MEMORANDUM

No 26/2000

Sharpened nonsmooth maximum principle for control problems in finite dimensional state space

By
Atle Seierstad

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 25	By Geir B. Asheim: Deriving Belief Operators from Preferences. 6 p.
No 24	By Finn R. Førsund and Nikias Sarafoglou: On The Origins of Data Envelopment Analysis. 33 p.
No 23	By Knut Røed and Tao Zhang: A Note on the Weibull Distribution and Time Aggregation Bias. 9 p.
No 22	By Atle Seierstad: Nonsmooth maximum principle for control problems in Banach state space. 24 p.
No 21	By Diderik Lund: Imperfect loss offset and the after-tax expected rate of return to equity, with an application to rent taxation. 20 p.
No 20	By Christian Brinch: Identification of Structural Duration Dependence and Unobserved Heterogeneity with Time-varying Covariates. 19 p.
No 19	By Knut Røed and Morten Nordberg: Have the Relative Employment Prospects for the Low-Skilled Deteriorated After All? 21 p.
No 18	By Jon Vislie: Environmental Regulation under Asymmetric Information with Type-dependent outside Option. 20 p.
No 17	By Tore Nilssen and Lars Sørgard: Strategic Informative Advertising in a TV-Advertising Duopoly. 21 p.
No 16	By Michael Hoel and Perry Shapiro: Transboundary Environmental Problems with a Mobile Population: Is There a Need for Central Policy? 19 p.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/index.shtml

Sharpened nonsmooth maximum principle for control problems in finite dimensional state space.

by

Atle Seierstad, University of Oslo.

Department of Economics, Box 1095, 0317 Oslo, Norway

Abstract In a standard free end nonsmooth control problem in finite dimensional state space, a nonsmooth maximum principle is proved, in which the adjoint inclusion is sharper than the usual one. For end constrained problems, the same result holds, provided conditions ensuring local controllability are satisfied. The adjoint inclusion is expressed by means of a type of generalized gradient of the pseudoHamiltonian smaller than the standard one (Clarke's generalized gradient). From the results in this paper, one can recover the standard Pontryagin maximum principle in case of (not necessarily continuous) differentiability with respect to the state. (In end constrained problems, this still holds only if local controllability prevails.)

Mathematics Subject Classification 1991:
Primary 49K15;Secondary 49J52

Key word: Nonsmooth optimal control, nonsmooth maximum principle, smaller generalized gradients.

Introduction Various forms of nonsmooth maximum principles are found in the references, (Hestenes (1966) and Seierstad (1995) excepted). The aim of this paper is to obtain a sharper nonsmooth maximum principle, in the sense that the generalized gradient in the standard adjoint inclusion $-\dot{p}(t) \in \partial_x H(t, x^*(t), u^*(t), p(t))$, for example found in Clarke (1983), is replaced by a smaller generalized gradient. The problems considered here do not have restrictions on the time development of the state. First, free end problems are treated, then end constrained problems are briefly discussed. From the results in this paper, one can recover the standard Pontryagin maximum principle in case of (not necessarily continuous) differentiability with respect to the state. (In end constrained problems, this still holds only if conditions for local controllability hold.)

Notation and terminology For any locally Lipschitz continuous real-valued function $k(x)$ on $\mathbb{R}^n$, let $\partial k(x) = \partial_x k(x)$ be the generalized gradient of Clarke (1983), let $d^0 k(x)(v) = d^0_x k(x)(v)$ (rather than $k^0(x; v)$) be his generalized directional derivative, while $dk(x)(v) = d_x k(x)(v)$ is the ordinary (one sided) directional derivative. Define

$$d^\alpha k(x)(v) := \limsup_{t\searrow 0,} \sup_{y\in B(x,\alpha)}\{k(x + tv + ty) - k(x + ty)\}/t,$$
$$\tilde{d}k(x)(v) := \sup_{\alpha>0} d^\alpha k(x)(v),$$

and

$$\tilde{\partial} k(x) = \{x^* \in \mathbb{R}^n, \langle v, x^*\rangle \le \tilde{d}k(x)(v) \text{ for all } v\}.$$

(Evidently, $\tilde{d}k(x)(v) \le d^0 k(x)(v)$, further comments on $\tilde{d}$ and $\tilde{\partial}$ are given in Remark 4 below.) If h takes values in $\mathbb{R}^m, \partial h(x) = \partial_x h(x)$ is the generalized Jacobian of Clarke (1983). A set F of functions $f(t, x) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ is switching closed if $f 1_M(t) + (1 - 1_M(t))g \in F$ whenever $f, g \in F$ and M is a measurable set in $[0, T]$. Here1_M is the indicator function of the set M. $B(a, \alpha)$ is an open ball of radius α around a. The symbol cl denotes closure. It should be noted that $\tilde{\partial}$ is larger than the generalized gradient ∂_B of Treiman (1990).

The control problem

Consider the following control problem:

$$\max_{x(.),f}\ a\cdot x(T), a \text{ a fixed vector in } \mathbb{R}^n, \tag{1}$$

(where $a \cdot x$, also written $\langle a, x\rangle = \langle x, a\rangle$, means scalar product), subject to

$$\dot{x}(t) = f(t, x(t)) \text{ a.e. in } J := [0, T], x(0) = x_0,\ f \in F, x(.) \text{ absolutely continuous.} \tag{2}$$

Here, F is a fixed switching closed family of functions $f(t, x) : J \times \mathbb{R}^n \to \mathbb{R}^n$, and $x_0 \in \mathbb{R}^n$ and $T > 0$ are fixed. Below, $(x^*(.), f^*)$ is a pair satisfying (2). It is assumed that there exist positive constants M and ς such that, for all $f \in F$, for all t, $|f(t, x)| \leq M$ for all $x \in B(x^*(t), \varsigma)$ and $x \to f(t, x)$ is Lipschitz continuous on $B(x^*(t), \varsigma)$ with rank M. We assume that for any pair $(x(.), f)$ satisfying (2) and $x(t) \in B(x^*(t), \varsigma)$ for all t, we have that $a \cdot x^*(T) \geq a \cdot x(T)$, ($x^*(.)$ is called locally optimal in this case). The following theorem holds:

Theorem 1 There exists an absolutely continuous function $p(.)$, with $p(T) = a$, such that

$$\int_J \langle f(t, x^*(t)), p(t)\rangle dt \leq \int_J \langle f^*(t, x^*(t)), p(t)\rangle dt \text{ for all } f \in F, \tag{3}$$

and

$$-\dot{p}(t) \in \tilde{\partial}_x \langle f^*(t, x^*(t)), p(t)\rangle \text{ a.e.} \tag{4}$$

Of course, standard results says that $-\dot{p}(t)$ belongs to the (larger) set $\partial_x\langle f^*(t, x^*(t)), p(t)\rangle$. The proof of the theorem, as well as of the remarks to follow, is presented later on.

Remark 1 If the linear criterion $a \cdot x(T)$ is replaced by the criterion $\phi(x(T))$, where ϕ is Lipschitz continuous in $B(x^*(t), \varsigma)$, then replace the condition $p(T) = a$ by $p(T) \in \tilde{\partial}\phi(x^*(T))$. ,

(A proof is obtained from Theorem 1, by enlarging the interval to $[0, T + 1]$, and adding a new state variable y, governed by $\dot{y} = \phi(x(t))1_{[T,T+1]}, y(0) = 0$, with all f equal to 0 on $(T, T+1]$, and with criterion $y(T+1)$.)

Remark 2 In the case $f^*(t,x)$ has directional derivatives at $x^*(t)$, then (4) can be replaced by the stricter condition $-\dot{p}(t) \in [\partial_v\{d_x f^*(t, x^*(t))(v)\}]_{v=0}$ a.e.

(This assertion follows from the fact that in Lemma 4 below, $\gamma(s) \in [\partial_v\{d_x g(s, x^*(s))(v), \int_{(s,T]} d\nu(t)\}]_{v=0}$ a.e. in case $g(t,x)$ has directional derivatives at $x^*(t)$.)

Remark 3 Can the above nonsmooth maximum principle be extended to problems with end constraints? The answer is no in the general case, and yes in particular cases. The "no" in general, because the entity $\tilde{\partial}_x$ does not have the stability (upper semicontinuity in x) needed for a proof in general to go through. (That is, a proof using Ekeland's theorem and perturbed solutions being optimal in a perturbed problem is not working.) The "yes in particular cases", because sometimes an exact penalization result pertaining to the penalization of the end condition can be invoked.

To be more precise, add the requirement

$$\Pi x(T) = x_1 \in \mathbb{R}^m, \Pi : (x_1, ..., x_n) \to (x_1, ..., x_m), m < n. \tag{5}$$

Then, for some $K' > 0$, f^* is optimal in the free end problem (penalization problem) of maximizing $a \cdot x^f(T) - K'|\Pi x^f(T) - x_1|$ in D, provided certain local controllability conditions are satisfied. Hence, (using Remark 1),

Theorem 2 In the case where (5) is added to the problem, and one of the two types of local controllability conditions mentioned below holds, then the conclusion of Theorem 1 still holds, with the modification that $p(T) = a$ is replaced by $p(T) = a + (b, 0), (b, 0) \in \mathbb{R}^n, b$ some vector in $\mathbb{R}^m$.

To describe the two local controllability conditions mentioned, assume that $F = \{h(t, x, u(t)) : u(t) \in U \subset \mathbb{R}^{m^*}, u(t) \text{ measurable}\}$, where h is piecewise and left continuous in t, and Lipschitz continuous in u, with a rank independent of (t, x). Let $(x^*(.), u^*(.))$ be optimal. The two conditions are:

1. $h(t, x, U) = \text{co}h(t, x, U)$ for all (t, x), U is compact, and pseudonormality in the sense of Clarke (1983), p.224 holds.

2. Condition (4) in Theorem 3 in Seierstad (1995) holds. If U is convex, even weak variations are allowed in the latter condition, (then U must be closed as well).

Som further explanations of the conditions are given:

Pseudonormality is defined as follows: For any control $u^{**}(t)$ yielding $x^*(t)$, no multiplier $p(.)$ exists, with $p(T) = (b, 0) \in \mathbb{R}^n, b \in \mathbb{R}^m$, such that for a.e. t, $\langle h(t, x^*(t), u), p(t)\rangle \leq \langle h(t, x^*(t), u^{**}(t)), p(t)\rangle$ for all $u \in U$, and

$-\dot{p}(t) \in \partial_x \langle h(t, x^*(t), u^{**}(t)), p(t)\rangle$ a.e.

In the simple case where $h(t,x,u)$ has (one-sided) directional derivatives with respect to x in $B(x^*(t), \varsigma)$ for any (t,u), Condition 2, when employing only strong variations, reads as follows: For some $K'' > 0$, some $\delta \in (0,\varsigma/2e^{MT})$, for any unit vector $c \in \mathbb{R}^m$, for any $\hat{u}(t) \in \{u(.) : \text{meas}\{s : u(s) \neq u^*(t)\} \leq \delta\}$, for some finite collection of points $u_i \in U$, some time points $t_i \in (0,T)$ being regulated points of $\hat{u}(t)$, and some positive numbers υ_i, the inequality $c \cdot \Pi q(T) > K''\{\sum_i \upsilon_i\}$ holds, where $q(t)$ is the piecewise continuous solution of $\dot{q}(t) = d_x h(t, \hat{x}(t), \hat{u}(t))(q(t))$ a.e., $q(t_i+) - q(t_i-) = \sum_{j\in\{j:t_j=t_i\}} \upsilon_j[h(t_j, \hat{x}(t_j), u_j) - h(t_j, \hat{x}(t_j), \hat{u}(t_j))], q(0) = 0$. Here, $\hat{x}(.)$ is the solution corresponding to $\hat{u}(.), q(.)$ is continuous at all $t \notin \{t_i\}$.

For both Conditions 1 and 2, the following local controllability property holds: For some $\delta' > 0$, for all $\hat{x} \in \text{cl}B(x_1, \delta') \subset \mathbb{R}^m$, for some $f \in D$, $\Pi x^f(T) = \hat{x}$.

(From this result, optimality of f^* in the penalization problem follows by arguments similar to those employed in proving Lemma 1 below).

Proofs

In all proofs below, for simplicity, $x_0 = 0$ will be assumed. Define $D = \{f \in F : \sup_{x(.)\in B(x^*(.),\varsigma)} \int_0^T |f(t,x(t))dt - f^*(t,x(t))|dt \leq \varsigma/2K\}$, where $K = e^{MT}$ and $B(x^*(.), \varsigma)$ is a ball in $C(J, \mathbb{R}^n)$. By Gronwall's inequality and an existence and continuation argument, for any $f \in D' := \text{co}D$, a unique solution $x(t) = x^f(t) \in \text{cl}B(x^*(t), \varsigma/2)$ of the equation in (2) exists. Note that f^* is even optimal in D', (solutions $x^f(.)$, $f \in D'$ can be approximated uniformly by solutions $x^f(.)$, $f \in D$). Some lemmas are needed.

Lemma 1 (Exact penalization) The pair $(x^*(.), f^*)$ maximizes

$$\psi(x(.), f)) := a \cdot x(T) - K|a| \sup_{t\in J} |x(t) - \textstyle\int_0^t f(s, x(s))ds|$$

for $(x(.), f)$ in $\text{cl}B(x^*(.), \varsigma/2) \times D'$.

Proof Assume of the pair $(x(.), f) \in \text{cl}B(x^*(.), \varsigma/2) \times D'$ that $\psi(x(.), f) > \psi(x^*(.), f^*)$. There exists an absolutely continuous function $z(t) \in \text{cl}B(x^*(t), \varsigma/2)$ such that $\dot{z}(s) - f(s, z(s)) = 0$ a.e. Let $y(t) = x(t) - \int_0^t f(s, x(s))ds$. Then, by Gronwall's inequality, $\sup_t|z(t)-x(t)| \le K\sup_t|y(t)|$. Hence, $a\cdot x^*(T) \ge a\cdot z(T) \ge a\cdot x(T) - K|a|\sup_t|y(t)| = \psi(x(.),f) > \psi(x^*(.),f) = a\cdot x^*(T)$, a contradiction. Thus, $\psi(x(.),f) \le \psi(x^*(.),f^*)$.

Lemma 2 Let $a \in \mathbb{R}^m, C$ be a closed convex subset of $C(J,\mathbb{R}^n), g(t,x) : J\times\mathbb{R}^n \to \mathbb{R}^m$ be, separately measurable in t and, on $B(x^*(t),\varsigma)$ bounded by M and Lipschitz continuous in x of rank M, where $x^*(.)$ is a given function in C. Let P^* be the set of vector valued Radon measures ν taking values in $\mathrm{cl}B(0,1)\subset\mathbb{R}^m$. Let $\xi(x(.)) := \sup_t|\int_0^t g(s,x(s))ds|, x(.)\in C(J,X^n)$. Assume that $x^*(.)$ maximizes $a\cdot x(T)-\xi(x(.))$ in C. Then, for some measurable $\beta(t)\in\partial_x g(t,x^*(t))$ a.e. and some $\nu\in P^*$, we have $\zeta(v(.))\le 0$ for all $v(.)\in C-x^*(.)$, where

$$\zeta(v(.)) := a\cdot v(T) + \int_J(\int_0^t\beta(s)v(s)ds)d\nu(t) = a\cdot v(T) + \int_J\langle\beta(t)v(t), \int_{(t,T]}d\nu(\tau)\rangle dt.$$

(Below, we shall, essentially, apply Lemma 2 to a situation in which $\sup_t|\int_0^t g(s,x^*(s))ds| = 0$, so we drop additional information on ν).

Proof Choose a sequence of mollifiers ψ^i with respect to x, (see e.g. Rockafellar and Wets (1998), pp. 254, 409), vanishing outside $B(0,\varsigma/2i)$, with corresponding C^1- functions $g^i(t,x')$ defined for x' in $B(x^*(t),\varsigma/2)$, averaging values of $g(t,x)$, $x\in B(x',\varsigma/2i)\subset B(x^*(t),\varsigma)$, such that $|g^i(t,x')-g(t,x')| \le M\varsigma/2i$, for $x'\in B(x^*(t),\varsigma/2), t\in J$. Then, for $\xi^i(x(.)) := \sup_t|\int_0^t g^i(s,x(s))ds|$, we have $|\xi^i(x(.))-\xi(x(.))| \le T'/i$, where $T' := TM\varsigma/2$, for $x(.)\in B(x^*(t),\varsigma/2)$. Let $C' = C\cap clB(x^*(t),\varsigma/4)$. Using Ekeland's variational principle, choose for each i, elements $x^i(.)\in C'$, $|x^*(.)-x^i(.)|^\infty \le (T'/i)^{1/2}$, maximizing

$$a\cdot x(T) - \xi^i(x(.)) - (T'/i)^{1/2}\cdot|x(.)-x^i(.)|^\infty$$

in C'. Let $\eta(y(.)) = \sup_t|y(.)|, y(.)\in C(J,\mathbb{R}^m)$. Given any $y^*(.)\in C(J,\mathbb{R}^m), \partial_{y(.)}\eta(y^*(.)) \subset \{w(.)\to\int_J\langle w(t),z^*(t)\rangle d\mu(t) : \mu$ is some probability Radon measure, $z^*(t)$ some function such that $z^*(t)\in\mathrm{cl}B(0,1)$, $z^*(.)$ measurable with respect to $\mu\}$, see, for example, Clarke (1983), 2.8.2, Corollary 1, combined with the fact that $[\partial|y|]_{y=0}\subset\mathrm{cl}B(0,1)\subset\mathbb{R}^m, y\in\mathbb{R}^m$,
$[\partial|y|]_{y\neq0} = d|y|/dy \in \mathrm{cl}B(0,1)$. Moreover, using Chain rule II, 2.3.10 in Clarke (1983), for $y(.) = \int_0^\cdot g^i(s,x(s))ds$, $y^*(.) = \int_0^\cdot g^i(s,x^i(s))ds$, we get

$$\partial_{x(.)}\eta(y^*(.)) = \partial_{y(.)}\eta(y^*(.))\circ\int_0^\cdot g^i_x(s,x^i(s))ds \subset \{v(.)\to\int_J\langle\int_0^t g^i_x(s,x^i(s))v(s)ds, z^i(t)\rangle d\mu^i(t) :$$

μ^i is some probability Radon measure, $z^i(t)$ some function such that $z^i(t)\in\mathrm{cl}B(0,1)$,

$z^i(.)$ measurable with respect to $\mu^i\}$.

Thus, using optimality of $x^i(.)$, then for some $\nu^i, \nu_i \in P^*$, $\zeta^i(v(.)) \leq 0$ for any $v(.) \in C'' := C - x^*(.)$, where $\zeta^i(v(.)) := a \cdot v(T) + \int_J (\int_0^t g_x^i(s, x^i(s)) v(s) ds) d\nu^i(t) + (T'/i)^{1/2} \cdot \int_J v(s) d\nu_i(t)$, (the second term is obtained by letting $\nu^i(A) := -\int_A z^i(t) d\mu^i(t)$, the last term arises similarly.)

Using the Lipschitz rank M, (a bound on g_x^i), choose a subsequence i_j such that $g_x^{i_j}(., x^{i_j}(.))$ converges weakly (L_1, L_∞) to some limit $\beta(s)$, and such that ν^{i_j} $(\in C(J, \mathbb{R}^m)^*)$ converges weakly* to some limit ν. For simplicity, assume $(g_x^i(., x^i(.)), \nu^i)$ to converge in this manner. Then, $\beta(s) := \lim_i g_x^i(s, x^i(s)) \in \partial_x g(s, x^*(s))$ a.e. To see this, let a sequence of convex combinations $h_j(s) := \sum_{n=1}^{n_j} \theta_n^j g_x^{i_j^n}(s, x^{i_j^n}(s))$, $i_j^n \geq j$, of elements in the sequence $g_x^i(s, x^i(s))$ be L_1-, and even a.e. $-$ convergent, to $\beta(.)$. Then, by Rademacher's theorem, (see 9(38) in Rockafellar and Wets (1998)), for a.e. t, for any j,

$$h_j(t) = \sum_{n=1}^{n_j} \theta_n^j \int_{\mathbb{R}^n} g_x(t, x^{i_j^n} - z) \psi^{i_j^n}(z) dz \in \mathrm{clco} \cup_{x \in B(x^*(t), (T/j)^{1/2} + \varsigma/2j)} \partial_x g(t, x).$$

By upper semicontinuity of $x \to \partial_x g(t, x) = \mathrm{clco} \partial_x g(t, x)$, $\beta(t) = \lim_j h_j(t) \in \partial_x g(t, x^*(t))$, for a.e. t.

Moreover, using the Lipschitz rank M, $\{t \to \int_0^t g_x^i(s, x^i(s)) v(s) ds\}_i$ is equiuniformly continuous, so $\int_0^t g_x^i(s, x^i(s)) v(s) ds \to \int_0^t \beta(s) v(s) ds$ uniformly in t, and $\int_J (\int_0^t g_x^i(s, x^i(s)) v(s) ds) d\nu^i(t)$
$\to \int_J (\int_0^t \beta(s) v(s) ds) d\nu(t)$. Hence, $0 \geq \lim_i \zeta^i(v(.)) = \zeta(v(.))$, where $\zeta(v(.)) := a \cdot v(T) + \int_J (\int_0^t \beta(s) v(s) ds) d\nu(t) = a \cdot v(T) + \int_J \langle \beta(s) v(s), \int_{(s,T]} d\nu(t) \rangle ds$.

Lemma 3 Let $k(x)$ be real-valued and locally Lipschitz continuous on $\mathbb{R}^n$. Let $\check{d}k(x)(v) := \lim_j \sum_{n=1}^{n_j} \theta_n^j [k(x + \lambda_n^j v) - k(x)]/\lambda_n^j$, $\lambda_n^j \in (0, 1/j]$, $\theta_n^j \geq 0$, $\sum_n \theta_n^j = 1$, θ_n^j and λ_n^j being entities for which this limit, by assumption, holds for all v. Then $[d_v^0 \{\check{d}k(x)(v)\}(w)]_{v=0} \leq \tilde{d}k(x)(w)$.

Proof For any m, there exists a pair $(y, \mu) := (y_m, \mu_m) \in B(0, 1/m) \times (0, 1/m)$ such that $[d_v^0 \{\check{d}k(x)(v)\}(w)]_{v=0} - 1/m \leq [\check{d}k(x)(y + \mu w) - \check{d}k(x)(y)]/\mu$. Next, for all j large enough, $(j \geq j_{m,y,w})$, $|\check{d}k(x)(y + \mu w) - \sum_{n=1}^{n_j} \theta_n^j [k(x + \lambda_n^j (y + \mu w)) - k(x)]/\lambda_n^j| \leq 1/m$ and $|\check{d}k(x)(y) - \sum_{n=1}^{n_j} \theta_n^j [k(x + \lambda_n^j y) - k(x)]/\lambda_n^j| \leq 1/m$. Write $y_m' := y' := y/\mu$. Then, $[d_v^0 \{\check{d}k(x)(v)\}(w)]_{v=0} - 3/m \leq$

$\{\sum_{n=1}^{n_j}\theta_n^j[k(x+\lambda_n^j(y+\mu w))-k(x)]/\lambda_n^j-\sum_{n=1}^{n_j}\theta_n^j[k(x+\lambda_n^j y)-k(x)]/\lambda_n^j]/\lambda_n^j\}/\mu =$
$\{\sum_{n=1}^{n_j}\theta_n^j[k(x+\lambda_n^j(y+\mu w))-k(x+\lambda_n^j y)]/\lambda_n^j\}/\mu$
$=\{\sum_{n=1}^{n_j}\theta_n^j[k(x+\lambda_n^j\mu(y'+\mu w))-k(x+\lambda_n^j\mu y')]/\lambda_n^j\}/\mu$
$\le \sum_{n=1}^{n_j}\theta_n^j \sup_{\lambda\in(0,1/j]}\sup_{y''\le|y'|}\{k(x+\lambda\mu y''+\lambda\mu w)-k(x+\lambda\mu y'')\}/\lambda\mu =$
$\sup_{\lambda\in(0,1/j]}\sup_{y''\le|y'|}\{k(x+\lambda\mu y''+\lambda\mu w)-k(x+\lambda\mu y'')\}/\lambda\mu.$

This holds for all large j, so $[d_v^0\{\check{d}k(x)(v)\}(w)]_{v=0}-3/m \le$
$\limsup_{\lambda\searrow 0}\sup_{y''\le|y'|}\{k(x+\lambda\mu y''+\lambda\mu w)-k(x+\lambda\mu y'')\}/\lambda\mu = d^{|y'_m|}k(x)(w)\le$
$\tilde{d}k(x)(w)$.
Since this holds for all m, $[d_v^0\{\check{d}k(x)(v)\}(w)]_{v=0}\le \tilde{d}k(x)(w)$.

Lemma 4 Let $\xi(x^*(.))=0$ in Lemma 2. Then the inequality $0\ge\zeta''(v(.))$ holds for $v(.)\in C-x^*(.)$, where $\zeta''(v(.)):=a\cdot v(T)+\int_J\langle\gamma(t),v(t)\rangle dt$, for some $\nu\in P^*$ and some measurable $\gamma(s)\in\tilde{\partial}_x\langle g(s,x^*(s)),\int_{(s,T]}d\nu(\tau)\rangle$ a.e.

Proof For any $v(.)\in C(J,\mathbb{R}^n)$, there exists a sequence of numbers $\lambda_n\searrow 0$, such that $\triangle g(.,\lambda_n,v(.)) := [g(.,x^*(.)+\lambda_n v(.))-g(.,x^*(.))]/\lambda_n$ converges weakly (L_1,L_∞) to some weak limit $h^{v(.)}(.)$ $(=t\to h^{v(.)}(t))$. By diagonal selection, the sequence may even be taken to be one and the same for all $v(.)$ in a countable dense set V in $C(J,\mathbb{R}^n)$. For each $v(.)\in V$, a sequence of convex combinations $h_j(.,v(.))=\sum_{i=1}^{i_j}\theta_i^j\triangle g(.,\lambda_{n_i}^j,v(.))$, $\lambda_{n_i}^j\le 1/j$, converges in L_1-norm, to the weak limit $h^{v(.)}(t)$. By diagonal selection, for $\theta_i^j,\lambda_{n_i}^j$ suitably chosen, one obtains a sequence $\sum_{i=1}^{i_j}\theta_i^j\triangle g(.,\lambda_{n_i}^j,v(.))$ of convex combinations of elements from the sequence $\triangle g(.,\lambda_n,v(.))$ that converges in L_1-norm to $h^{v(.)}(.)$, for all $v(.)\in V$. In fact, we can even obtain pointwise convergence to $h^{v(.)}(.)$ on a set of full measure J', and for all $v(.)\in C(J,\mathbb{R}^n)$, J' independent of $v(.)$. For $t\in J'$, define $h(t,v)=h^{v(.)}(t)$, for $v(.)\equiv v\in\mathbb{R}^n$, and note that $h(t,v)$ is Lipschitz continuous in v of rank M. Moreover, by optimality of $x^*(t)$ and $\xi(x^*(.))=0$, for $v(.)\in C-x^*(.)$, $\lambda\in(0,1]$, we have $0\ge a\cdot(\lambda v(T)+x^*(T))-\sup_t|\int_0^t g(s,x^*(s)+\lambda v(s))ds|-\{a\cdot x^*(T)-\xi(x^*(.))\}=$ $\lambda a\cdot v(T)-\lambda\sup_t|\int_0^t\triangle g(s,\lambda,v(s))ds|$, or $0\ge a\cdot v(T)-\sup_t|\int_0^t\triangle g(s,\lambda,v(s))ds|$. Since, by the Lipschitz rank M, $\{t\to\int_0^t\triangle g(s,\lambda_n,v(s))ds\}_n$ is equiuniformly continuous, then
$\lim_{\lambda_n}\sup_t|\int_0^t\triangle g(s,\lambda_n,v(s))ds| = \sup_t|\int_0^t h(s,v(s))ds|$. Hence, $v(.)=0$ is optimal in the problem

$$\max_{v(.)\in C-x^*(.)} a\cdot v(T)-\sup_t|\int_0^t h(s,v(s))ds|.$$

Applying Lemma 2 gives $\zeta^*(v(.))\le 0$, $v(.)\in C-x^*(.)$, for $\zeta^*(v(.)):=a\cdot v(T)+$

$\int_J \langle \beta(t)v(t), (\int_{(t,T]} d\nu(s)\rangle dt = a \cdot v(T) + \int_J \langle \int_{(t,T]} d\nu(s), \beta(t)v(t)\rangle dt$,
$\beta(s) \in [\partial_v h(s, v(s))]_{v=0}$ a.e. Write $w(t) = \int_{(t,T]} d\nu(s)$. Then, a.e., $w(t)\beta(t) \in [\partial_v \langle w(t), h(t,v)\rangle]_{v=0} = [\partial_v \langle h(t,v), w(t)\rangle]_{v=0}$. Now, $\langle h(t,v), w(t)\rangle = \check{d}_x \langle g(t, x^*(t)), w(t)\rangle(v)$, where $\check{d}\langle g(t, x^*(t)), w(t)\rangle(v) = \lim_j \langle \sum_{i=1}^{i_j} \theta_i^j \triangle g(t, \lambda_{n_i}^j, v), w(t)\rangle$. Hence, by Lemma 3,
$[d_v^0\{\langle h(t,v), w(t)\rangle\}(v')]_{v=0} = [d_v^0\{\check{d}_x \langle g(t, x^*(t)), w(t)\rangle(v)\}(v')]_{v=0} \leq \tilde{d}_x \langle g(t, x^*(t))(w(t)\rangle(v')$. Thus, a.e., $\gamma(t) := w(t)\beta(t) \in \tilde{\partial}_x \langle g(t, x^*(t)), w(t)\rangle$.

Lemma 5 In Lemmas 2 and 4, change $\xi(x(.))$ to $\xi(x(.)) = \sup_t |\int_0^t g(s, x(s))ds - \Pi x(t)|$ where $\Pi : (x_1, ..., x_n) \to (x_1, ..., x_m), m < n$. If $x^*(.)$ is optimal in C for this definition of $\xi(x(.))$, and $\xi(x^*(.)) = 0$, then $0 \geq \hat{\zeta}(v(.))$ holds for $v(.) \in C - x^*(.)$, where $\hat{\zeta}(v(.)) := a \cdot v(T) + \int_J \langle \gamma(t), v(t)\rangle dt - \int_J \Pi v(s) d\nu(s)$, for some $\nu \in P^*$ and some measurable $\gamma(s) \in \tilde{\partial}_x \langle g(s, x^*(s)), \int_{(s,T]} d\nu(\tau)\rangle$ a.e.

Proof The proof is an obvious modification of the proofs of lemmas 2 and 4, (note that for the present definition of $\xi(x(.))$, the functions $\zeta^i(v(.))$ in the proof of Lemma 2 would become equal to $a \cdot v(T) + \int_J [(\int_0^t g_x^i(s, x^i(s))v(s)ds) - \Pi v(t)] d\nu^i(t) + (T'/i)^{1/2} \int_J v(s) d\nu_i(t))$.

Lemma 6 Let the locally Lipschitz continuous real-valued function $a(x, y)$, $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ be linear in y. Then $\tilde{d}_{x,y} a(\hat{x}, \hat{y})(v, v') = \tilde{d}_x a(\hat{x}, \hat{y})(v) + a(\hat{x}, v')$.

Proof Let $(v, v') \in \mathbb{R}^n \times \mathbb{R}^m, (w, w') \in B(0, \alpha) \subset \mathbb{R}^n \times \mathbb{R}^m$. Note that
$[a(\hat{x} + \lambda w + \lambda v, \hat{y} + \lambda w' + \lambda v') - a(\hat{x} + \lambda w, \hat{y} + \lambda w')]/\lambda =$
$[a(\hat{x} + \lambda w + \lambda v, \hat{y}) - a(\hat{x} + \lambda w, \hat{y})]/\lambda + a(\hat{x}, \lambda v')/\lambda +$
$[a(\hat{x} + \lambda w + \lambda v, \lambda w')/\lambda - a(\hat{x} + \lambda w, \lambda w')/\lambda] + [a(\hat{x} + \lambda w + \lambda v, \lambda v')/\lambda - a(\hat{x}, \lambda v')/\lambda]$
where the two last square bracket terms are smaller than $K^*\lambda|v|$ and $K^*\lambda|w + v|$, respectively, K^* being the Lipschitz rank of $a(x, y)$ near $(\hat{x}, \hat{y})$. From this the assertion immediately follows.

Continued proof of Theorem 1 Let $F^0 = \{f^i(t, x)\}$ be a finite subset of D', with m members. Define $\Lambda^m := \{\lambda = (\lambda_1, ..., \lambda_m) : \lambda_i \in [0, \varsigma/4m]$. Define $g(t, x, \lambda) := f^*(t, x) + \sum_i \lambda_i [f^i(t, x) - f^*(t, x)]$, $C = \{\hat{v}(.) \in (C(J, \mathbb{R}^{n+m}) : \hat{v}_i(.) \in \mathrm{cl} B(x_i^*(t), \varsigma/4n), i \leq n$,
$(\hat{v}_{n+1}(t), ..., \hat{v}_{n+m}(t)) \in \Lambda^m, \hat{v}_i(.)$ constant for $i > n\}$. By Lemma 1, for $(x(.), \lambda) \in C, (x^*(.), 0)$ maximizes $a \cdot x(T) - K|a| \sup_t |\int_0^t g(s, x(s), \lambda)ds - x(t)|$, or $a' \cdot x(T) - \sup_t |\int_0^t g(s, x(s), \lambda)ds - x(t)|$, where $a' = a/K|a|$. By Lemma 5, for some $\nu \in P^*$,(with $m = n$), and some $\gamma(t) = (\gamma'(t), \gamma''(t)) \in \tilde{\partial}_{x,\lambda} \langle g(t, x^*(t), 0), w(t)\rangle$, $w(t) = \int_{(t,T]} d\nu(s)$, we have, for $\hat{v}(.) = (v(.), \lambda') \in$

$C-(x^*(.),0)$, that $0 \geq \tilde{\zeta}(v(.),\lambda')$, where $\tilde{\zeta}(v(.),\lambda') = a'\cdot v(T)+\int_J(\langle\gamma'(t),v(t)\rangle+\langle\gamma''(t),\lambda'\rangle)dt - \int_J v(s)d\nu(s)$. Now, by Lemma 6, and the obvious rule $\tilde{d}(h_1+h_2) \leq \tilde{d}h_1+\tilde{d}h_2$, at $x=x^*(t), \lambda=0$, we get

$$\tilde{d}_{x,\lambda}\langle g(t,x^*(t),0),w(t)\rangle(v,\lambda') \leq \tilde{d}_x\langle f^*(t,x^*(t)),w(t)\rangle(v)+\langle \mathbf{f}(t,x^*(t))\lambda',w(t)\rangle,$$

where $\mathbf{f}(t,x) = (f^1(t,x)-f^*(t,x),...,f^m(t,x)-f^*(t,x))$ and $\mathbf{f}(t,x)\lambda' = \sum_i \lambda'_i(f^i(t,x)-f^*(t,x))$, $([\tilde{d}_{x,\lambda}\sum_i \lambda_i\{f^i(t,x)-f^*(t,x)\}]_{x=x^*(t),\lambda=0}(v,\lambda') = \mathbf{f}(t,x)\lambda'$, by Lemma 6). Thus, a.e.

$$(\gamma'(t),\gamma''(t)) \in \tilde{\partial}_{x,\lambda}\langle g(t,x^*(t),0),w(t)\rangle \subset (\tilde{\partial}_x\langle f^*(t,x^*(t)),w(t)\rangle,\langle \mathbf{f}(t,x^*(t)),w(t)\rangle).$$

As $0 \geq \tilde{\zeta}(v(.),0)$ is satisfied for $\pm v(.)$, $0 = \tilde{\zeta}(v(.),0) = a'\cdot v(T)+\int_J\langle\gamma'(t),v(t)\rangle dt - \int_J v(s)d\nu(s)$. In particular, for $v(t) = \int_0^t \check{v}(s)ds, v_i(t) \in \mathrm{cl}B(0,\varsigma/4n)$, ($\check{v}_i(.)$ otherwise arbitrary, integrable), this equality is satisfied. Inserting this $v(t)$ and partially integrating the last term yield

$$0 = \int_J[\langle a'+w(t),\check{v}(t)\rangle+\langle\gamma'(t),v(t)\rangle]dt. \qquad (6).$$

But then, by duBois Reymond's lemma, see Hestenes (1966), p.50, for a.e. t, $a'+w(t) = \int_0^t\gamma'(s)dt$. Now, $w(t)$ is absolutely continuous in $(0,T)$, by the last equality. In fact, $\dot{w}(t) = \gamma'(t)$ a.e. Since $w(t) = \int_{(t,T]}d\nu(t)$, $\dot{w} = -\dot{\nu}(t)$, $t \in (0,1)$, and $w(t) = \int_{\{T\}}d\nu(t)+\int_{(t,T)}\dot{\nu}(s)ds = \int_{\{T\}}d\nu(t)-\int_{(t,T)}\dot{w}(s)ds$. Inserting this expression for $w(t)$ in (6) and partially integrating the term containing $\check{v}(.)$, we get $0 = (a'-w(T-))\cdot v(T) = 0$, hence $w(T-) = a'$. Then, for $p(t) = K|a|w(t-)$, we get the conclusions $p(T) = a$ and (4) in Theorem 1. Moreover, for $(0,\lambda') \in C-(x^*(.),0)$, $0 \geq \tilde{\zeta}(0,\lambda') = \int_J\langle\gamma''(t),\lambda'\rangle dt = \int_J\langle\sum_i\lambda'_i(f^i(t,x^*(t))-f^*(t,x^*(t))),w(t)\rangle dt$, so the maximum condition (3) follows for $f \in F^0$.

Let P^{F^0} be the nonempty set of absolutely continuous functions $p(.)$ satisfying $p(T) = a$, (4), and (3) for F replaced by F^0. This set is compact in the $|.|^\infty-$ topology: Given any sequence $p_n(.)$, a subsequence n_j exists such that $\{\dot{p}_{n_j}(.)\}_j \subset L_1(J,\mathbb{R}^n)$ is weakly convergent with limit $\dot{p}(.) \in L_1(J,\mathbb{R}^n)$ and $p_{n_j}(t)$ is $|.|^\infty-$ convergent to $\lim_j p_{n_j}(T)+\int_T^t\dot{p}(s)ds$. By $\tilde{\partial}_x\langle f^*(t,x^*(t)),p(t)\rangle = \mathrm{clco}\tilde{\partial}_x\langle f^*(t,x^*(t)),p(t)\rangle$ and upper semicontinuity of $y^* \to \tilde{\partial}_x\langle f^*(t,x^*(t)),y^*\rangle$, the weak limit $p(.)$ satisfies (4). (To see this, consider again a L_1-convergent sequence of convex combinations of elements from $\{\dot{p}_{n_j}(.)\}_j$.) Let $\mathcal{F}$ be the set of finite collections F^0. The sets P^{F^0}, $F^0 \in \mathcal{F}$, obviously have the finite intersection property. Hence, $\cap_{F^0\in\mathcal{F}}P^{F^0}$ is nonempty. An element $p(.) \in \cap_{F^0 \in \mathcal{F}} P^{F^0}$ evidently satisfies $p(T) = a$ and (3) and (4).

Remark 4 For any locally Lipschitz continuous real-valued function $k(x)$, we have $\tilde{d}k(x)(v) \leq d^0 k(x)(v)$ and $\tilde{\partial}k(x) \subset \partial k(x)$. If $k(x)$ is differentiable at $\hat{x}$, then $\tilde{\partial}k(\hat{x})$ is the one point set $\{dk(\hat{x})/dx\}$. These properties as well as all properties mentioned below hold even if x belongs to a real normed space X. (Then differentiability must be in the Frechet sense.)

The function $v \to \tilde{d}k(x)(v)$ is finite, positively linearly homogeneous, subadditive, and is Lipschitz continuous with rank K if $k(.)$ has this Lipschitz rank near x. Moreover, $\tilde{d}k(x)(-v) = \tilde{d}(-k(x))(v)$, and $\tilde{d}(k(x) + k^*(x))(v) \leq \tilde{d}k(x)(v) + \tilde{d}k^*(x)(v)$ if $k^*(.)$ is another locally Lipschitz continuous function. For any real $\lambda, \tilde{\partial}(\lambda k(x)) = \lambda \tilde{\partial}k(x)$. If x is a local maximum or a local minimum of $k(x)$, then $0 \in \tilde{\partial}k(x)$. Moreover, $\tilde{\partial}k(x)(v) \neq \emptyset$, by the Hahn-Banach theorem. From the theory of support functions, $x^* \in \tilde{\partial}k(x) \Leftrightarrow \langle v, x^* \rangle \leq \tilde{d}k(x)(v)$ for all v.

Lebourg's mean value theorem is satisfied even for ∂ replaced by $\tilde{\partial}$, (i.e. $k(y) - k(x) \in \langle \tilde{\partial}k(u), y - x \rangle$ for some $u \in [x : y]$,(the segment between x and y). We have the following chain rule: Let $F : X \to \mathbb{R}^n$ be Lipschitz continuous near x, let g: $\mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous near $F(x)$, and let $f(x) = g(F(x))$. Then $\tilde{\partial}f(x) \subset \text{cl}^*\text{co}\{\tilde{\partial}\langle F(x), y^* \rangle : y^* \in \partial g(F(x))\}$, cl* being weak* closure.

Of course, upper semicontinuity of $x \to \tilde{\partial}k(x)$ fails to hold. At least in finite dimensions, $\cap_{\varepsilon > 0} \cup_{x \in B(\hat{x}, \varepsilon)} \tilde{\partial}k(x) = \partial k(\hat{x})$, (cf. Clarke (1983), 2.5.1).

Proofs are given in the Appendix.

Appendix

Proofs of assertions in Remark 4 The proof of subadditivity of $\tilde{d}k$ is as follows: Let $v, w \in X$. For any α, $d^\alpha k(x)(v+w) := \limsup_{t \searrow 0,} \sup_{y \in B(0,\alpha)} \{k(x+ t(v + w) + ty) - k(x + ty)\}/t \leq$
$\limsup_{t \searrow 0} \sup_{y \in B(0,\alpha)} \{k(x + tv + t(w + y)) - k(x + t(w + y))\}/t+$
$\limsup_{t \searrow 0} \sup_{y \in B(0,\alpha)} \{k(x + t(w + y)) - k(x + ty)\}/t \leq$
$\limsup_{t \searrow 0} \sup_{\breve{y} \in B(0,\alpha+|w|)} \{k(x + tv + t\breve{y}) - k(x + t\breve{y})\}/t+$
$\limsup_{t \searrow 0} \sup_{y \in B(0,\alpha)} \{k(x + t(w + y)) - k(x + ty)\}/t \leq$

$\tilde{d}k(x)(v) + \tilde{d}k(x)(w)$

Positive linear homogeneity of $\tilde{d}$: For $\lambda > 0, \limsup_{t\searrow 0} \sup_{y\in B(0,\alpha)}\{k(x+t\lambda v+ty)-k(x+ty)\}/t = \limsup_{t\searrow 0} \sup_{y\in B(0,\alpha/\lambda)} \lambda\{k(x+t\lambda v+t\lambda y)-k(x+t\lambda y)\}/t\lambda$. Taking $\sup_\alpha$ on both sides yields the homogeneity.

Proof of $\tilde{d}(-k(x))(v) = \tilde{d}k(x))(-v)$. We have $d^\alpha k(x)(-v) =$
$\limsup_{t\searrow 0} \sup_{y\in B(0,\alpha)}\{k(x+t(-v)+ty) - k(x+ty)\}/t =$
$\limsup_{t\searrow 0} \sup_{y\in B(0,\alpha)}\{-k(x+tv+t(y-v)) + k(x+t(y-v))\}/t \leq$
$\limsup_{t\searrow 0} \sup_{\breve{y}\in B(0,\alpha+|v|)}\{-k(x+tv+t\breve{y})) - [-k(x+t\breve{y})]\}/t \leq \tilde{d}(-k(x))(v)$.
Hence, $\tilde{d}k(x)(-v) \leq \tilde{d}(-k(x))(v)$. By symmetry, $\tilde{d}(-k(x))(v) \leq \tilde{d}k(x)(-v)$.

A local extreme point of k is a local extreme point of $-k$, and if x is a local minimum of one of them, the generalized direction derivative $\tilde{d}$ of it is ≥ 0 for all v, which gives that 0 belongs to the generalized gradient $\tilde{\partial}$ of it.

For any α, $\sup_{t\in(0,\varepsilon),} \sup_{y\in B(0,\alpha)}\{k(x+tv+ty) - k(x+ty)\}/t \leq$
$\sup_{t\in(0,\varepsilon),\breve{y}\in B(0,\varepsilon\alpha)}\{k(x+tv+\breve{y}) - k(x+\breve{y})\}/t$, $(t\in(0,\varepsilon), y\in B(0,\alpha) \Rightarrow \breve{y} = \varepsilon y \in B(0,\varepsilon\alpha))$. Taking $\inf_{\varepsilon>0}$ on both sides yields $\tilde{d}k(x)(v) \leq d^0k(x)(v)$.

To show Lebourg's theorem, first it is proved that if $\phi(\tau) = k(x+\tau(y-x))$ and $v\in\mathbb{R}$, then $\tilde{d}\phi(\tau)(v) \leq \tilde{d}k(x+\tau(y-x))(v(y-x))$. To show this, note that

$\limsup_{t\searrow 0} \sup_{\breve{y}\in B(0,\alpha)}\{\phi(\tau+tv+t\breve{y}) - \phi(x+t\breve{y})\}/t =$
$\limsup_{t\searrow 0} \sup_{\breve{y}\in B(0,\alpha)}\{k(x+\tau(y-x)+tv(y-x)+t\breve{y}(y-x)) - k(x+\tau(y-x)+t\breve{y}(y-x))\}/t \leq$
$\limsup_{t\searrow 0} \sup_{\breve{y}\in B(0,\alpha|y-x|)}\{k(x+\tau(y-x)+tv(y-x)+t\breve{y}) - k(x+\tau(y-x)+t\tilde{y})\}/t \leq$
$\tilde{d}k(x+\tau(y-x))(v(y-x))$,

(The first two balls are balls in $\mathbb{R}$, the last one is a ball in X.) Next, define $\theta(\tau) = \phi(\tau) + \tau(\phi(0)-\phi(1))$. Then $\theta(0) = \theta(1)$ and there is an extreme point $\tau^* \in (0,1)$ of $\theta(.)$. Then $0 \in \tilde{\partial}\theta(\tau^*)$. Now, $\tilde{\partial}\phi(\tau^*) \subset \{\zeta\in\mathbb{R} : v\zeta \leq \tilde{d}k(x+\tau^*(y-x))(v(y-x))$ for all $v\in\mathbb{R}\}$. Now, for any ζ in the last set, let ζ^* be a linear functional defined on linspan$\{(y-x)\}$ by $\langle w,\zeta^*\rangle := v\zeta$ if $w = v(y-x)$. By Hahn-Banach's theorem, ζ^* can be extended to all X, with $w\zeta^* \leq \tilde{d}k(x+\tau^*(y-x))(w)$ holding for all $w\in X$, hence the extended ζ^* belongs to $\tilde{\partial}k(x+\tau^*(y-x))$. Thus, $\tilde{\partial}\phi(\tau^*) \subset \langle\tilde{\partial}k(x+\tau^*(y-x)), y-x\rangle$ and $0 \in \tilde{\partial}\theta(\tau^*) \subset \langle\tilde{\partial}k(x+\tau^*(y-x)), y-x\rangle + \phi(0) - \phi(1)$.

The chain rule is proved as follows. For each $i = 1, 2, ...$, let $t_i \in (0, 1/i)$, $y_i \in B(0, \alpha)$, have the property that $d^\alpha f(x)(v) + 1/i \geq \{f(x + t_i v + t_i y_i) - f(x + t_i y_i)\}/t_i \geq d^\alpha f(x)(v) - 1/i$. The mean value theorem furnishes a pair (u_i, y_i^*), $y_i^* \in (\mathbb{R}^n)^* = \mathbb{R}^n$, $y_i^* \in \partial g(u_i)$, $u_i \in [F(x + t_i y_i) : F(x + t_i v + t_i y_i)]$, the segment between $F(x + t_i y_i)$ and $F(x + t_i v + t_i y_i)$, such that

$$\{f(x+t_i v+t_i y_i) - f(x+t_i y_i)\}/t_i = \langle \{F(x+t_i v+t_i y_i) - F(x+t_i y_i)\}/t_i, y_i^*\rangle.$$

A subsequence $y_{i_j}^*$ of y_i converges to some limit $y_\alpha^* \in \partial g(F(x))$, $(u_i \to F(x))$. Lipschitz continuity gives the boundedness needed to conclude that
$d^\alpha f(x)(v) =$
$\lim_i \langle \{F(x + t_i v + t_i y_i) - F(x + t_i y_i)\}/t_i, y_i^*\rangle =$
$\lim_i \langle \{F(x + t_i v + t_i y_i) - F(x + t_i y_i)\}/t_i, y_\alpha^*\rangle \leq$
$\limsup_{t \searrow 0} \sup_{y \in B(0,\alpha)} \langle \{F(x + tv + ty) - F(x + ty)\}/t, y_\alpha^*\rangle = d^\alpha \langle F(x), y_\alpha^*\rangle(v)$.
Hence, $\tilde{d} f(x)(v) \leq \sup_{y^* \in \partial g(F(x))} \tilde{d}\langle F(x), y^*\rangle(v) = \sup_{y^* \in \partial g(F(x))} \sup\langle v, \tilde{\partial}\langle F(x), y^*\rangle\rangle$.
Let $\kappa(v) := \sup\langle v, K\rangle$, $K := \cup_{y^* \in \partial g(F(x))} \tilde{\partial}\langle F(x), y^*\rangle$. As $\tilde{d} f(x)(v) \leq \kappa(v)$, then
$\tilde{\partial} f(x)(v) \subset \mathrm{cl}^* \mathrm{co} K$.

Proof of $k(x)$ differentiable $\Rightarrow \tilde{\partial} k(x) \subset \{dk(x)/dx\}$. We have, for all $w \in \mathrm{cl}B(0, \alpha)$, that $k(x + \lambda v + \lambda w) - k(x + \lambda w) = k(x + \lambda v + \lambda w) - k(x) - [k(x + \lambda w) - k(x)] = k'(x)[\lambda v + \lambda w] - k'(x)[\lambda w]$ plus two second order term in λ, the sizes of which are independent of w. From this it follows that $d^\alpha k(x)(v) = k'(x)[v]$ and $\tilde{d} k(x)(v) = k'(x)[v]$.

More details of certain proofs

a. The existence in Lemma 4 of $h_j(t, v(.)) := \sum_{i=1}^{i_j} \theta_i^j \triangle g(., \lambda_{n_i}^j, v(.))$, L_1−convergent to $\mathrm{h}^{v(.)}(.)$ for each $v(.) \in V := \{v_k(.)\}$. There exists a sequence of convex combinations $h_{j_1}(., v_1(.))$, $j_1 = 1, 2, ...$, $h_{j_1}(., v(.)) = \sum_{i=1}^{i_{j_1}} \theta_i^{j_1} \triangle g(., \lambda_{n_i}^{j_1}, v(.))$, all $n_i \geq j_1 \geq 1$, $h_{j_1}(., v_1(.))$ converging in L_1- norm to $h^{v_1(.)}(.)$. In fact, we can arrange it so that $|h_{j_1}(., v_1(.)) - h^{v_1(.)}(.)|_1 \leq 1/j_1$, for all j_1. Evidently, $h_{j_1}(., v_2(.))$, $j_1 = 1, 2...$ converges weakly to $h^{v_2(.)}(.)$. Hence, a convex combination $h_{j_2}(., v_2(.)) = \sum_{i=1}^{i_{j_2}} \theta_i^{j_2} h_{j_{1_i}}(., v_2(.))$, all $j_{1_i} \geq j_2$, $j_2 = 2, 3, ...$, satisfies $|h_{j_2}(., v_2(.)) - h^{v_2(.)}(.)|_1 \leq 1/j_2$. Evidently, also $v_1(.)$ satisfies the same inequality. Moreover, we can write $h_{j_2}(., v(.)) = \sum_{i=1}^{i^{j_2}} \hat{\theta}_i^{j_2} \triangle g(., \lambda_{n^i}^{j_2}, v(.))$, $\hat{\theta}_i^{j_2} \geq 0$, $\sum_i \hat{\theta}_i^{j_2} = 1$, $n^i \geq j_2$. Continuing in this manner, for any m, for any $j_m = m, m{+}1, ...$, we can find a convex combination $h_{j_m}(., v(.)) = \sum_{i=1}^{i^{j_m}} \hat{\theta}_i^{j_m} \triangle g(., \lambda_{n^i}^{j_m}, v(.))$ such that $|h_{j_m}(., v_{m'}(.)) - h^{v_{m'}(.)}(.)|_1 \leq 1/j_m$ for all $m' \leq m$, $\lambda_{n^i} = \lambda_{n^i}^{j_m}$

satisfying $n^i \geq j_m$. Then, in particular for all m, for $j_m = m$, we get $|h_m(., v_{m'}(.)) - h^{v_{m'}(.)}(.)|_1 \leq 1/m$ for $m' \leq m$. By diagonal selection, a subsequence m_j has the property that for each $m', h_{m_j}(t, v_{m'}(t)) \to h^{v_{m'}(.)}(t)$, for all $t \in J'$, J' having full measure and being independent of m'.

b. In Lemma 4, $h_{m_j}(t, v(t)) \to h^{v(.)}(t)$ for all $v(.) \in C(J, \mathbb{R}^n)$, not only for $v(.) \in V$: Let $v(.) \in C(J, \mathbb{R}^n), \varepsilon > 0$, and let $t \in J'$. Choose $v_k(.) \in V$ such $|v(.) - v_k(.)|^\infty < \varepsilon/3M$. Next, choose m_j such that $|h_{m_j}(t, v_k(t)) - h^{v_k(.)}(t)| < \varepsilon/3$. By Lipschitz continuity of $v \to h_{m_j}(t, v)$ and $v(.) \to h^{v(.)}(t)$ of rank M, $|h_{m_j}(t, v(t)) - h_{m_j}(t, v_k(t))| < \varepsilon/3$, and $|h^{v(.)}(t) - h^{v_k(.)}(t)| < \varepsilon/3$. Hence, $|h_{m_j}(t, v(t)) - h^{v(.)}(t)| < \varepsilon$.

c. Upper semicontinuity of $y^* \to \tilde{\partial}\langle f^*(t, x^*(t)), y^*\rangle$, (used in Continued proof of theorem 1). It suffices to prove the closed graph property: Let $y_n \in \tilde{\partial}\langle f^*(t, x^*(t)), y_n^*\rangle$, $y_n \to y, y_n^* \to y^*$. We need to show that $y \in \tilde{\partial}\langle f^*(t, x^*(t)), y^*\rangle$. This follows if we prove $\langle v, y\rangle \leq \tilde{d}\langle f^*(t, x^*(t)), y^*\rangle(v)$ for all v. By the Lipschitz rank M, for any $\lambda > 0$, $|[f^*(t, x^*(t) + \lambda v + \lambda w) - f^*(t, x^*(t) + \lambda w)]/\lambda| \leq M|v|$, so for any $z^*, \check{z}^* \in \mathbb{R}^n$,

$$\langle [f^*(t, x^*(t) + \lambda v + \lambda w) - f^*(t, x^*(t) + \lambda w)]/\lambda, z^*\rangle \leq$$
$$\langle [f^*(t, x^*(t) + \lambda v + \lambda w) - f^*(t, x^*(t) + \lambda w)]/\lambda, \check{z}^*\rangle + M|v||z^* - \check{z}^*|.$$

Taking $\sup_a \limsup_{\lambda \searrow 0} \sup_{w \in B(0,\alpha)}$ on both sides yields $\tilde{d}_x\langle f^*(t, x^*(t)), z^*\rangle(v) \leq \tilde{d}_x\langle f^*(t, x^*(t)), \check{z}^*\rangle(v) + M|v||z^* - \check{z}^*|$. By symmetry, $\tilde{d}_x\langle f^*(t, x^*(t)), \check{z}^*\rangle(v) \leq \tilde{d}_x\langle f^*(t, x^*(t)), z^*\rangle(v) + M|v||z^* - \check{z}^*|$. Hence, $|\tilde{d}_x\langle f^*(t, x^*(t)), z^*\rangle(v) - \tilde{d}_x\langle f^*(t, x^*(t)), \check{z}^*\rangle(v)| \leq M|v||z^* - \check{z}^*|$.

Then, evidently, $\tilde{d}_x\langle f^*(t, x^*(t)), y_n^*\rangle(v) \to \tilde{d}_x\langle f^*(t, x^*(t)), y^*\rangle(v)$. Thus, from $\langle v, y_n^*\rangle \leq$
$\tilde{d}\langle f^*(t, x^*(t)), y_n^*\rangle(v)$, we obtain $\langle v, y^*\rangle \leq \tilde{d}\langle f^*(t, x^*(t)), y^*\rangle(v)$.

Proof of $\partial_B \subset \tilde{\partial}$: Using Theorem 2.2 in Treiman (1990), it suffices to show

$\sup_{\lambda > 0} \limsup_{x' \to x}[f(x' + \lambda|x' - x|h - f(x')]/\lambda|x - x'| \leq \tilde{d}(x)(v)$. Now, letting $\alpha = 1/\lambda$ and $t = \lambda|x' - x|$, $\sup_{x' \in B(x,\varepsilon)}[f(x' + \lambda|x' - x|h - f(x')]/\lambda|x - x'| =$
$\sup_{t \in (0,\varepsilon/\alpha)} \sup_{x' \in \{y:||y-x||=\alpha t\}}[f(x' + th - f(x')]/t \leq$
$\sup_{t \in (0,\varepsilon/\alpha)} \sup_{x' \in B(x,\alpha t)}[f(x' + th - f(x')]/t.$

From this the asserted inequality easily follows.

References

Clarke, F.H. (1983) *Optimization and Nonsmooth Analysis,* John Wiley & Sons, Toronto.

Clarke, F.H. (1989) *Methods of Dynamic and Nonsmooth Optimization,* CBMS *57,* SIAM, Philadelphia.

Clarke, F.H. et. al. (1998) *Nonsmooth Analysis and Control Theory,* Springer-Verlag,New York.

Mordukhovich, B. (1980) Metric approximations and necessary conditions for general classes of nonsmooth extremal problems, *Dokl. Akad. Nauk SSSR*, 254 pp. 1072-1076.

Hestenes M. R. (1966) *Calculus of Variations and Optimal Control Theory,* John Wiley & Sons, New York.

Seierstad, A. (1995) Directional derivatives and the control of nonsmooth systems, *Optimization*, 35, pp. 61-75.

Treiman.J.S. (1990) Optimal control with small generalized gradients. *SIAM J. Control and Optimization Vol 28, pp.720-732.*