Operations Research Letters 44 (2016) 61–66

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Rate of convergence analysis of dual-based variables decomposition methods for strongly convex problems

Amir Beck*, Luba Tetruashvili, Yakov Vaisbourd, Ariel Shemtov

Faculty of Industrial Engineering and Management, Technion, Haifa, Israel

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form
15 November 2015
Accepted 15 November 2015
Available online 2 December 2015

Keywords:
Dual based methods
Block variables decomposition
Total variation denoising

ABSTRACT

We consider the problem of minimizing the sum of a strongly convex function and a term comprising the sum of extended real-valued proper closed convex functions. We derive the primal representation of dual-based block descent methods and establish a relation between primal and dual rates of convergence, allowing to compute the efficiency estimates of different methods. We illustrate the effectiveness of the methods by numerical experiments on total variation-based denoising problems.

1. Introduction

1.1. The basic setting

In this paper, our aim is to devise simple methods for solving minimization problems of the form

$$\text{(P)}\ \min_{\mathbf{x}\in\mathbb{E}} \left\{ f(\mathbf{x}) + \sum_{i=1}^{m} \psi_i(\mathbf{x}) \right\},$$

with $\mathbb{E}$ being a given final dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and associated Euclidean norm $\|\mathbf{x}\| \equiv \sqrt{\langle \mathbf{x},\mathbf{x}\rangle}$. The functions f and ψ_i satisfy the following conditions that are summarized in one assumption.

Assumption 1.
- $f : \mathbb{E} \to (-\infty,\infty]$ is a closed, proper extended valued σ-strongly convex function.
- $\psi_i : \mathbb{E} \to (-\infty,\infty]$ $(i = 1, 2, \ldots, m)$ is a closed, proper extended real-valued convex function.
- $\mathrm{ri}(\mathrm{dom} f) \cap \left(\bigcap_{i=1}^{m} \mathrm{ri}(\mathrm{dom}\,\psi_i)\right) \neq \emptyset$.

Under the latter assumption, problem (P) has a unique minimizer that we denote by $\mathbf{x}^*$; the optimal value is denoted by $f_{\mathrm{opt}} = f(\mathbf{x}^*)$. The dual problem of (P) is given by

$$\text{(D)}\ \max_{\mathbf{y}} \left\{ q(\mathbf{y}) \equiv -f^*\left(-\sum_{j=1}^{m}\mathbf{y}_j\right) - \sum_{j=1}^{m} \psi_j^*(\mathbf{y}_j) \right\}, \quad (1.1)$$

where $f^*(\cdot) = \sup_{\mathbf{x}\in\mathbb{E}} \langle\cdot,\mathbf{x}\rangle - f(\mathbf{x})$ and $\psi_i^*(\cdot) = \sup_{\mathbf{x}\in\mathbb{E}} \langle\cdot,\mathbf{x}\rangle - \psi_i(\mathbf{x})$ are the corresponding conjugate functions. The duality between (P) and (D) is obviously a simple application of Fenchel's (as well as Lagrangian) duality [18]. In this specific form, it is also known as the duality between the regularized consensus problem and the sharing problem (see Section 7 of [7]).

Since Slater's condition is satisfied, and since the primal problem is bounded below, strong duality holds, which means that the optimal solution of the dual problem is attained and the optimal value of the dual problem, which we denote by q_{opt}, coincides with the primal optimal value:

$$f_{\mathrm{opt}} = \mathrm{val}\,(P) = \mathrm{val}\,(D) = q_{\mathrm{opt}}.$$

Using the notation $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$, the dual problem (D) in minimization form can be written as

$$\min_{\mathbf{y}\in\mathbb{E}^m} \left\{ H(\mathbf{y}) \equiv F(\mathbf{y}) + \sum_{i=1}^{m} \Psi_i(\mathbf{y}_i) \right\} \quad (1.2)$$

with

$$F(\mathbf{y}) \equiv f^*\left(-\sum_{j=1}^{m}\mathbf{y}_j\right), \qquad \Psi_j(\mathbf{y}_j) \equiv \psi_j^*(\mathbf{y}_j). \quad (1.3)$$

Under Assumption 1, $\Psi_1, \Psi_2, \ldots, \Psi_m$ are closed, proper and convex and, by the well-known Baillon–Haddad Lemma (see

* Corresponding author.
E-mail addresses: becka@ie.technion.ac.il (A. Beck), lubate@campus.technion.ac.il (L. Tetruashvili), yakovv@campus.technion.ac.il (Y. Vaisbourd), ariel.shemtov@ensta-paristech.fr (A. Shemtov).

http://dx.doi.org/10.1016/j.orl.2015.11.007
0167-6377/

[19, Section 12H]), F is an L-smooth function with $L = \frac{m}{\sigma}$, meaning that $\|\nabla F(\mathbf{y}) - \nabla F(\mathbf{w})\| \leq L\|\mathbf{y} - \mathbf{w}\|$ for any $\mathbf{y}, \mathbf{w} \in \mathbb{E}$. In addition, for any i, $\nabla_i F$ is Lipschitz continuous with constant $\frac{1}{\sigma}$. The optimal solution set of the dual problem will be denoted by Y^*.

1.2. Paper layout

The main objective of the paper is to present a convergence analysis of dual-based decomposition methods for solving (P), where the basic step in the dual algorithm will either consist of a well known exact minimization [10] or a proximal gradient step [3,9] with respect to the corresponding block of dual variables. We begin in Section 2 by deriving a primal representation of both dual block proximal gradient and dual alternating minimization methods. We then establish in Section 3 a relation between certain primal and dual distances to optimality that allows to automatically translate any rate of convergence result in the dual space into a rate of convergence result in the primal space. We then utilize known results on rates of convergence for variables decomposition methods in order to establish new corresponding results for dual-based decomposition methods. Finally, we demonstrate in Section 4 the potential of the derived methods in the context of total variation-based denoising problems.

1.3. Literature review

Variables decomposition methods such as the alternating minimization method were extensively studied for many years, see e.g., [1,6,16,10]. Rate of convergence results under certain strong convexity and/or error bound assumptions were established in [10,13]. The first rate of convergence result in the deterministic setting without any strong convexity/error bounds assumption was established in [5], where an $O(1/k)$ rate convergence of the block coordinate gradient projection method was shown. In the unconstrained case, it was shown that the method can be accelerated to a rate of $O(1/k^2)$. The work [5] also established an $O(1/k)$ rate of convergence for the alternating minimization method with two blocks in the smooth unconstrained case with a multiplicative constant that depends on the minimum of the block Lipschitz constants. The latter was later generalized in [2] to the case of a composite objective function with a separable nonsmooth paper. Recently, it was shown in [11] that a sublinear rate of convergence can also be established for the block proximal gradient and alternating minimization methods with arbitrary number of blocks. Randomized methods in which the blocks are not picked by a deterministic rule, but rather by some random distribution on the indices set are also the topic of an extensive research [15,17,12].

The idea of solving a problem of the form (P) via a dual-based block decomposition method for the case $m = 2$ was studied in [8].

2. Dual-based block descent methods

2.1. Step types

We begin by describing the two types of minimization operations that will be employed on a given block $i \in \{1, 2, \dots, m\}$. We assume that the dual variables are given by $\mathbf{y}_j = \bar{\mathbf{y}}_j$, $j \in \{1, 2, \dots, m\}$, and show how to compute the new value of $\mathbf{y}_i$, which we denote by $\mathbf{y}_i^{\text{new}}$. We consider two options for the dual step employed on the ith block:

- **dual exact minimization step**.

$$\mathbf{y}_i^{\text{new}} \in \underset{\mathbf{y}_i}{\operatorname{argmin}} \left\{ f^*\left(-\sum_{j=1, j\neq i}^{m} \bar{\mathbf{y}}_j - \mathbf{y}_i \right) + \psi_i^*(\mathbf{y}_i) \right\}. \tag{2.1}$$

Note that for this minimization step, the value of $\bar{\mathbf{y}}_i$ is not being used.

- **dual proximal gradient step**.

$$\mathbf{y}_i^{\text{new}} = \operatorname{prox}_{\sigma\psi_i^*}\left(\bar{\mathbf{y}}_i + \sigma \nabla f^*\left(-\sum_{j=1}^{m} \bar{\mathbf{y}}_j \right) \right). \tag{2.2}$$

2.1.1. Primal representation of the dual exact minimization step

To derive a primal representation of (2.1), let us write it as

$$\min_{\mathbf{y}_i, \mathbf{w}} \left\{ f^*(\mathbf{w}) + \psi_i^*(\mathbf{y}_i) : \mathbf{w} + \mathbf{y}_i = -\tilde{\mathbf{y}}_i \right\}, \tag{2.3}$$

where $\tilde{\mathbf{y}}_i = \sum_{j=1,\, j\neq i}^{m} \bar{\mathbf{y}}_j$. The dual problem of (2.3) is

$$\begin{aligned}
&\max_{\mathbf{x}} \min_{\mathbf{w}, \mathbf{y}_i} \left\{ f^*(\mathbf{w}) + \psi_i^*(\mathbf{y}_i) - \langle \mathbf{x}, \mathbf{w} + \mathbf{y}_i + \tilde{\mathbf{y}}_i \rangle \right\} \\
&\quad = \max_{\mathbf{x}} \left\{ \left[\min_{\mathbf{w}} (f^*(\mathbf{w}) - \langle \mathbf{x}, \mathbf{w} \rangle) \right] \right. \\
&\qquad \left. + \left[\min_{\mathbf{y}_i} (\psi_i^*(\mathbf{y}_i) - \langle \mathbf{x}, \mathbf{y}_i \rangle) \right] - \langle \mathbf{x}, \tilde{\mathbf{y}}_i \rangle \right\} \\
&\quad = \max_{\mathbf{x}} \left\{ -f(\mathbf{x}) - \psi_i(\mathbf{x}) - \langle \mathbf{x}, \tilde{\mathbf{y}}_i \rangle \right\},
\end{aligned}$$

where in the last equality we used the fact that $f = f^{**}$ and $\psi_i = \psi_i^{**}$ (since f and ψ_i are closed, proper and convex). We can thus conclude that $\mathbf{y}_i^{\text{new}}$ can be determined by first computing $\bar{\mathbf{x}} \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{E}} \left\{ f(\mathbf{x}) + \psi_i(\mathbf{x}) + \langle \tilde{\mathbf{y}}_i, \mathbf{x} \rangle \right\}$, and then choosing $\mathbf{y}_i^{\text{new}} \in \operatorname{argmax}_{\mathbf{y}_i} \left\{ \langle \mathbf{y}_i, \bar{\mathbf{x}} \rangle - \psi_i^*(\mathbf{y}_i) \right\}$, which is exactly the same as $\mathbf{y}_i^{\text{new}} \in \partial\psi_i(\bar{\mathbf{x}})$. Therefore, step (2.1) is equivalent to

Primal representation of the dual exact minimization step:

$$\bar{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{E}}{\operatorname{argmin}} \left\{ f(\mathbf{x}) + \psi_i(\mathbf{x}) + \langle \tilde{\mathbf{y}}_i, \mathbf{x} \rangle \right\}, \quad \left(\tilde{\mathbf{y}}_i = \sum_{j\neq i} \bar{\mathbf{y}}_j \right) \tag{2.4}$$

$$\mathbf{y}_i^{\text{new}} \in \partial\psi_i(\bar{\mathbf{x}}). \tag{2.5}$$

When f is also assumed to be continuously differentiable over $\mathbb{E}$, we can use the first-order optimality condition on problem (2.4) to conclude that $-\nabla f(\bar{\mathbf{x}}) - \tilde{\mathbf{y}}_i \in \partial\psi_i(\bar{\mathbf{x}})$. Therefore, step (2.5) can be replaced (in this setting) with $\mathbf{y}_i^{\text{new}} = -\nabla f(\bar{\mathbf{x}}) - \tilde{\mathbf{y}}_i$.

2.1.2. Primal representation of the dual proximal gradient step

To find a primal representation of (2.2), first note that

$$\begin{aligned}
\nabla f^*\left(-\sum_{j=1}^{m} \bar{\mathbf{y}}_j \right) &= \underset{\mathbf{x}\in\mathbb{E}}{\operatorname{argmax}} \left\{ \left\langle -\sum_{j=1}^{m} \bar{\mathbf{y}}_j, \mathbf{x} \right\rangle - f(\mathbf{x}) \right\} \\
&= \underset{\mathbf{x}\in\mathbb{E}}{\operatorname{argmin}} \left\{ f(\mathbf{x}) + \left\langle \sum_{j=1}^{m} \bar{\mathbf{y}}_j, \mathbf{x} \right\rangle \right\}.
\end{aligned}$$

By denoting the above argmin/argmax by

$$\bar{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{E}}{\operatorname{argmin}} \left\{ f(\mathbf{x}) + \left\langle \sum_{j=1}^{m} \bar{\mathbf{y}}_j, \mathbf{x} \right\rangle \right\},$$

we obtain that the proximal gradient step takes the form $\mathbf{y}_i^{\text{new}} = \operatorname{prox}_{\sigma\psi_i^*}(\bar{\mathbf{y}}_i + \sigma\bar{\mathbf{x}})$. Using the Moreau decomposition formula [14], $\operatorname{prox}_{\sigma\psi_i^*}(\mathbf{z}) = \mathbf{z} - \sigma \operatorname{prox}_{\psi_i/\sigma}(\mathbf{z}/\sigma)$, and hence,

$$\mathbf{y}_i^{\text{new}} = \bar{\mathbf{y}}_i + \sigma\bar{\mathbf{x}} - \sigma \operatorname{prox}_{\psi_i/\sigma}\left(\frac{\bar{\mathbf{y}}_i}{\sigma} + \bar{\mathbf{x}} \right).$$