ON THE CONVERGENCE RATES OF PROXIMAL SPLITTING ALGORITHMS

Jingwei Liang[1]*, Jalal M. Fadili*[1] *and Gabriel Peyré*[2]

[1] GREYC, CNRS-ENSICAEN [2] CEREMADE, CNRS-Paris-Dauphine

ABSTRACT

In this work, we first provide iteration–complexity bounds (pointwise and ergodic) for the inexact Krasnosel'skiĭ–Mann iteration built from nonexpansive operators. Moreover, under an appropriate regularity assumption on the fixed point operator, local linear convergence rate is also established. These results are then applied to analyze the convergence rate of various proximal splitting methods in the literature, which includes the Forward–Backward, generalized Forward–Backward, Douglas–Rachford, ADMM and some primal–dual splitting methods. For these algorithms, we develop easily verifiable termination criteria for finding an approximate solution, which is a generalization of the termination criterion for the classical gradient descent method. We illustrate the usefulness of our results on a large class of problems in signal and image processing.

Index Terms— Convex optimization, Proximal splitting, Convergence rates, Inverse problems.

1. INTRODUCTION

Problem statement and overview Many convex optimization problems in image and signal processing can be solved via the inexact Krasnosel'skiĭ–Mann iteration which is defined by

$$x^{k+1} = x^k + \lambda_k(Tx^k + \varepsilon^k - x^k), \tag{1.1}$$

where $T : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator on a Hilbert space $\mathcal{H}$, $(\lambda_k)_{k\in\mathbb{N}} \in]0,1[$, and for $x^k \in \mathcal{H}$, $\varepsilon^k \in \mathcal{H}$ is the error when computing Tx^k.

One simple instance of such an algorithm is that of relaxed inexact gradient descent for solving

$$\min_{x\in\mathcal{H}} \ f(x),$$

where f is proper convex and has β^{-1}–Lipschitz continuous gradient, in which case $T = \mathrm{Id} - \gamma\nabla f$, for $\gamma \in]0, 2\beta[$. In this scenario, the error ε^k is that when evaluating ∇f at x^k.

Many structured convex optimization problems boil down to implementing an iteration that can be cast in the form of (1.1) for an appropriate T. Consider for example

$$\min_{x\in\mathcal{H}} \big\{\Phi(x) = f(x) + \textstyle\sum_{i=1}^n h_i \circ L_i(x)\big\}, \tag{1.2}$$

where $f \in \Gamma_0(\mathcal{H})$ has β^{-1}–Lipschitz continuous gradient, $h_i \in \Gamma_0(\mathcal{H})$, L_i is a bounded linear operator, $\Gamma_0(\mathcal{H})$ is the class of lower semicontinuous, proper, convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$. Assume that some appropriate domain qualification conditions are verified for (1.2) to be well–posed. Problem (1.2) has been considered in e.g. [1, 2, 3], who proposed iterative schemes in the form (1.1).

In the last decades, based on the notion of the proximity operator [4], and assuming that the functions h_i are simple (*i.e.* their proximity operators are easily computable), a wide range of proximal splitting algorithms have been proposed to solve problems of the form (1.2). One can cite for instance the Forward–Backward splitting method (FBS) [5] valid for $n = 1$ and $L_1 = \mathrm{Id}$, the Douglas–Rachford splitting method (DRS) [6] that applies for $f = 0$ and $L_i = \mathrm{Id}$, generalized Forward–Backward splitting method (GFB) [3] when $L_i = \mathrm{Id}$, or primal–dual splitting methods. See [7] for a comprehensive account.

Contributions In this paper, we first establish iteration complexity bounds of the inexact relaxed fixed point iteration (1.1). Under a regularity assumption on $\mathrm{Id} - T$, a local linear convergence rate is established. We then build upon these results to provide rates for several proximal splitting algorithms, from which an easily verifiable termination criterion for finding an approximate solution will be given. For space limitations, the proofs of the results can be found in the long version of the paper [8].

2. ITERATION COMPLEXITY BOUNDS

Denote $T' = \mathrm{Id} - T$, $\mathrm{fix}T$ is the set of fixed points of T, and

$$e^k = T'x^k.$$

Let d_0 be the distance from a starting point x^0 to the solution set $\mathrm{fix}T$. Denote $\tau_k = \lambda_k(1-\lambda_k)$, $\underline{\tau} = \inf_{k\in\mathbb{N}} \tau_k$, $\overline{\tau} = \sup_{k\in\mathbb{N}} \tau_k$, $\nu_1 = 2\sup_{k\in\mathbb{N}} \|T_k x^k - x^\star\| + \sup_{k\in\mathbb{N}} \lambda_k \|\varepsilon^k\|$, $\nu_2 = 2\sup_{k\in\mathbb{N}} \|e^k - e^{k+1}\|$, where $x^\star \in \mathrm{fix}T$. Denote ℓ^1_+ the set of summable sequences in $[0, +\infty[$.

This work has been supported by the ERC project SIGMA-Vision and l'Institut Universitaire de France.

Theorem 2.1 (Pointwise iteration complexity bound). *Assume that*

(a) $\mathrm{fix}T \neq \emptyset$;

(b) $0 < \inf_{k\in\mathbb{N}} \lambda_k \leq \sup_{k\in\mathbb{N}} \lambda_k < 1$;

(c) $\big((k+1)\|\varepsilon^k\|\big)_{k\in\mathbb{N}} \in \ell^1_+$.

Let $C_1 = \nu_1 \sum_{j\in\mathbb{N}} \lambda_j \|\varepsilon^j\| + \nu_2 \overline{\tau} \sum_{\ell\in\mathbb{N}} (\ell+1)\|\varepsilon^\ell\| < +\infty$, *then*

$$\|e^k\| \leq \sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}}.$$

Remark 2.2. When the fixed point iteration (1.1) is exact, we get

$$\|e^k\| \leq \frac{d_0}{\sqrt{\sum_{j=0}^k \tau_j}},$$

which recovers the result of [9, Propositon 11].

Next we present the ergodic iteration complexity bound of (1.1), define $\Lambda_k = \sum_{j=0}^k \lambda_j$, and $\bar{e}^k = \frac{1}{\Lambda_k}\sum_{j=0}^k \lambda_j e^j$.

Theorem 2.3 (Ergodic iteration complexity bound). *Assume that condition* (a) *in Theorem 2.1 holds, and* $C_3 = \sum_{j\in\mathbb{N}} \lambda_j \|\varepsilon^j\| < +\infty$. *Then,*

$$\|\bar{e}^k\| \leq \frac{2(d_0 + C_3)}{\Lambda_k}.$$

If $\inf_{k\in\mathbb{N}} \lambda_k > 0$, then $\Lambda_k = c(k+1)$ for some constant $c > 0$, and we get $O(1/k)$ rate.

A special class of nonexpansive operators is the α–averaged operators [7] for $\alpha \in]0,1[$. The above two complexity bounds obviously apply, where now $\lambda_k \in]0, \frac{1}{\alpha}[$ and condition Theorem (b) is changed accordingly.

3. LOCAL LINEAR RATE

We now turn to a local convergence analysis of (1.1).

Definition 3.1 (Metric subregularity [10]). A set–valued mapping $F : \mathcal{H} \to 2^{\mathcal{H}}$ is called metrically subregular at $\tilde{x}$ for $\tilde{u} \in F(\tilde{x})$ if there exists $\kappa \geq 0$ along with neighbourhood $\mathcal{X}$ of $\tilde{x}$ such that

$$d(x, F^{-1}\tilde{u}) \leq \kappa\, d(\tilde{u}, Fx), \quad \forall x \in \mathcal{X}. \tag{3.1}$$

The infimum of κ for which this holds is the modulus of metric subregularity, denoted by $\mathrm{subreg}(F; \tilde{x}|\tilde{u})$. The absence of metric regularity is signaled by $\mathrm{subreg}(F; \tilde{x}|\tilde{u}) = +\infty$.

Metric subregularity implies that, for any $x \in \mathcal{X}$, $d(\tilde{u}, Fx)$ is bounded below. The metric (sub)regularity of multifunctions plays a crucial role in modern variational analysis and optimization. These properties are a key to study the stability of solutions of generalized equations, see the dedicated monograph [10].

Let's specialize this definition to the operator T' and $\tilde{u} = 0$. T' is single–valued and $T'^{-1}(0) = \mathrm{fix}T$. Thus if T' is metrically subregular at some $x^\star \in \mathrm{fix}T$ for 0, then from (3.1) we have

$$d(x, \mathrm{fix}T) \leq \kappa \|T'x\|, \quad \forall x \in \mathcal{X}.$$

Metric subregularity implies that (3.1) gives an estimate for how far a point x is from being the fixed point set of T in terms of the residual $\|x - Tx\|$. This is the rationale behind using such a regularity assumption on the operator T' to quantify the convergence rate on $d(x^k, \mathrm{fix}T)$. Thus, starting from $x^0 \in \mathcal{H}$, and by virtue of Theorem 2.1, one can recover a $O(1/\sqrt{k})$ rate on $d(x^k, \mathrm{fix}T)$. In fact, we can do even better as is shown in the following result. We use the shorthand notation $d_k = d(x^k, \mathrm{fix}T)$.

Theorem 3.2 (Local linear rate). *Assume that conditions* (a)-(b) *in Theorem 2.1 hold, and* T' *is metrically subregular at* $x^\star \in \mathrm{fix}T$ *for* 0, *with* $\kappa > \mathrm{subreg}(T'; x^\star|0)$. *If* C_3 *is sufficiently small and there exists a ball* $\mathbb{B}_a(x^\star)$ *such that*

$$\mathbb{B}_{(a+C_3)}(x^\star) \subseteq \mathcal{X}.$$

Then, for any starting point $x^0 \in \mathbb{B}_a(x^\star)$, *there holds:*

(i) $\big(d_k^2\big)_{k\in\mathbb{N}} \in \ell^1_+$ *and*

$$d_{k+1}^2 \leq \zeta_k\, d_k^2 + c_k,$$

where $\zeta_k = \begin{cases} 1 - \frac{\tau_k}{\kappa^2}, & \text{if } \tau_k/\kappa^2 \in]0,1] \\ \frac{\kappa^2}{\kappa^2 + \tau_k}, & \text{otherwise} \end{cases} \in [0,1[$, $c_k = \nu_1 \lambda_k \|\varepsilon^k\|$.

(ii) *If* $\varepsilon^k = 0$, *then* $\lim_{k\to+\infty} \sqrt[k]{d_k} < 1$.

(iii) *If* $\mathrm{fix}T = \{x^\star\}$ *and* $\varepsilon^k = 0$, *then* x^k *converges linearly to* $x^\star$.

Remark 3.3. For simplicity, suppose the iteration is exact, and let $x^\star \in \mathrm{fix}T$ such that $d_k = \|x^k - x^\star\|$. Then we have

$$\|e^k\|^2 = \|x^k - x^\star + Tx^\star - Tx^k\|^2 \leq 4d_k^2 \leq 4\zeta^k d_0^2,$$

which means that locally, $\|e^k\|$ also converges linearly to 0.

Again, if T is α–averaged, then we can afford $\lambda_k \in]0, \frac{1}{\alpha}[$ and all statements of Theorem 3.2 remain valid with $\zeta_k = \kappa^2/\big(\kappa^2 + \alpha\lambda_k(1 - \alpha\lambda_k)\big)$.

Since metric regularity [10] implies metric subregularity, equivalent characterizations of the latter and its modulus can be given, for instance in terms of derivative criteria. In particular, as T' is single–valued, its metric regularity holds if it is differentiable on a neighbourhood of $x^\star$ with nonsingular derivatives at x around $x^\star$, and the operator norms of their inverses are uniformly bounded and serve as an estimate of the (sub)regularity modulus κ [11, Theorem 1.2]. Computing κ in practice is however far from obvious in general even in for the differentiable case and these details will be left to a future work.

4. APPLICATIONS TO PROXIMAL SPLITTING

FBS Suppose that $n = 1$ and $L_1 = \mathrm{Id}$ in (1.2). Then FBS with a fixed step–size corresponds to $T = \mathrm{Prox}_{\gamma h_1} \circ (\mathrm{Id} - \gamma \nabla f)$, $\gamma \in]0, 2\beta[$, $\lambda_k \in]0, \frac{4\beta-\gamma}{2\beta}[$, and ε^k is the error when evaluating both the proximity operator and the gradient. In this case, setting $g^k = \frac{1}{\gamma}(x^{k-1} - x^k) - \nabla f(x^{k-1})$, it then follows from Theorem 2.1 that

$$g^k \in \partial h_1(x^k), \text{ and } \|g^k + \nabla f(x^k)\|^2 = O(1/k).$$

In plain words, this means that $O(1/\epsilon)$ iterations are needed to find a pair $\big(x, g \in \partial h_1(x)\big)$ with the termination criterion $\|g + \nabla f(x)\|^2 \leq \epsilon$. This is the best–known complexity bound possessed by first–order methods to solve general non-linear problems (*e.g.* the gradient descent method [12]). From Theorem 2.3, this iteration–complexity improves to $O(1/\sqrt{\epsilon})$ in ergodic sense for the same termination criterion.

GFB Consider now $n > 1$ and $L_i = \mathrm{Id}$ in (1.2). [3] proposed a generalization in a product space $\mathcal{H}^n$ of the FBS scheme to solve (1.2). For the GFB method, T takes a more intricate form, omitted here for space limitation, and ε^k absorbs the error when computing both the proximity operators $\mathrm{Prox}_{\gamma h_i}$ and the gradient. Let γ and λ_k be chosen as for FBS. One can show that $O(1/\epsilon)$ iterations are needed to find a pair $\big((u_i)_{1 \leq i \leq n}, g\big)$ with the termination criterion $\|g + \nabla f(\sum_i u_i/n)\|^2 \leq \epsilon$, where $g \in \sum_{i=1}^n \partial h_i(u_i)$, see [8] for details. Again, ergodic iteration complexity is $O(1/\sqrt{\epsilon})$.

DRS, ADMM When $f = 0$ and $n = 2$, (1.2) can be solved by DRS for $L_1 = L_2 = \mathrm{Id}$, or the Alternating Direction method of Multipliers (ADMM) [13] if e.g. $L_1 = \mathrm{Id}$. ADMM is known to be equivalent to DRS applied to the dual problem [14]. It can be shown that in at most $O(1/\epsilon)$ iterations, DRS finds a subgradient g of $h_1 + h_2$ at x with $\|g\|^2 \leq \epsilon$.

Other splitting schemes Similar complexity bounds can be established for several primal–dual splitting methods, e.g. [1], for solving (1.2) (and even more general). One crucial property of these methods, is that they can be reformulated into the form of (1.1), and the corresponding fixed point operator is α–averaged; see [8] for details.

5. NUMERICAL EXPERIMENTS

In this section, we illustrate the obtained convergence results on two applications, the nonnegative matrix completion (NMC) problem [15], and the principal component pursuit (PCP) problem [16].

5.1. Nonnegative matrix completion

Suppose we observe measurements $y \in \mathbb{R}^p$ of a low rank matrix $X_0 \in \mathbb{R}^{m \times n}$ with nonnegative entries

$$y = \mathcal{A}(X_0) + w$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a measurement operator, and w is the noise. In our experiment here, $\mathcal{A}$ selects p entries of its argument uniformly at random. The matrix completion problem consists in recovering X_0, or an approximation of it, by solving a convex optimization problem, namely the minimization of the nuclear norm [17, 18, 15]. In penalized form, the problem reads

$$\min_{X \in \in \mathbb{R}^{m \times n}} \tfrac{1}{2}\|y - \mathcal{A}(X)\|^2 + \mu \|X\|_* + \iota_+(X), \qquad (5.1)$$

where ι_+ is the indicator function of the nonnegative orthant to account for the nonnegativity constraint, and μ is a regularization parameter chosen proportional to the noise level. Identifying $f(X) = \frac{1}{2}\|y - \mathcal{A}(X)\|^2$, $h_1(X) = \mu\|X\|_*$ and $h_2(X) = \iota_+(X)$, (5.1) is nothing but an instance of (1.2). Both h_1 and h_2 are simple, since $\mathrm{Prox}_{\gamma h_1}(X)$ amounts to soft–thresholding the singular values of X, and $\mathrm{Prox}_{\gamma h_2}(X) = \big(\max(X_{ij}, 0)_{i,j}\big)$ is the projector on the nonnegative orthant. Thus, (5.1) can be solved using e.g. GFB or the primal-dual (PD) scheme of [1].

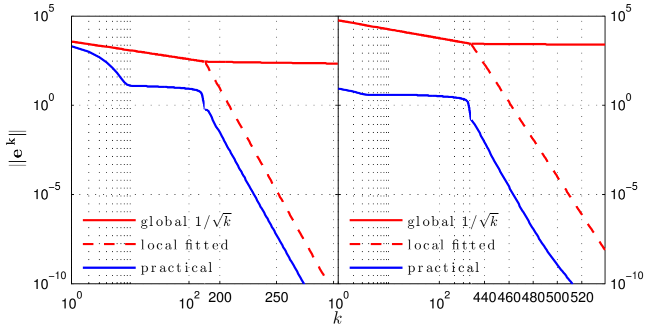

Fig. 1: Observed convergence curve (solid blue) of $\|e^k\|$ for GFB (left) and PD (right) to solve (5.1), theoretical global bound (solid red), and fitted local linear curve.

Fig. 1 displays the observed pointwise convergence rate of $\|e^k\|$ and the theoretical bound computed from Theorem 2.1 for GFB and PD. Note that the first half of the plot is in *log–log* scale while the second one is *log* scale on the ordinate. As expected, $O(1/\sqrt{k})$ convergence rate is observed on the first half. For sufficiently high iteration counter, a linear convergence regime takes over as clearly seen from the second half of the plot. This local behaviour is in consistent with the result of Theorem 3.2. Let us mention that the local linear convergence curve (dashed line) was fitted to the observed one, since the regularity modulus necessary to compute the theoretical rate in Theorem 3.2 is not easy to estimate.

5.2. Principal component pursuit

In this part, we consider the PCP problem [16], and apply it to decompose a video sequence into its background and foreground components. The rationale behind this is that since the background is virtually the same in all frames, if the latter are stacked as columns of a matrix, it is likely to be low–rank (even of rank 1 for perfectly constant background). On the other hand, moving objects appear occasionally on each frame and occupy only a small fraction of it. Thus the corresponding component would be sparse.

Assume that a real matrix M can be written as

$$M = X_{l,0} + X_{s,0} + W,$$

where a $X_{l,0}$ is low–rank, $X_{s,0}$ is sparse and W is a perturbation matrix that accounts for model imperfection. The PCP proposed in [16] attempts to provably recover $(X_{l,0}, X_{s,0})$, to a good approximation, by solving a convex optimization. Here, toward an application to video decomposition, we also add a non-negativity constraint to the low–rank component, which leads to the convex problem ($\|\cdot\|_F$ is the Frobenius norm)

$$\min_{X_l, X_s} \ \tfrac{1}{2}\|M - X_l - X_s\|_F^2 + \mu_1\|X_s\|_1 + \mu_2\|X_l\|_* + \iota_+(X_l), \tag{5.2}$$

One can observe that for fixed X_l, the minimizer of (5.2) is $X_s^\star = \mathrm{Prox}_{\mu_1\|\cdot\|_1}(M - X_l)$. Thus, (5.2) is equivalent to

$$\min_{X_l} \ {}^1(\mu_1\|\cdot\|_1)(M - X_l) + \mu_2\|X_l\|_* + \iota_+(X_l), \tag{5.3}$$

where ${}^1(\mu_1\|\cdot\|_1)(M - X_l) = \min_Z \frac{1}{2}\|M - X_l - Z\|_F^2 + \mu_1\|Z\|_1$ is the Moreau Envelope of $\mu_1\|\cdot\|_1$ of index 1. Since the Moreau envelope is differentiable with a 1–Lipschitz continuous gradient [19], (5.3) is a special instance of (1.2) and can be solved using GFB and PD schemes.

We first used a synthetic example to illustrate the convergence property of GFB and PD. Fig. 2 displays the observed pointwise convergence rate of $\|e^k\|$ and the theoretical one predicted by Theorem 2.1. Both the global and local convergence behaviours are similar to those observed in Fig. 1.

Now we consider a video sequence introduced in [20], whose resolution is 128 × 160, each frame is stacked as a column of the matrix M and 300 frames in total. Hence M is of size 20480 × 300. We then solve (5.3) to decompose the video into its foreground and background.

Fig. 3 displays the observed pointwise convergence rate and the one predicted by Theorem 2.1 but only for GFB for space limitation. Fig. 4 demonstrates the decomposition result of the method. The first column shows 3 frames from the video, the second and third column are the corresponding columns of low–rank component X_l and the sparse component X_s. Notice that X_l correctly recovers the background, while X_s correctly identifies the moving pedestrians or the change of illumination.

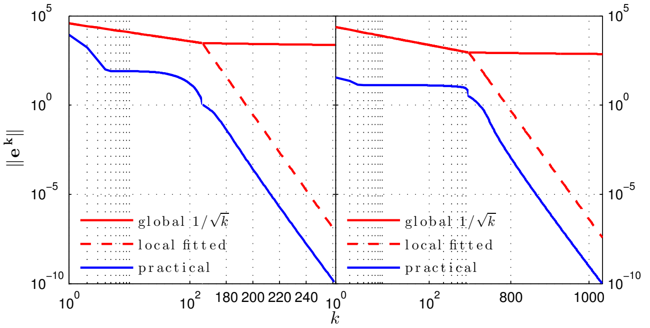

Fig. 2: Observed convergence curve (solid blue) of $\|e^k\|$ for GFB (left) and PD (right) to solve (5.3) with synthetic data, theoretical global bound (solid red), and fitted local linear curve.

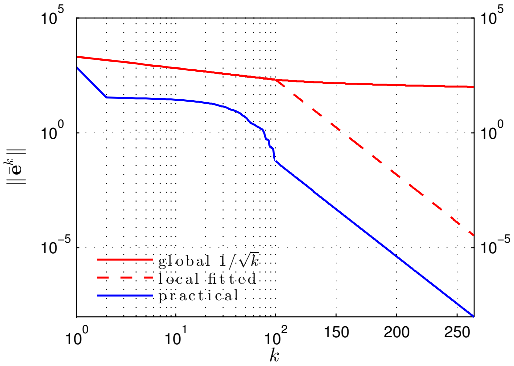

Fig. 3: Observed convergence curve (solid blue) of $\|e^k\|$ for the GFB to solve (5.3) with a real video sequence, theoretical global bound (solid red), and fitted local linear curve.

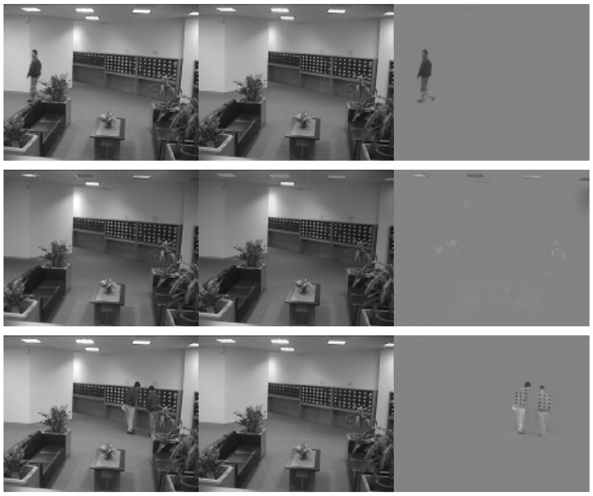

Fig. 4: Left: Original frames of a video sequence (300 frames). Middle and Right: recovered background (low–rank) and foreground (sparse) components.

6. REFERENCES

[1] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[2] P. L. Combettes and J. C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators," *Set-Valued and variational analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[3] H. Raguet, J. M. Fadili, and G. Peyré, "Generalized forward-backward splitting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1199–1226, 2013.

[4] J. J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.

[5] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward–backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[6] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979, 1979.

[7] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2011.

[8] J. Liang, J. M. Fadili, and G. Peyré, "Convergence rates with inexact nonexpansive operators," *arXiv preprint arXiv:1404.4837*, 2014.

[9] R. Cominetti, J. A. Soto, and J. Vaisman, "On the rate of convergence of krasnoselski-mann iterations and their connection with sums of bernoullis," *Israel Journal of Mathematics*, pp. 1–16, 2012.

[10] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings: A view from variational analysis*, Springer, 2009.

[11] A. L. Dontchev, M. Quincampoix, and N. Zlateva, "Aubin criterion for metric regularity," *Journal of Convex Analysis*, vol. 13, pp. 281–297, 2006.

[12] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer, 2004.

[13] D. Gabay, "Applications of the method of multipliers to variational inequalities," in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski, Eds., chapter 15, pp. 299–331. Elsevier, North-Holland, Amsterdam, 1983.

[14] J. Eckstein and D. P. Bertsekas, "On the ddouglas–rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[16] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[17] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[18] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2053–2080, 2010.

[19] J. J. Moreau, "Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires," *CR Acad. Sci. Paris*, vol. 255, pp. 238–240, 1962.

[20] L. Li, W. Huang, Y. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004.