53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Douglas-Rachford Splitting: Complexity Estimates and Accelerated Variants

Panagiotis Patrinos and Lorenzo Stella and Alberto Bemporad

Abstract— **We propose a new approach for analyzing convergence of the Douglas-Rachford splitting method for solving convex composite optimization problems. The approach is based on a continuously differentiable function, the *Douglas-Rachford Envelope (DRE)*, whose stationary points correspond to the solutions of the original (possibly nonsmooth) problem. By proving the equivalence between the Douglas-Rachford splitting method and a scaled gradient method applied to the DRE, results from smooth unconstrained optimization are employed to analyze convergence properties of DRS, to tune the method and to derive an accelerated version of it.**

I. Introduction

In this paper we consider convex optimization problems of the form

$$\text{minimize } F(x) = f(x) + g(x), \tag{1}$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper closed convex functions with easily computable *proximal mappings* [1]. We recall that for a convex function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and positive scalar γ, the proximal mapping is defined as

$$\operatorname{prox}_{\gamma h}(x) = \operatorname*{argmin}_z \left\{ h(z) + \tfrac{1}{2\gamma}\|z - x\|^2 \right\}. \tag{2}$$

A well known algorithm for solving (1) is the Douglas-Rachford splitting (DRS) method [2]. In fact, DRS can be applied to solve the more general problem of finding the zero of two maximal monotone operators. In the special case where the corresponding operators are the subdifferentials of f and g, DRS amounts to the following iterations

$$y^k = \operatorname{prox}_{\gamma f}(x^k), \tag{3a}$$
$$z^k = \operatorname{prox}_{\gamma g}(2y^k - x^k), \tag{3b}$$
$$x^{k+1} = x^k + \lambda_k(z^k - y^k), \tag{3c}$$

where $\gamma > 0$ and the stepsizes $\lambda_k \in [0, 2]$ satisfy $\sum_{k\in\mathbb{N}} \lambda_k(2 - \lambda_k) = +\infty$. A typical choice for λ_k is to be set equal to 1 for all k. If the minimum in (1) is attained and the relative interiors of the effective domains of f and g have a point in common, then it is well known that $\{z^k - y^k\}$ converges to 0, and $\{x^k\}$ converges to x such that $\operatorname{prox}_{\gamma f}(x) \in \operatorname{argmin} F$ [3]–[5]. Therefore $\{y^k\}$ and $\{z^k\}$ converge to a solution of (1). This general form of DRS was proposed by [3], [4], where it was shown that DRS is a particular case of the proximal point algorithm [1]. Thus DRS converges under very general assumptions. For example, unlike forward-backward splitting (FBS) [6], it does not require differentiability of one of the two summands and parameter γ can take any positive value.

Another well-known application of DRS is for solving problems of the form

$$\begin{aligned} &\text{minimize } f(x) + g(z), \\ &\text{subject to } Ax + Bz = b. \end{aligned} \tag{4}$$

Applying DRS to the dual of problem (4) leads to the alternating direction method of multipliers (ADMM) [3], [4], [7]. This method has recently received a lot of attention, especially because of its properties with respect to separable objective functions, that make it favorable for large-scale problems and distributed applications [8], [9].

However, when applied to (1), the behavior of DRS is quite different compared to standard optimization methods. For example, unlike FBS, DRS is not a descent method, in that the sequence of cost values $\{F(x^k)\}$ may not be monotone decreasing. This is perhaps one of the main reasons why the convergence rate of DRS has not been well understood and convergence rate results were scarce, until very recently. The first convergence result for DRS appeared in [2]. Translated to the setting of solving (1), under strong convexity and Lipschitz continuity assumptions for f, the sequence $\{x^k\}$ was shown to converge Q-linearly to the (unique) optimal solution of (1). More recently, it was shown that if f is differentiable then the squared residual $\|x^k - \operatorname{prox}_{\gamma g}(x^k - \gamma\nabla f(x^k))\|^2$ converges to zero with sublinear rate of $1/k$ [10]. In [11] convergence rates of order $1/k$ for the objective values are provided implicitly for DRS under the assumption that both f and g have Lipschitz continuous gradients. Under the additional assumption that f is quadratic, the authors of [11] give an accelerated version with convergence rate $1/k^2$. In [12] the authors show global linear convergence for ADMM under a variety of scenarios. Translated in the DRS setting, they require at least f to be strongly convex with Lipschitz continuous gradient. In [13] R-linear convergence of the duality gap and primal cost for multiple splitting ADMM under less stringent assumptions is shown, provided that the stepsizes λ_k are sufficiently small. However, the form of the convergence rate is not very informative, since the bound on the stepsizes depends on constants that are very hard to compute. In [14] it is shown that ADMM converges linearly for quadratic programs with the constraint matrix being full rank. However explicit complexity estimates are only provided for the (infrequent) case where the constraint matrix is full row rank.

The authors are with IMT Institute for Advanced Studies Lucca, Piazza S. Ponziano 6, 55100 Lucca, Italy. Email: {panagiotis.patrinos, lorenzo.stella, alberto.bemporad}@imtlucca.it

This work was partially supported by the European Research Project FP7-EFFINET (Grant no. 318556).

Convergence rates of DRS and ADMM are analyzed under various assumptions in the recent paper [15].

A. Our contribution

In this paper we follow a new approach to the analysis of the convergence properties and complexity estimates of DRS. We show that when f is twice continuously differentiable, then problem (1) is equivalent to computing a stationary point of a continuously differentiable function, the *Douglas-Rachford Envelope (DRE)*. Specifically, DRS is shown to be nothing more than a (scaled) gradient method applied to the DRE. This kind of interpretation is similar to the one offered by the Moreau envelope for the proximal point algorithm and paves the way for deriving new algorithms based on the Douglas-Rachford splitting approach.

A similar idea has been exploited in [16], [17] in order to express another splitting method, the forward-backward splitting, as a gradient method applied to the so-called Forward-Backward Envelope (FBE). There the purpose was use the FBE as a merit function on which to perform Newton-like methods with superlinear local convergence rates to solve non differentiable problems. Here the purpose is instead to analyze the convergence rate properties of Douglas-Rachford splitting by expressing it as a gradient method. Specifically, we show that if f is convex quadratic (but g can still be any convex nonsmooth function) then the DRE is convex with Lipschitz continuous gradient, provided that γ is sufficiently small. This covers a wide variety of problems such as quadratic programs, ℓ_1 least squares, nuclear norm regularized least squares, image restoration/denoising problems involving total variation minimization norm, etc. This observation makes convergence rate analysis of DRS extremely easy, since it allows us to directly apply the well known complexity estimates of the gradient method. Furthermore, we discuss the optimal choice of the parameter γ and of the stepsize λ_k defining the method, and devise a method with faster convergence rates by exploiting the acceleration techniques introduced by Nesterov [18], [19, Sec. 2.2].

The paper is structured as follows. In Section II we define the Douglas-Rachford envelope and analyze its properties, illustrating how DRS is equivalent to a scaled gradient method applied to the DRE. Section III discusses the convergence of Douglas-Rachford splitting in the particular but important case in which f is convex quadratic, where the DRE turns out to be convex. Section IV considers the application of accelerated gradient methods to the DRE to achieve faster convergence rates. Finally, Section V shows experimental results obtained with the proposed methods.

II. Douglas-Rachford Envelope

We indicate by $X_\star$ the set of optimal solutions to problem (1), which we assume to be nonempty. Then $x_\star \in X_\star$ if and only if [5, Cor. 26.3] $x_\star = \operatorname{prox}_{\gamma f}(\tilde{x})$, where $\tilde{x}$ is a solution of

$$\operatorname{prox}_{\gamma g}(2\operatorname{prox}_{\gamma f}(x) - x) - \operatorname{prox}_{\gamma f}(x) = 0. \tag{5}$$

Let $\tilde{X}$ be the set of solutions to (5). Our goal is to find a continuously differentiable function whose set of stationary points is equal to $\tilde{X}$.

Given a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, consider its *Moreau envelope*

$$h^\gamma(x) = \inf_z \left\{ h(z) + \tfrac{1}{2\gamma}\|z - x\|^2 \right\}.$$

It is well known that $h^\gamma : \mathbb{R}^n \to \mathbb{R}$ is differentiable (even if h is nonsmooth) with $(1/\gamma)$-Lipschitz continuous gradient

$$\nabla h^\gamma(x) = \gamma^{-1}(x - \operatorname{prox}_{\gamma h}(x)). \tag{6}$$

By using (6) we can rewrite (5) as

$$\nabla f^\gamma(x) + \nabla g^\gamma(x - 2\gamma\nabla f^\gamma(x)) = 0. \tag{7}$$

From now on we make the extra assumption that f is twice continuously differentiable, with L_f-Lipschitz continuous gradient. We also assume that f has strong convexity modulus equal to $\mu_f \geq 0$, i.e., function $f(x) - \frac{\mu_f}{2}\|x\|^2$ is convex. Notice that we allow μ_f to be equal to zero, including also the case where f is not strongly convex. Due to these assumptions we have

$$\|\nabla^2 f(x)\| \leq L_f, \text{ for all } x \in \mathbb{R}^n. \tag{8}$$

Moreover, from [20, Prop. 4.1, Th. 4.7] the Jacobian of $\operatorname{prox}_{\gamma f}$ and the Hessian of f^γ exist everywhere and are related to each other as follows:

$$\nabla \operatorname{prox}_{\gamma f}(x) = (I + \gamma\nabla^2 f(\operatorname{prox}_{\gamma f}(x)))^{-1}, \tag{9}$$
$$\nabla^2 f^\gamma(x) = \gamma^{-1}(I - \nabla \operatorname{prox}_{\gamma f}(x)). \tag{10}$$

Using (8)-(10) one can easily show that for any $d \in \mathbb{R}^n$

$$\tfrac{\mu_f}{1+\gamma\mu_f}\|d\|^2 \leq d'\nabla^2 f^\gamma(x)d \leq \tfrac{L_f}{1+\gamma L_f}\|d\|^2. \tag{11}$$

In other words, if f is twice continuously differentiable with L_f-Lipschitz continuous gradient then the eigenvalues of the Hessian of its Moreau envelope are bounded uniformly for every $x \in \mathbb{R}^n$.

Next, we premultiply (7) by $(I - 2\gamma\nabla^2 f^\gamma(x))$ to obtain the gradient of what we call the *Douglas-Rachford Envelope (DRE)*:

$$F_\gamma^{\mathrm{DR}}(x) = f^\gamma(x) - \gamma\|\nabla f^\gamma(x)\|^2 + g^\gamma(x - 2\gamma\nabla f^\gamma(x)). \tag{12}$$

If $(I - 2\gamma\nabla^2 f^\gamma(x))$ is nonsingular for every x, then every stationary point of F_γ^{DR} is also an element of $\tilde{X}$, and vice versa. From (11) we obtain

$$\tfrac{1-\gamma L_f}{1+\gamma L_f}\|d\|^2 \leq d'(I - 2\gamma\nabla^2 f^\gamma(x))d \leq \tfrac{1-\gamma\mu_f}{1+\gamma\mu_f}\|d\|^2. \tag{13}$$

Therefore whenever $\gamma < 1/L_f$ or $\gamma > 1/\mu_f$ (in case where $\mu_f > 0$), finding a stationary point of the DRE (12) is equivalent to solving (5).

It is convenient now to introduce the following notation:

$$\begin{aligned} P_\gamma(x) &= \operatorname{prox}_{\gamma f}(x), \\ G_\gamma(x) &= \operatorname{prox}_{\gamma g}(2P_\gamma(x) - x), \\ Z_\gamma(x) &= P_\gamma(x) - G_\gamma(x), \end{aligned}$$

so that condition (5) is expressed as $Z_\gamma(x) = 0$. By (10) we can rewrite $I - 2\gamma\nabla^2 f^\gamma(x) = 2\nabla P_\gamma(x) - I$, therefore the gradient of the DRE can be expressed as

$$\nabla F_\gamma^{\mathrm{DR}}(x) = \gamma^{-1}(2\nabla P_\gamma(x) - I)Z_\gamma(x). \tag{14}$$

The following proposition is instrumental in establishing an equivalence between problem (1) and that of minimizing the DRE.

Proposition 1: The following inequalities hold for any $\gamma > 0$ and $x \in \mathbb{R}^n$:

$$F_\gamma^{\mathrm{DR}}(x) \le F(P_\gamma(x)) - \tfrac{1}{2\gamma}\|Z_\gamma(x)\|^2, \tag{15a}$$

$$F_\gamma^{\mathrm{DR}}(x) \ge F(G_\gamma(x)) + \tfrac{1-\gamma L_f}{2\gamma}\|Z_\gamma(x)\|^2. \tag{15b}$$

Proof: See [21]. ∎

The following fundamental result shows, under the assumption of γ being sufficiently small, that minimizing the DRE, which is real-valued and smooth, is completely equivalent to solving the nonsmooth problem (1). Furthermore, the set of stationary points of the DRE, which may not be convex, coincide with the set of its minimizers.

Theorem 1: If $\gamma \in (0, 1/L_f)$ then

$$\inf\ F = \inf\ F_\gamma^{\mathrm{DR}},$$

$$\operatorname{argmin}\ F = P_\gamma(\operatorname{argmin}\ F_\gamma^{\mathrm{DR}}).$$

Proof: By [5, Cor. 26.3] we know that $x_\star \in X_\star$ if and only if $x_\star = P_\gamma(\tilde{x})$, for some $\tilde{x} \in \tilde{X}$, *i.e.*, with $P_\gamma(\tilde{x}) = G_\gamma(\tilde{x})$. Putting $x = \tilde{x}$ in (15a), (15b) one obtains

$$F_\gamma^{\mathrm{DR}}(\tilde{x}) = F(x_\star).$$

When $\gamma < 1/L_f$, Eq. (15b) implies that for all $x \in \mathbb{R}^n$

$$F_\gamma^{\mathrm{DR}}(x) \ge F(G_\gamma(x)) \ge F(x_\star) = F_\gamma^{\mathrm{DR}}(\tilde{x}), \tag{16}$$

where the last inequality follows from optimality of $x_\star$. Therefore the elements of $\tilde{X}$ are minimizers of F_γ^{DR} and $\inf F = \inf F_\gamma^{\mathrm{DR}}$. They are indeed the only minimizers, for if $x \notin \tilde{X}$ then $Z_\gamma(x) \neq 0$ in (15b), and the first inequality in (16) is strict. ∎

A. DRS as a variable-metric gradient method

In simple words, Theorem 1 tells us that under suitable assumptions on γ, one can employ whichever smooth unconstrained optimization technique for minimizing the DRE and thus solve (1). The resulting algorithm will of course bear a close relationship to DRS since the gradient of the DRE, cf. (14), is inherently related to a step of DRS, cf. (3).

In particular, from the expression (14) for $\nabla F_\gamma^{\mathrm{DR}}$, one observes that Douglas-Rachford splitting can be interpreted as a variable-metric gradient method for minimizing F_γ^{DR}. Specifically, we have that the x-iterates defined by (3) correspond to

$$x^{k+1} = x^k - \lambda_k D^k \nabla F_\gamma^{\mathrm{DR}}(x^k), \tag{17}$$

where

$$D^k = \gamma(2\nabla P_\gamma(x^k) - I)^{-1}. \tag{18}$$

We can then exploit all the well known convergence results of gradient methods to analyze the properties of DRS or propose alternative schemes of it.

B. Connection between DRS and FBS

The DRE reveals an interesting link between Douglas-Rachford splitting and forward-backward splitting, that has remained unnoticed at least to our knowledge. Let us first derive an alternative way of expressing the DRE. Since $P_\gamma(x) = \operatorname{argmin}_z\{f(z) + \frac{1}{2}\|z - x\|^2\}$ satisfies

$$\nabla f(P_\gamma(x)) + \gamma^{-1}(P_\gamma(x) - x) = 0, \tag{19}$$

the gradient of the Moreau envelope of f becomes

$$\nabla f^\gamma(x) = \gamma^{-1}(x - P_\gamma(x)) = \nabla f(P_\gamma(x)). \tag{20}$$

Using (19), (20) in (12) we obtain the following alternative expression for the DRE

$$F_\gamma^{\mathrm{DR}} = f(P_\gamma(x)) - \tfrac{\gamma}{2}\|\nabla f(P_\gamma(x))\|^2 + g^\gamma(2P_\gamma(x) - x), \tag{21}$$

Next, using the definition of g^γ in (21), it is possible to express

$$\begin{aligned} F_\gamma^{\mathrm{DR}}(x) = \min_{z\in\mathbb{R}^n}\{&f(P_\gamma(x)) + \nabla f(P_\gamma(x))'(z - P_\gamma(x)) \\ &+ g(z) + \tfrac{1}{2\gamma}\|z - P_\gamma(x)\|^2\}. \end{aligned} \tag{22}$$

Comparing this with the definition of the forward-backward envelope (FBE) introduced in [16]

$$F_\gamma^{\mathrm{FB}}(x) = \min_{z\in\mathbb{R}^n}\{f(x) + \nabla f(x)'(z - x) + g(z) + \tfrac{1}{2\gamma}\|z - x\|^2\},$$

it is apparent that the DRE at x is equal to the FBE evaluated at $P_\gamma(x)$:

$$F_\gamma^{\mathrm{DR}}(x) = F_\gamma^{\mathrm{FB}}(P_\gamma(x)).$$

Let us recall here that iterates x^{k+1} of FBS are obtained by solving the optimization problem appearing in the definition of FBE for $x = x^k$. Therefore, it can be easily seen that an iteration of DRS corresponds to a forward-backward step applied to $\operatorname{prox}_{\gamma f}(x^k)$ (instead of x^k, as in FBS).

III. Douglas-Rachford Splitting

In case f is convex quadratic, *i.e.*,

$$f(x) = \tfrac{1}{2}x'Qx + q'x,$$

with $Q \in \mathbb{R}^{n\times n}$ symmetric and positive semidefinite and $q \in \mathbb{R}^n$, we have

$$P_\gamma(x) = (I + \gamma Q)^{-1}(x - \gamma q), \tag{23}$$

$$\nabla P_\gamma(x) = (I + \gamma Q)^{-1}. \tag{24}$$

We now have $\mu_f = \lambda_{\min}(Q)$ and $L_f = \lambda_{\max}(Q)$. It turns out that in this case, under the already mentioned assumption $\gamma < 1/L_f$, the DRE is convex.

Theorem 2: Suppose that f is convex quadratic. If $\gamma < 1/L_f$, then F_γ^{DR} is convex with $L_{F_\gamma^{\mathrm{DR}}}$-Lipschitz continuous gradient and convexity modulus $\mu_{F_\gamma^{\mathrm{DR}}}$ given by

$$L_{F_\gamma^{\mathrm{DR}}} = \frac{1 - \gamma\mu_f}{1 + \gamma\mu_f}\gamma^{-1}, \tag{25}$$

$$\mu_{F_\gamma^{\mathrm{DR}}} = \min\left\{\frac{(1 - \gamma\mu_f)\mu_f}{(1 + \gamma\mu_f)^2}, \frac{(1 - \gamma L_f)L_f}{(1 + \gamma L_f)^2}\right\}. \tag{26}$$

Proof: See [21]. ∎

Therefore, under the assumptions of Theorem 2, we can exploit the well known results on the convergence of the gradient method for convex problems. To do so, note that when f is quadratic, P_γ is linear and the scaling matrix D^k defined in (18) is constant, *i.e.*,

$$D^k \equiv D = \gamma(2(I+\gamma Q)^{-1} - I)^{-1}.$$

Consider the linear change of variables $x = Sw$, where $S = D^{1/2}$. Note that

$$\lambda_{\min}(D) = \gamma\frac{1+\gamma\mu_f}{1-\gamma\mu_f}, \quad \lambda_{\max}(D) = \gamma\frac{1+\gamma L_f}{1-\gamma L_f}, \quad (27)$$

so if $\gamma < 1/L_f \le 1/\mu_f$ then matrix D is positive definite and S is well defined.

In the new variable w, the scaled gradient iterations (17) correspond to the (unscaled) gradient method applied to the preconditioned problem

$$\text{minimize } h(w) = F_\gamma^{\mathrm{DR}}(Sw).$$

Indeed, the gradient method applied on h is

$$w^{k+1} = w^k - \lambda_k \nabla h(w^k) \quad (28)$$

Multiplying by S and using $\nabla h(w^k) = S\nabla F_\gamma^{\mathrm{DR}}(Sw^k)$, we obtain

$$x^{k+1} = x^k - \lambda_k D\nabla F_\gamma^{\mathrm{DR}}(x^k).$$

Recalling (14), this becomes

$$x^{k+1} = x^k - \lambda_k Z_\gamma(x^k),$$

which is exactly DRS, cf. (3). From now on we will indicate by $\tilde{w}$ a minimizer of h, so that $\tilde{w} = S\tilde{x}$ for some $\tilde{x} \in \tilde{X}$. From Theorem 2 we know that if $\gamma < 1/L_f$ then F_γ^{DR} is convex with Lipschitz continuous gradient, and so is h. In particular,

$$\mu_h = \lambda_{\min}(D)\mu_{F_\gamma^{\mathrm{DR}}}, \quad (29)$$

$$L_h = \lambda_{\max}(D)L_{F_\gamma^{\mathrm{DR}}} = \frac{1+\gamma L_f}{1-\gamma L_f}. \quad (30)$$

Theorem 3: For convex quadratic f, if $\gamma < 1/L_f$ and

$$\lambda_k = \lambda = (1-\gamma L_f)/(1+\gamma L_f) \quad (31)$$

then the sequence of iterates generated by (3a)-(3c) satisfies

$$F(z^{k+1}) - F_\star \le \frac{1}{(2\gamma\lambda)k}\|x^0 - \tilde{x}\|^2.$$

Proof: Douglas-Rachford splitting (3) corresponds to the gradient descent iterations (28). So by setting $\lambda = 1/L_h$ one has:

$$h(w^k) - h(\tilde{w}) \le \frac{L_h}{2k}\|w^0 - \tilde{w}\|^2,$$

see for example [22, Prop. 6.10.2]. Applying the substitution $x = Sw$, and considering that

$$\lambda_{\max}^{-1}(D)\|x\|^2 \le \|x\|_{D^{-1}}^2 \le \lambda_{\min}^{-1}(D)\|x\|^2, \ \forall x \in \mathbb{R}^n \quad (32)$$

one obtains

$$\begin{aligned} F_\gamma^{\mathrm{DR}}(x^k) - F_\gamma^{\mathrm{DR}}(\tilde{x}) &\le \frac{L_h}{2k}\|x^0 - \tilde{x}\|_{D^{-1}}^2 \\ &\le \frac{1}{2k}\frac{1+\gamma L_f}{(1-\gamma L_f)}\frac{1}{\lambda_{\min}(D)}\|x^0 - \tilde{x}\|^2 \\ &= \frac{1}{2k}\frac{1+\gamma L_f}{\gamma(1-\gamma L_f)}\|x^0 - \tilde{x}\|^2, \end{aligned}$$

where the last equality holds considering (27). The claim follows by $z^k = G_\gamma(x^k)$, Theorem 1 and inequality (15b). ∎

From Theorem 3 we easily obtain the following optimal value of γ:

$$\gamma_\star = \underset{\gamma}{\operatorname{argmin}} \ \frac{1+\gamma L_f}{\gamma(1-\gamma L_f)} = \frac{\sqrt{2}-1}{L_f}. \quad (33)$$

For this particular value of $\gamma_\star$ the stepsize becomes equal to $\lambda_k = \sqrt{2}-1$. In the strongly convex case we instead obtain the following stronger result.

Theorem 4: If $\mu_f > 0$ and $\lambda_k = \lambda \in (0, 2/(L_h + \mu_h)]$ then

$$\|y^k - x_\star\|^2 \le \frac{\lambda_{\max}(D)}{\lambda_{\min}(D)}\left(1 - \frac{2\lambda\mu_h L_h}{\mu_h + L_h}\right)^k \|x^0 - \tilde{x}\|^2.$$

Proof: Just like in the proof of Theorem 3, iteration (28) is the standard gradient method applied to h. If f is strongly convex then we have, using (26) and (29), that also h is strongly convex. From [19, Th. 2.1.15] we have

$$\|w^k - \tilde{w}\|^2 \le \left(1 - \frac{2\lambda\mu_h L_h}{\mu_h + L_h}\right)^k \|w^0 - \tilde{w}\|^2.$$

Applying the substitution $x = Sw$ we get

$$\|x^k - \tilde{x}\|_{D^{-1}}^2 \le \left(1 - \frac{2\lambda\mu_h L_h}{\mu_h + L_h}\right)^k \|x^0 - \tilde{x}\|_{D^{-1}}^2.$$

The thesis follows considering (32) and that

$$\|y^k - x_\star\|^2 = \|\mathrm{prox}_{\gamma f}(x^k) - \mathrm{prox}_{\gamma f}(\tilde{x})\|^2 \le \|x^k - \tilde{x}\|^2,$$

where the equality holds since $x_\star = \mathrm{prox}_{\gamma f}(\tilde{x})$, and the inequality by nonexpansiveness of $\mathrm{prox}_{\gamma f}$. ∎

IV. Fast Douglas-Rachford splitting

We have shown that DRS is equivalent to the gradient method minimizing $h(w) = F_\gamma^{\mathrm{DR}}(Sw)$. In the quadratic case, since for $\gamma < 1/L_f$ we know that $F_\gamma^{\mathrm{DR}}(x)$ is convex, we can as well apply the optimal first order methods due to Nesterov [18], [19, Sec. 2.2] to the same problem. This way we obtain a *fast Douglas-Rachford splitting* method. The scheme is as follows: given $u^0 = x^0 \in \mathbb{R}^n$, iterate

$$y^k = \mathrm{prox}_{\gamma f}(u^k), \quad (34a)$$

$$z^k = \mathrm{prox}_{\gamma g}(2y^k - u^k), \quad (34b)$$

$$x^{k+1} = u^k + \lambda_k(z^k - y^k), \quad (34c)$$

$$u^{k+1} = x^{k+1} + \beta_k(x^{k+1} - x^k). \quad (34d)$$

We have the following estimates regarding the convergence rate of iterations (34a)-(34d), whose proofs are based on [19].

Theorem 5: For convex quadratic f, if $\gamma < 1/L_f$, λ_k are given by (31) and

$$\beta_k = \begin{cases} 0 & \text{if } k = 0, \\ \frac{k-1}{k+2} & \text{if } k > 0, \end{cases}$$

then the sequence of iterates generated by (34a)-(34d) satisfies

$$F(z^k) - F_\star \leq \frac{2}{\gamma\lambda(k+2)^2}\|x^0 - \tilde{x}\|^2.$$

Proof: See [21]. ■

The optimal choice for γ is again $\gamma_\star = (\sqrt{2}-1)/L_f$. We similarly obtain complexity bounds for the strongly convex case, as described in the following result.

Theorem 6: If f is strongly convex quadratic, $\gamma < 1/L_f$, λ_k are given by (31) and

$$\beta_k = \frac{1-\sqrt{\mu_h/L_h}}{1+\sqrt{\mu_h/L_h}},$$

then the sequence of iterates generated by (34a)-(34d) satisfies

$$F(z^k) - F_\star \leq \frac{L_h}{\lambda_{\min(D)}}\left(1-\sqrt{\frac{\mu_h}{L_h}}\right)^k \|x^0 - x_\star\|^2.$$

Proof: See [21]. ■

V. Simulations

A. Box-constrained QP

We tested our analysis against numerical results obtained by applying the considered methods to the following box-constrained convex quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x'Qx + q'x \\ \text{subject to} & l \leq x \leq u, \end{array}$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite, while $q, l, u \in \mathbb{R}^n$. The problem is expressed in composite form by setting

$$f(x) = \tfrac{1}{2}x'Qx + q'x, \quad g(x) = \delta_{[l,u]}(x),$$

where δ_C is the indicator function of the convex set C. As it was pointed out in Section III, the proximal mapping associated with f is linear

$$\mathrm{prox}_{\gamma f}(x) = (I + \gamma Q)^{-1}(x - \gamma q).$$

The proximal mapping associated with g is simply the projection onto the $[l, u]$ box, $\mathrm{prox}_{\gamma g}(x) = \Pi_{[l,u](x)}$. Tests were performed on problems generated randomly as described in [23]. In Figure 1 we illustrate the performance of DRS for different choices of the parameter γ. Figure 2 compares the standard DRS and the accelerated method (34a)-(34d).

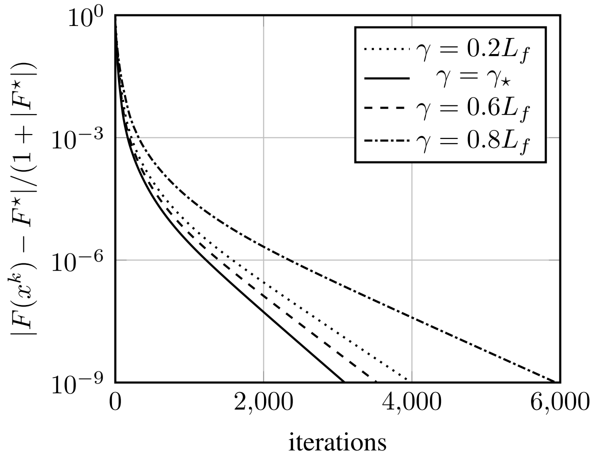

Fig. 1: DRS applied to a randomly generated box-constrained QP, with $n = 500$, for different values of γ.

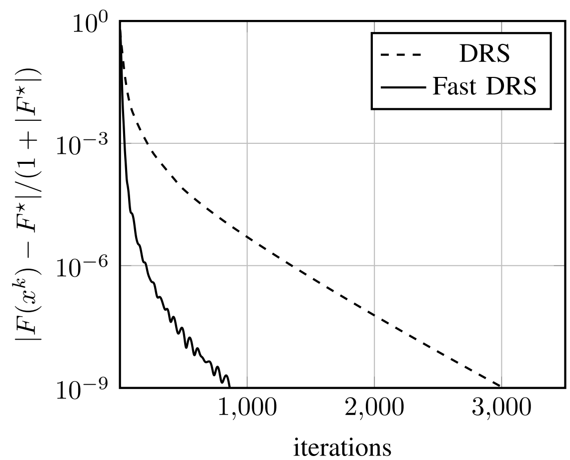

Fig. 2: Comparison between DRS and its accelerated variant, for $\gamma = \gamma_\star$, applied to a randomly generated box-constrained QP with $n = 500$.

B. Sparse least squares

The well known ℓ_1-regularized least squares problem consists of finding a sparse solution to an underdetermined linear system. The goal is achieved by solving

$$\text{minimize} \;\; \tfrac{1}{2}\|Ax - b\|_2^2 + \rho\|x\|_1,$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The regularization parameter ρ modulates between a low residual $\|Ax - b\|_2^2$ and a sparse solution. In this case the proximal mapping with respect to f is

$$\mathrm{prox}_{\gamma f}(x) = (A'A + \gamma^{-1}I)^{-1}(A'b + \gamma^{-1}x),$$

while $\mathrm{prox}_{\gamma g}$ is the following *soft-thresholding* operator,

$$\left[\mathrm{prox}_{\gamma g}(x)\right]_i = \mathrm{sign}(x_i)\cdot\max\{0, |x_i| - \gamma\rho\}, \; i = 1,\ldots n.$$

Random problems were generated according to [24], and the results are shown in Figure 3 and 4, where we compare different choices for γ and the fast Douglas-Rachford iterations.

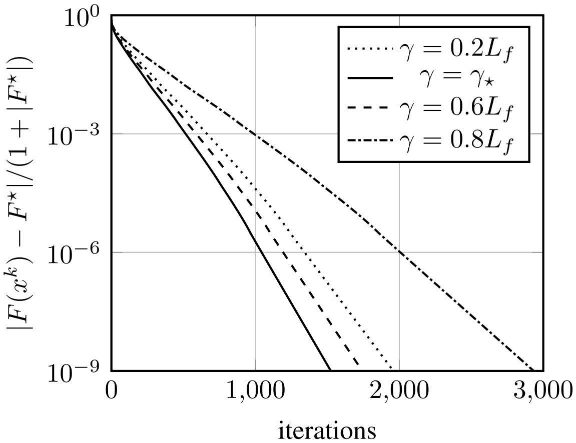

Fig. 3: Comparison of different choices of γ for a random ℓ_1 least squares problem, with $m = 100, n = 1000$.

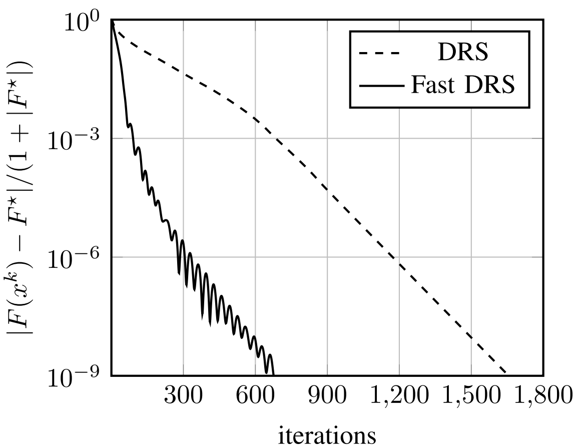

Fig. 4: DRS and its accelerated variant, with $\gamma = \gamma_\star$, applied to a random sparse least squares problem of size $m = 100, n = 1000$.

VI. Conclusions & Future Work

In this paper we dealt with convex composite minimization problems. We introduced a continuously differentiable function, namely the Douglas-Rachford Envelope (DRE). Its minimizers, under suitable assumptions, are in a one-to-one correspondence with the solutions of the original convex composite optimization problem. We observed how the DRS iterations, for finding zeros of the sum of two maximal monotone operators A and B, are equivalent to a scaled unconstrained gradient method applied to the DRE, when $A = \partial f$ and $B = \partial g$ and f is twice continuously differentiable with Lipschitz continuous gradient. This allowed us to to apply well-known results of smooth unconstrained optimization to analyze the convergence of DRS in the particular case of f being convex quadratic. Moreover, we have been able to apply and analyze optimal first-order methods and obtain a fast Douglas-Rachford splitting method. Ongoing work on this topic include exploiting the illustrated results to study convergence properties of ADMM.

References

[1] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 14, no. 5, pp. 877–898, 1976.

[2] 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.

[3] J. Eckstein, "Splitting methods for monotone operators with applications to parallel optimization," PhD Thesis, Massachusetts Institute of Technology, 1989.

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

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

[6] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, 2011.

[7] 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. North-Holland: Amsterdam, 1983.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[9] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[10] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, Jan. 2012.

[11] T. Goldstein, B. O'Donoghue, and S. Setzer, "Fast alternating direction optimization methods," *CAM report*, pp. 12–35, 2012. [Online]. Available: http://www.mia.uni-saarland.de/Publications/goldstein-cam12-35.pdf

[12] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," DTIC Document, Tech. Rep., 2012.

[13] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *arXiv:1208.3922 [math.OC]*, Aug. 2012.

[14] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems," *arXiv:1306.2454 [math.OC]*, 2013.

[15] D. Davis and W. Yin, "Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions," *arXiv:1407.5210 [math.OC]*, 2014.

[16] P. Patrinos and A. Bemporad, "Proximal Newton methods for convex composite optimization," in *IEEE Conference on Decision and Control*, 2013, pp. 2358–2363.

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

[18] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[19] ——, *Introductory lectures on convex optimization: A basic course*. Springer, 2003, vol. 87.

[20] C. Lemaréchal and C. Sagastizábal, "Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 367–385, 1997.

[21] P. Patrinos, L. Stella, and A. Bemporad, "Douglas-Rachford splitting: complexity estimates and accelerated variants," *arXiv:1407.6723 [math.OC]*, 2014.

[22] D. P. Bertsekas, *Convex Optimization Theory*. Athena Scientific, 2009.

[23] C. C. Gonzaga, E. W. Karas, and D. R. Rossetto, "An optimal algorithm for constrained differentiable convex optimization," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 1939–1955, 2013.

[24] D. Lorenz, "Constructing test instances for basis pursuit denoising," *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1210–1214, 2013.