A Primal-Dual Line Search Method and Applications in Image Processing

Pantelis Sopasakis*, Andreas Themelis*, Johan Suykens* and Panagiotis Patrinos*
*KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS) & OPTEC, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

***Abstract*—Operator splitting algorithms are enjoying wide acceptance in signal processing for their ability to solve generic convex optimization problems exploiting their structure and leading to efficient implementations. These algorithms are instances of the Krasnosel'skiĭ-Mann scheme for finding fixed points of averaged operators. Despite their popularity, however, operator splitting algorithms are sensitive to ill conditioning and often converge slowly. In this paper we propose a line search primal-dual method to accelerate and robustify the Chambolle-Pock algorithm based on SuperMann: a recent extension of the Krasnosel'skiĭ-Mann algorithmic scheme. We discuss the convergence properties of this new algorithm and we showcase its strengths on the problem of image denoising using the anisotropic total variation regularization.**

I. Introduction

A. Background and Motivation

Operator splitting methods have become popular in numerical optimization for their ability to handle abstract linear operators and nonsmooth terms and to lead to algorithmic formulations which require only simple steps without the need to perform matrix factorizations or solve linear systems [1]. As a result they scale gracefully with the problem dimension and they are applicable to large-scale and huge-scale problems as they are amenble to parallelization (such as on graphics processing units) [2]. Because of these advantages, they have attracted remarkable attention in signal processing [3]–[5].

Their main limitation, however, is that they are sensitive to ill conditioning and although under certain conditions they converge linearly, in practice they often perform poorly — as a result, they are only suitable for small-to-medium-accuracy solutions. Moreover, their tuning parameters are selected prior to the execution of the algorithm.

In this paper we propose a line search method to accelerate the popular Chambolle-Pock optimization method, we discuss its convergence properties and apply it for the solution of an image denoising problem [6].

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information. Research Council KUL: CoE PFV/10/002 (OPTEC); PhD/Postdoc grants. Flemish Government: FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grants. IWT: projects: SBO POM (100031); PhD/Postdoc grants. iMinds Medical Information Technologies SBO 2014. Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017). The work of P. Patrinos was supported by KU Leuven Research Council BOF/STG-15-043.

B. Mathematical preliminaries

Throughout the paper, $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\mathcal{K}, \langle\cdot,\cdot\rangle_{\mathcal{K}})$ are two Hilbert spaces. We denote by $\mathcal{H}\oplus\mathcal{K}$ their direct sum, endowed with the inner product $\langle(x,y),(\xi,\eta)\rangle_{\mathcal{H}\oplus\mathcal{K}} = \langle x,\xi\rangle_{\mathcal{H}} + \langle y,\eta\rangle_{\mathcal{K}}$. We indicate with $\mathcal{B}(\mathcal{H};\mathcal{K})$ the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{K}$, writing simply $\mathcal{B}(\mathcal{H})$ if $\mathcal{K} = \mathcal{H}$. With $\|L\| \coloneqq \mathbf{sup}_{x\in\mathcal{H}} \frac{\|Lx\|_{\mathcal{K}}}{\|x\|_{\mathcal{H}}}$ we denote the *norm* of $L \in \mathcal{B}(\mathcal{H};\mathcal{K})$, whereas with L^* its *adjoint*. We say that $L \in \mathcal{B}(\mathcal{H})$ is *self-adjoint* if $L = L^*$, and *skew-adjoint* if $L = -L^*$.

The extended real line is denoted as $\overline{\mathbb{R}} = \mathbb{R}\cup\{\infty\}$. The *Fenchel conjugate* of a proper, closed, convex function $h : \mathcal{H} \to \overline{\mathbb{R}}$ is $h^* : \mathcal{H} \to \overline{\mathbb{R}}$, defined as $h^*(y) = \mathbf{sup}_{x\in\mathcal{H}}\{\langle x,y\rangle - h(x)\}$. Properties of conjugate functions are well described for example in [7]–[9]. Among these we recall that f^* is also proper, closed and convex, and $y \in \partial h(x) \Leftrightarrow x \in \partial h^*(y)$ [7, Thm. 23.5].

The identity operator is denoted as I. Given an operator $T : \mathcal{H} \to \mathcal{H}$, $\mathbf{fix}\, T \coloneqq \{x \in \mathcal{H} \mid Tx = x\}$ and $\mathbf{zer}\, R \coloneqq \{x \in \mathcal{H} \mid Tx = 0\}$ are the sets of its *fixed points* and *zeros*, respectively. Moreover, we say that T is *firmly nonexpansive* if for every $x, y \in \mathcal{H}$

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(I - T)x - (I - T)y\|^2.$$

The projector on a nonempty closed convex set $C \subseteq \mathcal{H}$, denoted as $\mathbf{\Pi}_C$, is firmly nonexpansive [8, Prop. 4.8].

The *graph* of a set-valued operator $F : \mathcal{H} \rightrightarrows \mathcal{H}$ is $\mathbf{gph}(F) \coloneqq \{(x,\xi) \mid \xi \in F(x)\}$. F is said to be *monotone* if $\langle \xi - \eta, x - y\rangle \geq 0$ for all $(x,\xi),(y,\eta) \in \mathbf{gph}(F)$. F is *maximally monotone* if it is monotone and there exists no monotone operator F' such that $\mathbf{gph}(F) \subsetneq \mathbf{gph}(F')$, in which case the *resolvent* $J_F \coloneqq (I + F)^{-1}$ is (single-valued and) firmly nonexpansive [8, Prop. 23.7]. This is the case of the *subdifferential* $\partial h(x) \coloneqq \{v \in \mathcal{H} \mid h(y) \geq h(x) + \langle v, y - x\rangle\ \forall y \in \mathcal{H}\}$ of any proper convex and lower semicontinuous function $h : \mathcal{H} \to \overline{\mathbb{R}}$, in which case, for any $\gamma > 0$ the resolvent of $\gamma\partial h$ is the *proximal mapping* of h with stepsize γ, namely

$$J_{\gamma\partial h} = \mathbf{prox}_{\gamma h} \coloneqq \underset{z\in\mathcal{H}}{\mathbf{argmin}}\left\{h(z) + \tfrac{1}{2\gamma}\|z - \cdot\,\|^2\right\} \quad (1)$$

see [8, Thm.s 12.27, 20.40 and Prop. 16.34].

For a set $C \subseteq \mathcal{H}$, we denote its *strong relative interior* as $\mathbf{sri}\, C$; the strong relative interior coincides with the standard relative interior in finite-dimensional spaces [8, Fact 6.14 (i)].

In what follows, for a (single-valued) operator T we use the convenient notation Tx instead of $T(x)$. Similarly, we shall denote the composition of two operators T_1 and T_2 as T_1T_2 instead of $T_1 \circ T_2$.

ISBN 978-0-9928626-7-1 © EURASIP 2017

II. The Chambolle-Pock Method

Given $L \in \mathcal{B}(\mathcal{H};\mathcal{K})$ and two proper, closed, convex and proximable functions $f : \mathcal{H} \to \overline{\mathbb{R}}$ and $g : \mathcal{K} \to \overline{\mathbb{R}}$ such that $\mathbf{dom}(f + g \circ L) \neq \emptyset$, consider the optimization problem

$$\underset{(x,z)\in\mathcal{H}\oplus\mathcal{K}}{\mathbf{minimize}}\, f(x) + g(z) \quad \text{s.t.} \quad Lx = z. \tag{P}$$

The Fenchel dual problem of (P) is

$$\underset{u\in\mathcal{K}}{\mathbf{minimize}}\, f^*(-L^*u) + g^*(u). \tag{D}$$

Under *strict feasibility*, *i.e.*, if $0 \in \mathbf{sri}(\mathbf{dom}\, g - L(\mathbf{dom}\, f))$, strong duality holds and the set of dual optima is nonempty, if $\mathcal{H}$ is finite-dimensional it is compact [9, Cor. 31.2.1] and any primal-dual solution $(x_\star, u_\star)$ to (P)-(D) is characterized by the optimality conditions

$$0 \in F(x_\star, u_\star), \quad \text{where} \quad F := \begin{bmatrix}\partial f & \\ & \partial g^*\end{bmatrix} + \begin{bmatrix} & L^* \\ -L & \end{bmatrix} \tag{2a}$$

as it follows from [8, Thm. 19.1].

Here, denoting the primal-dual space as $\mathcal{Z} := \mathcal{H} \oplus \mathcal{K}$, $F : \mathcal{Z} \to \mathcal{Z}$ is the sum of a maximally monotone and a skew-adjoint operator. Although J_F may be hard to compute, for some invertible $P \in \mathcal{B}(\mathcal{Z})$ the resolvent of the *preconditioned operator* $P^{-1}F$ leads to a simple algorithmic scheme. To this end, define the self-adjoint operator $P : \mathcal{Z} \to \mathcal{Z}$ as

$$P := \begin{bmatrix}\frac{1}{\alpha_1}I & -L^* \\ -L & \frac{1}{\alpha_2}I\end{bmatrix}, \tag{2b}$$

which is positive definite provided that $\alpha_1\alpha_2\|L\|^2 < 1$. In this case, P induces on $\mathcal{Z}$ the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_P := \langle\,\cdot\,, P\,\cdot\,\rangle_{\mathcal{Z}}$. In what follows, the space $\mathcal{Z}$ is equipped with this inner product and the corresponding norm $\|z\|_P = \sqrt{\langle z, z\rangle_P}$.

The monotone inclusion $F(z) \ni 0$ can be equivalently written as $P^{-1}F(z) \ni 0$. The application of the *proximal point algorithm* on $P^{-1}F$, namely fixed-point iterations of its resolvent, yields the *preconditioned proximal point method* (PPPM) which uses the mapping $T : \mathcal{Z} \ni z \mapsto \bar{z} \in \mathcal{Z}$ implicitly defined via $(I + P^{-1}F)\bar{z} \ni z$ or, equivalently,

$$(P + F)\bar{z} \ni Pz. \tag{3}$$

In the metric induced by P, the PPPM operator T is firmly nonexpansive because it is the resolvent of a maximally monotone operator.

Using the definitions of F and P, the PPPM iterates become

$$(I + \alpha_1\partial f)\bar{x} \ni x - \alpha_1 L^*u, \tag{4a}$$

$$(I + \alpha_1\partial g^*)\bar{u} \ni u + \alpha_2 L(2\bar{x} - x). \tag{4b}$$

This is the Chambolle-Pock method, which prescribes fixed-point iterations $z^+ = z + \lambda(Tz - z)$ of the (firmly nonexpansive) operator $T = (P + F)^{-1}P$; using (1), this is easily seen to be equivalent to the steps of Algorithm 1. Notice that due to the Moreau identity [8, Thm. 14.3], $\mathbf{prox}_{g^*}$ can be computed in terms of $\mathbf{prox}_g$.

Note that the zeros of F are exactly the fixed points of T, that is $F(z) \ni 0$ if and only if $T(z) = z$. Similarly, defining the *residual* operator $R : \mathcal{Z} \to \mathcal{Z}$ associated with T

$$R = I - T, \tag{5}$$

which is also firmly nonexpansive, the problem of determining a fixed point of T can be seen as the problem of finding a zero of its residual R.

Algorithm 1 Chambolle-Pock

Require : $\alpha_1, \alpha_2 > 0$ s.t. $\alpha_1\alpha_2\|L\|^2 < 1$, $\lambda \in (0, 2)$, $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{K}$

1: **for** $k = 0, 1, \ldots$ **do**
2: $\bar{x}_k \leftarrow \mathbf{prox}_{\alpha_1 f}(x_k - \alpha_1 L^* u_k)$
3: $\bar{u}_k \leftarrow \mathbf{prox}_{\alpha_2 g^*}(u_k + \alpha_2 L(2\bar{x}_k - x_k))$
4: $(x_{k+1}, u_{k+1}) \leftarrow (1 - \lambda)(x_k, u_k) + \lambda(\bar{x}_k, \bar{u}_k)$

III. From Krasnosel'skiĭ-Mann to SuperMann

Let $T : \mathcal{H} \to \mathcal{H}$ be a firmly nonexpansive operator with $\mathbf{fix}\, T \neq \emptyset$. Given $\lambda \in (0, 2)$, the Krasnosel'skiĭ-Mann (KM) algorithm for finding a fixed point of T is

$$z^+ = z + \lambda(Tz - z). \tag{6}$$

The KM algorithm has been the locomotive of numerical convex optimization and encompasses all operator-based methods such as the proximal point algorithm, the forward-backward and forward-backward-forward splittings and three-term splittings such as the Combettes-Pesquet and Vũ-Condat and the all-embracing asymmetric forward-backward algorithms [8], [10], [11]. Despite its simplicity and popularity, the convergence rate of this scheme is — at best — Q-linear, let alone it is sensitive to ill-conditioning and likely to exhibit slow convergence.

Recently, [12] proposed *SuperMann*: an algorithmic framework based on a modification of (6) which exploits the interpretation of the KM step as a (relaxed) projection, namely

$$z^+ = (1 - \lambda)z + \lambda\, \mathbf{\Pi}_{C_z} z, \tag{7}$$

where C_z is the halfspace

$$C_z = \left\{y \in \mathcal{H} \mid \|Rz\|^2 - \langle Rz, z - y\rangle \leq 0\right\}.$$

The key idea is the replacement of the halfspace C_z with a different C_w in (7) which leads to generalized KM (GKM) steps. More precisely, given a candidate update direction $d \in \mathcal{H}$, w is taken as $w = z + \tau d$ where $\tau > 0$ is such that

$$\rho := \langle Rw, Rw - \tau d\rangle \geq \sigma\|Rw\|\|Rx\|. \tag{8a}$$

The GKM step can be explicitly written as

$$z^+ = z - \lambda\frac{\rho}{\|Rw\|^2}Rw. \tag{8b}$$

At the same, to encourage favorable updates, an *educated update* of the form $z^+ = z + \tau d$ is accepted if the norm of the candidate residual $\|Rz\|$ is *sufficiently smaller* than the norm of the current one, that is $\|Rw\| \leq c\|Rz\|$ for some $c \in (0, 1)$. This combination of GKM and educated updates gives rise to the *SuperMann algorithm*, where GKM steps are used as globalization strategy for fast iterative methods $z^+ = z + d$ for solving the nonlinear equation $Rz = 0$. As we shall see in Section VI, when "good" update directions d are employed, SuperMann leads to faster convergence and allows

ISBN 978-0-9928626-7-1 © EURASIP 2017

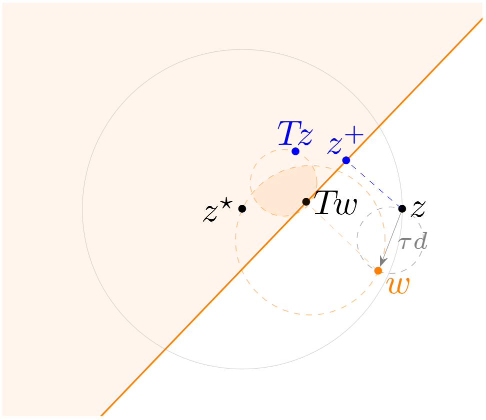

Figure 1: *Starting from a point z, we move along a direction d with step size τ arriving at a point $w = z + \tau d$. This point defines the halfspace C_w. Because of firm nonexpansiveness of T, Tw lies within the intersection of the two dashed orange disks. For adequately small τ, as shown here, $z \notin C_w$. The point $z^+ = \mathbf{\Pi}_{C_w} z$ is then closer to $z^\star \in \mathbf{fix}\, T$ than z.*

the attainment of higher precision. We will elaborate on a possible choice of d in Section V.

In the SuperMann scheme the step length τ is determined via a simple backtracking line search and is selected so that either (i) the fixed-point residual Rw decreases sufficiently or (ii) the next step approaches $\mathbf{fix}\, T$ sufficiently as shown in Figure 1. The former enforces fast convergence whereas the latter, which is referred to as a *safeguard step*, guarantees global convergence by enforcing a quasi-Fejér monotonicity condition.

IV. SuperMann applied to Chambolle-Pock

In Section II we showed that Chambolle-Pock algorithm consists in fixed-point iterations of the operator $T = (P + F)^{-1}P$, where F and P are as in (2), which is firmly nonexpansive with respect to the metric $\langle \cdot\,, \cdot \rangle_P$. As such, it fits into the framework of KM schemes and can be robustified and enhanced with the SuperMann scheme [12]. This is the goal of this section, namely applying SuperMann to find a zero of the Chambolle-Pock residual R (5). Throughout this section we operate on space $(\mathcal{Z}, \langle \cdot\,, \cdot \rangle_P)$.

SuperMann uses exactly the same oracle as the original algorithm, that is, it requires evaluations of $\mathbf{prox}_{\alpha_1 f}$, $\mathbf{prox}_{\alpha_2 g^*}$ and of the linear operator L and its adjoint (see Algorithm 2).

Throughout lines 2 to 4 we compute a step $(\bar{x}_k, \bar{u}_k)$ of the Chambolle-Pock operator T and the corresponding fixed-point residual $r_k = Rz_k$. In order to preclude certain invocations to L and L^*, prior to the linear search (lines 7 to 17) we may precompute the quantities $l_k^x = Lx_k$, $l_k^u = L^* u_k$, $\delta_k^u = L^* d_k^u$, $\delta_k^x = L d_k^x$ and $\tilde{\delta}_k = \delta_k^x - \alpha_1 \delta_k^u$. Then, $\bar{w}_k$ can be evaluated as

$$\bar{w}_k = \mathbf{prox}_{\alpha_1 f}(x_k - \alpha_1 l_k^u + \tau_k \tilde{\delta}_k).$$

Similarly, $\bar{v}_k$ can be evaluated as

$$\bar{v}_k = \mathbf{prox}_{\alpha_2 g^*}(w_k^u + \alpha_2(2\bar{l}_k^x - l_k^x - \tau_k \delta_k^x)),$$

where $\bar{l}_k^x = L\bar{w}_k^x$. In the special yet frequent case when f is quadratic, $\mathbf{prox}_{\alpha_1 f}$ is linear and further computational savings can be obtained by precomputing $\mathbf{prox}_{\alpha_1 f}(\tilde{\delta}_k)$ and $\mathbf{prox}_{\alpha_1 f}(x_k - \alpha_1 l_k^u)$.

Algorithm 2 SuperMann on Chambolle-Pock

Require : $\alpha_1, \alpha_2 > 0$ s.t. $\alpha_1\alpha_2\|L\|^2 < 1$, $\lambda \in (0, 2)$, $c, q, \sigma \in (0, 1)$, $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{K}$
Initialize: $r_{\text{safe}} \leftarrow \infty$
1: **for** $k = 0, 1, \ldots$ **do**
2: $\bar{x}_k \leftarrow \mathbf{prox}_{\alpha_1 f}(x_k - \alpha_1 L^* u_k)$
3: $\bar{u}_k \leftarrow \mathbf{prox}_{\alpha_2 g^*}(u_k + \alpha_2 L(2\bar{x}_k - x_k))$
4: $r_k \leftarrow (x_k - \bar{x}_k, u_k - \bar{u}_k)$
5: Choose $d_k^x \in \mathcal{H}$, $d_k^u \in \mathcal{K}$ and $\tau_k \leftarrow 1$, loop $\leftarrow$ true
6: **while** loop **do**
7: $(w_k, v_k) \leftarrow (x_k, u_k) + \tau_k(d_k^x, d_k^u)$
8: $\bar{w}_k \leftarrow \mathbf{prox}_{\alpha_1 f}(w_k - \alpha_1 L^* v_k)$
9: $\bar{v}_k \leftarrow \mathbf{prox}_{\alpha_2 g^*}(v_k + \alpha_2 L(2\bar{w}_k - w_k))$
10: $\tilde{r}_k \leftarrow (w_k, v_k) - (\bar{w}_k, \bar{v}_k)$
11: **if** $\|r_k\|_P \le r_{\text{safe}}$ **and** $\|\tilde{r}_k\|_P \le c\|r_k\|_P$ **then**
12: $(x_{k+1}, u_{k+1}) \leftarrow (w_k, v_k)$
13: $r_{\text{safe}} \leftarrow \|\tilde{r}_k\|_P + q^k$, loop $\leftarrow$ false
14: **else if** $\underbrace{\langle \tilde{r}_k, \tilde{r}_k - \tau_k \binom{d_k^x}{d_k^u} \rangle_P}_{:=\rho_k} \ge \sigma \|r_k\|_P \|\tilde{r}_k\|_P$ **then**
15: $\eta_k \leftarrow \lambda\rho_k / \|\tilde{r}_k\|_P^2$, loop $\leftarrow$ false
16: $(x_{k+1}, u_{k+1}) \leftarrow (x_k, u_k) - \eta_k \tilde{r}_k$
else
17: $\tau_k \leftarrow \tau_k/2$

Lastly, for the sake of computing scalar products and norms in the required metric, the number of calls to the operator P can be optimized by storing two additional vectors, namely Pr_k and $P\tilde{r}_k$.

In line 11 we accept an *educated update* (w_k, v_k) provided that the norm of its residual $\tilde{r}_k$ is adequately smaller that the norm of r_k and that the norm of the latter has not significantly increased compared to that at the previous iteration. In line 14, we accept a Fejérian update following (8).

For any choice of direction $d_k = (d_k^x, d_k^u) \in \mathcal{Z}$, the line search in Algorithm 2 is guaranteed to finish in finitely many iterations, the resulting sequence converges weakly to a fixed point $z^\star = (x^\star, u^\star)$ in $\mathbf{fix}\, T$, and $(Rz_k)_{k\in\mathbb{N}}$ is square-summable.

V. Quasi-Newtonian directions

The choice of good directions is essential for the fast convergence of the algorithm. As suggested in [12], a good selection consists in d^k being computed with a modified Broyden's method. Namely, letting $u \otimes v$ denote the rank-one operator $x \mapsto \langle v, x \rangle_P u$, starting from an invertible $B_0 \in \mathcal{B}(\mathcal{Z})$

$$B_{k+1} = B_k + \tfrac{\vartheta_k}{\|s_k\|_P^2}(y_k - B_k s_k) \otimes s_k.$$

Here,

$$\begin{aligned} s_k &= (\bar{w}_k, \bar{v}_k) - (x_k, u_k) \\ y_k &= R(\bar{w}_k, \bar{v}_k) - R(x_k, u_k) \\ \gamma_k &= \tfrac{\langle B_k^{-1} y_k, s_k \rangle_P}{\|s_k\|_P^2} \end{aligned} \tag{9a}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

and

$$\vartheta_k = \begin{cases} 1 & \text{if } |\gamma_k| \geq \bar{\vartheta} \\ \frac{1-\mathbf{sgn}(\gamma_k)\bar{\vartheta}}{1-\gamma_k} & \text{otherwise} \end{cases} \tag{9b}$$

with the convention $\mathbf{sgn}(0) = 1$. Alternatively, using the Sherman-Morrison-Woodbury identity we can directly compute $H_k = B_k^{-1}$ as

$$H_{k+1} = H_k + \tfrac{1}{\langle \tilde{s}_k, s_k \rangle_P}(s_k - \tilde{s}_k) \otimes (H_k^* s_k) \tag{9c}$$

where

$$\tilde{s}_k = (1-\vartheta_k)s_k + \vartheta_k H_k y_k. \tag{9d}$$

This obviates the storage and inversion of B_k as we can directly operate with their inverses H_k. We now have all the ingredients to prove the efficiency of Algorithm 2.

Theorem V.1 (see [12])**.** *Suppose that $\mathcal{H}$ and $\mathcal{K}$ are finite dimensional, and consider the iterates generated by Algorithm 2 applied to* (P)*, with directions $(d_k^x, d_k^u) = -H_k r_k$, $(H_k)_{k\in\mathbb{N}}$ being selected with Broyden's method* (9)*. Suppose that the sequence of Broyden's operators $(H_k)_{k\in\mathbb{N}}$ remains bounded. Then,*

(i) (x_k, u_k) converge to a primal-dual solution $(x_\star, u_\star)$ and the residuals r_k converge to 0 square-summably;

(ii) if R is metrically subregular at $(x_\star, u_\star)$, i.e., there exist $\epsilon, \kappa > 0$ such that $\mathbf{dist}((x,u), \mathbf{zer}\, R) \leq \kappa \|R(x,u)\|$ for all $\|(x,u) - (x^\star, u^\star)\| \leq \epsilon$, then the convergence is linear;

(iii) if, additionally, the residual R is calmly semidifferentiable at $(x_\star, u_\star)$, then the convergence is superlinear.

In image processing applications, problem sizes prohibit the use of full Broyden methods where one needs to store and update linear operators H_k. At the expense of losing certain theoretical properties of full-memory Broyden methods — such as superlinear convergence under certain assumptions — limited-memory variants, where one needs to store only m past pairs (s_k, y_k), lead to a considerable decrease in memory requirements.

In Algorithm 3 we propose a *restarted* limited-memory Broyden method tailored for the updates (9). A buffer of fixed maximum capacity M is required, where we store the pairs $(s_k, \tilde{s}_k)$. A similar remark regarding the minimization of calls to P as discussed for Algorithm 2 applies to this inner procedure. Specifically, the number of calls to operator P can be reduced to one per execution of Algorithm 3 by simply including the vectors $P\tilde{s}_k$ in the memory buffer.

VI. Image Denoising

A common problem in image processing is that of retrieving an unknown image $x \in \mathbb{R}^{m\times n}$ (of height m and width n pixels) from an observed image y which has been distorted by noise [13]. Such problems can be formulated as optimization problems of the form

$$\underset{x\in\Omega}{\mathbf{minimize}}\ \tfrac{1}{2}\|x-y\|^2 + \mu \mathbf{TV}_{\ell_1}(x), \tag{10}$$

where $\Omega = [0, 255]^{m\times n}$ and $\mathbf{TV}_{\ell_1}$ is the *anisotropic total variation* regularizer defined as $\mathbf{TV}_{\ell_1}(x) = \|Lx\|_1$, where

Algorithm 3 Restarted Broyden for the computation of directions $d_k \in \mathcal{Z}$

$d_k \leftarrow -Rz_k,\ \ \tilde{s}_{k-1} \leftarrow y_{k-1}$

$M' = k \ \ \mathbf{mod}\ M$

for $i = k - M' \ldots k-2$ **do**

  $\tilde{s}_{k-1} \leftarrow \tilde{s}_{k-1} + \frac{\langle s_i, \tilde{s}_{k-1}\rangle_P}{\langle s_i, \tilde{s}_i\rangle_P}(s_i - \tilde{s}_i)$

  $d_k \leftarrow d_k + \frac{\langle s_i, d_k\rangle_P}{\langle s_i, \tilde{s}_i\rangle_P}(s_i - \tilde{s}_i)$

Compute ϑ_{k-1} as in (9b)

$\tilde{s}_{k-1} \leftarrow (1-\vartheta_{k-1})s_{k-1} + \vartheta_{k-1}\tilde{s}_{k-1}$

$d_k \leftarrow d_k + \frac{\langle s_{k-1}, d_k\rangle_P}{\langle s_{k-1}, \tilde{s}_{k-1}\rangle_P}(s_{k-1} - \tilde{s}_{k-1})$

if $M' = M$ **then**

  Empty the buffer

else

  Append $(s_k, \tilde{s}_k)$ into the buffer

L is the linear operator $L : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times 2n}$ with $Lx = (L_h x, L_v x)$, L_h and L_v are the *horizontal* and *vertical discrete gradient* operators and $\|\cdot\|_1$ is the ℓ_1 norm [14]. The use of $\mathbf{TV}_{\ell_1}$ as a regularizer is based on the principle that noisy images exhibit larger changes in the values of adjacent pixels. For (10), operator F defined in (2a) has a polyhedral graph, therefore it satisfies the metric subregularity condition required by Theorem V.1 [15], so SuperMann converges R-linearly.

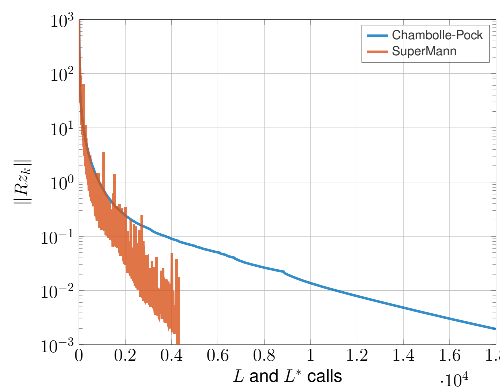

Figure 2: *Convergence of Chambolle-Pock and SuperMann: $\|Rz_k\|$ vs number of calls of operators L and L^* with $\mu = 0.05$.*

In (10) we look for an image x which is close to the given noisy image y (in the squared Euclidean distance) and has a low total variation. The regularization weight μ can be chosen via statistical methods [16]. For $x \in \mathbb{R}^{m\times n}$, operators L_h and L_v are defined as

$$(L_h x)_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j} & \text{for } j = 1\ldots n-1 \\ 0 & \text{for } j = n \end{cases}$$

for $i = 1 \ldots m$, and

$$(L_v x)_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j} & \text{for } i = 1\ldots m-1 \\ 0 & \text{for } i = m \end{cases}$$

for $j = 1 \ldots n$. It is known that $\|L\| = \sqrt{8}$ and that L^* is the *discrete divergence* operator [17].

For the problem in (10) we define

ISBN 978-0-9928626-7-1 © EURASIP 2017

Figure 3: *(Left) Original image* (640×480 *pixels), (Middle) Image distorted with zero-mean Gaussian noise with variance* 0.025 *(PSNR* -31dB*), (Right) Denoised image with* $\mu = 24.5$.

1) the primal Hilbert space $\mathcal{H} := \mathbb{R}^{m \times n}$ and the dual space $\mathcal{K} := \mathbb{R}^{m \times 2n}$,
2) the term $f(x) = \frac{1}{2}\|x - y\|^2 + \delta_{[0,255]^{m\times n}}(x)$ whose proximal map is $\mathbf{prox}_{\alpha_1 f}(v) = \mathbf{\Pi}_\Omega[(1+\alpha_1)^{-1}(v + \alpha_1 y)]$, and
3) the term $g(z) = \mu\|z\|_1$ with $\mathbf{prox}_{\gamma g}(v)_i = \mathbf{sgn}(v_i)[|v_i| - \gamma\mu]_+$.

We apply the aforementioned methodology for the filtering of Gaussian noise which has been added to the image shown in Figure 3 (Left) leading to a distorted image (Middle). Parameters α_1 and α_2 are taken equal to $0.95/\sqrt{8} \approx 0.3359$ and $\lambda = 1$. For the restarted Broyden method, we chose $\bar{\vartheta} = 0.5$ and memory $M = 10$. For the line search in Algorithm 2 we set $\sigma = 1 - c = 10^{-4}$ and $q = 10^{-1}$.

As shown in Figure 2 for $\mu = 24.5$, the proposed algorithm converges considerably faster than Chambolle-Pock with the former converging with termination criterion $\|Rz_k\| < 10^{-3}$ in 1129 iterations (4302 calls of L and L^*) and the latter converging in 10527 iterations (21054 calls of L and L^*). In Figure 4 (Left) we show the number of calls to L and L^* for different values of μ — SuperMann is consistenty faster than Chambolle-Pock. In order to evaluate how μ affects the quality of the produced image, computing solutions of (10) for several values of μ is often desired. In Figure 4 (Right) we show how μ affects the PSNR of the denoised image with respect to the original image.

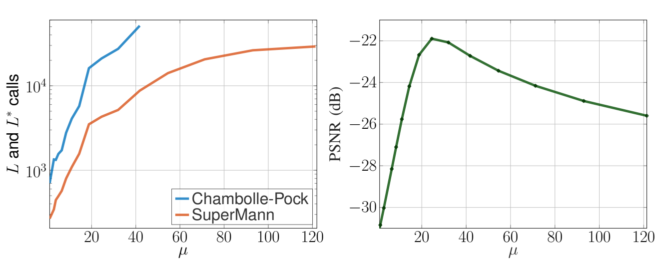

Figure 4: *(Left) Number of calls to* L *and* L^* *vs* μ*; for values of* μ *larger than 42 the Chambolle-Pock algorithm did not converge within* $5 \cdot 10^4$ *iterations, (Right) Peak signal-to-noise ratio (PSNR) vs* μ.

VII. Conclusions

We proposed a primal-dual line search algorithm to accelerate the Chambolle-Pock method which only involves invocations to $\mathbf{prox}_{\alpha_1 f}$, $\mathbf{prox}_{\alpha_2 g^*}$, L and L^*. We tested the proposed method on the problem of image denoising using the anisotropic total variation regularization demonstrating that the new algorithm exhibits considerably faster convergence.

In future work we will further exploit the structure of operator T to compute semi-smooth Newton directions to achieve even faster convergence results in the spirit of [18], [19].

References

[1] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, jan 2014.

[2] A. K. Sampathirao, P. Sopasakis, A. Bemporad, and P. Patrinos, "GPU-accelerated stochastic predictive control of drinking water networks," *IEEE TCST*, 2017. [Online]. Available: http://arxiv.org/abs/1604.01074

[3] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. New York, NY: Springer, 2011, pp. 185–212.

[4] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multisc. Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, jan 2005.

[5] M. Fadili and J. Starck, "Monotone operator splitting for optimization problems in sparse recovery," in *IEEE ICIP*, 2009, pp. 1461–1464.

[6] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[7] R. T. Rockafellar, *Convex Analysis*. Princeton university press, 1997.

[8] H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[9] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer, 2011, vol. 317.

[10] E. Ryu and S. Boyd, "A primer on monotone operator methods," *Appl. Comp. Math. Int. J.*, vol. 15, no. 1, 2016.

[11] P. Latafat and P. Patrinos, "Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators," *Computational Optimization and Applications*, pp. 1–37, 2017. [Online]. Available: http://dx.doi.org/10.1007/s10589-017-9909-6

[12] A. Themelis and P. Patrinos, "SuperMann: a superlinearly convergent algorithm for finding fixed points of nonexpansive operators," *arXiv preprint arxiv:1609.06955*, 2016.

[13] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[14] M. Grasmair and F. Lenzen, "Anisotropic total variation filtering," *Appl. Math. Optim.*, vol. 62, no. 3, pp. 323–339, 2010.

[15] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings: A View from Variational Analysis*. Springer, 2014.

[16] M. Lucchese, I. Frosio, and N. Borghese, "Optimal choice of regularization parameter in image denoising," in *ICIAP, Lect. Not. Comp. Sci.*, vol. 6978. Springer, 2011.

[17] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[18] P. Patrinos, L. Stella, and A. Bemporad, "Forward-backward truncated Newton methods for convex composite optimization," *arXiv:1402.6655v2 [math.OC]*, Feb. 2014.

[19] P. Sopasakis, N. Freris, and P. Patrinos, "Accelerated reconstruction of a compressively sampled data stream," in *24th European Signal Processing conference (EUSIPCO)*, Budapest, Hugary, 2016.

ISBN 978-0-9928626-7-1