s)} (\varsigma_a \cdot y + \varsigma_s \cdot x),$$

Here, in the last expression, the supremum may not be attained since D' (defined in the previous theorem) is not necessarily bounded. Thus, $\forall \epsilon > 0$, choose $y(x, \epsilon)$ in D' such that:

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

Consider $\overline{y} = y(x, \epsilon) + \delta \widetilde{y}$ where δ is arbitrarily small and $\widetilde{y}$ satisfies MFCQ (that is, $\nabla_a \overline{g}(a^*(s), s) \cdot \widetilde{y} < 0$). We have:

$$\begin{aligned} & \nabla_a \overline{g}(a^*(s), s) \cdot \overline{y} + \nabla_s \overline{g}(a^*(s), s) \cdot x \\ = \; & (\nabla_a \overline{g}(a^*(s), s) \cdot y + \nabla_s \overline{g}(a^*(s), s) \cdot x) + \delta \nabla_a \overline{g}(a^*(s), s) \cdot \widetilde{y} \\ < \; & 0 \end{aligned}$$

Further:

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

Choose δ small enough such that $\delta \min(\varsigma_a \cdot \widetilde{y}) > -\epsilon/2$. Then, we obtain for all $(\varsigma_a, \varsigma_s) \in \partial_a f(a^*(s), s) \times \partial_s f(a^*(s), s)$:

$$\nabla_a \overline{g} \cdot \overline{y}(a^*(s), s) + \nabla_s \overline{g}(a^*(s), s) \cdot x < 0$$

and:

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

■

Now we proceed to show the Dini derivatives of the value function are bounded in the next subsection.

3.4.1 Differential Bounds

For differential bounds, we assume MFCQ (the weakest constraint qualification we consider). Under this constraint qualification, assuming a locally Lipschitz objective, the calculation of the differential bounds of the value function can be obtained. We do this by first providing method to compute both the lower and upper bounds for $V(s)$ using the Lagrangian $L_s^o(a^*(s), s; \lambda; x)$, and then we immediately have a global bound, and our main result on differential bounds follows immediately.

Theorem 5 *Under Assumptions 1-4, for problem (3.0.1), if, (i) f locally Lipschitz, (ii) $D(s)$ is nonempty and uniformly compact near s, and (iii) MFCQ hold for every optimal solution $a^*(s) \in A^*(s)$, then for any direction $x \in \mathbf{R}^m$ of perturbation we have the following:*

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

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

$$\begin{aligned} (iii) \qquad & \sup_{a^*(s) \in A^*(s)} \min_{\lambda \in K(a^*(s),s)} \{L -_s^o (a^*(s), s; \lambda) \cdot x\} \\ \leq \quad & D_+ V(s; x) \leq D^+ V(s; x) \\ \leq \quad & \sup_{a^*(s) \in A^*(s)} \max_{\lambda \in K(a^*(s),s)} \{L_s^o (a^*(s), s; \lambda) \cdot x\} \end{aligned}$$

where

$$\begin{aligned} L_s^{-o}(a^*(s), s; \lambda; x) \quad &= \quad \min_{\varsigma_s \in \partial_s f(a^*(s),s)} \left[\left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s), s) \right) \cdot x \right] \\ L_s^{o}(a^*(s), s; \lambda; x) \quad &= \quad \max_{\varsigma_s \in \partial_s f(a^*(s),s)} \left[\left(\varsigma_s - \lambda^T \nabla_s \overline{g}(a^*(s), s) \right) \cdot x \right] \end{aligned}$$

Proof. Appendix. ■

The Dini derivatives are bounded above and below, so the value function is locally Lipschitz in s. It is well known that a locally Lipschitz function is calm. Theorem (5) is an very important intermediary step to show the Classical Lagrangian serves as an exact penalization function for a nonconvex optimization problem under very mild conditions. However, it is pertinent to point out that Theorem (5) is interesting in by it's own, as in basically shows that the Clarke envelopes exist. The Clarke gradient of the value function for problem (3.0.1) is strictly included in the Clarke gradient of the Lagrangian.

Also the Clarke upper and lower generalized directional derivatives of the value function are bounded above and below respectively by the upper and lower generalized directional derivatives of the Lagrangian. Thus, a lot of interesting economic problems with nonconvexities can be solved by the Clarke envelope. Further, by strengthening the hypothesis of Theorem (5), directional differentiable and continuously differentiable envelopes can be obtained. This is discussed in length in Morand Reffett and Tarafdar ([45]).

For a $s \in S$, the problem (3.0.1) is calm at s of module $k^c(s)$ if $\forall\ s'$ in the neighborhood of s, $N(s, e)$

$$\left\|V(s) - V(s')\right\| \leq k^c(s) \left\|s - s'\right\|$$

Thus, under the hypothesis of Theorem (5) the value function is locally Lipschitz and consequently calm.

Now we state the most important contribution of the paper. The following theorem gives condition under which the classical Lagrange is a saddle function and satisfies zero duality gap. Thus the classical Lagrange procedure can be applied to solve economic optimization problems with nonconvexities.

Theorem 6 *Under Assumptions 1-4, if (i) f locally Lipschitz, (ii) $D(s)$ is nonempty and uniformly compact near s, and (iii) MFCQ holds at all optimal solution $a^*(s) \in A^*(s)$, then, for any direction of perturbation $x \in R^m$, at $s \in S$ and any $y \in R^n$, a zero duality gap is obtained for a Classical Lagrangian L, and $V(s) = \min_{\lambda,\mu} \max_a L(a, \lambda, \mu, s)$ is a saddle-value.*

Proof. Under assumption 1-4, f locally Lipschitz and Clarke regular, $D(s)$ is nonempty and uniformly compact near s, and MFCQ holds at all optimal solution $a^*(s) \in A^*(s)$ from Theorem (5), the value function is locally Lipschitz, hence calm. Appealing to Bonnans and Shapiro (Theorem 3.4(i)) calmness implies the Classical Lagrangian (3.1.1) of problem (3.0.1), has is a saddle-value and zero duality gap. ■

This theorem lays down the mathematical groundwork necessary to solve a optimization problem for economies with nonconvexities via the classical Lagrangian procedure. Thus we unify the approach to solving convex and nonconvex optimization problems.

4 Applications

4.1 Nonsmooth Envelope Theorems

Morand Reffett and Tarafdar [45] develop a nonsmooth approach to envelope theorems that unifies the result across nonconvex and convex parameterized nonlinear optimization problems important in economics as an application of our theory in the last section. In the literature by envelope theorem, one means the once continuously differentiable envelope (or smooth envelopes). However, there are many instances in economics where nonconvexities very naturally arise. For example quantity discounts make the budget set nonconvex for the consumer utility maximization problem. Secondly, the production function might be nonconvex due to indivisibility of some inputs, increasing returns, or existence of externalities. In a finite or infinite period dynamic programming problem, the Bellman equation can be guaranteed to be continuously differentiable for convex optimization problems under some conditions. But in the presence of nonconvexities the value function of an optimization problems looses smoothness very easily, thus the Bellman equation need not be smooth, rendering the classical envelope theorem inapplicable. Thus the nonsmooth envelopes discussed in length in Morand Reffett and Tarafdar [45] is very useful in solving nonconvex dynamic optimization problems among others. Theorem (5) gives sufficient conditions for the existence of Clarke envelopes. By imposing stronger conditions the next corollary provides sufficient conditions for the existence of directionally differentiable envelops:

Corollary 7 *Under Assumption 1-4, in Problem (3.0.1), if, (i) f is C^1, (ii) $D(s)$ is nonempty and uniformly compact near s, and (iii) SMFCQ hold for every optimal solution $a^*(s) \in A^*(s)$, then for any direction $x \in \mathbf{R}^m$*

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

Proof. Morand Reffett and Tarafdar ([45], Theorem 17) ■

4.2 Two Sector Growth with Nonconvexities in Labor Services

We consider and example of a two sector Uzawa growth model with nonconvexities in labor services (e.g., Prescott, Rogerson, and Wallenius [49]). The representative household has period preferences $u(c) = \ln c$, and discounts the future at the rate δ. The economy has two sectors, one for consumption goods, the other for investment goods. More specifically, in each period, sector 1 produces consumption goods $c_t \in \mathbf{K} \subset \mathbf{R}_+$, and sector 2 produces (next period) capital goods $k_{t+1} \in \mathbf{K} \subset \mathbf{R}_+$. The production functions of the two sectors are given by $f_i : \mathbf{R}_+^2 \to \mathbf{R}$, where $f_i(k_i, L_i)$ is assumed to be Cobb-Douglas in capital k_i and labor services L_i. The initial capital stock for the economy is $k_0 > 0$ and total labor endowment each period is normalized to unity. Capital at the beginning of each period can be allocated costlessly in amounts x_{1t} and x_{2t}, and let l_{1t} and l_{2t} be labor supply allocated to sectors 1 and 2 respectively.

We allow labor services in each period be given by $g_i : [0,1] \to \mathbf{R}$. We shall assume, for simplicity, $g_1(l_1) = L_1$ is convex-concave, but $g_2(l_2) = L_2$ is linear. In particular, we assume $g(l)$ is given by:

$$\begin{aligned} g_1(l_{1t}) &= l_{1t}^2, \quad \text{if } l_{1t} \le 0.25 \\ &= \frac{l_{1t}^{1/2}}{8}, \text{ if } l_{1t} \ge 0.25 \\ g_2(l_{2t}) &= l_{2t} \end{aligned}$$

Following Benhabib and Nishimura [8], we study the optimal decisions of this problem with a two-stage procedure. First, in any period t, the planner solves the following for k_t given,

$$V_t(k_t) = \max_{0 \le k_{t+1} \le f_2(} \{U(k_t, k_{t+1}) + \delta V_{t+1}(k_{t+1})\} \tag{4.2.1}$$

where,

$$U(k_t, k_{t+1}) = \max_{x,l} \{u(f_1(x_{1t}, g_1(l_{1t})))\}$$

subject to

$$\begin{aligned} x_{1t} + x_{2t} &\le k_t \\ l_{1t} + l_{2t} &\le 1 \\ k_{t+1} &\le f_2(x_{2t}, g_2(l_{2t})) \end{aligned}$$

By backward induction the planners problem first solves,

$$U(k_t, k_{t+1}) = \max \left\{ \begin{array}{ll} \left\{\frac{1}{2}\ln k_{1t} + \ln l_{1t}\right\}, & \text{if } l_{1t} \le 0.25 \\ \left\{\frac{1}{2}\ln k_{1t} + \frac{1}{4}\ln l_{1t} - \frac{3}{2}\ln 2\right\}, & \text{if } l_{1t} \ge 0.25 \end{array} \right\}$$

subject to

$$k_{t+1} \le (k_t - k_{1t})^{1/2}(1 - l_{1t})^{1/2}$$

Let λ be the multiplier. Here the constraint is continuously differentiable and the objective is Clarke regular. The value function of the second stage is directionally differentiable. To illustrate that $U(k_t, k_{t+1})$ is not C^1 we solve the above problem at $k_t = 1$. The optimal solution is given by, for $1 > k_{t+1} \geq \frac{3}{14^{1/2}}$

$$
\begin{aligned}
l_{1t}^* &= \frac{4 - k_{t+1}^2 - k_{t+1}(8 + k_{t+1}^2)^{1/2}}{4} \\
k_{1t}^* &= \frac{l_{1t}^*}{2 - l_{1t}^*} \\
\lambda &= \frac{(1 - l_{1t}^*)^{1/2}}{l_{1t}^*(1 - k_{1t}^*)^{1/2}}
\end{aligned}
$$

For $0.67082 \leq k_{t+1} \leq \frac{3}{14^{1/2}}$

$$
\begin{aligned}
l_{1t}^* &= 0.25 \\
k_{1t}^* &= 0.80 \\
\lambda &= \frac{1.25}{15^{1/2}}
\end{aligned}
$$

As is clear from the optimal solutions, $U(1, k_{t+1})$ is not continuously differentiable but directional derivatives exist. The directional derivative of $U(1, k_{t+1})$ with respect to k_{t+1} is given by,

$$
\begin{aligned}
U'_{k_{t+1}}(k_{t+1}; d) &= -\frac{1.25d}{15^{1/2}}, \text{ if } d \leq 0 \\
&= -2(21)^{1/2}d, \qquad \text{if } d \geq 0
\end{aligned}
$$

Thus to solve Problem (1), we need the envelope theorem for directional differentiability.

4.3 Stackelberg Models

This is a simple duopoly game where the leader's marginal cost is lower than the followers. In this game the second period best response function is kinked and therefore we cannot apply the standard first order condition to solve the leader's problem.

Example: the leader, firm 1 chooses quantity q_1 in period 1 and the follower, firm 2 chooses quantity q_2 in period 2. Both firms face a constant marginal cost $c_1 = 2$ and $c_2 = 3$. The inverse demand function is given by,

$$p = 5 - q_1 - q_2$$

We solve the game by backward induction. In the second period firm 2 maximizes:

$$\Pi_2(q_2; q_1) = (5 - q_1 - q_2)q_2 - 3q_2$$

The first order condition is,

$$\begin{aligned} 5 - q_1 - 2q_2^* - 3 &= 0 \text{ if } q_2 > 0 \\ &\leq 0 \text{ if } q_2 = 0 \end{aligned}$$

Solving,

$$\begin{aligned} q_2^* &= \frac{2 - q_1}{2} \text{ if } q_1 \leq 2 \\ &= 0 \qquad\quad \text{if } q_1 > 2 \end{aligned}$$

Second period value function is given by,

$$\begin{aligned} \Pi_2^*(q_1) &= \frac{(2 - q_1)^2}{4} \text{ if } q_1 \leq 2 \\ &= 0 \qquad\qquad \text{if } q_1 > 2 \end{aligned}$$

The best response function and the value function of period 2 are not C^1 but Lipschitz continuous, Clarke regular. This leads to just a directionally differentiable objective function in the first period,

$$\begin{aligned} \Pi_1(q_1) &= (5 - q_1 - q_2^*)q_1 - 2q_1 \\ &= (3 - q_1 - q_2^*)q_1 \end{aligned}$$

Substituting q_2^* from above,

$$\begin{aligned} \Pi_1(q_1) &= \frac{(4 - q_1)q_1}{2} \text{ if } q_1 \leq 2 \\ &= (3 - q_1)q_1 \quad \text{if } q_1 > 2 \end{aligned}$$

Since the objective function is just directionally differentiable the first order necessary condition is $\Pi'^{1}(q_1; d) \leq 0$ for $d \geq 0$ and $d \leq 0$.

Here $\Pi_1(2) = 2$. Now at $q_1 = 2$ for $d \leq 0$

$$\Pi_1'(q_1 = 2; d) = 0$$

For for $d \geq 0$

$$\begin{aligned} \Pi_1'(q_1 &= 2; d) = \lim_{t \to 0} \frac{\Pi^1(2 + td) - \Pi^1(2)}{t} \\ &= \lim_{t \to 0} \frac{(3 - (2 + td))(2 + td) - 2}{t} \\ & \lim_{t \to 0} \frac{6 + 3td - 4 - 4td - t^2d^2 - 2}{t} \\ &= -d \leq 0 \end{aligned}$$

Thus the optimal solutions are,

$$\begin{aligned} q_1^* &= 2 \\ q_2^* &= 0 \end{aligned}$$

4.4 Entry Deterrence Model

We consider a two stage duopoly game with an incumbent and a potential entrant. In the first stage the incumbent chooses an investment level that reduces cost (increases demand). In the second stage the potential entrant either enters and the two firms compete in quantity or does not enter and the incumbent is the only active firm in the market. A higher investment by the incumbent in the first stage in cost reduction (increasing demand) makes the incumbent aggressive (passive) in the second stage, which results in softer (tougher) action or reduction in (increasing) quantity by the entrant. In quantity competition game a lower output by a rival is beneficial. Thus, for strategic reason the incumbent over-invests (under-invests) in stage 1. According to Fudenberg and Tirole (1984, AER) this strategy is called Top dog (Lean and Hungry look). Whether the incumbent wants to accommodate or deter entry it will always over-invest (under-invest) to signal tough (soft) competition to the rival when investment is in reducing cost (increasing demand). In the following two examples the incumbent (or firm 1) finds deterring entry to be it's best strategy. In these simple examples the best response function of the potential entrant (firm 2) is kinked shape at quantity zero. Thus the first period objective function is not continuously differentiable and we apply the directional differentiable envelope theorem to calculate the optimal investment level.

Example: (Cost reduction). Let the inverse demand function faced by firm 1 (incumbent) and 2 (potential entrant) in the second period be given by, $p = 10 - q_1 - q_2$, where p is the market price, q_1 is the quantity of the incumbent and q_2 is the quantity produced by the potential entrant. The constant marginal cost of the potential entrant is $c_2 = 6$. The incumbent can reduce it's marginal cost by investing in stage 1. The investment reduces the marginal cost by,

$$\begin{aligned} c_1(k) &= 3 - k^{3/16} \quad \text{if } 0 \leq k \leq 100 \\ &= 3 - 100^{3/16} \text{ if } k > 100 \end{aligned}$$

We solve the game by backward induction. In second period both firms maximize their profit by choosing optimal quantities with marginal cost c_1 and $c_2 = 6$. Thus the firms solve the following:

$$\max_{q_1} (10 - q_1 - q_2)\, q_1 - c_1 q_1$$
$$\max_{q_1} (10 - q_1 - q_2)\, q_2 - 6 q_2$$

The best response functions are given by

$$\begin{aligned} q_1 &= \frac{10 - c_1}{2} - \frac{q_2}{2} \text{ if } q_2 \leq 10 - c_1 \\ &= 0 \qquad\qquad\qquad \text{otherwise} \end{aligned}$$

$$\begin{aligned} q_2 &= 2 - \frac{q_1}{2} \text{ if } q_1 \leq 4 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

The equilibrium outcomes are,

$$\begin{aligned} q_1^* &= \frac{16 - 2c_1}{3} \text{ if } q_2 \leq 10 - c_1 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

$$\begin{aligned} q_2^* &= \frac{c_1 - 2}{3} \text{ if } q_1 \leq 4 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

The first period profit function of the incumbent firm is given by, (we restrict the problem to $k \leq 100$, this constraint is not binding so does not change the optimal solution)

$$\pi_{1I}(q_1^*(k), q_2^*(k), k) = (10 - q_1^*(k) - q_2^*(k))\, q_1^*(k) - \left(3 - k^{3/16}\right) q_1^*(k) - k$$

Here π_{1I} maps $\mathbf{R}_+^3 \to \mathbf{R}_+$ and is C^1. q_1^*, q_2^* are locally Lipschitz and Clarke regular function in $\mathbf{R}_+ \to \mathbf{R}_+$. Therefore from (Theorem 2.3.9, Clarke 1983) π_{1I} is Clarke regular and the generalized gradient of π_{1I} is given by,

$$\partial \pi_{1I} = \overline{CO}\left\{ \frac{\partial \pi_{1I}}{\partial q_1^*(k)} \zeta_{q_1^*(k)} + \frac{\partial \pi_{1I}}{\partial q_2^*(k)} \zeta_{q_2^*(k)} + \frac{\partial \pi_{1I}}{\partial c_1(k)} \frac{\partial c_1(k)}{\partial k} d + \frac{\partial \pi_{1I}}{\partial k} d \right\}$$

where $\zeta_{q_1^*(k)}$, $\zeta_{q_2^*(k)}$ are elements of the Clarke gradient of q_1^*, q_2^*. Therefore the directional derivative of π_{1I} and generalized directional derivative of π_{1I}

in direction d are equal and is given by,

$$\begin{aligned}\pi'_{1I}(q_1^*(k), q_2^*(k), k; d) &= \frac{\partial \pi_{1I}}{\partial q_1^*(k)} q_1^*(k; d) + \frac{\partial \pi_{1I}}{\partial q_2^*(k)} q_2^*(k; d) + \frac{\partial \pi_{1I}}{\partial c_1(k)} \frac{\partial c_1(k)}{\partial k} d \\ &\quad + \frac{\partial \pi_{1I}}{\partial k} d \\ &= 0 - q_1^*(k) q_2^*(k; d) + \frac{3k^{-13/16}}{16} q_1^*(k) d - d \qquad (4.4.1)\end{aligned}$$

The first term is zero from the second period first order condition. To evaluate above first we substitute c_1 in the equilibrium quantities $q_1^*(k)$ and $q_2^*(k)$ we get for $k \in [0, 100]$,

$$\begin{aligned}q_1^*(k) &= \frac{16 - 2\left(3 - k^{3/16}\right)}{3} \text{ if } q_2 \leq 10 - c_1 \\ &= 0 \qquad \text{otherwise}\end{aligned}$$

$$\begin{aligned}q_2^*(k) &= \frac{\left(3 - k^{3/16}\right) - 2}{3} \text{ if } q_1 \leq 4 \\ &= 0 \qquad \text{otherwise}\end{aligned}$$

Simplifying,

$$\begin{aligned}q_1^*(k) &= \frac{10 + 2k^{3/16}}{3} \text{ if } q_2 \leq 10 - c_1 \\ &= 0 \qquad \text{otherwise}\end{aligned}$$

$$\begin{aligned}q_2^*(k) &= \frac{1 - k^{3/16}}{3} \text{ if } q_1 \leq 4 \\ &= 0 \qquad \text{otherwise}\end{aligned}$$

Now we calculate the directional derivative of firm 2 equilibrium quantity of the first stage for all directions at $k = 1$. For $d > 0$

$$q_2'^*(1; d) = 0 \qquad (4.4.2)$$

For $d < 0$

$$q_2'^*(1; d) = -\frac{d}{16} \geq 0 \qquad (4.4.3)$$

Evaluating the directional derivative (4.4.1) for $k = 1$,

$$\pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) = \left\{0 - q_1^*(1) q_2^*(1; d) + \frac{3d}{16} q_1^*(1) - d\right\}$$

Substituting $q_1^*(1) = 4$, $q_2^*(1) = 0$ we get

$$\pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) = \left\{-4q_2^*(1; d) + \frac{3d}{4} - d\right\}$$

For $d > 0$, from expression (4.4.2)

$$\begin{aligned}\pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) &= -4(0) + \frac{3d}{4} - d \\ &= -\frac{d}{4} < 0\end{aligned}$$

For $d < 0$, from expression (4.4.3)

$$\begin{aligned}\pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) &= -4(-\frac{d}{16}) + \frac{3d}{4} - d \\ &= 0\end{aligned}$$

Therefore $k = 1$ satisfies the first order condition Thus optimal solutions are given by,

$$\begin{aligned}k &= 1 \\ q_1^* &= 4 \\ q_2^* &= 0\end{aligned}$$

Example: (Advertisements). Let the inverse demand function faced by firm 1 and 2 in the second period be given by, $p = a(k) - q_1 - q_2$, where p is the market price, q_1 is the quantity of the incumbent and q_2 is the quantity produced by the potential entrant. The constant marginal costs are $c_1 = 1$ and $c_2 = 3$. The incumbent can increase the demand of stage 2 by investing in stage 1. The investment increases the intercept of the demand curve by,

$$a(k) = 3 + 2k^{3/8}$$

For any investment in stage 1, the stage 2 best responses are,

$$\begin{aligned}q_1 &= \frac{a(k)-1}{2} - \frac{q_2}{2} \text{ if } q_2 \leq a(k) - 1 \\ &= 0 \qquad\qquad\qquad \text{otherwise}\end{aligned}$$

$$\begin{aligned}q_2 &= \frac{a(k)-3}{2} - \frac{q_1}{2} \text{ if } q_1 \leq a(k) - 3 \\ &= 0 \qquad\qquad\qquad \text{otherwise}\end{aligned}$$

The best response functions are Lipschitz continuous and Clarke regular. The equilibrium outcomes are,

$$\begin{aligned} q_1^* &= \frac{a(k)+1}{3} \text{ if } q_2 \leq a(k)-1 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

$$\begin{aligned} q_2^* &= \frac{a(k)-5}{3} \text{ if } q_1 \leq a(k)-3 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

The first period profit function of the incumbent firm is given by,

$$\pi_{1I}(q_1^*(k), q_2^*(k), k) = \left(3 + 2k^{3/8} - q_1^*(k) - q_2^*(k)\right) q_1^*(k) - c_1 q_1^*(k) - k$$

As in the last example the first stage incumbent profit function is Clarke regular. Thus the directional derivative is given by,

$$\begin{aligned} \pi'_{1I}(q_1^*(k), q_2^*(k), k; d) &= \left\{ \frac{\partial \pi_{1I}}{\partial q_1^*(k)} q_1^*(k; d) + \frac{\partial \pi_{1I}}{\partial q_2^*(k)} q_2^*(k; d) + \frac{\partial \pi_{1I}}{\partial a(k)} \frac{\partial a(k)}{\partial k} d + \frac{\partial \pi_{1I}}{\partial k} d \right\} \\ &= \left\{ 0 - q_1^*(k) q_2^*(k; d) + \frac{3dk^{-5/8}}{4} q_1^*(k) - d \right\} \end{aligned} \tag{4.4.4}$$

The first term is zero from the first order condition of stage 2. At $k = 1$, $q_1^*(1) = 2$ and

$$q_2^*(1; d) = \frac{1}{4} d \qquad \text{if } d > 0 \tag{4.4.5}$$

$$= 0 \leq 0 \qquad \text{if } d < 0 \tag{4.4.6}$$

Evaluating the directional derivative of the incumbent's first stage profit for $k = 1$,

$$\begin{aligned} \pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) &= -2(\frac{d}{4}) + \frac{3d}{4}(2) - d \\ &= 0 \qquad \text{if } d > 0 \end{aligned}$$

from expression (4.4.5)

$$\begin{aligned} \pi'_{1I}(q_1^*(1), q_2^*(1), 1; d) &= -2(0) + \frac{3d}{4}(2) - d \\ &= \frac{d}{2} < 0 \qquad \text{if } d < 0 \end{aligned}$$

by expression (4.4.6). Hence the optimal solution is given by,

$$\begin{aligned} k &= 1 \\ q_1^* &= 2 \\ q_2^* &= 0 \end{aligned}$$

5 Appendix

Proof of Lemma 2:

Proof. (i) The Lagrangian, $L(a, s, \lambda, \mu)$ corresponding to the optimization problem (3.0.1) is given by equation (3.1.1).

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

From ([34], Theorem 3.1) the first order condition with directional differentiable objective and constraints is given as, $\forall\ x_a \in \mathbf{R}^n$

$$f'_a(a^*(s), s; x_a) - \lambda^T g'_a a^*(s), s; x_a) - \mu^T h'_a a^*(s), s; x_a) \leq 0$$

If the constraint is C^1, we can sharpen the first order condition to $\forall\ x_a \in \mathbf{R}^n$,

$$\begin{aligned} & f'_a(a^*(s), s; x_a) - \big(\lambda^T \nabla g_a(a^*(s), s) + \mu^T \nabla h_a(a^*(s), s)\big) \cdot x_a \leq 0 \\ \Longrightarrow\ & f'_a(a^*(s), s; x_a) \leq \big(\lambda^T \nabla g_a(a^*(s), s) + \mu^T \nabla h_a(a^*(s), s)\big) \cdot x_a \end{aligned} \quad (5.0.7)$$

(ii) If the objective function $f(a, s)$ and the inequality constraints are locally Lipschitz, the Clarke generalized directional derivative of the Lagrangian exists. Thus, the first order necessary condition for each s satisfies,

$$0 \in \partial L_a(a^*(s), s, \lambda)$$

From ([5], Theorem 2.2),

$$0 \in \partial_a f(a^*(s), s, \lambda) - \lambda^T \partial g_a(a^*(s), s) - \mu^T \partial h_a(a^*(s), s)$$

When the constraints are C^1 in a the first order necessary condition reduces to,

$$0 \in \partial_a f(a^*(s), s) - \lambda^T \nabla g_a(a^*(s), s) - \mu^T \nabla h_a(a^*(s), s)$$

Thus there exist $\varsigma_{a^*(s)}(a^*(s), s) \in \partial_a f(a^*(s), s)$ such that

$$\varsigma_{a^*(s)}(a^*(s), s) = \lambda^T \nabla g_a(a^*(s), s) + \mu^T \nabla h_a(a^*(s), s) \quad (5.0.8)$$

(iii) Note by definition, $f'_a(a^*(s), s; x_a) \geq f_a^{-o}(a^*(s), s; x_a)\ \forall\ x_a \in \mathbf{R}^n$, hence the result follows. ■

Proof of Theorem 5:

Proof. (i) Given any ϵ and a direction x of perturbation, consider $\overline{y}$ satisfying the fundamental lemma. By the mean value theorem:

$$f(a^*(s) + t\overline{y}, s + tx) - f(a^*(s), s) = t(\varsigma_a(t), \varsigma_s(t)) \cdot (\overline{y}, x), \tag{5.0.9}$$

where $(\varsigma_a(t), \varsigma_s(t)) \in \overline{co}\{\cup_{x \in T} \partial f(x)\}$ with $T = [(a^*(s), s), (a^*(s)+t\overline{y}, s+tx]$, and:

$$\overline{g}(a^*(s)+t\overline{y}, s+tx) - \overline{g}(a^*(s), s) = t(\nabla_a \overline{g}(z(t)), \nabla_s \overline{g}(z(t))) \cdot (\overline{y}, x) \tag{5.0.10}$$

where $z(t) \in T$. By upper hemicontinuity of the generalized gradient, as $t \downarrow 0$, $(\nabla_a \overline{g}(z(t)), \nabla_s \overline{g}(z(t)))$ converges to $(\nabla_a \overline{g}(a^*(s), s), \nabla_s \overline{g}(a^*(s), s))$ and $(\varsigma_a(t), \varsigma_s(t))$ converges to cluster points in $\partial f(a^*(s), s)$. Let $(\varsigma_a, \varsigma_s)$ be any one of them; by our choice of $\overline{y}$ we have:

$$\nabla_a \overline{g}(a^*(s), s).\overline{y} + \nabla_s \overline{g}(a^*(s), s) \cdot x < 0$$

and

$$(\varsigma_a, \varsigma_s).(\overline{y}, x) > \inf_{\lambda \in K(a^*(s), s)} L_s^{-o}(a^*(s), s; \lambda; x)$$

Therefore for t small enough, by substituting above in (5.0.9) and (5.0.10) we have

$$\nabla_a \overline{g}(z(t)).\overline{y} + \nabla_s \overline{g}(z(t)) \cdot x < 0$$

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

This implies that $\overline{g}(a^*(s)+t\overline{y}, s+tx) = \overline{g}(a^*(s), s)+t(\nabla_a \overline{g}(z(t)), \nabla_s \overline{g}(z(t))).(\overline{y}, x) < 0$. Therefore, $a^*(s) + t\overline{y} \in D(s + tx)$, and also that

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

Since $a^*(s) + t\overline{y} \in D(s + tx)$, then it must be the case that $V(s + tx) \geq f(a^*(s) + t\overline{y}, s + tx)$, and we get that:

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

Therefore:

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

(ii) Here

$$\inf_{\lambda\in K} L_s^{-o}(a^*(s),s;x) = \inf_{\lambda\in K(a^*(s),s)} \left(\min_{\varsigma_s} \varsigma_s \cdot x - [\lambda^T \nabla_s \overline{g}(a^*(s),s)\cdot x]\right)$$

And note by definition of a max and min,

$$L_s^o(a^*(s),s;x) = -L_s^{-o}(a^*(s),s;-x)$$

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 always exists $a_n^*(s) \in D(s+t_nx)$ such that $V(s+t_nx) = f(a_n^*(s), s+t_nx)$. Since the sequence $\{a_n^*(s)\}$ is in a compact domain, there exists a convergent subsequence. So without loss of generality, assume that $a_n^*(s) \to a^*(s)$, and necessarily $a^*(s) \in D(s)$. By continuity of V, $V(s) = f(a^*(s), s)$.Next, let $\overline{y} = \overline{y}(-x, \epsilon)$ satisfy the fundamental lemma above at $a^*(s)$ for a direction $-x$ of perturbation, and let $a_n(s) = a_n^*(s)+t_n\overline{y}$. By the mean value theorem we have:

$$f(a_n(s),s) - f(a_n^*(s), s+t_nx) = t_n(\varsigma_a(t_n), \varsigma_s(t_n))\cdot(\overline{y}, -x) \tag{5.0.11}$$

where $(\varsigma_a(t_n), \varsigma_s(t_n)) \in \overline{co}\{\cup_{x\in T}\partial f(x)\}$ and $T = [(a_n(s), s), (a_n^*(s), s+t_n\overline{y})]$, and also:

$$\overline{g}(a_n(s),s) - \overline{g}(a_n^*(s), s+t_nx) = t_n(\nabla_a\overline{g}(z(t_n)), \nabla_s\overline{g}(z(t_n)))\cdot(\overline{y}, -x) \tag{5.0.12}$$

where $z(t_n) \in T$. As $n \to \infty$, $t_n \to 0$ and $a_n^*(s) \to a^*(s)$, by continuity of the gradient (smooth constraints), $(\nabla_a\overline{g}(z(t_n)), \nabla_s\overline{g}(z(t_n)))$ converges to $(\nabla_a\overline{g}(a^*(s), s), \nabla_s\overline{g}(a^*(s), s))$. Also, $(\varsigma_a(t_n), \varsigma_s(t_n))$ converges to cluster points in $\partial f(a^*(s), s)$. Let $(\varsigma_a, \varsigma_s)$ be any one of them; by our choice of $\overline{y}$ satisfying the fundamental lemma for direction $-x$, and where $\overline{y} = \overline{y}(\epsilon, -x)$, we have:

$$\nabla_a\overline{g}(a^*(s),s)\cdot\overline{y} - \nabla_s\overline{g}(a^*(s),s)\cdot x < 0$$

and:

$$(\varsigma_a, \varsigma_s).(\overline{y}, -x) > \inf_{\lambda} L_s^{-o}(a^*(s), s; -x) - \epsilon$$

Hence, for n large enough, we have:

$$\nabla_a \overline{g}(a_n^*(s), s) \cdot \overline{y} - \nabla_s \overline{g}(a_n^*(s), s) \cdot x < 0 \tag{5.0.13}$$

and:

$$(\varsigma_a(t_n), \varsigma_s(t_n)) \cdot (\overline{y}, -x) > \inf_{\lambda \in K} L_s^{-o}(a^*(s), s; -x) - \epsilon. \tag{5.0.14}$$

Inequalities (5.0.12) and (5.0.13) implies that $\overline{g}(a_n(s), s) - \overline{g}(a_n^*(s), s + t_n x) < 0$ implies that $a_n(s) \in D(s)$, and inequalities (5.0.11) and (5.0.14) imply:

$$f(a_n(s), s) - f(a_n^*(s), s + t_n x) > \inf_{\lambda \in K} L_s^{-o}(a^*(s), s; -x) - \epsilon.$$

Since $a_n(s) \in D(s)$ it must be that $V(s) \geq f(a_n(s), s)$ so we have that:

$$\begin{aligned}
\limsup_{t \to 0} \frac{V(s + tx) - V(s)}{t} &\leq \lim_{n \to \infty} \frac{f(a_n^*(s), s + t_n x) - f(a_n(s), s)}{t_n} \\
&\leq -\left[\inf_{\lambda \in K} L_s^{-o}(a^*(s), s; -x) - \epsilon \right] \qquad (5.0.15) \\
&= -\left[\begin{array}{l} \min_{\varsigma_s} \varsigma_s.(-x) - \\ \inf_{\lambda \in K(a^*(s), s)} [\lambda^T \nabla_s \overline{g}(a^*(s), s) \cdot (-x)] - \epsilon \end{array} \right] \\
&= \max_{\varsigma_s} \varsigma_s.(x) - \sup_{\lambda \in K(a^*(s), s)} [\lambda^T \nabla_s \overline{g}(a^*(s), s) \cdot x] + \epsilon \qquad (5.0.16)
\end{aligned}$$

Since ϵ is arbitrary small, we have:

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

(iii) Since MFCQ holds for all $a^*(s) \in D(s)$, from Theorem (5), we have:

$$\begin{aligned}
& \sup_{a^*(s) \in A^*(s)} \min_{(\lambda) \in K(a^*(s), s)} \left\{ L_s^{-o}(a^*(s), s; \lambda) x \right\} \\
\leq \; & D_+ V(s; x) \leq D^+ V(s; x) \\
\leq \; & \sup_{a^*(s) \in A^*(s)} \max_{(\lambda) \in K(a^*(s), s)} \{ L_s^o(a^*(s), s; \lambda) x \}
\end{aligned}$$

■

References

[1] Aliprantis, C, and K. Border. 1999. *Infinite Dimensional Analysis: A Hitchhiker's Guide,* Springer-Verlag Press.

[2] Amir, R. L. Mirman, and W. Perkins. 1991, One-sector nonclassical optimal growth: optimality conditions and comparative dynamics. *International Economic Review,* 32, 625-644.

[3] Amir, R. 1996. Sensitivity analysis of multisector optimal economic dynamics. *Journal of Mathematical Economics,* 25, 123-141

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

[5] Auslender, A. 1979.Differentiable Stability in Non Convex and Non Differentiable Programming. *Mathematical Programming Study.* 10, 29-41.

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

[7] Bonnans, J.F and A. Shapiro. 200. Perturbation Analysis of Optimization Problems.

[8] Benhabib, J. and K. Nishimura. 1985. Competitive equilibrium cycles. *Journal of Economic Theory,* 35, 284-306.

[9] Berge, C. 1963. *Topological Spaces,* MacMillan Press.

[10] Brown, D., W. Heller, and R. Starr. 1992. Two-part marginal cost pricing equilibrium: existence and efficiency. *Journal of Economic Theory,* 57(1), 52-72.

[11] Clarke, F. 1975. Generalized Gradient and Application. *Trans. American Mathematical Society, 205, 247-62.*

[12] Clarke, F. 1983. *Optimization and Nonsmooth Analysis.* SIAM Press.

[13] Cornet, B. 1983. Sensitivity analysis in Optimization. *Core Discussion paper No. 8322, Universite Catholique de Louvain, Louvain-la-Neuve, Belgiom.*

[14] Danksin, J. 1967. *The Theory of Max-Min.* Springer-Verlag, New York.

[15] Dechert, W. and K. Nishimura. 1983. A complete characterization of optimal growth paths in an aggregative model with a non-concave production function. *Journal of Economic Theory,* 31, 332-354.

[16] Dontchev, A. and R.T.Rockafellar. 2009. Implicit Functions and Solution Mappings.

[17] Edlin, A. and C. Shannon. 1998. Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics. *Econometrica,* 66, 1417-1425.

[18] Edlin, A. and C. Shannon. 1998. Strict monotonicity in comparative statics. *Journal of Economic Theory,* 81, 201-219.

[19] Farrell, M. 1959. The convexity assumption in the theory of competitive equilibrium. *Journal of Political Economy,* 1959, 67(4), 377-391.

[20] Fiacco, A.V. and J. Kyparisis. 1986. Convexity and Concavity of the Optimal Value Function in Parametric Nonlinear Programming. *Journal Of Optimization Theory and Application,* 48(1), 95-126.

[21] Fontanie, G. 1980. Subdifferential stability in Lipschitz programming. MS, Operations Research and Systems Analysis Center, University of North Carolina.

[22] Gauvin, J. and F. Dubeau. 1982. Differential properties of the marginal function in mathematical programming. *Mathematical Programming Studies,* 19, 101-119.

[23] Gauvin, J. and F. Dubeau. 1983. Some examples and counterexamples for the stability of nonlinear programming problems. *Mathematical Programming Studies,* 21, 69-78.

[24] Gauvin, J and J.W. Tolle. 1977. Differential Stability in Nonlinear Programming. *SIAM Journal of Control and Optimization,* 15, 294-311.

[25] Giorgi,G and S. Komlosi. 1992. Dini derivatives in Optimization -Part 1. *Decisions in Economics and Finance,* 15, 3-30

[26] Hinderer, K. 2005. Lipschitz Continuity of Value Functions in Markovian Decision Processes. *Math. Meth. Oper. Res.*, 62, 3–22.

[27] Hopenhayn, H. and E. Prescott. 1992. Stochastic Monotonicity and Stationary Distributions for Dynamic Economies, *Econometrica*, 60(6), 1387-1406.

[28] Jeyakumar,V., D.T Luc and S. Schaible. 1998. Characterizations of Generalized Monotone Nonsmooth Continuous Maps using Approximate Jacobians. *Journal of Convex Analysis* 5(1) 119-32.

[29] Kehoe, T.K., D.K. Levine and P.M. Romer 1990. Determinacy of equilibria in dynamic models with finitely many consumers. *Journal of Economic Theory,* 50(1), 1-21

[30] Kamihigashi, T. and S. Roy. 2006. Dynamic optimization with a nonsmooth, nonconvex technology: the linear objective case. *Economic Theory,* 29, 325-340.

[31] Kamihigashi, T. and S. Roy. 2007. A nonsmooth, nonconvex, model of optimal growth. *Journal of Economic Theory,* 132, 435-460.

[32] Kelley, J. 1955. *General Topology.* Van Nostrand Press.

[33] Khan, A. and J. Thomas. 2008. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica,* 76(2), 396-436.

[34] Khanh,P.Q and N.D Tuan. 2007, Optimality Conditions for Nonsmooth multiobjective Optimization Using Hadamard Directional Derivatives. *Journal of Optimization Theory and Application,* 133, 341-57.

[35] Klatte, D. and B. Kummer. 2002. Nonsmooth equations in optimization : regularity, calculus, methods, and applications.

[36] Koopmans, T. 1961. Convexity assumptions, allocative efficiency, and competitive equilibrium. *Journal of Political Economy,* 69(5), 478-479.

[37] Kuratowski, K. 1968. *Topology,* Academic Press.

[38] Kyparisis, J. 1985. On the uniqueness of Kuhn-Tucker Multiplier in Nonlinear Programming. *Mathematical Programming, 32, 242-246.*

[39] Laraki, R. and W. Sudderth. 2004. The preservation of continuity and Lipschitz continuity of optimal reward operators. *Mathematics of Operations Research,* 29, 672-685.

[40] Li Calzi, M. and A. Veinott, Jr. 1992. Subextremal functions and lattice programming. MS. Stanford University.

[41] Milgrom, P. and I. Segal. 2002. Envelope theorems for arbitrary choices. *Econometrica,* 70, 583-601.

[42] Milgrom, P. and C. Shannon. 1994. Monotone comparative statics. *Econometrica,* 62, 157-180.

[43] Mirman, L., O. Morand, and K. Reffett. 2008. A qualitative approach to Markovian equilibrium in infinite horizon economies with capital. *Journal of Economic Theory, 139(1), 75-98.*

[44] Mirman, L. and I. Zilcha. 1975. On optimal growth under uncertainty. *Journal of Economic Theory,* 11, 329-339.

[45] Morand, O, K. Reffett, and S. Tarafdar, 2009b. A Nonsmooth Approach to Envelope Theorems. MS, Arizona State University.

[46] Morand, O, K. Reffett, and S. Tarafdar, 2009b. Lipschitzian Stochastic Dynamic Programming. MS, Arizona State University.

[47] Nishimura, K. and J. Stachurski 2005. Stability of stochastic optimal growth models: a new approach. *Journal of Economic Theory,* 122(1), 100-118.

[48] Nishimura, K. R. Rudnicki, and J. Stachurski. 2006. Stochastic optimal growth with nonconvexities. *Journal of Mathematical Economics.* 42(1), 74-96.

[49] Prescott, E., R. Rogerson, and J. Wallenius. 2009. Lifetime aggregate labor supply with endogeneous workweek lengthn. *Review of Economic Dynamics,* 12(1), 23-36.

[50] Quah, J. 2007. The comparative statics of constrained optimization problems. *Econometrica.*

[51] Reiter, S. 1961. A note on convexity of the aggregate production set. *Journal of Political Economy,* 69(4), 386-387.

[52] Rincon-Zapatero, J. and M. Santos. 2009. Differentiability of the value function without interiority assumptions. *Journal of Economic Theory, 144(5), 1948-1964.*

[53] Rockafellar, R.T. *Convex Analysis.* Princeton Press.

[54] Rockafellar, R.T. and R. Wets. 1998. *Variational Analysis,* Springer.

[55] Rogerson, R. and J. Wallenius. 2008. Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory,*

[56] Romer, P. 1990. Endogeneous Technical Change. *Journal of Political Economy, 98(5), S71-S102*

[57] Romer, P. 1990. Are nonconvexities important for understanding growth? *The American Economic Review,* Papers and Proceedings of the Hundred and Second Meeting of the American Economic Association,80(2), 97-103.

[58] Rothemberg, J. 1960. Nonconvexity, aggregation, and Pareto optimality. *Journal of Political Economy,* 1960, 454.

[59] Rubinov, A and X Yang. 2003. Lagrange-Type Functions in Constrained Non-Convex Optimization.

[60] Samuelson, P. 1947. *Foundations of Economic Analysis,* Cambridge Press.

[61] An Ordinal Theory of Games with Strategic Complementarities, *Working Paper , June 1990.*

[62] Topkis, D. 1998. *Supermodularity and Complementarity.* Princeton Press.

[63] Veinott, A. 1992. Lattice programming: qualitative optimization and equilibria. MS. Stanford.

[64] Viner, J. 1931. Cost curves and supply curves. *Zeitschrift fur Nationalokonomie 3:* Reprinted in Readings in price theory. Homewood, Il. Richard D. Irwin, 1951