Proximal Primal-Dual Optimization Methods

Audrey Repetti, Emilie Chouzenoux, and Jean-Christophe Pesquet
Université Paris-Est, Laboratoire d'Informatique Gaspard Monge – CNRS UMR 8049
77454 Marne la Vallée Cedex 2, France

***Abstract*—In the field of inverse problems, one of the main benefits which can be drawn from primal-dual optimization approaches is that they do not require any linear operator inversion. In addition, they allow to split a convex objective function in a sum of simpler terms which can be dealt with individually either through their proximity operator or through their gradient if they correspond to smooth functions. Proximity operators constitute powerful tools in nonsmooth functional analysis which have been at the core of many advances in sparsity aware data processing. Using monotone operator theory, the convergence of the resulting algorithms can be shown to be theoretically guaranteed.**

In this paper, we provide a survey of the existing proximal primal-dual approaches which have been proposed in the recent literature. We will also present new developments based on a randomization of these methods, which allow them to be applied block-coordinatewise or in a distributed fashion.

I. Introduction

Recently, there has been a growing interest in primal-dual methods for optimizing functions of a high number of variables. A great majority of such optimization problems arising in computer vision, inverse problems, or machine learning can be formulated as follows:

$$\underset{\mathsf{x}_1\in\mathsf{H}_1,\ldots,\mathsf{x}_p\in\mathsf{H}_p}{\text{minimize}} \quad \sum_{j=1}^{p}\big(\mathsf{f}_j(\mathsf{x}_j)+\mathsf{h}_j(\mathsf{x}_j)\big)+\sum_{k=1}^{q}(\mathsf{g}_k\,\square\,\mathsf{l}_k)\Big(\sum_{j=1}^{p}\mathsf{L}_{k,j}\mathsf{x}_j\Big) \tag{1}$$

where, for every $j\in\{1,\ldots,p\}$, H_j is a separable real Hilbert space, $\mathsf{f}_j\in\Gamma_0(\mathsf{H}_j)$, $\mathsf{h}_j\in\Gamma_0(\mathsf{H}_j)$ is Lipschitz-differentiable, and, for every $k\in\{1,\ldots,q\}$, G_k is a separable real Hilbert space, $\mathsf{g}_k\in\Gamma_0(\mathsf{G}_k)$, $\mathsf{l}_k\in\Gamma_0(\mathsf{G}_k)$ is strongly convex, and $\mathsf{L}_{k,j}\colon\mathsf{H}_j\to\mathsf{G}_k$ is a bounded linear operator.[1] The symbol $\square$ here denotes the inf-convolution operation, that is

$$(\forall k\in\{1,\ldots,q\})(\forall \mathsf{v}_k\in\mathsf{G}_k)\qquad (\mathsf{g}_k\,\square\,\mathsf{l}_k)(\mathsf{v}_k)=\inf_{\mathsf{v}'_k\in\mathsf{G}_k}\mathsf{g}_k(\mathsf{v}'_k)+\mathsf{l}_k(\mathsf{v}_k-\mathsf{v}'_k). \tag{2}$$

In particular, if l_k reduces to the indicator function of $\{0\}$, $\mathsf{g}_k\,\square\,\mathsf{l}_k=\mathsf{g}_k$, which corresponds to a case frequently encountered in practical applications. One of the main interests in considering the general formulation in (1) is that the dual problem has a fully symmetric form:

$$\underset{\mathsf{v}_1\in\mathsf{G}_1,\ldots,\mathsf{v}_q\in\mathsf{G}_q}{\text{minimize}} \quad \sum_{j=1}^{p}(\mathsf{f}_j^*\,\square\,\mathsf{h}_j^*)\Big(-\sum_{k=1}^{q}\mathsf{L}_{k,j}^*\mathsf{v}_k\Big)+\sum_{k=1}^{q}\big(\mathsf{g}_k^*(\mathsf{v}_k)+\mathsf{l}_k^*(\mathsf{v}_k)\big), \tag{3}$$

where φ^* designates the Fenchel-Legendre conjugate of a function φ. Solving (1) raises a number of difficulties due to the facts that (*i*) the space $\mathbf{H}=\mathsf{H}_1\oplus\cdots\oplus\mathsf{H}_p$ in which the unknown solution lies is often very high-dimensional, (*ii*) the functions $(\mathsf{f}_j)_{1\leqslant j\leqslant p}$ and $(\mathsf{g}_k)_{1\leqslant k\leqslant q}$ may be nonsmooth so as to promote the sparsity of the solution or to impose some hard constraints on it, (*iii*) the number of functions may be large, and (*iv*) the involved linear operators may be difficult to (pseudo-)invert because of their size and lack of structure. To overcome these difficulties, primal-dual algorithms have been recently developed which aim at jointly solving the primal and the dual problems instead of dealing with the primal problem only.

[1] $\Gamma_0(\mathsf{H})$ denotes the class of lower-semicontinuous convex functions with nonempty domain defined on H and taking their values in $]-\infty,+\infty]$.

II. An overview of recent algorithms

Proximal primal-dual methods for solving (1) and (3) (or specific instances of them) can be decomposed in three main families (see [1] and the references therein for more details):

- forward-backward based methods: these methods are based on the forward-backward iteration, which means that they combine proximal steps and gradient steps. Two main variants of these methods have been proposed in the literature.
- forward-backward-forward based methods: compared with the previous approaches, these techniques include an extra-gradient step. They were the first proposed methods allowing to address Problem (1) under its general split form.
- projection-based methods: these quite recent approaches present the potential advantage of requiring no knowledge on the norms of the linear operators (or an upper bound of it). They have been however little used in practice, up to now.

Parallel versions of all these algorithms are available, but they require to activate all the primal and dual variables at each iteration which may be intensive in terms of computation and memory requirements. In the spirit of the work performed for designing stochastic versions of forward-backward-like variational methods [2], efforts have been made in order to develop block-coordinate versions of the primal-dual algorithms belonging to the first family. These novel algorithms are grounded on recent theoretical results concerning the randomization of proximal algorithms, and more generally fixed point algorithms for solving monotone inclusion problems [3].

Interestingly, by using consensus techniques, it is possible to deduce from these block-coordinate primal-dual algorithms, distributed algorithms involving linear operators where a splitting is performed on the functions instead of the variables [4].

References

[1] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," 2014, http://www.optimization-online.org/DB_HTML/2014/06/4398.html.

[2] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Math. Program.*, vol. 144, no. 1–2, pp. 1–38, Apr. 2014.

[3] P. L. Combettes and J.-C. Pesquet, "Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping," 2014, http://www.optimization-online.org/DB_HTML/2014/04/4333.html.

[4] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," 2014, http://arxiv.org/abs/1406.6404.