J Optim Theory Appl (2018) 177:413–447
https://doi.org/10.1007/s10957-018-1265-x

Lagrangian Penalization Scheme with Parallel Forward–Backward Splitting

Cesare Molinari[1,2] · **Juan Peypouquet**[3]

Received: 11 September 2017 / Accepted: 16 March 2018 / Published online: 22 March 2018

Abstract We propose a new iterative algorithm for the numerical approximation of the solutions to convex optimization problems and constrained variational inequalities, especially when the functions and operators involved have a separable structure on a product space, and exhibit some dissymmetry in terms of their component-wise regularity. Our method combines Lagrangian techniques and a penalization scheme with bounded parameters, with parallel forward–backward iterations. Conveniently combined, these techniques allow us to take advantage of the particular structure of the problem. We prove the weak convergence of the sequence generated by this scheme, along with worst-case convergence rates in the convex optimization setting, and for the strongly non-degenerate monotone operator case. Implementation issues related to the penalization of the constraint set are discussed, as well as applications in image recovery and non-Newtonian fluids modeling. A numerical illustration is also given, in order to prove the performance of the algorithm.

Keywords Convex programming · Forward–backward · Lagrange multipliers · Penalization

✉ Juan Peypouquet
jpeypou@dim.uchile.cl

Cesare Molinari
cecio.molinari@gmail.com

1 Departamento de Matemática, Universidad Técnica Federico Santa María, Av España 1680, Valparaiso, Chile

2 GREYC CNRS UMR 6072, ENSICAEN, Normandie Université, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France

3 Departamento de Ingeniería Matemática & Centro de Modelamiento Matemático (CNRS UMI2807), FCFM, Universidad de Chile, Beauchef 851, Santiago, Chile

Mathematics Subject Classification 49M37 · 90C25

1 Introduction

This paper is concerned with the numerical approximation of the solutions to convex optimization problems and constrained variational inequalities in Hilbert spaces, especially when the functions and operators involved have a separable structure and exhibit some dissymmetry in terms of their regularity. For instance, in the optimization setting, we are interested in the minimization of the sum of two convex functions over a linear-convex constraint, where a function is differentiable and the other is just lower-semicontinuous. Such problems arise frequently in economics, image processing, statistics, management, inverse problems, optimal control, partial differential equations, among others (see Sect. 4.2 for more details and references).

In convex optimization algorithms, constraints are frequently dealt with in a purely primal fashion by means of projections (as in the projected gradient method [1,2]), or by introducing penalties or barriers [3,4], the latter giving rise to many interior-point methods. Another way consists in using Lagrange multiplier theory and approximating both the primal variables and the multipliers, which are interpreted as dual variables. These methods are related to the (possibly augmented) Lagrangian function, and many are more or less sophisticated variants of the alternating direction method of multipliers [5,6]. Further details can be found in the classical references [7–10], and also [11]. For structured problems on product spaces, where the objective function has an additively separable structure (as the ones considered here), there exist sequential splitting methods, with [12,13] or without [14,15] linearization, and parallel methods [16]. For a comprehensive review of now classical iterative algorithms in continuous optimization, the reader may consult [17]. Finally, there are some methods (for instance, [18–21]) that combine the penalization and Lagrangian approaches, and our algorithm follows this philosophy. As in many first-order methods, we use gradient [22] and proximal [23–25] iterations, and their generalization to the monotone operator context, as building blocks.

The main purpose of this work is to present a methodology where several techniques, when conveniently combined, allow us to take advantage of the particular structure of the problem, and produce convergent methods that are applicable in diverse situations. As mentioned above, we are concerned with the case where the functions and operators involved have a separable structure on a product space and exhibit some dissymmetry in terms of their component-wise regularity. An observation of the state variable, given by the image of a linear operator, is required to lie in a given closed convex set. Our algorithm is based on the following ideas:

– *Parallelized forward–backward primal iterations.* The primal variables (one on each factor space) are treated in parallel, independently of each other, and of the *module* corresponding to the dual variables. The corresponding sub-iterations are of a different nature in order to capture and exploit the underlying dissymmetry. Roughly speaking, we use a backward step on the non-regular part and a forward step on the regular one, which reduces the iteration complexity. To our knowledge, the existing algorithms that handle constraints and split the primal variables always treat the latter in a symmetric way. We propose here a parallel, yet asymmetric approach, in order to keep the benefits of parallelization while taking advantage of the *partial* regularity of the problem.

– *Lagrangian penalization approach.* The linear part of the constraint is treated in a classical fashion introducing a Lagrange multiplier and performing a prediction-correction sub-iterations on this dual variable. For the nonlinear part of the constraint —corresponding to the feasible set—we use a general exterior penalization function. However, the penalization parameter is updated as a Lagrange multiplier, so this gives a Lagrangian approach to deal with nonlinear constraints. The main consequence of this device is that the parameter does not tend to $+\infty$ as the algorithm evolves, as occurs typically in penalization schemes, and limits the possible inherent numerical stabilities.

– *Relaxation.* The multiplier steps account for under- or over-relaxation that can favor the progress of either the primal or the dual variables, which results in a faster or more convenient evolution of the algorithm.

The paper is organized as follows: In Sect. 2, we describe the convex optimization setting. The primal-dual sequence is proved to converge weakly to a saddle point of the Lagrangian and the function evaluations are shown to converge to the optimal value. Despite the Lagrangian penalization character, the worst-case convergence rates are consistent with those for descent methods. An inertial interpretation, which can be used to design an accelerated method, is also discussed. Section 3 deals with the more general maximally monotone framework. Similar results are obtained except that a variational description replaces the Lagrangian optimality conditions. Sharper results are obtained in the strongly non-degenerate case, where the iterates converge linearly, as expected. Some illustrations, including implementation and applicability issues, are given in Sect. 4. Section 6 contains conclusions and possible lines of future development. To simplify the reading, the most technical aspects are gathered in the appendices.

2 Potential Setting

Let X, Y, Z be real Hilbert spaces and let $f : X \to \mathbb{R}\cup\{+\infty\}$ and $g : Y \to \mathbb{R}\cup\{+\infty\}$ be proper, lower-semicontinuous and convex. Suppose $A : X \to Z$ and $B : Y \to Z$ are bounded linear operators, and consider a non-empty, closed and convex set $C \subset Z$. We aim at solving the structured convex minimization problem

$$\text{Find} \quad (\bar{x}, \bar{y}) \in \underset{(x,y)\in X\times Y}{\arg\min} \{f(x) + g(y) : Ax + By \in C\}. \tag{1}$$

The function $f+g$ is minimized subject to the constraint that an observation $Ax+By$ of the decision variable (x, y) belongs to the set C. We shall assume that f is differentiable and ∇f is Lipschitz continuous with constant $L_{\nabla f}$, while g is arbitrary. This is where the dissymmetry comes into play, and the proposed method is intended to capture and exploit this fact.

Following [19], we introduce a convex and L_p-Lipschitz continuous penalization function $p :\ Z \to \mathbb{R}_+$ with

$$C = p^{-1}(0) = \arg\min\{p(z) : z \in Z\}.$$

Problem (1) can be written as

$$(\bar{x}, \bar{y}, \bar{z}) \in \underset{(x,y,z)\in X\times Y\times Z}{\arg\min} \{f(x) + g(y) : Ax + By = z \text{ and } p(z) = 0\}. \tag{2}$$

The associated Lagrangian $\mathcal{L} : X \times Y \times Z \times Z \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\mathcal{L}(x, y, z, \mu, \nu) = f(x) + g(y) + \langle \mu, Ax + By - z\rangle + \nu\, p(z). \tag{3}$$

Under qualification conditions (Moreau–Rockafellar, Attouch–Brézis), $(\bar{x}, \bar{y}, \bar{z}) \in X \times Y \times Z$ is a solution for problem (2) if, and only if, there exists $(\bar{\mu}, \bar{\nu})$ in $Z \times \mathbb{R}$ such that $(\bar{x}, \bar{y}, \bar{z}, \bar{\mu}, \bar{\nu})$ is a saddle point for the Lagrangian (see, for example, [26]), which means that

$$\mathcal{L}(\bar{x}, \bar{y}, \bar{z}, \mu, \nu) \le \mathcal{L}(\bar{x}, \bar{y}, \bar{z}, \bar{\mu}, \bar{\nu}) \le \mathcal{L}(x, y, z, \bar{\mu}, \bar{\nu})$$

for every $(x, y, z) \in X \times Y \times Z$ and for every $(\mu, \nu) \in Z \times \mathbb{R}$. We denote by $\mathcal{S}$ the set of such saddle points and, from now on, we suppose it is not empty.

2.1 Proposed Algorithm and Main Result

We propose an iterative algorithm for the numerical solution of (2). Given $\varepsilon \ge 0$, the ε-subdifferential of a function $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\partial_\varepsilon \Phi(u) = \left\{u^* \in H : \Phi(v) + \varepsilon \ge \Phi(u) + \langle u^*, v - u\rangle \quad \forall v \in H\right\}.$$

Take a positive sequence (λ_k) of step sizes, two nonnegative sequences (α_k) and (β_k), to be interpreted as computational error tolerances for the proximal sub-iterations, and two positive relaxation parameters γ and δ. Given an initial point $w_0 := (x_0, y_0, z_0, \mu_0, \nu_0) \in X \times Y \times Z \times Z \times \mathbb{R}_+$.

Lagrangian Penalization Splitting Algorithm (LaPSA).
For each $k \ge 0$, we do as follows:
Multiplier prediction:

(A1) Write $\tilde{\mu}_{k+1} = \mu_k + \gamma\lambda_k (Ax_k + By_k - z_k)$;
(A2) Write $\tilde{\nu}_{k+1} = \nu_k + \delta\lambda_k p(z_k)$.

Primal variable update:

(B1) Write $x_{k+1} = x_k - \lambda_k \left[A^*\tilde{\mu}_{k+1} + \nabla f(x_k)\right]$;
(B2) Set $\tilde{y}_{k+1} = y_k - \lambda_k B^*\tilde{\mu}_{k+1}$ and compute y_{k+1} such that

$$-\frac{y_{k+1} - \tilde{y}_{k+1}}{\lambda_k} \in \partial_{\alpha_k} g(y_{k+1});$$

(B3) Set $\tilde{z}_{k+1} = z_k + \lambda_k \tilde{\mu}_{k+1}$ and compute

$$-\frac{z_{k+1} - \tilde{z}_{k+1}}{\lambda_k \tilde{\nu}_{k+1}} \in \partial_{\beta_k} p\left(z_{k+1}\right);$$

Multiplier correction:

(C1) Write $\mu_{k+1} = \mu_k + \gamma\lambda_k\left(Ax_{k+1} + By_{k+1} - z_{k+1}\right)$;
(C2) Write $\nu_{k+1} = \nu_k + \delta\lambda_k p\left(z_{k+1}\right)$.

Despite the similarities between LaPSA and the methods studied in [12,19], the former takes advantage of the separable and "smooth + non-smooth" structure. Steps (B2) and (B3) are the only possibly computationally expensive ones. Nevertheless, they are performed in parallel, and independently of (B1). Moreover, there are many relevant instances where these proximal steps can be explicitly and cheaply computed. For instance, when ℓ^1 norms are involved. We shall come back to this point in Sect. 4.2. Finally, notice that the sequence (ν_k) is non-decreasing.

Definition 2.1 To simplify the notation, we write $u = (x, y, z)$, $v = (\mu, \nu)$ and $w = (u, v)$. We define the scalar products in $U = X \times Y \times Z$, $V = Z \times \mathbb{R}$ and $W = U \times V$, respectively, by

$$\langle u,\ u'\rangle = \langle x,\ x'\rangle + \langle y,\ y'\rangle + \langle z,\ z'\rangle,$$
$$\langle v,\ v'\rangle = \frac{1}{\delta}\langle \mu,\ \mu'\rangle + \frac{1}{\gamma}\nu\nu' \quad \text{and} \quad \langle w,\ w'\rangle = \langle u,\ u'\rangle + \langle v,\ v'\rangle.$$

The induced norms are given by $\|u\|^2 = \|x\|^2 + \|y\|^2 + \|z\|^2$, $\|v\|^2 = \frac{1}{\delta}\|\mu\|^2 + \frac{1}{\gamma}|\nu|^2$ and $\|w\|^2 = \|u\|^2 + \|v\|^2$.

With this notation, steps (B1)–(B3) can be written more succinctly as

$$u_{k+1} \in \operatorname*{arg\,min}_{u\in U}\left\{\mathcal{L}_k\left(u, \tilde{v}_{k+1}\right) + \frac{1}{2\lambda}\|u - u_k\|^2\right\},$$

where

$$\mathcal{L}_k\left(w\right) := f(x_k) + \langle \nabla f\left(x_k\right),\ x - x_k\rangle + g(y) + \langle \mu, Ax + By - z\rangle + \nu\ p\left(z\right).$$

When combined, the multiplier prediction and correction steps reveal an inertial aspect of this algorithm. This idea is further explained in Sect. 2.5, where some possible ideas for future research are outlined.

The convergence of this algorithm will be granted under the following hypothesis:

Hypothesis (H):

(H1) The sequences (α_k) and (β_k) belong to ℓ^1.
(H2) The parameters are chosen in the following manner:
  (i) First, fix $0 < \lambda \leq \Lambda < \frac{1}{L_{\nabla f}}$, and pick $\lambda_k \in [\lambda, \Lambda]$ for all (sufficiently large) k;

(ii) Next, set $\Gamma = \frac{1}{3\Lambda^2} \min \left\{ 1, \frac{1-\Lambda L_{\nabla f}}{\|A\|^2}, \frac{1}{\|B\|^2} \right\} > 0$, and pick $\gamma \in (0, \Gamma)$.

(iii) Finally, set $\Delta = \frac{1}{\Lambda L_p} \sqrt{1 - 3\gamma\Lambda^2} > 0$ and pick $\delta \in (0, \Delta)$.

The main result of this section is the following.

Theorem 2.1 *Under Hypothesis* **(H)***, for every starting point $w_0 \in X \times Y \times Z \times Z \times \mathbb{R}_+$, the sequence generated by LaPSA converges weakly to a saddle point of the Lagrangian.*

2.2 Convergence Analysis

The convergence analysis relies on the following estimation:

Lemma 2.1 *Let (w_k) be generated by LaPSA with $0 < \lambda_k \leq \frac{1}{L_{\nabla f}}$ and let $\bar{w} \in \mathcal{S}$. Then, for every $k \geq 0$, we have*

$$\|w_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + 2\lambda_k \left[\mathcal{L}(u_{k+1}, \bar{v}) - \mathcal{L}(\bar{u}, \bar{v}) \right]$$
$$+ E_k \leq 2\lambda_k (\alpha_k + \beta_k) \leq \frac{2}{L_{\nabla f}} (\alpha_k + \beta_k),$$

where

$$E_k = \left(1 - L_{\nabla f}\lambda_k - 3\gamma\lambda_k^2\|A\|^2\right) \|x_{k+1} - x_k\|^2 + \left(1 - 3\gamma\lambda_k^2\|B\|^2\right) \|y_{k+1} - y_k\|^2$$
$$+ \left(1 - \lambda_k^2 \left(3\gamma + L_p^2\delta\right)\right) \|z_{k+1} - z_k\|^2 + \frac{1}{\gamma} \|\tilde{\mu}_{k+1} - \mu_k\|^2 + \frac{1}{\delta} |\tilde{\nu}_{k+1} - \nu_k|^2.$$

Moreover, if Hypothesis **(H2)** *holds, there is $\sigma > 0$ such that*

$$E_k \geq \sigma \|u_{k+1} - u_k\|^2 + \|\tilde{v}_{k+1} - v_k\|^2.$$

The proof of this lemma is technical and given in "Appendix B". It can be omitted in a first reading. As a consequence of Lemma 2.1, we have the following:

Proposition 2.1 *Let (w_k) be generated by LaPSA and assume Hypothesis* **(H)** *holds. We have the following:*

(i) *for every $\bar{w}$ in $\mathcal{S}$, $\lim_k \|w_k - \bar{w}\|$ exists finite;*
(ii) *the sequences $\left(\|u_{k+1} - u_k\|^2\right)$, $\left(\|\tilde{v}_{k+1} - v_k\|^2\right)$ and $\left(\mathcal{L}(u_k, \bar{v}) - \mathcal{L}(\bar{u}, \bar{v})\right)$ belong to ℓ^1;*
(iii) *the sequence $\left(Ax_k + By_k - z_k, p(z_k)\right)$ converges (strongly) to $0 \in Z \times \mathbb{R}_+$; and*
(iv) *the sequence $\left(\mathcal{L}(w_k)\right)$ converges to $\mathcal{L}(\bar{w})$.*

Proof By the saddle point property, $\mathcal{L}(\bar{u}, \bar{v}) \leq \mathcal{L}(u_{k+1}, \bar{v})$, so items (i) and (ii) are consequences of Lemmas 2.1 and A.1. In particular, the sequences $\tilde{\mu}_{k+1} - \mu_k$ and $\tilde{\nu}_{k+1} - \nu_k$ converge (strongly) to zero. Using the definition of $\tilde{\mu}_{k+1}, \tilde{\nu}_{k+1}$ and the boundedness of λ_k, we obtain (iii). Finally, $\mathcal{L}(w_k)$ converges to $\mathcal{L}(\bar{w})$ in view of items (ii) and (iii), since μ_k and ν_k are bounded. □

We are now in a position to prove the main results of this section.

Proof of Theorem 2.1. We shall use Opial's Lemma A.2 with $\omega_k = w_k$ and $F = \mathcal{S}$. Item (i) in Proposition 2.1 gives the first condition in Lemma A.2. It remains to show that the second one is satisfied. In other words, that every weak cluster point of the sequence (w_k) belongs to $\mathcal{S}$. Let w_{k_j} converge weakly, as $j \to \infty$, to $w^\infty \in X \times Y \times Z \times Z \times \mathbb{R}_+$. First, since $u_{k_j} \rightharpoonup u^\infty$,

$$Ax_{k_j} + By_{k_j} - z_{k_j} \rightharpoonup Ax^\infty + By^\infty - z^\infty.$$

But $Ax_k + By_k - z_k$ converges strongly to zero, by Proposition 2.1. It follows that

$$Ax^\infty + By^\infty - z^\infty = 0.$$

Next, since $z_{k_j} \rightharpoonup z^\infty$, we have

$$0 \leq p\left(z^\infty\right) \leq \liminf_j p\left(z_{k_j}\right) = \lim_k p\left(z_k\right) = 0.$$

Hence, the point $(x^\infty, y^\infty, z^\infty)$ satisfies the constraints. So for every v in V we have that

$$\mathcal{L}\left(u^\infty, v\right) = \mathcal{L}\left(u^\infty, v^\infty\right), \tag{4}$$

which implies the first inequality of the saddle point property.

Let us prove the second inequality. Using the (sub)gradient inequality and simple algebraic manipulations (the same used to obtain inequalities (b1), (b2) and (b3) in the proof of Lemma 2.1), we obtain

$$\begin{aligned}
2\lambda_k\left[f(x_{k+1}) - f(x) + \langle\tilde{\mu}_{k+1}, A(x_{k+1} - x)\rangle\right] &\leq \|x_k - x\|^2 - \|x_{k+1} - x\|^2 \\
&\quad - \left(1 - L_{\nabla f}\lambda_k\right)\|x_{k+1} - x_k\|^2 \\
2\lambda_k\left[g(y_{k+1}) - g(y) + \langle\tilde{\mu}_{k+1}, B(y_{k+1} - y)\rangle\right] &\leq \|y_k - y\|^2 - \|y_{k+1} - y\|^2 \\
&\quad - \|y_{k+1} - y_k\|^2 + 2\lambda_k\alpha_k \\
2\lambda_k\left[\tilde{v}_{k+1}\left[p(z_{k+1}) - p(z)\right] + \langle\tilde{\mu}_{k+1}, -(z_{k+1} - z)\right] &\leq \|z_k - z\|^2 - \|z_{k+1} - z\|^2 \\
&\quad - \|z_{k+1} - z_k\|^2 + 2\lambda_k\beta_k
\end{aligned}$$

for each $(x, y, z) \in X \times Y \times Z$. Adding the three inequalities, and using the hypothesis on the step size λ_k to estimate the right-hand side, we obtain

$$2\lambda_k\left[\mathcal{L}(u_{k+1}, \tilde{v}_{k+1}) - \mathcal{L}(u, \tilde{v}_{k+1})\right] \leq \|u_k - u\|^2 - \|u_{k+1} - u\|^2 + 2\lambda_k(\alpha_k + \beta_k).$$

Summing for $k = 0, \ldots, K$, we have

$$2\lambda\sum_{k=0}^{K}\left[\mathcal{L}(u_{k+1}, \tilde{v}_{k+1}) - \mathcal{L}(u, \tilde{v}_{k+1})\right] \leq \|u_0 - u\|^2 + 2\Lambda\sum_{k=1}^{\infty}(\alpha_k + \beta_k) < \infty.$$

It follows that

$$\liminf_k \left[\mathcal{L}\left(u_{k+1}, \tilde{v}_{k+1}\right) - \mathcal{L}\left(u, \tilde{v}_{k+1}\right)\right] \leq 0.$$

But items (ii) and (iv) of Proposition 2.1 imply that $\lim_k \ \mathcal{L}(u_{k+1}, \tilde{v}_{k+1})$ exists. We also know that $\tilde{v}_{k_j+1} \rightharpoonup v^\infty$, because $v_{k_j} \rightharpoonup v^\infty$ and $\|\tilde{v}_{k+1} - v_k\| \to 0$ (again by Proposition 2.1). Then, by the definition of $\mathcal{L}$, we have

$$\lim_j \mathcal{L}\left(u, \tilde{v}_{k_j+1}\right) = \mathcal{L}\left(u, v^\infty\right).$$

We deduce that

$$\lim_j \mathcal{L}\left(u_{k_j+1}, \tilde{v}_{k_j+1}\right) \leq \mathcal{L}\left(u, v^\infty\right). \tag{5}$$

By the weak lower semicontinuity of f and g, and using that $Ax^\infty + By^\infty = z^\infty$ and $p\left(z^\infty\right) = 0$, we get

$$\begin{aligned}\mathcal{L}\left(u^\infty, v^\infty\right) &\leq \liminf_j f\left(x_{k_j+1}\right) + \liminf_j g\left(y_{k_j+1}\right)\\ &\leq \liminf_j \left[f\left(x_{k_j+1}\right) + g\left(y_{k_j+1}\right)\right]\\ &= \liminf_j \mathcal{L}\left(u_{k_j+1}, \tilde{v}_{k_j+1}\right),\end{aligned} \tag{6}$$

since $\langle\tilde{\mu}_{k_j+1}, Ax_{k_j} + By_{k_j} - a_{k_j}\rangle \to 0$ and $\tilde{\nu}_{k_j+1}\ p\left(z_{k_j}\right) \to 0$. Finally, combining (5) and (6), we conclude that

$$\mathcal{L}\left(u^\infty, v^\infty\right) \leq \mathcal{L}\left(u, v^\infty\right). \tag{7}$$

In view of (4) and (7), w^∞ is a saddle point for the Lagrangian. □

2.3 Convergence Rate

Let $\bar{w}$ be any saddle point for the Lagrangian $\mathcal{L}$. Define $D^{\bar{w}} : U \to \mathbb{R}$ by

$$\begin{aligned}D^{\bar{w}}(u) = \mathcal{L}\left(u, \bar{v}\right) - \mathcal{L}\left(\bar{u}, \bar{v}\right) &= \left[\left(f\left(x\right) - f\left(\bar{x}\right)\right) + \left(g\left(x\right) - g\left(\bar{x}\right)\right)\right]\\ &\quad + \langle\bar{\mu}, Ax + By - z\rangle + \bar{\nu} p\left(z\right).\end{aligned} \tag{8}$$

It is easy to see that $D^{\bar{w}}(u) \geq 0$ for every u in U. Moreover, $\hat{u}$ is a (primal) solution if, and only if, $\hat{u}$ is feasible and $D^{\bar{w}}(\hat{u}) = 0$ for some (thus all) $\bar{w} \in \mathcal{S}$. We interpret the quantity $D^{\bar{w}}(u)$ as a measure of the quality of an approximate solution u. The following result establishes that the worst-case convergence rate for LaPSA matches that of first-order descent methods. It is surprising to observe this behavior in the absence of the descent property.

Theorem 2.2 *Let* (w_k) *be generated by LaPSA under Hypotheses* **(H1)–(H2)***, and set* $\bar{w} \in \mathcal{S}$*. For each k, set* $\sigma_k = \sum_{n=0}^{k-1} \lambda_n$*, and define the average up to the kth primal iterate as*

$$\check{u}_k = \frac{1}{\sigma_k} \sum_{n=0}^{k-1} \lambda_n u_{n+1}.$$

Then,

$$D^{\bar{w}}\left(\check{u}_k\right) \leq \frac{\bar{C}}{\sigma_k}, \qquad \textit{where} \qquad \bar{C} = \frac{1}{2}\|w_0 - \bar{w}\|^2 + \sum_{n \geq 0} \lambda_n\left(\alpha_n + \beta_n\right) < +\infty.$$

In particular, $D^{\bar{w}}\left(\check{u}_k\right) = \mathcal{O}(k^{-1})$*. For the best primal iterate, we have* $\min_{1 \leq n \leq k} D^{\bar{w}}(u_n) = o(k^{-1})$*.*

Proof From Lemma 2.1 and Hypothesis **(H1)–(H2)**, we have

$$2\lambda_k D^{\bar{w}}(u_{k+1}) \leq \|w_k - \bar{w}\|^2 - \|w_{k+1} - \bar{w}\|^2 + 2\lambda_k\left(\alpha_k + \beta_k\right).$$

Summing up, and using the telescopic property on the right-hand side, we obtain

$$\sum_{n=0}^{k-1} \lambda_n D^{\bar{w}}(u_{n+1}) \leq \bar{C},$$

for all k. Dividing by σ_k and using the convexity of $D^{\bar{w}}$, we deduce that $D^{\bar{w}}\left(\check{u}_k\right) \leq \bar{C}\sigma_k^{-1} = O(k^{-1})$. For the last part, set $a_k = \min_{1 \leq n \leq k} D^{\bar{w}}(u_n)$. The sequence (a_k) is nonnegative, non-increasing, and summable, in view of

$$\lambda \sum_{n=0}^{k} a_n \leq \sum_{n=0}^{k} \lambda_n D^{\bar{w}}(u_{n+1}) \leq \bar{C}.$$

By Lemma A.3, we have $a_k = o(k^{-1})$. □

2.4 An Alternative Error Criterion

The inexact computation of the iterates may be defined in terms of approximate minimizers, instead of approximate subdifferentials. Given $\varepsilon \geq 0$ and $\Phi : H \to \mathbb{R} \cup \{+\infty\}$, define

$$\varepsilon - \operatorname*{arg\,min}_{h \in H} \Phi(h) = \left\{h \in H : \Phi(h) \leq \inf_H \Phi + \varepsilon\right\}.$$

In the same setting as before, for each $k = 0, 1, \ldots,$ we do as follows:

LaPSA*.

Steps (A1), (A2), (B1), (C1) and (C2) are the same as in LaPSA, but we replace (B2) and (B3) by

(B2*) Set $\tilde{y}_{k+1} = y_k - \lambda_k B^* \tilde{\mu}_{k+1}$ and compute y_{k+1} such that

$$y_{k+1} \in \alpha_k - \operatorname*{arg\,min}_{y \in Y} \left\{ g(y) + \frac{1}{2\lambda_k} \|y - \tilde{y}_{k+1}\|^2 \right\}, \text{ and}$$

(B3*) Set $\tilde{z}_{k+1} = z_k + \lambda_k \tilde{\mu}_{k+1}$ and compute

$$z_{k+1} \in \beta_k - \operatorname*{arg\,min}_{z \in Z} \left\{ p\,(z) + \frac{1}{2\lambda_k \tilde{\nu}_{k+1}} \|z - \tilde{z}_{k+1}\|^2 \right\},$$

respectively. In turn, state Hypothesis **(H*)** by replacing (H1) in Hypothesis **(H)** by

(H1*) The sequences (α_k) and (β_k) belong to $\ell^{\frac{1}{2}}$.

Remark 1 Conditions (B2*) and (B3*) are weaker than (B2) and (B3), respectively. On the other hand, Hypothesis (H1*) for the errors is stronger than (H1).

We have essentially the same convergence result, namely:

Theorem 2.3 *Under Hypothesis* **(H*)***, for every starting point* $w_0 \in X \times Y \times Z \times Z \times \mathbb{R}_+$*, the sequence generated by LaPSA* converges weakly to a saddle point of the Lagrangian.*

The proof is similar to that of Theorem 2.1 and is given in "Appendix C".

2.5 Inertial Interpretation and Acceleration

The multiplier prediction steps can be combined with the correction steps of the previous iteration using the variable $(\tilde{u}, \tilde{v})$ and introducing auxiliary primary variables in a sort of inertial or extrapolation step. More precisely, one may set

$$\begin{aligned} \tilde{u}_k &= u_k + \theta_k (u_k - u_{k-1}), \\ \tilde{p}_k &= p(z_k) + \theta_k (p(z_k) - p(z_{k-1})), \end{aligned}$$

with $\theta_k = \frac{\lambda_{k-1}}{\lambda_k}$ and then perform the dual update

$$\begin{aligned} \tilde{\mu}_{k+1} &= \tilde{\mu}_k + \gamma \lambda_k \left(A\tilde{x}_k + B\tilde{y}_k - \tilde{z}_k \right), \\ \tilde{\nu}_{k+1} &= \tilde{\nu}_k + \delta \lambda_k \tilde{p}_k, \end{aligned}$$

and the primal one

$$u_{k+1} \in \operatorname*{arg\,min}_{u \in U} \left\{ \mathcal{L}_k \left(u, \tilde{\nu}_{k+1} \right) + \frac{1}{2\lambda} \|u - u_k\|^2 \right\},$$

where

$$\mathcal{L}_k(w) := f(x_k) + \langle \nabla f(x_k),\ x - x_k\rangle + g(y) + \langle \mu, Ax + By - z\rangle + \nu\, p(z)\,.$$

The first step is not exactly an inertial step on the dual variables, but an inertial step on the primal variables for the update of the dual ones. This suggests a new scheme with an inertial update for the primal variables, namely

$$u_{k+1} \in \operatorname*{arg\,min}_{u\in U} \left\{ \tilde{\mathcal{L}}_k(u, \tilde{v}_{k+1}) + \frac{1}{2\lambda}\|u - \tilde{u}_k\|^2 \right\},$$

where

$$\tilde{\mathcal{L}}_k(w) := f(\tilde{x}_k) + \langle \nabla f(\tilde{x}_k),\ x - \tilde{x}_k\rangle + g(y) + \langle \mu, Ax + By - z\rangle + \nu\, p(z)\,.$$

On the one hand, the resulting algorithm would have the same computational complexity of the one presented here. On the other hand, it would open the possibility to accelerate the method, following the ideas of Nesterov [27]. To this end, it is reasonable to prescribe the step size selection rule

$$\lambda_k = \frac{k}{k-\alpha}\lambda_{k-1}$$

for some $\alpha > 3$ all sufficiently large k, which would give $\theta_k = 1 - \frac{\alpha}{k}$. This choice for the extrapolation is known to preserve the convergence properties of the proximal gradient method, while improving the worst-case convergence rate (see [28–30]). In view of its complexity, we shall not explore this line of research in this paper.

3 Maximal Monotone Operator Setting

With the same notation of the previous section, we focus now on solving the variational inequality

$$\begin{aligned}&\text{Find}(\bar{x}, \bar{y}, \bar{z}) \in X \times Y \times Z \text{ such that } \quad (0, 0, \bar{z}) \in \big(M\bar{x} + A^*\bar{z}, N\bar{y}\\ &\qquad + B^*\bar{z}, \mathcal{N}_C(A\bar{x} + B\bar{y})\big),\end{aligned} \tag{9}$$

where $M : X \rightrightarrows X$ and $N : Y \rightrightarrows Y$ are maximally monotone operators and $\mathcal{N}_C$ is the normal cone of convex analysis, namely $\mathcal{N}_C(\bar{z}) := \{z \in Z :\ \langle z, z - c\rangle \leq 0\ \forall c \in C\}$ if $\bar{z} \in C$ and the empty set otherwise. Observe that, if $M = \nabla f$ and $N = \partial g$ (the setting of Sect. 2), then every solution of (9) is a solution of (1). By analogy, we shall also consider a lack of symmetry and assume that the operator M is cocoercive with constant θ (in particular, it is single-valued), while N is arbitrary. This will allow us to design a specific algorithm for the numerical approximation of the solutions of (9), *with an explicit step on the x variable*.

The spaces U, V and W were introduced in Definition 2.1. Define the monotone operator $\mathcal{M} : W \rightrightarrows W$ by

$$\mathcal{M}(w) = \big(Mx + A^*\mu, Ny + B^*\mu, \partial^z[\nu p(z)] - \mu, z - (Ax + By), -p(z)\big). \tag{10}$$

Remark 2 Notice that the domain of $\mathcal{M}$ is contained in $X \times Y \times Z \times Z \times \mathbb{R}_+$. Indeed, as p is convex and nonnegative on Z, for $\nu < 0$ we have $\partial^z[\nu p(z)] = \emptyset$ for all $z \in Z$.

Remark 3 In the setting of Sect. 2, we have $\mathcal{M} = \big(\partial_x \mathcal{L}, \partial_y \mathcal{L}, \partial_z \mathcal{L}, \partial_\mu(-\mathcal{L}), \partial_\nu(-\mathcal{L})\big)$.

We assume (9) is well posed, in the sense that there exists

$$(\bar{x}, \bar{y}, \bar{z}, \bar{\mu}, \bar{\nu}) \in \mathcal{S}_{\mathcal{M}} := \mathcal{M}^{-1}(0).$$

In other words, that $(\bar{x}, \bar{y}, \bar{z}, \bar{\mu}, \bar{\nu})$ satisfies

$$\begin{aligned} -A^*\bar{\mu} = M\bar{x}; \quad -B^*\bar{\mu} \in N\bar{y}; \quad \bar{\mu} \in \partial^z[\bar{\nu} p(\bar{z})]; \\ A\bar{x} + B\bar{y} = \bar{z}; \quad \text{and} \quad p(\bar{z}) = 0. \end{aligned} \tag{11}$$

Since $\bar{\nu} \in \mathbb{R}_+$, we have

$$\partial^z[\bar{\nu} p(\bar{z})] = \bar{\nu}\partial p(z) \subseteq \mathcal{N}_C(z)$$

for every $z \in C$. In particular, as $\bar{z} \in C$ by (11), the corresponding pair $(\bar{x}, \bar{y})$ is a solution for (9).

3.1 Proposed Algorithm and Main Result

We propose and study an iterative algorithm for the numerical approximation of solutions for (11). As in the previous section, we allow the implicit steps to be computed with some approximation errors, given by the positive sequences α_k and β_k. To this purpose, for $\varepsilon > 0$ we recall the definition of ε-enlargement of a maximally monotone operator $\mathcal{A} : H \rightrightarrows H$ at a point $u \in H$:

$$\mathcal{A}_\varepsilon(u) = \left\{u^* \in H : \langle u^* - v^*, u - v\rangle \geq -\varepsilon \quad \forall v \in H \text{ and } v^* \in \mathcal{A}(v)\right\}.$$

Observe that the standard ε-approximate subdifferential of a function $F : H \to \mathbb{R} \cup \{+\infty\}$ satisfies $\partial_\varepsilon F \subseteq (\partial F)_\varepsilon$ (see, for instance, [31] and [32]).

Take a positive sequence λ_k of step sizes, and two relaxation parameters $\gamma > 0$ and $\delta > 0$. Given an initial point $w_0 := (x_0, y_0, z_0, \mu_0, \nu_0) \in X \times Y \times Z \times Z \times R_+$.

LaPSA':

For each $k \geq 0$, we do as follows:
Multiplier prediction:

(A1) Write $\tilde{\mu}_{k+1} = \mu_k + \gamma\lambda_k(Ax_k + By_k - z_k)$;
(A2) Write $\tilde{\nu}_{k+1} = \nu_k + \delta\lambda_k p(z_k)$.

Primal variable update:

(B1') Write $x_{k+1} = x_k - \lambda_k \left[A^* \tilde{\mu}_{k+1} + M(x_k)\right]$;

(B2') Compute y_{k+1} satisfying

$$-\frac{y_{k+1} - y_k}{\lambda_k} - B^* \tilde{\mu}_{k+1} \in N_{\alpha_k}(y_{k+1});$$

(B3') Compute z_{k+1} satisfying

$$-\frac{z_{k+1} - z_k}{\lambda_k} + \tilde{\mu}_{k+1} \in \tilde{\nu}_{k+1}\, \partial_{\beta_k} p(z_{k+1}).$$

Multiplier correction:

(C1) Write $\mu_{k+1} = \mu_k + \gamma \lambda_k (A x_{k+1} + B y_{k+1} - z_{k+1})$;

(C2) Write $\nu_{k+1} = \nu_k + \delta \lambda_k p(z_{k+1})$.

We shall assume the following hypothesis hold:

Hypothesis (H'):

(H1) The sequences (α_k) and (β_k) belong to ℓ^1;

(H2') The parameters are chosen in the following manner:

(i) First, fix $0 < \lambda \leq \Lambda < 2\theta$, and pick $\lambda_k \in [\lambda, \Lambda]$ for all (sufficiently large) k;

(ii) Next, set $\Gamma = \frac{1}{3\Lambda^2} \min\left\{1, \frac{2\theta - \Lambda}{2\theta \|A\|^2}, \frac{1}{\|B\|^2}\right\} > 0$, and pick $\gamma \in (0, \Gamma)$.

(iii) Finally, set $\Delta = \frac{1}{\Lambda L_p}\sqrt{1 - 3\gamma\Lambda^2} > 0$ and pick $\delta \in (0, \Delta)$.

The main result of this section establishes the global weak convergence for the sequence generated by LaPSA' to a solution of (11):

Theorem 3.1 *Suppose that Z is finite dimensional. Under Hypothesis* **(H')**, *for every starting point $w_0 \in X \times Y \times Z \times Z \times \mathbb{R}_+$, the sequence generated by LaPSA' converges weakly to a point in $\mathcal{S}_{\mathcal{M}} = \mathcal{M}^{-1}(0)$.*

Remark 4 In the context of the previous section, the dimension of Z may be arbitrary.

3.2 Convergence Analysis

We begin by showing the following estimation, which is similar (both in its spirit and its proof) to Lemma 2.1. Recall the spaces and norms introduced in Definition 2.1.

Lemma 3.1 *Let $\gamma > 0$, $\delta > 0$ and $0 < \lambda_k \leq \Lambda < 2\theta$. Suppose that $\bar{w}$ belongs to $\mathcal{S}_{\mathcal{M}}$. Then, we have*

$$\|w_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + \bar{E}_k \leq 2\lambda_k (\alpha_k + \beta_k),$$

where

$$\bar{E}_k = \left(\eta - 3\gamma\lambda_k^2 \|A\|^2\right) \|x_{k+1} - x_k\|^2 + \left(1 - 3\gamma\lambda_k^2 \|B\|^2\right) \|y_{k+1} - y_k\|^2 + \left(1 - \lambda_k^2 \left(3\gamma + L_p^2 \delta\right)\right) \|z_{k+1} - z_k\|^2 + \frac{1}{\gamma} \|\tilde{\mu}_{k+1} - \mu_k\|^2 + \frac{1}{\delta} \|\tilde{\nu}_{k+1} - \nu_k\|^2.$$

Moreover, if Hypothesis (H2') *holds, there is* $\sigma > 0$ *such that*

$$\bar{E}_k \geq \sigma \|u_{k+1} - u_k\|^2 + \|\tilde{v}_{k+1} - v_k\|^2.$$

The proof is very close to that of Lemma 2.1, but we include it in "Appendix D" for the sake of completeness. As a consequence of Lemmas A.1 and 3.1, we obtain the following result:

Proposition 3.1 *Let Hypothesis* **(H')** *hold. Then,*

(i) *for every* $\bar{w}$ *in* $\mathcal{S}_{\mathcal{M}} = \mathcal{M}^{-1}(0)$, $\lim_k \|w_k - \bar{w}\|$ *exists in* $\mathbb{R}$*;*
(ii) *the sequences* $\left(\|u_{k+1} - u_k\|^2\right)$, $\left(\|\tilde{v}_{k+1} - v_k\|^2\right)$, $\left(\|Ax_k + By_k - z_k\|^2\right)$ *and* $\left(\|p(z_k)\|^2\right)$ *belong to* ℓ^1.

We are now in a position to prove the main result of this section.
Proof of Theorem 3.1. As in the proof of Theorem 2.1 we use Opial's Lemma A.2. The first condition is given in item i) of Proposition 3.1. It remains to show that every weak limit point of the sequence (w_k) belongs to $\mathcal{S}_{\mathcal{M}}$. Let w_{k_j} converge weakly, as $k \to \infty$, to w^∞. Item ii) in Proposition 3.1 implies

$$Ax_k + By_k - z_k \to 0 \quad \text{and} \quad p(z_k) \to 0.$$

Since Z is finite dimensional and λ_k is bounded, we have

$$\tilde{\mu}_{k_j+1} \to \mu^\infty \quad \text{and} \quad \tilde{\nu}_{k_j+1} \to \nu^\infty.$$

Moreover, as $Ax_{k_j} + By_{k_j} - z_{k_j} \to Ax^\infty + By^\infty - z^\infty$ and $Ax_k + By_k - z_k \to 0$, we have

$$Ax^\infty + By^\infty = z^\infty,$$

which is the forth condition in (11). By lower semicontinuity, we also have $p(z^\infty) = 0$ (fifth condition in (11)) and so that z^∞ belongs to C. Now take $x \in X$ and $\tilde{x} \in Mx$. The monotonicity of M gives

$$\langle -\frac{x_{k+1} - x_k}{\lambda_k} - A^*\tilde{\mu}_{k+1} - \tilde{x},\ x_k - x\rangle \geq 0. \tag{12}$$

But $-\frac{x_{k+1}-x_k}{\lambda_k} \to 0$, because $\|x_{k+1} - x_k\|^2 \to 0$ and $\lambda_k \geq \lambda > 0$. Also, $A^*\tilde{\mu}_{k_j+1} \to A^*\mu^\infty$ and $x_{k_j} \rightharpoonup x^\infty$. So, passing to the limit for $k \to +\infty$ in (12), we obtain

$$\langle -A^*\mu^\infty - \tilde{x},\ x^\infty - x\rangle \geq 0. \tag{13}$$

The maximality of M gives $-A^*\mu^\infty = Mx^\infty$, which is the first condition in (11). In a similar fashion, we verify the second one using the definition of N_{α_k} and the hypothesis $\alpha_k \in \ell^1$ (more specifically, use that $\alpha_k \to 0$). Finally, take $z \in Z$ and $z^* \in \nu^\infty \partial p(z)$. We have

$$\nu^\infty\ p(z_{k+1}) \geq \nu^\infty\ p(z) + \langle z^*,\ z_{k+1} - z\rangle.$$

Using (B3'), we obtain

$$\tilde{\nu}_{k+1}\ p(z) \geq \tilde{\nu}_{k+1}\ p(z_{k+1}) + \left\langle -\frac{z_{k+1} - z_k}{\lambda_k} + \tilde{\mu}_{k+1},\ z - z_{k+1} \right\rangle - \beta_k.$$

The two previous inequalities yield

$$\left(\nu^\infty - \tilde{\nu}_{k+1}\right)\left[p(z_{k+1}) - p(z)\right] \geq \left\langle z^* + \frac{z_{k+1} - z_k}{\lambda_k} - \tilde{\mu}_{k+1},\ z_{k+1} - z \right\rangle - \beta_k.$$

Passing to the limit in the subsequence, we conclude that

$$\langle \mu^\infty - z^*,\ z^\infty - z\rangle \geq 0.$$

Again, by maximality, it follows that $\mu^\infty \in \partial^z \left[\nu^\infty p(z^\infty)\right]$, which is the third condition in (11). □

3.3 Linear Convergence in a Strongly Non-degenerate Setting

To finish this section, we derive the linear convergence of LaPSA' in a *strongly non-degenerate* setting, namely when the operator $\mathcal{M}$ defined in (10) is locally Lipschitz continuous at the origin. Although the analysis is rather standard in the proximal case, we believe it is important to show that it can be applied in this parallel forward–backward setting. Therefore, we include an abridged proof for the sake of completeness. We begin by recalling the following auxiliary lemma (see, for instance, [25]):

Lemma 3.2 *Let H be a Hilbert space and let $\mathcal{T} : H \rightrightarrows H$. Consider $\bar{\omega} \in \mathcal{T}^{-1}(0)$. Assume there exists $\tau > 0$ and $\zeta > 0$ such that*

$$\zeta\ \|\omega - \bar{\omega}\| \leq \|\pi\|$$

whenever $\pi \in \mathcal{T}(\omega)$ and $\|\pi\| < \tau$. Let (ω_k) be a sequence in H and suppose that, for all sufficiently large $k \in \mathbf{N}$, there is $\xi > 0$ such that

$$\|\omega_{k+1} - \bar{\omega}\|^2 + \xi\,\|\pi_k\|^2 \leq \|\omega_k - \bar{\omega}\|^2 \tag{14}$$

for some $\pi_k \in \mathcal{T}(\omega_{k+1})$. Then, there exists $\bar{k} \in \mathbf{N}$ such that

$$\|\omega_{k+1} - \bar{\omega}\| \leq \frac{1}{\sqrt{1 + \zeta^2\xi}}\ \|\omega_k - \bar{\omega}\|$$

for every $k \geq \bar{k}$. In particular, ω_k converges linearly (and strongly) to $\bar{\omega}$.

We have the following

Theorem 3.2 *Let Hypothesis* **(H')** *hold and let $\bar{w} \in \mathcal{M}^{-1}0$. Assume there exist $\tau > 0$ and $\zeta > 0$ such that*

$$\zeta \, \|w - \bar{w}\| \leq \|\pi\|$$

whenever $\pi \in \mathcal{M}(w)$ and $\|\pi\| < \tau$. Then, every sequence generated by LaPSA', with $\alpha_k = \beta_k = 0$, converges linearly (and strongly) to $\bar{w}$.

Proof Take

$$\pi_k = \left(\pi_k^x, \pi_k^y, \pi_k^z, \pi_k^\mu, \pi_k^\nu\right) \in \mathcal{M}\left(w_{k+1}\right).$$

After rather long but otherwise straightforward computations, we obtain

$$\begin{aligned}
\|\pi_k^x\|^2 &\leq \left(\frac{4}{\theta^2} + 8\gamma^2\Lambda^2\|A\|^4 + \frac{8}{\lambda^2}\right)\|x_{k+1} - x_k\|^2 + 4\gamma^2\Lambda^2\|A\|^2\|B\|^2\|y_{k+1} - y_k\|^2 \\
&\quad + 4\gamma^2\Lambda^2\|A\|^2\|z_{k+1} - z_k\|^2, \\
\|\pi_k^y\|^2 &\leq 3\gamma^2\Lambda^2\|B\|^2\|A\|^2\|x_{k+1} - x_k\|^2 + \left(6\gamma^2\Lambda^2\|B\|^4 + \frac{6}{\lambda^2}\right)\|y_{k+1} - y_k\|^2 \\
&\quad + 3\gamma^2\Lambda^2\|B\|^2\|z_{k+1} - z_k\|^2, \\
\|\pi_k^z\|^2 &\leq 9\gamma^2\Lambda^2\|A\|^2\|x_{k+1} - x_k\|^2 + 9\gamma^2\Lambda^2\|B\|^2\|y_{k+1} - y_k\|^2 \\
&\quad + \left(3L_p^4\delta^2\Lambda^2 + 9\gamma^2\Lambda^2 + \frac{3}{\lambda^2}\right)\|z_{k+1} - z_k\|^2, \\
\|\pi_k^\mu\|^2 &\leq 6\|A\|^2\|x_{k+1} - x_k\|^2 + 6\|B\|^2\|y_{k+1} - y_k\|^2 \\
&\quad + 6\|z_{k+1} - z_k\|^2 + \frac{2}{\gamma^2\lambda^2}\|\tilde{\mu}_{k+1} - \mu_k\|^2, \\
\|\pi_k^\nu\|^2 &\leq 2L_p^2\|z_{k+1} - z_k\|^2 + \frac{2}{\delta^2\lambda^2}\|\tilde{\nu}_{k+1} - \nu_k\|^2.
\end{aligned}$$

It follows that

$$\|\pi_k\|^2 \leq D \,\|\left(u_{k+1}, \tilde{v}_{k+1}\right) - \left(u_k, v_k\right)\|^2$$

for some constant $D > 0$. Using this inequality, along with Lemma 3.1, we obtain

$$\|w_{k+1} - \bar{w}\|^2 + \xi\|\pi_k\|^2 \leq \|w_k - \bar{w}\|^2$$

with $\xi = \frac{1}{D}\min\{1, \sigma\}$. The result then follows from Lemma 3.2. □

The hypotheses of the previous result hold in the following simple example:

Proposition 3.2 *Let $M : H \to H$ be strongly monotone and cocoercive, let $A : H \to Z$ be linear, bounded and surjective, and fix $c \in Z$. Define*

$$\mathcal{M}(x, \mu) = \left(M(x) + A^* \mu,\ c - Ax\right).$$

Then $\mathcal{M}^{-1}$ is Lipschitz continuous at the origin.

Proof Let $(\bar{x}, \bar{\mu}) \in \mathcal{M}^{-1}(0)$ and $(x, \mu) \in \mathcal{M}^{-1}(\pi^x, \pi^\mu)$. In other words,

$$\begin{aligned} -A^* \bar{\mu} &= M(\bar{x}); \\ A\bar{x} &= c; \\ \pi^x &= M(x) + A^* \mu; \\ \pi^\mu &= c - Ax. \end{aligned}$$

Since M is strongly monotone and cocoercive, we deduce that

$$\rho \|x - \bar{x}\|^2 + \theta \|\pi^x - A^*(\mu - \bar{\mu})\|^2 \leq 2\langle \pi^x - A^* \mu + A^* \bar{\mu}, x - \bar{x} \rangle$$

for some constants $\rho > 0$ and $\theta > 0$. The previous inequality implies

$$\begin{aligned} \rho \|x - \bar{x}\|^2 + \theta \|A^*(\mu - \bar{\mu})\|^2 &\leq 2\langle \pi^x, x - \bar{x}\rangle - 2\langle \mu - \bar{\mu}, A(x - \bar{x})\rangle \\ &\quad + 2\theta \langle \pi^x, A^*(\mu - \bar{\mu})\rangle \\ &= 2\langle \pi^x, x - \bar{x}\rangle + 2\langle \pi^\mu, \mu - \bar{\mu}\rangle \\ &\quad + 2\theta \langle \pi^x, A^*(\mu - \bar{\mu})\rangle \\ &\leq 2 \max\{1, \theta \|A^*\|\} \cdot \left[\|\pi^x\| + \|\pi^\mu\|\right] \\ &\quad \cdot \left[\|x - \bar{x}\| + \|\mu - \bar{\mu}\|\right], \\ &\leq 8 \max\{1, \theta \|A^*\|\} \\ &\quad \cdot \|(\pi^x, \pi^\mu)\|_2 \cdot \|(x - \bar{x}, \mu - \bar{\mu})\|_2. \end{aligned}$$

Finally, A is surjective if, and only if, there is a constant $D > 0$ such that $\|A^* p\| \geq D\|p\|$ for all $p \in H$ (see [33, Theorem 2.20]). We deduce that

$$\begin{aligned} \min\{\rho, \theta D^2\} \cdot \|(x - \bar{x}, \mu - \bar{\mu})\|_2^2 &\leq 8 \max\{1, \theta \|A^*\|\} \cdot \|(\pi^x, \pi^\mu)\|_2 \\ &\quad \cdot \|(x - \bar{x}, \mu - \bar{\mu})\|_2. \end{aligned}$$

It suffices to define

$$\zeta = \frac{\min\{\rho, \theta D^2\}}{8 \max\{1, \theta \|A^*\|\}}$$

to conclude. □

4 Implementation Issues and Areas of Application

In this section, we briefly discuss some implementation issues and outline some potential applications.

4.1 The Feasible Set *C* and Its Representation

For various choices of the constraint subset C and a related penalization function p, we focus on sub-step (B3) of LaPSA, which consists in finding the unique solution $\hat{z}$ for a problem of the form

$$\min_{z \in Z} \left\{ p(z) + \frac{1}{2\lambda} \|z - \bar{z}\|^2 \right\}, \tag{15}$$

where $\lambda > 0$ and $\bar{z} \in Z$ are given. In many relevant cases, this step is explicit.

4.1.1 Ball-Like Sets

Ball-like sets have simple structures that allow for simple penalization functions for which the proximal step can be explicitly computed.

(i) *Singletons.* Let $C = \{c\}$ for $c \in Z$, and set $p(z) = \frac{1}{2}\|z - c\|^2$. Then, trivially, $\hat{z} = \frac{\bar{z}+\lambda c}{1+\lambda}$. Another option is to set $p(z) = \|z - c\|$. In that case, $\hat{z} = \bar{z} - \lambda \|\bar{z} - c\|^{-1}(\bar{z} - c)$ if $\|c - \bar{z}\| > \lambda$, and $\hat{z} = c$ otherwise. (ii) *Ellipsoids.* For $Z = \mathbb{R}^N$, let $C \subseteq Z$ be defined as

$$C = \left\{ z \in Z : \sum_{i=1}^{N} a_i z_i^2 \leq 1 \right\}$$

for (a_i) vector of positive real numbers. Such sets arise naturally under *bounded variance* constraints in stochastic optimization problems. Moreover, define

$$p(z) = \left(\sum_{i=1}^{N} a_i z_i^2 - 1 \right)_+ ,$$

where, for $a \in \mathbb{R}$, the positive part is given by $(a)_+ = \max(0, a)$. Then, for every $i = 1, \ldots, N$, we have

$$\hat{z}_i = \begin{cases} [\Pi_C(\bar{z})]_i & \text{if } \sum_{i=1}^{N} \frac{a_i \bar{z}_i^2}{(1+2\lambda a_i)^2} \leq 1, \\ \frac{\bar{z}_i}{1+2\lambda a_i} & \text{otherwise,} \end{cases}$$

where Π_C denotes the projection operator on the set C. In particular, when $a_i = 1$ for every $i = 1, \ldots, N$, then C is the unit Euclidean ball centred at the origin, and

$$\hat{z}_i = \begin{cases} \bar{z}_i & \text{if } \|\bar{z}\|_2 \leq 1, \\ \frac{\bar{z}_i}{\|\bar{z}\|_2} & \text{if } 1 \leq \|\bar{z}\|_2 \leq 1 + 2\lambda, \\ \frac{\bar{z}_i}{1+2\lambda} & \text{otherwise.} \end{cases}$$

(iii) ℓ^1 emph balls. Again for $Z = \mathbb{R}^N$, let

$$C = \{ z \in Z : \ \|z\|_1 \leq 1 \},$$

where $\|z\|_1 = \sum_{i=1}^N |z_i|$. Choose the penalization function as

$$p(z) = \left(\sum_{i=1}^N |z_i| - 1 \right)_+ .$$

Now define the *shrink* (or *soft-thresholding*) operator: for $i = 1, \dots, N$

$$\left[shrink\left(\hat{z}, \lambda\right) \right]_i = shrink_1(\hat{z}_i, \lambda), \tag{16}$$

where, for $t \in \mathbb{R}$,

$$shrink_1(t, \lambda) := \arg\min_{s \in \mathbb{R}} \left\{ \lambda |s| + \frac{1}{2} |s - t|^2 \right\} = \begin{cases} t + \lambda & \text{if } t < -\lambda \\ 0 & \text{if } -\lambda \leq t \leq \lambda \\ t - \lambda & \text{if } t > \lambda. \end{cases} \tag{17}$$

Notice the similarity with the second case of i). Then we can write the solution of problem (15) as

$$\hat{z} = \begin{cases} \Pi_C(\bar{z}) & \text{if } \|shrink(\bar{z}, \lambda)\|_1 \leq 1 \\ shrink(\bar{z}, \lambda) & \text{otherwise.} \end{cases}$$

(iv) *Boxes* (*or* ℓ^∞ *balls*). For $Z = \mathbb{R}^N$, let

$$C = \times_{i=1}^N [a_i, b_i],$$

where (a_i) and (b_i) are vectors of real numbers. If we choose $p(z) = \sum_{i=1}^N p_i(z_i)$ with

$$p_i(z_i) = \begin{cases} (a_i - z_i)^2 & \text{if } z_i < a_i, \\ 0 & \text{if } a_i \leq z_i \leq b_i \\ (z_i - b_i)^2 & \text{else,} \end{cases}$$

then the components of the solution are given by

$$\hat{z}_i = \begin{cases} \frac{\bar{z}_i + 2\lambda a_i}{1+2\lambda} & \text{if } \bar{z}_i < a_i \\ \bar{z}_i & \text{if } a_i \leq \bar{z}_i \leq b_i \\ \frac{\bar{z}_i + 2\lambda b_i}{1+2\lambda} & \text{else.} \end{cases}$$

As in the first example, another option is to set

$$p_i(z_i) = \begin{cases} a_i - z_i & \text{if } z_i < a_i \\ 0 & \text{if } a_i \leq z_i \leq b_i \\ z_i - b_i & \text{otherwise.} \end{cases}$$

In this case,

$$\hat{z}_i = \begin{cases} \bar{z}_i + \lambda & \text{if } \bar{z}_i < a_i - \lambda \\ \Pi_{[a_i, b_i]}(\bar{z}_i) & \text{if } a_i - \lambda \leq \bar{z}_i \leq b_i + \lambda \\ \bar{z}_i - \lambda & \text{otherwise.} \end{cases}$$

4.1.2 Polyhedra

Polyhedra are intersections of half-spaces. They can be characterized as the solutions of systems of linear inequalities. Let

$$C = \{ z \in Z : \langle a_i, z \rangle \leq \alpha_i, \quad i = 1, \ldots, N \}$$

for $a_i \in Z$ and $\alpha_i \in \mathbb{R}$ for $i = 1, \ldots, N$. For any $\epsilon > 0$ (its importance will be clarified later), we have

$$C = \{ z \in Z : \langle \epsilon a_i, z \rangle \leq \epsilon \alpha_i, \quad i = 1, \ldots, N \}.$$

In general, the projection on these polyhedra is a non-trivial combinatorial problem, especially for big values of N (in [34, Chapter 28], you can find the cases $N = 1, 2$). The penalization approach can simplify the resolution of this problem. Indeed, it can be easier to compute a proximity operator of a penalization function for C than computing directly the projection onto the set. For instance, set

$$p(z) = \frac{1}{2} \sum_{i=1}^{N} [\langle \epsilon a_i, z \rangle - \epsilon \alpha_i]_+^2 .$$

The optimality condition for (15) is given by

$$0 = \epsilon \sum_{i=1}^{N} \left[\langle \epsilon a_i, \hat{z} \rangle - \epsilon \alpha_i \right]_+ a_i \; + \; \frac{1}{\lambda} \left(\hat{z} - \bar{z} \right). \tag{18}$$

Now, for $i = 1, \ldots, N$, define

$$\tau_i = \left[\langle \epsilon a_i, \hat{z}\rangle - \epsilon \alpha_i\right]_+ , \tag{19}$$

and rewrite (18) as

$$\hat{z} = \bar{z} - \epsilon\lambda \sum_{i=1}^{N} \tau_i a_i . \tag{20}$$

Substitute (20) in (19), to obtain an equation on the only scalar variables $(\tau_i)_{i=1}^N$:

$$\tau_i = \left[\epsilon\langle a_i, \bar{z}\rangle - \epsilon^2\lambda \sum_{j=1}^{N} \langle a_i, a_j\rangle \tau_j - \epsilon\alpha_i\right]_+ , \quad \text{for } i = 1, \ldots, N. \tag{21}$$

In order to pass to the vectorial notation, define

$$A_{ij} = \langle a_i, a_j\rangle, \qquad B_{ij} = -\epsilon^2\lambda A_{ij} \qquad \text{and} \qquad \beta_i = \epsilon\left(\langle a_i, \bar{z}\rangle - \alpha_i\right).$$

Moreover, introduce the map $T : \mathbb{R}^N \to \mathbb{R}^N$ defined by

$$T(\tau) = [B\tau + \beta]_+ .$$

From (21), the vector τ is a fixed point of the map T, which is Lipschitz continuous with constant $\|B\| = \epsilon^2\lambda\|A\|$. For each $\epsilon < (\lambda\|A\|)^{-1/2}$, the map T is (strictly) contractive. Thus, τ can be approximated using fixed point iterations:

$$\tau^k = T^k(\tau_0)$$

from some initial guess τ_0 until some $K \in \mathbf{N}$ that gives a desired precision. The solution of (15) can be approximated by

$$\hat{z} \approx \bar{z} - \epsilon\lambda \sum_{i=1}^{N} \tau_i^K a_i .$$

This analysis can be extended to nonlinear constraints defined as intersections of sublevel sets of convex functions.

4.1.3 Skipping the Penalization

If the constraint set C has a simple projector, such as balls, boxes, the positive orthant, to mention a few (see also [34, Chapter 28] or [35, Chapter 8]), the penalization procedure can be replaced by directly projecting onto it. This can be done by considering the space $\mathbf{Y} = Y \times Z$, and writing $G(\mathbf{y}) = g(y) + \delta_C(t)$ for $\mathbf{y} = (y, t)$, where δ_C is the indicator function for the set C. So, the resulting problem is

$$\min_{(x,\mathbf{y}) \in X \times \mathbf{Y}} \{ f(x) + G(\mathbf{y}) : \; Ax + \mathbf{B}\mathbf{y} = 0 \}, \tag{22}$$

where $\mathbf{B}\mathbf{y} = By - t$. The proximal step (B2) admits a separable form. Indeed, $\mathbf{y}_{k+1} = (y_{k+1}, t_{k+1})$ where, for $\tilde{y}_{k+1} = y_k - \lambda_k B^* \tilde{\mu}_{k+1}$,

$$y_{k+1} = \arg\min_{y \in X} \left\{ g(y) + \frac{1}{2\lambda_k} \|y - \tilde{y}_{k+1}\|^2 \right\}$$

and t_{k+1} is just the projection on C of $\tilde{t}_{k+1} = t_k + \lambda_k \tilde{\mu}_{k+1}$.

4.1.4 Parallel Splitting

Consider the following structured problem:

$$\min_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} g_i(x) \right\}, \tag{23}$$

where $f : X \to \mathbb{R}$ is again a proper, convex, differentiable function with ∇f Lipschitz continuous and (g_i) are proper, convex and lower-semicontinuous functions on X (see [36] and the references therein). Introducing positive weights (ω_i) such that $\sum_{i=1}^m \omega_i = 1$, an equivalent formulation of (23) in the product space is

$$\min_{x \in X, \; \mathbf{y} \in X^m} \left\{ f(x) + \sum_{i=1}^{m} g_i(y^i) : \quad x = \sum_{i=1}^{m} \omega_i y^i \; \text{ and } \; y^1 = y^2 = \ldots = y^m \right\}.$$

Now define the function g on $Y = X^m$ as $g(\mathbf{y}) = \sum_{i=1}^m g_i(y^i)$ and the linear operators

$$A = \begin{pmatrix} \mathbb{1}_X \\ 0_X \\ 0_X \\ \ldots \\ 0_X \end{pmatrix}, \quad B = \begin{pmatrix} -\omega_1 & -\omega_2 & -\omega_3 & \ldots & -\omega_{m-1} & -\omega_m \\ \mathbb{1}_X & -\mathbb{1}_X & 0_X & \ldots & 0_X & 0_X \\ 0_X & \mathbb{1}_X & -\mathbb{1}_X & \ldots & 0_X & 0_X \\ & & & \ldots & & \\ 0_X & 0_X & 0_X & \ldots & \mathbb{1}_X & -\mathbb{1}_X \end{pmatrix}.$$

Problem (23) can be reformulated as:

$$\min_{x \in X, \; \mathbf{y} \in Y} \{ f(x) + g(\mathbf{y}) : \quad Ax + B\mathbf{y} = 0 \}.$$

LaPSA does not require the computation of the proximal operator for the sum $\sum_{i=1}^{m} g_i$, but just of each function g_i *separately*. More precisely, the proximal step (B2), namely

$$\mathbf{y}_{k+1} \in \underset{\mathbf{y}\in Y}{\arg\min} \left\{ g(\mathbf{y}) + \langle \tilde{\mu}_{k+1}, \; Ax_k + B\mathbf{y}_k \rangle_Y + \frac{1}{2\lambda_k} \|\mathbf{y} - \mathbf{y}_k\|_Y^2 \right\},$$

can be written component-wise as:

$$y_{k+1}^i = \underset{y\in X}{\arg\min} \left\{ g_i(y) + \langle \left(B^* \tilde{\mu}_{k+1}\right)^i, \; y \rangle + \frac{1}{2\lambda_k} \|y - y_k^i\|^2 \right\}.$$

4.2 Some Particular Models

We now describe some examples for which the application of LaPSA is convenient.

4.2.1 Image Recovery

In what follows, $\hat{x} \in \mathbb{R}^{N_d}$ a vector containing the information of a damaged image, $H \in \mathbb{R}^{N_d \times N_x}$ is a blurring/compressing (linear) operator (we shall denote its ith row by H_i) and $\varepsilon \in \mathbb{R}^{N_d}$ is a vector representing some noise. We are interested in the reconstruction of the original image $\bar{x}$ in $\mathbb{R}^{N_x}$ that satisfies the relation $\hat{x} = H\bar{x} + \varepsilon$, assuming that, for some given matrix $G \in \mathbb{R}^{N_z \times N_y}$, the vector Gx is *sparse*. Then it is possible to approximate $\bar{x}$ by solving the minimization problem

$$\min_{x\in\mathbb{R}^{N_x}} \{ r\|Gx\|_1 + f(x) \}, \tag{24}$$

where the term $\|G(\cdot)\|_1$ is known as the ℓ^1-*regularization* and f is a smooth function representing the *fidelity* to the data $\hat{x}$. The positive parameter r accounts for a trade-off between the regularization and the fidelity. We mention here two possible choices for G, namely

- If G is a transformation to a wavelet basis, (24) is the well-known *wavelet-based restoration model* (see, for instance, [37] and the numerical examples in [38]);
- On the other hand, it corresponds to the *Total Variation* minimization, also called *ROF model*, if the matrix G is a discrete gradient operator (first proposed in [39], see also [40] and [41]).

Notice that, while G is easily invertible in the first example, it is not the case in the second. For the data fidelity f, two possible choices are the following:

- The quadratic function

$$f(x) = \frac{1}{2}\|Hx - \hat{x}\|_2^2 \tag{25}$$

  is typically used when the noise is Gaussian.

– In turn, when the noise has a Poisson distribution (see, for example, [42], [43] and [44]), the preferred choice for f tends to be the *Kullback–Leibler distance* (also called *relative entropy*), given by

$$f(x) = \sum_{i=1}^{N_d} \left[(Hx)_i - \hat{x}_i \ln (Hx)_i \right]. \tag{26}$$

Introducing the auxiliary variable $y = Gx$, both wavelet-based restoration and the ROF model operate in the setting of problem (1) with $X = \mathbb{R}^{N_x}$, $Y = \mathbb{R}^{N_y}$, $Z = \mathbb{R}^{N_y}$, $g(y) = r\|y\|_1$, $A = G$, $B = -I$ and $C = \{0\} \subset \mathbb{R}^{N_z}$. The proximal step corresponding to the non-differentiable function g has an explicit solution in terms of the $shrink$ operator (see equations (16) and (17)). As commented earlier step $(B3)$ is also explicit if the penalization function is quadratic. Moreover, it is clear that the function in (25) enters in our setting, since its gradient is globally Lipschitz continuous. For the Kullback–Leibler distance (26), it is possible to localize the problem in a region containing the solution and such that the gradient of f is well defined and Lipschitz continuous (using *a priori* information). Moreover, the gradients of functions (25) and (26) are easily computed, while the computation of the proximal point iterations is, in general, expensive. As a consequence, the *proximal gradient* splitting in LaPSA is preferable to the *proximal–proximal* splitting proposed in other algorithms, such as [12].

4.2.2 Non-Newtonian Fluids

The following problem arises, for example, in the modeling of non-Newtonian fluids (see [45] and the references therein):

$$\min_{u \in W_0^{1,p}(\Omega)} \left\{ \frac{1}{p} \int_\Omega |\nabla u|^p \, \mathrm{d}x - \int_\Omega hu \, \mathrm{d}x + r \int_\Omega |\nabla u| \ x \right\}$$

for $p \geq 1$ and $r > 0$. If $D \in \mathbb{R}^{N \times N}$ is a suitable approximation matrix for the gradient, the corresponding discrete version is given by

$$\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{p} \|Dx\|_p^p - \langle h, \ x \rangle + r\|Dx\|_1 \right\}. \tag{27}$$

To deal with this problem, we add, as a constraint, a possibly rough estimation $\|Dx\|_\square \leq M$ for the localization of the gradient $D\bar{x}$ of a solution $\bar{x}$, where M is an appropriate positive constant, and $\|\cdot\|_\square$ is a *convenient* norm in $\mathbb{R}^N$ (for instance, whose balls have a simple projector). This guarantees that the function

$$f(x) = \frac{1}{p} \|Dx\|_p^p - \langle h, x \rangle$$

has Lipschitz continuous gradient in the problem domain. Writing $y = Dx$, this falls in the setting of (1):

$$\min_{x\in\mathbb{R}^N}\left\{\frac{1}{p}\|Dx\|_p^p - \langle h,\ x\rangle + r\|y\|_1 \ :\ \ Dx = y \text{ and } \|y\|_\square \le M\right\}. \tag{28}$$

If $\|\cdot\|_\square = \|\cdot\|_\infty$, it is not necessary to introduce the auxiliary variable z in order to obtain a fully explicit scheme. Indeed, (28) can be rewritten as

$$\min_{x\in\mathbb{R}^N}\left\{\frac{1}{p}\|Dx\|_p^p - \langle h,\ x\rangle + r\|y\|_1 + \delta_{[-M,M]^N}(y)\ :\ \ Dx = y\right\},$$

with the following Lagrangian functional:

$$\mathcal{L}(x, y, \mu) = \frac{1}{p}\|Dx\|_p^p - \langle h,\ x\rangle + r\|y\|_1 + \delta_{[-M,M]^N}(y) + \langle \mu,\ Dx - y\rangle.$$

The algorithm is reduced to

$$\begin{aligned}
\tilde{\mu}_{k+1} &= \mu_k + \gamma\lambda_k\left(Dx_k - y_k\right);\\
(x_{k+1})_q &= (x_k)_q - \lambda_k\left[\left(D^*\tilde{\mu}_{k+1}\right)_q - (h)_q + \sum_{i=1}^{N} D_{ij}\left(\sum_{j=1}^{N} D_{ij}\,(x_k)_j\right)^{p-1}\right],\\
&\qquad q = 1,\ldots,N;\\
y_{k+1} &= \operatorname*{arg\,min}_{w\in\mathbb{R}^N}\left\{r\|y\|_1 + \delta_{[-M,M]^N}(y) - \langle\tilde{\mu}_{k+1},\ y\rangle + \frac{1}{2\lambda_k}\|y - y_k\|_2^2\right\};\\
\mu_{k+1} &= \mu_k + \gamma\lambda_k\left(Dx_{k+1} - y_{k+1}\right).
\end{aligned}$$

The proximal step can be computed explicitly for $q = 1,\ldots,N$ by

$$(y_{k+1})_q = \Pi_M^\infty\left(shrink\left((y_k + \lambda_k\tilde{\mu}_k)_q\ ,\ r\lambda_k\right)\right),$$

where the *shrink* operator is defined by (16)–(17), and Π_M^∞ is the projection operator onto $[-M, M]^N$.

4.2.3 Logistic Regression

Finally, we mention the *logistic regression* statistical model, with feature selection given by 1-norm and penalization of the 2−norm:

$$\min_{x\in\mathbb{R}^{N_x}}\left\{\ f(x) + r_1\|x\|_1 + \frac{r_2}{2}\|x\|_2\ \right\},$$

$$\text{where}\qquad f(x) = \sum_{i=1}^{N_d}\ln\left[1 + \exp\left(-\bar{x}_i\langle h_i,\ x\rangle\right)\right],$$

where and h_i are given vectors in $\mathbb{R}^n_x$. In this case, the model fits directly in our setting because f is everywhere differentiable with Lipschitz continuous gradient. Observe that LaPSA allows to include linear constraints to this model in a very simple way.

Remark 5 In some of the examples given above, the computation of the proximal point operator for the differentiable function may be complicated or costly and may not have separable structure in the standard coordinates. LaPSA offers a computationally inexpensive alternative to other methods, such as its relatives [12] and [19].

5 Numerical Illustration

For the numerical illustration of the algorithm, we consider the following optimization problem:

$$\min_{(x,y)\in\mathbb{R}^{N_x}\times\mathbb{R}^{N_y}} \left\{ \frac{1}{3}\|x\|_p^p + \alpha\|y\|_1 + \frac{\beta}{2}\|y\|_2^2 : \quad Ax + By = c \right\}, \tag{29}$$

where A and B are random generated matrices, respectively, in $\mathbb{R}^{N_x\times M}$ and $\mathbb{R}^{N_y\times M}$, and the vector $c \in \mathbb{R}^M$ is chosen so that the feasible set is not empty neither a singleton. In the presented experiment, the dimensions of the spaces are given by $N_x = 2^{12}$, $N_y = 2^{11}$ and $M = \frac{3}{4}\left(N_x + N_y\right)$, and the constants take the values $p = 3.5$, $\alpha = 0.05$ and $\beta = 0.0001$. The starting points x_0, y_0 and μ_0 are the zero vectors in the correspondent spaces. In Fig. 1, we show the values of the objective $(f(x_k) + g(y_k))$ and the gap $(\|Ax_k + By_k - c\|_2^2)$ along the first 4000 iterations of the

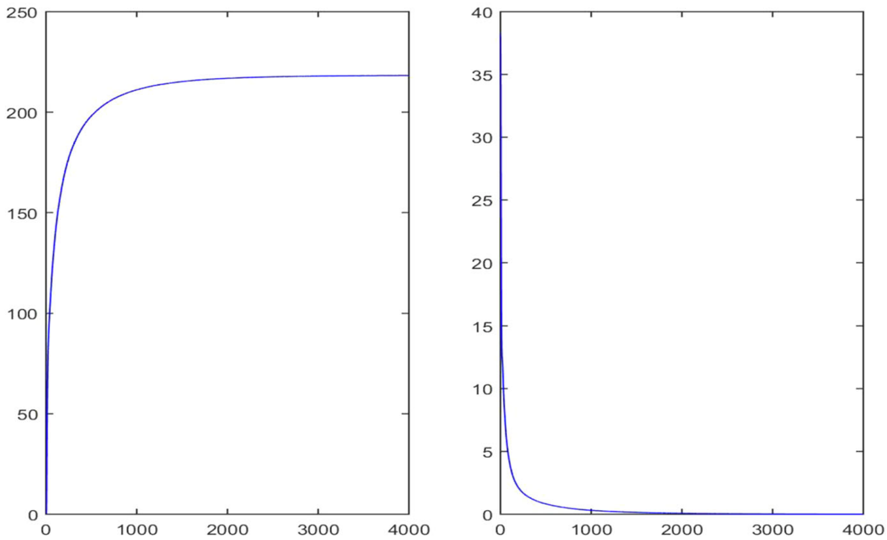

Fig. 1 Energy (left) and the gap (right) along the first 4000 iterations

algorithm. It is clear that both the quantities converge and that the gap approximates zero.

Remark 6 Notice that it is not possible to apply directly the forward–backward algorithm (ISTA) to problem (29). Moreover, the computation of the prox-operator of the function $\frac{1}{3}\|x\|_3^3$ is not straightforward in general, because it requires the resolution of a nonlinear equation. On the other side, the iterations of the proposed algorithm are again completely explicit.

6 Conclusions

We have presented a class of algorithms to deal with structured convex minimization problems and constrained variational inequalities. As a distinctive mark, they feature parallelized forward–backward primal iterations, which allow us to capture and exploit possible dissymmetries in the regularity of the objective functions or operators involved. They also combine Lagrangian and penalization approaches, with the penalization parameters being understood as Lagrange multipliers, rendering the penalization exact. Finally, they include relaxation parameters that can be tuned to favor the reduction of either the primal or the dual gap. Despite the presence of a penalization function and the primal-dual character of this method, its convergence rate is consistent with that of first-order descent methods.

It is possible to provide a measure of optimality (8) involving the primal and dual variables, where the objective function can be interpreted as a potential for a Bregman divergence. We believe this can lead to a better understanding of (Euclidean) proximal gradient primal-dual methods under the light of Bregman theory, which is an idea for future research. Also observe that the proximal steps may be modified in terms of the metric, by either using a Bregman or a Riemannian approach. On the other hand, in Sect. 2.5, we outlined a variant of this algorithm with a hidden inertial property that should permit the use of Nesterov's acceleration procedure [27]. This would lead to a versatile method with a faster convergence rate.

Acknowledgements Supported by Fondecyt Grant 1140829 and Basal Project CMM Universidad de Chile. The first author was also supported by CONICYT scholarship CONICYT-PCHA/Doctorado Nacional/2016.

Appendix A: Auxiliary Results

The following results will be used in some of the proofs:

Lemma A.1 ([26, Lemma 6.18]) *Let (a_k) and (b_k) be real nonnegative sequences, and (η_k) a nonnegative sequence in ℓ^1. If $a_{k+1} - a_k + b_k \leq \eta_k$ for all (sufficiently large) $k \in \mathbf{N}$, then (a_k) is convergent and (b_k) belongs to ℓ^1.*

Lemma A.2 (Opial's Lemma [46], see [26, Lemma 5.2]) *Let F be a non-empty subset of a Hilbert space H. Suppose that (ω_k) is a sequence in H such that i) for each ω in F, $\lim_k \|\omega_k - \omega\|$ exists; and ii) every weak cluster point of ω_k belongs to F. Then, ω_k converges weakly, as $k \to \infty$, to some $\bar{\omega} \in F$.*

Lemma A.3 ([47, Theorem 3.3.1]) *Let $(a_k)_k$ be nonnegative, non-increasing and summable. Then $a_k = o\left(k^{-1}\right)$.*

We use also the two following results on the inexact computation of proximal iterations.

Lemma A.4 *Let $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower-semicontinuous function on a Hilbert space H. Set an error tolerance $\varepsilon > 0$ and a step size $\lambda > 0$. Given $u_k \in H$, let u_{k+1} satisfy*

$$-\frac{u_{k+1} - u_k}{\lambda} \in \partial_\varepsilon \Phi(u_{k+1}). \tag{30}$$

Then, for every $u \in H$, we have

$$2\lambda \left[\Phi(u_{k+1}) - \Phi(u)\right] + \|u_{k+1} - u\|^2 - \|u_k - u\|^2 + \|u_{k+1} - u_k\|^2 \leq 2\lambda\varepsilon.$$

Proof From the definition of $\partial_\varepsilon \Phi$ and inclusion (30), we have $2\lambda \left[\Phi(u_{k+1}) - \Phi(u)\right] + 2\langle u_k - u_{k+1},\ u - u_{k+1}\rangle \leq 2\lambda\varepsilon$. It suffices to observe that $2\langle u_k - u_{k+1},\ u - u_{k+1}\rangle = \|u_{k+1} - u\|^2 - \|u_k - u\|^2 + \|u_{k+1} - u_k\|^2$ to conclude. □

Lemma A.5 ([12, Lemma 3.1]) *Let $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ be proper, convex and lower-semicontinuous. Set an error tolerance $\varepsilon > 0$ and a step size $\lambda > 0$. Given $u_k \in H$, let*

$$u_{k+1} \in \varepsilon - \operatorname*{arg\,min}_{u \in H} \left\{\Phi(u) + \frac{1}{2\lambda}\|u - u_k\|^2\right\}, \tag{31}$$

and let u^e_{k+1} denote the solution to problem (31) that would be obtained with $\varepsilon = 0$. Then $\|u_{k+1} - u^e_{k+1}\| \leq \sqrt{2\lambda\varepsilon}$.

Appendix B: Proof of Lemma 2.1

We begin by showing the following inequalities:

(a1) $\|\mu_{k+1} - \tilde{\mu}_{k+1}\|^2 + \|\tilde{\mu}_{k+1} - \mu_k\|^2 + 2\gamma\lambda_k \langle \mu_{k+1} - \tilde{\mu}_{k+1}, Ax_k + By_k - z_k\rangle$
$\leq \|\mu_{k+1} - \mu_k\|^2$;

(a2) $|\nu_{k+1} - \tilde{\nu}_{k+1}|^2 + |\tilde{\nu}_{k+1} - \nu_k|^2 + 2\delta\lambda_k\left(\nu_{k+1} - \tilde{\nu}_{k+1}\right)\ p\left(z_k\right) \leq |\nu_{k+1} - \nu_k|^2$;

(b1) $\|x_{k+1} - \bar{x}\|^2 + \left(1 - L_{\nabla f}\lambda_k\right)\|x_{k+1} - x_k\|^2$
$+ 2\lambda_k\left[f(x_{k+1}) - f(\bar{x}) + \langle\tilde{\mu}_{k+1},\ A\left(x_{k+1} - \bar{x}\right)\rangle\right] \leq \|x_k - \bar{x}\|^2$;

(b2) $\|y_{k+1} - \bar{y}\|^2 + \|y_{k+1} - y_k\|^2 + 2\lambda_k\left[g(y_{k+1}) - g(\bar{y}) + \langle\tilde{\mu}_{k+1},\ B\left(y_{k+1} - \bar{y}\right)\rangle\right]$
$\leq \|y_k - \bar{y}\|^2 + 2\lambda_k\alpha_k$;

(b3) $\|z_{k+1}-\bar{z}\|^2+\|z_{k+1}-z_k\|^2$
$+2\lambda_k\left[\tilde{v}_{k+1}\left[p\left(z_{k+1}\right)-p\left(\bar{z}\right)\right]+\langle\tilde{\mu}_{k+1},\ -\left(z_{k+1}-\bar{z}\right)\right]\leq\|z_k-\bar{z}\|^2+2\lambda_k\beta_k;$

(c1) $\|\mu_{k+1}-\bar{\mu}\|^2+\|\mu_{k+1}-\mu_k\|^2+2\gamma\lambda_k\langle\bar{\mu}-\mu_{k+1},Ax_{k+1}+By_{k+1}-z_{k+1}\rangle$
$\leq\|\mu_k-\bar{\mu}\|^2;$

(c2) $|v_{k+1}-\bar{v}|^2+|v_{k+1}-v_k|^2+2\delta\lambda_k\left(\bar{v}-v_{k+1}\right)\ p\left(z_{k+1}\right)\leq|v_k-\bar{v}|^2.$

First, observe that (a1)–(a2), as well as (c1)–(c2), are straightforward consequences of the definitions and simple properties of the norm. To show (b1), rewrite step (B1) as

$$\nabla f(x_k)=-\left[\frac{x_{k+1}-x_k}{\lambda_k}+A^*\tilde{\mu}_{k+1}\right].$$

Using the gradient inequality at the point $\bar{x}$, we obtain

$$2\lambda_k\left[f(x_k)-f(\bar{x})\right]\leq 2\langle x_{k+1}-x_k,\bar{x}-x_k\rangle+2\lambda_k\langle\tilde{\mu}_{k+1},A\left(\bar{x}-x_k\right)\rangle. \tag{32}$$

Moreover, the Descent Lemma (see, for instance, [26, Lemma 1.30]) at x_{k+1} gives

$$f(x_{k+1})-f(x_k)\leq-\left\langle\frac{x_{k+1}-x_k}{\lambda_k}+A^*\tilde{\mu}_{k+1},x_{k+1}-x_k\right\rangle+\frac{L_{\nabla f}}{2}\|x_{k+1}-x_k\|^2. \tag{33}$$

Multiply (33) by $2\lambda_k$, add (32), and rearrange the terms to obtain

$$\begin{aligned}2\lambda_k\left[f(x_{k+1})-f(\bar{x})\right]&\leq 2\lambda_k\langle\tilde{\mu}_{k+1},A\left(\bar{x}-x_{k+1}\right)\rangle\\&\quad+2\langle x_{k+1}-x_k,\bar{x}-x_k\rangle+\left(\lambda_k L_{\nabla f}-2\right)\|x_{k+1}-x_k\|^2.\end{aligned}$$

Finally, we use the equality

$$\|x_{k+1}-\bar{x}\|^2=\|x_{k+1}-x_k\|^2+\|\bar{x}-x_k\|^2-2\langle x_{k+1}-x_k,\bar{x}-x_k\rangle,$$

to conclude. To prove inequalities (b2) and (b3), we use steps (B2) and (B3) of the algorithm and apply Lemma A.4 at the points $\bar{y}$ and $\bar{z}$, respectively.

Recall that the spaces U, V and W were introduced in Definition 2.1. Summing up inequalities (a1) through (c2), we obtain

$$\begin{aligned}&\|w_{k+1}-\bar{w}\|^2-\|w_k-\bar{w}\|^2+\|v_{k+1}-\tilde{v}_{k+1}\|^2+\|\tilde{v}_{k+1}-v_k\|^2\\&\quad+\left(1-L_{\nabla f}\lambda_k\right)\|x_{k+1}-x_k\|^2+\|y_{k+1}-y_k\|^2+\|z_{k+1}-z_k\|^2\\&\quad+2\lambda_k\left[\langle\mu_{k+1}-\tilde{\mu}_{k+1},\ Ax_k+By_k-z_k\rangle+\langle\bar{\mu}-\mu_{k+1},Ax_{k+1}+By_{k+1}-z_{k+1}\rangle\right]\\&\quad+2\lambda_k\left[\left(v_{k+1}-\tilde{v}_{k+1}\right)p\left(z_k\right)+\left(\bar{v}-v_{k+1}\right)p\left(z_{k+1}\right)+\mathcal{L}\left(u_{k+1},\tilde{v}_{k+1}\right)-\mathcal{L}\left(\bar{u},\tilde{v}_{k+1}\right)\right]\\&\leq 2\lambda_k\left(\alpha_k+\beta_k\right).\end{aligned}$$

As $A\bar{x} + B\bar{y} = \bar{z}$ and $p(\bar{z}) = 0$, we have $\mathcal{L}(\bar{u}, \tilde{v}_{k+1}) = \mathcal{L}(\bar{u}, \bar{v})$. Also, notice that

$$\begin{aligned}\mathcal{L}(u_{k+1}, \tilde{v}_{k+1}) = \mathcal{L}(u_{k+1}, \bar{v}) + \langle \tilde{\mu}_{k+1} - \bar{\mu},\ Ax_{k+1} + By_{k+1} - z_{k+1}\rangle \\ + (\tilde{\nu}_{k+1} - \bar{\nu})\, p(z_{k+1}).\end{aligned}$$

Therefore, the inequality above can be rewritten as

$$\begin{aligned}&\|w_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + \|v_{k+1} - \tilde{v}_{k+1}\|^2 + \|\tilde{v}_{k+1} - v_k\|^2 \\ &\quad + \left(1 - L_{\nabla f}\lambda_k\right)\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 + \|z_{k+1} - z_k\|^2 \\ &\quad -2\lambda_k\langle \mu_{k+1} - \tilde{\mu}_{k+1},\ A(x_{k+1} - x_k) + B(y_{k+1} - y_k) - (z_{k+1} - z_k)\rangle \\ &\quad -2\lambda_k(\nu_{k+1} - \tilde{\nu}_{k+1})\left[p(z_{k+1}) - p(z_k)\right] + 2\lambda_k\left[\mathcal{L}(u_{k+1}, \bar{v}) - \mathcal{L}(\bar{u}, \bar{v})\right] \\ &\le 2\lambda_k(\alpha_k + \beta_k).\end{aligned}$$

Since

$$\begin{aligned}A(x_{k+1} - x_k) + B(y_{k+1} - y_k) - (z_{k+1} - z_k) &= \frac{1}{\gamma\lambda_k}(\mu_{k+1} - \tilde{\mu}_{k+1}) \\ p(z_{k+1}) - p(z_k) &= \frac{1}{\delta\lambda_k}(\nu_{k+1} - \tilde{\nu}_{k+1}),\end{aligned}$$

we have

$$\begin{aligned}&\|w_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + \|\tilde{v}_{k+1} - v_k\|^2 \\ &\quad + \left(1 - L_{\nabla f}\lambda_k\right)\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 \\ &\quad + \|z_{k+1} - z_k\|^2 + 2\lambda_k\left[\mathcal{L}(u_{k+1}, \bar{v}) - \mathcal{L}(\bar{u}, \bar{v})\right] \\ &\le 2\lambda_k(\alpha_k + \beta_k) + \|v_{k+1} - \tilde{v}_{k+1}\|^2.\end{aligned}$$

The estimation in Lemma 2.1 is obtained by observing that the last term on the right-hand side can be bounded as follows:

$$\begin{aligned}&\frac{1}{\gamma}\|\mu_{k+1} - \tilde{\mu}_{k+1}\|^2 \\ &\quad = \gamma\lambda_k^2\|A(x_{k+1} - x_k) + B(y_{k+1} - y_k) - (z_{k+1} - z_k)\| \\ &\quad \le 3\,\lambda_k^2\,\gamma\,\left(\|A\|^2\|x_{k+1} - x_k\|^2 + \|B\|^2\|y_{k+1} - y_k\|^2 + \|z_{k+1} - z_k\|^2\right)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\delta}|\nu_{k+1} - \tilde{\nu}_{k+1}|^2 &= \delta\lambda_k^2|p(z_{k+1}) - p(z_k)|^2 \\ &\le \delta\, L_p^2\lambda_k^2\|z_{k+1} - z_k\|^2.\end{aligned}$$

For the last part, it suffices to observe that the coefficients in E_k are bounded from below away from zero under (H2). □

Appendix C: Proof of Theorem 2.3

Given w_k, we denote by w^e_{k+1} the iterate we would obtain from w_k if we had $\alpha_k = \beta_k = 0$. From Lemma 2.1 and Hypothesis (H2), there exists $\sigma > 0$ such that

$$\|w^e_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + 2\lambda_k \left[\mathcal{L}\left(u^e_{k+1}, \bar{v}\right) - \mathcal{L}\left(\bar{u}, \bar{v}\right)\right] + \sigma\|u^e_{k+1} - u_k\|^2 + \|\tilde{v}_{k+1} - v_k\|^2 \leq 0$$

for every $\bar{w} \in \mathcal{S}$. Since $\mathcal{L}(\bar{u}, \bar{v}) \leq \mathcal{L}(u, \bar{v})$ for every u, we obtain

$$\|w^e_{k+1} - \bar{w}\|^2 + \sigma\|u^e_{k+1} - u_k\|^2 + \|\tilde{v}_{k+1} - v_k\|^2 \leq \|w_k - \bar{w}\|^2. \tag{34}$$

In particular, we have

$$\|w_{k+1} - \bar{w}\| \leq \|w_{k+1} - w^e_{k+1}\| + \|w^e_{k+1} - \bar{w}\| \leq \|w_{k+1} - w^e_{k+1}\| + \|w_k - \bar{w}\|. \tag{35}$$

But

$$\begin{aligned}
\|w_{k+1} - w^e_{k+1}\|^2 &= \|y_{k+1} - y^e_{k+1}\|^2 + \|z_{k+1} - z^e_{k+1}\|^2 + \frac{1}{\gamma}\|\mu_{k+1} - \mu^e_{k+1}\|^2 \\
&\quad + \frac{1}{\delta}\|\nu_{k+1} - \nu^e_{k+1}\|^2 \\
&= \|y_{k+1} - y^e_{k+1}\|^2 + \|z_{k+1} - z^e_{k+1}\|^2 + \gamma\lambda_k^2\|B(y_{k+1} - y^e_{k+1}) \\
&\quad - (z_{k+1} - z^e_{k+1})\|^2 + \delta\lambda_k^2\|p(z_{k+1}) - p(z^e_{k+1})\|^2 \\
&\leq \left(1 + 2\gamma\Lambda^2\|B\|^2\right)\|y_{k+1} - y^e_{k+1}\|^2 \\
&\quad + \left(1 + 2\gamma\Lambda^2 + \delta\Lambda^2 L_p^2\right)\|z_{k+1} - z^e_{k+1}\|^2 \\
&\leq 2\left(1 + 2\gamma\Lambda^2\|B\|^2\right)\lambda_k\alpha_k + 2\left(1 + 2\gamma\Lambda^2 + \delta\Lambda^2 L_p^2\right)\lambda_k\beta_k,
\end{aligned}$$

by Lemma A.5. It follows that

$$\|w_{k+1} - w^e_{k+1}\| \leq C\left(\sqrt{\alpha_k} + \sqrt{\beta_k}\right) \tag{36}$$

for some $C > 0$. From (35), it ensues that

$$\|w_{k+1} - \bar{w}\| \leq \|w_k - \bar{w}\| + C\left(\sqrt{\alpha_k} + \sqrt{\beta_k}\right). \tag{37}$$

By hypothesis, $(\alpha_k), (\beta_k) \in \ell^{\frac{1}{2}}$. From (37) and Lemma A.1, we deduce that $\|w_k - \bar{w}\|$ converges as $k \to +\infty$. Using (36) and that $\alpha_k, \beta_k \to 0$, we have also that

$$\|w_{k+1} - w^e_{k+1}\| \to 0.$$

The rest of the proof is essentially the same as that of Theorem 2.1. □

Appendix D: Proof of Lemma 3.1

First, by (11), (B1') and the cocoercivity of M, we obtain

$$\left\langle -\frac{x_{k+1}-x_k}{\lambda_k} - A^*\tilde{\mu}_{k+1} + A^*\bar{\mu},\ x_k - \bar{x} \right\rangle \geq \theta \left\| -\frac{x_{k+1}-x_k}{\lambda_k} - A^*\tilde{\mu}_{k+1} + A^*\bar{\mu} \right\|^2 .$$

After some manipulations, we have

$$\begin{aligned} &\|x_{k+1}-\bar{x}\|^2 - \|x_k-\bar{x}\|^2 + 2\lambda_k \langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_k-\bar{x}\right)\rangle \\ &\quad \leq \|x_{k+1}-x_k\|^2 - \frac{2\theta}{\lambda_k}\|x_{k+1}-x_k+\lambda_k A^*\left(\tilde{\mu}_{k+1}-\bar{\mu}\right)\|^2, \end{aligned}$$

and then

$$\begin{aligned} &\|x_{k+1}-\bar{x}\|^2 - \|x_k-\bar{x}\|^2 + \left(\frac{2\theta}{\lambda_k}-1\right)\|x_{k+1}-x_k\|^2 \\ &\quad + 2\lambda_k\langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_{k+1}-\bar{x}\right)\rangle - 2\lambda_k\langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_{k+1}-x_k\right)\rangle \\ &\quad \leq -4\theta\langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_{k+1}-x_k\right)\rangle - 2\theta\lambda_k\|A^*\left(\tilde{\mu}_{k+1}-\bar{\mu}\right)\|^2. \end{aligned}$$

So, for every $\epsilon > 0$, by Young's inequality, we obtain

$$\begin{aligned} &\|x_{k+1}-\bar{x}\|^2 - \|x_k-\bar{x}\|^2 + \left(\frac{2\theta}{\lambda_k}-1\right)\|x_{k+1}-x_k\|^2 + 2\lambda_k\langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_{k+1}-\bar{x}\right)\rangle \\ &\quad \leq (4\theta-2\lambda_k)\left[\frac{1}{2\epsilon}\|x_{k+1}-x_k\|^2 + \frac{\epsilon}{2}\|A^*\left(\tilde{\mu}_{k+1}-\bar{\mu}\right)\|^2\right] - 2\theta\lambda_k\|A^*\left(\tilde{\mu}_{k+1}-\bar{\mu}\right)\|^2. \end{aligned}$$

Now choose $\epsilon = \frac{4\theta\lambda_k}{4\theta-2\lambda_k} > 0$. Defining $\eta = \frac{2\theta-\Lambda}{2\theta} > 0$, we have

$$\frac{2\theta}{\lambda_k} - 1 - \frac{(4\theta-2\lambda_k)^2}{8\theta\lambda_k} = \frac{2\theta-\lambda_k}{2\theta} \geq \eta.$$

It follows that

$$\begin{aligned} &\|x_{k+1}-\bar{x}\|^2 - \|x_k-\bar{x}\|^2 + \eta\|x_{k+1}-x_k\|^2 \\ &\quad + 2\lambda_k\langle \tilde{\mu}_{k+1}-\bar{\mu}, A\left(x_{k+1}-\bar{x}\right)\rangle \leq 0. \end{aligned} \tag{38}$$

Next, use (11), (B2') and the definition of N_{α_k} to obtain

$$\begin{aligned} &\|y_{k+1}-\bar{y}\|^2 - \|y_k-\bar{y}\|^2 + \|y_{k+1}-y_k\|^2 + 2\lambda_k\langle \tilde{\mu}_{k+1}-\bar{\mu}, \\ &B\left(y_{k+1}-\bar{y}\right)\rangle \leq 2\lambda_k\alpha_k. \end{aligned} \tag{39}$$

Further, from (11), we deduce that

$$2\lambda_k\langle -\bar{\mu}, \bar{z}-z\rangle_m - 2\lambda_k\bar{\nu}\left[p\left(z\right)-p\left(\bar{z}\right)\right] \leq 0$$

for every $z \in Z$. Using (B3'), we obtain

$$\|z_{k+1} - \bar{z}\|^2 - \|z_k - \bar{z}\|^2 + \|z_{k+1} - z_k\|^2 + 2\lambda_k \langle \tilde{\mu}_{k+1}, -(z_{k+1} - \bar{z})\rangle \\ +2\lambda_k \tilde{\nu}_{k+1} \left[p(z_{k+1}) - p(\bar{z})\right] \leq 2\lambda_k \beta_k.$$

This inequality, along with (38) and (39) together give

$$\|u_{k+1} - \bar{u}\|^2 - \|u_k - \bar{u}\|^2 + \eta \|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 + \|z_{k+1} - z_k\|^2 \\ +2\lambda_k \langle \tilde{\mu}_{k+1} - \bar{\mu},\ A(x_{k+1} - \bar{x}) + B(y_{k+1} - \bar{y}) - (z_{k+1} - \bar{z})\rangle \\ +2\lambda_k (\tilde{\nu}_{k+1} - \bar{\nu}) \left[p(z_{k+1}) - p(\bar{z})\right] \leq 2\lambda_k (\alpha_k + \beta_k). \tag{40}$$

Now we focus on the terms involving the inner products. On the one hand, we have

$$\langle \tilde{\mu}_{k+1} - \bar{\mu},\ A(x_{k+1} - \bar{x}) + B(y_{k+1} - \bar{y}) - (z_{k+1} - \bar{z})\rangle \\ = \frac{1}{2\gamma\lambda_k} \left[\|\tilde{\mu}_{k+1} - \mu_k\|^2 - \|\tilde{\mu}_{k+1} - \mu_{k+1}\|^2 + \|\mu_{k+1} - \bar{\mu}\|^2 - \|\mu_k - \bar{\mu}\|^2\right].$$

On the other,

$$(\tilde{\nu}_{k+1} - \bar{\nu}) \left[p(z_{k+1}) - p(\bar{z})\right] = \frac{1}{2\delta\lambda_k} \Big[\|\tilde{\nu}_{k+1} - \nu_k\|^2 - \|\tilde{\nu}_{k+1} - \nu_{k+1}\|^2 \\ + \|\nu_{k+1} - \bar{\nu}\|^2 - \|\nu_k - \bar{\nu}\|^2\Big].$$

So (40) can be rewritten as

$$\|w_{k+1} - \bar{w}\|^2 - \|w_k - \bar{w}\|^2 + \|\tilde{v}_{k+1} - v_k\|^2 + \eta \|x_{k+1} - x_k\|^2 \\ +\|y_{k+1} - y_k\|^2 + \|z_{k+1} - z_k\|^2 \leq 2\lambda_k (\alpha_k + \beta_k) + \|v_{k+1} - \tilde{v}_{k+1}\|^2.$$

The last term on the right-hand side is estimated as at the end of the proof of Lemma 2.1. The additional estimation using (H2') is also straightforward. □

References

1. Goldstein, A.A.: Convex programming in Hilbert space. Bull. Am. Math. Soc. **70**, 709–710 (1964)
2. Levitin, E.S., Polyak, B.T.: Constrained minimization problems. USSR Comput. Math. Math. Phys. **6**, 1–50 (1966)
3. Auslender, A., Crouzeix, J.-P., Fedit, P.: Penalty-proximal methods in convex programming. J. Optim. Theory Appl. **55**(1), 1–21 (1987)
4. Cominetti, R., Courdurier, M.: Coupling general penalty schemes for convex programming with the steepest descent method and the proximal point algorithm. SIAM J. Optim. **13**, 745–765 (2002)
5. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite elements approximations. Comp. Math. Appl. **2**, 17–40 (1976)
6. Glowinski, R., Marrocco, A.: Sur l'approximation par éléments finis d'ordre un, et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. RAIRO **2**, 41–76 (1975)
7. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrange Methods: Applications to the Solution of Boundary Valued Problems, pp. 299–331. North-Holland, Amsterdam (1983)

8. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer Series in Computational Physics. Springer, New York (1984)
9. Glowinski, R., Le Tallec, P.: Augmented lagrangian and operator-splitting methods in nonlinear mechanics. SIAM Stud. Appl. Math. **9** (1989)
10. Gol'shtein, E.G., Tret'yakov, N.V.: Modified Lagrangians in convex programming and their generalizations. Point-to-set maps and mathematical programming. Math. Program. Stud. **10**, 86–97 (1979)
11. Rockafellar, R.T.: Augmented lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116 (1976)
12. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Program. **64**, 81–101 (1994)
13. Tseng, P.: Alternating projection-proximal methods for convex programming and variational inequalities. SIAM J. Optim. **7**, 951–965 (1997)
14. Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. Pac. J. Optim. **5**(1), 17–37 (2009)
15. Xu, M.H.: Proximal alternating directions method for structured variational inequalities. J. Optim. Theory Appl. **134**, 107–117 (2007)
16. Han, D., Yuan, X., Zhang, W.: An augmented Lagrangian based parallel splitting method for separable convex minimization with applications to image processing. Math. Comp. **83**(289), 2263–2291 (2014)
17. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2000)
18. Ben-Tal, A., Zibulevsky, M.: Penalty/barrier multiplier methods for convex programming problems. SIAM J. Optim. **7**(2), 347–366 (1997)
19. Frankel, P., Peypouquet, J.: Lagrangian-penalization algorithm for constrained optimization and variational inequalities. Set-Valued Var. Anal. **20**(2), 169–185 (2012)
20. Necoara, I., Suykens, J.A.K.: Interior-point Lagrangian decomposition method for separable convex optimization. J. Optim. Theory Appl. **143**(3), 567–588 (2009)
21. Sarmiento, O., Papa Quiroz, E.A., Oliveira, P.R.: A proximal multiplier method for separable convex minimization. Optimization **65**(2), 501–537 (2016)
22. Cauchy, A.-L.: Méthode générale pour la résolution des systèmes d'équations simultanées. C. R. Acad. Sci. Paris **25**, 536–538 (1847)
23. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Franaise Informat. Recherche Oprationnelle **4**, 154–158 (1970)
24. Brézis, H., Lions, P.L.: Produits infinis de résolvantes. Isr. J. Math. **29**, 329–345 (1978)
25. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
26. Peypouquet, J.: Convex Optimization in Normed Spaces: Theory, Methods and Examples. Springer, New York (2015)
27. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Sov. Math. Dokl. **27**, 372–376 (1983)
28. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824–1834 (2016)
29. Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**(3), 968–982 (2015)
30. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. NIPS **27**, 2510–2518 (2014)
31. Martinez-Legaz, J.E., Théra, M.: ε-subdifferentials in terms of subdifferentials. Set-Valued Anal. **4**(4), 327–332 (1996)
32. Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. **5**(2), 159–180 (1997)
33. Brézis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext, Springer, New York (2011)
34. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, CMS Books in Mathematics, New York (2011)
35. Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
36. Raguet, H., Fadili, J., Peyré, G.: Generalized forward–backward splitting. SIAM J. Imaging Sci. **6**(3), 1199–1226 (2013)

37. Hansen, P.C., Nagy, J.G., O'Leary, D.P.: Deblurring images: matrices, spectra, and filtering. Fundam. Algorithms 3, SIAM (2006)
38. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
39. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**, 259–268 (1992)
40. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Proc. Numer. Math. **76**, 167–188 (1997)
41. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Proc. **18**(11), 2419–2434 (2009)
42. Le, T., Chartrand, R., Asaki, T.J.: A variational approach to reconstructing images corrupted by Poisson noise. J. Math. Imaging Vision **27**(3), 257–263 (2007)
43. Sawatzky, A., Brune, C., Muller, J., Burger, M.: Total variation processing of images with poisson statistics. In: International Conference on Computer Analysis of Images and Patterns, pp. 533–540 (2009)
44. De los Reyes, J.C., Schönlieb, C.B.: Image denoising: learning the noise model via nonsmooth PDE-constrained optimization. Inverse Probl. Imaging **7**(4), 1183–1214 (2013)
45. Gonzlez-Andrade, S.: A preconditioned descent algorithm for variational inequalities of the second kind involving the p-Laplacian operator. Comput. Optim. Appl. **66**(1), 123–162 (2017)
46. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math Soc. **73**(4), 591–597 (1967)
47. Knopp, K.: Theory and Application of Infinite Series. Courier Corporation (2013)