Local Linear Convergence of Douglas–Rachford/ADMM for Low Complexity Regularization

Jingwei Liang*, Jalal M. Fadili*, Gabriel Peyré† and Russell Luke‡
*GREYC, CNRS, ENSICAEN, Université de Caen, Email: {Jingwei.Liang, Jalal.Fadili}@ensicaen.fr
†CNRS, Ceremade, Université Paris-Dauphine, Email: Gabriel.Peyre@ceremade.dauphine.fr
‡Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Email: r.luke@math.uni-goettingen.de

***Abstract*—The Douglas–Rachford (DR) and ADMM algorithms have become popular to solve sparse recovery problems and beyond. The goal of this work is to understand the local convergence behaviour of DR/ADMM which have been observed in practice to exhibit local linear convergence. We show that when the involved functions (resp. their Legendre-Fenchel conjugates) are partly smooth, the DR (resp. ADMM) method identifies their associated active manifolds in finite time. Moreover, when these functions are partly polyhedral, we prove that DR (resp. ADMM) is locally linearly convergent with a rate in terms of the cosine of the Friedrichs angle between the tangent spaces of the two active manifolds. This is illustrated by several concrete examples and supported by numerical experiments.**

I. Introduction

In this work, we consider the problem of solving

$$\min_{x\in\mathbb{R}^n} J(x)+G(x), \tag{1}$$

where both J and G are in $\Gamma_0(\mathbb{R}^n)$, the class of proper, lower semi-continuous and convex functions. We assume that $\mathrm{ri}(\mathrm{dom}J)\cap\mathrm{ri}(\mathrm{dom}G)\neq\emptyset$, where $\mathrm{ri}(C)$ is the relative interior of the nonempty convex set C, and $\mathrm{dom}F$ is the domain of the function F. We also assume that the set of minimizers of (1) is non-empty, and the two functions are simple ($\mathrm{prox}_{\gamma J},\mathrm{prox}_{\gamma G},\gamma>0$, are easy to compute), where the proximity operator is defined, for $\gamma>0$, as $\mathrm{prox}_{\gamma J}(z)=\mathrm{argmin}_{x\in\mathbb{R}^n}\frac{1}{2}\|x-z\|^2+\gamma J(x)$.

An efficient and provably convergent algorithm for solving (1) is the DR method [1], which reads, in its relaxed form,

$$\begin{cases} z^{k+1}=(1-\lambda_k)z^k+\lambda_k B_{\mathrm{DR}}z^k,\\ x^{k+1}=\mathrm{prox}_{\gamma J}z^{k+1}, \end{cases} \tag{2}$$

where $B_{\mathrm{DR}}\overset{\text{def.}}{=}\frac{1}{2}\big((2\mathrm{prox}_{\gamma J}-\mathrm{Id})\circ(2\mathrm{prox}_{\gamma G}-\mathrm{Id})+\mathrm{Id}\big)$, for $\gamma>0$, $\lambda_k\in]0,2]$ with $\sum_{k\in\mathbb{N}}\lambda_k(2-\lambda_k)=+\infty$.

Since the set of minimizers of (1) is non-empty, so is the set of fixed points $\mathrm{fix}(B_{\mathrm{DR}})$ since the former is nothing but $\mathrm{prox}_{\gamma J}(\mathrm{fix}(B_{\mathrm{DR}}))$. See [2] for a more detailed account on DR. Though we focus in the following on DR, all our results readily apply to ADMM since it is the DR applied to the Fenchel dual problem of (1).

II. Partly Smooth Functions and Finite Identification

Beside $J, G\in\Gamma_0(\mathbb{R}^n)$, our central assumption is that J, G are partly smooth functions. Partial smoothness was originally defined in [3]. Here we specialize it to the case of functions in $\Gamma_0(\mathbb{R}^n)$. Denote $\mathrm{par}(C)$ the subspace parallel to the non-empty convex set $C\subset\mathbb{R}^n$.

Definition II.1. Let $J\in\Gamma_0(\mathbb{R}^n)$ and $x\in\mathbb{R}^n$ such that $\partial J(x)\neq\emptyset$. J is *partly smooth at x relative to a set $\mathcal{M}$* containing x if

(Smoothness) $\mathcal{M}$ is a C^2-manifold, $J|_{\mathcal{M}}$ is C^2 around x;
(Sharpness) The tangent space $\mathcal{T}_x(\mathcal{M})=T_x\overset{\text{def.}}{=}\mathrm{par}(\partial J(x))$;
(Continuity) The ∂J is continuous at x relative to $\mathcal{M}$.

The class of partly smooth functions at x relative to $\mathcal{M}$ is denoted as $\mathrm{PSF}_x(\mathcal{M})$.

The class of PSF is very large. Popular examples in signal processing and machine learning include ℓ_1, $\ell_{1,2}$, ℓ_∞ norms, TV semi-norm and nuclear norm, see also [4].

Now define the variable $v^k=\mathrm{prox}_{\gamma G}(2\mathrm{prox}_{\gamma J}-\mathrm{Id})z^k$.

Theorem II.2 (Finite activity identification). *Let the DR scheme* (2) *be used to create a sequence (z^k,x^k,v^k). Then $(z^k,x^k,v^k)\to(z^\star,x^\star,x^\star)$, where $z^\star\in\mathrm{fix}(B_{\mathrm{DR}})$ and $x^\star$ is a global minimizer of* (1)*. Assume that $J\in\mathrm{PSF}_{x^\star}(\mathcal{M}^J_{x^\star})$, $G\in\mathrm{PSF}_{x^\star}(\mathcal{M}^G_{x^\star})$, and*

$$z^\star\in x^\star+\gamma\big(\mathrm{ri}(\partial J(x^\star))\cap\mathrm{ri}(-\partial G(x^\star))\big). \tag{3}$$

Then, the DR scheme has the finite activity identification property, i.e. for all k sufficiently large, $(x^k,v^k)\in\mathcal{M}^J_{x^\star}\times\mathcal{M}^G_{x^\star}$.

Condition (3) implies that $0\in\mathrm{ri}(\partial J(x^\star)+\partial G(x^\star))$, which can be viewed as a geometric generalization of the strict complementarity of non-linear programming. In a compressed sensing scenario, it can be guaranteed for a sufficiently large number of measurements.

III. Local Linear Convergence of DR

We now turn to local linear convergence properties of DR for the case of locally polyhedral functions. This is a subclass of partly smooth functions, whose epigraphs look locally like a polyhedron. In the following, we will refer to the Friedrichs angle between two subspaces V and W, denoted $\theta_F(V,W)$. In fact, $\theta_F(V,W)$ is the $(d+1)$-th principal angle between V and W, where $d=\dim(V\cap W)$, see also [5].

Theorem III.1. *Assume that J and G are locally polyhedral, and the conditions of Theorem II.2 hold with $\lambda_k\equiv\lambda$. Then there exists $K>0$ such that for all $k\geqslant K$,*

$$\|z^k-z^\star\|\leqslant\rho^k\|z_0-z^\star\|, \tag{4}$$

where $\rho=\sqrt{(1-\lambda)^2+\lambda(2-\lambda)\cos^2\theta_F(T^J_{x^\star},T^G_{x^\star})}\in[0,1[$.

This rate is optimal. For the special case of basis pursuit, we recover the result of [6], but with less stringent assumptions.

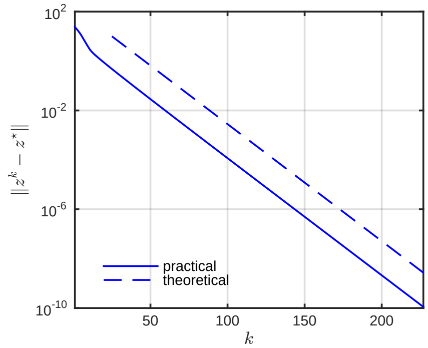
IV. Numerical Experiments

As examples, we consider the ℓ_1, ℓ_∞ norms and the anisotropic TV semi-norm which are all polyhedral, hence partly smooth relative the following subspaces

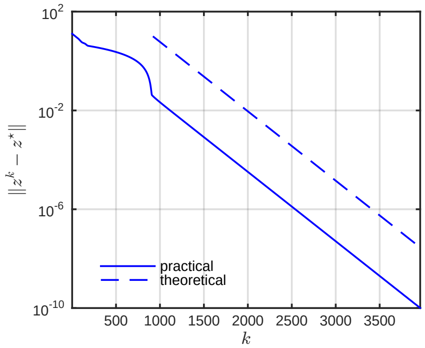
$$\begin{aligned}
&\ell_1: T_x=\{u\in\mathbb{R}^n:\mathrm{supp}(u)\subseteq\mathrm{supp}(x)\},\\
&\ell_\infty: T_x=\big\{u:u_I=rs_I,\ r\in\mathbb{R}\big\},\ s=\mathrm{sign}(x),\ I=\{i:|x_i|=\|x\|_\infty\},\\
&\mathrm{TV}: T_x=\big\{u\in\mathbb{R}^n:\mathrm{supp}(\nabla u)\subseteq I\big\},\ I=\mathrm{supp}(\nabla x),
\end{aligned}$$

where ∇ is the gradient operator.

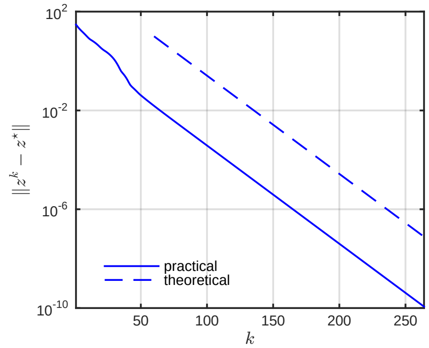
Figure1 displays the observed and predicted convergence profiles of DR when solving several problem instances, including compressed sensing, denoising and inpainting.



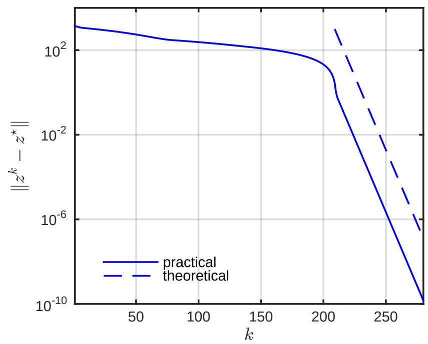
(a) CS ℓ_1-norm



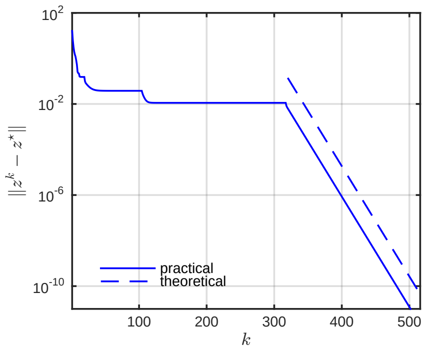
(b) CS ℓ_∞-norm



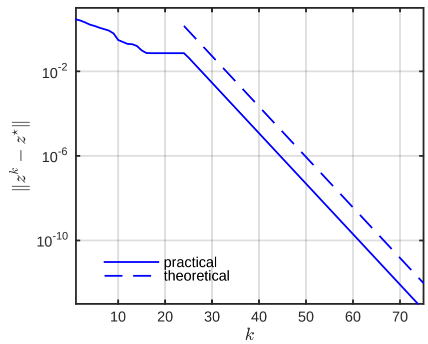
(c) CS TV semi-norm



(d) TV image inpainting



(e) Uniform noise removal



(f) Outliers removal

Fig. 1. Observed (solid) and predicted (dashed) convergence profiles of DR (2) in terms of $\|z^k - z^\star\|$. For the first 4 subfigures, we solve a problem of the form $\min_{x\in\mathbb{R}^n} J(x)$ s.t. $Ax = y$, where A is either drawn randomly from the standard Gaussian ensemble (CS), or random binary (inpainting). (a) CS with $J = \|\cdot\|_1$, $A \in \mathbb{R}^{48\times 128}$. (b) CS with $J = \|\cdot\|_\infty$, $A \in \mathbb{R}^{120\times 128}$. (c) CS with $J = \|\cdot\|_{\mathrm{TV}}$, $A \in \mathbb{R}^{48\times 128}$. (d) TV image inpainting, $A \in \mathbb{R}^{512\times 1024}$. (e) Uniform noise removal by solving $\min_{x\in\mathbb{R}^{128}} \|x\|_{\mathrm{TV}}$ s.t. $\|x - y\|_\infty \leqslant \delta$. (f) Outliers removal by solving $\min_{x\in\mathbb{R}^{128}} \lambda\|x\|_{\mathrm{TV}} + \|x - y\|_1$. The starting points of the dashed lines are the iteration at which the active subspaces are identified.

Acknowledgment

This work has been partly supported by the European Research Council (ERC project SIGMA-Vision). JF was partly supported by Institut Universitaire de France.

References

[1] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16(6):964–979, 1979.

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

[3] A. S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM Journal on Optimization, 13(3):702–725, 2003.

[4] S. Vaiter, G. Peyré, and M. J. Fadili, *Partly smooth regularization of inverse problems*, arXiv:1405.1004, 2014.

[5] H. H. Bauschke, J.Y. Bello Cruz, T.A. Nghia, H. M. Phan, and X. Wang, *Optimal rates of convergence of matrices with applications*, arxiv:1407.0671, 2014.

[6] L. Demanet and X. Zhang, *Eventual linear convergence of the Douglas–Rachford iteration for basis pursuit*, Mathematical Programming, 2013.