Control and Cybernetics

vol. **36** (2007) No. 3

Composite semi-infinite optimization*

by

Darinka Dentcheva[1] and Andrzej Ruszczyński[2]

[1]Stevens Institute of Technology, Department of Mathematical Sciences
Hoboken, NJ, USA

[2]Rutgers University,
Department of Management Science and Information Systems
Piscataway, NJ 08854, USA

e-mail: darinka.dentcheva@stevens.edu, rusz@business.rutgers.edu

Abstract: We consider a semi-infinite optimization problem in Banach spaces, where both the objective functional and the constraint operator are compositions of convex nonsmooth mappings and differentiable mappings. We derive necessary optimality conditions for these problems. Finally, we apply these results to nonconvex stochastic optimization problems with stochastic dominance constraints, generalizing earlier results.

Keywords: semi-infinite optimization, nonsmooth optimization, composite optimization, stochastic programming, stochastic dominance.

1. Introduction

Let $\mathscr{X}$ and $\mathscr{Z}$ be Banach spaces, $\mathscr{Y}$ be a separable Banach space, and let T be a compact Hausdorff space. We consider the mappings: $F : \mathscr{X} \to \mathscr{Z}$, $\varphi : \mathscr{Z} \to \mathbb{R}$, $H : \mathscr{X} \to \mathscr{Y}$ and $G : \mathscr{Y} \times T \to \mathbb{R}$ and a set X_0 in $\mathscr{X}$. We focus on the optimization problem

$$\min \ \varphi(F(x)) \tag{1}$$

$$\text{s.t.} \ G(H(x), t) \leq 0 \quad \text{for all} \quad t \in T, \tag{2}$$

$$x \in X_0. \tag{3}$$

We assume that X_0 is convex, φ is convex and continuous, F is continuously Fréchet differentiable, G is continuous, $G(\cdot, t)$ is convex for all $t \in T$, and H is continuously Fréchet differentiable.

*This research was supported by the NSF awards DMS-0603728, DMS-0604060, DMI-0354500 and DMI-0354678.

Our objective is to develop necessary and sufficient conditions of optimality for this problem.

Motivation for our research comes from stochastic optimization models with stochastic dominance constraints, introduced in Dentcheva and Ruszczyński (2003, 2004). In these problems, $\mathscr{Y} = \mathscr{L}_1(\Omega, \mathscr{F}, P)$ for some probability space $(\Omega, \mathscr{F}, P)$. The mapping H assigns to a decision vector x an integrable random variable $Y = H(x)$. The operator G is defined as follows:

$$G(Y,t) = \int_{\Omega} \max(0, t - Y(\omega))\, P(d\omega) - v(t), \quad t \in [a,b] \subset \mathbb{R}, \tag{4}$$

where $v(\cdot)$ is a given continuous function. In particular, we may use

$$v(t) = \int_{\Omega} \max(0, t - Y_0(\omega))\, P(d\omega)$$

for some benchmark random variable $Y_0 \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$. In this case, constraint (2) takes on the form of the second order *stochastic dominance relation*:

$$\int_{\Omega} \max(0, t - [H(x)](\omega))\, P(d\omega) \le \int_{\Omega} \max(0, t - Y_0(\omega))\, P(d\omega), \quad t \in [a,b].$$

This relation plays a crucial role in modeling risk averse preferences. It can be shown that it is equivalent to the following inequality:

$$\mathbb{E}[u(H(x))] \ge \mathbb{E}[u(Y_0)],$$

for all concave nondecreasing utility functions $u : \mathbb{R} \to \mathbb{R}$, satisfying a linear growth condition. The symbol $\mathbb{E}[\cdot]$ denotes the expected value, $\mathbb{E}[Y] = \int_{\Omega} Y(\omega)\, P(d\omega)$. For further details, see Dentcheva and Ruszczyński (2003) and the references therein.

Problem (1)–(3) is related to two well-established structures in optimization theory: semi-infinite optimization and composite optimization. In the analysis of semi-infinite problems, it is usually assumed that the semi-infinite constraints (2) are defined by a linear or continuously differentiable function $G(\cdot, t)$ and the space $\mathscr{X}$ is finite dimensional (see, *inter alia*, Bonnans and Cominetti, 1996; Bonnans and Shapiro, 2000, Section 5.4; Canovas et al., 2005; Goberna and Lopez, 1998; and Klatte and Henrion, 1998). In our case, these assumptions are not satisfied. The research on composite optimization focuses on the composite structure of the objective functional, as in (1) (see, *inter alia*, Bonnans and Shapiro, 2000, Section 3.4.1; Penot, 1994; Studniarski and Jeyakumar, 1995; and Yang, 1998).

In our setting, the main difficulty is associated with the infinite system of inequalities (2) of composite structure, which has not been investigated before.

2. Optimality conditions. The convex case

In this section we assume that the mappings F and H are affine:

$$F(x) = Ax + a, \qquad H(x) = Bx + b,$$

with bounded linear operators $A : \mathscr{X} \to \mathscr{Z}$ and $B : \mathscr{X} \to \mathscr{Y}$. This makes problem (1)–(3) convex:

$$\begin{aligned} \min\ & \varphi(Ax + a) && (5)\\ \text{s.t.}\ & G(Bx + b, t) \le 0 \quad \text{for all} \quad t \in T, && (6)\\ & x \in X_0. && (7) \end{aligned}$$

Denote by $\mathscr{M}(T)$ the space of regular countably additive measures on T having finite variation, and by $\mathscr{M}_+(T)$ its subset of nonnegative measures. Furthermore, we denote the adjoint operators of A and B by A^* and B^*, respectively. The distance of a point $x \in \mathscr{X}$ to a set $X \subset \mathscr{X}$ will be denoted by $\operatorname{dist}(x, X)$, i.e.,

$$\operatorname{dist}(x, X) = \inf_{y \in X} \|x - y\|.$$

Recall that the Bouligand tangent cone to a set $X \subset \mathscr{X}$ at a point $x_0 \in X$ is defined as follows:

$$T_X(x_0) = \Big\{ d \in \mathscr{X} : \exists\, \tau_n \downarrow 0,\ \frac{1}{\tau_n} \operatorname{dist}(x_0 + \tau_n d, X) \to 0 \text{ as } n \to \infty \Big\}. \quad (8)$$

The normal cone to X at x_0, denoted by $N_X(x_0)$, is defined as the polar of the tangent cone:

$$N_X(x_0) = \Big\{ s \in \mathscr{X}^* : \langle s, d \rangle \le 0, \quad \forall d \in T_X(x_0) \Big\}.$$

In this formula, $\langle \cdot, \cdot \rangle$ refers to the dual pairing. For a convex set X and $x_0 \in X$, we have

$$N_X(x_0) = \big\{ s \in \mathscr{X}^* : \langle s, y - x_0 \rangle \le 0, \quad \forall y \in X \big\}.$$

By definition $N_X(x_0) = \emptyset$ if $x_0 \notin X$.

The following property plays the role of a constraint qualification condition.

DEFINITION 1 *Problem (5)–(7) satisfies the* constraint qualification condition *if there exists a point $\tilde{x} \in X_0$ such that*

$$\max_{t \in T} G(B\tilde{x} + b, t) < 0.$$

We introduce the Lagrangian $L : \mathscr{X} \times \mathscr{M}_+(T) \to \mathbb{R}$

$$L(x,\mu) = \varphi(Ax+a) + \int_T G(Bx+b,t)\,\mu(dt).$$

It is well defined, because $G(Bx+b,\cdot)$ is a continuous function.

We use the symbol $\partial\varphi(y)$ to denote the subdifferential of a convex function $\varphi(\cdot)$ at y. Similarly, $\partial G(y,t)$ denotes the subdifferential of $G(\cdot,t)$ at y.

Theorem 1 *Assume that the constraint qualification condition is satisfied. If $\hat{x}$ is an optimal solution of* (5)–(7) *then there exist a measure $\hat{\mu} \in \mathscr{M}_+(T)$ such that*

$$0 \in A^*\partial\varphi(A\hat{x}+a) + B^*\int_T \partial G(B\hat{x}+b,t)\,\hat{\mu}(dt) + N_{X_0}(\hat{x}), \tag{9}$$

$$\int_T G(B\hat{x}+b,t)\,\hat{\mu}(dt) = 0. \tag{10}$$

Conversely, if for some measure $\hat{\mu} \in \mathscr{M}_+(T)$ a point $\hat{x}$ satisfies (6), (9) *and* (10)*, then $\hat{x}$ is an optimal solution of* (5)–(7).

Proof. Define the operator Γ from $\mathscr{X}$ to the space of continuous functions $\mathscr{C}(T)$ as follows:

$$[\Gamma(x)](t) = G(Bx+b,t), \quad t \in T.$$

It is convex with respect to the cone K of nonpositive functions in $\mathscr{C}(T)$:

$$\Gamma(\lambda x_1 + (1-\lambda)x_2) - [\lambda\Gamma(x_1) + (1-\lambda)\Gamma(x_2)] \in K,$$

for all x_1, x_2 in X_0 and all $\lambda \in [0,1]$.

The constraint (6) can be abstractly written as $\Gamma(x) \in K$. Let us observe that the constraint qualification condition is equivalent to the following generalized Slater condition: There exists a point $\tilde{x} \in X_0$ such that: $\Gamma(\tilde{x}) \in \operatorname{int} K$. By Bonnans and Shapiro (2000, Proposition 2.106), this is equivalent to the regularity condition: $0 \in \operatorname{int}[\Gamma(X_0) - K]$. Therefore we can use the necessary and sufficient conditions of optimality in Banach spaces (see, e.g., Bonnans and Shapiro 2000, Theorem 3.4). We conclude that there exists a nonnegative measure $\hat{\mu} \in \mathscr{M}_+(T)$ such that the following condition holds true:

$$L(\hat{x},\hat{\mu}) = \min_{x \in X_0} L(x,\hat{\mu}), \tag{11}$$

along with the complementarity condition (10). Since $L(\cdot,\hat{\mu})$ is a continuous convex functional, condition (11) is equivalent to the inclusion:

$$0 \in \partial L(\hat{x},\hat{\mu}) + N_{X_0}(\hat{x}). \tag{12}$$

We note that the subdifferential of $L(\cdot, \hat{\mu})$ can be calculated as

$$\partial L(\hat{x}, \hat{\mu}) = \partial[\varphi(A\hat{x} + a)] + \partial \int_T G(B\hat{x} + b, t)\, \hat{\mu}(dt).$$

Furthermore, we have $\partial[\varphi(A\hat{x} + a)] = A^*\partial\varphi(A\hat{x} + a)$. For every $t \in T$, the function $G(Bx + b, t)$ is continuous and convex with respect to x, and its subdifferential at $\hat{x}$ is given by $\partial[G(Bx + b, t)] = B^*\partial G(B\hat{x} + b, t)$. Using Castaing and Valadier (1977, Theorem VII-16) we observe that for a separable $\mathscr{Y}$,

$$\partial \int_T G(B\hat{x} + b, t)\, \hat{\mu}(dt) = B^* \int_T \partial G(B\hat{x} + b, t)\, \hat{\mu}(dt).$$

Applying these calculations to condition (12), we obtain that (11) is equivalent to (9). ■

Let us define the dual functional $D : \mathscr{M}_+(T) \to \overline{\mathbb{R}}$ associated with problem (5)–(7) as follows:

$$D(\mu) = \inf_{x \in X_0} L(x, \mu). \tag{13}$$

We also define the dual problem:

$$\max \left\{ D(\mu) : \mu \in \mathscr{M}_+(T) \right\}. \tag{14}$$

As a direct consequence of Theorem 1, we obtain the duality theorem.

THEOREM 2 *Assume that the constraint qualification condition is satisfied. If problem* (5)–(7) *has an optimal solution, then the dual problem* (14) *has an optimal solution as well, and the optimal values of both problems coincide. Furthermore, for every solution $\hat{\mu}$ of the dual problem, every point $\hat{x}$ satisfying* (6), (9)*, and* (10)*, is an optimal solution of the primal problem* (5)–(7).

3. The nonconvex case

Now we return to problem (1)–(3). We denote the Fréchet derivatives of F and H by F' and H', respectively. We also make an additional assumption that the mapping $G(\cdot, t)$ is Lipschitz continuous with some constant $\varkappa$, for all $t \in T$.

Define the set of feasible solutions:

$$X = \left\{ x \in X_0 : G(H(x), t) \le 0,\ \forall\, t \in T \right\}.$$

Let

$$I_0(x) = \left\{ t \in T : G(H(x), t) = 0 \right\}.$$

Definition 2 *The set X satisfies the* differential constraint qualification condition *at the point $x_0 \in X$ if there exists a point $x_S \in X_0$ and a constant $\delta > 0$ such that for all $t \in I_0(x_0)$*

$$\sup_{\lambda \in \partial G(H(x_0),t)} \big\langle \lambda, H'(x_0)(x_S - x_0) \big\rangle \le -\delta. \tag{15}$$

The above condition may be interpreted as a generalization to the nonsmooth composite case of Robinson's constraint qualification condition introduced in Robinson (1976).

Now we can characterize the tangent cone $T_X(x_0)$.

Theorem 3 *Assume that the differential constraint qualification condition is satisfied at the point $x_0 \in X$. Then $T_X(x_0)$ is the set of vectors $d \in T_{X_0}(x_0)$ satisfying the inequalities*

$$\langle \lambda, H'(x_0)d \rangle \le 0, \tag{16}$$

for all $\lambda \in \partial G(H(x_0), t)$ and all $t \in I_0(x_0)$.

Proof. Suppose that $d \in T_X(x_0)$. This means that there exists a sequence of points $x_k \in X$ and scalars $\tau_k \downarrow 0$ such that:

$$\lim_{k\to\infty} \frac{1}{\tau_k}(x_k - x_0) = d. \tag{17}$$

Consider $\lambda(t) \in \partial G(H(x_0), t)$, for $t \in T$. Using the feasibility of x_k and convexity of $G(\cdot, t)$ we obtain

$$\begin{aligned} 0 \ge G(H(x_k), t) &\ge G(H(x_0), t) + \langle \lambda(t), H(x_k) - H(x_0) \rangle \\ &= G(H(x_0), t) + \langle \lambda(t), H'(x_0)(x_k - x_0) + o(x_k, x_0) \rangle. \end{aligned} \tag{18}$$

Here $\|o(x_k, x_0)\| / \|x_k - x_0\| \to 0$, as $k \to \infty$.

Consider $t \in I_0(x_0)$. From (18) we obtain the inequality

$$0 \ge \langle \lambda(t), H'(x_0)(x_k - x_0) \rangle + \langle \lambda(t), o(x_k, x_0) \rangle.$$

Dividing by τ_k and passing to the limit we get

$$0 \ge \lim_{k\to\infty} \langle \lambda(t), H'(x_0) \frac{x_k - x_0}{\tau_k} \rangle + \lim_{k\to\infty} \langle \lambda(t), \frac{o(x_k, x_0)}{\tau_k} \rangle.$$

Using (17) we obtain (16).

To prove the converse implication, define $d_0 = x_S - x_0$. Consider a direction $d \in T_{X_0}(x_0)$ satisfying (16). We shall prove that $d \in T_X(x_0)$. As d is tangent to the convex set X_0, for every $t > 0$ we can find $\varrho(t) \in \mathscr{X}$ such that

$$x_0 + td + \varrho(t) \in X_0 \quad \text{and} \quad \lim_{t\downarrow 0} \frac{1}{t} \varrho(t) = 0.$$

Choose $\alpha \in (0,1)$. By taking the convex combination of the above point and of $x_0 + td_0$ with coefficients $1-\alpha$ and α, respectively, we obtain the point

$$x(t) = x_0 + (1-\alpha)td + (1-\alpha)\varrho(t) + \alpha t d_0.$$

By convexity of X_0 the point $x(t)$ belongs to X_0. Let $\tau_k \downarrow 0$ and let $t_k = \tau_k/(1-\alpha)$, $\varepsilon = \alpha/(1-\alpha)$. The last formula defines a sequence of points

$$x_k = x_0 + \tau_k d + \varrho_k + \varepsilon\tau_k d_0,$$

with $\varrho_k = (1-\alpha)\varrho(t_k)$. Clearly, $\|\varrho_k\|/\tau_k \to 0$, as $k \to \infty$. Our aim is to show that for all sufficiently large k the points x_k belong to the set X, provided that ε is sufficiently small.

By the differentiability of H and by the fact that $\|\varrho_k\|$ is infinitely smaller than τ_k we have

$$\|H(x_k) - H(x_0) - \tau_k H'(x_0)(d + \varepsilon d_0)\| \le o(\tau_k).$$

Then, using Lipschitz continuity of $G(\cdot, t)$ with a constant $\varkappa$, we get

$$\begin{aligned} &G(H(x_k), t) - G(H(x_0) + \tau_k H'(x_0)(d + \varepsilon d_0)), t) \\ &\qquad \le \varkappa \big\| H(x_k) - H(x_0) - \tau_k H'(x_0)(d + \varepsilon d_0)) \big\| \le \varkappa o(\tau_k). \end{aligned}$$

Taking the maximum over $t \in T$ we conclude that

$$\max_{t\in T} G(H(x_k), t) \le \max_{t\in T} G(H(x_0) + \tau_k H'(x_0)(d + \varepsilon d_0), t) + \varkappa o(\tau_k). \tag{19}$$

Define the function $\psi : \mathscr{Y} \to \mathbb{R}$ as follows:

$$\psi(y) = \max_{t\in T} G(H(x_0) + y, t).$$

It is convex and continuous, and, therefore, directionally differentiable. Inequality (19) implies that

$$\begin{aligned} \max_{t\in T} G(H(x_k), t) &\le \psi(\tau_k H'(x_0)(d + \varepsilon d_0)) + \varkappa o(\tau_k) \\ &\le \psi(0) + \tau_k \psi'(0; H'(x_0)(d + \varepsilon d_0)) + o_1(\tau_k), \end{aligned} \tag{20}$$

with $\psi'(0; y)$ denoting the directional derivative of ψ at 0 in the direction y. Obviously, $o_1(\tau_k)/\tau_k \to 0$, as $\tau_k \downarrow 0$.

Observe that

$$\psi(0) = \max_{t\in T} G(H(x_0), t) \le 0, \tag{21}$$

because the point x_0 is feasible. If $\psi(0) < 0$, then the feasibility of x_k for large k is obvious. It remains to consider the case of $\psi(0) = 0$. In this case, the maximum in (21) is attained on the set $I_0(x_0)$. The directional derivative has the form (see Levin, 1985, Theorem 1.6):

$$\psi'(0;y) = \sup_{t\in I_0(x_0)} \sup_{\lambda\in\partial G(H(x_0),t)} \langle \lambda, y\rangle.$$

Our estimate (20) can thus be continued as follows:

$$\max_{t\in T} G(H(x_k),t) \le \tau_k \sup_{t\in I_0(x_0)} \sup_{\lambda\in\partial G(H(x_0),t)} \langle \lambda, H'(x_0)(d+\varepsilon d_0)\rangle + o_1(\tau_k).$$

Using (16) and (15) in the last inequality, we obtain

$$\max_{t\in T} G(H(x_k),t) \le -\tau_k\varepsilon\delta + o_1(\tau_k).$$

For k large enough, $\varepsilon \ge o_1(\tau_k)/(\tau_k\delta)$. Therefore, we have

$$G(H(x_k),t) \le 0 \quad \text{for all} \quad t\in T.$$

We conclude that for all $\varepsilon > 0$ there exists $\bar{k}(\varepsilon)$ such that for all $k \ge \bar{k}(\varepsilon)$ the points x_k are elements of X. We fix a sequence $\varepsilon_i \to 0$, define $k_i = \bar{k}(\varepsilon_i)$ and consider points

$$x_{k_i} = z_0 + \tau_{k_i} d + \varrho_{k_i} + \varepsilon_i \tau_{k_i} d_0,$$

They are feasible by construction. Moreover,

$$\frac{x_{k_i} - x_0}{\tau_{k_i}} = d + \varepsilon_i d_0 + \frac{1}{\tau_{k_i}} \varrho_{k_i} \to d \quad \text{as} \quad i\to\infty.$$

Consequently, d is indeed a tangent direction to X. ■

Observe that our characterization of the tangent cone involves only the derivative of the mapping H. This allows us to work with a convex approximation $X^{c}(x_0)$ of the feasible set X, obtained by replacing $H(x)$ with its linearization $H(x_0) + H'(x_0)(x - x_0)$. It is defined as follows:

$$X^{c}(x_0) = \big\{x\in X_0 : G(H(x_0) + H'(x_0)(x-x_0)),t) \le 0,\ \forall\, t\in T\big\}.$$

We observe that the tangent cones to X and $X^{c}(x_0)$ at the point x_0 are identical.

COROLLARY 1 *Assume the conditions of Theorem 3. Then*

$$T_X(x_0) = T_{X^{c}(x_0)}(x_0).$$

We can also characterize the cone of descent directions of the composition $f = \varphi\circ F$. Under our assumptions, f is directionally differentiable in every direction d. Its directional derivative $f'(x_0;d)$ has the form:

$$\begin{aligned} f'(x_0;d) &= \varphi'(F(x_0);F'(x_0)d) \\ &= \max_{s\in\partial\varphi(F(x_0))} \langle s, F'(x_0)d\rangle = \max_{s\in\partial\varphi(F(x_0))} \langle [F'(x_0)]^* s, d\rangle. \end{aligned}$$

Using the directional derivative we can characterize the cone of directions of descent as follows:

$$\begin{aligned} K(x_0) &= \big\{d \in \mathscr{X} : f'(x_0; d) < 0\big\} \\ &= \big\{d \in \mathscr{X} : \langle [F'(x_0)]^* s, d\rangle < 0 \text{ for all } s \in \partial\varphi(F(x_0))\big\}. \end{aligned}$$

Denoting by $K^{\mathrm{c}}(x_0)$ the cone of directions of descent of the convex function

$$f^{\mathrm{c}}(x) = \varphi\big(F(x_0) + F'(x_0)(x - x_0)\big)$$

we note that $K(x_0) = K^{\mathrm{c}}(x_0)$.

Let $\hat{x}$ be a local minimum of the nonconvex problem (1)–(3). Our idea is to use the optimality conditions for the convex problem

$$\min \varphi\big(F(\hat{x}) + F'(\hat{x})(x - \hat{x})\big) \tag{22}$$

$$\text{s.t. } G(H(\hat{x}) + H'(\hat{x})(x - \hat{x})), t) \le 0 \quad \text{for all} \quad t \in T, \tag{23}$$

$$x \in X_0. \tag{24}$$

We can now use the results of Section 2 to characterize the local minimum $\hat{x}$.

We introduce the functional $L^{\mathrm{c}} : \mathscr{X} \times \mathscr{M}_+(T) \to \mathbb{R}$ associated with problem (22)-(24):

$$\begin{aligned} L^{\mathrm{c}}(x, \mu; \hat{x}) &= \varphi\big(F(\hat{x}) + F'(\hat{x})(x - \hat{x})\big) \\ &\quad + \int_T G(H(\hat{x}) + H'(\hat{x})(x - \hat{x})), t)\, \mu(dt). \end{aligned}$$

THEOREM 4 *Assume that the point $\hat{x}$ is a local minimum of* (1)–(3) *and the differential constraint qualification condition is satisfied at $\hat{x}$. Then there exist a measure $\hat{\mu} \in \mathscr{M}_+(T)$ such that*

$$0 \in [F'(\hat{x})]^* \partial\varphi(F(\hat{x})) + [H'(\hat{x})]^* \int_T \partial G(H(\hat{x}), t)\, \hat{\mu}(dt) + N_{X_0}(\hat{x}), \tag{25}$$

$$\int_T G(H(\hat{x}), t)\, \hat{\mu}(dt) = 0. \tag{26}$$

Proof. If $\hat{x}$ is a local minimum of problem (1)–(3), then $f'(\hat{x}; d) \ge 0$ for all $d \in T_X(\hat{x})$. Thus, $\hat{x}$ is also a global minimum of problem (22)–(24).

We shall verify the constraint qualification condition for problem (22)–(24), according to Definition 1. It follows from Definition 2 that there exists a point $x_{\mathrm{S}} \in X_0$ such that

$$\sup_{\lambda \in \partial G(H(\hat{x})), t)} \langle \lambda, H'(\hat{x})(x_{\mathrm{S}} - \hat{x})\rangle \le -\delta,$$

for all $t \in I_0(\hat{x})$.

Consider the convex continuous function

$$\psi(y) = \max_{t \in T} G(H(\hat{x}) + y), t).$$

Observe that its value at 0 is nonpositive, because $\hat{x}$ is feasible for (22)-(24). If $\psi(0) < 0$, the uniform dominance condition of Definition 1 is satisfied at $\hat{x}$.

If $\psi(0) = 0$, the active set of parameters t in the calculation of $\psi(0)$ is exactly $I_0(\hat{x})$. Therefore the directional derivative at 0 in the direction $H'(\hat{x})(x_{\mathrm{S}} - \hat{x})$ has the form (see Levin, 1985, Theorem 1.6):

$$\psi'(0; H'(\hat{x})(x_{\mathrm{S}} - \hat{x})) = \sup_{t \in I_0(\hat{x})} \sup_{\lambda \in \partial G(H(\hat{x}),t)} \langle \lambda, H'(\hat{x})(x_{\mathrm{S}} - \hat{x}) \rangle.$$

It follows from the differential constraint qualification condition that

$$\psi'(0; H'(\hat{x})(x_{\mathrm{S}} - \hat{x})) \leq -\delta < 0.$$

For a small $\tau > 0$ we define the point $x(\tau) = (1 - \tau)\hat{x} + \tau x_{\mathrm{S}}$ and obtain

$$\begin{aligned} \max_{t \in T} G(H(\hat{x}) + H'(\hat{x})(x(\tau) - \hat{x})), t) &= \psi(H'(\hat{x})(x(\tau) - \hat{x})) \\ &= \psi(\tau H'(\hat{x})(x_{\mathrm{S}} - \hat{x})) < 0, \end{aligned}$$

provided $\tau > 0$ is sufficiently small. Consequently, the point $x(\tau)$ satisfies the uniform dominance condition of Definition 1 for problem (22)–(24).

We can thus apply Theorem 1 to problem (22)–(24) and obtain the statement of the theorem. ■

Let us introduce the Lagrangian functional for the nonconvex problem (1)–(3), $L : \mathscr{X} \times \mathscr{M}_+(T) \to \mathbb{R}$:

$$L(x, \mu) = \varphi(F(x)) + \int_T G(H(x), t)\, \mu(dt). \tag{27}$$

The functional $L(\cdot, \mu)$ is Lipschitz continuous and, therefore, we can use the Mordukhovich calculus. Using Mordukhovich (2006, Proposition 1.112) we observe that the Mordukhovich subdifferential of $L(\cdot, \hat{\mu})$ evaluated at $\hat{x}$ equals

$$\hat{\partial} L(\hat{x}, \hat{\mu}) = [F'(\hat{x})]^* \partial \varphi(F(\hat{x})) + [H'(\hat{x})]^* \int_T \partial G(H(\hat{x}), t)\, \hat{\mu}(dt).$$

Therefore, condition (25) can be equivalently stated as

$$0 \in \hat{\partial} L(\hat{x}, \hat{\mu}) + N_{X_0}(\hat{x}).$$

If F and H are linear, then $L(\cdot, \hat{\mu})$ is convex and achieves its minimum at $\hat{x}$ over $x \in X_0$.

4. Application to optimization problems with stochastic dominance constraints

Let us return to the stochastic optimization problem with stochastic dominance constraints, mentioned in the introduction.

In this problem, $\mathscr{Y} = \mathscr{L}_1(\Omega, \mathscr{F}, P)$, $T = [a, b] \subset \mathbb{R}$, and the operator G is defined as follows:

$$G(Y,t) = \int_{\Omega} \max(0, t - Y(\omega))\, P(d\omega) - \int_{\Omega} \max(0, t - Y_0(\omega))\, P(d\omega),\ t \in [a,b]. \quad (28)$$

For all t, the function $G(\cdot, t)$ is convex and Lipschitz continuous with constant $\varkappa = 1$. The operator $H : \mathscr{X} \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ is assumed to be continuously Fréchet differentiable. Its derivative $H'(x)$ is a continuous linear operator from $\mathscr{X}$ to $\mathscr{L}_1(\Omega, \mathscr{F}, P)$. For $d \in \mathscr{X}$, the value of $H'(x)d$ at $\omega \in \Omega$ is denoted by $[H'(x)d](\omega)$.

The optimization problem takes on the form

$$\min\ \varphi(F(x)) \quad (29)$$

$$\text{s.t.} \int_{\Omega} \max(0, t - [H(x)](\omega))\, P(d\omega) \le \int_{\Omega} \max(0, t - Y_0(\omega))\, P(d\omega),\ \forall\, t \in T, \quad (30)$$

$$x \in X_0. \quad (31)$$

In particular, we may have here $F(x) = H(x)$ and $\varphi(Y) = -\int Y(\omega)\ P(d\omega)$. Then, problem (29)–(31) can be interpreted as the maximization of the expected value $\mathbb{E}[H(x)]$, under the condition that $H(x)$ stochastically dominates the benchmark Y_0.

In order to obtain a more explicit formulation of the optimality conditions, we calculate the subdifferential of the operator $G(\cdot, t)$ in (28). To this end we define the multifunction $DG : \mathscr{L}_1(\Omega, \mathscr{F}, P) \times [a, b] \times \Omega \rightrightarrows \mathbb{R}$ as follows:

$$DG(Y, t, \omega) = \begin{cases} \{-1\} & \text{if } Y(\omega) < t, \\ [-1, 0] & \text{if } Y(\omega) = t, \\ \{0\} & \text{if } Y(\omega) > t. \end{cases}$$

For a fixed t, the mapping $Y \to \int_{\Omega} \max(0, t - Y(\omega))\, P(d\omega)$ is a convex integral functional on $\mathscr{L}_1(\Omega, \mathscr{F}, P)$, with the convex normal integrand $y \to \max(0, t-y)$. Using Castaing and Valadier (1977, Theorem VII-7) we see that the subdifferential of G with respect to the first argument at the point $Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$, $\partial G(Y, t)$, is the collection of all measurable selections $\lambda(\cdot)$ of $DG(Y, t, \cdot)$. Observe that each $\lambda \in \mathscr{L}_\infty(\Omega, \mathscr{F}, P) = \mathscr{Y}^*$.

Using the explicit form of $\partial G(\cdot,t)$, we can calculate the supremum in the constraint qualification condition (15) as follows:

$$\begin{aligned}
&\sup_{\lambda\in\partial G(H(x_0),t)} \big\langle \lambda, H'(x_0)(x_{\mathrm{S}}-x_0)\big\rangle \\
&\quad= \sup_{\lambda\in\partial G(H(x_0),t)} \int_{\Omega} \lambda(\omega)[H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\; P(d\omega) \\
&\quad= \int_{\Omega} \sup_{\lambda(\omega)\in DG(H(x_0),t,\omega)} \lambda(\omega)[H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\; P(d\omega) \\
&\quad= -\int_{\{H(x_0)<t\}} [H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\; P(d\omega) \\
&\qquad+ \int_{\{H(x_0)=t\}} \max\big(0, -[H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\big)\; P(d\omega).
\end{aligned}$$

The second equation is due to Rockafellar and Wets (1998, Theorem 14.60), and the last one follows from the structure of the multifunction DG.

The constraint qualification condition can now be written as follows:

DEFINITION 3 *The set X satisfies the* differential constraint qualification condition *at the point $x_0 \in X$ if there exists a point $x_{\mathrm{S}} \in X_0$ and a constant $\delta > 0$ such that for all $t \in I_0(x_0)$*

$$\begin{aligned}
&\int_{\{H(x_0)<t\}} [H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\; P(d\omega) \\
&\geq \int_{\{H(x_0)=t\}} \max\big(0, -[H'(x_0)(x_{\mathrm{S}}-x_0)](\omega)\big)\; P(d\omega) + \delta. \qquad (32)
\end{aligned}$$

The Lagrangian (27) associated with problem (29)-(31) takes on the form:

$$\begin{aligned}
L(x,\mu) = \varphi(F(x)) &+ \int_T \int_\Omega \max(0, t - [H(x)](\omega))\, P(d\omega)\, \mu(dt) \\
&- \int_T \int_\Omega \max(0, t - Y_0(\omega))\, P(d\omega)\, \mu(dt).
\end{aligned} \qquad (33)$$

The specific form of the Lagrangian allows us to represent it in a more transparent way.

We define the set $\mathscr{U}$ of functions $u(\cdot)$ satisfying the following conditions:

$u(\cdot)$ is concave and nondecreasing;

$u(t) = 0$ for all $t \geq b$;

$u(t) = u(a) + c(t-a)$, with some $c > 0$, for all $t \leq a$.

It is evident that $\mathscr{U}$ is a convex cone. Moreover, the subgradients of each function $u \in \mathscr{U}$ are bounded for all $t \in \mathbb{R}$.

Every measure μ on $[a, b]$ can be extended to the whole real line by assigning measure 0 to Borel sets not intersecting $[a, b]$. A function $u : \mathbb{R} \to \mathbb{R}$ can be associated with every nonnegative measure μ as follows:

$$u(t) = \begin{cases} -\int_t^b \mu([\tau, b])\, d\tau & t < b, \\ 0 & t \geq b. \end{cases}$$

Since $\mu \geq 0$, the function $\mu([\cdot, b])$ is nonnegative and nonincreasing, which implies that $u(\cdot)$ is nondecreasing and concave.

Conversely, if $u \in \mathscr{U}$ then the left derivative of u,

$$u'_-(t) = \lim_{\tau \uparrow t} [u(t) - u(\tau)]/(t - \tau),$$

is well-defined, nonincreasing and continuous from the left. By Billingsley (1995, Theorem 12.4), after an obvious adaptation, there exists a unique regular nonnegative measure μ satisfying $\mu([t, b]) = u'_-(t)$. Thus, the correspondence between nonnegative measures in $\mathscr{M}([a, b])$ and functions in $\mathscr{U}$ is a bijection.

In Dentcheva and Ruszczyński (2003), we have proved that for every random variable Y,

$$\int_T \int_\Omega \max(0, t - Y(\omega))\, P(d\omega)\, \mu(dt) = -\int_\Omega u(Y)\, P(d\omega),$$

where $u(\cdot)$ is derived from μ in the way described above.

This correspondence entails a correspondence of the Lagrangian (33) to another functional $\Lambda : \mathscr{X} \times \mathscr{U} \to \mathbb{R}$ defined as follows:

$$\Lambda(x, u) = \varphi(F(x)) - \int_\Omega u(H(x))\, P(d\omega) + \int_\Omega u(Y_0)\, P(d\omega).$$

Using Theorem 4, we obtain the following optimality conditions for problem (29)–(31).

Theorem 5 *Assume that the point $\hat{x}$ is a local minimum of* (29)–(31) *and the differential constraint qualification condition (32) is satisfied at $\hat{x}$. Then, there exists a function $\hat{u} \in \mathscr{U}$ such that*

$$0 \in [F'(\hat{x})]^* \partial\varphi(F(\hat{x})) - [H'(\hat{x})]^* \int_\Omega \partial u(H(\hat{x}))\, P(d\omega) + N_{X_0}(\hat{x}),$$

$$\int_\Omega u(H(\hat{x}))\, P(d\omega) = \int_\Omega u(Y_0)\, P(d\omega).$$

The integral $\int_\Omega \partial u(Y)\,P(d\omega)$ is understood as a collection of integrals of all measurable selections of the multifunction $\omega \to \partial u(Y(\omega))$.

Theorem 5 generalizes to the nonconvex case the optimality conditions of Dentcheva and Ruszczyński (2004, Theorem 2).

References

Billingsley, P. (1995) *Probability and Measure.* John Wiley & Sons, New York.

Bonnans, J.F. and Cominetti, R. (1996) Perturbed optimization in Banach spaces. III. Semi-infinite optimization. *SIAM J. Control Optim.* **34**, 1555–1567.

Bonnans, J.F. and Shapiro, A. (2000) *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York.

Castaing, C. and Valadier, M. (1977) *Convex Analysis and Measurable Multifunctions.* Springer-Verlag, Berlin.

Canovas, M.J., Dontchev, A.L., Lopez, M.A. and Parra, J. (2005) Metric regularity of semi-infinite constraint systems. *Mathematical Programming* **104**, 329–346.

Dentcheva, D. and Ruszczyński, A. (2003) Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* **14**, 548–566.

Dentcheva, D. and Ruszczyński, A. (2004) Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming* **99**, 329–350.

Goberna, M.A. and Lopez, M.A. (1998) *Linear Semi-Infinite Optimization.* Wiley, Chichester.

Klatte, D. and Henrion, R. (1998) Regularity and stability in nonlinear semi-infinite optimization. *Nonconvex Optimization and Applications* **25**, 69–102.

Levin, V.L. (1985) *Convex Analysis in Spaces of Measurable Functions and its Applications in Economics.* Nauka, Moscow, (in Russian).

Mordukhovich, B. (2006) *Variational Analysis and Generalized Differentiation.* Springer-Verlag, Berlin.

Penot, J.P. (1994) Optimality conditions in mathematical programming and composite optimization. *Mathematical Programming* **67**, 225–246.

Robinson, S.M. (1976) Stability theory for systems of inequalities. II: Differentiable nonlinear systems. *SIAM J. Numer. Anal.* **13**, 497–513.

Rockafellar, R.T. and Wets, R.J.-B. (1998) *Variational Analysis.* Springer-Verlag, Berlin.

Studniarski, M. and Jeyakumar, V. (1995) A generalized mean-value theorem and optimality conditions in composite nonsmooth minimization. *Nonlinear Analysis* **24**, 883–894.

Yang, X.Q. (1998) Second-order global optimality conditions for convex composite optimization. *Mathematical Programming* **81**, 327–347.