Department of Economics Working Paper Series
401 Sunset Avenue
Windsor, Ontario, Canada N9B 3P4
Administrator of Working Paper Series: Christian Trudeau
Contact: trudeauc@uwindsor.ca

Envelope Theorem without Differentiability

Yuntong Wang (University of Windsor)

Working paper 14 - 04

Envelope Theorem without Differentiability

Yuntong Wang*

October 14, 2014

*Department of Economics, University of Windsor, Windsor, Ontario, Canada N9B 3P4, E-mail: yuntong@uwindsor.ca, Fax: (519) 973-7096.

Abstract

Summary. We establish a new *envelope theorem* in which the choice variables are discrete and the objective function and the constraints are Lipschitz continuous with respect to the parameters. The parameters can be finite or infinite dimensional vectors in a Banach space. In an application, we revisit the principal-agent problem and derive a weaker first-order condition than the traditional one in the literature. In an insurance example, we use the condition to show an insurance contract that is discontinuous at some level of the loss.

Keywords and Phrases: Envelope theorem, Discrete choice set, Lipschitz continuity, Generalized gradients, Principal-agent problem.

JEL Classification Numbers: C61.

1 Introduction

Envelope theorem states that under certain conditions the optimal value function of a parametric optimization problem with or without constraints is differentiable and its derivatives (or partial derivatives) with respect to the parameters can be computed by the corresponding derivatives of the associated Lagrangian function of the problem. The theorem has many applications in the demand theory, the theory of production, and the general comparative analysis in economics.[1]

Recently, the theorem has been applied in information economics, especially in mechanism design (e.g., Milgrom and Roberts, 1988, Milgrom and Shannon, 1994, Athey, Milgrom, and Roberts, 2000). In these applications, the theorem has been generalized in various ways, for example, by relaxing some regularity properties on the choice set (Sah and Zhao, 1998, Milgrom and Segal, 2002). These generalizations are useful because they make the theorem more applicable in a wider range of problems.

[1]See Löfgren (2011) for a review on how Envelope Theorems have been used in economics. For a review on the recent applications in optimal control theory, see Caputo (1990). For a review on earlier applications since Samuelson (1947) and Viner (1931), see Silberberg (1974).

In these generalizations, they all require that the objective function and the constraints are differentiable with respect to the parameters of the problem. However, it is not hard to see that the differentiability of these functions is neither necessary nor sufficient to guarantee the differentiability of the value function.

In this paper, we develop a new envelope theorem. We continue to assume that the choice set is a discrete set[2]as in Sah and Zhao (1998) and Milgrom and Segal (2002). But more importantly, we do not assume that the objective function and the constraints are differentiable with respect to the parameters. Instead, we assume that these functions are *Lipschitz* continuous. Moreover, the parameters can be either finite or infinite dimensional vectors in a Banach space.

As an application, we revisit the well studied *principal-agent problem*. We use the new envelope theorem to derive a weaker first-order condition than the traditional one in the literature. Roughly speaking, for an optimal contract, the first-order condition is satisfied for almost all outcomes (e.g, losses in an insurance problem) except a few, the set of which has a measure of zero.[3] In other words, the condition can be violated on some outcomes. Thus, the optimal incentive-contract can be a discontinuous function of outcomes rather than it has to be a continuous function implied by the traditional first-order condition.[4]

The concept of *generalized gradient*, first introduced by Clarke (1983), plays an important role in our generalization. This concept has been used before in the literature (e.g., Wang, 2007; Clausen and Strub, 2012). However, this paper is the first to allow non-differentiable functions in the envelope theorem.[5]

[2]We could easily extend our result to arbitrary set but we choose not to do it here.

[3]See Section 4 for the meaning of zero measure set.

[4]On the other hand, Mirrlees (1999) points out that there may exist many contracts that satisfy the first-order condition but are not optimal. Thus, replacing the incentive-compatibility condition by the first-order condition creates additional problems since it alters the set of feasible contracts. It is worth noting that there has been an interesting development in the literature of bilevel programming that is closely related to the principal-agent problem. For example, Ye and Zhu (2010) combine the first-order condition approach with the value function approach in a bilevel programming problem of which the principal-agent problem is a special case.

[5]In many generalizations of the envelope theorem, it is assumed that all functions, often including the value function, are differentiable. We argue that the differentiability of the

2 The Generalized Gradients

It is helpful to briefly introduce the Lipschitz functions and the generalized gradients of Lipschitz functions that we use in the paper. The reader can find more details on them in Clarke (1983).

Let X be a Banach space, whose elements x are called vectors or points and whose norm are denoted $\|x\|$, and whose open ball is denoted B. Let Y be a subset of X. A function $f : Y \to R$ is said to satisfy a Lipschitz condition (on Y) provided that, for some positive scalar L, one has

$$|f(y) - f(y')| \leq L\|y - y'\|, \forall y, y' \in Y. \tag{1}$$

We will say that f is Lipschitz of rank L near x if $Y = B(x)$, where $B(x)$ is an open ball centered at x.

Let f be Lipschitz near a given point x, and let v be any other vector in X. The *generalized directional derivative* of f at x in the direction v, denoted $f^\circ(x; v)$, is defined as

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \to 0+}} \frac{f(y + tv) - f(y)}{t}. \tag{2}$$

Proposition 1 *(Clarke, 1983, P.25) Let f be Lipschitz of rank L near x. Then, the function $v \to f^\circ(x; v)$ is finite, positively homogeneous, and subadditive on X, and satisfies*

$$|f^\circ(x; v)| \leq L\|v\|.$$

For convenience, for the rest of the paper we assume that the Banach space X is a Hilbert space. The well-known Separating-Hyperplane Theorem (Taylor, 1958; Balakrishnan, 1981) implies that any positively homogeneous

value function should be derived rather than assumed. Obviously, the differentiability of the value function is related to the differentiability of the functions in the problem but is not guaranteed by their differentiability. Additional constraints qualifications are needed. On the other hand, it is also easy to find parametric optimization problems in which all functions are not differentiable but the value function is differentiable.

and subadditive functional on X majorizes some linear functional on X.[6] Thus, under Proposition 1, there is at least one linear functional $\xi : X \to R$ such that, for all v in X, one has $f^\circ(x;v) \geq \langle \xi, v \rangle$. It follows that ξ is bounded, and belongs to the dual space X^* of continuous linear functionals on X. Now we can define the generalized gradient of f at x, denoted $\partial f(x)$, as the subset of X^* given by

$$\{\xi \in X^* : f^\circ(x;v) \geq \langle \xi, v \rangle, \forall v \in X\}. \tag{3}$$

Denote by $\|\xi\|_*$ the norm in X^*. We have the following proposition.

Proposition 2 *(Clarke, 1983, P.27) Let f be Lipschitz of rank L near x. Then*

1 $\partial f(x)$ is a nonempty, convex, weak-compact subset of X^* and $\|\xi\|_* \leq L$ for every ξ in $\partial f(x)$.*

2 For every v in X, one has

$$f^\circ(x;v) = \max\{\langle \xi, v \rangle : \xi \in \partial f(x)\}.$$

Example 1. Consider the Lipschitz function $f(x_1, x_2) = |x_1|+|x_2|, (x_1, x_2) \in R^2$. It is easy to check that

$$\partial f(0) = [-1, 1] \times [-1, 1].$$

Recall the *directional derivative* $f'(x;v)$ defined by

$$f'(x;v) = \lim_{t \to 0+} \frac{f(x+tv) - f(x)}{t}.$$

We say f is regular at x if

1. For all v, $f'(x;v)$ exists.
2. For all v, $f'(x;v) = f^\circ(x;v)$.

[6] This has been considered in the literature as a easy corollary of a related or equivalent Hahn-Banach theorem (Taylor, 1958). However, the author is not aware of any available proof. For completeness, here we sketch one below. Define a convex set $epi(f^\circ) = \{(y, v) : y \geq f^\circ(x, v)\}$. Consider the point $(y^*, v^*) = (0, 0) \in \partial epi(f^\circ)$ (the boundary set of the graph $epi(f^\circ)$). By the Separating Hyperplane Theorem, there exists a vector (λ, ξ) where $\lambda \neq 0$ and $\xi \neq 0$ such that $0 \leq \lambda f^\circ(x, v) + \langle \xi, v \rangle$. Rearranging terms completes the proof.

3 The Envelope Theorem

Now we establish our envelope theorem in which the choice variables are discrete vectors and the objective function and constraints are only assumed to be Lipschitz continuous in the parameter vector in a Hilbert space X (either finite or infinite dimensional). Formally, consider the following problem:

$$V(a) = \max f(x, a) \text{ s.t. } g^j(x, a) \leq 0, x \in D, j = 1, ..., m. \tag{4}$$

where x is a vector of choice variable in the discrete set $D = \{x_1, x_2, ..., x_d\} \subset R^n$, and a is a vector of parameters in X. Assume that, for any $x \in D$, $f(x, a)$ and $g^j(x, a), j = 1, ..., m$ are Lipschitz continuous functions of a with a common rank L.[7] Let $D(a)$ be the set of the maxima of (4). First, we have the following lemma.

Lemma 1 *For problem (4), for small change Δa in a, we have*[8]

$$D(a + \Delta a) = D(a), \text{ or } D(a + \Delta a) \subset D(a).$$

Proof. It suffices to show that, for any $a^k \to a, k \to \infty$ and $y^k \in D(a^k)$, $y^k \to y$, we have

$$y \in D(a).$$

By assumption, $y^k \in D(a^k)$, we have $y^k \in D$ and

$$g^j(y^k, a^k) \leq 0, j = 1, ..., m,$$

and

$$f(y^k, a^k) \geq f(x, a^k),$$

for all $x \in D$, s.t.

$$g^j(x, a^k) \leq 0, j = 1, ..., m.$$

Because $y^k \to y$, by the discrete nature of the set D, there must exist a K such that for all $k \geq K$,

$$y^k = y.$$

Therefore, we have

$$f(y, a^k) \geq f(x, a^k),$$

[7] If they all have different ranks, then choose the largest one as their common rank.

[8] We call the mapping $D(a)$ *upper semi-continuous*. See Berge (1963).

for all $x \in D$ and $g^j(x, a^k) \le 0, j = 1, ..., m, \ k \ge K$.

Taking limit $k \to \infty$, by the (Lipschitz) continuity of f and $g^j, j = 1, ..., m$ we obtain

$$f(y, a) \ge f(x, a),$$

for all $x \in D, g^j(x, a) \le 0, j = 1, ..., m$. This implies that

$$y \in D(a).$$

This proves the lemma.

Lemma 2 *Problem (4) is equivalent to the following problem*

$$\begin{aligned} V(a) &= \max_{\mu} \sum_{i \in I} \mu_i f(x_i, a), \qquad (5) \\ s.t. \quad & \sum_{i \in I} \mu_i g^j(x_i, a) \le 0, j = 1, ..., m, \\ & \sum_{i \in I} \mu_i = 1, \\ & \mu_i \ge 0, i \in I, \end{aligned}$$

where $I = \{i : x_i \in D(a)\}$.

Proof. First, note that in Lemma 1, it is shown that set I either remains the same or becomes smaller (subset) in a small neighborhood of a given a. This implies that the problem (5) is well-defined since we can always use the same I at a (I implicitly depends on a) if a varies little. Thus, we can consider problem (5) as an equivalent formulation of problem (4). This completes the proof.

Now we are ready to prove our main theorem.

Theorem 1 *Given problem (4), we have*

$$\partial V(a) \subseteq co\{\partial f(x, a), x \in D(a)\} - \cup_{j=1}^{m} cone\{\partial g^j(x, a), x \in D(a)\}, \qquad (6)$$

where "co" means the convex hull of a set and "cone" means the convex cone of a set, $\partial f(x, a)$ *and* $\partial g^j(x, a)$ *are the generalized gradients of* f *and* g^j *with respect to the parameter vector* a*. If* $f(x, a), g^j(x, a)$ *are regular at* a *for each* x *in* $D(a)$*, then equality holds and* V *is regular at* a *as well.*

Proof. Define

$$C = \{a : \sum_{i \in I} \mu_i g^j(x_i, a) \leq 0, j = 1, ..., m\},$$

and the indicator function of the set C,

$$\psi_C(a) = \begin{cases} 0 & \text{if} \quad a \in C \\ +\infty & \quad \text{otherwise} \end{cases}$$

Note that by the definition of (5), we have

$$V(a) = \sum_{i \in I} \mu_i f(x_i, a) - \psi_C(a),$$

where μ is any vector such that $\mu_i \geq 0, i \in I$ and $\sum_{i \in I} \mu_i = 1$.

Therefore,

$$\partial V(a) \subseteq \sum_{i \in I} \mu_i \partial f(x_i, a) - \partial \psi_C(a), \tag{7}$$

Since[9]

$$\partial \psi_C(a) = \{\sum_{j=1}^{m} \lambda_j \sum_{i \in I} \mu_i \partial g^j(x_i, a), \lambda_j \geq 0, j = 1, ..., m\},$$

replacing it into (7) proves the first part of the theorem.

The second part on the case when f and g^j are regular at a, can be proved using a similar argument as in Theorem 2.3.9 in Clarke (1983). We omit it.

This completes our proof.

It is easy to check that the following Sah and Zhao (1998) and Milgrow and Segal (2002) are corollaries of Theorem 1. Note that their envelope theorems consider only parametric optimization problems *without* constraints.

Corollary 1 *(Sah and Zhao, 1998) Consider the problem*

$$e(\theta) = f(n(\theta), \theta) = \max\{f(n.\theta) | n \in D\}, \tag{8}$$

where $\theta \in R$ is a continuous parameter and D is finite set of consecutive integers, and $f(n, \theta)$ is differentiable with respect to θ.

[9]See Clarke (1983).

Let $N(\theta)$ be the set of optimal integers for (8) for a given θ. Then, we have

$$\partial e(\theta) \subseteq co\{f_\theta(n,\theta) : n \in N(\theta)\}, \tag{9}$$

where $f_\theta(n,\theta)$ is the derivative with respect to θ. Moreover, if $f_\theta(n',\theta) = f_\theta(n'',\theta)$ for any $n', n'' \in N(\theta)$, we have $e_\theta(\theta) = f_\theta(n,\theta)|_{n \in N(\theta)}$.

Corollary 2 *(Milgrow and Segal, 2002) Let D be the choice set and $t \in [0,1]$ be the parameter. Let $f : D \times [0,1] \to R$ be the objective function. The value function V and the optimal choice set X^* are defined below:*

(1) $V(t) = \max_{x \in D} f(x,t)$.

(2) $X^*(t) = \{x \in D : f(x,t) = V(t)\}$.

Then,

$$[V'(t-), V'(t+)] \text{ or } [V'(t+), V'(t-)] \subset [\min_{x \in X^*(t)} f_t(x,t), \max_{x \in X^*(t)} f_t(x,t)]. \tag{10}$$

4 Revisiting the Principal-Agent Problem

We now apply Theorem 1 to the principal-agent problem. We use an insurance problem as an example.

Consider a single insurance company and a single consumer. The consumer might incur an accident resulting in a varying amount of loss, ranging from 0 dollar through $\overline{L}$ dollars, depending on the severity of the accident incurred. It is also possible that an accident is avoided altogether. It is convenient to refer to this latter possibility as an accident resulting in a loss of 0 dollars.

The probability distribution of incurring an accident resulting in a loss of l or less is given by $F(l,e) = \int_0^l f(l,e)dl$, where $f(l,e) > 0$ for all $l \in [0, \overline{L}]$ is the density function in which e is the amount of effort attempted to avoid the accident. Assume that there a finite number of effort levels from lowest $e = 0$ to the highest $e = M$.[10]

Assume that the consumer has a strictly increasing, strictly concave, von Neumann-Morgenstern utility function, $u(\cdot)$, over wealth, and the initial

[10] Holmstrom and Milgrom (1991) have considered multi-dimensional effort.

wealth equal to $w > \overline{L}$. In addition, $d(e)$ denotes the consumer's disutility of effort, which we assume is an increasing function in e. Thus, for a given effort level e, the consumer's von Neumann-Morgenstern utility over wealth is $u(\cdot) - d(e)$.

We assume that the insurance company can observe the amount of loss, l, due to an accident, but not the amount of accident avoidance effort, e. Consequently, the insurance company can only tie the benefit amount to the amount of loss. Let $B(l)$ denote the benefit paid by the insurance company to the consumer when accident loss is l. Thus, a *policy* is a pair $(p, B(l))$, where p denotes the price paid to the insurance company in return for guaranteeing the consumer $B(l)$ dollars if an accident loss of l dollars occurs. Note that a policy can also be expressed as a net benefit function $B^0(l) = B(l) - p$.

Our main question is: *What kind of policy will the insurance company offer the consumer*? Specifically, what are the optimal conditions that an optimal policy should satisfy?

Following the standard approach (e.g., Holmstrom [1979]), the insurance company's problem is

$$\max_{e,p,B(l)} p - \int_0^{\overline{L}} f(l,e)B(l)dl, \tag{11}$$

$$\text{s. t.} \int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) \geq \overline{u}, \tag{12}$$

$$\int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) \geq \int_0^{\overline{L}} f(l,e')u(w-p-l+B(l))dl - d(e'),$$
$$\forall e, e' \in \{0, 1, ..., M\} \text{ and } e \neq e', \tag{13}$$

where $\overline{u}$ is the consumer's reservation utility.

The above problem is a Bilevel Programming problem[11] and can be dealt with in the following two steps.[12]

[11] See Dempe (2002) for a survey on Bilevel Programming.

[12] This two-step approach reflects more clearly the moral hazard issue in the principal-agent problem, in which agent chooses his action after seeing the contract proposal from the principal. The traditional one-step approach in which the principal chooses both the target effort level on behalf of the agent and the contract that induces the targeted effort level blurs the implicit timing issue in contract design. Apparently, how we frame the problem might affect the solutions. See Mirrlees [1999, p.6] for an example on this issue. Our approach is more in line with Mirrlees.

First, for any given policy $(p, B(l))$, consider the following problem

$$\begin{aligned} V(p, B(l)) = & \max_{e \in \{0,1,...,M\}} p - \int_0^{\overline{L}} f(l,e)B(l)dl, \\ \text{s. t.} & \int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) \geq \overline{u}, \\ & \int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) \geq \int_0^{\overline{L}} f(l,e')u(w-p-l+B(l))dl - d(e'), \\ & \text{where } e, e' \in \{0,1,...,M\} \text{ and } e \neq e'. \end{aligned}$$

The Lagrangian of this problem is then

$$\begin{aligned} L &= p - \int_0^{\overline{L}} f(l,e)B(l)dl + \lambda[\int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) - \overline{u}], \\ &+ \sum_{e' \in \{0,1,...,M\}} \beta(e')[\int_0^{\overline{L}} f(l,e)u(w-p-l+B(l))dl - d(e) \\ &- \int_0^{\overline{L}} f(l,e')u(w-p-l+B(l))dl - d(e')], \end{aligned}$$

where λ, $\beta(e')$ are the multipliers corresponding to the constraints (12) and (13), respectively.

Next, the Insurance Company solves the following problem,

$$\max_{p, B(l)} V(p, B(l)). \tag{14}$$

The first-order conditions are

$$\begin{aligned} 0 &\in \partial_p V(p, B(l)) \\ 0 &\in \partial_B V(p, B(l)), \end{aligned}$$

where $\partial_p V$ and $\partial_B V$ denote the generalized gradients of V with respect to p and the function $B(\cdot)$, respectively.

For simplicity, assume that there exists a unique optimal effort level $e(p, B(l))$ for each policy $(p, B(l))$. Then, applying the envelope theorem (Theorem 1), we have

$$\begin{aligned} \partial_p V(p, B(l)) &= \partial_p L, \\ \partial_B V(p, B(l)) &= \partial_B L. \end{aligned}$$

Therefore,

$$0 \in \partial_p L,$$
$$0 \in \partial_B L.$$

Using the result of generalized gradients on integral functions[13], we have, for almost every $l \in [0, \overline{L}]$ except a set of l's with zero Lebesgue measure, that

$$0 = 1 - \lambda[\int_0^{\overline{L}} (f(l,e) + \sum_{e' \in \{0,1,\ldots,M\}} \beta(e')(f(l,e) - f(l,e')))u'(w - p - l + B(l))dl], \tag{15}$$

$$0 = -f(l,e) + [\lambda f(l,e) + \sum_{e' \in \{0,1,\ldots,M\}} \beta(e')(f(l,e) - f(l,e'))]u'(w - p - l + B(l)). \tag{16}$$

Note that the first of these conditions (15) is implied by (16). Thus, we focus on the latter.

To better compare with the traditional first-order condition, we look at the special case where there are only two effort levels 0 and 1. Then, the above first-order condition simplifies to the following familiar but slightly different result.

$$\frac{1}{u'(w - p + B(l) - l)} = \lambda + \beta[1 - \frac{f(l,0)}{f(l,1)}], \text{ L-a.e. for } l \in [0, \overline{L}]. \tag{17}$$

The traditional version of the above condition (i.e., for all outcomes l) is well-understood (see Jehle and Reny (2011) for an interpretation of the condition). Below, we show in an example that the new weaker condition allow contracts that do not satisfy the traditional first-order condition.[14]

[13] Consider an integral functional F on a Banach space X defined by $F(x) = \int f(l,x)dl$. Clarke (1983, pp. 75-76) shows that $\partial F(x) \subset \int \partial_x f(l,x)dl$ in the sense that, to every $\xi \in \partial F(x)$ there corresponds a mapping $l \to \xi_l$ from $[0, \overline{L}]$ to X^* with $\xi_l \in \partial_x f(l,x)$ almost everywhere relative to the Lebesgue measure (denoted L-a.e.) such that $\langle \xi, v \rangle = \int \langle \xi_l, v \rangle dl$ for all $v \in X$.

[14] Recall that the traditional first-order condition would restrict the solution $B(l)$ to be a differentiable function of l by the implicit function theorem. But in practice, many contracts are often only piece-wisely differentiable or even not continuous. For example, a wage contract could be a fixed salary plus a performance bonus after certain target has been met.

Consider the following example. Suppose the probability density function is as follows.

$$f(l,0) = \begin{cases} \frac{\delta}{l^*} & l \in [0, l^*) \\ \frac{1-\delta}{\overline{L}-l^*} & l \in [l^*, \overline{L}] \end{cases}$$

$$f(l,1) = \begin{cases} \frac{1-\epsilon}{l^*} & l \in [0, l^*) \\ \frac{\epsilon}{\overline{L}-l^*} & l \in [l^*, \overline{L}] \end{cases}$$

where $0 < l^* < \overline{L}$ and δ and ϵ are two small numbers such that

$$\frac{\delta}{1-\epsilon} < \frac{1-\delta}{\epsilon}.$$

Note first that $f(l,0)/f(l,1)$ is not continuous at l^*. By the first-order condition (17), we can show that there exist two constants d_1 and d_2 such that $d_1 < d_2$ and

$$B^0(l) = B(l) - p = \begin{cases} l - d_1, & \text{for} \quad l \in [0, l^*), \\ l - d_2, & \text{for} \quad l \in [l^*, \overline{L}]. \end{cases}$$

In words, with the full coverage contract, higher deductible is applied for losses higher than the threshold level l^*.[15] Note that the benefit function is not monotonic.

Unlike the traditional first-order condition which allows only continuous contract within the outcome domain, our condition allows discontinuous or even non-monotone contracts. Ever since Mirrlees (1999), many efforts[16] have been made to strengthen the first-order condition. One drawback of the traditional approach is that it may rule out meaningful contracts, some of which might depend on outcomes in a discontinuous way or non-monotonically.

Our "weaker" condition allows contracts that are not admissible under the traditional first-order approach. This is important since for the principal-agent problem, it is not about which contract should be selected, rather it is more about what *types* of contracts we can *possibly* choose from by ruling out those contracts that can never be optimal. For an optimization problem, trying to make the necessary first-order condition become also sufficient helps to find a solution. But, for the principal-agent problem, the question we

[15]It is assumed that $f(l,e)$ and l^* are common knowledge.

[16]See the recent Kirkegaard (2014).

should ask about is what types of contracts should never be considered. Therefore, we should "weaken" the necessary condition. The question of which contract we shall actually choose is of secondary importance.

References

[1] Athey, S., Milgrom, P., and Roberts, J.: Robust Comparative Statics. Princeton, Princeton University Press 2000

[2] Balakrishnan, A. V.: Applied Functional Analysis. Springer-Verlag, New York, Heidelberg, Berlin 1981

[3] Berge, C.: Topological spaces. Oliver and Boyd, London 1963

[4] Caputo, M. R.: How to Do Comparative Dynamics on the Back of an Envelope in Optimal Control Theory. J. Econ. Dynamics and Control **14**, 655-683 (1990)

[5] Clausen, A., Strub, C.: Envelope theorems for non-smooth and non-concanve optimization. Working Paper Series, Department of Economics, University of Zurich, No. **62**, (2012)

[6] Aubin, J. P.: Applied Functional Analysis. Wiley-Interscience, New York 1978

[7] Clarke, F. H.: A new approach to Lagrange multipliers. Math. Oper. Res. **1**, 165-174 (1976)

[8] Clarke, F. H.: Optimization and Nonsmooth Analysis. Wiley-Interscience Publication, New York 1983

[9] Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht 2002

[10] Grossman, S. J., Hart, O. D.: An analysis of the principal-agent problem. Econometrica **51**, 7-45 (1983)

[11] Holmstrom, B.: Moral hazard and observability. Bell J. Econ. **10**, 74-91 (1979)

[12] Holmstrom, B., Milgrom, P.: Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. J. Law, Econ. Organ. **7**, 24-52 (1987)

[13] Jehle, G. A., Reny, P. J.: Advanced Microeconomic Theory. Third Edition, Prentice Hall 2011

[14] Kirkegaard, R.: A Unifying Approach to Incentive Compatibility in Moral Hazard Problems. Working Paper, University of Guelph 2014

[15] Löfgren, K. G.: On Envelope Theorems In Economics: Inspired by a Revival of a Forgotten Lecture. Working Paper, University of Umeå2011

[16] Milgrom, P., Roberts, J.: Communication and Inventories as Substitutes in Organizing Production. Scandinavian J. Econ. **90**, 275-289 (1988)

[17] Milgrom, P., Segal, I.: Envelope Theorems For Arbitrary Choice Sets. Econometrica **70**, 583-601 (2002)

[18] Milgrom, P., Shannon, C.: Monotone Comparative Statics. Econometrica **62**, 157-180 (1994)

[19] Mirrlees, J. A.: The Theory of Moral Hazard and Unobservable Behaviour: Part I. Rev. Econ. Stud. **66**, 3-21 (1999)

[20] Sah, R., Zhao, J.: Some Envelope Theorems for Integer and Discrete Choice Variables. Int. Econ. Rev. **39**, 623-634 (1998)

[21] Samuelson, P. A.: Foundations of Economic Analysis. Cambridge: Harvard University Press 1947

[22] Silberberg, E.: A Revision of Comparative Statics Methodology in Economics, or, How to Do Comparative Statics on the Back of an Envelope. J. Econ. Theory **7**, 159-172 (1974)

[23] Taylor, A. E.: Introduction to Functional Analysis. John Wiley & Sons, In., New York 1958

[24] Viner, J.: Cost Curves and Supply Curves. Zeischrift für Nationalökonomie **3.1**, 26-46 (1931)

[25] Wang, Y.: Subgradients of Convex Games and Public Good Games. J. Convex Analysis **14**, 13-26 (2007)

[26] Ye, J. J., Zhu, D.: New Necessary Optimality Conditions for Bilevel Programs by Combining the MPEC and Value Function Approach. SIAM J. Optim. **20**, 1885-1905 (2010)