Multiple cuts in separating plane algorithms

Evgeni Nurminski*

Far Eastern Federal University, School of Natural Sciences,
Ajax St., Vladivostok, Russky Island, Russia

Abstract. This paper presents an extended version of the separation plane algorithm for subgradient-based finite-dimensional nondifferentiable convex optimization. The extension introduces additional cuts for epigraph of the conjugate of objective function which improve the convergence of the algorithm. The case of affine cuts is considered in more details and it is shown that it requires solution of an auxiliary convex subproblem the dimensionality of which depends on the number of additional cuts and can be kept arbitrary low. Therefore algorithm can make use of the efficient algorithms of low-dimensional nondifferentiable convex optimization which overcome known computational complexity bounds for the general case.

Keywords: convex optimization, conjugate function, cutting plane, separating plane, center of gravity algorithm

Introduction and Notations

We consider a finite-dimensional nondifferentiable convex optimization (NCO) problem

$$\min_{x \in E} f(x) = f_\star = f(x^\star), x^\star \in X_\star\,, \tag{1}$$

where E denotes a finite-dimensional space of primal variables and $f : E \to \mathbb{R}$ is a finite convex function, not necessarily differentiable. As we are interested in computational issues related to solving (1) mainly we assume that this problem is solvable and has nonempty set of solutions $X_\star$.

This problem enjoys a considerable popularity due to its important theoretical properties and numerous applications in large-scale structured optimization, Lagrange relaxation in discrete optimization, exact penalization in constrained optimization, and others. This led to the development of different algorithmic ideas, starting with the subgradient algorithm due to Shor [1] and Polyak [2] and followed by cutting plane [3], conjugate subgradient [4], bundle methods [13], ellipsoid and space dilatation [5–7], ϵ-subgradient methods [8, 9], VU-methods [10] and many others. This paper describes an extended version of the separation plane algorithm (SPA) [14] which differs from the original idea in that it introduces several additional cuts for epigraph of the conjugate of objective function. The simplest form of SPA with just one additional cut was considered

* This work is supported by RFBR grant 13-07-1210

in all details including computational experiments in [15–17]. The positive experience with this algorithm raised some hopes that introduction of more cuts will improve the computational efficiency further on.

Throughout the paper we use the following notations: $\dim(E)$ is the dimensionality of E, $|I|$ is the cardinality of a finite set I, xy is the inner product of x, y from E, $\|x\| = \sqrt{xx}$. The set of nonnegative vectors of E is denoted as E_+ or E^n_+ if the dimensionality n of E has to be specified.

We use also the distance function $\mathrm{dist}(X,Y) = \inf_{x\in X, y\in Y} \|x-y\| = \mathrm{dist}(Y,X)$ between $X \subset E, Y \subset E$. If X is a singleton $\{x\}$ we will write just $\mathrm{dist}(x,Y)$.

A vector of ones of a suitable dimensionality is denoted by $e = (1,1,\dots,1)$. A standard simplex $\{x : x \geq 0, xe = 1\}$ with $x \in E, \dim(E) = n$ is denoted by Δ_n.

1 Separating Plane Algorithms

One of the ways to represent the popular bundle [13] and the other methods of NCO is to view them as a projection algorithms for computing

$$f^\star(0) = -\min_x f(x) = -f_\star = -\inf_{(0,\mu)\in \mathrm{epi}\, f^\star} \mu\,,$$

where $f^\star(g) = \sup_x\{xg - f(x)\}$ is a Fenchel-Moreau conjugate of f, $\mathrm{epi}\, f^\star = \{(g,\mu') : \mu \geq f^\star(g)\} \subset E^\star \times \mathbb{R}$ is the epigraph of $f^\star(g)$, and $g \in E^\star$, the space of conjugate variables (gradients). This idea, presented originally in [14], unifies a number of known NCO techniques and suggests some new computational ideas.

The general idea of SPA is to bound the epigraph $\mathrm{epi}\, f^\star$ of the conjugate function $f^\star$ from below and above (in terms of set-theoretical inclusion) by the approximations L_f and U_f:

$$L_f \subset \mathrm{epi}\, f^\star \subset U_f\,.$$

These approximations provide lower and upper estimates for $f^\star(0)$:

$$\inf_{(0,\mu)\in U_f} \mu = v_U \leq -f^\star(0) \leq \inf_{(0,\mu)\in L_f} \mu = v_L \tag{2}$$

and are gradually refined in the vicinity of the vertical axis $\{0\} \times \mathbb{R} \subset E^\star \times \mathbb{R}$ to make at least one of v_U or v_L converge to $f^\star(0)$.

The iterations of SPA consist in recursive application of the update procedure to L_f and U_f which is given in more details further on. This procedure is based on computed values of conjugate function $f^\star$ at certain points of the conjugate space, determined by the procedure itself. As a result at k-th iteration of SPA we have the bundle of accumulated information on $\mathrm{epi}\, f^\star$ which consists of pairs of conjugate variables and values of conjugate function at these points. This bundle will be denoted as $\mathcal{B}^\star_I = \{(g^i, f^\star(g^i)), i \in I\}$ where $I = \{1,2,\dots,k\}$ and $g^i, f^\star(g^i)$ are conjugate variables and the value of conjugate function, computed at i-th iteration. In other words $\mathcal{B}^\star_I$ contains all information available up to the current iteration k, however some selection can be performed to save memory.

For technical reasons we assume also that $\mathcal{B}_I^\star$ contains a special pair $(0, \alpha)$ with $\alpha > f^\star(0)$. In terms of the original problem (1) it means that we assume a certain lower bound $-\alpha$ for $f_\star$ to be known. It may be a very crude estimate and introduced mainly for formal reasons, but it is necessary to avoid in a simplest way certain ill-defined subproblems in the algorithm. Notice that by construction $(0, \alpha) \in \operatorname{epi} f^\star$.

The points in the bundle $\mathcal{B}_I^\star$ have their natural counterparts $\{(x^i, f(x^i)), i \in I\}$ in the extended space of primal variables $E \times \mathbb{R}$ with $g^i \in \partial f(x^i), f^\star(g^i) = x^i g^i - f(x^i)$. In fact the algorithms based on the bundle $\mathcal{B}_I^\star$ can be considered as based on the primal bundle $\mathcal{B}_I = \{(x^i, f(x^i)), i \in I\}$ and operating on the primal variables and the original objective function. Notice that the bundle $\mathcal{B}_I$ provides information on the support function of $\operatorname{epi} f^\star$, that is the hyperplane

$$P_i = \{(g, \mu) : g\hat{x}^i - \mu = f(\hat{x}^i) = \sup_{(g,\mu)\in \operatorname{epi} f^\star} \{g\hat{x}^i - \mu\}\} \tag{3}$$

is a supporting plane of $\operatorname{epi} f^\star$ at the point $(g^i, f^\star(g^i))$.

Due to convexity the natural way to construct L_f and U_f is to use the inner and outer approximations:

$$L_f = \operatorname{co}\{(g^i, f^\star(g^i)), i \in I\} + \{0\} \times \mathbb{R}_+ \subset \operatorname{epi} f^\star \tag{4}$$

and

$$U_f = \cap H_i,\ i \in I \supset \operatorname{epi} f^\star \tag{5}$$

where

$$H_i = \{(g, \mu) : \mu \geq f^\star(g^i) + x^i(g - g^i), x^i \in \partial(g^i)\} \supset \operatorname{epi} f^\star$$

are the half-spaces, generated by supporting planes P_i (3) to $\operatorname{epi} f^\star$ at the points $(g^i, f^\star(g^i))$.

The general scheme to update L_f and U_f at k-th iteration with $I = \{1, 2, \ldots, k\}$ is described in the Algorithm 1.

For better understanding the sequence of major steps in the update process is illustrated on Fig. 1–4.

From computational point of view the separating plane $H_{\hat{x}}$ in the **Step 2 (Separate)** can be obtained for the finite value of v_U by solving the projection problem

$$\min_{(z,\mu)\in L_{f^\star}} \|z\|^2 + (v_U - \mu)^2 = \|\hat{z}\|^2 + (v_U - \hat{\mu})^2 \tag{7}$$

and appropriate normalization: $\hat{x} = -\hat{z}/(v_U - \hat{\mu})$.

The **Support** step of the algorithm is just the computation of the objective function and its subgradient at the point $\hat{x}$ as demonstrated by (3).

Notice that after the update of U_f in any way we obtain a new upper estimate for $f_\star$ which is not worse that the previous:

$$v'_U = \inf_{(0,\mu)\in U_f \cap S^\star_{\hat{x}}} \mu \geq \max\{ \inf_{(0,\mu)\in U_f} \mu,\ \inf_{(0,\mu)\in S^\star_{\hat{x}}} \mu\} = \max\{v_U, -f(\hat{x})\} \geq v_U$$

Data: The bundle $\mathcal{B}_I^\star$, the upper and low approximations U_f, L_f of $\operatorname{epi} f^\star$.
Result: The updated: set I, approximations L_f, U_f and the bundle $B_f^\star$.
Step 1. Estimate: estimate the lower bound for $f^\star(0)$. Compute

$$v_U = \inf_{(0,\mu)\in U_f} \mu \leq= \inf_{(0,\mu)\in \operatorname{epi} f^\star} \mu = f^\star(0)\,.$$

It can be set to $-\infty$ if U_f is taken to be the trivial upper approximation $E \times \mathbb{R}$ at the start of SPA.
Step 2. Separate: strictly separate $(0, v_U)$ from L_f with a separating plane $S_{\hat{x}} = \{(g,\mu) : g\hat{x} - \mu = -\hat{v}_U\}$, parameterized by the support vector $(\hat{x}, -1)$ and $\hat{v}_U$ to be found. If $v_U = -\infty$ just take an arbitrary $\hat{x}$. If strict separability is impossible, that is $\hat{v}_U = f^\star(0) = -f_\star$, then we are done, otherwise continue.
Step 3. Support: for a given $\hat{x}$, found at the previous step, find the supporting hyperplane $P_{\hat{x}}^\star$ for $\operatorname{epi} f^\star$:

$$\begin{array}{c} P_{\hat{x}}^\star = \{(g,\mu) : g\hat{x} - \mu = \sup_{(g,\epsilon)\in \operatorname{epi} f^\star}\{\hat{x}g - \epsilon\} = \\ \sup_g\{\hat{x}g - f^\star(g)\} = \hat{x}\hat{g} - f^\star(\hat{g}) = f(\hat{x})\} \end{array} \tag{6}$$

with $\hat{g} \in \partial f(\hat{x})$. Notice, that this is just the calculation of $f(\hat{x})$ and $\hat{g} \in \partial f(\hat{x})$.
The hyperplane $P_{\hat{x}}^\star$ defines the "upper" half-space $H_{\hat{x}}^\star$ which contains $\operatorname{epi} f^\star$:

$$\begin{array}{c} H_{\hat{x}}^\star = \{(g,\mu) : \mu \geq g\hat{x} - f(\hat{x})\} \supset \{(g,\mu) : \mu \geq \sup_x\{gx - f(x)\}\} = \\ \{(g,\mu) : \mu \geq f^\star(g)\} = \operatorname{epi} f^\star \end{array}$$

and hence $H_{\hat{x}}^\star$ can be safely added to the cuts of the upper approximation U_f.
Step 4. Update: perform the update of the basic data structures of SPA:
the bundle: $\mathcal{B}_I^\star \to \mathcal{B}_I^\star \cap \{(\hat{g}, f^\star(\hat{g})\}$,
the approximations: redefine L_f and U_f according to (4) and (5)

$$L_f \to \operatorname{co}(L_f, (\hat{g}, \hat{\epsilon})), \quad U_f \to U_f \cap S_{\hat{x}}^\star$$

the index set: $I \to I \cup \{k+1\}$.

Algorithm 1: The generic structure of update step for the upper and low approximations of $\operatorname{epi} f^\star$

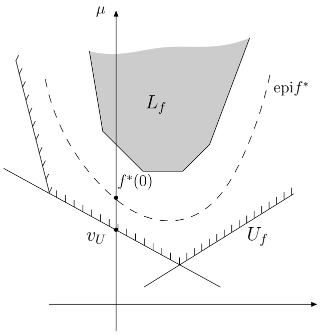

Fig. 1. Basic algorithm objects: L_f, U_f are lower and upper approximations, v_U approximates $f^\star(0)$ from below.

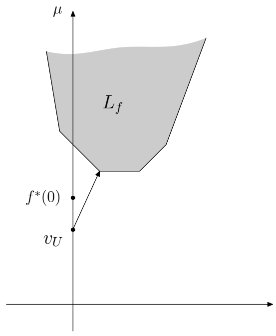

Fig. 2. Projection: determines the (normalized) vector $(\hat{x}, -1)$ such that $g\hat{x} - \mu \leq -v_U$ for any $(g, \mu) \in \operatorname{epi} f^\star$.

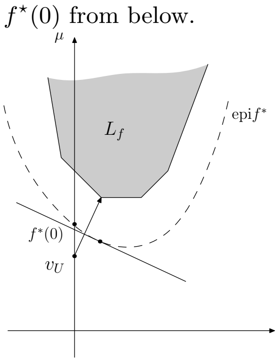

Fig. 3. Support: compute $\sup_{(g,\mu)\in \operatorname{epi} f^\star}\{\hat{x}g - \mu\} = f(\hat{x})$ and the corresponding subgradient $\hat{g} \in \partial f(\hat{x})$.

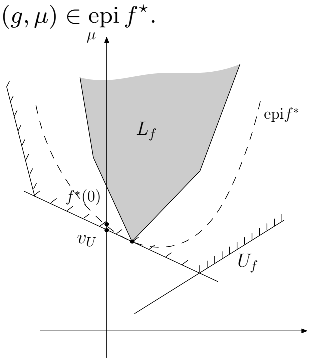

Fig. 4. Update: the lower L_f and the upper U_f approximations are updated with the help of a new $(g, f^\star(g))$ and cutting support plane at $(g, f^\star(g))$.

and may be better if $f(\hat{x})$ sets a new record. Unfortunately we can not guarantee that this will be just the case and so the algorithm is not monotone in terms of the objective function. This may be one of the factors which slows down the practical convergence of SPA, and it seems to be possible to improve it by adding an additional cut or cuts on epi $f^\star$.

That was the original idea, tested with positive results in [15, 16] when just the single extra cut generated by the auxiliary subproblem of cutting plane method was added. Here we consider some aspects of adding several extra cuts.

2 Multiple Additional Cuts

From the formal point of view the additional cuts for epi $f^\star$ can be considered as a a certain subset Q of $E \times \mathbb{R}$ which is superimposed on epi $f^\star$. It means that now instead of epi $f^\star$ in the **Support** step of the Algorithm 1. we are going to use epi $f^\star \cap Q$

In this case a new supporting hyperplane $\bar{H}^\star_{\hat{x}} = \{(g, \mu) : g\hat{x} - \mu = \bar{\mu}\}$ will have $\bar{\mu} \geq \hat{\mu}$:

$$-\bar{\mu} = \sup_{(g,\mu)\in \mathrm{epi}\, f^\star \cap Q} \{g\hat{x} - \mu\} \leq \sup_{(g,\mu)\in \mathrm{epi}\, f^\star} \{g\hat{x} - \mu\} = -\hat{\mu} = f(\hat{x})$$

and therefore we have a better chance to improve v'_U:

$$\bar{v}'_U = \max\{v_U, \bar{\mu}\} \geq \max\{v_U, \hat{\mu}\} = v'_U$$

There is a great flexibility in the choice of Q, the only essential requirement is to ensure that the solution $(0, -f_\star)$ still belongs to epi $f^\star \cap Q$.

The updated iteration of the separating plane algorithms with cuts is represented in Algorithm 2.

From practical point of view it is convenient to have Q described by a system of convex inequalities $Q = \{(g, \mu) : h_i(g, \mu) \leq 0, i = 1, 2, \dots, m\}$, each of which can be considered as a separate cut, applied to epi $f^\star$. Therefore we call this type of algorithms as separating plane algorithm with multiple cuts (SPA-MC).

In the simplest case all $h_i(g, \mu)$ are affine functions:

$$h_i(g, \mu) = \hat{x}^i g + \mu - \bar{\mu}_i\,, \tag{9}$$

where $\hat{x}^i$ represent some trial points in the space of the original primal variables.

The support problem of the **Step 3** in SPA-MC for the case of affine cuts can be written as

$$\begin{gathered} w_U = \sup\{xg - \mu] \\ \mu \geq f^\star(g) \\ \hat{x}^i g + \mu \leq \bar{\mu}_i, i = 1, 2, \dot{,} m \end{gathered} \tag{10}$$

which can be transformed into the dual form

$$w_U = \sup_{\mu \geq f^\star(g)} \inf_{\lambda \geq 0} \{xg - \mu - \sum_{i=1}^{m} \lambda_i(\hat{x}^i g + \mu - \bar{\mu}_i)\}\,, \tag{11}$$

Data: The bundle $\mathcal{B}^\star_I$, the upper and low approximations U_f, L_f of $\operatorname{epi} f^\star$, and the cut $Q \subset E \times \mathbb{R}$.
Result: The updated index set I, approximations L_f, U_f and the bundle $B^\star_f$.
Step 1. Estimate: Unchanged.
Step 2. Separate: Unchanged.
Step 3. Support: Modified to include the cut Q. For a given $\hat{x}$, found at the previous step, find the supporting hyperplane $H^\star_{\hat{x}}$ for $\operatorname{epi} f^\star \cap Q$:

$$H^\star_{\hat{x}} = \{(g,\mu) : g\hat{x} - \mu = \sup_{\substack{(g,\epsilon) \in \operatorname{epi} f^\star \\ (g,\epsilon) \in Q}} \{\hat{x}g - \epsilon\}\} \,. \tag{8}$$

The details of these calculations depend upon the definition of the cut set Q and are discussed further on.
The hyperplane $H^\star_{\hat{x}}$ defines the "upper" half-space $S^\star_{\hat{x}}$ which contains $\operatorname{epi} f^\star$:

$$S^\star_{\hat{x}} = \{(g,\mu) : \mu \ge g\hat{x} - f(\hat{x})\} \supset \{(g,\mu) : \mu \ge \sup_x\{gx - f(x)\}\} = \{(g,\mu) : \mu \ge f^\star(g)\} = \operatorname{epi} f^\star$$

and hence $S^\star_{\hat{x}}$ can be safely added to the cuts of the upper approximation U_f.
Step 4. Update: Unchanged.

Algorithm 2: The generic structure of update step for the upper and low approximations of $\operatorname{epi} f^\star$ in SPA with multiple cuts.

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is a nonnegative vector of Lagrange multipliers.

By convexity

$$\begin{aligned}
w_U = &\inf_{\lambda \ge 0} \sup_{\mu \ge f^\star(g)} \{xg - \mu - \textstyle\sum_{i=1}^m \lambda_i(\hat{x}^i g + \mu - \bar{\mu}_i)\} = \\
&\inf_{\lambda \ge 0} \{\textstyle\sum_{i=1}^m \lambda_i\bar{\mu}_i + \sup_{\mu \ge f^\star(g)} \{(x - \sum_{i=1}^m \lambda_i\hat{x}^i)g - (1 + \sum_{i=1}^m \lambda_i)\mu)\} = \\
&\inf_{\lambda \ge 0} \{\textstyle\sum_{i=1}^m \lambda_i\bar{\mu}_i + (1 + \sum_{i=1}^m \lambda_i) \sup_{\mu \ge f^\star(g)} \{\dfrac{x - \sum_{i=1}^m \lambda_i\hat{x}^i}{1 + \sum_{i=1}^m \lambda_i} g - \mu)\}\} = \\
&\inf_{\lambda \ge 0} \{\textstyle\sum_{i=1}^m \lambda_i\bar{\mu}_i + (1 + \sum_{i=1}^m \lambda_i) \sup_g \{\dfrac{x - \sum_{i=1}^m \lambda_i\hat{x}^i}{1 + \sum_{i=1}^m \lambda_i} g - f^\star(g)\} = \\
&\inf_{\lambda \ge 0} \{\textstyle\sum_{i=1}^m \lambda_i\bar{\mu}_i + (1 + \sum_{i=1}^m \lambda_i) f(\dfrac{x - \sum_{i=1}^m \lambda_i\hat{x}^i}{1 + \sum_{i=1}^m \lambda_i})\} = \inf_{\lambda \ge 0} \varXi(\lambda)\,,
\end{aligned}$$

where

$$\varXi(\lambda) = \sum_{i=1}^m \lambda_i\bar{\mu}_i + (1 + \sum_{i=1}^m \lambda_i) f(\frac{x - \sum_{i=1}^m \lambda_i\hat{x}^i}{1 + \sum_{i=1}^m \lambda_i})$$

has a controllable dimensionality m which is determined by the number of additional cuts and can be set to any value.

Therefore $\varXi(\lambda)$ can be minimized by specific algorithms, tailored to this particular dimensionality. An appropriate example of such algorithms is the center of gravity method (CGM) by Levin [11] and Newmann [12] which is easily implemented at least in 2-dimensional case and provides a geometric rate of convergence independent of properties of objective function and feasibility set. Hopefully the efficient and practical methods may appear or already exist, unknown to the author, in higher dimensions.

The essential part of $\Xi(\lambda)$ which may create different problems with the following minimization is the nonlinear term $(1+\sum_{i=1}^m \lambda_i) f((x-\sum_{i=1}^m \lambda_i \hat{x}^i)/(1+\sum_{i=1}^m \lambda_i))$. Fortunately it inherits a convexity of the original problem which follows from its definition as a supremum of linear forms in λ. Nevertheless it is useful for the further maximization to consider the nonlinear part of $\Xi(\lambda)$ as a generic function

$$\phi(\theta) = \left(\sum_{i=1}^m \theta_i\right) f\left(\frac{\sum_{i=1}^m \theta_i \hat{x}^i}{\sum_{i=1}^m \theta_i}\right) \tag{12}$$

for $\theta = (\theta_1, \theta_2, \ldots, \theta_m) \in E_+^m$ and $\theta \neq 0$. It makes sense to complement the definition of $\phi(\cdot)$ at 0 as $\phi(0) = 0$ without loosing the continuity. Then ϕ becomes defined on the whole E_+^m and its convexity properties are covered by the following lemma which might be of a separate interest.

Lemma 1. *Let $f : E \to \mathbb{R}$ is a convex finite function, $\hat{x}^i, i = 1, 2, \ldots m$ — a collection of m points in E, and $\theta = (\theta_1, \theta_2, \ldots, \theta_m) \in E_+^m$ — a vector of nonnegative variables. Then $\phi(\theta)$ defined by (12) is a convex function of θ on E_+^m.*

Proof. Denote $\sum_{i=1}^m \theta_i = \sigma(\theta)$. Then

$$\phi(\theta) = \sigma(\theta) f\left((\sum_{i=1}^m \theta_i \hat{x}^i)/\sigma(\theta)\right)$$

for $\sigma(\theta) > 0$ and $\phi(0) = 0$ by definition. Let $\alpha \in [0, 1]$ and $\theta', \theta'' \in E_+^m$. Next we show that $\phi(\cdot)$ satisfies the Jensen inequality $\phi(\alpha\theta' + (1-\alpha)\theta'') \leq \alpha\phi(\theta') + (1-\alpha)\phi(\theta'')$.

Notice first, that ϕ is positive homogeneous of degree 1: $\phi(\nu\theta) = \nu\phi(\theta)$ for $\nu \geq 0$ hence the case when either $\theta' = 0$ or $\theta'' = 0$ is trivial.

Assume further on that $\sigma(\theta')\sigma(\theta'') > 0$. Let us fix α and denote $\kappa = \alpha\sigma(\theta') + (1-\alpha)\sigma(\theta'') > 0$. Then

$$\begin{gathered} \phi(\alpha\theta' + (1-\alpha)\theta'') = \kappa f\left((\alpha \textstyle\sum_{i=1}^m \theta_i' x^i + (1-\alpha)\sum_{i=1}^m \theta_i'' x^i)/\kappa\right) = \\ \kappa f\left(\alpha(\textstyle\sum_{i=1}^m \theta_i' x^i)/\kappa + (1-\alpha)(\sum_{i=1}^m \theta_i'' x^i)/\kappa\right) = \\ \kappa f\left(\alpha \frac{\sum_{i=1}^m \theta_i' x^i}{\sigma(\theta')} \frac{\sigma(\theta')}{\kappa} + (1-\alpha) \frac{\sum_{i=1}^m \theta_i'' x^i}{\sigma(\theta'')} \frac{\sigma(\theta'')}{\kappa}\right) = \kappa f(\gamma' \bar{x}' + \gamma'' \bar{x}''), \end{gathered}$$

where

$$\gamma' = \alpha\sigma(\theta')/\kappa\,, \quad \gamma'' = \alpha\sigma(\theta'')/\kappa\,,$$

and

$$\bar{x}' = \sum_{i=1}^m \theta_i' x^i/\sigma(\theta')\,, \quad \bar{x}'' = \sum_{i=1}^m \theta_i'' x^i/\sigma(\theta'')\,,$$

As $\gamma' + \gamma'' = \alpha\sigma(\theta')/\kappa + (1-\alpha)\sigma(\theta'')/\kappa = 1$ and $\gamma', \gamma'' \geq 0$. then

$$\phi(\alpha\theta' + (1-\alpha)\theta'') \leq \kappa f(\gamma'\bar{x}' + \gamma''\bar{x}'')) \leq \kappa(\gamma' f(\bar{x}') + \gamma'' f(\bar{x}'')) =$$
$$\alpha\sigma(\theta')\kappa f(\bar{x}')/\kappa + (1-\alpha)\sigma(\theta'')\kappa f(\bar{x}'')/\sigma(\theta'')/\kappa = \alpha\phi(\theta') + (1-\alpha)\phi(\theta'') \,.$$

which completes the proof.

By setting $z^1 = x$, $z^{i+1} = -\hat{x}^i, i = 1, 2, \ldots, m$ and applying lemma 1 to $\phi(\theta) = \sigma(\theta) f\left((\sum_{i=1}^{m+1} \theta_i z^i)/\sigma(\theta)\right)$ with $\theta_1 = 1$ we obtain convexity of $\Xi(\theta)$.

3 Convergence

The following theorem establishes the convergence of SPA-MC.

Theorem 1. *Let $\{v_U^k\}$ be the sequence of lower estimates*

$$v_U^k = \inf_{(0,\mu)\in U_f^k} \mu \leq f^\star(0)\,,$$

of the optimal value in the problem (1) which are generated by SPA-MC as prescribed by Algorithms 1-2. Then

$$\lim_{k\to\infty} v_U^k = f^\star(0) = -\min_x f(x)\,.$$

Proof. Let $k = 1, 2, \ldots$ number the sequence of the update iterations of SPA-MC which are prescribed by Algorithms 1 and 2, and let U_f^k, L_f^k are the corresponding upper and lower approximations of $\operatorname{epi} f^\star$ at the beginning of k-th iteration. Naturally, updated U_f^k, L_f^k become U_f^{k+1}, L_f^{k+1}.

By construction $U_f^k \supset \operatorname{epi} f^\star \supset L_f^k$ and

$$\operatorname{epi} f^\star \subset U_f^{k+1} \subset U_f^k, \quad L_f^k \subset L_f^{k+1} \subset \operatorname{epi} f^\star$$

so both these sequences have Kuratovski limits, which we denote as $U_f^\bullet, L_f^\bullet$ respectively.

Observe that $v_U^k \leq v_U^{k+1} \leq f^\star(0)$ hence the sequence $\{v_U^k\}$ has a limit which we denote as $v_U^\bullet$.

Convergence of SPA-MC means that $v_U^k \to f^\star(0)$ or, equivalently, $\operatorname{dist}(\bar{v}_U^k, \operatorname{epi} f^\star) = \operatorname{dist}((0, v_U^k), \operatorname{epi} f^\star) \to 0$ when $k \to \infty$. As $\operatorname{dist}(\bar{v}_U^k, \operatorname{epi} f^\star) \leq \operatorname{dist}(\bar{v}_U^k, L_f^k)$ it is sufficient to show that $\operatorname{dist}(\bar{v}_U^k, L_f^k) \to 0$.

Denote $\bar{V}_U^k = v_U^k - \{0\} \times \mathbb{R}_+$ and notice that

$$\operatorname{dist}(\bar{v}_U^k, L_f^k) = \operatorname{dist}(\bar{v}_U^k - 0 \times R_+, L_f^k) = \operatorname{dist}(\bar{V}_U^k, L_f^k)\,.$$

As $\bar{V}_U^{k+1} \supset \bar{V}_U^k$ and $L_f^{k+1} \supset L_f^k$ the distance $\operatorname{dist}(\bar{V}_U^k, L_f^k)$ is non-increasing:

$$\operatorname{dist}(\bar{V}_U^{k+1}, L_f^{k+1}) \leq \operatorname{dist}(\bar{V}_U^k, L_f^k) \leq \operatorname{dist}(\bar{V}_U^0, (0,\kappa) + 0 \times \mathbb{R}_+) = \|v_U^0 - \kappa\|$$

hence the norms of all vectors $z^k = \Pi_{L_f^k}(\bar{V}_U^k) - \bar{V}_U^k)$ are uniformally bounded and have the same limit $\rho_z = \lim_{k\to\infty} \|z^k\|$. The key question is however what is the value of ρ_z. If $\rho_z = 0$ then $\mathrm{dist}(\bar{V}_U^k, \mathrm{epi}\, f^\star) \to 0$ and

$$\lim_{k\to\infty} v_U^k = v_U^\bullet = f^\star(0)$$

which establishes convergence of SPA-MC.

To show that this is just the case assume contrary: $\rho_z > 0$. Then the sequence $\{z^k\}$ due to its boundness has at least one limit point, which we denote as $z^\bullet$ with a certain subsequence $\{z^{k_t}, t = 1, 2, \dots\} \to z^\bullet$.

The **Support** and **Update** steps of the Algorithms 1-2 redefines v_U^k in a following way:

1. Solve
$$\inf_{\bar{g}=(g,\mu)\in \mathrm{epi}\, f^\star \cap Q_k} z^k \bar{g} = z^k \bar{g}^k = \gamma_k \,,$$
where $\bar{g}^k = (g^k, \mu_k)$.
2. If $\gamma_k > v_U^k$ redefine $v_U^{k+1} = \gamma_k$. Otherwise $v_U^{k+1} = v_U^k$.

In any case $\bar{g}^k z^k \leq \bar{v}_U^{k+1} z^k$ and passing in this inequality to the limit along the subsequence where all limits exist obtain

$$\bar{g}^\bullet z^\bullet \leq v_U^\bullet z^\bullet \,. \tag{13}$$

On the other hand as z^k is obtained by projection of $(0, v_U^k)$ on L_f^k and taking into account that $\bar{g}^k \in L_f^{k+1}$ we have $(\bar{g}^k - \bar{v}^{k+1})z^{k+1} \geq \|z^{k+1}\|^2$. Passing to the limit gives $(\bar{g}^\bullet - \bar{v}^\bullet)z^\bullet \geq \|z^\bullet\|^2 \geq \gamma > 0$ or

$$\bar{g}^\bullet z^\bullet \geq \bar{v}_U^\bullet z^\bullet + \gamma > \bar{v}_U^\bullet z^\bullet \,. \tag{14}$$

Obviously (13) and (14) contradict each other and it proves the theorem.

Conclusion

We present in this work the general scheme for modification of separating plane algorithms which provides additional possibilities for improving relaxational properties of algorithms of nonsmooth optimization. It is based on imposing additional cuts in the dual space of conjugate variables which resrict the test area and may additionally localize the extremum. The scheme allows also more sophisticated low-dimensional local search procedures to be applied on each iteration to speed up convergency.

References

1. Shor, N.Z.: Primenenije metoda gradientnogo spuska dlya reshenija setevoj transportnoj zadachi. Materialy nauchn. seminara po teor. i prikl. vopr. kibernetiki i issledovanija operacij, 9–17 (1962)

2. Polyak, B.T.: Minimization of Unsmooth Functionals. USSR Computational Mathematics and Mathematical Physics, 9, 14-–29 (1969)
3. Kelley, J.E.: The Cutting-Plane Method for Solving Convex Programs. Journal of the Society for Industrial and Applied Mathematics, 8(4), 703–712 (1960)
4. Wolfe, P.: A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions. Mathematical Programming Study, 3, 145-–173 (1975)
5. Shor, N.Z.: Utilization of the Operation of Space Dilatation in the Minimization of Convex Functions. Cybernetics, 6, 7-–15 (1970)
6. Shor, N.Z., Zhurbenko, N.G.: The Minimization Method Using Space Dilatation in Direction of Difference of Two Sequential Gradients. Kibernetika, No. 3, 51-59 (1971).
7. Shor, N.Z., Zhurbenko, N.G., Likhovid, A.P., & Stetsyuk, P.I.: Algorithms of Nondifferentiable Optimization: Development and Application. Cybernetics and Systems Analysis, 39(4), 537-–548 (2003)
8. Kiwiel, K.C.: An Algorithm for Nonsmooth Convex Minimization With Errors. Mathematics of Computation, 45(171), 173–180 (1985)
9. Rzhevskiy, S.V.: ϵ-Subgradient Method for the Solution of a Convex Programming Problem. USSR Computational Mathematics and Mathematical Physics, 21(5), 51-–57 (1981)
10. Mifflin, R., Sagastizábal, C.A.: A VU-algorithm for convex minimization. Mathematical Programming, 104, 583–606 (2005)
11. Levin A.Yu.: Ob odnom algoritme miniminzacii vypuklykh funkcij. DAN SSSR, 160(6), 1244–1247 (1965)
12. Newman, D.J.: Location of the Maximum on Unimodal Surfaces. Journal of the Association for Computing Machinery, 12(3), 395–398 (1965)
13. Lemarechal, C.: An extension of Davidon methods to non-differentiable problems. Mathematical Programming Study, 3, 95–109 (1975).
14. Nurminski, E.A.: Separating Plane Algorithms for Convex Optimization. Mathematical Programming, 76, 375—391 (1997)
15. Vorontsova, E.A.: A Projective Separating Plane Method with Additional Clipping for Non-smooth Optimization. WSEAS Transactions on Mathematics, 13, 115-–121 (2014)
16. Vorontsova, E.A., Nurminski, E.A.: Synthesis of Cutting and Separating Planes in a Nonsmooth Optimization Method. Cybernetics and Systems Analysis, 51(4), (2015)
17. Vorontsova, E.A.: Extended Separating Plane Algorithm and NSO-Solutions of PageRank Problem. In: Kochetov, Yu. et all (eds.) DOOR-2016. LNCS, vol. 9999, pp. YYY–ZZZ. Springer, Heidelberg (2016)