Computational Experience and Challenges with the Conjugate Epi-Projection Algorithms for Non-Smooth Optimization *

E.A. Nurminski †, N.B. Shamray ‡

20 June 2019

Abstract

This paper considers implementable versions of a conceptual convex optimization algorithm which provides a high-speed (superlinear, quadratic and finite) convergence for the broad classes of convex optimization problems.

Keywords: convex optimization, conjugate function, approximate sub-differential, superlinear convergence, quadratic convergence, finite convergence, projection, epigraph

Introduction

This work considers different issues of solving the fundamental convex optimization problem

$$\min_x f(x) \tag{1}$$

where the objective function f neither need to be a finite and/or differentiable in a classical sence.

The main idea is to consider the equivalent problem in the congugate (subgradient) space of of computing the value and subgradient of a conjugate function at zero. Convexity allows to guarantee a number of attaractive features of such approach [1]: uniform treatment of conditional and unconditional optimization problems, development of projection-type algorithms with superlinear convergence in the general case, quadratic rate of convergence in sub-quadratic case and finite convergence in the case of sharp minima.

It was further suggested in [2, 3] to impose certain additional cuts to improve the relaxation properties of the algorithm. Convergence of the resulting algoriths was proved under very general conditions however the computational efficiency of these algorithms remained under the question. Here we intend to study it at least experimentally.

1 Notations and Preliminaries

Throughout the paper we use the following notations: E is a finite dimensional euclidean space of primal variables of any dimensionality. The inner product of vectors x, y from E is denoted as xy. The cone of non-negative vectors of E is denoted as E_+. The set of real numbers is denoted as $\mathbb{R}$ and $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$.

*This work is supported by RF Ministry of Education and Science, project 1.7658.2017/6.7 and RFBR grant 18-29-03071

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

‡Institute for Automation and Control Processes, ul. Radio, 5, Vladivostok, Russia

The norm in E is defined in a standard way: $\|x\| = \sqrt{xx}$ and for $X \subset E$ $\|X\| = \sup_{x\in X} \|x\|$. This norm defines of course the standard topology on E with the common definitions of open and closed sets and closure and interior of subsets of E. The interior of a set X is denoted as $\mathrm{int}(X)$.

The unit ball in E is denoted as $B = \{x : \|x\| \leq 1\}$. The support function of a set $Z \subset E$ is denoted and defined as $(Z)_x = \sup_{z\in Z} xz$.

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

We use the standard definitions of convex analysis (see f.i. [6]) related mainly to functions $f : E \to \mathbb{R}_\infty$: the domain $\mathrm{dom}\, f$ of a function f is the set $\mathrm{dom}\, f = \{x : f(x) < \infty\}$, the epigraph $\mathrm{epi}\, f$ of a function f is a set $\mathrm{epi}\, f = \{(\mu, x) : \mu \geq f(x)\} \subset \mathbb{R}_\infty \times E$.

Further on all functions are closed convex in a sense that their epigraphs are *closed convex* subsets of $\mathbb{R}_\infty \times E$.

Definition 1 *For a convex function $f : E \to \mathbb{R}$ and fixed $x \in E$ the set $\partial f(x) = \{g : f(y) - f(x) \geq g(y - x)$ for all $y \in \mathrm{dom}\, f\}$ is called a sub-differential of f at the point x.*

The sub-differential $\partial f(x)$ of f at point x is well-defined and is a closed bounded convex set for all $x \in \mathrm{int}(\mathrm{dom}\, f)$. At the boundary of $\mathrm{dom}\, f$ it may or may not exists. The sub-differential $\partial f(x)$ is also upper semi-continuous as a multi-function of x when exists.

Definition 2 *The directional derivative of a finite convex function f at point x in direction d is denoted and defined as $\partial f(x; d) = \lim_{\delta\to+0}(f(x + \delta d) - f(x))/\delta$.*

It is well-known from convex analysis that $\partial f(x; d) = \sup_{g\in\partial f(x)} gd = (\partial f(x))_d$.

Definition 3 *For a convex function $f : E \to \mathbb{R}_\infty$ the function*

$$f^\star(g) = \sup_x\{gx - f(x)\} = (\mathrm{epi}\, f)_{\bar g}, \text{ where } \bar g = (-1, g) \in \mathbb{R}_\infty \times E \tag{2}$$

is called a conjugate function of f.

The key result of convex analysis is that for a closed convex function f

$$\sup_g\{gx - f^\star(g)\} = (\mathrm{epi}\, f^\star)_{\bar x} = f(x), \tag{3}$$

where $\bar x = (-1, x) \in \mathbb{R}_\infty \times E$.

It is also easy to see that if $(\mathrm{epi}\, f^\star)_{\bar x} = g_x x - f^\star(g_x)$ then $g_x \in \partial f(x)$ and the other way around: for $\bar g = (-1, g)$ if $(\mathrm{epi}\, f)_{\bar g} = gx_g - f(x_g)$ then $x_g \in \partial f^\star(g)$.

The trivial consequence of the Definition 3 is that $f^\star(0) = -\inf_x f(x)$ which is the key correspondence used by the conjugate epi-projection algorithm, considered further on. As the conjugate epi-projection algorithm operates in the conjugate space its convergence properties depend upon the properties of the conjugate function of the objective. Therefore we introduce some additional classes of primal functions to ensure the desired behavior of the conjugates.

Definition 4 *Convex function f is called sup-quadratic with respect to a point $x \in \mathrm{int}(\mathrm{dom}\, f)$ if there exists a constant $\tau > 0$ such that*

$$f(y) - f(x) \geq g(y - x) + \frac{1}{2}\tau\|y - x\|^2 \tag{4}$$

for any $g \in \partial f(x)$ and any y.

We will call τ the sup-quadratic characteristic of f at x. Notice that strongly convex functions are sup-quadratic at any x from their domains, however a function f, sup-quadratic at some x, need not to be strongly convex.

A symmetric definition can be given for *sub-quadratic* functions.

Definition 5 *Convex function f is called sub-quadratic with respect to a point $x \in \mathrm{int}(\mathrm{dom}\, f)$ if there exists a constant $\tau > 0$ such that*

$$f(y) - f(x) \leq g(y-x) + \frac{1}{2}\tau^{-1}\|y-x\|^2 \tag{5}$$

for any $y \in \mathrm{dom}\, f$ and some $g \in \partial f(x)$.

Notice that it follows from this definition that the function f, sub-quadratic at point x is in fact differentiable at this point. Of course not all functions differentiable at x are sub-quadratic.

From the point of view of non-smooth optimization namely sup-quadratic functions are of particular interest, and the definitions 4 and 5 establish important properties of conjugates functions for sup-quadratic primals.

Lemma 1 *Let $f : E \to \mathbb{R}$ attains its minimum value $f_\star$ at the point $x^\star$ and f is sup-quadratic at point $x^\star$ with the positive sup-quadratic characteristic τ. Then $f^\star(g)$ is sub-quadratic at $g = 0$ with the corresponding sub-quadratic characteristic not lower then τ^{-1}.*

Proof. By definition for any x

$$\frac{1}{2}\tau\|x^\star - x\|^2 \leq f(x) - f_\star = f(x) + f^\star(0) \tag{6}$$

and hence

$$f^\star(g) - f^\star(0) = x_g g - (f(x_g) + f^\star(0)) \leq x_g g - \frac{1}{2}\tau\|x^\star - x_g\|^2 \tag{7}$$

for any $x_g \in \partial f^\star(g)$. Hence

$$\begin{array}{c} f^\star(g) - f^\star(0) \leq x^\star g + (x_g - x^\star)g - \frac{1}{2}\tau\|x^\star - x_g\|^2 \leq \\ x^\star g + \sup_z\{zg - \frac{1}{2}\tau\|z\|^2\} = x^\star g + \frac{1}{2}\tau^{-1}\|g\|^2. \end{array} \tag{8}$$

■

Another interesting subclass of convex functions are those which have zero in the interior of the subdifferential at the solution $x^\star$ of (1), that is $0 \in \mathrm{int}(\partial f(x^\star))$. This condition is also known as "sharp minimum" and extended further on in [7] and others. The special attraction of this case is that the well-known proximal method has then a finite termination [8] for such problems. The conjugate epi-projection optimization algorithm has the same property which is based on the fact that the conjugate functions for the primal functions with sharp minimum have very simple behavior in the vicinity of zero.

Lemma 2 *If solution $x^\star$ of (1) is such that $0 \in \mathrm{int}(\partial f(x^\star))$ then there is $\rho > 0$ such that $f^\star(g) = gx^\star - f(x^\star)$ for $\|g\| < \rho$.*

Proof. If ρ is small enough then sharp minimum condition implies $0 \in \partial(f(x^\star) - gx^\star) = \partial f(x^\star) - g$ for any $g \in \rho B$ and therefore

$$f^\star(g) = \sup_x\{gx - f(x)\} = gx^\star - f(x^\star)$$

is a linear function of g. ■

Namely this property guarantees the finite termination of the conjugate epi-projection optimization algorithm.

For additional results on connections between sharp minimum and properties of conjugate functions see also [9].

2 Conjugate Epi-Projection Algorithm

As it was already mentioned the basic idea of the conjugate epi-projection algorithms consists in considering the convex problem (1) as the problem of computing the conjugate function of the objective at the origin:

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

We suggest to use for computing $f^\star(0)$ the algorithms based on projection onto the epigraph $\operatorname{epi} f^\star$. This idea demonstrates some promises for effective solution of (1) and suggests some new computational ideas.

The algorithms considered here consist in execution of an infinite sequence of iterations, which generates the corresponding sequence of points $\{(\xi_k, 0) \in \mathbb{R} \times E, k = 0, 1, \dots\}$ with $\xi_k \to f^\star(0)$ when $k \to \infty$. For each of these iterations they call a subgradient oracle which for any $x \in E$ computes $f(x)$ and some arbitrary $g \in \partial f(x)$. Also they require solution of nonlinear projection problems which make the algorithms, strictly speaking, unimplementable. However the analysis of the algorithm demonstrate its potential and can show the ways to its practical implementations.

We give here first the original version of a conceptual conjugate epi-projection algorithm and cite here the key results about its convergence. This is followed by a few simple numerical experiments just to provide a reference point for further modifications and to indicate some numerical problems which can arise in its straightforward implementation.

2.1 Basic computational scheme

The principal details of the iteration of the conjugate epi-projection algorithm are given on the Fig. Algorithm 1. Convergence of the Algorithm 1 is confirmed by the following theorem.

Theorem 1 *Let f be a finite convex function with the finite minimum $f_\star = \min_x f(x) = -f^\star(0)$ and $\xi_k, k = 1, 2, \dots$ are defined by the Algorithm 1 with $\xi_0 < f^\star(0)$. Then*

$$\lim_{k\to\infty} \xi_k = f^\star(0) = -f_\star$$

and

$$f^\star(0) - \xi_{k+1} \le \lambda_k (f^\star(0) - \xi_k)$$

with $\lambda_k \to 0$ when $k \to \infty$.

It means that Algorithm 1 in general case has at least superlinear rate of convergence.

Next we consider the problem (1) with sup-quadratic objective function f where we can claim global convergence of the conceptual conjugate epi-projection algorithm and asymptotic quadratic rate of convergence.

Theorem 2 *Let objective function f in problem (1) is locally sup-quadratic with sup-quadratic characteristic τ and $\xi_k, k = 1, 2, \dots$ are defined by the Algorithm 1 with $\xi_0 < -f_\star$. Then $\lim_{k\to\infty} \xi_k = f^\star(0)$ (Algorithm 1 converges) and for k large enough $f^\star(0) - \xi_{k+1} \le \tau^{-1}(f^\star(0) - \xi_k)^2$ (that is convergence is quadratic).*

Data: The convex function $f : E \to \mathbb{R}$, the epigraph epi $f^\star$, the current iteration number k and the current approximation $\xi_k \le f^\star(0)$.

Result: The next approximation ξ_{k+1} such that $\xi_k \le \xi_{k+1} \le f^\star(0)$

Each iteration consists of two basic operations: **Project** and **Support-Update**

Project. Solve the projection problem of the point $(\xi_k, 0)$ onto epi $f^\star$:

$$\min_{(\xi,g)\in \operatorname{epi} f^\star} \{(\xi - \xi_k)^2 + \|g\|^2\} = (\xi_k^p - \xi_k)^2 + \|g_p^k\|^2 \tag{9}$$

with the corresponding solution $(\xi_k^p, g_p^k) = (f^\star(g_p^k), g_p^k) \in \operatorname{epi} f^\star$. We demonstrate in the analysis of the algorithm convergence that $f^\star(0) \ge \xi_k^p > \xi_k$ if $\xi_k < f^\star(0)$.

Support-Update Compute support function of epi $f^\star$ with the support vector $z^k = -(\xi_k^p - \xi_k, g_p^k) \in \mathbb{R} \times E$

$$(\operatorname{epi} f^\star)_{z^k} = \sup\nolimits_{(\mu,g)\in \operatorname{epi} f^\star} \{-(\xi_k^p - \xi_k)\mu + g_p^k g)\} =$$
$$(\xi_k^p - \xi_k) \sup\nolimits_{(\mu,g)\in \operatorname{epi} f^\star} \{-\mu + \frac{g_p^k}{(\xi_k^p - \xi_k)} g\} = (\xi_k^p - \xi_k) \sup\nolimits_{(\mu,g)\in \operatorname{epi} f^\star} \{-\mu + x_p^k g\} =$$
$$(\xi_k^p - \xi_k)(x_p^k g_p^k - f^\star(g_p^k)\} = (\xi_k^p - \xi_k) f(x_p^k),$$

where $x_p^k = g_p^k / (\xi_k^p - \xi_k)$. Notice that as f is assumed to be a finite function this operation is well-defined.

Finally we update the approximate solution with ξ_{k+1} using the relationship

$$\bar{\xi}_{k+1} z^k = (\operatorname{epi} f^\star)_{z^k}, \text{ where } \bar{\xi}_{k+1} = (\xi_{k+1}, 0) \in \mathbb{R} \times E,$$

which actually amounts to $\xi_{k+1} = -f(x_p^k)$, increment iteration counter $k \to k+1$, etc.

Algorithm 1: The basic iteration of the conceptual conjugate epi-projection algorithm algorithm

Finite convergence of this algorithm for sharp minimum is established by the following theorem.

Theorem 3 *Let the objective function of (1) has a sharp minimum at solution point $x^\star$, all assumptions of the theorem 1 are satisfied and $\xi_k, k = 1, 2, \dots$ are defined by the Algorithm 1 with $\xi_0 < -f_\star$. Then there exists $k^\star$ such that $\xi_{k^\star} = f^\star(0) = -f_\star$.*

Notice that in all cases convergence is global and does not requir any additional assumptions.

2.2 Implementation issues

The critical part in implementation of Algorithm 1 is the projection step (9), where the point $(\xi_k, 0)$ is projected onto epi $f^\star$. The set epi $f^\star$ is given implicitely only, however due to Fenchel-Morou duality we can easily compute the supremum of any linear function $\bar{p}\bar{z}$ on it where $\bar{z} = (\mu, z), \mu \geq f^\star(z)$, $\bar{p} = (\pi, p)$. This supremum is finite when $\pi < 0$ and then

$$(\sup_{\bar{z} \in \text{epi}\, f^\star} \{\pi\mu + pz\} = |\pi| \sup_{z, \mu \geq f^\star(z)} \{pz/|\pi| - \mu\} = |\pi| \sup_z \{pz/|\pi| - f^\star(z)\} = |\pi| f(pz/|\pi|).$$

It gives a chance to suggest simple iteration-like algorithms, using implementable projection onto inner approximation P_k of epi $f^\star$ which is represented on Fig. 2. This algorithm in practice is interrupted

Data: The epigraph epi $f^\star$, its polyhedral approximation, the point $\bar{q} = (\xi, 0) \notin$ epi $f^\star$
Result: The sequence $\{\bar{g}^k = (\xi_k, g^k) \in \text{epi}\, f^\star, k = 1, 2, \dots\}$ such that $\bar{g}^k \to \bar{g}^\star \in \text{epi}\, f^\star$ and $\|\bar{g}^\star - \bar{q}\| = \min_{\bar{g} \in \text{epi}\, f^\star} \|\bar{g} - \bar{q}\|$
Initialize; $P_0 = g^0, k = 0$
While;
Solve quadratic optimization problem:

$$\min_{g \in P_k} \|g - \bar{q}\|^2 = \|g^k - \bar{q}\|^2 \tag{10}$$

Upgrade:

$$P_{k+1} = \text{co}\{P_k, g^k\}, \quad k \to k + 1$$

end while;

Algorithm 2: Iterative algorithm for projection on epi $f^\star$

when desirable accuracy is acheived. The quadratic optimization problem (10) can be solved by many off-the-shelf quadratic solvers, however our experience is that the specialized algorithms like [4] outperforms them. One can find the OCTAVE-version of the code as `DOI: 10.13140/RG.2.2.21281.86882` at [5].

2.3 Numerical example

The most interesting and difficult tests of nonsmooth optimization consist in minimization of piecewise quadratic problems which are constructed as finite maximum of convex quadratics. We demonstrate performance of the implementable version of Algorithm 1 with iterative Algorithm 2 for approximate solution of the auxiliary projection problem (9) on a simple problem (1) with $f(x) = \max_{i=1,2}(x - a^i)A_i(x - a^i)$ with $a^1 = (0, 0, 0), a^2 = (2, 3, 9)$ and A_i are diagonal matrices: $A_1 =$

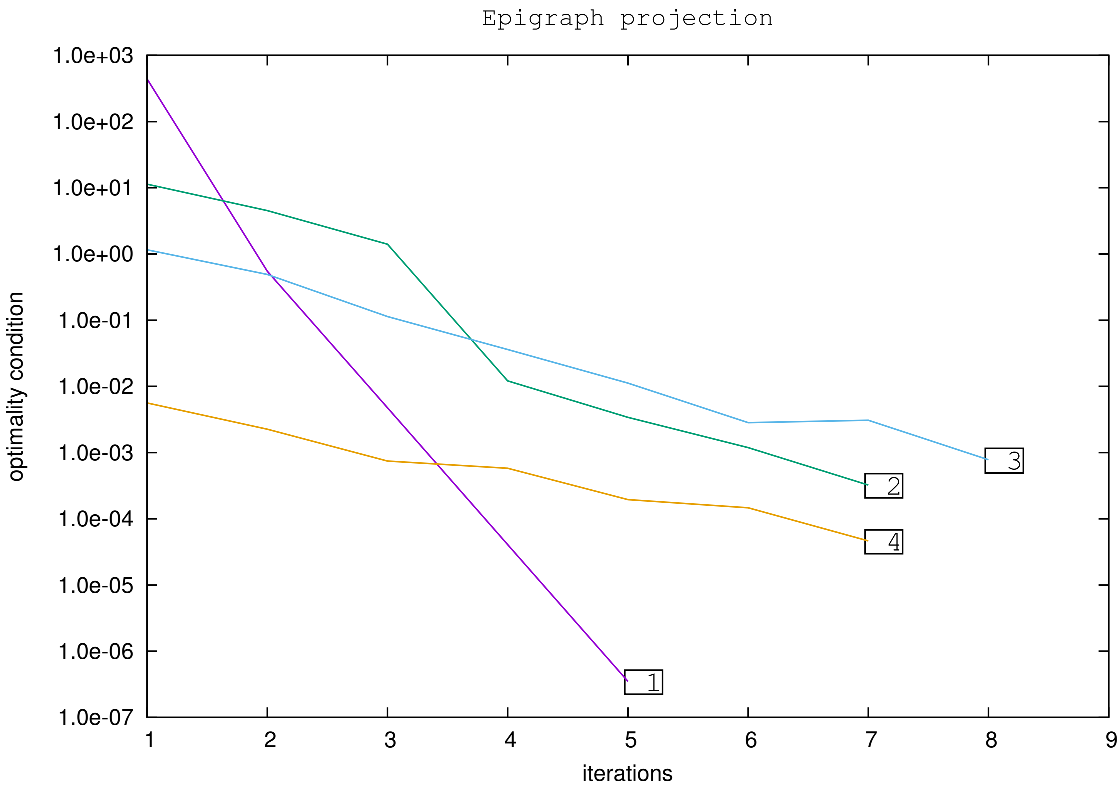

Figure 1: Projection operation on epi $f^{\star}$ on different iterations of Algorithm 1. Boxed numbers on the Figure denotes the major iterations of Algorithm 1

diag$(9,4,1), A_2 =$ diag$(1,4,9)$. The optimization solver CONDOR 1.06, running on NEOS optimization solver [10] reported succesfull complition after 63 function evaluations with the objective value of `0.4348696068`. Our solver attained slightly worse `0.43673` with 27 function evaluations.

The loss in the value of objective function can be probably explained by the numerial instability of projection problems (9) at the final iterations of optimization process. The Figure 1 demonstrates the peculiar features of SU-step during solution of minimization problem. It shows convergence of the simple projection Algorithm 2 in solution of the projection problem (9) in terms of optimality condition $\delta_k = \|z^k\|^2 - \inf_{z \in \text{epi } f^{\star}} z z^k$ where $z^k \in \text{epi } f^{\star}$ — an approximate solution of (2) obtained on k-th iteration of this algorithm. For any k the value of δ_k is non-negative and if $\delta_k = 0$ then z^k is the solution of (2).

It can be seen from the Fig. 1 that in all cases the auxiliary projection problem was solved sufficiently quickly with at least the linear rate of covergence. However, it also can be seen that the projection Algorithm 2 slows down when projected point approaches the epigraph epi $f^{\star}$. This was expected behavior of the algorithm and there are known technics to improve solution of 2 in this case, but this issue requires additional investigation.

3 Conjugate Epi-Projection Algorithm with a Skew Cut

One of the other possible ways to improve computational behavior of the conjugate epi-projection algorithm is to introduce additional condtraints in **Support-Update** (SU) step of this algoritm. Namely, if we assume that there is an additional condition $(\mu, g) \in Q \subset E \times \mathbb{R}$ with $(f^\star(0), 0) \in Q$ then

$$\omega_x = \sup_{(\mu,g)\in \text{epi}\, f^\star \cap Q} \{xg - \mu\} = xg_x - f^\star(g_x) \leq xg_x - xg_x + f(x) = f(x)$$

so ω_x will provide better (lower) upper estimates for $\min_x f(x) = -f^\star(0)$. Of course it will be necessary to ensure that an aditional constraint $(\mu, g) \in Q$ does not cut off the solution $(f^\star(0), 0)$. It implies that $(f^\star(0), 0) \in Q$ which can be ensured in different ways.

The corresponding modification of SU-step is shown as Algorithm 3.

Data: The epigraph epi $f^\star$, the current iteration number k, the current approximation $\xi_k \leq f^\star(0)$, and projection vector z^k obtained from **Project** step.
Result: The next approximation ξ_{k+1} such that $\xi_k \leq \xi_{k+1} \leq f^\star(0)$.
Modified Support-Update
Compute support function of $G_k = \text{epi}\, f^\star \cap Q_k$ with the support vector $z^k = -(\xi^p_k - \xi_k, g^k_p) \in \mathbb{R} \times E$

$$(G_k)_{z^k} = \sup_{(\mu,g)\in G_k} \{-(\xi^p_k - \xi_k)\mu + g^k_p g)\} =$$
$$(\xi^p_k - \xi_k) \sup_{(\mu,g)\in G_k} \{-\mu + \frac{g^k_p}{(\xi^p_k - \xi_k)} g\} = (\xi^p_k - \xi_k) \sup_{(\mu,g)\in G_k} \{-\mu + x^k_p g\} =$$
$$(\xi^p_k - \xi_k)(x^k_p \tilde{g^k_p} - f^\star(g^k_p)\}.$$

where $x^k_p = g^k_p / (\xi^p_k - \xi_k)$.
Notice that now $g^k_p \notin \partial f(x^k_p)$ and we need an additional operation to recover the support vector to epi $f^\star$ at the point $(f^\star(g^k_p), g^k_p)$.

Algorithm 3: Modified **Support-Update** (MSU) step

Convergence of the Algorithm 3 is confirmed by the following theorem.

Theorem 4 *Let f be a finite convex function with the finite minimum $f_\star = \min_x f(x) = -f^\star(0)$ and $\xi_k, k = 1, 2, \ldots$ are defined by the Algorithm 3 with $\xi_0 < f^\star(0)$. Then $\lim_{k\to\infty} \xi_k = f^\star(0) = -f_\star$, that is the algorithm converges;*

Proof. Assume that on k-th iteration we have $\xi_k < f^\star(0)$ as the approximation of $f^\star(0)$. According to Algorithm 3 to construct the next ($k + 1$-th) approximation ξ_{k+1} the point $(\xi_k, 0) \in \mathbb{R} \times E$ is to be projected onto epi $f^\star \cap Q_k$ first:

$$\min_{(\xi,g)\in \text{epi}\, f^\star \cap Q_k} \{(\xi - \xi_k)^2 + \|g\|^2\} = (\xi^p_k - \xi_k)^2 + \|g^k_p\|^2 \tag{11}$$

The solution $(\xi^p_k, g^k_p) = (f^\star(g^k_p), g^k_p) \in \text{epi}\, f^\star$ of this problem satisfies optimality conditions

$$(f^\star(g^k_p) - \xi_k)(\xi - \xi^p_k) + g^k_p(g - g^k_p) \geq 0 \tag{12}$$

for any $(\xi, g) \in \text{epi}\, f^\star \cap Q_k$.

It is easy to see that $\xi^p_k > \xi_k$. Indeed the opposite strict inequality $\xi^p_k < \xi_k$ contradicts the optimality of (ξ^p_k, g^k_p) as in this case

$$(\xi_k, g^k_p) = (\xi^p_k + (\xi_k - \xi^p_k), g^k_p) \in \text{epi}\, f^\star \cap Q_k \subset \text{epi}\, f^\star,$$

and

$$(\xi_k - \xi_k)^2 + \|g^k_p\|^2 < (\xi_k - \xi^p_k)^2 + \|g^k_p\|^2 = \min_{(\xi,g)\in \text{epi}\, f^\star} \{(\xi_k - \xi)^2 + \|g\|^2\}.$$

If $\xi_k^p = \xi_k$ then $\mathbb{R} \times \{0\}$ is strictly separable from epi $f^\star$:

$$\xi(\xi_k - \xi_k^p) + 0g_p^k = 0 < \|g_p^k\|^2 \leq \mu(\xi_k - \xi_k^p) + gg_p^k$$

for any $(\mu, g) \in \text{epi}\, f^\star$ as it follows from projection conditions. Hence $0 \notin \text{dom}(f^\star)$ which contradicts the assumptions of the theorem.

According to Algorithm 3 the next approximation ξ_{k+1} is determined from the equality

$$(\xi_k^p - \xi_k)(\xi_{k+1} - \xi_k)) - \|g_p^k\|^2 = (\xi_k^p) - \xi_k)^2 + \|g_p^k\|^2$$

which gives the following expression for ξ_{k+1}:

$$\xi_{k+1} = \xi_k + \|g_p^k\|^2/(\xi_k^p - \xi_k) \geq \xi_k,$$

and $\xi_{k+1} = \xi_k$ if and only if $g_p^k = 0$ which means that we already obtained the solution.

Repeating this operation we obtain the monotone sequence $\xi_k, k = 0, 1, \ldots$ such that

$$\xi_k \leq \xi_{k+1} \leq f^\star(0), k = 0, 1, \ldots$$

where inequalities turn into equalities only if either $\xi_k = f^\star(0)$ or $\xi_{k+1} = f^\star(0)$ which of course makes no difference. Under these conditions $\lim_{k\to\infty} \xi_k = f^\star(0)$ which proves the convergence of the algorithm 1. ■

3.1 Projection in Modified Support-Update Step

The key step in MSU step of Algorithm 3 is the computation of projection of a given point, say z, on $G_k = \text{epi}\, f^\star \cap Q_k$ where Q_k is a cutting set. It can be approximately solved by the iterative Algorithm 2 which in turn requiers computing of the support function $(G_k)_{z^k}$ of the set $G_k = \text{epi}\, f^\star \cap Q_k$ with the given support vector $z^k \in \mathbb{R} \times E$. By dropping for simplicity the iteration index k we face the following problem

$$(G)_z = \sup_{\substack{g \in \text{epi}\, f^\star \\ g \in Q}} zg = zg^\star$$

the computational difficulty of which critically depends on cutting set Q. To begin with something constructive we consider here the simplest choice of $Q = H_{p,\beta}$ where $H_{p,\beta}$ is the half-space, described by linear inequality $H_{p,\beta} = \{(\mu, g) : pg - \mu \geq \beta\}$, where $p \in E$ and $\beta \geq f^\star(0)$ to guarantee that $(f^\star(0), 0) \in H_{p,\beta}$. Such β is easy to obtain from the inner approximation D of epi $f^\star$ if available. If the vertical line $\mathbb{R} \times \{0\}$ intersects D at some point $(-\beta, 0)$. Then $-\beta \geq f^\star(0)$ and therefore $(f^\star(0), 0) \in H_{p,\beta}$.

For the choice of vector p we have almost unlimited freedom and choice of the best p might be an interesting subject for further consideration.

Then

$$\sup_{\substack{\mu \geq f^\star(g) \\ pg - \mu \geq \beta}} \{xg - \mu\} = \inf_{\theta \geq 0} \sup_{\mu \geq f^\star(g)} \{xg - \mu + \theta(pg - \mu - \beta)\} =$$

$$\inf_{\theta \geq 0} \sup_{\mu \geq f^*(g)} \{g(x + \theta p) - \mu(1 + \theta)\} - \beta\theta =$$

$$\inf_{\theta \geq 0} \{-\beta\theta + (1 + \theta) \sup_{\mu \geq f^*(g)} \{g\frac{x + \theta p}{1 + \theta} - \mu\}\} =$$

$$\inf_{\theta \geq 0} \{-\beta\theta + (1 + \theta)f(\frac{x + \theta p}{1 + \theta})\} = \inf_{\theta \geq 0} \{-\beta\theta + (1 + \theta)f(x + \frac{\theta}{1 + \theta}(p - x))\}.$$

By introduction of new variable $\gamma = \dfrac{\theta}{1+\theta}$ the last expression can be transformed in

$$\inf_{\gamma \in [0,1)} \{-\frac{\gamma}{1-\gamma}\beta + \frac{1}{1-\gamma} f(x+\gamma(p-x))\} = \inf_{\gamma \in [0,1)} (1-\gamma)^{-1}\{f(x_\gamma) - \gamma\beta\} = \psi(\gamma),$$

where $x_\gamma = x + \gamma(p-x)$ and so the support problem is reduced to one-dimensional minimization.

Conclusion

The conceptual version of the dual epi-projection algorithm has promising computational properties which makes it a viable candidate for developing implementable versions. First of all it guarantees global super-linear convergence to the optimum for any solvable convex optimization problem. Second, it provides quadratic convergence and even finite termination without any changes in the algorithm for quite common types of convex optimization problems: sup-quadratic, which strictly contain strongly convex, and convex optimization problems with sharp minimum. It is worth to notice that the algorithm is absolutely parameter-free, use the first-order subgradient oracle only, and requires no specific knowledge of any specific characteristics of convex optimization problem, like Lipshitz constants, strong convexity parameter or close enough starting point.

The implementation perspectives for the algorithm depend upon the possibility to produce practical version of the projection operator on $\operatorname{epi} f^\star$. From the theoretical point of view it is easy to derive accuracy estimates for its termination so it can be finitely solved for any required accuracy. It can be used to preserve the overall rates of convergence in terms of Algorithm 1 iterations, however the resulting computational complexity requires further investigations.

References

[1] Nurminski, E.A.: A Conceptual Conjugate Epi-Projection Algorithm of Convex Optimization: Superlinear, Quadratic and Finite Convergence. Optim Lett(2019) 13:23-34 ISSN 1862-4472 DOI 10.1007/s11590-018-1269-3

[2] Vorontsova, E.A., Nurminski, E.A.: Synthesis of Cutting and Separating Planes in a Nonsmooth Optimization Method. Cybernetics and Systems Analysis, 51(4), 619–631 (2015)

[3] Nurminski, E.A.: Multiple cuts in separating planes algorithms. In: Kochetov, Y., Khachay, M., Beresnev, V., Nurminski, E., Pardalos, P. (eds) Discrete Optimization and Operations Research. 9th International Conference, DOOR 2016 Vladivostok, Russia, September 19-23, 2016. Proceedings. Lecture Notes in Computer Science, vol. 9869 pp. 430-440. Springer, Heidelberg (2016) DOI; 10.1007/978-3-319-44914-2.34

[4] Nurminski, E.A.: Convergence of the Suitable Affine Subspace Method for Finding the Least Distance to a Simplex. Computational Mathematics and Mathematical Physics, 45(11), 1915–1922 (2005)

[5] Nurminski, E.A.: Orthogonal projection on convex hull of a finite set of points of a finite-dimensional euclidean space. Version 1.6, updates 1.5. DOI: 10.13140/RG.2.2.21281.86882

[6] Hirriart-Uruty, J.-B.; Lemarechal, C.: Convex Analysis and Minimization Algorithms II Advanced Theory and Bundle Methods. A Series of Comprehensive Studies in Mathematics, 306. Springer-Verlag Berlin Heidelberg (1993)

[7] Ferris, M.C.: Weak Sharp Minima and Penalty Functions in Mathematical Programming. PhD Dissertation, University of Cambridge, Cambridge, UK, (1988)

[8] Ferris, M.C.: Finite termination of the proximal point algorithm. Math. Program, vol. 50, 359-366 (1991)

[9] Zhou, J.; Wang, C. New characterizations of weak sharp minima Optim Lett 6: 1773. doi:10.1007/s11590-011-0369-0 (2012)

[10] NEOS Server: State-of-the-Art Solvers for Numerical Optimization. `https://neos-server.org/neos/`