Excessive gap technique in non-smooth convex minimization

Yu. Nesterov *

First version: January 30, 2003
Final version: April 23, 2003

Abstract

In this paper we introduce a new primal-dual technique for convergence analysis of gradient schemes for non-smooth convex optimization. As an example of its application, we derive a primal-dual gradient method for a special class of structured non-smooth optimization problems, which ensures a rate of convergence of the order $O(\frac{1}{k})$, where k is the iteration count. Another example is a gradient scheme which minimizes a non-smooth strongly convex function with known structure with the rate of convergence $O(\frac{1}{k^2})$. In both cases the efficiency of the methods is higher than the corresponding black-box lower complexity bounds by an order of magnitude.

Keywords: Convex optimization, non-smooth optimization, complexity theory, black-box oracle, optimal methods, structural optimization.

*CORE, Catholic University of Louvain, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be

1 Introduction

Motivation. This paper continues the research started in [3], where it was shown that some structured non-smooth optimization problems can be solved with efficiency estimates $O(\frac{1}{\epsilon})$, where ϵ is the desired accuracy of the solution. This complexity is much better than the theoretical lower complexity bound $O(\frac{1}{\epsilon^2})$ (see [2]). This improvement, of course, is possible because of certain relaxation of the standard black box assumption. Instead, it was assumed that our problem has an explicit and quite simple minimax structure. The numerical scheme proposed in [3] has a drawback, which decreases its practical efficiency. In this scheme the number of steps must be fixed in advance. It is chosen in accordance to a worst-case complexity analysis.

In this paper we propose several new primal-dual gradient schemes for the same class of problems as in [3]. However, our schemes now are free from the above deficiency. They are derived from a new primal-dual symmetric technique for convergence analysis, which we call the *excessive gap condition*.

The paper is organized as follows. In Section 2 we introduce our *model* of optimization problem and recall several useful facts from [3]. In Section 3 we describe the excessive gap condition. In the next two sections we present two different strategies for maintaining the condition during the optimization process. In Section 6 we give the convergence result of the order $O(\frac{1}{k})$, where k is the iteration counter. This convergence result is valid for all non-smooth functions described by our model. However, if we assume more, (namely, the strong convexity of the primal objective), then the convergence can be improved up to $O(\frac{1}{k^2})$. This improvement is presented in the last Section 7. Note that both complexity results improve the corresponding general lower complexity bound by an order of magnitude.

Notation. In what follows we work with different primal and dual spaces equipped by corresponding norms. For sake of notation, we apply the following convention. The (primal) finite-dimensional real vector space is always denoted by E, possibly with an index. This space is endowed with a norm $\|\cdot\|$, which has the same index as the corresponding space. The space of linear functions on E is denoted by E^*. For $s \in E^*$ and $x \in E$ we denote $\langle s, x\rangle$ the value of s at x. The *scalar product* $\langle\cdot,\cdot\rangle$ is marked by the same index as E. The norm for the dual space is defined in the standard way:

$$\|s\|^* = \max_x\{\langle s, x\rangle : \ \|x\| = 1\}.$$

For an operator $A : E_1 \to E_2^*$ we define *adjoint* operator $A^* : E_2 \to E_1^*$ in the following way:

$$\langle Ax, u\rangle_2 = \langle A^*u, x\rangle_1 \quad \forall x \in E_1, \ u \in E_2.$$

The *norm* of such operator is defined as follows:

$$\|A\|_{1,2} = \max_{x,u}\{\langle Ax, u\rangle_2 : \ \|x\|_1 = 1, \ \|u\|_2 = 1\}.$$

Clearly,

$$\|A\|_{1,2} = \|A^*\|_{2,1} = \max_x\{\|Ax\|_2^* : \ \|x\|_1 = 1\} = \max_u\{\|A^*u\|_1^* : \ \|u\|_2 = 1\}.$$

Hence, for any $u \in E_2$ we have

$$\|A^* u\|_1^* \leq \|A\|_{1,2} \cdot \|u\|_2. \tag{1.1}$$

Acknowledgement. The author is very thankful to F.Glineur for his constructive comments on the text.

2 Model of the problem

In this paper we are interested in the following minimization problem:

$$\text{Find } f^* = \min_x \{ f(x) : \; x \in Q_1 \}, \tag{2.1}$$

where Q_1 is a bounded closed convex set in a finite-dimensional real vector space E_1 and $f(x)$ is a continuous convex function on Q_1. We do not assume f to be differentiable.

Very often, the *structure* of the objective function in (2.1) is known. Let us assume that this structure can be described by the following *model* (see [3] for different examples):

$$f(x) = \hat{f}(x) + \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) : \; u \in Q_2 \}, \tag{2.2}$$

where function $\hat{f}(x)$ is continuous and convex on Q_1, Q_2 is a closed convex bounded set in a finite-dimensional real vector space E_2, $\hat{\phi}(u)$ is a continuous convex function on Q_2 and the linear operator A maps E_1 to E_2^*. In this case the problem (2.1) can be written in an *adjoint* form:

$$\begin{aligned} &\max_u \{ \phi(u) : \; u \in Q_2 \}, \\ &\phi(u) = -\hat{\phi}(u) + \min_x \{ \langle Ax, u \rangle_2 + \hat{f}(x) : \; x \in Q_1 \}. \end{aligned} \tag{2.3}$$

We assume that this representation is completely similar to (2.1) in the following sense. The methods described in this paper are implementable only if the optimization problems involved in the definitions of functions $f(x)$ and $\phi(u)$ can be solved in a closed form. So, we assume that the structure of the objects $\hat{f}$, $\hat{\phi}$, Q_1 and Q_2 is simple enough. We also assume that the functions $\hat{f}$ and $\hat{\phi}$ have Lipschitz-continuous gradients with Lipschitz constants $L_1(\hat{f})$ and $L_2(\hat{\phi})$ respectively. In some situations we allow these constants to be infinitely big.

Let us show that the knowledge of structure (2.2) can help in solving problems (2.1) and (2.3). Same as in [3], we are going to use this structure for constructing a smooth approximation of the objective functions.

Consider a *prox-function* $d_2(u)$ of the set Q_2. We assume that $d_2(u)$ is continuous and strongly convex on Q_2 with the convexity parameter σ_2. Denote by

$$u_0 = \arg\min_u \{ d_2(u) : \; u \in Q_2 \}$$

the *prox-center* of the function $d_2(\cdot)$. Without loss of generality we assume that $d_2(u_0) = 0$. Thus, for any $u \in Q_2$ we have

$$d_2(u) \geq \tfrac{1}{2} \sigma_2 \|u - u_0\|_2^2. \tag{2.4}$$

Let μ_2 be a positive *smoothness* parameter. Consider the following function:

$$f_{\mu_2}(x) = \hat{f}(x) + \max_u \{\langle Ax, u\rangle_2 - \hat{\phi}(u) - \mu_2 d_2(u) : \ u \in Q_2\}. \tag{2.5}$$

Denote by $u_{\mu_2}(x)$ the optimal solution of above problem. Since function $d_2(u)$ is strongly convex, this solution is unique. In accordance to Theorem 1 ([3]), the gradient

$$\nabla f_{\mu_2}(x) = \nabla \hat{f}(x) + A^* u_{\mu_2}(x)$$

is Lipschitz-continuous with the constant

$$L_1(f_{\mu_2}) = L_1(\hat{f}) + \frac{1}{\sigma_2 \mu_2}\|A\|_{1,2}^2.$$

Similarly, let us consider a prox-function $d_1(x)$ of the set Q_1, which has convexity parameter σ_1 and the prox-center x_0 with $d_1(x_0) = 0$. Thus, for any $x \in Q_1$ we have

$$d_1(x) \geq \tfrac{1}{2}\sigma_1 \|x - x_0\|_1^2. \tag{2.6}$$

Let μ_1 be a positive smoothness parameter. Consider

$$\phi_{\mu_1}(u) = -\hat{\phi}(u) + \min_x \{\langle Ax, u\rangle_2 + \hat{f}(x) + \mu_1 d_1(x) : \ x \in Q_1\}. \tag{2.7}$$

Clearly, since the second term in the above definition is a minimum of linear functions, $\phi_{\mu_1}(u)$ is concave. Denote by $x_{\mu_1}(u)$ the unique optimal solution of the above problem. In accordance to Theorem 1 ([3]), the gradient

$$\nabla \phi_{\mu_1}(u) = -\nabla \hat{\phi}(u) + A x_{\mu_1}(u)$$

is Lipschitz-continuous with the constant

$$L_2(\phi_{\mu_1}) = L_2(\hat{\phi}) + \frac{1}{\sigma_1 \mu_1}\|A\|_{1,2}^2.$$

3 Excessive gap condition

Note that for any $x \in Q_1$ and $u \in Q_2$ we have

$$\phi(u) \leq f(x)$$

However, $f_{\mu_2}(x) \leq f(x)$ and $\phi(u) \leq \phi_{\mu_1}(u)$. That opens a possibility to satisfy the following *excessive gap condition*:

$$\boxed{f_{\mu_2}(\bar{x}) \ \leq \ \phi_{\mu_1}(\bar{u})} \tag{3.1}$$

by certain $\bar{x} \in Q_1$ and $\bar{u} \in Q_2$. This inequality is the main relation we are going to maintain recursively in our optimization schemes. Let us justify this strategy by convergence results.

Denote $D_1 = \max_x \{d_1(x) : \ x \in Q_1\}$ and $D_2 = \max_u \{d_2(u) : \ u \in Q_2\}$. As we will see later, the size of these bounds significantly affects our complexity bounds.

Lemma 1 *Let $\bar{x}$ from Q_1 and $\bar{u}$ from Q_2 satisfy (3.1). Then*

$$f(\bar{x}) - \phi(\bar{u}) \le \mu_1 D_1 + \mu_2 D_2. \tag{3.2}$$

Proof:
Indeed, for any $x \in Q_1$, $u \in Q_2$ we have $f_{\mu_2}(x) \ge f(x) - \mu_2 D_2$ and $\phi_{\mu_1}(u) \le \phi(u) + \mu_1 D_1$. □

Thus, our goal is to maintain inequality (3.1) as μ_1 and μ_2 go to zero. This can be done in two different ways, which correspond to two different auxiliary problems we must be ready to solve at each iteration.

Before we start our analysis, let us prove one useful inequality.

Lemma 2 *For any x and $\bar{y}$ from Q_1 we have:*

$$\hat{f}(x) + \langle Ax, u_{\mu_2}(\bar{y})\rangle_2 - \hat{\phi}(u_{\mu_2}(\bar{y})) \ge f_{\mu_2}(\bar{y}) + \langle \nabla f_{\mu_2}(\bar{y}), x - \bar{y}\rangle_1. \tag{3.3}$$

Proof:
Let us take arbitrary x and $\bar{y}$ from Q_1. Denote $\bar{u} = u_{\mu_2}(\bar{y})$. Then

$$\begin{aligned} & f_{\mu_2}(\bar{y}) + \langle \nabla f_{\mu_2}(\bar{y}), x - \bar{y}\rangle_1 \\ = \; & \hat{f}(\bar{y}) + \langle A\bar{y}, \bar{u}\rangle_2 - \hat{\phi}(\bar{u}) - \mu_2 d_2(\bar{u}) + \langle \nabla \hat{f}(\bar{y}) + A^*\bar{u}, x - \bar{y}\rangle_1 \\ \le \; & \hat{f}(x) + \langle Ax, \bar{u}\rangle_2 - \hat{\phi}(\bar{u}). \end{aligned}$$

□

4 Gradient mapping

Let us justify first a possibility to launch the process. Define the *primal gradient mapping*:

$$T_{\mu_2}(x) = \arg\min_y \left\{ \langle \nabla f_{\mu_2}(x), y - x\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \|y - x\|_1^2 : \; y \in Q_1 \right\}, \quad x \in Q_1.$$

Lemma 3 *Let us choose an arbitrary $\mu_2 > 0$. For prox-center x_0 define*

$$\bar{x} = T_{\mu_2}(x_0), \quad \bar{u} = u_{\mu_2}(x_0). \tag{4.1}$$

Then the excessive gap condition (3.1) is satisfied for any $\mu_1 \ge \frac{1}{\sigma_1} L_1(f_{\mu_2})$.

Proof:
Indeed, in view of inequality (3.3), for $\mu_1 = \frac{1}{\sigma_1} L_1(f_{\mu_2})$ we get the following relations:

$$\begin{aligned} f_{\mu_2}(T_{\mu_2}(x_0)) & \le \min_x \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \|x - x_0\|_1^2 : \; x \in Q_1 \right\} \\ & \le \min_x \left\{ \hat{f}(x) + \langle Ax, u_{\mu_2}(x_0)\rangle_2 - \hat{\phi}(u_{\mu_2}(x_0)) + \tfrac{1}{2}\mu_1\sigma_1 \|x - x_0\|_1^2 : x \in Q_1 \right\} \\ & \le -\hat{\phi}(u_{\mu_2}(x_0)) + \min_x \left\{ \hat{f}(x) + \langle Ax, u_{\mu_2}(x_0)\rangle_2 + \mu_1 d_1(x) : x \in Q_1 \right\} \\ & \equiv \phi_{\mu_1}(u_{\mu_2}(x_0)). \end{aligned}$$

□

Thus, the condition (3.1) can be satisfied at the first step of the algorithm. Let us show how we can update the points $\bar{x}$ and $\bar{u}$ in order to get (3.1) for smaller values of μ_1 and μ_2. Note that in view of absolute symmetry of the situation, at the first stage of the algorithm we can try to decrease only μ_1 keeping μ_2 unchanged. After that, at the second stage, we update μ_2 and keep μ_1. The main advantage of such a strategy is that we need to find a justification only for the first stage. The proof for the second one will be absolutely symmetric.

Theorem 1 *Let points $\bar{x} \in Q_1$ and $\bar{u} \in Q_2$ satisfy the excessive gap condition (3.1) for some positive μ_1 and μ_2. Let us fix $\tau \in (0,1)$ and choose $\mu_1^+ = (1-\tau)\mu_1$,*

$$\begin{aligned} \hat{x} &= (1-\tau)\bar{x} + \tau x_{\mu_1}(\bar{u}), \\ \bar{u}_+ &= (1-\tau)\bar{u} + \tau u_{\mu_2}(\hat{x}), \\ \bar{x}_+ &= T_{\mu_2}(\hat{x}). \end{aligned} \tag{4.2}$$

Then the pair $(\bar{x}_+, \bar{u}_+)$ satisfies condition (3.1) with smoothness parameters μ_1^+ and μ_2 provided that τ is chosen in accordance to the following relation:

$$\frac{\tau^2}{1-\tau} \leq \frac{\mu_1 \sigma_1}{L_1(f_{\mu_2})}. \tag{4.3}$$

Proof:
Denote $\hat{u} = u_{\mu_2}(\hat{x})$. In view of rules (4.2) and inequality (3.3), we have:

$$\begin{aligned} \phi_{\mu_1^+}(\bar{u}_+) &= \min_x \left\{ (1-\tau)\mu_1 d_1(x) + \langle Ax, (1-\tau)\bar{u} + \tau\hat{u}\rangle_2 + \hat{f}(x) : \ x \in Q_1 \right\} - \hat{\phi}(\bar{u}_+) \\ &\geq \min_x \left\{ (1-\tau) \left[\mu_1 d_1(x) + \langle Ax, \bar{u}\rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u}) \right] \right. \\ &\qquad \left. + \tau [\hat{f}(x) + \langle Ax, \hat{u}\rangle_2 - \hat{\phi}(\hat{u})] : \ x \in Q_1 \right\} \\ &\geq \min_x \left\{ (1-\tau) \left[\phi_{\mu_1}(\bar{u}) + \tfrac{1}{2}\mu_1\sigma_1 \|x - x_{\mu_1}(\bar{u})\|_1^2 \right] \right. \\ &\qquad \left. + \tau [f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), x - \hat{x}\rangle_1] : \ x \in Q_1 \right\}. \end{aligned}$$

Note that

$$\phi_{\mu_1}(\bar{u}) \geq f_{\mu_2}(\bar{x}) \geq f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - \hat{x}\rangle_1 = f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - x_{\mu_1}(\bar{u})\rangle_1.$$

Hence, using the condition (4.3), we accomplish the proof as follows:

$$\begin{aligned}
& \phi_{\mu_1^+}(\bar{u}_+) \\
\geq \;& \min_x \left\{ f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), x - x_{\mu_1}(\bar{u}) \rangle_1 + \tfrac{1}{2}(1-\tau)\mu_1 \sigma_1 \| x - x_{\mu_1}(\bar{u}) \|_1^2 : \; x \in Q_1 \right\} \\
\geq \;& \min_x \left\{ f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), x - x_{\mu_1}(\bar{u}) \rangle_1 + \tfrac{1}{2}\tau^2 L_1(f_{\mu_2}) \| x - x_{\mu_1}(\bar{u}) \|_1^2 : \; x \in Q_1 \right\} \\
\geq \;& \min_x \left\{ f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), x - \hat{x} \rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2}) \| x - \hat{x} \|_1^2 : \; x \in Q_1 \right\} \\
\geq \;& f_{\mu_2}(\bar{x}_+).
\end{aligned}$$

□

5 Bregman projection

Let us assume for simplicity that $d_1(x)$ is differentiable. Then for any $x \in Q_1$ we have

$$\langle \nabla d_1(x_0), x - x_0 \rangle_1 \geq 0. \tag{5.1}$$

For x and z from Q_1 denote by

$$\xi_1(z, x) = d_1(x) - d_1(z) - \langle \nabla d_1(z), x - z \rangle_1$$

the *Bregman distance* between z and x. If z is fixed, then $\xi(z, x)$ is strongly convex in x. Moreover,

$$\xi_1(z, x) \geq \tfrac{1}{2}\sigma_1 \| x - z \|_1^2. \tag{5.2}$$

Define the Bregman projection of some $g \in E_1^*$ onto the set Q_1 as follows:

$$V_1(z, g) = \arg\min_x \{ \langle g, x - z \rangle_1 + \xi_1(z, x) : \; x \in Q_1 \}. \tag{5.3}$$

As compared with the gradient mapping, the main advantage of the Bregman projection is that the optimization problem in (5.3) involves the same objects as (2.7). So, we have more chances to have it solvable in an easy way (see Section 5.3 in [3] for an example).

Let us show first how we can find the starting points of the process.

Lemma 4 *Let us choose an arbitrary $\mu_2 > 0$. Denote $\gamma = \frac{\sigma_1}{L_1(f_{\mu_2})}$ and set*

$$\bar{x} = V_1(x_0, \gamma \nabla f_{\mu_2}(x_0)), \quad \bar{u} = u_{\mu_2}(x_0). \tag{5.4}$$

Then the excessive gap condition is satisfied for any $\mu_1 \geq \frac{1}{\gamma}$.

Proof:

Indeed,

$$\begin{aligned}
f_{\mu_2}(\bar x) &\le f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), \bar x - x_0\rangle_1 + \tfrac{1}{2} L_1(f_{\mu_2})\|\bar x - x_0\|_1^2 \\
&= f_{\mu_2}(x_0) + \tfrac{1}{\gamma}\left[\gamma\langle \nabla f_{\mu_2}(x_0), \bar x - x_0\rangle_1 + \tfrac{1}{2}\sigma_1\|\bar x - x_0\|_1^2\right] \\
\text{(by (5.2))} \quad &\le f_{\mu_2}(x_0) + \tfrac{1}{\gamma}\left[\langle \gamma\nabla f_{\mu_2}(x_0), \bar x - x_0\rangle_1 + \xi_1(x_0, \bar x)\right] \\
&= \min_x \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0\rangle_1 + \tfrac{1}{\gamma}\xi_1(x_0, x) :\ x \in Q_1\right\} \\
\text{(by (5.1))} \quad &\le \min_x \left\{ f_{\mu_2}(x_0) + \langle \nabla f_{\mu_2}(x_0), x - x_0\rangle_1 + \tfrac{1}{\gamma} d_1(x) :\ x \in Q_1\right\} \\
\text{(using (3.3))} \quad &\le \min_x \left\{ \hat f(x) + \langle Ax, u_{\mu_2}(x_0) - \hat\phi(u_{\mu_2}(x_0)) + \tfrac{1}{\gamma} d_1(x) :\ x \in Q_1\right\} \\
&= \phi_{\frac{1}{\gamma}}(u_{\mu_2}(x_0)) \le \phi_{\mu_1}(u_{\mu_2}(x_0)).
\end{aligned}$$

□

Let us present now one iteration of this variant of the algorithm.

Theorem 2 *Let points $\bar x \in Q_1$ and $\bar u \in Q_2$ satisfy the excessive gap condition (3.1) for some positive μ_1 and μ_2. Let us choose $\tau \in (0,1)$ in accordance with (4.3) and set*

$$\begin{aligned}
\hat x &= (1-\tau)\bar x + \tau x_{\mu_1}(\bar u), \\
\bar u_+ &= (1-\tau)\bar u + \tau u_{\mu_2}(\hat x), \\
\tilde x &= V_1(x_{\mu_1}(\bar u), \tfrac{\tau}{(1-\tau)\mu_1}\nabla f_{\mu_2}(\hat x)), \\
\bar x_+ &= (1-\tau)\bar x + \tau \tilde x, \\
\mu_1^+ &= (1-\tau)\mu_1.
\end{aligned} \tag{5.5}$$

Then the pair $(\bar x_+, \bar u_+)$ satisfies condition (3.1) with the smoothness parameters μ_1^+ and μ_2.

Proof:
Denote $\hat u = u_{\mu_2}(\hat x)$. In view of the rules (5.5) and inequality (3.3), we have:

$$(1-\tau)\mu_1 d_1(x) + \langle Ax, (1-\tau)\bar u + \tau\hat u\rangle_2 + \hat f(x) - \hat\phi(\bar u_+)$$

$$\ge (1-\tau)\left[\mu_1 d_1(x) + \langle Ax, \bar u\rangle_2 + \hat f(x) - \hat\phi(\bar u)\right] + \tau[\hat f(x) + \langle Ax, \hat u\rangle_2 - \hat\phi(\hat u)]$$

$$\ge (1-\tau)\left[\mu_1 d_1(x) + \langle Ax, \bar u\rangle_2 + \hat f(x) - \hat\phi(\bar u)\right]_1 + \tau[f_{\mu_2}(\hat x) + \langle \nabla f_{\mu_2}(\hat x), x - \hat x\rangle_1]_2.$$

Note that for any $x \in Q_1$ we have:

$$\langle \mu_1 \nabla d_1(x_{\mu_1}(\bar{u})) + A^*\bar{u} + \nabla \hat{f}(x_{\mu_1}(\bar{u})), x - x_{\mu_1}(\bar{u}) \rangle_1 \geq 0.$$

Therefore we can estimate the term $[\cdot]_1$ as follows:

$$\begin{aligned}
[\cdot]_1 = \;& \mu_1 \left(\xi(x_{\mu_1}(\bar{u}), x) + d_1(x_{\mu_1}(\bar{u})) + \langle \nabla d_1(x_{\mu_1}(\bar{u})), x - x_{\mu_1}(\bar{u}) \rangle_1 \right) \\
& + \langle Ax, \bar{u} \rangle_2 + \hat{f}(x) - \hat{\phi}(\bar{u}) \\
\geq \;& \mu_1 \xi(x_{\mu_1}(\bar{u}), x) + \mu_1 d_1(x_{\mu_1}(\bar{u})) + \langle A x_{\mu_1}(\bar{u}), \bar{u} \rangle_2 \\
& + \hat{f}(x) - \langle \nabla \hat{f}(x_{\mu_1}(\bar{u})), x - x_{\mu_1}(\bar{u}) \rangle_1 - \hat{\phi}(\bar{u}) \\
\geq \;& \mu_1 \xi(x_{\mu_1}(\bar{u}), x) + \mu_1 d_1(x_{\mu_1}(\bar{u})) + \langle A x_{\mu_1}(\bar{u}), \bar{u} \rangle_2 + \hat{f}(x_{\mu_1}(\bar{u})) - \hat{\phi}(\bar{u}) \\
= \;& \mu_1 \xi(x_{\mu_1}(\bar{u}), x) + \phi_{\mu_1}(\bar{u}) \geq \mu_1 \xi(x_{\mu_1}(\bar{u}), x) + f_{\mu_2}(\bar{x}) \\
\geq \;& \mu_1 \xi(x_{\mu_1}(\bar{u}), x) + f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x} - \hat{x} \rangle_1.
\end{aligned}$$

Thus, using rules (5.5) and the relation (4.3), we can continue:

$$\begin{aligned}
\phi_{\mu_1^+}(\bar{u}_+) \;&=\; \min_{x \in Q_1} \left\{ (1-\tau)\mu_1 d_1(x) + \langle Ax, (1-\tau)\bar{u} + \tau \hat{u} \rangle_2 + \hat{f}(x) \right\} - \hat{\phi}(\bar{u}_+) \\
&\geq\; \min_{x \in Q_1} \{ (1-\tau)\mu_1 \xi(x_{\mu_1}(\bar{u}), x) + f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), (1-\tau)\bar{x} + \tau x - \hat{x} \rangle_1 \} \\
&=\; \min_{x \in Q_1} \{ (1-\tau)\mu_1 \xi(x_{\mu_1}(\bar{u}), x) + f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), x - x_{\mu_1}(\bar{u}) \rangle_1 \} \\
&=\; (1-\tau)\mu_1 \xi(x_{\mu_1}(\bar{u}), \tilde{x}) + f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), \tilde{x} - x_{\mu_1}(\bar{u}) \rangle_1 \\
&\geq\; \tfrac{1}{2}(1-\tau)\mu_1 \sigma_1 \| \tilde{x} - x_{\mu_1}(\bar{u}) \|_1^2 + f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), \tilde{x} - x_{\mu_1}(\bar{u}) \rangle_1 \\
&\geq\; \tfrac{1}{2}\tau^2 L_1(f_{\mu_2}) \| \tilde{x} - x_{\mu_1}(\bar{u}) \|_1^2 + f_{\mu_2}(\hat{x}) + \tau \langle \nabla f_{\mu_2}(\hat{x}), \tilde{x} - x_{\mu_1}(\bar{u}) \rangle_1 \\
&=\; \tfrac{1}{2} L_1(f_{\mu_2}) \| \bar{x}_+ - \hat{x} \|_1^2 + f_{\mu_2}(\hat{x}) + \langle \nabla f_{\mu_2}(\hat{x}), \bar{x}_+ - \hat{x} \rangle_1 \geq f_{\mu_2}(\bar{x}_+).
\end{aligned}$$

□

6 Convergence analysis

In Sections 4, 5 we have seen that the smoothness parameters μ_1 and μ_2 can be decreased by a switching strategy. Thus, in order to convert the results of Theorems 1, 2 into an algorithmic scheme we only need to point out a strategy for updating these parameters, which is compatible with the condition (4.3). In this section we do that for an important case $L_1(\hat{f}) = L_2(\hat{\phi}) = 0$.

It is convenient to represent the smoothness parameters as follows:

$$\mu_1 = \lambda_1 \cdot \|A\|_{1,2} \cdot \sqrt{\frac{D_2}{\sigma_1\sigma_2 D_1}}, \quad \mu_2 = \lambda_2 \cdot \|A\|_{1,2} \cdot \sqrt{\frac{D_1}{\sigma_1\sigma_2 D_2}}. \tag{6.1}$$

Then the estimate (3.2) for the duality gap becomes symmetric:

$$f(\bar{x}) - \phi(\bar{u}) \le (\lambda_1 + \lambda_2) \cdot \|A\|_{1,2} \cdot \sqrt{\frac{D_1 D_2}{\sigma_1\sigma_2}}. \tag{6.2}$$

At the same time the condition (4.3) becomes problem independent:

$$\frac{\tau^2}{1-\tau} \le \mu_1\mu_2 \cdot \frac{\sigma_1\sigma_2}{\|A\|_{1,2}^2} = \lambda_1\lambda_2. \tag{6.3}$$

Let us write down the switching algorithmic scheme in an explicit form. It is convenient to have a permanent iteration counter. In this case at even iterations we apply the primal update (4.2) (or (5.5)), and at odd iterations we apply the corresponding dual update. Since at even iterations λ_2 is not changing, and at odd iterations λ_1 is not changing, it is convenient to put their new values in same sequence $\{\alpha_k\}_{k=-1}^{\infty}$. Let us fix the following relations between the sequences:

$$\begin{array}{lcl} k = 2l & : & \lambda_{1,k} = \alpha_{k-1}, \quad \lambda_{2,k} = \alpha_k, \\ k = 2l+1 & : & \lambda_{1,k} = \alpha_k, \quad \lambda_{2,k} = \alpha_{k-1}. \end{array} \tag{6.4}$$

Then the parameters τ_k get the following sense.

Lemma 5 *For all $k \ge 0$ we have $\alpha_{k+1} = (1-\tau_k)\alpha_{k-1}$.*

Proof:
Indeed, in accordance to (6.4), if $k = 2l$ then

$$\alpha_{k+1} = \lambda_{1,k+1} = (1-\tau_k)\lambda_{1,k} = (1-\tau_k)\alpha_{k-1}.$$

And if $k = 2l+1$ then $\alpha_{k+1} = \lambda_{2,k+1} = (1-\tau_k)\lambda_{2,k} = (1-\tau_k)\alpha_{k-1}$. □

Corollary 1 *In terms of the sequence $\{\alpha_k\}_{k=-1}^{\infty}$ the condition (6.3) looks as follows:*

$$(\alpha_{k+1} - \alpha_{k-1})^2 \le \alpha_{k+1}\alpha_k\alpha_{k-1}^2, \quad k \ge 0. \tag{6.5}$$

Proof:
Indeed, in view of (6.4) we always have $\lambda_{1,k}\lambda_{2,k} = \alpha_k\alpha_{k-1}$. It remains to use $\tau_k = 1 - \frac{\alpha_{k+1}}{\alpha_{k-1}}$.
□

Clearly, condition (6.5) is satisfied by

$$\alpha_k = \frac{2}{k+2}, \quad k \ge -1. \tag{6.6}$$

Then

$$\tau_k = 1 - \frac{\alpha_{k+1}}{\alpha_{k-1}} = \frac{2}{k+3}, \quad k \geq 0. \tag{6.7}$$

Now we are ready to write down the algorithmic scheme. Let us do that for the gradient mapping update (4.2). In this scheme we use the sequences $\{\mu_{1,k}\}_{k=-1}^{\infty}$ and $\{\mu_{2,k}\}_{k=-1}^{\infty}$, generated in accordance to the rules (6.1), (6.4) and (6.6).

1. **Initialization:**
   Choose $\bar{x}_0$ and $\bar{u}_0$ in accordance to (4.1) with $\mu_1 = \mu_{1,0}$ and $\mu_2 = \mu_{2,0}$.

2. **Iterations ($k \geq 0$):**

   a) Set $\tau_k = \frac{2}{k+3}$. (6.8)

   b) If k is even then generate $(\bar{x}_{k+1}, \bar{u}_{k+1})$ from $(\bar{x}_k, \bar{u}_k)$ using (4.2).

   c) If k is odd then generate $(\bar{x}_{k+1}, \bar{u}_{k+1})$ from $(\bar{x}_k, \bar{u}_k)$ using the symmetric dual variant of (4.2).

Theorem 3 *Let the sequences $\{\bar{x}_k\}_{k=0}^{\infty}$ and $\{\bar{u}_k\}_{k=0}^{\infty}$ be generated by the method (6.8). Then each pair of points satisfy the excessive gap condition. Therefore*

$$f(\bar{x}_k) - \phi(\bar{u}_k) \leq \frac{4\|A\|_{1,2}}{k+1}\sqrt{\frac{D_1 D_2}{\sigma_1 \sigma_2}}. \tag{6.9}$$

Proof:
In accordance to our choice of parameters,

$$\mu_{1,0}\mu_{2,0} = \lambda_{1,0}\lambda_{2,0} \cdot \frac{\|A\|_{1,2}^2}{\sigma_1 \sigma_2} = \frac{2\mu_{2,0}}{\sigma_1} L_1(f_{\mu_2}) > \frac{\mu_{2,0}}{\sigma_1} L_1(f_{\mu_2}).$$

Hence, in view of Lemma 3 the pair $(\bar{x}_0, \bar{u}_0)$ satisfies the excessive gap condition. We have already checked that the sequence $\{\tau_k\}_{k=0}^{\infty}$ defined by (6.7) satisfies the conditions of Theorem 1. Therefore the excessive gap conditions will be valid for the sequences generated by (6.8). It remains to use inequality (6.2). □

Clearly, the same statement is valid for the method based on the updating scheme (5.5).

7 Minimizing a strongly convex function

Consider now the model (2.2), which satisfies the following assumption.

Assumption 1 *In representation (2.2) function $\hat{f}(x)$ is strongly convex with convexity parameter $\hat{\sigma}$.*

Lemma 6 *Under Assumption 1, function $\phi(u)$ defined by (2.3) is differentiable. Moreover, its gradient*

$$\nabla\phi(u) = -\nabla\hat{\phi}(u) + Ax_0(u),$$

with $x_0(u)$ defined in (2.7), is Lipschitz-continuous with the constant

$$L_2(\phi) = \frac{\|A\|_{1,2}^2}{\hat{\sigma}} + L_2(\hat{\phi}). \tag{7.1}$$

Proof:
Denote $\tilde{\phi}(u) = \min\limits_x\{\langle Ax, u\rangle_2 + \hat{f}(x) : \ x \in Q_1\}$. This function is concave as a minimum of linear functions. Since $\hat{f}$ is strongly convex, the solution of the above minimization problem is unique. Therefore $\tilde{\phi}(u)$ is differentiable and $\nabla\tilde{\phi}(u) = Ax_0(u)$.

Consider two points u_1 and u_2. From the first-order optimality conditions in (2.3) we have

$$\langle A^* u_1 + \nabla\hat{f}(x_0(u_1)), x_0(u_2) - x_0(u_1)\rangle_1 \geq 0,$$

$$\langle A^* u_2 + \nabla\hat{f}(x_0(u_2)), x_0(u_1) - x_0(u_2)\rangle_1 \geq 0.$$

Adding these inequalities and using strong convexity of $\hat{f}(\cdot)$, we continue as follows:

$$\langle Ax_0(u_2) - Ax_0(u_1), u_1 - u_2\rangle_2 \geq \langle \nabla\hat{f}(x_0(u_1)) - \nabla\hat{f}(x_0(u_2)), x_0(u_1) - x_0(u_2)\rangle_1$$

$$\geq \hat{\sigma}\|x_0(u_1) - x_0(u_2)\|_1^2 \geq \tfrac{\hat{\sigma}}{\|A\|_{1,2}^2}\left(\|\nabla\tilde{\phi}(u_1) - \nabla\tilde{\phi}(u_2)\|_2^*\right)^2. \qquad \Box$$

Lemma 7 *For any u and $\hat{u}$ from Q_2 we have:*

$$\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2 \geq -\hat{\phi}(u) + \langle Ax_0(\hat{u}), u\rangle_2 + \hat{f}(x_0(\hat{u})). \tag{7.2}$$

Proof:
Let us take arbitrary u and $\hat{u}$ from Q_2. Denote $\hat{x} = x_0(\hat{u})$. Then

$$\begin{aligned}\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2 \ &= -\hat{\phi}(\hat{u}) + \langle A\hat{x}, \hat{u}\rangle_2 + \hat{f}(\hat{x}) + \langle -\nabla\hat{\phi}(\hat{u}) + A\hat{x}, u - \hat{u}\rangle_2 \\ &\geq -\hat{\phi}(u) + \langle A\hat{x}, u\rangle_2 + \hat{f}(\hat{x}).\end{aligned}$$

$\Box$

In this section we derive an optimization scheme from the following variant of the excessive gap condition:

$$f_{\mu_2}(\bar{u}) \leq \phi(\bar{u}) \tag{7.3}$$

for some $\bar{x} \in Q_1$ and $\bar{u}$ in Q_2.

This condition can be seen as a variant of condition (3.1) with $\mu_1 = 0$. However, in this section we prefer do not use the results of the previous sections since now our assumptions are different. For example, we do not need anymore the set Q_1 to be bounded.

Lemma 8 *Let points $\bar{x}$ from Q_1 and $\bar{u}$ from Q_2 satisfy (7.3). Then*

$$0 \leq f(\bar{x}) - \phi(\bar{u}) \leq \mu_2 D_2. \tag{7.4}$$

Proof:
Indeed, for any $x \in Q_1$ we have $f_{\mu_2}(x) \geq f(x) - \mu_2 D_2$. □

Define the adjoint gradient mapping as follows:

$$V(u) = \arg\max_v \{\langle \nabla\phi(u), v - u\rangle_2 - \tfrac{1}{2}L_2(\phi)\|v - u\|_2^2\}. \tag{7.5}$$

Lemma 9 *Let us choose $\mu_2 = \frac{1}{\sigma_2}L_2(\phi)$. Then the excessive gap condition (7.3) is satisfied for*

$$\bar{u} = V(u_0), \quad \bar{x} = x_0(u_0). \tag{7.6}$$

Proof:
Indeed, in view of Lemma 6 we get the following relations:

$$\begin{aligned}
\phi(V(u_0)) \geq\ & \max_u \Big\{\phi(u_0) + \langle \nabla\phi(u_0), u - u_0\rangle_2 - \tfrac{1}{2}L_2(\phi)\|u - u_0\|_2^2 :\ u \in Q_2\Big\} \\
=\ & \max_u \Big\{-\hat{\phi}(u_0) + \langle Ax_0(u_0), u_0\rangle_2 + \hat{f}(x_0(u_0)) \\
& +\langle Ax_0(u_0) - \nabla\hat{\phi}(u_0), u - u_0\rangle_2 - \tfrac{1}{2}\mu_2\sigma_2\|u - u_0\|_2^2 :\ u \in Q_2\Big\} \\
\geq\ & \max_u \Big\{-\hat{\phi}(u) + \hat{f}(x_0(u_0)) + \langle Ax_0(u_0), u\rangle_2 - \mu_2 d_2(u) :\ u \in Q_2\Big\} \\
=\ & f_{\mu_2}(x_0(u_0)).
\end{aligned}$$

□

Theorem 4 *Let points $\bar{x} \in Q_1$ and $\bar{u} \in Q_2$ satisfy the excessive gap condition (7.3) for some positive μ_2. Let us fix $\tau \in (0,1)$ and choose $\mu_2^+ = (1-\tau)\mu_2$,*

$$\begin{aligned}
\hat{u} &= (1-\tau)\bar{u} + \tau u_{\mu_2}(\bar{x}), \\
\bar{x}_+ &= (1-\tau)\bar{x} + \tau x_0(\hat{u}), \\
\bar{u}_+ &= V(\hat{u}).
\end{aligned} \tag{7.7}$$

Then the pair $(\bar{x}_+, \bar{u}_+)$ satisfies condition (7.3) with smoothness parameter μ_2^+, provided that τ is chosen in accordance to the following relation:

$$\frac{\tau^2}{1-\tau} \leq \frac{\mu_2\sigma_2}{L_2(\phi)}. \tag{7.8}$$

Proof:

Denote $\hat{x} = x_0(\hat{u})$. In view of inequality (7.2) and the rules (7.7) we have:

$$
\begin{aligned}
f_{\mu_2^+}(\bar{x}_+) &= \max_u \Big\{(1-\tau)\mu_2 d_2(u) + \langle A((1-\tau)\bar{x} + \tau\hat{x}), u\rangle_2 - \hat{\phi}(u) :\ u \in Q_2\Big\} + \hat{f}(\bar{x}_+) \\
&\le \max_u \Big\{(1-\tau)\left[\mu_2 d_2(u) + \langle A\bar{x}, u\rangle_2 - \hat{\phi}(u) + \hat{f}(\bar{x})\right] \\
&\quad +\tau[-\hat{\phi}(u) + \langle A\hat{x}, u\rangle_2 + \hat{f}(\hat{x})] :\ u \in Q_2\Big\} \\
&\le \max_u \Big\{(1-\tau)\left[f_{\mu_2}(\bar{x}) - \tfrac{1}{2}\mu_2\sigma_2\|u - u_{\mu_2}(\bar{x})\|_2^2\right] \\
&\quad +\tau[\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2] :\ u \in Q_2\} .
\end{aligned}
$$

Note that

$$
f_{\mu_2}(\bar{x}) \le \phi(\bar{u}) \le \phi(\hat{u}) + \langle\nabla\phi(\hat{u}), \bar{u} - \hat{u}\rangle_2 = \phi(\hat{u}) + \tau\langle\nabla\phi(\hat{u}), \bar{u} - u_{\mu_2}(\bar{x})\rangle_2.
$$

Hence, using the condition (7.8), we can finish the proof as follows:

$$
\begin{aligned}
& f_{\mu_2^+}(\bar{x}_+) \\
\le\ & \max_u \Big\{\phi(\hat{u}) + \tau\langle\nabla\phi(\hat{u}), u - u_{\mu_2}(\bar{x})\rangle_2 - \tfrac{1}{2}(1-\tau)\mu_2\sigma_2\|u - u_{\mu_2}(\bar{x})\|_2^2 :\ u \in Q_2\Big\} \\
\le\ & \max_u \Big\{\phi(\hat{u}) + \tau\langle\nabla\phi(\hat{u}), u - u_{\mu_2}(\bar{x})\rangle_2 - \tfrac{1}{2}\tau^2 L_2(\phi)\|u - u_{\mu_2}(\bar{x})\|_2^2 :\ u \in Q_2\Big\} \\
\le\ & \min_u \Big\{\phi(\hat{u}) + \langle\nabla\phi(\hat{u}), u - \hat{u}\rangle_2 - \tfrac{1}{2}L_2(\phi)\|u - \hat{u}\|_2^2 :\ u \in Q_2\Big\} \\
\le\ & \phi(\bar{u}_+).
\end{aligned}
$$

□

Now we can justify the following minimization scheme.

1. Initialization:
Set $\mu_{2,0} = 2\frac{L_2(\phi)}{\sigma_2}$, $\bar{x}_0 = x_0(u_0)$ and $\bar{u}_0 = V(u_0)$.

2. For $k \ge 0$ iterate :

$$
\begin{aligned}
&\text{Set } \tau_k = \tfrac{2}{k+3} \text{ and } \hat{u}_k = (1-\tau_k)\bar{u}_k + \tau_k u_{\mu_{2,k}}(\bar{x}_k). \\
&\text{Update} \quad \mu_{2,k+1} = (1-\tau_k)\mu_{2,k}, \\
&\qquad\qquad \bar{x}_{k+1} = (1-\tau_k)\bar{x}_k + \tau_k x_0(\hat{u}_k), \\
&\qquad\qquad \bar{u}_{k+1} = V(\hat{u}_k).
\end{aligned} \tag{7.9}
$$

□

Theorem 5 *Let problem (2.1) satisfies Assumption 1. Then the pairs $(\bar{x}_k, \bar{u}_k)$ generated by scheme (7.9) satisfy the following inequality:*

$$f(\bar{x}_k) - \phi(\bar{u}_k) \le \frac{4L_2(\phi)D_2}{(k+1)(k+2)\sigma_2}, \tag{7.10}$$

where $L_2(\phi)$ is given by (7.1).

Proof:
Indeed, in view of Theorem 4 and Lemma 9 we need only to justify that the sequences $\{\mu_{2,k}\}_{k=0}^{\infty}$ and $\{\tau_k\}_{k=0}^{\infty}$ satisfy relation (7.8). That is straightforward because of relation

$$\mu_{2,k} = \frac{4L_2(\phi)}{(k+1)(k+2)\sigma_2},$$

which is valid for all $k \ge 0$. □

Let us conclude the paper with an example. Consider the problem

$$f(x) = \tfrac{1}{2}\|x\|_1^2 + \max_{1\le j\le m}[f_j + \langle g_j, x - x_j\rangle_1] \quad \min: \ x \in E_1. \tag{7.11}$$

The problems of this type arise, for example, at each iteration of Bundle Method [1]. Let $E_1 = R^n$ and we choose

$$\|x\|_1^2 = \sum_{i=1}^{n} (x^{(i)})^2, \quad x \in E_1.$$

Then this problem can be solved by the scheme (7.9).

Indeed, we can represent the objective function in (7.11) in the form (2.2) using the following objects:

$$E_2 = R^m, \quad Q_2 = \Delta_m = \{u \in R^m_+ : \sum_{j=1}^{m} u^{(j)} = 1\},$$

$$\hat{f}(x) = \tfrac{1}{2}\|x\|_1^2, \quad \hat{\phi}(u) = \langle b, u\rangle_2, \quad b^{(j)} = \langle g_j, x_j\rangle_1 - f_j, \ j = 1, \dots, m,$$

$$A^T = (a_1, \dots, a_m).$$

Thus, $\hat{\sigma} = 1$ and $L_2(\hat{\phi}) = 0$. Let us choose for E_2 the following norm:

$$\|u\|_2 = \sum_{j=1}^{m} |u^{(j)}|.$$

Then we can use the entropy distance function (see [3]):

$$d_2(u) = \ln m + \sum_{j=1}^{m} u^{(j)} \ln u^{(j)}, \quad u_0 = (\tfrac{1}{m}, \dots, \tfrac{1}{m}),$$

for which $\sigma_2 = 1$ and $D_2 = \ln m$ (see [3]). Note that in this case

$$\|A\|_{1,2} = \max_{1\le j\le m} \|g_j\|_1^*.$$

Thus, method (7.9) as applied to the problem (7.11) converges with the following rate:

$$f(\bar{x}_k) - \phi(\bar{u}_k) \le \frac{4 \ln m \max\limits_{1 \le j \le m} (\| g_j \|_1^*)^2}{(k+1)(k+2)}.$$

Let us study the complexity of this scheme for our example. At each iteration we need to compute the following objects.

1. **Computation of** $u_{\mu_2}(\bar{x})$. This is the solution of the following problem:

$$\max_u \left\{ \sum_{j=1}^m u^{(j)} s^{(j)}(\bar{x}) - \mu_2 d_2(u) : \; u \in Q_2 \right\}$$

   with $s^{(j)}(\bar{x}) = f_j + \langle g_j, \bar{x} - x_j \rangle$, $j = 1, \dots, m$. In accordance to (4.14) in Lemma 4 [3], this solution can be found in a closed form:

$$u_{\mu_2}^{(j)}(\bar{x}) = \frac{e^{s^{(j)}(\bar{x})/\mu_2}}{\sum\limits_{l=1}^m e^{s^{(l)}(\bar{x})/\mu_2}}, \quad j = 1, \dots, m.$$

2. **Computation of** $x_0(\hat{u})$. In our case this is a solution to the problem

$$\min_x \{ \langle Ax, \hat{u} \rangle_2 + \tfrac{1}{2} \| x \|_1^2 : \; x \in E_1 \}.$$

   Hence, the answer is very simple: $x_0(\hat{u}) = -A^T \hat{u}$.

3. **Computation of** $V(\hat{u})$. In our case

$$\begin{aligned} \phi(\hat{u}) &= \min_{x \in E_1} \left\{ \sum_{j=1}^m \hat{u}^{(j)} [f_j + \langle g_j, x - x_j \rangle_1] + \tfrac{1}{2} \| x \|_1^2 \right\} \\ &= -\langle b, \hat{u} \rangle_2 - \tfrac{1}{2} \left(\| A^T \hat{u} \|_1^* \right)^2. \end{aligned}$$

   Thus, $\nabla \phi(\hat{u}) = -b - AA^T \hat{u}$. Now we can compute $V(\hat{u})$ by (7.5). In [3], Section 5.1, it is shown that the complexity of such a computation is of the order $O(m \ln m)$.

Thus, we have seen that all computations at each iteration of the method (7.9) as applied to the problem (7.11) are very cheap. The most expensive part of the iteration is the multiplication of the matrix A by a vector. In a straightforward implementation we need three such multiplications per iteration. However, a simple modification of the order of operations can reduce this amount up to two.

References

[1] J.-B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*, vols. I and II. Springer-Verlag, 1993.

[2] A. Nemirovsky and D. Yudin. *Informational Complexity and Efficient Methods for Solution of Convex Extremal Problems*, J. Wiley & Sons, New York, 1983

[3] Yu.Nesterov. Smooth minimization of non-smooth functions. CORE DP 2003/12. Submitted to *Mathematical Programming*, Series A.