BUNDLE–BASED DECOMPOSITION: CONDITIONS FOR CONVERGENCE

Stephen M. Robinson

Septpember 1987
WP-87-80

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Bundle-based decomposition uses the bundle method for finding approximate solutions of linearly constrained cource optimization problems whose structure allows for decomposition to subproblems of smaller dimension. Together with the general features of the bundle method this has contributed to encouraging numerical results. The detailed analysis contained in the presented paper adds important results on a priori conditions needed for the convergence of the approximate solutions and of the approximate optimal values to the actual ones.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

BUNDLE–BASED DECOMPOSITION: CONDITIONS FOR CONVERGENCE

Stephen M. Robinson
University of Wisconsin—Madison

ABSTRACT

Bundle–based decomposition is a recently proposed method for decentralized convex optimization. Computational tests indicate that it is very fast. In this paper we exhibit conditions for convergence of the method. In the process we study conditions for linearly–constrained approximate minimization of a convex function.

Keywords: Decomposition, bundle method, convergence conditions, epsilon–subdifferential, local boundedness, epsilon–first–order conditions.

AMS(MOS) Subject classification: 90C30.

Sponsored by the National Science Foundation under Grant No. CCR-8502202, and by the International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

1. Introduction.

Bundle–based decomposition (BBD) is a recently proposed method for solving the convex optimization problem

$$\begin{aligned} &\text{Minimize} && \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to} && \sum_{i=1}^{n} A_i x_i = a, \end{aligned} \tag{1.1}$$

where the f_i are closed proper convex functions on $\mathbb{R}^{n_i}, a \in \mathbb{R}^m$, and each A_i is a linear transformation from $\mathbb{R}^{n_i}$ to $\mathbb{R}^m$. The problem (1.1) represents a decentralized optimization with certain overall constraints connecting the individual problems. The method in question was described in [11], and extensive computational tests are reported in [9]. These tests showed the method to be very fast compared both to MINOS 5.0 [10] and to the Ho–Loute "advanced implementation" of Dantzig–Wolfe decomposition [3,4].

After the user prescribes certain parameters the BBD method produces, in a finite number of steps, approximate primal and dual solutions of (1.1). In this paper we identify conditions on the problem (1.1) under which the method is *convergent*: that is, under which the parameters can, in principle, be set so that the computed approximate solutions will lie within any preassigned tolerance of an actual pair of primal and dual solutions of (1.1). Thus, the analysis here contributes *a priori* convergence conditions, whereas in [9, Theorem 3.7] Medhi develops *a posteriori* error information.

The rest of the paper consists of three sections. In Section 2 we analyze the BBD method to establish properties of the approximate solutions it produces. We show that they satisfy certain "ϵ–first–order" optimality conditions given by Strodiot, Nguyen, and Heukemes [14], and we characterize points satisfying those conditions in terms of approximate optimization of a certain perturbed dual pair of convex programming problems.

In Section 3 we introduce a simple characterization of local boundedness for multifunctions, and use it to show that the inverse of the multifunction associated with the ϵ–first–order conditions is Hausdorff upper semicontinuous at interior points of its image. Further, we obtain an expression for the interior of that image and we show that it is independent of the tolerance ϵ.

In Section 4 we translate the interiority information obtained in Section 3 into a pair of simple conditions on the optimization problem (1.1): these amount to a Slater condition plus a compactness assumption on the level sets of the essential objective function. Then we show that under these two conditions the BBD method is convergent in the sense described above.

2. The BBD method and the ϵ–first–order conditions.

The BBD method solves (1.1) by dualizing with respect to the equality constraint to produce a concave dual objective function

$$g(p^*) = \langle p^*, a\rangle - \sum_{i=1}^{n} f_i^*(A_i^* p^*).$$

Under the technical assumptions that

$$a \in \sum_{i=1}^{n} A_i(\text{ri dom} f_i) \tag{2.1}$$

and that there exists p_0^* with

$$A_i^* p_0^* \in \text{ ri dom } f_i^*, \quad i = 1, \ldots, n, \tag{2.2}$$

we have

$$\partial g(p^*) = a - \sum_{i=1}^{n} A_i x_i(p^*) \tag{2.3}$$

where $x_i(p^*)$ is the set of points solving the decentralized subproblem

$$\text{minimize}\, \{f_i(x_i) - \langle A_i^* p^*,\, x_i\rangle\}. \tag{2.4}$$

The BBD method uses the bundle method [7] to find an approximate maximizer of g, using (2.3) and (2.4) to compute subgradients of g. Since the way in which the method uses this information is important to our analysis, we shall describe it in enough detail to develop the facts that we need later.

The use of the method prescribes two small tolerances, ϵ and δ. At the termination of the bundle algorithm one has dual elements $p_1^*, \ldots, p_k^*$ and associated primal elements $\{x_{ji} \mid i = 1, \ldots, n;\ j = 1, \ldots, k\}$ having the following properties:

(1) x_{ji} minimizes $f_i(\cdot) - \langle A_i^* p_j^*, \cdot\rangle$ for each i and j: that is,

$$A_i^* p_j^* \in \partial f_i(x_{ji}),\ i = 1, \ldots, n;\ j = 1, \ldots, k. \tag{2.5}$$

(2) With $d_j := a - \sum_{i=1}^{n} A_i x_{ji}$, we have from (2.3)

$$d_j \in \partial g(p_j^*) \qquad j = 1, \ldots, k. \tag{2.6}$$

(3) There exist $\lambda_1, \ldots, \lambda_k$ all non-negative with $\sum_{j=1}^{k} \lambda_j = 1$ and such that with $d := \sum_{j=1}^{k} \lambda_j d_j$ we have

$$\|d\| \le \delta \tag{2.7}$$

and

$$\sum_{j=1}^{n} \lambda_j \epsilon_j \le \epsilon, \tag{2.8}$$

where

$$\epsilon_j := g(p_j^*) - g(p_k^*) - \langle p_j^* - p_k^*, d_j \rangle; \tag{2.9}$$

one has $\epsilon_j \ge 0$ by (2.6).

The method takes $\hat{p}^* = p_k^*$ to be the approximate dual solution for (1.1). To construct an approximate primal solution $(\hat{x}_1, \ldots, \hat{x}_n)$ it sets

$$\hat{x}_i = \sum_{j=1}^{k} \lambda_j x_{ji}, \quad i = 1, \ldots, n; \tag{2.10}$$

note that

$$\sum_{i=1}^{n} A_i \hat{x}_i = \sum_{j=1}^{k} \lambda_j \left(\sum_{i=1}^{n} A_i x_{ji} \right) = a - \sum_{j=1}^{k} \lambda_j d_j,$$

so that (2.7) implies $\|\sum_{i=1}^{n} A_i \hat{x}_i - a\| \le \delta$. Thus if δ is small then $(\hat{x}_1, \ldots, \hat{x}_n)$ is nearly feasible for (1.1).

The objective of this paper can now be precisely stated as follows: exhibit conditions on the problem (1.1) under which for each positive η there exists a positive γ so that whenever $\max\{\delta, \epsilon\} < \gamma$ there are points $(\bar{x}_1, \ldots, \bar{x}_n)$ solving (1.1) and $\bar{p}^*$ maximizing the dual objective g, such that

$$\max\{\|\hat{x}_1 - \bar{x}_1\|, \ldots, \|\hat{x}_n - \bar{x}_n\|, \|\hat{p}^* - \bar{p}^*\|\} < \eta,$$

where $(\hat{x}_1, \ldots, \hat{x}_n)$ and $\hat{p}^*$ are the points produced by the algorithm as described above. These conditions will be obtained in Section 4; they turn out to be strengthened versions of the technical assumptions (2.1) and (2.2).

In the remainder of this section we rewrite the information in (2.5) through (2.10) in a more manageable form. To do so we let $x := (x_1, \dots, x_n) \in \mathbb{R}^N$, where $N = \sum_{i=1}^{n} n_i$, and we define $f(x)$ to be $\sum_{i=1}^{n} f_i(x_i)$ and

$$A := [A_1 A_2 \cdots A_n],$$

so that $A : \mathbb{R}^N \to \mathbb{R}^m$, and $Ax = \sum_{i=1}^{n} A_i x_i$. We use a similar convention for $\hat{x}$ and $\bar{x}$.

PROPOSITION 2.1. *The approximate solutions $\hat{x}$ and $\hat{p}^*$ produced by the BBD method satisfy*

$$\begin{pmatrix} 0 \\ -d \end{pmatrix} \in \begin{pmatrix} \partial_\epsilon f & -A^* \\ A & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{p}^* \end{pmatrix} + \begin{pmatrix} 0 \\ -a \end{pmatrix}; \tag{2.11}$$

that is,

$$0 \in \partial_\epsilon f(\hat{x}) - A^* \hat{p}^* \tag{2.12}$$

and

$$-d = A\hat{x} - a. \tag{2.13}$$

PROOF: We have $\partial f(x) = \times_{i=1}^{n} \partial f_i(x_i)$, so we can rewrite (2.5) as

$$A^* p_j^* \in \partial f(x_j), \quad j = 1, \dots, k$$

and so $x_j \in \partial f^*(A^* p_j^*)$ for each j. Hence for each z^* and each j,

$$\begin{aligned} f^*(z^*) &\geq f^*(A^* p_j^*) + \langle z^* - A^* p_j^*, x_j \rangle \\ &= f^*(A^* \hat{p}^*) + \langle z^* - A^* \hat{p}^*, x_j \rangle \\ &\quad - \left[f^*(A^* \hat{p}^*) - f^*(A^* p_j^*) - \langle A^* \hat{p}^* - A^* p_j^*, x_j \rangle \right]. \end{aligned}$$

The quantity in brackets can be rewritten as

$$g(p_j^*) - g(\hat{p}^*) - \langle p_j^* - \hat{p}^*,\, a - Ax_j \rangle.$$

Comparing this with (2.9) and using $\hat{p}^* = p_k^*$ and $d_j = a - Ax_j$, we see that this is just ϵ_j, so we have

$$f^*(z^*) \;\geq\; f^*(A^* \hat{p}^*) + \langle z^* - A^* \hat{p}^*,\, x_j \rangle - \epsilon_j.$$

Now multiplying this inequality by λ_j and summing over j, we obtain

$$f^*(z^*) \geq f^*(A^*\hat{p}^*) + \langle z^* - A^*\hat{p}^*, \hat{x}\rangle - \epsilon;$$

i.e., $\hat{x} \in \partial_\epsilon f^*(A^*\hat{p}^*)$, which is equivalent to (2.12). The proof of (2.13) amounts to multiplying the definition $d_j = a - Ax_j$ by λ_j and summing over j. ∎

The form in which (2.11) is written emphasizes its closeness to the standard first–order optimality conditions. In fact, (2.11) amounts to a slight perturbation of the "ϵ-first–order" optimality conditions of Strodiot, Nguyen, and Heukemes [14], specialized to the present case: the perturbation consists in the replacement of $\binom{0}{0}$ on the left side of the inclusion by $\binom{0}{-d}$.

The analysis in [14] emphasized establishing necessary and sufficient conditions for ϵ-optimality in the presence of a constraint qualification. For the simpler problem with which we are concerned here, the conditions (2.11) have a very clear and direct interpretation, which we give in the following proposition. In it, we consider the pair of optimization problems

$$\inf\{f(x) \mid Ax = a - d\}, \tag{2.14}$$

and

$$\sup g_d(p^*), \tag{2.15}$$

where

$$g_d(p^*) := \langle p^*, a - d\rangle - f^*(A^*p^*).$$

Note that (2.14) and (2.15) are dual to each other under the duality structure generated by

$$F(x,p) := \begin{cases} f(x) & \text{if } Ax = a - d - p, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.16}$$

which is a slight perturbation (by d) of that used to generate the dual objective g of the BBD method. The function g_d is $g - \langle \cdot, d\rangle$.

PROPOSITION 2.2. *The following are equivalent:*

(i) x *and* p^* *satisfy (2.11).*

(ii) $Ax = a - d$ *and* $f(x) - g_d(p^*) \leq \epsilon$.

PROOF: x and p^* satisfy (2.11) if and only if $Ax = a - d$ and $A^*p^* \in \partial_\epsilon f(x)$.

The second of these relations can be written

$$\begin{aligned}\epsilon &\geq f(x) + f^*(A^*p^*) - \langle A^*p^*, x\rangle \\ &= f(x) - \{\langle p^*, a-d\rangle - f^*(A^*p^*)\} \\ &= f(x) - g_d(p^*),\end{aligned}$$

so (ii) holds. Reversing the argument shows that (ii) implies (i). ∎

Now define a multifunction M with arguments (ϵ, r^*, s) by

$$M(\epsilon, r^*, s) := \left\{(x, p^*) \,\middle|\, \begin{pmatrix} r^* \\ s \end{pmatrix} \in \begin{pmatrix} \partial_\epsilon f & -A^* \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ p^* \end{pmatrix} + \begin{pmatrix} 0 \\ -a \end{pmatrix}\right\}; \tag{2.17}$$

that is, for each ϵ $M(\epsilon, ., .)$ is the multifunction inverse to that on the right side of (2.11). With this notation $M(0,0,0)$ is the product of the primal and dual solution sets of (1.1), where the duality structure is that used in the BBD method: i.e., (2.16) with $d = 0$. Therefore our aim of proving the BBD method convergent will be achieved if we can show that when ϵ, r^*, and s are sufficiently close to 0, *each* point of $M(\epsilon, r^*, s)$ will lie within a predetermined distance of *some* point of $M(0, 0.0)$. This amounts to proving that M is *Hausdorff upper semicontinuous* (H-usc) at $(0,0,0)$. In the next section we exhibit conditions under which this will be true.

3. Semicontinuity of solutions to the ϵ–first–order conditions.

In Section 2 we observed that the critical issue in proving convergence of the BBD method was to show that the operator M, expressing solutions of the perturbed ϵ-first-order conditions in terms of the perturbations and the tolerance ϵ, was Hausdorff usc at (0,0,0). In this section we prove this by showing that M is locally bounded under certain assumptions. We then conclude that M is actually Hausdorff usc. Then in Section 4 we analyze the required assumptions and relate them to properties of the minimization problem (1.1), thus developing conditions on (1.1) under which the BBD method will converge.

To begin the analysis of local boundedness, we consider a multifunction G from $\mathbb{R}^k$ to $\mathbb{R}^\ell$. By definition, G is locally bounded at a point $x_0 \in \mathbb{R}^k$ if there is some neighborhood N of x_0 such that $G(N)$ $\left(:= \bigcup \left\{G(x) \,\middle|\, x \epsilon N\right\}\right)$ is a bounded set. The following simple proposition characterizes local boundedness.

PROPOSITION 3.1. *Let G be a multifunction from $\mathbb{R}^k$ to $\mathbb{R}^\ell$. Then G is locally bounded at x_0 if and only if for each y near x_0*

$$\limsup_{\substack{x \to x_0 \\ x^* \in G(x)}} \langle x^*, y - x \rangle < +\infty. \tag{3.1}$$

PROOF (only if): Choose a neighborhood V of x_0 small enough so that $G(V) \subset \eta B$ for some η, where B is the unit ball. Let $y \epsilon \mathbb{R}^k$. Then for each $x \in V$ and each $x^* \in G(x)$, $\langle x^*, y - x \rangle \leq \eta \|y - x\|$. Hence

$$\limsup_{\substack{x \to x_0 \\ x^* \in G(x)}} \langle x^*, y - x \rangle \leq \limsup_{x \to x_0} \eta \|y - x\| = \eta \|y - x_0\|,$$

and thus (3.1) holds. (Note that if $G(V) = \phi$ the limit superior is $-\infty$ by definition.)

(if): Assume that (3.1) holds for each y near x_0. If G is not locally bounded at x_0 then there is a sequence $\{x_n\}$ converging to x_0, with $x_n^* \in \mathrm{G}(x_n)$ such that $\|x_n^*\| \geq n$ for $n = 1, 2, \ldots$. There is no loss in assuming that $x_n^*/\|x_n^*\|$ converges to some point z_0. Now choose any y near x_0. By (3.1) there is some γ such that for each n, $\langle x_n^*, y - x_n \rangle \leq \gamma$. Dividing this inequality by $\|x_n^*\|$ and taking the limit, we find that

$$\langle z_0, y - x_0 \rangle \leq 0. \tag{3.2}$$

However, $\|z_0\| = 1$, so (3.2) cannot hold for every such y. Therefore G is locally bounded at x_0. ∎

We consider briefly some classes of multifunctions that satisfy (3.1). First, consider monotone operators: that is, multifunctions $G : \mathbb{R}^k \to \mathbb{R}^k$ having the property that for each x_1 and x_2 in $\operatorname{dom} G$ $\left(:= \{x \epsilon \mathbb{R}^k | G(x) \neq \phi\}\right)$ and each $y_1^* \in G(x_1)$ and $y_2^* \in G(x_2)$, one has

$$\langle y_1^* - y_2^*, x_1 - x_2 \rangle \geq 0.$$

For such an operator G, if $x_0 \in \operatorname{int} \operatorname{dom} G$ then for any y near x_0, any fixed $y^* \in G(y)$, any x near x_0 and any $x^* \in G(x)$, we have

$$\langle x^*, y - x \rangle \leq \langle y^*, y - x \rangle;$$

therefore

$$\limsup_{\substack{x \to x_0 \\ x^* \in G(x)}} \langle x^*, y - x \rangle \leq \langle y^*, y - x_0 \rangle < +\infty,$$

and (3.1) holds. In this case the result of Proposition 3.1 is a special case of Rockafellar's theorem on the local boundedness of monotone operators [12], and of Kato's earlier results and (3.1) holds. In this case the result of Proposition 3.1 is a special case of Rockafellar's theorem on the local boundedness of monotone operators [12], and of Kato's earlier results [5,6]. These results hold in much more general spaces and, as might be expected, their proofs are much more substantial than that of Proposition 3.1.

Next, consider for some fixed $\epsilon \geq 0$ the multifunction G_ϵ defined by

$$\begin{aligned} G_\epsilon(x, p^*) &:= \begin{pmatrix} \partial_\epsilon f & -A^* \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ p^* \end{pmatrix} + \begin{pmatrix} 0 \\ -a \end{pmatrix} \\ &= \left\{ (r^*, s) \,\middle|\, r^* + A^* p^* \in \partial_\epsilon f(x),\ s = Ax - a \right\}. \end{aligned} \tag{3.3}$$

Suppose that x_1 and x_2 belong to $\operatorname{dom} \partial_\epsilon f$; let p_1^* and p_2^* be arbitrary, and let $(r_i^*, s_i) \in G_\epsilon(x_i, p_i^*)$ for $i = 1, 2$. Then

$$f(x_2) \geq f(x_1) + \langle r_1^* + A^* p_1^*,\, x_2 - x_1 \rangle - \epsilon$$

and

$$f(x_1) \geq f(x_2) + \langle r_2^* + A^* p_2^*,\, x_1 - x_2 \rangle - \epsilon,$$

so by addition we find that

$$\begin{aligned} -2\epsilon &\leq \langle r_1^* - r_2^*,\, x_1 - x_2 \rangle + \langle p_1^* - p_2^*,\, A(x_1 - x_2) \rangle \\ &= \langle (r_1^*, s_1) - (r_2^*, s_2),\, (x_1, p_1^*) - (x_2, p_2^*) \rangle, \end{aligned}$$

where we have used the obvious definition of the inner product on $\mathbb{R}^{N+m}$. Since this multifunction G_ϵ satisfies an inequality similar to that satisfied by monotone operators, we can use an argument similar to the one just made to show that G_ϵ is locally bounded at each point of $\operatorname{int} \operatorname{dom} G_\epsilon$.

Observe that since the key inequalities used above for monotone operators and for the operator G_ϵ are symmetric in arguments and values, the local boundedness conclusions hold also for the inverses of those operators, where the inverse of a multifunction $F : \mathbb{R}^k \to \mathbb{R}^\ell$ is the multifunction $F^{-1} : \mathbb{R}^\ell \to \mathbb{R}^k$ defined by

$$F^{-1}(y) := \{ x \mid y \in F(x) \}.$$

Since the effective domain of F^{-1} is then the image of F (written $\operatorname{im} F$, this is the set $\{ y \mid \text{for some } x, y \in F(x) \}$), the local boundedness assertions for the inverses hold at interior points of the images of the original multifunctions.

Also, note that the graph of the operator G_ϵ defined by (3.3) can be written as

$$\left\{(x, p^*, r^*, s) \,\middle|\, s = Ax - a,\ (x, r^* + A^*p^*) \in \partial_\epsilon f\right\},$$

where $\partial_\epsilon f$ represents the set $\left\{(x, x^*) \,\middle|\, x^* \in \partial_\epsilon f(x)\right\}$, the graph of $\partial_\epsilon f$. As $\partial_\epsilon f \supset \partial_\eta f$ when $\epsilon \geq \eta$, the same isotonicity holds for the graph of G_ϵ. In particular, for any sets U and V, if $\epsilon \geq \eta$ then $G_\epsilon(U) \supset G_\eta(U)$ and $G_\epsilon^{-1}(V) \supset G_\eta^{-1}(V)$. Thus if G_ϵ^{-1} is locally bounded somewhere, then the same bound applies to G_η^{-1}.

We can summarize these observations in the following corollary.

COROLLARY 3.2. *Let $\epsilon \geq 0$ and let G_ϵ be defined by (3.3). If (r_0^*, s_0) belongs to the interior of $\operatorname{im} G_\epsilon$, then there exist a neighborhood N of (r_0^*, s_0) and a bounded set V, such that for each $\eta \in [0, \epsilon]$, $G_\eta^{-1}(N) \subset V$.*

We can see from the results already proved that we will need to identify points in the interior of $\operatorname{im} G_\epsilon$. The following theorem characterizes such points: in fact, it characterizes the closure and interior of $\operatorname{im}(\partial_\epsilon g + H)$ where g is any closed proper convex function and H is a monotone operator. In this sense it extends the fact that $\operatorname{im} \partial_\epsilon g \cong \operatorname{im} \partial g$, where we write $C \cong D$ to indicate that the sets C and D have the same closure and the same interior.

THEOREM 3.3. *Let g be a closed proper convex function on $\mathbb{R}^k$, and H a monotone operator from $\mathbb{R}^k$ to itself such that $\partial g + H$ is maximal monotone. Then for each $\epsilon \geq 0$,*

$$\operatorname{im}(\partial_\epsilon g + H) \cong \operatorname{im}(\partial g + H) \cong (\operatorname{im} \partial g) + H(\operatorname{dom} \partial g).$$

PROOF: Denote by $\tilde{H}$ the restriction of H to dom ∂g. Then $\tilde{H}$ is monotone, dom $\partial g \supset \operatorname{dom} \tilde{H}$, and $\partial g + H = \partial g + \tilde{H}$ is maximal monotone. By the theorem of Brezis and Haraux [2, Th. 4] one has $\operatorname{im}(\partial g + H) \cong \operatorname{im} \partial g + \operatorname{im} \tilde{H}$. But $\operatorname{im} \tilde{H} = H(\operatorname{dom} \partial g)$, so this proves the second "$\cong$" claim.

For the first, note that the graph inclusion property implies $\operatorname{im}(\partial_\epsilon g + H) \supset \operatorname{im}(\partial g + H)$, and therefore this inclusion holds also for the closures and the interiors of these sets. Write S_ϵ for $\operatorname{im}(\partial_\epsilon g + H)$ and S for $\operatorname{im}(\partial g + H)$, and suppose that we could prove $\operatorname{cl} S_\epsilon \subset \operatorname{cl} S$. We know that $\operatorname{int} S = \operatorname{int} \operatorname{cl} S$ [1, p. 33], and therefore we would have $\operatorname{int} S_\epsilon \supset \operatorname{int} S = \operatorname{int} \operatorname{cl} S = \operatorname{int} \operatorname{cl} S_\epsilon \supset \operatorname{int} S_\epsilon$, implying that all of the sets in this chain of inclusions are the same. Thus we will have finished the proof if we can show that $\operatorname{cl} S_\epsilon \subset \operatorname{cl} S$.

Since $\operatorname{im}\partial_\epsilon g \subset \operatorname{cl}\operatorname{im}\partial g$ and $\operatorname{dom}\partial_\epsilon g \subset \operatorname{cl}\operatorname{dom}\partial g$, we have

$$\begin{aligned}\operatorname{im}(\partial_\epsilon g + H) &\subset \operatorname{im}\partial_\epsilon g + H(\operatorname{cl}\operatorname{dom}\partial g)\\ &\subset \operatorname{cl}\ \operatorname{im}\partial g + \operatorname{cl} H(\operatorname{dom}\partial g)\\ &\subset \operatorname{cl}[\operatorname{im}\partial g + H(\operatorname{dom}\partial g)]\\ &= \operatorname{cl}\ \operatorname{im}(\partial g + H),\end{aligned}$$

where we have used the second " $\cong$ " relation, already proved. Now by taking the closure of the left side above, we obtain $\operatorname{cl} S_\epsilon \subset \operatorname{cl} S$ as required. ∎

It is worth remarking that we do not in general have equality, even when $H = 0$, as the example $g(x) = e^{-x}$ shows. Here $\operatorname{im}\partial g = (-\infty, 0)$, but for $\epsilon > 0$ $\operatorname{im}\partial_\epsilon g = (-\infty, 0]$.

Now recall that at the end of Section 2 we pointed out that the convergence property we wanted amounted to Hausdorff upper semicontinuity of a certain multifunction. For a multifunction F from $\mathbb{R}^k$ to $\mathbb{R}^\ell$, we say F is Hausdorff upper semicontinuous (H-usc) at $x_0 \epsilon \mathbb{R}^k$ if for each $\eta > 0$ there is some neighborhood N of x_0 such that $F(N) \subset F(x_0) + \eta B$, where B is the unit ball. As might be expected, this property is closely related to local boundedness. Specifically, we say that F is *closed at* x_0 if

$$F(x_0) = \bigcap_{N \in N(x_0)} \operatorname{cl} F(N),$$

where $N(x_0)$ is the neighborhood system at x_0. This amounts to saying that if $x_n \to x_0$ and $y_n \in F(x_n)$ for each n, with $y_n \to y_0$, then $y_0 \in F(x_0)$. Now it is easy to show that if F is closed at x_0 and locally bounded there, then it is Hausdorff usc at x_0. This fact, together with what we have proved up to now, leads to the following continuity result for solutions of the ϵ-first-order conditions.

THEOREM 3.4. *Let M be defined by (2.17) and let $\epsilon \geq 0$. Then M is Hausdorff usc at (ϵ, r^*, s), relative to $\mathbb{R}_+ \times \mathbb{R}^N \times \mathbb{R}^m$, whenever*

$$r^* \in \operatorname{int}[\operatorname{dom} f^* + \operatorname{im} A^*] \tag{3.4}$$

and

$$s \in \operatorname{int}[A(\operatorname{dom} f) - a]. \tag{3.5}$$

PROOF: We are going to show that the (r^*, s) satisfying (3.4) and (3.5) are those belonging to the interior of the image of the operator G_ϵ defined by (3.3). By Theorem 3.3

this is also the interior of $\operatorname{im} G_\sigma$ for some $\sigma > \epsilon$. Then Corollary 3.2 shows that for some neighborhood N of (r^*, s) and all $\eta \in [0, \sigma]$, $G_\eta^{-1}(N)$ is contained in some bounded set V. It follows that the image under M of a neighborhood of (ϵ, r^*, s) in $I\!R_+ \times I\!R^N \times I\!R^m$ is bounded; thus M is locally bounded at (ϵ, r^*, s). If we consider (ϵ_n, r_n^*, s_n) converging to (ϵ, r^*, s) and let $(x_n, p_n^*) \in M(\epsilon_n, r_n^*, s_n)$ with (x_n, p_n^*) converging to (x_0, p_0^*), then for each n we have

$$r_n^* + A^* p_n^* \in \partial_{\epsilon_n} f(x_n), \tag{3.6}$$

and

$$s_n = A x_n - a. \tag{3.7}$$

Now (3.6) can be rewritten as

$$\epsilon_n \geq f(x_n) + f^*(r_n^* + A^* p_n^*) - \langle r_n^* + A^* p_n^*, x_n \rangle;$$

taking the limit and using the lower semicontinuity of f and f^* we find that

$$\epsilon \geq f(x_0) + f^*(r^* + A^* p_0^*) - \langle r^* + A^* p_0^*, x_0 \rangle;$$

that is, $r^* + A^* p_0^* \in \partial_\epsilon f(x_0)$, while we have $s = A x_0 - a$ from (3.7). Hence $(x_0,\, p_0^*) \in M(\epsilon, r^*, s)$, and therefore M is closed at (ϵ, r^*, s). But this shows that M is Hausdorff usc at (ϵ, r^*, s), as claimed.

Thus it remains to show that (3.4) and (3.5) describe the pairs (r^*, s) in $\operatorname{int} \operatorname{im} G_\epsilon$. Applying Theorem 3.3 with

$$H\begin{pmatrix} x \\ p^* \end{pmatrix} := \begin{pmatrix} 0 & -A^* \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ p^* \end{pmatrix} + \begin{pmatrix} 0 \\ -a \end{pmatrix},$$

we find that

$$\operatorname{int} \operatorname{im} G_\epsilon = \operatorname{int} \{ (\operatorname{im} \partial g) + H(\operatorname{dom} \partial g) \},$$

where $g(x, p^*) := f(x)$. Now

$$\operatorname{im} \partial g = \operatorname{im}[(\partial f) \times \{0\}] = (\operatorname{im} \partial f) \times \{0\} = (\operatorname{dom} \partial f^*) \times \{0\},$$

and

$$H(\operatorname{dom} \partial g) = \begin{pmatrix} 0 & -A^* \\ A & 0 \end{pmatrix} \begin{bmatrix} \operatorname{dom} \partial f \\ I\!R^m \end{bmatrix} + \begin{pmatrix} 0 \\ -a \end{pmatrix}$$

$$(\operatorname{im} A^*) \times [A(\operatorname{dom} \partial f) - a].$$

Therefore

$$\begin{aligned}\text{int im}\, G_\epsilon &= \text{int}\,\{[\text{dom}\,\partial f^* + \text{im}\, A^*] \times [A(\text{dom}\,\partial f) - a]\}\\ &= \{\text{int}[\text{dom}\,\partial f^* + \text{im}\, A^*]\} \times \{\text{int}[A(\text{dom}\,\partial f) - a]\}\,.\end{aligned}$$

Now we always have

$$\begin{aligned}\text{ri}[\text{dom}\,\partial f^* + \text{im}\, A^*] &= \text{ri dom}\,\partial f^* + \text{im}\, A^*\\ &= \text{ri dom}\, f^* + \text{im}\, A^*\\ &= \text{ri}[\text{dom}\, f^* + \text{im}\, A^*],\end{aligned}$$

so these two sets have the same affine hull. Thus int $[\text{dom}\,\partial f^* + \text{im}\, A^*] =$ int $[\text{dom}\, f^* + \text{im}\, A^*]$. A similar argument using the relation ri $A(C) = A(\text{ri}\ C)$ establishes that int $[A(\text{dom}\,\partial f) - a] =$ int $[A(\text{dom}\, f) - a]$. Therefore,

$$\text{int im}\ G_\epsilon = \text{int}\ [\text{dom}\, f^* + \text{im}\, A^*] \times \text{int}\ [A(\text{dom}\, f) - a],$$

as required. ∎

Theorem 3.4 gives a general criterion for Hausdorff usc of the solutions to the ϵ–first–order optimality conditions. In the next section we apply this criterion to establish conditions for convergence of bundle–based decomposition.

4. Application: convergence of the BBD method.

In this section we apply Theorem 3.4 to prove convergence of the bundle–based decomposition method discussed in Section 2. In terms of the notation of that theorem, we want to prove that M is Hausdorff usc at (0,0,0) relative to $\mathbb{R}_+ \times \mathbb{R}^N \times \mathbb{R}^m$. Therefore we need to verify (3.4) for $r^* = 0$ and (3.5) for $s = 0$. Condition (3.4) says that

$$\begin{aligned}0 &\in \text{int}[\text{dom}\, f^* \,+\, \text{im}\, A^*]\\ &= \text{int}[\text{dom}\, f^* \,-\, \text{dom}\, I_{L^*}]\end{aligned}$$

where L^* is the subspace $\text{im}\, A^*$ and I denotes the indicator function. This is equivalent (e.g., by [8, Lemma 6]) to:

$$(\text{rec}\ f)(v) \,+\, (\text{rec}\ I^*_{L^*})(v) \,>\, 0 \quad \text{if} \quad v \neq 0,$$

where rec f denotes the recession function of f. Since $I^*_{L^*}$ is positively homogeneous, it is its own recession function; as it also equals I_L, where $L = \ker A$, we see that (3.4) with $r^* = 0$ is equivalent to the assertion that f has no directions of recession in $\ker A$. From [13, Th. 8.7] we find that this is equivalent to the following compact–level–set condition:

$$\begin{array}{l}\text{For each real } \gamma, \text{ the} \\ \text{set } \{x \mid Ax = a, f(x) \le \gamma\} \text{ is} \\ \text{compact.}\end{array} \tag{4.1}$$

Condition (3.5) with $s = 0$ is directly interpretable as the following Slater–type condition:

$$\begin{array}{l}\text{For any } d \text{ near } 0, \text{ the system} \\ Ax = a - d \text{ has a solution } x \in \operatorname{dom} f.\end{array} \tag{4.2}$$

It is worth noting that (4.1) and (4.2) are strengthened forms of, respectively, the conditions (2.2) and (2.1) used in development of the BBD method; essentially, "ri" has been replaced by "int". The following theorem shows that this strengthening enables us to conclude *a priori* that the method is convergent.

THEOREM 4.1. *Let $f_i\,(i = 1, \ldots, n)$ be closed proper convex functions from $\mathbb{R}^{n_i}$ to $(-\infty, +\infty]$ and let A_i be linear transformations from $\mathbb{R}^{n_i}$ to $\mathbb{R}^m$, with $a \in \mathbb{R}^m$. Assume the following:*

(i) For each d near 0 in $\mathbb{R}^m$, the system

$$\sum_{i=1}^{n} A_i x_i = a - d, \quad x_i \in \operatorname{dom} f_i\,(i = 1, \ldots, n)$$

is solvable.

(ii) For each real γ the set

$$\left\{(x_1, \ldots, x_n) \,\middle|\, \sum_{i=1}^{n} A_i x_i = a,\ \sum_{i=1}^{n} f_i(x_i) \le \gamma\right\}$$

is bounded.

Then for each $\eta > 0$ there exist $\delta > 0$ and $\epsilon > 0$ such that if $\hat{x}_1, \ldots, \hat{x}_n, \hat{p}^*$, and d satisfy (2.5) – (2.10), then there exist $\bar{x}_1, \ldots, \bar{x}_n$ and $\bar{p}^*$ such that $(\bar{x}_1, \ldots, \bar{x}_n)$ minimizes $\sum_{i=1}^{n} f_i(x_i)$ on the set $\left\{(x_1, \ldots, x_n) \,\middle|\, \sum_{i=1}^{n} A_i x_i = a\right\}$, and $\bar{p}^*$ maximizes the function

$$g(p^*) := \langle p^*, a\rangle - \sum_{i=1}^{n} f_i^*(A_i^* p^*),$$

and such that

$$\|\hat{x}_1 - \bar{x}_1\| < \eta, \ldots, \|\hat{x}_n - \bar{x}_n\| < \eta, \; \|\hat{p}^* - \bar{p}^*\| < \eta.$$

PROOF: (i) and (ii) are equivalent to (4.2) and (4.1) respectively, and we have shown these to be equivalent to (3.5) with $s = 0$ and (3.4) with $r^* = 0$. Applying Theorem 3.4 with $\epsilon = 0$, $r^* = 0$, and $s = 0$, we find that the multifunction M defined by (2.17) is Hausdorff usc at (0,0,0) relative to $\mathbb{R}_+ \times \mathbb{R}^N \times \mathbb{R}^m$. This means that if ϵ and δ are taken to be small enough positive numbers, and if $\|d\| \leq \delta$ as required by (2.7), then each point of $M(\epsilon, 0, -d)$ will lie within any preassigned positive distance from the set $M(0,0,0)$. But $M(0,0,0)$ is the set $\{(\bar{x}_1, \ldots, \bar{x}_n), \bar{p}^*\}$ having the optimality properties claimed in the statement of Theorem 4.1, and $M(\epsilon, 0, -d)$ contains, by Proposition 2.1, all $\{(\hat{x}_1, \ldots, \hat{x}_n), \hat{p}^*\}$ satisfying (2.5) - (2.10). ∎

References

[1] H. Brezis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Mathematics Studies No. 5 (North-Holland, Amsterdam, 1973).

[2] H. Brezis and A. Haraux, "Image d'une somme d'opérateurs monotones et applications", *Israel Journal of Mathematics* 23 (1976) 165-186.

[3] J. K. Ho and E. Loute, "An advanced implementation of the Dantzig-Wolfe decomposition algorithm for linear programming". in: G. B. Dantzig, M. A. H. Dempster, and M. Kallio, eds., *Large-Scale Linear Programming* (International Institute for Applied Systems Analysis, Laxenburg, Austria, 1981) pp. 425-460.

[4] J. K. Ho and E. Loute, "DECOMP User's Guide", unpublished manuscript.

[5] T. Kato, "Demicontinuity, hemicontinuity, and monotonicity". *Bull. Amer. Math. Soc.* 70 (1964) 548-550.

[6] T. Kato, "Demicontinuity, hemicontinuity, and monotonicity. II.". *Bull. Amer. Math. Soc.* 73 (1967) 886-889.

[7] C. Lemaréchal, J.-J. Strodiot, and A. Bihain, "On a bundle algorithm for nonsmooth optimization", in: O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., *Nonlinear Programming 4* (Academic Press, New York, 1981) pp. 245-282.

[8] L. McLinden and R. C. Bergstrom, "Preservation of convergence of convex sets and functions in finite dimensions", *Trans. Amer. Math. Soc.* 268 (1981) 127-142.

[9] D. Medhi, "Decomposition of structured large scale optimization problems and parallel computing". Dissertation, Computer Sciences Department, University of Wisconsin—Madison (Madison, WI 53706, 1987).

[10] B. A. Murtagh and M. A. Saunders, "MINOS 5.0 User's Guide". Technical Report No. SOL 83-20, System Optimization Laboratory, Department of Operations Research, Stanford University (Stanford, CA 94305, December 1983).

[11] S. M. Robinson, "Bundle-based decomposition: Description and preliminary results". in: A. Prékopa, J. Szelezsán, and B. Strazicky, eds., *System Modelling and Optimization* (Springer-Verlag, Berlin, 1986) pp. 751-756.

[12] R. T. Rockafellar, "Local boundedness of nonlinear, monotone operators", *Michigan Mathematical Journal* 16 (1969) 397-407.

[13] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[14] J.-J. Strodiot, V. H. Nguyen, and N. Heukemes, " ϵ –optimal solutions in nondifferentiable convex programming and some related questions", *Mathematical Programming* 25 (1983) 307-328.