Notes on The Primal-Dual Method

Yuxuan Qiao

Department of Mathmetics, Nanjing University.

Abstract

In this note, we briefly talk about the basic materials for the primal-dual method. In Sect 2, we introduce the primal-dual problem from the image restoration background. In Sect 3, we deduce the optimality conditions for the primal-dual problem. In Sect 4, we present the initial Primal-Dual Hybrid Gradient by M. Zhu. in 2008 and show its non-convergence for some cases. In Sect 5, we discuss the modified Primal-Dual Hybrid Gradient with better convergence by Chambolle and Pock. In Sect 6, we think about the non-convex cases not included in courses, and we present some non-convex primal-dual methods for this type of saddle point problem.

1 Introduction

The primal-dual method is an optimization algorithm commonly used in image processing and computer vision applications. It is particularly effective for solving variational problems that involve minimizing an objective function consisting of a data fidelity term and a regularization term.

The primal-dual method is derived from the primal-dual formulation of the variational problem, which establishes a duality relationship between the primal variables (e.g., the image to be reconstructed) and the dual variables (associated with the regularization term). By leveraging this duality, the primal-dual method offers an efficient and effective way to solve variational problems.

In each iteration of the primal-dual method, primal and dual variables are updated alternately using specific update rules derived from the variational problem. The primal update typically aims to minimize the data fidelity term, while the dual update encourages regularization and enforces constraints. This alternating update scheme ensures that the solution converges to an optimal trade-off between fidelity and regularization.

The primal-dual method is particularly well-suited for problems involving non-smooth and non-convex regularization terms, such as total variation (TV) regularization. TV regularization promotes piecewise smoothness and edge preservation, making it valuable for denoising, image inpainting, and image reconstruction tasks.

The advantages of the primal-dual method include computational efficiency, robustness to noise, and the ability to handle large-scale problems. It also provides flexibility in incorporating additional constraints and priors into the optimization framework.

In conclusion, the primal-dual method is a powerful optimization algorithm based on variational methods. It is widely used in image processing and computer vision for solving variational problems with non-smooth regularization terms, such as total variation. By balancing data fidelity and regularization, the primal-dual method enables effective image restoration, inpainting, and reconstruction.

2 Problem

We will introduce the primal-dual problem from an image restoration example to present the background of the primal-dual method.

Let $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ be an original image describing a real scene by gray intensity, and let u_0 be the observed image of the same scene(a degradation of u). We assume that

$$u_0 = Ru + \eta \tag{1}$$

where R is a linear operator usually representing the blur and where η stands for a white additive Gaussian noise. Given an observed image u_0, the restoration problem is to reconstruct u with the knowledge (1). Recovering u from u_0 knowing (1) is a typical example of an inverse problem. By supposing that η is a white Gaussian noise, and according to the maximum likelihood principle, we can find an approximation of u by solving the least-square problem

$$\inf_u \int_\Omega |u_0 - Ru|^2 dx \tag{2}$$

where Ω is the domain of the image and $|\cdot|$ stands for the Euclidean norm. The problem (2) admits a unique solution characterized by the Euler-Lagrange equation

$$R^* u_0 - R^* Ru = 0 \tag{3}$$

where R^* is the adjoint of R. However, solving(3) is an ill-posed problem in general, since R^*R is not always one-to-one, and even if R^*R is one-to-one, the eigenvalues may be very small, causing numerical instability. In consequence, there is a natural idea of regularization. Rudin, Osher, and Fatemi proposed to consider the following minimization problem which is also called total variation

$$F(u) = \int_\Omega |u_0 - Ru|^2 dx + \lambda \int_\Omega |\nabla u| dx \tag{4}$$

The first term in $F(u)$ measures the fidelity to the data. The second is a smoothing term that provides regularity and guarantees that its gradient is low (so that noise will be removed). The parameter λ is a weighting constant. By the functional spaces theory and variational method, if the Euler-Lagrange equation of (4) were given some specific boundary condition, the total variation problem (4) admits a unique solution. However, there is another issue we are concerned about, that is how to obtain the minimum point numerically. In other words, we desire to approximate the real image u by (4).

By discretization, minimization of the problem (4) can be written as

$$\min_x f(x) + g(Ax) \tag{5}$$

In general, let f, g be closed convex functions, and A be a linear transformation represented by a matrix. The problem (5) is called the primal problem. We can vary the form of (5) with conjugate functions

$$\max_z -g^*(z) - f^*\left(-A^\intercal z\right) \tag{6}$$

$$\min_x \max_z f(x) + z^\intercal Ax - g^*(z) \tag{7}$$

The problem (6) is called the Fenchel dual problem, which is Lagrange dual of the following reformulated problem

$$\min_{x,y} f(x) + g(y) \quad \text{ s.t. } Ax = y \tag{8}$$

The saddle point problem (7) is called the primal-dual problem, which is the issue we discuss in this note.

3 Primal-Dual Optimality Conditions

The optimality conditions play a crucial role in the analysis and solution of optimization problems. By examining the optimality conditions, we can verify the validity and quality of a solution, thereby determining whether it is indeed the optimal solution to the problem. These conditions not only provide a theoretical foundation for optimization problems but also guide the development of optimization algorithms and their applications.

As mentioned in Sect 2, the minimization problem has four different formulations (5)(6)(7)(8), and naturally, their optimality conditions also have different manifestations. However, they are all equivalent to each other. In fact, we can give the following optimality conditions

Optimality conditions(Karush-Kuhn-Tucker conditions)

- primal feasibility: $x \in \operatorname{dom} f$ and $Ax = y \in \operatorname{dom} g$
- dual feasibility: $-A^T z \in \partial f(x)$ and $z \in \partial g(y)$

Remark 1. *the primal feasibility derives from the reformulated problem (8), and the dual feasibility comes from that (x, y) is a minimizer of the Lagrangian $f(x) +$*

$g(x) + z^{\intercal}(Ax - y)$. *According to the subgradient property of the optimal point of the unconstrained optimization problem, we have*

$$\begin{aligned} 0 &\in \partial f(x) + A^{\intercal} z \\ 0 &\in \partial g(y) - z \end{aligned} \tag{9}$$

thus deriving $-A^T z \in \partial f(x)$ *and* $z \in \partial g(y)$, *equivalent to* $x \in \partial f^*(-A^{\intercal} z)$ *and* $y \in \partial g^*(z)$.

By $Ax = y$ in the primal feasibility, the dual feasibility can be written as

$$-A^{\intercal} z \in \partial f(x), \quad Ax \in \partial g^*(z)$$

$$\text{i.e.} \quad \begin{aligned} 0 &\in A^{\intercal} z + \partial f(x) \\ 0 &\in -A^{\intercal} x + \partial g^*(z) \end{aligned}$$

Considering matrix forms, the optimality conditions can be written symmetrically as the following formulation

$$0 \in \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{pmatrix} x \\ z \end{pmatrix} + \begin{pmatrix} \partial f(x) \\ \partial g^*(z) \end{pmatrix} \tag{10}$$

where

$$\begin{pmatrix} \partial f(x) \\ \partial g^*(z) \end{pmatrix} = \left\{ \begin{pmatrix} u \\ v \end{pmatrix} \middle| \begin{aligned} u \in \partial f(x) \\ v \in \partial g^*(z) \end{aligned} \right\}$$

Solutions to the primal-dual problem (7) are saddle points of the convex-concave function $f(x) + z^{\intercal} Ax - g^*(z)$. Here we assume that the optimality conditions are solvable, i.e. there exists a primal-dual optimal point. This assumption is of course subject to conditions. In fact, if primal is solvable and following Slater's condition holds, the optimality conditions are solvable.

Definition 1 (Slater's condition)**.** *There exists* $\hat{x} \in \operatorname{int} \operatorname{dom} f, \text{s. t.} A\hat{x} \in \operatorname{int} \operatorname{dom} g$, *which is Slater's condition for the problem* $\min_x f(x) + g(Ax)$.

Remark 2. *In the case of Slater's condition holding, strong duality holds and dual optimality is attained.*

4 Initial Primal-Dual Hybrid Gradient Algorithm

In Sect 2, we have mentioned that the total variation problem can be written as (5) by discretization. In this section, we consider an easier and more specific primal-dual saddle point problem by setting R as the identity operator in the total variation problem

$$\min_x \max_z 2\|x - l\|^2 + \lambda x^{\intercal} A z \tag{11}$$

where l comes from observation and λ is a weighting constant. In this section, we denote $F(x, z) = 2\|x - l\|^2 + \lambda x^{\intercal} A z$.

To solve the problem (8), in 2008, M. Zhu and T. Chan developed a simple yet efficient algorithm which we refer to as the initial primal-dual hybrid gradient(PDHG) algorithm. It is explicit so the memory requirement is low and each iteration only takes $O(N)$ operations. Before their work, developed gradient-descent type methods were so-called the primal gradient descent algorithm and dual gradient descent algorithm. In fact, they are both special cases of the initial PDHG algorithm.

Initial Primal-Dual Hybrid Gradient Algorithm Given any intermediate solution (y^k, x^k) at iteration step k, the initial PDHG algorithm updates the solution as follows.

1. **DUAL STEP** Fix $x = x^k$, apply one step of the projected gradient ascent method to the maximization problem

$$\max_z F\left(x^k, z\right) \tag{12}$$

The ascent direction is $\nabla_z F(x^k, z) = A^\intercal x^k$, so we update z as

$$z^{k+1} = \text{Proj}\left(z^k + \frac{\tau_k}{\lambda} A^\intercal x^k\right) \tag{13}$$

where τ_k is the dual stepsize and $Proj$ denotes the projection onto the corresponding set

$$\text{Proj}(x) = \arg\min_z \|x - z\|$$

The factor λ is used in (13) so that the stepsize τ_k will not be sensitive to different problems or scales of gray levels.

2. **PRIMAL STEP** Fix $z = z^{k+1}$, apply one step of the gradient descent method to the minimization problem

$$\max_x F\left(x, z^{k+1}\right) \tag{14}$$

The ascent direction is $\nabla_x F\left(x, z^{k+1}\right) = \lambda A z^{k+1} + \left(x^k - l\right)$, so we update x as

$$x^{k+1} = x^k - \theta_k\left(\lambda A z^{k+1} + x^k - l\right) \tag{15}$$

By alternate iterations, we have the following algorithm

Algorithm 1 Initial PDHG Algorithm

1: Initialization. Pick y_0 and a feasible x_0, set $k \Leftarrow 0$.
2: Choose stepsize τ_k and θ_k.
3: Updating

$$z^{k+1} = \text{Proj}\left(z^k + \frac{\tau_k}{\lambda} A^\intercal x^k\right)$$
$$x^{k+1} = x^k - \theta_k\left(\lambda A z^{k+1} + x^k - l\right)$$

4: Terminate if a stopping criterion is satisfied; otherwise set $k \Leftarrow k + 1$ and return to step 2.

Moreover, the initial PDHG algorithm can also be developed as a primal-dual proximal-point method

$$
\begin{aligned}
z^{k+1} &= \arg\max_{z} F\left(x^k, z\right) - \frac{\lambda}{\tau_k}\left\|z - z^k\right\|^2 \\
x^{k+1} &= \arg\min_{x} F\left(x, z^{k+1}\right) + \frac{(1-\theta_k)}{\lambda\theta_k}\left\|x - x^k\right\|^2
\end{aligned}
\tag{16}
$$

The idea here is that when using the dual variable to update the primal variable, since the dual variable is not optimal yet, we do not want to solve the primal minimization problem exactly, instead, we add a penalty term to force the new update close to the previous value, and the same applies to the dual step. A similar derivation will be presented in the next section.

Remark 3. *The primal subgradient descent method and the dual projected gradient descent method are two special cases of our algorithm, which correspond to taking special stepsizes $\tau_k = \infty$ and $\theta_k = 1$ respectively in* (13)(15).

The initial PDHG algorithm provides an original PDHG scheme that can be extended to other saddle point problems, but we should notice that in some cases the convergence of the initial PDHG algorithm cannot be guaranteed, and here is an example.

Example 1 (The nonconvergence of the initial PDHG algorithm)**.** *Let us consider the linear programming in* $\mathbb{R}$

$$\min x \quad \text{s.t. } x \geq 1, x \geq 0$$

the dual problem is

$$\max y \quad \text{s.t. } y \leq 1, y \leq 0$$

Obviously, the unique optimal solutions to the problems above are $x^* = 1$ and $y^* = 1$, and the Lagrange function of the primal problem is

$$L(x, y) = x - y(x - 1)$$

and $(x^*, y*) = (1, 1)$ is the unique saddle point of the Lagrange function.

Finding the saddle point of $L(x, y)$ by use of the original PDHG scheme is the iteration of the following format

$$
\begin{aligned}
y^{k+1} &= \arg\max_{y>0} L\left(x^k, y\right) - \frac{s}{2}\left\|y - y^k\right\|^2 \\
x^{k+1} &= \arg\min_{x>0} L\left(x, y^{k+1}\right) + \frac{r}{2}\left\|x - x^k\right\|^2
\end{aligned}
\tag{17}
$$

Then, setting $r = s = 1$, starting with $(x^0, y^0) = (0, 0)$, the sequence $\{(x^k, y^k)\}$ generated by (17) is

$$\begin{pmatrix} x^0 \\ y^0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} x^7 \\ y^7 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x^1 \\ y^1 \end{pmatrix}, \quad \ldots$$

which is depicted in Figure 1. This is actually a cyclic sequence.

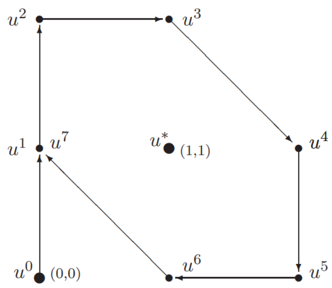

Fig. 1 The illustration of Example 1

Hence, the initial PDHG scheme for the saddle point problem $\min\max L(x, y)$ does not converge. Example 1 comes from the work of B. He, Y. You, and X. Yuan in 2014. We follow their statements and figures.

5 Modified Primal-Dual Hybrid Gradient Algorithm

In Sect 4, we discuss the initial primal-dual hybrid gradient algorithm proposed by M. Zhu and T. Chan. However, in Example 1, we notice that the initial PDHG scheme for some saddle point problems does not converge. Therefore, we have sufficient motivation to propose a modified PDHG algorithm to provide a framework with wider applicability. In this section, we will talk about the modified PDHG algorithm proposed by Chambolle and Pock in 2011.

The general problem they considered in their paper is the generic saddle point problem

$$\min_{x\in X}\max_{y\in Y}\langle Kx, y\rangle + G(x) - F^*(y)$$

where $\langle\cdot,\cdot\rangle$ represents an inner product, and K is a continuous linear operator with the induced norm. We consider the special case of the natural inner product in Euclidean spaces, which is the problem (7) indeed.

$$\min_x\max_z f(x) + z^\intercal Ax - g^*(z)$$

Similar to the dual step and the primal step in the initial PDHG algorithm, the modified PDHG algorithm also has corresponding iteration formats

$$\begin{aligned} x^{k+1} &= \mathrm{Prox}_{\tau f}\left(x^k - \tau A^\intercal z^k\right) \\ z^{k+1} &= \mathrm{Prox}_{\sigma g^*}\left(z^k + \sigma A\left(2x^{k+1} - x^k\right)\right) \end{aligned} \tag{18}$$

Here we give a certain explanation of the iteration formats (18), and we still denote $F(x, z) = f(x) + z^\intercal Ax - g^*(z)$

1. **For the iteration format above in (18)** Fix $z = z^k$, apply the proximal point method to the primal-dual problem, and we get

$$\min_x F\left(x, z^k\right) + \frac{1}{2\tau}\left\|x - x^k\right\|^2$$

where τ stands for the stepsize, and we get the iteration sequence

$$x^{k+1} = \arg\min_x F\left(x, z^k\right) + \frac{1}{2\tau}\left\|x - x^k\right\|^2$$

According to the subgradient property of the optimal point of the unconstrained optimization problem, we have

$$0 \in \partial_x F\left(x^{k+1}, z^k\right) + \frac{1}{\tau}\left(x^{k+1} - x^k\right)$$

By computing the subgradient $\partial_x F$, we get

$$0 \in \partial f\left(x^{k+1}\right) + A^{\intercal} z^k + \frac{1}{\tau}\left(x^{k+1} - x^k\right)$$

$$\text{i.e.} \quad 0 \in (\tau \partial f + I)x^{k+1} + \tau A^{\intercal} z^k - x^k$$

which can be written as

$$x^k - \tau A^{\intercal} z^k \in (I + \tau \partial f)x^{k+1} \tag{19}$$

The relation (19) means the iteration format above we want, i.e.

$$x^{k+1} = \mathrm{Prox}_{\tau f}\left(x^k - \tau A^{\intercal} z^k\right)$$

2. **For the iteration format below in (18)** First, we denote that $\hat{x}^{k+1} = 2x^{k+1} - x^k = x^{k+1} + \left(x^{k+1} - x^k\right)$, representing the point on the extension line connecting x^k and x^{k+1}. Fix $x = \hat{x}^{k+1}$, apply the proximal point method to the primal-dual problem, and we get

$$\max_z F\left(\hat{x}^{k+1}, z\right) - \frac{1}{2\sigma}\|z - z^k\|^2$$

where σ stands for the stepsize, and we get the iteration sequence

$$z^{k+1} = \arg\max_z F\left(\hat{x}^{k+1}, z\right) - \frac{1}{2\sigma}\|z - z^k\|^2$$

Similarly, according to the subgradient property, we have

$$0 \in \partial_z F\left(\hat{x}^{k+1}, z^{k+1}\right) - \frac{1}{\sigma}\left(z^{k+1} - z^k\right)$$

By computing the subgradient $\partial_z F$, we get

$$0 \in A\hat{x}^{k+1} - \partial g^*\left(z^{k+1}\right) - \frac{1}{\sigma}\left(z^{k+1} - z^k\right)$$

which can be written as

$$\sigma A\hat{x}^{k+1} + z^k \in (I + \sigma\partial g^*)\, z^{k+1} \tag{20}$$

The relation (20) means the iteration format below we want, i.e.

$$z^{k+1} = \mathrm{Prox}_{\sigma g^*}\left(z^k + \sigma A\hat{x}^{k+1}\right)$$

What should be mentioned here is that in the iteration format below, we fix $x = \hat{x}^{k+1}$ instead of $x = x^{k+1}$ like (16). In fact, fixing x on the extension line can guarantee the convergence of the algorithm in some sense. Due to this operation, applying the modified PDHG algorithm to Example 1 results in the convergence to the optimal point $(1, 1)$.

Remark 4. *When $z = z^k$ is fixed, the primal-dual problem seems to become $\min\limits_x F(x, z^k)$. However, $F(x, z^k)$ may be not strongly convex, causing problems in the existence and uniqueness of the minimizer. The robust way is to add the term of proximal point term.*

Now we present the modified PDHG algorithm below by alternate iterations

---

Algorithm 2 Modifeid PDHG Algorithm

1: Initialization. Pick y_0 and a feasible x_0, set $k \Leftarrow 0$.
2: Choose stepsize τ and σ.
3: Updating

$$x^{k+1} = \mathrm{Prox}_{\tau f}\left(x^k - \tau A^\intercal z^k\right)$$
$$z^{k+1} = \mathrm{Prox}_{\sigma g^*}\left(z^k + \sigma A\left(2x^{k+1} - x^k\right)\right)$$

4: Terminate if a stopping criterion is satisfied; otherwise set $k \Leftarrow k + 1$ and return to step 2.

During the updating of variables, each iteration requires evaluations of proximal mappings of f and g^* and multiplications with A and $A^\intercal$, with no need to compute solutions to linear equations. Therefore, the PDHG algorithm has the ability to handle large-scale problems with computational efficiency.

6 Extension: The Primal-Dual Scheme In Some Non-convex Cases

In Sect 2, in consequence of the linear property of R, we write the minimization of (4) as the problem (5) $\min_x f(x) + g(Ax)$, where f and g are both closed and convex.

Then, we can apply the so-called PDHG algorithm naturally to solve the primal-dual problem. However, if we replace the linear operator R representing blurring with a more general operator T, the model can be constructed as

$$\min_x \Phi(z - Tx) + G(Ax) \tag{21}$$

where Φ represents the model for data fidelity term, usually L^2-norm, and $G \circ A$ comes from the regularization term. Indeed, the operator T here can be nonlinear in some cases, such as diffusion tensor imaging. For these cases, the corresponding primal-dual problems are non-convex-concave. Despite this, we want to get the approximation of the saddle points(probably not unique).

Although the PDHG is designed for the convex problem (i.e. the convex-concave primal-dual problem), its scheme can also be used in the non-convex problem to search for local optimal solutions or approximate optimal solutions, while these solutions are probably not the global optimal solution. Due to the appearance of many local optimal points, the analysis of convergence and efficiency in these cases can be more complex and troublesome. In numerical applications, there needs to adopt suitable stepsizes, initialization, and stop criteria. Moreover, we can introduce stochastic methods when dealing with non-convex problems to avoid iterations falling into local optimality. By adaptation, the PDHG scheme still has certain robustness and compatibility.

In addition to the PDHG method we mentioned, there exists more work on the problem (21). In fact, if the noise model Φ is convex, proper, and lower semicontinuous. We can write the problem (21) as the primal-dual problem

$$\min_x \max_{(y_1,y_2)} K_{TA}\left(x,(y_1,y_2)\right) - \Phi^*\left(y_1\right) - G_*\left(y_2\right) \tag{22}$$

where $K_{TA}\left(x,(y_1,y_2)\right) = \langle z - T(x) \mid y_1\rangle + \langle Ax \mid y_2\rangle$, $\langle\cdot|\cdot\rangle$ denotes the dual product, and G_* denotes the preconjugate meaning $G = \left(G_*\right)^*$. We will find that the problem (22) is a special case of the following problem

$$\min_{x\in X}\min_{y\in Y} F(x) + K\left(x,y\right) + G_*(y) \tag{23}$$

where $K \in C^1\left(X,Y\right)$, and F and G_* are convex, proper, lower semicontinuous functions. Our goal is to solve the saddle problem (23) at present.

When K is affine, we have many available methods. Among them, there are the primal-dual proximal splitting algorithm (PDPS) and the primal-dual Bregman-proximal splitting algorithm (PDBS) based on Bregman divergence. We present the iteration formats of them.

Given Gâteaux-differentiable functions J_X and J_Y, and We denote D for Gâteaux derivative. Iteratively over $k \in \mathbb{N}$, solve for x_{k+1} and y_{k+1}:

Primal–Dual Bregman-Proximal Splitting (PDBS)

$$\begin{aligned}&DJ_X\left(x^k\right) - D_xK\left(x^k,y^k\right) \in DJ_X\left(x^{k+1}\right) + \partial F\left(x^{k+1}\right) \text{ and}\\&DJ_Y\left(y^k\right) - D_yK\left(x^k,y^k\right) \in DJ_Y\left(y^{k+1}\right) + \partial G_*\left(y^{k+1}\right) - 2D_yK\left(x^{k+1},y^{k+1}\right)\end{aligned} \tag{24}$$

Given specific conditions, the PDBS method can be arranged as the PDPS method below

Primal–Dual Proximal Splitting (PDPS)

$$\begin{aligned} x^{k+1} &:= \operatorname{prox}_{\tau F}\left(x^k - \tau \nabla_x K\left(x^k, y^k\right)\right) \\ y^{k+1} &:= \operatorname{prox}_{\sigma\left[G_* - 2K\left(x^{k+1}, \cdot\right)\right]}\left(y^k - \sigma \nabla_y K\left(x^k, y^k\right)\right) \end{aligned} \tag{25}$$

If K were bilinear, the convergence of iteration formats can follow from Rockafellar in consequence of the monotone operator theory. However, in the problem (23), K does not have to be affine, causing troublesome convergence of the PDBS method and the PDPS method. Now we present the modified primal-dual methods by T. Valkonen based the PDBS method and the PDPS method, dealing with the nonlinear cases.

Modified Primal–Dual Bregman-Proximal Splitting (PDBS)

$$\begin{gathered} DJ_X\left(x^k\right) - D_x K\left(x^k, y^k\right) \in DJ_X\left(x^{k+1}\right) + \partial F\left(x^{k+1}\right) \quad \text{and} \\ DJ_Y\left(y^k\right) + \left[2D_y K\left(x^{k+1}, y^k\right) + D_y K\left(x^k, y^k\right) - 2D_y\left(x^k, y^{k-1}\right)\right] \\ \in DJ_Y\left(y^{k+1}\right) + \partial G_*\left(y^{k+1}\right) \end{gathered} \tag{26}$$

The method reduces to the basic PDBS (24) when K is affine. Similarly, the PDBS method can be arranged as the PDPS method

Modified Primal–Dual Proximal Splitting (PDPS)

$$\begin{aligned} x^{k+1} =& \operatorname{prox}_{\mathrm{r}F}\left(x^k - \tau \nabla_x K\left(x^k, y^k\right)\right) \\ y^{k+1} =& \operatorname{prox}_{\sigma G_*}\left(y^k + \sigma\left[2\nabla_y K\left(x^{k+1}, y^k\right) + \right.\right. \\ & \left.\left. \nabla_y K\left(x^k, y^k\right) - 2\nabla_y K\left(x^k, y^{k-1}\right)\right]\right) \end{aligned} \tag{27}$$

The convergence is presented by T. Valkonen in this related article: First-Order Primal–Dual Methods for Nonsmooth Non-convex Optimisation. For more details including convergence, acceleration, and alternative methods, one can check the original article by T. Valkonen.

7 Conclusion

The primal-dual method illustrates a new way to solve the primal problem which balances the data fidelity with regularization. By updating the primal variable and the dual variable, the method approximates the optimal point of the saddle point problem when alternatively iterating. Many algorithms for application derive from the primal-dual method, including PDHG, PDPS, PDBS, and so on. Although they are originally proposed to solve convex problems, we offer some thoughts on applying them to the non-convex case and demonstrate concrete non-convex methods by T. Valkonen based on PDPS and PDBS. By virtue of its efficiency and compatibility, the primal-dual method is widely adopted in image restoration, inpainting, and reconstruction, producing profound impact in image modeling.