

Accelerated ADMM based on Accelerated Douglas-Rachford Splitting

Ivan Pejcic¹ and Colin N. Jones¹

Abstract—Alternating direction method of multipliers (ADMM) is a form of augmented Lagrangian optimisation algorithm that found its place in many new applications in recent years. This paper explores a possibility for an upgrade of the ADMM by extrapolation-based acceleration, which has been successfully utilised for a long time in case of accelerated gradient method. The development uses a recently proposed accelerated Douglas-Rachford splitting by applying it on Fenchel dual problem, resulting in a method that replaces the classical proximal point convergence mechanism of ADMM with the accelerated gradient. The obtained method requires that the second function involved in the cost is strongly convex quadratic, as well as an upper bound on the penalty parameter. A heuristic modification of the derived method is described, and numerical experiments are performed by solving a randomly generated quadratic programming (QP) problem.

I. INTRODUCTION

The alternating direction method of multipliers (ADMM) is a convex optimisation algorithm that has received an intensive attention in recent years due to its applicability to large-scale machine learning and image processing problems [1]. Even though developed a long time ago, the reasons for its renewed attention lie in its form conducive to distributed-memory implementation, its possibility of formulating closed form solutions for subproblems involved in the algorithm, and its practical performance, which produces solutions with accuracy sufficiently high for many applications of interest [2].

One of the two ADMM optimisation models that commonly appear in the literature is the Fenchel primal:

$$\begin{aligned} & \text{minimize} && f_1(x) + f_2(Ax) \\ & \text{subject to} && x \in \mathbb{R}^n, \end{aligned} \quad (1)$$

where $f_1 : \mathbb{R}^n \rightarrow (-\infty, \infty]$ and $f_2 : \mathbb{R}^m \rightarrow (-\infty, \infty]$ are closed proper convex functions, A is an $m \times n$ matrix, and a feasible solution x is assumed to exist.

¹Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland {ivan.pejcic, colin.jones}@epfl.ch. This work was supported by the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013), ERC Grant Agreement 307608 (BuildNet).

The augmented Lagrangian for (1) is

$$L_c(x, z, \lambda) = f_1(x) + f_2(z) + \langle \lambda, Ax - z \rangle + \frac{c}{2} \|Ax - z\|^2,$$

where $\lambda \in \mathbb{R}^m$ is a dual variable, $c > 0$ is a penalty parameter, and $\langle a, b \rangle$ denotes the inner product $a'b$ with a' being the transpose of a . The ADMM takes the form

$$\begin{aligned} x_{k+1} & \in \arg \min_{x \in \mathbb{R}^n} L_c(x, z_k, \lambda_k), \\ z_{k+1} & \in \arg \min_{z \in \mathbb{R}^m} L_c(x_{k+1}, z, \lambda_k), \\ \lambda_{k+1} & = \lambda_k + c(Ax_{k+1} - z_{k+1}). \end{aligned} \quad (2)$$

A general property of first order methods, including the gradient method and ADMM, is their slow convergence rate in comparison to advanced second order methods. A substantial contribution to the performance of gradient method is achieved through its acceleration based on a sophisticated extrapolation rule between subsequent gradient steps. In the case of a gradient projection method which applies to minimisation of a convex differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ over a closed convex set $X \subset \mathbb{R}^n$ with f having a Lipschitz continuous gradient, the version accelerated by extrapolation [3], [4] involves an extrapolation step $y_k = x_k + \beta_k(x_k - x_{k-1})$ followed by a gradient step $x_{k+1} = P_X(y_k - \nabla f(y_k))$, where $P_X : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes projection on the set X , value $x_{-1} = x_0$, and $\{\beta_k\}$ represents a sequence of extrapolation parameters. Under some very particular choices of the extrapolation $\{\beta_k\}$, [3], [4], one of which is defined as

$$\beta_0 = 0, \quad \beta_k = \frac{k-1}{k+2}, \quad \forall k \geq 1, \quad (3)$$

the $O(1/k^2)$ iteration complexity is attained, improving the $O(1/k)$ complexity of the non-accelerated gradient method obtained when $\{\beta_k\}$ is $\beta_k \equiv 0$.

The acceleration technique of the gradient method has been successfully extended to acceleration of the alternating minimisation algorithm (AMA), an algorithm with a form similar to ADMM, and this resulted in the fast alternating minimisation algorithm (FAMA) [5], [6]. An upgrade of ADMM by an extrapolation technique

has been analysed in [5], where it has been shown that if both functions involved in the cost are strongly convex with the second one being quadratic, the dual function values converge with $O(1/k^2)$ iteration complexity provided that the penalty parameter c is small enough.

An important insight into the ADMM convergence mechanism represents its interpretation as a special case of proximal point algorithm applied for finding a zero of a maximal monotone operator [7]. By replacing this mechanism with a generalised version, a generalised ADMM is obtained with a potential for acceleration of convergence through a choice of overrelaxation factors.

This document derives an accelerated version of ADMM by using a recently proposed accelerated Douglas-Rachford (DR) splitting [8] on Fenchel dual problem. The obtained method replaces the internal proximal point algorithm of classical ADMM by the accelerated gradient method applied on a specially constructed scaled DR envelope function [8]. The derived algorithm addresses the optimisation model (1) with an assumption that the function $f_2(z)$ is strongly convex quadratic, and involves an $O(1/k^2)$ bound on the values of the scaled DR envelope function when an upper bound on the penalty parameter is satisfied. A heuristic modification of the obtained method which can potentially extend the benefit of extrapolation for penalties beyond the upper bound is provided and tested in the numerical results section.

The paper is structured as follows. Section II summarises some theoretical results that will be used in the later analysis. Section III derives the accelerated ADMM algorithm based on accelerated DR splitting. Section IV contains numerical experiments. Section V presents conclusion and suggests directions for future research.

II. THEORETICAL TOOLS

A. Accelerated DR splitting

This section summarises the results of [8] by expressing them in a notation which will be used in the later developments. The DR splitting addresses problems of the form

$$\begin{aligned} & \text{minimize} && d(\lambda) = d_1(\lambda) + d_2(\lambda) \\ & \text{subject to} && \lambda \in \mathbb{R}^m, \end{aligned} \quad (4)$$

where $d_1 : \mathbb{R}^m \rightarrow (-\infty, \infty]$ and $d_2 : \mathbb{R}^m \rightarrow (-\infty, \infty]$ are closed proper convex functions. Given $v_0 \in \mathbb{R}^m$, the DR splitting algorithm consists of the steps

$$\lambda_k = \text{prox}_{cd_2}(v_k), \quad (5)$$

$$\mu_k = \text{prox}_{cd_1}(2\lambda_k - v_k), \quad (6)$$

$$v_{k+1} = v_k + \rho_k(\mu_k - \lambda_k), \quad (7)$$

where $c > 0$ is a penalty parameter, $\{\rho_k\} \subset [0, 2]$ is a sequence of relaxation factors, and the expression

$$\text{prox}_{ch}(v) = \arg \min_{z \in \mathbb{R}^m} \left\{ h(z) + \frac{1}{2c} \|z - v\|^2 \right\} \quad (8)$$

defines the proximal operator $\text{prox}_{ch} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ which is single valued and whose existence is guaranteed when the function $h : \mathbb{R}^m \rightarrow (-\infty, \infty]$ is closed proper convex. Assuming that the set of optimal solutions of (4) is nonempty, that $0 < \inf_k \{\rho_k\} \leq \sup_k \{\rho_k\} < 2$, and that the relative interiors of $\text{dom}(d_1)$ and $\text{dom}(d_2)$ have a point in common where $\text{dom}(d) = \{\lambda \mid d(\lambda) < \infty\}$ represents the effective domain of a function d , the sequence $\{v_k\}$ converges to a fixed point of the DR splitting v^* which is related to $\lambda^* \in \arg \min\{d(\lambda)\}$ as $\lambda^* = \text{prox}_{cd_2}(v^*)$. Because of the relation between the fixed point of the DR splitting v^* and the optimal solution λ^* , finding a fixed point v^* is essentially the same as finding an optimal solution λ^* of problem (4).

The following supplementary assumption ensures the existence of a convex differentiable function referred to as the DR envelope, which plays the key role in development of the accelerated DR splitting.

Assumption 1. *The function $d_2(\lambda)$ and the penalty c satisfy $d_2(\lambda) = \frac{1}{2}\lambda'Q\lambda + q'\lambda$ and $c < \frac{1}{L_{d_2}}$, where $Q \in \mathbb{R}^{m \times m}$ is symmetric positive semidefinite, $q \in \mathbb{R}^m$, and L_{d_2} is the Lipschitz constant of the function d_2 (i.e., the maximal eigenvalue of Q).*

The DR envelope $F_c^{DR}(v)$ whose existence is guaranteed under this additional assumption takes the form

$$F_c^{DR}(v) = d_2^c(v) - c\|\nabla d_2^c(v)\|^2 + d_1^c(v - 2c\nabla d_2^c(v)),$$

with $v \in \mathbb{R}^m$, and $h^c(v)$ denoting a Moreau envelope of a function $h(v)$ defined by

$$h^c(v) = \inf_{x \in \mathbb{R}^m} \left\{ h(x) + \frac{1}{2c} \|x - v\|^2 \right\}. \quad (9)$$

The set of stationary points of the DR envelope $F_c^{DR}(v)$ (i.e., points $v^* \in \arg \min\{F_c^{DR}(v)\}$) coincides with the set of fixed points of the DR splitting.

The significance of the DR envelope is that one iteration of the scaled gradient method

$$v_{k+1} = v_k - \rho_k D \nabla F_c^{DR}(v_k) \quad (10)$$

is equivalent to one iteration of DR splitting (5)-(7), provided that $D = c(2(I + cQ)^{-1} - I)^{-1}$ and the employed $\{\rho_k\}$ is the same. By introducing a scaled variable w defined by $v = Sw$ using $S = D^{\frac{1}{2}}$, the iteration (10) can be written as

$$w_{k+1} = w_k - \rho_k \nabla h(w_k) \quad (11)$$

where $h(w) = F_c^{DR}(Sw)$. Since it can be shown that the Lipschitz constant of $\nabla h(w)$ is

$$L_h = \frac{1 + cL_{d_2}}{1 - cL_{d_2}}, \quad (12)$$

it follows by convergence theory of the gradient method with constant stepsize [3] that the algorithm converges by choosing $\rho_k \in (0, 2/L_h)$.

The established equivalence of one DR splitting cycle with one iteration of the gradient method applied on $h(w)$ allows introduction of accelerated DR splitting, which is obtained by applying the accelerated gradient method on the function $h(w)$. Given $y_0 = v_0 \in \mathbb{R}^m$, the resulting accelerated DR splitting algorithm takes the form

$$\lambda_k = \text{prox}_{cd_2}(y_k), \quad (13)$$

$$\mu_k = \text{prox}_{cd_1}(2\lambda_k - y_k), \quad (14)$$

$$v_{k+1} = y_k + \rho_k(\mu_k - \lambda_k), \quad (15)$$

$$y_{k+1} = v_{k+1} + \beta_k(v_{k+1} - v_k). \quad (16)$$

It can be shown that under the extrapolation rule $\{\beta_k\}$ given in (3), the accelerated version has an $O(1/k^2)$ iteration complexity guaranteeing that

$$d(\mu_k) - d^* \leq F_c^{DR}(v_k) - F_c^{DR*} \leq \frac{2}{c\rho(k+2)^2} \|v_0 - v^*\|^2, \quad (17)$$

where $F_c^{DR*} = F_c^{DR}(v^*) = d^* = \inf_{\lambda \in \mathbb{R}^m} \{d(\lambda)\}$ and $\rho := \rho_k \equiv 1/L_h$. By considering the strong convexity properties of $F_c^{DR}(v)$ [8] and utilising the extrapolation rule $\{\beta_k\}$ from [4], a linear convergence rate can as well be established. For the details related to development of the previous results, the reader is referred to [8].

B. Augmented Lagrangian and proximal algorithm

The proximal iteration

$$\lambda = \text{prox}_{cd}(\mu), \quad \lambda, \mu \in \mathbb{R}^m, \quad (18)$$

where $d: \mathbb{R}^m \rightarrow (-\infty, \infty]$ is closed proper convex, can be shown to be equivalent to finding a decomposition [3] of the form

$$\mu = \lambda + cm, \quad m \in \partial d(\lambda), \quad (19)$$

where $\partial d(\lambda)$ denotes the subdifferential of the function d at λ , which represents the set of all subgradients $g \in \mathbb{R}^m$ at λ satisfying $d(\gamma) \geq d(\lambda) + g'(\gamma - \lambda)$, $\forall \gamma \in \mathbb{R}^m$.

In case where the function $d(\lambda)$ is defined by

$$d(\lambda) = - \inf_{x \in \mathbb{R}^n} \{h(x) + \langle \lambda, Ax - b \rangle\}, \quad (20)$$

where $h: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is closed, proper and convex, the proximal iteration on the function $d(\lambda)$ can

be evaluated in a way which involves minimisation of the augmented Lagrangian function, as described in, for example, [2]:

Proposition 1. ([2], Proposition 9) *Given any $\mu \in \mathbb{R}^m$, consider the problem*

$$\inf_{x \in \mathbb{R}^n} \left\{ h(x) + \langle \mu, Ax - b \rangle + \frac{c}{2} \|Ax - b\|^2 \right\}. \quad (21)$$

If \bar{x} is an optimal solution to this problem, then setting $\lambda = \mu + c(A\bar{x} - b)$ and $m = b - A\bar{x}$ yields $\lambda, m \in \mathbb{R}^m$ such that $\mu = \lambda + cm$ and $m \in \partial d(\lambda)$, where $d(\lambda)$ is as defined in (20).

III. ACCELERATED ADMM

The derivation of the accelerated ADMM algorithm based on accelerated DR splitting (13)-(16) from [8] will be performed in a way inspired by the development of standard ADMM in [2]. The outcome will be a modified ADMM algorithm whose underlying working mechanism is characterised by the $O(1/k^2)$ iteration complexity given in (17). The optimisation model addressed by the method to be developed is the Fenchel primal form given in (1), which can be reformulated as

$$\begin{aligned} & \text{minimize} && f_1(x) + f_2(z) \\ & \text{subject to} && Ax = z. \end{aligned} \quad (22)$$

By applying the standard Lagrange duality for equality constrained problems [9], one obtains the Lagrange dual function:

$$\begin{aligned} q(\lambda) &= \inf_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} \{f_1(x) + f_2(z) + \lambda'(Ax - z)\} \\ &= \inf_{x \in \mathbb{R}^n} \{f_1(x) + (A'\lambda)'x\} + \inf_{z \in \mathbb{R}^m} \{f_2(z) - \lambda'z\} \\ &= -f_1^*(-A'\lambda) - f_2^*(\lambda) = -d_1(\lambda) - d_2(\lambda), \end{aligned}$$

in which $d_1(\lambda) = f_1^*(-A'\lambda)$, $d_2(\lambda) = f_2^*(\lambda)$, and the $f^*(\lambda) = \sup_{z \in \mathbb{R}^m} \{\lambda'z - f(z)\}$ represents the conjugate of a function f . This leads to a Fenchel dual problem

$$\begin{aligned} & \text{minimize} && d(\lambda) = d_1(\lambda) + d_2(\lambda) \\ & \text{subject to} && \lambda \in \mathbb{R}^m, \end{aligned} \quad (23)$$

which minimises $d(\lambda) = -q(\lambda)$ (i.e., maximises the dual function $q(\lambda)$ defined above).

Accelerated ADMM is derived by applying the accelerated DR splitting (13)-(16) to the Fenchel dual problem (23). For this purpose, (23) should satisfy the assumptions of Section II-A guaranteeing the existence of the DR envelope $F_c^{DR}(v)$.

Proposition 2. *The functions $d_1(\lambda)$ and $d_2(\lambda)$ are closed, convex and proper. Under the assumption that*

the function $f_2(z)$ is strongly convex quadratic and the penalty c is small enough:

$$f_2(z) = \frac{1}{2}z'Pz + p'z, \quad c < \frac{1}{L_{d_2}}, \quad (24)$$

where $P \in \mathbb{R}^{m \times m}$ is symmetric positive definite, $p \in \mathbb{R}^m$ and where L_{d_2} is a Lipschitz constant of the function d_2 , the conditions of the Assumption 1 are satisfied, ensuring the existence of the DR envelope $F_c^{DR}(v)$.

Proof: The closedness and convexity of $d_1(\lambda)$ and $d_2(\lambda)$ follow from Lemma 8 in [2], and their properness from Prop. 1.6.1(b) in [9]. Since by the expression for conjugate of a strongly convex quadratic [10] there holds $d_2(\lambda) = f_2^*(\lambda) = \frac{1}{2}\lambda'P^{-1}\lambda - p'P^{-1}\lambda$, the function $d_2(\lambda)$ is strongly convex quadratic as P^{-1} of a positive definite matrix P is positive definite. Considering as well the bound imposed on the penalty c , the conditions of Assumption 1 are satisfied. \square

The following lemma derives a form of the accelerated ADMM based on the accelerated DR splitting. For this purpose, an additional assumption is introduced providing two sufficient conditions that ensure the existence of the optimal solution in the minimisation (25) below.

Assumption 2. *It holds that the effective domain of the function $f_1(x)$ is bounded and/or that $A'A$ is invertible.*

Lemma 1. *Given $z_0, \lambda_0 \in \mathbb{R}^m$, $c > 0$, a sequence $\{\rho_k\} \subset [0, 2]$ and a sequence $\{\beta_k\}$, the sequence of recursions applied to the Fenchel primal (1):*

$$x_{k+1} \in \arg \min_{x \in \mathbb{R}^n} \{f_1(x) + \langle \lambda_k, Ax \rangle + \frac{c}{2}\|Ax - z_k\|^2\}, \quad (25)$$

$$z_{k+1} \in \arg \min_{z \in \mathbb{R}^m} \{f_2(z) - \langle \lambda_k + E_k, z \rangle + \frac{1}{2}\|\xi_k - z\|^2\}, \quad (26)$$

$$\lambda_{k+1} = \lambda_k + E_k + \xi_k - cz_{k+1}, \quad (27)$$

with ξ_k, E_k calculated as

$$\xi_k = c(\rho_k Ax_{k+1} + (1 - \rho_k)z_k), \quad (28)$$

$$E_k = \beta_k(\lambda_k - \lambda_{k-1}) + \beta_k c A(\rho_k x_{k+1} - \rho_{k-1} x_k) + \beta_k c((1 - \rho_k)z_k - (1 - \rho_{k-1})z_{k-1}), \quad (29)$$

where $E_0 := 0$, is equivalent to the accelerated DR splitting (13)–(16) applied to the Fenchel dual (23), and therefore equivalent to the accelerated gradient method applied to $h(w) = F_c^{DR}(S w)$.

Proof: The starting point for the development is the accelerated DR splitting (13)–(16). Given the initial

$\lambda_0 \in \mathbb{R}^m$, $y_{-1} := \lambda_0 + cz_0$ and $v_{-1} := v_0$, (13)–(16) can be rewritten as

$$\mu_k = \text{prox}_{cd_1}(2\lambda_k - y_{k-1}), \quad (30)$$

$$v_k = y_{k-1} + \rho_k(\mu_k - \lambda_k), \quad (31)$$

$$y_k = v_k + \beta_k(v_k - v_{k-1}), \quad (32)$$

$$\lambda_{k+1} = \text{prox}_{cd_2}(y_k), \quad (33)$$

which in comparison to (13)–(16) has the λ_k update moved to the end, and the indices of v and y variable shifted from $k+1$ to k . By the equivalence (18)–(19), the sequence (30)–(33) can be equivalently written as

$$\mu_k + cw_k = \lambda_k - cm_k, \quad w_k \in \partial d_1(\mu_k), \quad (34)$$

$$v_k = y_{k-1} + \rho_k(\mu_k - \lambda_k), \quad (35)$$

$$y_k = v_k + \beta_k(v_k - v_{k-1}), \quad (36)$$

$$\lambda_{k+1} + cm_{k+1} = y_k, \quad m_{k+1} \in \partial d_2(\lambda_{k+1}), \quad (37)$$

where $m_0 = z_0$, and $\lambda_k - cm_k$ in (34) is obtained by substituting the y_k in (37) for y_{k-1} in (30). By applying the evaluation of proximal iterate by augmented Lagrangian from Proposition 1, (34) can be written as

$$x_{k+1} \in \arg \min_{x \in \mathbb{R}^n} \{f_1(x) + \langle \lambda_k - cm_k, Ax \rangle + \frac{c}{2}\|Ax\|^2\}, \quad (38)$$

$$\mu_k = \lambda_k - cm_k + cAx_{k+1}, \quad (39)$$

$$w_k = -Ax_{k+1}, \quad (40)$$

as well as (37) which becomes

$$z_{k+1} \in \arg \min_{z \in \mathbb{R}^m} \{f_2(z) + \langle y_k, -z \rangle + \frac{c}{2}\|z\|^2\}, \quad (41)$$

$$\lambda_{k+1} = y_k - cz_{k+1}, \quad (42)$$

$$m_{k+1} = z_{k+1}. \quad (43)$$

By Weierstrass theorem (Prop. 3.2.1, [3]), the existence of the solution of (38) is ensured by the Assumption 2, and in case of (41) the existence is ensured by the presence of the quadratic term $\|z\|^2$.

By using (42) to express y_{k-1} and (39), (43) to express $\mu_k - \lambda_k$, (35) can be written as

$$v_k = \lambda_k + cz_k + \rho_k c(Ax_{k+1} - z_k) = \lambda_k + c(\rho_k Ax_{k+1} + (1 - \rho_k)z_k). \quad (44)$$

By introducing (44) into (36) for v_k and v_{k-1} , one obtains

$$y_k = \lambda_k + \xi_k + E_k, \quad (45)$$

where ξ_k, E_k are defined in (28)–(29). By specifying $E_0 := 0$, the compliance of (45) with $v_{-1} := v_0$ is obtained.

By using the expressions (38)-(45), the cycle (34)-(37) gets expressed as

$$x_{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ f_1(x) + \langle \lambda_k - cz_k, Ax \rangle + \frac{c}{2} \|Ax\|^2 \right\}, \quad (46)$$

$$y_k = \lambda_k + \xi_k + E_k, \quad (47)$$

$$z_{k+1} \in \arg \min_{z \in \mathbb{R}^m} \left\{ f_2(z) + \langle y_k, -z \rangle + \frac{c}{2} \|-z\|^2 \right\}, \quad (48)$$

$$\lambda_{k+1} = y_k - cz_{k+1}, \quad (49)$$

with ξ_k, E_k defined as in (28)-(29). By substituting (47) into (48) and (49), and by adding the constant terms $\frac{c}{2} \|z_k\|^2$ and $\frac{c}{2} \|\frac{1}{c} \xi_k\|^2$ to (46) and (48), respectively, the equations (46)-(49) take the form (25)-(27). \square

It can be noticed that by setting the extrapolation term $\beta_k \equiv 0$, the accelerated ADMM (25)-(27) reduces to the generalized ADMM [7] with relaxations, and by choosing as well the relaxations $\rho_k \equiv 1$ we obtain the classical ADMM given in (2). The following proposition summarises the preceding development, and provides a proof which is an adapted version of Prop. 15 in [2].

Proposition 3. *Consider the optimisation model (1) in its equivalent form (22). Let the Assumption 2 hold, and let the function $f_2(z)$ be strongly convex quadratic $f_2(z) = \frac{1}{2} z' P z + p' z$ with $P \in \mathbb{R}^{m \times m}$ symmetric positive definite and $p \in \mathbb{R}^m$. Assume that all subgradients of the function $d_1(\lambda) = -\inf_{x \in \mathbb{R}^n} \{f_1(x) + (A'\lambda)'x\}$ at each point $\lambda \in \mathbb{R}^m$ take the form $-A\bar{x}$ where \bar{x} attains the stated minimum over x . Then, there exists a primal-dual optimal solution pair $((x^*, z^*), \lambda^*)$, and if the sequences $\{x_k\} \subset \mathbb{R}^n$, $\{z_k\} \subset \mathbb{R}^m$ and $\{\lambda_k\} \subset \mathbb{R}^m$ conform to the recursion (25)-(27) under the assumptions of Lemma 1 using (3) for $\{\beta_k\}$ and*

$$\rho_k \equiv \rho = \frac{1 - cL_{d_2}}{1 + cL_{d_2}}, \quad c < \frac{1}{L_{d_2}}, \quad (50)$$

where L_{d_2} is the maximal eigenvalue of P^{-1} , then $\lambda_k \rightarrow \lambda^\infty$, $z_k \rightarrow z^\infty$ and $Ax_k \rightarrow Ax^\infty = z^\infty$ where x^∞ is a limit point of $\{x_k\}$ and $((x^\infty, z^\infty), \lambda^\infty)$ corresponds to a primal-dual solution pair $((x^*, z^*), \lambda^*)$.

Proof: The existence of the primal-dual optimal solution pair $((x^*, z^*), \lambda^*)$ follows from Prop. 1.2.1(a)-(b) in [3], where the relative interior conditions are satisfied due to the quadratic form of $f_2(z)$ and $d_2(\lambda)$.

The previous development shows that the recursion (25)-(27) is equivalent to the sequence $y_{k-1} = \lambda_k + cm_k = \lambda_k + cz_k$ from (42)-(43) produced by the accelerated DR splitting (30)-(33), which is furthermore equivalent to the application of the fast gradient method to the scaled DR envelope $h(w)$. Since the relaxation

factor ρ is by (50) chosen in accordance with the Lipschitz constant of $\nabla h(w)$ given in (12), the sequence $\{v_k\}$ of (30)-(33) converges to the DR fixed point v^∞ and by (32) the sequence $\{y_k\} \rightarrow v^\infty$, if the point v^∞ exists. Since the primal-dual optimal solution pair exists, the point $\lambda^* + cz^*$ is just such a point v^∞ , so it exists. Therefore the sequence $\{y_k\}$ converges to a DR splitting fixed point $v^\infty = y^\infty$, and by Lemma 14 in [2] it has the form $y^\infty = \lambda^\infty + cz^\infty$ where $z^\infty \in \partial d_2(\lambda^\infty)$ and $-z^\infty \in \partial d_1(\lambda^\infty)$. By the assumption regarding the subgradients of d_1 , there exists some x^∞ such that $-Ax^\infty = -z^\infty$, or equivalently $Ax^\infty = z^\infty$.

Since the proximal mapping $\text{prox}_{cd_2}(y)$ is nonexpansive (Prop. 5.1.8 in [3]), it is also continuous. We have $\text{prox}_{cd_2}(y^\infty) = \lambda^\infty$ and $\text{prox}_{cd_2}(y_k) = \lambda_k$, and because of continuity of $\text{prox}_{cd_2}(y)$ we also have $\lambda_k = \text{prox}_{cd_2}(y_k) \rightarrow \text{prox}_{cd_2}(y^\infty) = \lambda^\infty$ and therefore $z_k = (y_{k-1} - \lambda_k)/c \rightarrow z^\infty = (y^\infty - \lambda^\infty)/c$. By using the first equation in (44) together with $y_{k-1} = \lambda_k + cz_k \rightarrow v^\infty$, since $\rho > 0$ due to (50) we have $Ax_{k+1} - z_k \rightarrow 0$, from where it follows that $Ax_{k+1} \rightarrow Ax^\infty = z^\infty$ with x^∞ being a limit point of $\{x_k\}$. \square

IV. NUMERICAL EXPERIMENTS

The algorithm is tested by solving a quadratic programming (QP) problem of the form

$$\begin{aligned} & \text{minimize} && \frac{1}{2} x' P x + p' x \\ & \text{subject to} && l_b \leq x \leq u_b, \end{aligned} \quad (51)$$

where $P \in \mathbb{R}^{n \times n}$ is symmetric positive definite, $p \in \mathbb{R}^n$, and the $l_b, u_b \in \mathbb{R}^n$ represent the lower and upper bound of the variable x , respectively. The QP data are generated randomly using Matlab commands, and are such that $n = 100$, the eigenvalues of the matrix P are equally spaced between 1 and 100, and $-1 \leq l_b \leq u_b \leq 1$.

The QP (51) is expressed in the form (22) by setting $f_1(x) = \delta_X(x)$, $f_2(z) = \frac{1}{2} z' P z + p' z$, where $\delta_X(x)$ is the indicator function of the set $X = \{x \mid l_b \leq x \leq u_b\}$.

The stopping criteria are derived as in Section 3.3 of [1], obtaining $r_{k+1} = Ax_{k+1} - z_{k+1}$ and $s_{k+1} = c(z_k - z_{k+1}) + E_k - c(1 - \rho_k)(Ax_{k+1} - z_k)$ as the primal and dual residual, respectively. The stopping criteria used in the experiments are $\|r_k\| \leq \varepsilon^{pri}$ and $\|s_k\| \leq \varepsilon^{dual}$, with ε^{pri} and ε^{dual} chosen using the absolute and relative criterion from Section 3.3.1 of [1] with $\varepsilon^{abs} = 10^{-4}$, $\varepsilon^{rel} = 10^{-2}$. The maximal number of iterations for which the experiments are run is $k_{max} = 10000$. The initial conditions are set to zero vectors of appropriate dimensions.

The comparison of classical ADMM (2) with the accelerated version over the range of penalties $c \in (0, 1]$ which ensures convergence of the accelerated version

by Prop. 3 is given in Fig. 1a. As the value of the penalty c approaches the upper bound $c_{\max} = 1/L_{d_2} = 1$ from (50), the relaxation ρ tends to 0 according to (50) and thus the accelerated version, characterised by the $O(1/k^2)$ complexity in (17), gradually worsens performance and eventually stops converging.

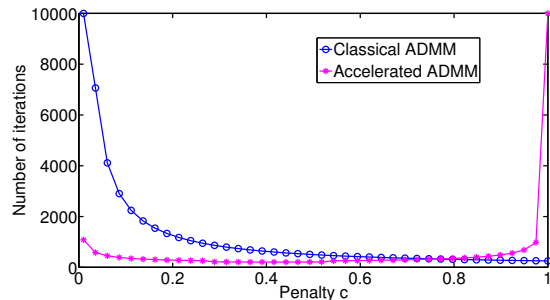
For this reason, a heuristic version which uses $\rho = 1$ for every value of c is introduced and tested. The results are given in Fig. 1b, where the heuristically accelerated ADMM (*i.e.*, the method with $\rho = 1$) is compared with the classical ADMM (2), the generalised ADMM [2] with overrelaxation set to $\rho = 1.9$, and the Fast ADMM with and without restarting ([5], Algorithms 7 and 8). Nevertheless, the heuristic modification $\rho = 1$ may as well cause divergence, as can be seen on Fig. 1c where a QP with a random matrix P containing the eigenvalues equally spaced between 1 and 500 is considered. These results indicate that the heuristic version could benefit by introducing a restarting scheme, like the one of the Fast ADMM with restarting [5], which would ensure convergence while keeping the accelerated behaviour.

V. CONCLUSION

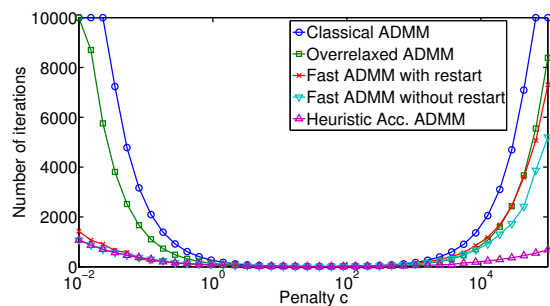
An accelerated version of ADMM based on accelerated Douglas-Rachford splitting is derived resulting in a method characterised by an $O(1/k^2)$ complexity of the internal convergence mechanism. In comparison to the classical ADMM, the derived algorithm involves an additional algebraic step which corresponds to the extrapolations in the underlying fast gradient method. The numerical results show that the method can improve the performance of classical ADMM over the allowed range of penalty parameters, and that a heuristic modification can potentially extend the benefits of acceleration beyond this allowed range. The future work will be oriented to exploration of restarting schemes which can enforce stability of the heuristic version, as well as to testing of the derived method on practical problems.

REFERENCES

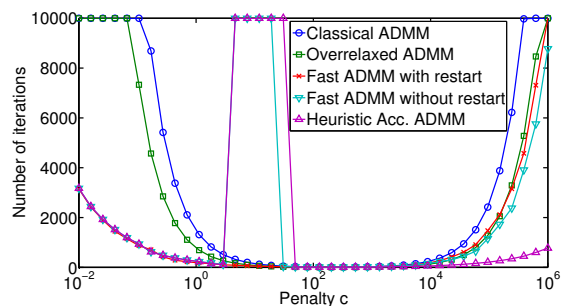
- [1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [2] J. Eckstein, "Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," *RUTCOR Research Reports*, 2012.
- [3] D. P. Bertsekas, *Convex Optimization Algorithms*. Athena Scientific, 2015.
- [4] Y. Nesterov, *Introductory lectures on convex optimization - A basic course*. Springer, 2004.
- [5] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.



(a) Classical and accelerated ADMM over the theoretically allowed range of penalties c . The condition number of QP is equal to 100.



(b) Heuristically accelerated ADMM compared with other versions of ADMM. The same random QP as in (a) is considered.



(c) The same methods as in (b), but applied on a QP whose condition number is 500. Divergence over a range of c can be observed.

Fig. 1: Accelerated ADMM and its heuristic modification, applied for solving a random QP of the form (51).

- [6] M. T. Amir Beck, "Gradient-based algorithms with applications to signal recovery problems," *Convex Optimization in Signal Processing and Communications*, 2009.
- [7] J. Eckstein and D. 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, 1992.
- [8] P. Patrinos, L. Stella, and A. Bemporad, "Douglas-rachford splitting: Complexity estimates and accelerated variants," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, Dec 2014, pp. 4234–4239.
- [9] D. P. Bertsekas, *Convex Optimization Theory*. Athena Scientific, 2009.
- [10] L. Vandenberghe, "Optimization methods for large-scale systems, course slides," *UCLA*, Spring 2013-14.