INEXACT VARIANTS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS AND THEIR ITERATION-COMPLEXITY ANALYSES

Doctoral Thesis by

Vando Antônio Adona

Supervised by

Jefferson D. G. Melo

Funded by
CAPES

IME - INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE FEDERAL DE GOIÁS
Goiânia, Goiás, Brazil

Março 2019

sistema de bibliotecas ufg

PRPG

PRÓ-REITORIA DE PÓS-GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[ ] Dissertação** **[x] Tese**

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Vando Antônio Adona

Título do trabalho: INEXACT VARIANTS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS AND THEIR ITERATION-COMPLEXITY ANALYSES

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Assinatura do(a) autor(a)[2]

Ciente e de acordo:

Jefferson D. G. de Melo

Assinatura do(a) orientador(a)[2]

Data: 09 / 04 / 2019

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

[2] A assinatura deve ser escaneada.

Vando Antônio Adona

INEXACT VARIANTS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS AND THEIR ITERATION-COMPLEXITY ANALYSES

Tese apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Doutor em Matemática.
Área de concentração: Otimização
Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo
Co-orientador: Prof. Dr. Max Leandro Nobre Gonçalves

Goiânia
2019

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Adona, Vando Antônio
INEXACT VARIANTS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS AND THEIR ITERATION-COMPLEXITY ANALYSES [manuscrito] / Vando Antônio Adona. - 2019.
xiii, 79 f.

Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo; co orientador Dr. Max Leandro Nobre Gonçalves.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2019.
Bibliografia.
Inclui siglas, abreviaturas, símbolos, gráfico, tabelas, algoritmos.

1. Alternating direction method of multipliers. 2. Convex program. 3. Hybrid extragradient method. 4. Relative error criterion. 5. Pointwise iteration-complexity. I. Melo, Jefferson Divino Gonçalves de, orient. II. Título.

CDU 51

UFG
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
Campus Samambaia – Caixa Postal 131 – CEP: 74.001-970 – Goiânia-GO.
Fones: (62) 3521-1208 e 3521-1137 www.ime.ufg.br

ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE TESE DE VANDO ANTÔNIO ADONA – Ao vigésimo sétimo dia do mês de março do ano de dois mil e dezenove (27/03/2019), às 14:30 horas, reuniram-se os componentes da Banca Examinadora: Prof. Jefferson Divino Gonçalves de Melo - Orientador, Prof. Max Leandro Nobre Gonçalves, Prof. Leandro da Fonseca Prudente, Prof. Luis Roman Lucambio Perez, Prof. Roberto Andreani e Prof. Gabriel Haeser, sob a presidência do primeiro, e em sessão pública realizada no auditório do Instituto de Matemática e Estatística, procederem a avaliação da defesa de tese intitulada: **"Inexact Variants of the Alternating Direction Method of Multipliers and their Iteration-Complexity Analyses"**, em nível de Doutorado, área de concentração em Otimização, de autoria de Vando Antônio Adona, discente do Programa de Pós-Graduação em Matemática da Universidade Federal de Goiás. A sessão foi aberta pelo Presidente da Banca, Prof. Jefferson Divino Gonçalves de Melo que fez a apresentação formal dos membros da Banca. A seguir, a palavra foi concedida ao autor da tese que, em 45 minutos procedeu a apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca arguiu o examinando, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1513 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Matemática e procedidas às correções recomendadas, a tese foi **APROVADA** por unanimidade, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **DOUTOR EM MATEMÁTICA**, na área de concentração em Otimização pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega na secretaria do PPGM da versão definitiva da tese, com as devidas correções supervisionadas e aprovadas pelo orientador. Cumpridas as formalidades de pauta, às 16:30 horas a presidência da mesa encerrou esta sessão de defesa de tese e para constar eu, Flávia Magalhães Freire, secretária do PPGM, lavrei a presente Ata que, depois de lida e aprovada, será assinada pelos membros da Banca Examinadora em quatro vias de igual teor.

Jefferson D. G. de Melo
Prof. Dr. Jefferson Divino Gonçalves de Melo
Presidente - IME/UFG

Prof. Dr. Leandro da Fonseca Prudente
Membro – IME/UFG

Max Leandro N. Gonçalves
Prof. Dr. Max Leandro Nobre Gonçalves
Membro – IME/UFG

Prof. Dr. Luis Roman Lucambio Perez
Membro – IME/UFG

Prof. Dr. Roberto Andreani
Membro – DM/UNICAMP

Prof. Dr. Gabriel Haeser
Membro – IME/USP

VANDO ANTÔNIO ADONA

INEXACT VARIANTS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS AND THEIR ITERATION-COMPLEXITY ANALYSES

Tese defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de doutor(a) em Matemática, aprovada em 27 de março de 2019, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Jefferson Divino Gonçalves de Melo
Instituto de Matemática e Estatística - UFG
Presidente da Banca

Prof. Dr. Leandro da Fonseca Prudente
Instituto de Matemática e Estatística - UFG

Prof. Max Leandro Nobre Gonçalves
Instituto de Matemática e Estatística - UFG

Prof. Dr. Roberto Andreani
Departamento de Matemática Aplicada - UNICAMP

Prof. Dr. Gabriel Haeser
Instituto de Matemática e Estatística - USP

Prof. Dr. Luis Roman Lucambio Perez
Instituto de Matemática e Estatística - UFG

Dedicado a:

Meu filho Vítor Gabriel

Minha sobrinha e afilhada Lívia Maria

Em memória de meu pai Dirceu (1949-2007)

Agradecimentos

Gostaria de agradecer primeiramente a Deus pela saúde, paz, por todas as graças alcançadas e pela força divina que tenho recebido durante toda minha vida.

Sou muito grato aos meus orientadores Dr. Jefferson D. G. de Melo e Dr. Max Leandro N. Gonçalves pela ajuda na elaboração deste trabalho, pelas idéias, sugestões, conselhos e amizade, também pela paciência, atenção e cordialidade com que sempre me receberam.

Agradeço a Universidade Federal de Goiás (UFG) pela oportunidade como aluno, a todos os professores e funcionários do Instituto de Matemática e Estatística (IME), em especial aos Professores do grupo de Otimização do IME-UFG: Glaydston C. Bento, Jefferson D. G. de Melo, Leandro F. Prudente, Luis R. L. Pérez, Max L. N. Gonçalves, Reinier D. Millán e Orizon P. Ferreira que contribuiram considerávelmente com a minha formação. Agradeço também aos Professores Gabriel Haeser, Leandro F. Prudente, Luis R. L. Pérez e Roberto Andreani por participarem da banca examinadora, pelas sugestões e comentários.

Muito obrigado a todos os meus amigos do curso. Não mais importante do que outros, mas provavelmente pela constante convivência e intimidade adquirida, agradeço especialmente aos amigos: Fabiana Rodrigues de Oliveira, Fabrícia Rodrigues de Oliveira, Lucas Vidal de Meireles e Maurício Silva Louzeiro.

Agradeço à minha mãe Neosa e meu falecido pai Dirceu, pela educação, amor e base familiar que me deram. Tenho muito respeito e admiração por eles. Muito obrigado a todos os meus familiares, às minhas irmãs Vanessa e Vânia sou grato principalmente pela motivação e entusiasmo transmitos a mim durante meus estudos.

Agradeço também a minha namorada Simone pela paciência, compreensão, carinho e por conceber nosso amado filho. Sou grato também a ele, meu filho Vítor Gabriel, pelo orgulho e felicidade que me proporciona.

À CAPES, agradeço pelo suporte financeiro.

Abstract

This thesis proposes and analyzes some variants of the alternating direction method of multipliers (ADMM) for solving separable linearly constrained convex optimization problems. This thesis is divided into three parts. First, we establish the iteration-complexity of a proximal generalized ADMM. This ADMM variant, proposed by Bertsekas and Eckstein, introduces a relaxation parameter α into the second ADMM subproblem in order to improve its computational performance. We show that, for a given tolerance $\rho > 0$, the proximal generalized ADMM with $\alpha \in (0,2)$ provides, in at most $\mathcal{O}(1/\rho^2)$ iterations, an approximate solution of the Lagrangian system associated to the optimization problem under consideration. It is further demonstrated that, in at most $\mathcal{O}(1/\rho)$ iterations, an approximate solution of the Lagrangian system can be obtained by means of an ergodic sequence associated to a sequence generated by the proximal generalized ADMM with $\alpha \in (0,2]$. Second, we propose and analyze an inexact variant of the aforementioned proximal generalized ADMM. In this variant, the first subproblem is approximately solved using a relative error condition whereas the second one is assumed to be easy to solve. It is important to mention that in many ADMM applications one of the subproblems has a closed-form solution; for instance, ℓ_1-regularized convex composite optimization problems. We show that the proposed method possesses iteration-complexity bounds similar to its exact version. Third, we develop an inexact proximal ADMM whose first subproblem is inexactly solved using an approximate relative error criterion similar to the aforementioned inexact proximal generalized ADMM. Pointwise and ergodic iteration-complexity bounds for the proposed method are established. Our approach consists of interpreting these ADMM variants as an instance of a hybrid proximal extragradient framework with some special properties. Finally, in order to show the applicability and advantage of the inexact ADMM variants proposed here, we present some numerical experiments performed on a setting of problems derived from real-life applications.

Keywords: Alternating direction method of multipliers, Convex program, Hybrid extragradient method, Relative error criterion, Pointwise iteration-complexity, Ergodic iteration-complexity.

Resumo

Esta tese propõe e analisa algumas variantes do método dos multiplicadores das direções alternadas (ADMM) para resolver problemas de otimização convexa com restrição linear. Esta tese é dividida em três partes. Primeiro, estabelecemos iteração complexidade de um ADMM generalizado proximal. Essa variante ADMM, proposta por Bertsekas e Eckstein, introduz um parâmetro de relaxação α no segundo subproblema do ADMM para melhorar seu desempenho computacional. Mostramos que, para uma determinada tolerância $\rho > 0$, o ADMM generalizado proximal com $\alpha \in (0, 2)$ fornece, em no máximo $\mathcal{O}(1/\rho^2)$ iterações, uma solução aproximada do sistema Lagrangiano associado ao problema de otimização considerado. É ainda demonstrado que, em no máximo $\mathcal{O}(1/\rho)$ iterações, uma solução aproximada do sistema Lagrangiano pode ser obtida por meio de uma sequência ergódica associada à sequência gerada pelo ADMM generalizado proximal com $\alpha \in (0, 2]$. Em segundo lugar, propomos e analisamos uma variante inexata do ADMM generalizado proximal acima mencionado. Nesta variante, o primeiro subproblema é aproximadamente resolvido usando uma condição de erro relativo, enquanto o segundo é considerado fácil de resolver. É importante mencionar que, em muitas aplicações do ADMM, um dos subproblemas tem uma solução em forma fechada; por exemplo, problemas de otimização convexos compostos ℓ_1-regularizados. Mostramos que o método proposto possui iteração complexidade semelhantes à sua versão exata. Terceiro, desenvolvemos um ADMM proximal inexato cujo primeiro subproblema é resolvido inexatamente usando um critério de erro relativo aproximado semelhante ao ADMM inexato generalizado proximal acima mencionado. Os limites de iteração complexidade pontual e ergódico para o método proposto são estabelecidos. Nossa abordagem consiste em interpretar essas variantes do ADMM como uma instância de um estrutura híbrida proximal extragradiente com algumas propriedades especiais. Finalmente, a fim de mostrar a aplicabilidade e vantagem das variantes inexatas do ADMM propostas aqui, apresentamos alguns experimentos numéricos realizados em um cenário de problemas derivados de aplicações da vida real.

Palavras-chave : Método dos multiplicadores das direções alternadas, Programa convexo, Método extragradiente híbrido, Critério de erro relativo, Iteração complexidade pontual, Iteração complexidade ergódica.

Basic notation and terminology

$\Re^n$: the n-dimensional Euclidean space,

$\Re_+$: the set of nonnegative real numbers,

$\mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}, \Gamma$: finite-dimensional real inner product vector spaces,

$Q^* : \mathcal{Y} \to \mathcal{X}$: the adjoint of a linear operator $Q : \mathcal{X} \to \mathcal{Y}$,

$\| \cdot \|_Q$: the seminorm induced by self-adjoint semidefinite linear operator Q,

$T : \mathcal{X} \rightrightarrows \mathcal{Y}$: a set-valued operator from $\mathcal{X}$ to $\mathcal{Y}$,

$\langle \cdot, \cdot \rangle$: inner product,

$\| \cdot \|$: norm induced by an inner product,

∂h: subdifferential set of a convex function h,

ADMM: abbreviation for alternating direction method of multipliers,

P-ADMM: abbreviation for proximal ADMM,

PG-ADMM: abbreviation for proximal generalized ADMM,

HPE: abbreviation for hybrid proximal extragradient,

Pointwise: a term which refers to the sequence directly generated by a method,

Ergodic: a term associated to an auxiliary sequence obtained from a sequence by means of an ergodic procedure.

Contents

Chapter 1

Introduction

Let $\mathcal{X}$, $\mathcal{Y}$ and Γ be finite-dimensional real vector spaces with inner products and associated norms denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ respectively. Consider the following linearly constrained optimization problem

$$\min\{f(x)+g(y) : Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{1.1}$$

where $f\colon\mathcal{X}\to(-\infty,\infty]$ and $g\colon\mathcal{Y}\to(-\infty,\infty]$ are proper, closed and convex functions, $A\colon\mathcal{X}\to\Gamma$ and $B\colon\mathcal{Y}\to\Gamma$ are linear operators, and $b\in\Gamma$. This problem naturally arises in many applications such as signal and image processing, statistics, compressive sensing and machine learning (see, for example, [7, 12, 58]). An important class of problems that can be fit into the above setting is the well-known composite convex optimization problems of the form

$$\min\{f(x)+g(Qx) : x\in\mathcal{X}\}, \tag{1.2}$$

where $Q\colon\mathcal{X}\to\mathcal{Y}$ is a linear operator. Indeed, this can be done by considering an artificial variable $y=Qx$ and setting $A=-Q$, $B=I$, and $b=0$. Special instances of (1.2) include: (i) LASSO [77, 78] and ℓ_1-regularized logistic regression [51], where $Q=I$; (ii) least absolute deviations [12, Sect. 6.1] and total variation denoising [67], where Q is associated to the least squares fitting model for the former application and the first-order finite difference for the latter.

The augmented Lagrangian function $L_\beta\colon\mathcal{X}\times\mathcal{Y}\times\Gamma\to(-\infty,\infty]$ associated with problem (1.1) is defined as

$$L_\beta(x,y,\gamma)=f(x)+g(y)-\langle\gamma,Ax+By-b\rangle+\frac{\beta}{2}\left\|Ax+By-b\right\|^2, \tag{1.3}$$

where $\beta>0$ is a penalty parameter and $\gamma\in\Gamma$ denotes the Lagrange multiplier.

Recently, there has been a growing interest in the study of the alternating direction method of multipliers (ADMM) and its variants, due to their efficiency for solving the aforementioned class of problems; see, for instance, [12] for a complete review. The ADMM is an augmented Lagrangian type method that explores the separable structure of problem (1.1) in such a way that the augmented Lagrangian subproblem is solved alternately. More specifically, the ADMM applied for solving (1.1) consists of the iterative scheme

$$x_k \in \arg\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 \right\}, \tag{1.4a}$$

$$y_k \in \arg\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|Ax_k + By - b\|^2 \right\}, \tag{1.4b}$$

$$\gamma_k = \gamma_{k-1} - \theta\beta \left(Ax_k + By_k - b \right), \tag{1.4c}$$

where $\beta > 0$. Note that (1.4a)-(1.4b) corresponds to minimize, respectively, the "partial" augmented Lagrangian functions $L_\beta(x, y_{k-1}, \gamma_{k-1})$ and $L_\beta(x_k, y, \gamma_{k-1})$, whereas (1.4c) is the Lagrange multiplier update rule with a relaxation factor θ which is frequently chosen in the interval $(0, (1+\sqrt{5})/2)$. The first ones to consider this scheme (or slight variant of it) were Glowinski and Marroco in [37] and Gabay and Mercier in [34]. Its convergence was established in [32, 33], see also [33, 35, 36] and [12, 28] for detailed discussions about this scheme. It has been observed that the use of the relaxation parameter θ, specially with $\theta \approx 1.6$, in the Lagrange multiplier update (1.4c) improves the numerical performance of the method, see [18,35,47]. Recently, many authors have proposed and studied some variants of this method; see, for example, [10, 13, 17, 21, 25, 31, 42, 43, 45, 48, 52, 61].

Among the aforementioned variants, one that has received a special attention is the so called proximal ADMM, which can be described as follows:

$$x_k \in \arg\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 + \frac{1}{2} \|x - x_{k-1}\|_G^2 \right\}, \tag{1.5a}$$

$$y_k \in \arg\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|Ax_k + By - b\|^2 + \frac{1}{2} \|y - y_{k-1}\|_H^2 \right\}, \tag{1.5b}$$

$$\gamma_k = \gamma_{k-1} - \theta\beta \left(Ax_k + By_k - b \right), \tag{1.5c}$$

where $G \colon \mathcal{X} \to \mathcal{X}$ and $H \colon \mathcal{Y} \to \mathcal{Y}$ are self-adjoint positive semidefinite linear operators. Note that the difference between the proximal and the standard ADMMs is the inclusion of the proximal terms in the associated subproblems. Indeed, the standard ADMM can be recovered by setting $(G, H) = (0, 0)$. In general, the inclusion of proximal terms as in (1.5a)-(1.5b) make the subproblems easier to solve or even to have closed-form solutions. This estrategy was first introduced by Eckstein in [25] and more recently considered in several papers; see for example, [5, 21, 40, 46, 48, 49, 81]. The standard ADMM (1.4) with $\theta = 1$ can be recovered by applying the Douglas-Rachford splitting method [23, 53] to the dual problem of (1.1) see, for example, [26, 33, 82]. In [26], Eckstein and Bertsekas also proposed the following generalized ADMM for solving (1.2): fixed two summable sequences $\{\mu_k\} \subset \Re_+$ and $\{\nu_k\} \subset \Re_+$, obtain (x_k, y_k, γ_k) as follows

$$x_k \approx \arg\min_{x\in\mathcal{X}} \left\{ f(x) + \langle \gamma_{k-1}, Qx\rangle + \frac{\beta}{2}\|Qx - y_{k-1}\|^2 \right\}, \tag{1.6a}$$

$$y_k \approx \arg\min_{y\in\mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, y\rangle + \frac{\beta}{2}\|y - \alpha Qx_k - (1-\alpha)y_{k-1}\|^2 \right\}, \tag{1.6b}$$

$$\gamma_k = \gamma_{k-1} - \beta\left[y_k - \alpha Qx_k - (1-\alpha)y_{k-1}\right], \tag{1.6c}$$

where $\alpha \in (0, 2)$ and the approximate solutions x_k and y_k are such that $\|x_k - x_k^e\| \le \mu_k$ and $\|y_k - y_k^e\| \le \nu_k$, with x_k^e and y_k^e being the exact solutions of (1.6a) and (1.6b), respectively. Note that if $\mu_k = \nu_k = 0$ for every k and $\alpha = 1$ the generalized ADMM (1.6) becomes the ADMM (1.4) with $\theta = 1$ applied to (1.2), i.e., with $A = -Q$, $B = I$, and $b = 0$. As has been observed by many authors (see, e.g., [2, 9, 24, 31, 59]), the use of the relaxation parameter $\alpha > 1$ in (1.6b)–(1.6c) may considerably improve the numerical performance of the method.

1.1 Main contributions

We propose and analyze some ADMM variants applied for solving the linearly constrained convex optimization problem (1.1). We are interested in establishing pointwise and ergodic iteration-complexities for these variants to obtain approximate solutions of the following Lagrangian system associated with problem (1.1)

$$0 \in \partial f(x) - A^*\gamma, \qquad 0 \in \partial g(y) - B^*\gamma, \qquad 0 = Ax + By - b. \tag{1.7}$$

Note that (x^*, y^*, γ^*) is a solution of the above system, if and only if, (x^*, y^*) is a solution to problem (1.1) and γ^* is an associated Lagrange multiplier. Here, for a given tolerance $\rho > 0$, we shall consider two concepts of approximate solutions of (1.7). A triple $(\hat{x}, \hat{y}, \hat{\gamma})$ is said to be a ρ-approximate solution of (1.7) with residue $(v_{\hat{x}}, v_{\hat{y}}, v_{\hat{\gamma}}) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$ if the following conditions hold

$$\begin{gathered} v_{\hat{x}} \in \partial f(\hat{x}) - A^*\hat{\gamma}, \qquad v_{\hat{y}} \in \partial g(\hat{y}) - B^*\hat{\gamma}, \qquad v_{\hat{\gamma}} = A\hat{x} + B\hat{y} - b \\ \max\{\|v_{\hat{x}}\|, \|v_{\hat{y}}\|, \|v_{\hat{\gamma}}\|\} \le \rho; \end{gathered} \tag{1.8}$$

whereas a triple $(\bar{x}, \bar{y}, \bar{\gamma})$ is said to be a relaxed ρ-approximate solution of (1.7) with residues $(v_{\bar{x}}, v_{\bar{y}}, v_{\bar{\gamma}}) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$ and $(\varepsilon_{\bar{x}}, \varepsilon_{\bar{y}}) \in \Re_+ \times \Re_+$ if the following conditions hold

$$\begin{gathered} v_{\bar{x}} \in \partial_{\varepsilon_{\bar{x}}} f(\bar{x}) - A^*\bar{\gamma}, \qquad v_{\bar{y}} \in \partial_{\varepsilon_{\bar{y}}} g(\bar{y}) - B^*\bar{\gamma}, \qquad v_{\bar{\gamma}} = A\bar{x} + B\bar{y} - b, \\ \max\left\{\|v_{\bar{x}}\|, \|v_{\bar{y}}\|, \|v_{\bar{\gamma}}\|, \varepsilon_{\bar{x}}, \varepsilon_{\bar{y}}\right\} \leq \rho. \end{gathered} \tag{1.9}$$

Note that the latter concept generalizes the former since $\partial h(\cdot) \subset \partial_\varepsilon h(\cdot)$ for any convex function h and $\varepsilon \geq 0$. Indeed, a ρ-approximate solution $(\hat{x}, \hat{y}, \hat{\gamma})$ of (1.7) with residue $(v_{\hat{x}}, v_{\hat{y}}, v_{\hat{\gamma}})$ is a relaxed ρ-approximate solution with residues $(v_{\bar{x}}, v_{\bar{y}}, v_{\bar{\gamma}}) = (v_{\hat{x}}, v_{\hat{y}}, v_{\hat{\gamma}})$ and $(\varepsilon_{\bar{x}}, \varepsilon_{\bar{y}}) = (0, 0)$.

Here, we first analyze an exact proximal generalized ADMM (a version of scheme (1.6) applied to problem (1.1) with proximal terms added to the associated subproblems). This analysis is essential to the subsequent study associated to two new inexact ADMM variants. The first proposed inexact ADMM variant consists of an inexact version of the aforementioned proximal generalized ADMM, whereas the second one is an inexact variant of the proximal ADMM (1.5). These variants are such that their first partial subproblems (corresponding to (1.5a) and (1.6a)) are approximately solved using relative error conditions based on the works of Solodov and Svaiter [71–74]. The proposed schemes are interesting in applications in which a solution to the first partial (proximal) ADMM subproblem can not be easily obtained, whereas the second one is relatively easy to solve. We mention that many real-life applications problems can be approached via ℓ_1-regularized convex composite optimization which in turn can be approximately solved by means of the inexact variants proposed here. In particular, a solution to the corresponding second proximal (generalized) ADMM subproblem can be explicitly computed, see Chapter 6. We mention that in many applications, a solution for the corresponding second (proximal) ADMM subproblem can be explicitly computed; for instance, this is the case for the large class of ℓ_1-regularized convex composite optimization problems.

We show that, for a given tolerance $\rho > 0$, the proposed ADMM variants generate ρ-approximate solutions of (1.7) in at most $\mathcal{O}(1/\rho^2)$ iterations. Moreover, we also show that relaxed ρ-approximate solutions of (1.7) can be obtained by means of auxiliary sequences (generated in an ergodic sense) associated to the proposed schemes in at most $\mathcal{O}(1/\rho)$ iterations. Note that, the latter iteration-complexity bound is better than the former by a factor of $\mathcal{O}(1/\rho)$; however, the inclusions in the ergodic case (see (1.9)) are, in general, weaker than those considered in the pointwise case (see (1.8)). It is worth mentioning that the residuals pairs $(v_{\hat{x}}, v_{\hat{y}})$, $(v_{\bar{x}}, v_{\bar{y}})$, and $(\varepsilon_{\bar{x}}, \varepsilon_{\bar{y}})$ in (1.8)-(1.9) are explicitly computed. Hence, the last condition in (1.8) (resp. (1.9)) can be used as a verifiable pointwise (resp. ergodic) stopping criterion. One of our goal is to show that aforementioned ADMM variants fall within the setting of a hybrid proximal extragradient (HPE) framework[1] since this is an interesting approach to establish iteration-complexity bounds for these schemes in order to obtain approximate solutions of (1.7) in the sense of (1.8)-(1.9).

The last part of this thesis is devoted to the computational study of the proposed inexact ADMM variants. Some numerical experiments performed on a setting of problems derived from real-life applications, such as LASSO and ℓ_1-regularized logistic regression, are considered in order to show the applicability and advantage of these schemes. In particular, we confirm that, similarly to the corresponding exact ADMM versions, the use of $\alpha \approx 1.9$ (rep. $\theta \approx 1.6$) in the inexact proximal generalized ADMM (resp. inexact proximal ADMM) can lead to a better numerical performance.

Finally, the material of this thesis originated three papers which were submitted for publication. Specifically, the material of Chapter 3 is associated to [2], whereas the materials of Chapters 4 and 5 are associated to [1] and [3], respectively.

1.2 Previous most related works

For convenience, we divide this literature review into three parts. First, we review the papers dealing with the standard ADMM and its proximal variants in the exact case. Second, we provide a survey of the literature related to the proximal generalized ADMM. Finally, we discuss the papers about inexact ADMMs.

Standard ADMM and its proximal variants: The first ones to establish iteration-complexity bounds for the standard ADMM (1.4) with $\theta = 1$ were Monteiro and Svaiter [57] (although their analysis assumes $\theta = 1$ and considers only the ergodic case, it can be easily adapted to cover the pointwise case and the use of $\theta < 1$). Subsequently, He and Yuan analyzed ergodic [48] and pointwise [49] iteration-complexities of a partial proximal ADMM (the proximal ADMM (1.5) with $H = 0$ and $\theta = 1$). Pointwise and ergodic iteration-complexity results for the proximal ADMM in its general form (1.5) were considered in [20, 41, 43]. In [40], the authors established iteration-complexity bounds for a variable metric proximal ADMM. It is worth mentioning that all of the aforementioned papers obtain iteration-complexity bounds of the same order than the ones obtained here, i.e., $\mathcal{O}(1/\rho^2)$ to the pointwise case and $\mathcal{O}(1/\rho)$ in the ergodic case. However, it should be mentioned that none of these papers deals with inexact ADMM. In [42], the authors proposed and analyzed

[1]The HPE framework considered here is a slight modification of the well-known HPE scheme first introduced by Solodov and Svaiter [71] for solving monotone inclusion problems. Iteration-complexity bounds for the latter HPE scheme was first established by Monteiro and Svaiter in [56].

a regularized ADMM whose pointwise iteration-complexity bound is better than the one obtained here by an $\mathcal{O}(\varepsilon \log(\varepsilon^{-1}))$ factor. The latter scheme was further explored in [39] by expanding the region in which a relaxation parameter used in the Lagrange multiplier update rule can be chosen. These regularized ADMMs consist of a combination of an inner and an outer procedures, where each of the inner procedure is itself an implementation of a proximal ADMM, whereas the outer one dynamically adjusts a regularization parameter. Although this method has an improved pointwise iteration-complexity, it still lacks of a computational study in order to improve its numerical performance since the aforementioned overall procedure is, in general, time consuming in practice. By assuming that function f in (1.1) is differentiable with Lipschitz continuous gradient, [61] proposed some accelerated ADMM schemes which improve previous convergence rate bounds in terms of the dependence on the Lipschitz constant of the gradient. Finally, under the latter assumptions along with strong convexity of f and certain rank conditions on the matrices A and B, paper [21] established linear convergence rate for a proximal variant of the ADMM. We refer the reader to [2,11,31,39,42,45,52,68] where iteration-complexities of other exact ADMM variants have been considered.

Proximal generalized ADMM: Convergence rates and iteration-complexity of the (exact) generalized ADMM have been recently studied in different contexts (see [19, 31, 59, 75, 76]). However, it should be mentioned that none of these papers is focused on approximately solving the Lagrangian system (1.7) in the sense of (1.8) or (1.9). Namely, paper [31] derived pointwise and ergodic iteration-complexity bounds for the generalized ADMM to obtain an approximate solution of (1.1) in the context of variational inequality, assuming that the matrix B has full column rank. Although its approach is different from ours, it can be shown that its pointwise iteration-complexity bound is similar to the one provided in this thesis. On the other hand, its ergodic iteration-complexity results are based on a termination criterion which can not be easily verifiable and is not directly related to the one considered here. Paper [19] proposed a generalized proximal point algorithm for finding roots of a maximal monotone operator in a Hilbert space and analyzed its convergence rates under different assumptions. In particular, for a given tolerance $\rho > 0$, the authors established an $\mathcal{O}(1/\rho^2)$ pointwise iteration-complexity bound to obtain an approximate solution based on the Yosida approximation of the operator. As a by-product, the same bound can be derived for a especial case of the proximal generalized ADMM considered here. It should be noted, however, that the residual based on the Yosida approximation of the operator is not easy to compute and, hence, it is not clear how their result can be used to obtain and/or identify approximation solutions of (1.7) in the sense of (1.8) or (1.9). The algorithm proposed in [19] was further explored in [76], where the authors established convergence (in the weak and strong topology) of the proposed scheme as well as linear convergence rate. Under the assumptions that A is invertible, B has full column rank, and f is a differentiable strongly convex function with parameter $m > 0$ whose gradient is L-Lipschitz continuous, paper [59] established the linear convergence of a special case of the proximal generalized ADMM studied here with penalty parameter β specifically chosen depending on m, L, and the smallest and largest singular values of the matrix A. Paper [75] analyzed the proximal generalized ADMM as a particular case of a general scheme in a Hilbert space and obtained $\mathcal{O}(1/k)$ ergodic convergence rate by measuring a partial primal-dual gap associated to the Lagrangian function of problem (1.1). The latter result was obtained under the assumption that the operators $(\partial f + \beta A^*A)^{-1}$ and $(\partial g + \beta B^*B)^{-1}$ exist and are Lipschitz continuous which is stronger than the assumption that f and g are convex. Moreover, contrary to our iteration-complexity analysis, the one presented in [75] does not provide any practical termination criterion.

Inexact ADMMs: Inexact variants of the ADMM considering different strategies to compute approximate solutions of its subproblems have been studied in the literature, see for example [26, 29, 30, 58, 80]. Bertsekas and Eckstein in [26] introduced an inexact generalized ADMM whose subproblems are approximately solved using absolute error conditions. In [58], Ng et al. proposed inexact variants of the proximal ADMM (1.4) (with $\theta = 1$ and proximal terms defined by the identity operator) in the setting of variational inequalities, where absolute and relative error criteria were considered. The aforementioned relative error criterion is closely related to the one proposed here. The main advantage of our criterion is that a parameter associated to the criterion is constant (see the parameters τ_1 and τ_2 in (5.2)) whereas the corresponding parameters in the former criterion needs to be square summable, in particular, they vanish asymptotically. This property is too stringent and makes their scheme quite slowly in practice. Most recently, Eckstein and Wang proposed and analyzed other inexact ADMM variants whose subproblems are approximately solved using relative and/or summable error criteria (see [29,30]). Specifically, [29] further developed to the ADMM setting the study of [27], where an inexact augmented Lagrangian method was proposed and analyzed. The main idea of the last references was to approximately solve the associated subproblems using a relative error condition based on the one introduced by Solodov and Svaiter [71–73] in the setting of proximal-point type methods. Numerical comparisons with the inexact ADMM variant proposed in [29] is presented in Chapter 6. Paper [30] proposed a relaxed Douglas–Rachford splitting method for solving (1.2) and derived, as a consequence, a variant of the ADMM which uses, in a special way, a relative error condition.

1.3 Thesis outline

This thesis is organized as follows. Chapter 2 contains preliminary results, notation, basic definitions as well as some assumptions. Chapter 3 is divided into two sections. The first one formally states the proximal generalized ADMM, whereas the second one establishes its pointwise and ergodic iteration-complexity bounds to obtain approximate solution of (1.1) in the sense of (1.8)-(1.9). Chapter 4 and Chapter 5 introduce two new inexact ADMM variants and present their iteration-complexity analysis. Specifically, Chapter 4 is devoted to an inexact proximal generalized ADMM, whereas Chapter 5 deals with an inexact proximal ADMM. Chapter 6 is devoted to numerical experiments. Finally, Chapter 7 contains some concluding remarks.

Chapter 2

Preliminary

This chapter is divided into three sections. The first one presents our notation and basic results. The second section describes a modified HPE framework and its corresponding pointwise and ergodic iteration-complexity bounds for approximately solving a monotone inclusion problem. The last section discuss some concepts of approximate solutions for a monotone inclusion problems as well to the linearly constrained optimization problem (1.1). Some assumptions that will be used throughout this thesis are also considered in this section.

2.1 Notation and basic definitions

In this thesis, $\Re^n$ denotes the usual n-dimensional Euclidean space. The coordinates of a vector $x \in \Re^n$ will be written as $x^1, \ldots, x^n$, i.e., $x = (x^1, \ldots, x^n)$. When $n = 1$, $\Re^1 := \Re$ is the set of real numbers. $\Re_+$ denotes the set of nonnegative real numbers. The p-norm $(p \geq 1)$ and maximum norm of $x \in \Re^n$ are denoted, respectively, by $\|x\|_p = (\sum_{i=1}^n |x^i|^p)^{1/p}$ and $\|x\|_\infty = \max\{|x^1|, \ldots, |x^n|\}$. The index p is omitted when $p = 2$.

Let $\mathcal{V}$ be a finite-dimensional real vector space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively, and let Q be a linear operator on $\mathcal{V}$. Recall that the adjoint of Q is the uniquely determined linear operator satisfying $\langle v, Q\tilde{v} \rangle = \langle Q^* v, \tilde{v} \rangle$, for every $v, \tilde{v} \in \mathcal{V}$. Q^*. When $Q^* = Q$, the operator Q is called self-adjoint. A self-adjoint linear operator $Q\colon \mathcal{V} \to \mathcal{V}$ is said to be positive semidefinite if and only if $\langle Qv, v \rangle \geq 0$, for all $v \in \mathcal{V}$. For a given self-adjoint positive semidefinite linear operator $Q\colon \mathcal{V} \to \mathcal{V}$, the seminorm induced by Q on $\mathcal{V}$ is defined by $\| \cdot \|_Q = \langle Q(\cdot), \cdot \rangle^{1/2}$. Since $\langle Q(\cdot), \cdot \rangle$ is symmetric and bilinear, for all $v, \tilde{v} \in \mathcal{V}$, we have

$$2 \langle Qv, \tilde{v} \rangle \leq \|v\|_Q^2 + \|\tilde{v}\|_Q^2, \qquad \|v + v'\|_Q^2 \leq 2 \left(\|v\|_Q^2 + \|v'\|_Q^2 \right). \tag{2.1}$$

We denote the identity operator on a vector space $\mathcal{V}$ by I.

Given a set-valued operator $T\colon \mathcal{V} \rightrightarrows \mathcal{V}$, its domain and graph are defined, respectively, as

$$\mathrm{Dom}\, T = \{v \in \mathcal{V} \,:\, T(v) \neq \emptyset\} \qquad \text{and} \qquad Gr(T) = \{(v, \tilde{v}) \in \mathcal{V} \times \mathcal{V} \,:\, \tilde{v} \in T(v)\}.$$

The operator T is said to be monotone if

$$\langle u - v, \tilde{u} - \tilde{v} \rangle \geq 0 \qquad \forall\, (u, \tilde{u}),\, (v, \tilde{v}) \,\in\, Gr(T).$$

Moreover, T is maximal monotone if it is monotone and there is no other monotone operator S such that $Gr(T) \subset Gr(S)$. Given a scalar $\varepsilon \geq 0$, the ε-enlargement $T^{[\varepsilon]}\colon \mathcal{V} \rightrightarrows \mathcal{V}$ of the operator T is defined as

$$T^{[\varepsilon]}(v) = \{\tilde{v} \in \mathcal{V} \,:\, \langle \tilde{v} - \tilde{u}, v - u \rangle \geq -\varepsilon,\;\; \forall (u, \tilde{u}) \in Gr(T)\} \quad \forall\, v \in \mathcal{V}. \tag{2.2}$$

The ε-subdifferential of a proper closed convex function $f\colon \mathcal{V} \to [-\infty, \infty]$ is defined by

$$\partial_\varepsilon f(v) = \{u \in \mathcal{V} \,:\, f(\tilde{v}) \geq f(v) + \langle u, \tilde{v} - v \rangle - \varepsilon,\;\; \forall\, \tilde{v} \in \mathcal{V}\} \qquad \forall\, v \in \mathcal{V}.$$

When $\varepsilon = 0$, $\partial_0 f(v)$ is denoted by $\partial f(v)$ and is called the subdifferential of f at v. It is well-known that the subdifferential operator of a proper closed convex function is maximal monotone [65].

The next result is a consequence of the transportation formula in [15, Theorem 2.3] combined with [14, Proposition 2(i)].

Proposition 2.1.1 *Suppose $T\colon \mathcal{V} \rightrightarrows \mathcal{V}$ is maximal monotone and let $\tilde{v}_i, v_i \in \mathcal{V}$, for $i = 1, \ldots, k$, be such that $v_i \in T(\tilde{v}_i)$ and define*

$$\tilde{v}_k^a = \frac{1}{k} \sum_{i=1}^{k} \tilde{v}_i, \qquad v_k^a = \frac{1}{k} \sum_{i=1}^{k} v_i, \qquad \varepsilon_k^a = \frac{1}{k} \sum_{i=1}^{k} \langle v_i, \tilde{v}_i - \tilde{v}_k^a \rangle.$$

Then, the following hold:

(a) *$\varepsilon_k^a \geq 0$ and $v_k^a \in T^{[\varepsilon_k^a]}(\tilde{v}_k^a)$;*

(b) *if, in addition, $T = \partial f$ for a proper closed and convex function f, then $v_k^a \in \partial_{\varepsilon_k^a} f(\tilde{v}_k^a)$.*

2.2 A modified HPE framework

Our problem of interest in this section is the monotone inclusion problem

$$0 \in T(z), \tag{2.3}$$

where $T\colon \mathcal{Z} \rightrightarrows \mathcal{Z}$ is a maximal monotone operator and $\mathcal{Z}$ is a finite-dimensional real vector space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We consider the following basic assumption:

Assumption 2.2.1 The solution set of (2.3), denoted by $T^{-1}(0)$, is nonempty.

A classic iterative scheme applied to solve (2.3) is the proximal point method [55], which, starting from an initial point $z_0 \in \mathcal{Z}$, generates a sequence $\{z_k\}$ satisfying

$$z_k = (I + \lambda_k T)^{-1} (z_{k-1}),$$

where $\lambda_k > 0$ is a parameter. Often, in applications, it can be difficult to explicitly obtain the resolvent operator $(I + \lambda T)^{-1}$, so that some inexact versions of the proximal point method were considered. In [66] Rockafellar proposed an inexact proximal point method which allows $\{z_k\}$ to be computed such that

$$\left\| z_k - (I + \lambda_k T)^{-1} (z_{k-1}) \right\| \leq e_k, \qquad \sum_{k=1}^{\infty} e_k < \infty,$$

where λ_k is bounded away from zero, and $\{e_k\}$ is a non-negative sequence of error tolerances. More recently, there is a growing interest in inexact versions that use relative error criteria instead of absolute error. In this sense, the hybrid proximal extragradient (HPE) method proposed by Solodov and Svaiter in [71] (see also [72–74]) suggests, in each iteration, to find a triple $(\tilde{z}_k, v_k, \varepsilon_k) \in \mathcal{Z} \times \mathcal{Z} \times \Re_+$ and $\lambda_k > 0$ such that

$$v_k \in T^{[\varepsilon_k]}(\tilde{z}_k), \qquad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma \|\tilde{z}_k - z_{k-1}\|^2, \tag{2.4}$$

where $\sigma \in [0, 1)$ is a error tolerance parameter. The new iteration z_k is then defined as $z_k = z_{k-1} - \lambda_k v_k$. If $\sigma = 0$, it follows easily that for every $k \geq 1$, $\varepsilon_k = 0$ and $\tilde{z}_k = z_k$, and hence

$$v_k \in T(z_k), \qquad \lambda_k v_k + z_k - z_{k-1} = 0,$$

which is equivalente to the exact iteration of the proximal point method. Thus, we can conclude that, by increasing the value of σ in the interval $[0, 1)$, the HPE method (2.4) allows a growing relaxation in inclusion and/or equation of the above system. Monteiro and Svaiter in [56] established iteration-complexity results for the HPE method (2.4). Since then, iteration-complexity of other HPE-type methods have been considered in the literature (see, e.g., [42, 50, 54, 57, 74]).

In the following, we formally describe a modified HPE framework for computing approximate solutions of (2.3) which will be essential to characterize and analyze the algorithms considered in this thesis.

Modified HPE framework.

0. Let $z_0 \in \mathcal{Z}$, $\eta_0 \in \Re_+$, $\sigma \in [0,1]$ and a self-adjoint positive semidefinite linear operator $M\colon \mathcal{Z} \to \mathcal{Z}$ be given, and set $k = 1$;

1. obtain $(z_k, \tilde{z}_k, \eta_k) \in \mathcal{Z} \times \mathcal{Z} \times \Re_+$ such that

$$M(z_{k-1} - z_k) \in T(\tilde{z}_k), \tag{2.5a}$$

$$\|\tilde{z}_k - z_k\|_M^2 + \eta_k \le \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 + \eta_{k-1}; \tag{2.5b}$$

2. set $k \leftarrow k+1$ and go to step 1.

end

Remark 2.2.2 Some remarks about the modified HPE framework are in order:

(a) It is an instance of the non-Euclidean HPE framework of [41] with $\lambda_k = 1, \varepsilon_k = 0$ and $(dw)_z(z') = (1/2)\|z - z'\|_M^2$, for every $z, z' \in \mathcal{Z}$. Note that, the distance generating function $w(\cdot) = (1/2)\|\cdot\|_M^2$ is a $(1,1)$-regular with respect to $(\mathcal{Z}, \|\cdot\|_M)$ in the sense of [41, Definition 3.2].

(b) The way to obtain $(z_k, \tilde{z}_k, \eta_k)$ will depend on the particular instance of the framework and properties of the operator T. In later chapters, we will show that the proposed variants of ADMM can be seen as an instance of the modified HPE framework specifying, in particular, how this triple $(z_k, \tilde{z}_k, \eta_k)$ can be obtained in this context.

(c) The inclusion in (2.5a) can be interpreted as a generalized proximal inclusion where the pair $(z_k, \tilde{z}_k)$ is controlled according to the relative error condition in (2.5b). Indeed, if M is positive definite and $\sigma = \eta_0 = 0$, then (2.5b) implies that $\eta_k = 0$ and $z_k = \tilde{z}_k$ for every $k \ge 1$, and hence that $M(z_{k-1} - z_k) \in T(z_k)$ in view of (2.5a). In particular, if $M = I$ and $\sigma = \eta_0 = 0$, then (2.5) implies that $\eta_k = 0$, $z_k = \tilde{z}_k$ and $0 \in z_k - z_{k-1} + T(z_k)$ for every $k \ge 1$, which corresponds to the proximal point method to solve problem (2.3). Therefore, the HPE error conditions (2.5) can be viewed as a relaxation of an iteration of the exact proximal point method. It is worth mentioning that the use of a positive semidefinite operator M instead of a positive definite is essential in the analysis discussed in the next chapters. More examples of algorithms which can be seen as special cases of HPE-type frameworks can be found in [56, 57, 71].

(d) In view of Assumption 2.2.1 and the first remark above, it follows from [41, Lemma 3.6(d)] that the sequence $\{z_k\}$ is bounded when M is positive definite. On the other hand, if the solution set of (2.3) is empty, then $\{z_k\}$ may be unbounded; see, for example, [71, Theorem 3.1], where is shown that the sequence generated by a special case of the modified HPE framework has this behavior.

It should be noted that all results given in this section are derived from [41] (incluing the modified HPE framework as mentioned in Remark 2.2.2(a)). Due to the relevance of this framework in the analysis of the ADMM variants considered here, and also for completeness and convenience of the reader, we formally present a simplified proofs of these facts.

The next result summarizes some useful properties about the sequence generated by the modified HPE framework (see [41, Lemma 3.6]).

Lemma 2.2.3 *Let $\{(z_k, \tilde{z}_k, \eta_k)\}$ be the sequence generated by the modified HPE framework. For every $k \geq 1$, the following statements hold:*

(a) *for every $z \in \mathcal{Z}$, we have*

$$\|z - z_k\|_M^2 + \eta_k \leq (\sigma - 1)\|\tilde{z}_k - z_{k-1}\|_M^2 + \|z - z_{k-1}\|_M^2 + 2\langle M(z_{k-1} - z_k), z - \tilde{z}_k\rangle + \eta_{k-1};$$

(b) *for every $z^* \in T^{-1}(0)$, we have*

$$\|z^* - z_k\|_M^2 + \eta_k \leq (\sigma - 1)\|\tilde{z}_k - z_{k-1}\|_M^2 + \|z^* - z_{k-1}\|_M^2 + \eta_{k-1} \leq \|z^* - z_{k-1}\|_M^2 + \eta_{k-1}.$$

Proof. (a) Note that, for every $z \in \mathcal{Z}$,

$$\begin{aligned}\|z - z_k\|_M^2 - \|z - z_{k-1}\|_M^2 &= \|(z - \tilde{z}_k) + (\tilde{z}_k - z_k)\|_M^2 - \|(z - \tilde{z}_k) + (\tilde{z}_k - z_{k-1})\|_M^2 \\ &= \|\tilde{z}_k - z_k\|_M^2 - \|\tilde{z}_k - z_{k-1}\|_M^2 + 2\langle M(z_{k-1} - z_k), z - \tilde{z}_k\rangle,\end{aligned}$$

which, combined with (2.5b), proves the desired inequaliy.
(b) Since $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ and $0 \in T(z^*)$, we have $\langle M(z_{k-1} - z_k), \tilde{z}_k - z^*\rangle \geq 0$. Hence, the first inequality in (b) follows from (a) with $z = z^*$. Now, the second inequality in (b) follows from the fact that $\sigma \leq 1$. ■

Lemma 2.2.3(b) is closely related to the well-known quasi-Fejér inequality which can be used to show that $\{z_k\}$ converges to a point in $T^{-1}(0)$ when M is positive definite.

2.2.1 Iteration-complexity of the modified HPE framework

In order to present pointwise and ergodic iteration-complexity results for the modified HPE framework, the following scalar needs to be defined

$$d_0 = \inf\{\|z^* - z_0\|_M^2 : z^* \in T^{-1}(0)\}, \tag{2.6}$$

where M is given in step 0 of the modified HPE framework.

We first consider the pointwise case (see [41, Theorem 3.3(b)]).

Theorem 2.2.4 *Consider the sequence $\{(z_k, \tilde{z}_k, \eta_k)\}$ generated by the modified HPE framework with $\sigma < 1$. Then, for every $k \geq 1$, there hold $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ and there exists $i \leq k$ such that*

$$\|z_i - z_{i-1}\|_M \leq \frac{1}{\sqrt{k}} \sqrt{\frac{2(1+\sigma)d_0 + 4\eta_0}{1-\sigma}},$$

where d_0 is as defined in (2.6).

Proof. The inclusion $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ holds due to (2.5a). It follows from the second property in (2.1) with $Q = M$ that, for every $j \geq 1$,

$$\|z_j - z_{j-1}\|_M^2 \leq 2(\|\tilde{z}_j - z_{j-1}\|_M^2 + \|\tilde{z}_j - z_j\|_M^2) \leq 2(1+\sigma)\|\tilde{z}_j - z_{j-1}\|_M^2 + 2(\eta_{j-1} - \eta_j)$$

where the last inequality is due to (2.5b). Now, if $z^* \in T^{-1}(0)$, we obtain from Lemma 2.2.3(b)

$$(1-\sigma)\|\tilde{z}_j - z_{j-1}\|_M^2 \leq \|z^* - z_{j-1}\|_M^2 - \|z^* - z_j\|_M^2 + \eta_{j-1} - \eta_j, \quad \forall j \geq 1.$$

Combining the last two estimates, we get

$$\begin{aligned}(1-\sigma)\sum_{j=1}^{k}\|z_j - z_{j-1}\|_M^2 &\leq 2(1+\sigma)\left(\|z^* - z_0\|_M^2 - \|z^* - z_k\|_M^2 + \eta_0 - \eta_k\right) + 2(1-\sigma)(\eta_0 - \eta_k)\\ &\leq 2(1+\sigma)\|z^* - z_0\|_M^2 + 4\eta_0.\end{aligned}$$

Hence, as $\sigma < 1$, we obtain

$$\min_{i=1,\ldots,k}\|z_i - z_{i-1}\|_M^2 \leq \frac{1}{k(1-\sigma)}\left(2(1+\sigma)\|z^* - z_0\|_M^2 + 4\eta_0\right).$$

Therefore, the desired inequality follows from the latter inequality and the definition of d_0 given in (2.6).

■

Corollary 2.2.5 *Consider the sequence $\{(z_k, \tilde{z}_k, \eta_k)\}$ generated by the modified HPE framework with $\sigma < 1$, and asssume that the sequence $\{\|z_k - z_{k-1}\|_M\}$ is nonincreasing. Then, for every $k \geq 1$, there hold $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ and*

$$\|z_k - z_{k-1}\|_M \leq \frac{1}{\sqrt{k}} \sqrt{\frac{2(1+\sigma)d_0 + 4\eta_0}{1-\sigma}},$$

where d_0 is as defined in (2.6).

Proof. This result follows immediately from Theorem 2.2.4 noting that, for every $k \geq 1$, $\min_{i=1,\ldots,k} \|z_i - z_{i-1}\|_M^2 = \|z_k - z_{k-1}\|_M^2$. ■

Remark 2.2.6 For a given tolerance $\bar{\rho} > 0$, it follows from Theorem 2.2.4 that in at most $\mathcal{O}(1/\bar{\rho}^2)$ iterations, the modified HPE framework computes an approximate solution $\tilde{z}$ of (2.3) and a residual r in the sense that $Mr \in T(\tilde{z})$ and $\|r\|_M \leq \bar{\rho}$. Although M is assumed to be only positive semidefinite, if $\|r\|_M = 0$, then $M^{1/2}r = 0$ which, in turn, implies that $Mr = 0$. Hence, the latter inclusion implies that $\tilde{z}$ is a solution of problem (2.3). Therefore, the aforementioned concept of approximate solutions makes sense.

Let $\{(z_k, \tilde{z}_k, \eta_k)\}$ be the sequence generated by the modified HPE framework. In order to present the ergodic case (see [41, Theorem 3.4]), consider the ergodic sequences $\{(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)\}$ defined by

$$\tilde{z}_k^a = \frac{1}{k}\sum_{i=1}^{k}\tilde{z}_i, \quad r_k^a = \frac{1}{k}\sum_{i=1}^{k}(z_{i-1} - z_i), \quad \varepsilon_k^a = \frac{1}{k}\sum_{i=1}^{k}\langle M(z_{i-1} - z_i), \tilde{z}_i - \tilde{z}_k^a\rangle, \quad \forall\, k \geq 1. \tag{2.7}$$

Theorem 2.2.7 *Let $\sigma \in [0,1]$ and consider the ergodic sequence $\{(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)\}$ as in (2.7). Then, for every $k \geq 1$, there hold $\varepsilon_k^a \geq 0$, $Mr_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$ and*

$$\|r_k^a\|_M \leq \frac{2\sqrt{d_0 + \eta_0}}{k}, \quad \varepsilon_k^a \leq \frac{3\,[3(d_0 + \eta_0) + \sigma\tilde{\rho}_k]}{2k}$$

where

$$\tilde{\rho}_k := \max_{i=1,\ldots,k} \|\tilde{z}_i - z_{i-1}\|_M^2, \tag{2.8}$$

and d_0 is as defined in (2.6). Moreover, the sequence $\{\tilde{\rho}_k\}$ is bounded under either one of the following situations:

(a) *$\sigma < 1$, in which case*

$$\tilde{\rho}_k \leq \frac{d_0 + \eta_0}{1 - \sigma}; \tag{2.9}$$

(b) *$\operatorname{Dom} T := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}$ is bounded, in which case*

$$\tilde{\rho}_k \leq 2[d_0 + \eta_0 + \widetilde{D}],$$

where $\widetilde{D} := \sup\{\|y' - y\|_M^2 : y, y' \in \operatorname{Dom} T\}$.

Proof. The inequality $\varepsilon_k^a \geq 0$ and inclusion $Mr_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$ follow from (2.5a), (2.7), and Theorems 2.1.1(a). Using (2.7), it is easy see that for any $z^* \in T^{-1}(0)$

$$kr_k^a = z_k - z_0 = (z^* - z_0) + (z_k - z^*).$$

Hence, from the second inequality in (2.1) with $Q = M$ and Lemma 2.2.3(b), we have

$$k^2 \|r_k^a\|_M^2 \leq 2(\|z^* - z_0\|_M^2 + \|z^* - z_k\|_M^2) \leq 4(\|z^* - z_0\|_M^2 + \eta_0).$$

Combining the above inequality with definition of d_0, we obtain the bound on $\|r_k^a\|_M$. Let us now to prove the bound on ε_k^a. From Lemma 2.2.3(a), we have

$$2\sum_{i=1}^{k} \langle M(z_{i-1} - z_i), \tilde{z}_i - z \rangle \leq \|z - z_0\|_M^2 - \|z - z_k\|_M^2 + \eta_0 - \eta_k \leq \|z - z_0\|_M^2 + \eta_0,$$

for every $z \in \mathcal{Z}$. Letting $z = \tilde{z}_k^a$ and using (2.7), we get

$$2k\varepsilon_k^a \leq \|\tilde{z}_k^a - z_0\|_M^2 + \eta_0 \leq \frac{1}{k}\sum_{i=1}^{k} \|\tilde{z}_i - z_0\|_M^2 + \eta_0 \leq \max_{i=1,\ldots,k} \|\tilde{z}_i - z_0\|_M^2 + \eta_0 \tag{2.10}$$

where the second inequality is due to convexity of the function $\|\cdot\|_M^2$, which also implies that, for every $i \geq 1$ and $z^* \in T^{-1}(0)$,

$$\|\tilde{z}_i - z_0\|_M^2 \leq 3\left[\|\tilde{z}_i - z_i\|_M^2 + \|z^* - z_i\|_M^2 + \|z^* - z_0\|_M^2\right].$$

Hence, using (2.5b) and twice Lemma 2.2.3(b), it follows, for every $i \geq 1$ and $z^* \in T^{-1}(0)$, that

$$\begin{aligned}
\|\tilde{z}_i - z_0\|_M^2 &\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|_M^2 + \eta_{i-1} + \|z^* - z_{i-1}\|_M^2 + \eta_{i-1} + \|z^* - z_0\|_M^2\right] \\
&\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|_M^2 + 2(\|z^* - z_{i-1}\|_M^2 + \eta_{i-1}) + \|z^* - z_0\|_M^2\right] \\
&\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|_M^2 + 3\|z^* - z_0\|_M^2 + 2\eta_0\right],
\end{aligned}$$

which, combined with (2.10) and definitions of $\tilde{\rho}_k$ in (2.8), yields

$$2k\varepsilon_k^a \leq 3\left[3\|z^* - z_0\|_M^2 + \sigma\tilde{\rho}_k\right] + 7\eta_0 \leq 3\left[3(\|z^* - z_0\|_M^2 + \eta_0) + \sigma\tilde{\rho}_k\right].$$

Thus, the bound on ε_k^a now follows from the definition of the d_0 in (2.6).

It remains to prove the second part of the theorem.

(a) if $\sigma < 1$, then it follows from Lemma 2.2.3(b), for every $i \geq 1$ and $z^* \in T^{-1}(0)$, that

$$(1 - \sigma)\|\tilde{z}_i - z_{i-1}\|_M^2 \leq \|z^* - z_{i-1}\|_M^2 + \eta_{i-1} \leq \|z^* - z_0\|_M^2 + \eta_0.$$

Hence, in view of definitions of $\tilde{\rho}_k$ and d_0, we obtain (2.9).

(b) If $\operatorname{Dom} T$ is bounded, then it follows from the second inequality in (2.1) with $Q = M$, and Lemma 2.2.3(b), for every $i \geq 1$ and $z^* \in T^{-1}(0)$, that

$$\|\tilde{z}_i - z_{i-1}\|_M^2 \leq 2\left[\|z^* - z_{i-1}\|_M^2 + \|\tilde{z}_i - z^*\|_M^2\right] \leq 2\left[\|z^* - z_0\|_M^2 + \eta_0 + \widetilde{D}\right]$$

which, combined with definitions of $\tilde{\rho}_k$ and d_0, proves the desired result. ■

If $\sigma < 1$ or $\operatorname{Dom} T$ is bounded, it follows from Theorem 2.2.7 that $\{\tilde{\rho}_k\}$ is bounded and hence $\max\{\|v_k^a\|_M, \varepsilon_k^a\} = \mathcal{O}(1/k)$. However, it may happen that the sequence $\{\tilde{\rho}_k\}$ is bounded even when $\sigma = 1$. Indeed, in Chapter 3, we will show that this is the case for the proximal generalized ADMM, which is an instance of the modified HPE framework.

Remark 2.2.8 For a given tolerance $\bar{\rho} > 0$, Theorem 2.2.7 ensures that in at most $\mathcal{O}(1/\bar{\rho})$ iterations of the modified HPE framework, the triple $(\tilde{z}, r, \varepsilon) := (\tilde{z}_k^a, r_k^a, \varepsilon_k^a)$ satisfies $Mr \in T^{\varepsilon}(\tilde{z})$ and $\max\{\|r\|_M, \varepsilon\} \leq \bar{\rho}$. Similarly to Remark 2.2.6, the point $\tilde{z}$ can be interpreted as an approximate solution of (2.3). Note that, the above ergodic complexity bound is better than the pointwise one by a factor of $\mathcal{O}(1/\bar{\rho})$; however, the above inclusion is, in general, weaker than that of the pointwise case.

2.3 Elementary concepts

In this section, we introduce a maximal monotone operator constructed from the Lagrangian system (1.7), which will be used throughout this thesis.

We assume that $\mathcal{Z} := \mathcal{X} \times \mathcal{Y} \times \Gamma$ and $T \colon \mathcal{Z} \rightrightarrows \mathcal{Z}$ is the operator defined as

$$T(x, y, \gamma) = \begin{bmatrix} \partial f(x) - A^*\gamma \\ \partial g(y) - B^*\gamma \\ Ax + By - b \end{bmatrix}. \tag{2.11}$$

Since f and g are proper, closed and convex functions, the operators ∂f and ∂g are maximal monotone (see [64]), hence the operator T is maximal monotone. Indeed, the maximal monotonicity of the operator T in (2.11) follows from the fact that T can be decomposed as $T = \widetilde{T} + \widehat{T}$, where $\widetilde{T} \colon \mathcal{Z} \rightrightarrows \mathcal{Z}$ is the multi-valued map given by

$$\widetilde{T}(x, y, \gamma) = \partial f(x) \times \partial g(y) \times \{-b\}$$

and $\widehat{T} \colon \mathcal{Z} \to \mathcal{Z}$ is the linear operator given by

$$\widehat{T}(x, y, \gamma) = (-A^*\gamma, -B^*\gamma, Ax + By)$$

(note that $\widehat{T}$ is skew-symmetric, i.e., $\langle \widehat{T}z, \tilde{z}\rangle = -\langle z, \widehat{T}\tilde{z}\rangle$ for all $z, \tilde{z} \in \mathcal{Z}$).

Throughout this thesis, we also consider the following basic assumption.

Assumption 2.3.1 The solution set of the Lagrangian system (1.7), denoted by Ω^*, is nonempty.

Note that $(x^*, y^*, \gamma^*) \in \Omega^*$ if and only if $0 \in T(z^*)$, where $z^* := (x^*, y^*, \gamma^*)$ and T is as defined above. Moreover, as previously mentioned, it is well-known that $(x^*, y^*, \gamma^*) \in \Omega^*$ if and only if (x^*, y^*) is a solution to problem (1.1) and γ^* is an associated Lagrange multiplier.

For convenience, we rewrite the concept of approximate solutions (1.8) of the Lagrangian system in terms of the operator T given in (2.11). This is convenient in order to obtain the pointwise iteration-complexity bounds of some ADMM variants in the setting of the modified HPE framework. Similarly, we could consider a concept of approximate solution closely related to the relaxed approximate solution (1.9) in terms of the enlargement $T^{[\varepsilon]}$ of the aforementioned operator T. However, this latter concept is a bit more general and is more useful when analyzing ergodic sequences derived from general instances of HPE-type methods. In the case of the ADMM variants considered in this thesis, we will be able to present a more refined analysis in order to avoid the use of this general enlargement operator, using instead the ε-subdifferential of the functions f and g. This will provide a sharper ergodic iteration-complexity bound for the ADMM variants studied here.

Definition 2.3.2 *Given a tolerance $\rho > 0$, a triple $(x, y, \gamma) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$ is said to be a ρ-approximate solution of (1.7) with residue r if*

$$r \in T(x, y, \gamma) \qquad \text{and} \qquad \|r\| \leq \rho, \tag{2.12}$$

where T is as in (2.11).

Obviously, a triple $(x^*, y^*, \gamma^*) \in \Omega^*$ if and only if $0 \in T(x^*, y^*, \gamma^*)$. Hence, for all $\rho > 0$, any element in Ω^* is a ρ-approximate solution with residue 0.

Chapter 3

Iteration-complexity analysis of the proximal generalized ADMM

This chapter is devoted to the iteration-complexity analysis of the proximal generalized ADMM and is related to paper [2]. In Section 3.1, we formally state the method (Algorithm 1). In Section 3.2, we present the iteration-complexity analysis of the method. This section is divided into two subsections. Subsection 3.2.1 presents some technical results and shows that the proximal generalized ADMM is an instance of the modified HPE framework, whereas Subsection 3.2.2 establishes its pointwise and ergodic iteration-complexity results.

3.1 Proximal generalized ADMM (PG-ADMM)

In the following, we formally state the proximal generalized ADMM for solving (1.1).

Algorithm 1: Proximal generalized ADMM

0. Let an initial point $(x_0, y_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$, a penalty parameter $\beta > 0$, a relaxation factor $\alpha \in (0, 2]$, and two self-adjoint positive semidefinite linear operators $G \colon \mathcal{X} \to \mathcal{X}$ and $H \colon \mathcal{Y} \to \mathcal{Y}$ be given, and set $k = 1$.

1. Compute an optimal solution $x_k \in \mathcal{X}$ of the subproblem

$$\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 + \frac{1}{2} \|x - x_{k-1}\|_G^2 \right\} \tag{3.1}$$

and compute an optimal solution $y_k \in \mathcal{Y}$ of the subproblem

$$\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|\alpha(Ax_k + By_{k-1} - b) + B(y - y_{k-1})\|^2 \right. \\ \left. + \frac{1}{2} \|y - y_{k-1}\|_H^2 \right\}. \tag{3.2}$$

2. Set

$$\gamma_k = \gamma_{k-1} - \beta[\alpha(Ax_k + By_{k-1} - b) + B(y_k - y_{k-1})] \tag{3.3}$$

and $k \leftarrow k + 1$, and go to step (1).

Remark 3.1.1 Algorithm 1 has different features depending on the choices of the matrices G, H, and the relaxation factor α. For instance, by taking $\alpha = 1$ and $(G, H) = (0, 0)$, it reduces to the standard ADMM (1.4). By choosing $(G, H) = (\tau_1 I - \beta A^* A, \tau_2 I - \beta B^* B)$ with $\tau_1 \geq \beta \|A^* A\|$ and $\tau_2 \geq \beta \|B^* B\|$, it reduces to a linearized ADMM with a relaxation parameter. The latter method cancels the quadratic terms $(\beta/2)\|Ax\|^2$ and $(\beta/2)\|By\|^2$ in (3.1) and (3.2), respectively. More specifically, the subproblems (3.1) and (3.2) become

$$\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1} - \beta(Ax_{k-1} + By_{k-1} - b), Ax \rangle + \frac{\tau_1}{2} \|x - x_{k-1}\|^2 \right\},$$

and

$$\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1} - \alpha\beta(Ax_k + By_{k-1} - b), By \rangle + \frac{\tau_2}{2} \|y - y_{k-1}\|^2 \right\}.$$

In many applications, the above subproblems are much easier to solve or even have closed-form solutions (see [48,79,83] for more details). We also mention that depending on the structure of problem (1.1), other choices of G and H may be recommended; see, for instance, [21] (although the latter reference considers $\alpha = 1$, it is clear that the same discussion regarding the choices of G and H holds for arbitrary $\alpha \in (0, 2)$). In some applications, the use of an over-relaxation parameter ($\alpha > 1$) leads to a better numerical performance than the standard ADMM; see, for example, [9,24,31] and Chapter 6, where some numerical experiments are reported in order to illustrate the performance of Algorithm 1 with different choices of the relaxation parameter α.

3.2 Iteration-complexity of the PG-ADMM

This section presents pointwise and ergodic iteration-complexity bounds for Algorithm 1. Our approach consists of interpreting Algorithm 1 as an instance of the modified HPE framework with a very special property, namely, a key parameter sequence $\{\tilde{\rho}_k\}$ associated to the sequence generated by the method is upper bounded by a multiple of d_0 (a parameter measuring, in some sense, the distance of the initial point to the solution set), see Lemma 3.2.7. This property is essential to obtain the ergodic iteration-complexity of Algorithm 1.

3.2.1 The PG-ADMM as an instance of the modified HPE framework

Our aim in this subsection is to show that the PG-ADMM is an instance of the modified HPE framework for solving problem (1.7).

Let us first introduce the elements required by the setting of Section 2.2. Consider the linear operator

$$M := \begin{bmatrix} G & 0 & 0 \\ 0 & (H + \frac{\beta}{\alpha} B^* B) & \frac{(1-\alpha)}{\alpha} B^* \\ 0 & \frac{(1-\alpha)}{\alpha} B & \frac{1}{\alpha\beta} I \end{bmatrix}, \tag{3.4}$$

and the quantity

$$d_0 := \inf_{(x,y,\gamma)\in T^{-1}(0)} \left\{ \|(x - x_0, y - y_0, \gamma - \gamma_0)\|_M^2 \right\}, \tag{3.5}$$

where T is as in (2.11). It is easy to verify that M is a self-adjoint positive semidefinite linear operator for every $\beta > 0$ and $\alpha \in (0, 2]$. Let $\{(x_k, y_k, \gamma_k)\}$ be the sequence generated by Algorithm 1. In order to simplify some relations in the results below, define the sequence $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k, \tilde{\gamma}_k)\}$ as

$$\begin{aligned} &\Delta x_k = x_k - x_{k-1}, \qquad \Delta y_k = y_k - y_{k-1}, \\ &\Delta\gamma_k = \gamma_k - \gamma_{k-1}, \qquad \tilde{\gamma}_k = \gamma_{k-1} - \beta(Ax_k + By_{k-1} - b), \qquad \forall\, k \geq 1. \end{aligned} \tag{3.6}$$

We next present two technical results.

Lemma 3.2.1 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1 and consider the sequence $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k, \tilde{\gamma}_k)\}$ as in (3.6). Then, for every $k \geq 1$,*

$$\tilde{\gamma}_k - \gamma_{k-1} = \frac{1}{\alpha}\left[\Delta\gamma_k + \beta B\Delta y_k\right], \tag{3.7}$$

$$0 \in G\Delta x_k + \left[\partial f(x_k) - A^*\tilde{\gamma}_k\right], \tag{3.8}$$

$$0 \in (H + \frac{\beta}{\alpha}B^*B)\Delta y_k + \frac{(1-\alpha)}{\alpha}B^*\Delta\gamma_k + \left[\partial g(y_k) - B^*\tilde{\gamma}_k\right], \tag{3.9}$$

$$0 = \frac{(1-\alpha)}{\alpha}B\Delta y_k + \frac{1}{\alpha\beta}\Delta\gamma_k + \left[Ax_k + By_k - b\right]. \tag{3.10}$$

As a consequence, $z_k := (x_k, y_k, \gamma_k)$ and $\tilde{z}_k := (x_k, y_k, \tilde{\gamma}_k)$ satisfy the inclusion (2.5a) with T and M as in (2.11) and (3.4), respectively.

Proof. It follows from the definitions of γ_k and $\tilde{\gamma}_k$ in (3.3) and (3.6), respectively, that

$$\frac{1}{\alpha}(\gamma_k - \gamma_{k-1}) + \frac{\beta}{\alpha}B(y_k - y_{k-1}) = -\beta(Ax_k + By_{k-1} - b) = \tilde{\gamma}_k - \gamma_{k-1},$$

which, combined with the definitions of Δy_k and $\Delta\gamma_k$ in (3.6), proves (3.7). From the optimality condition for (3.1), we have

$$0 \in \partial f(x_k) - A^*(\gamma_{k-1} - \beta(Ax_k + By_{k-1} - b)) + G(x_k - x_{k-1}),$$

which, combined with the definitions of $\tilde{\gamma}_k$ and Δx_k in (3.6), yields (3.8). Similarly, from the optimality condition for (3.2) and definitions of γ_k and Δy_k in (3.3) and (3.8), respectively, we obtain

$$\begin{aligned} 0 &\in \partial g(y_k) - B^*\left[\gamma_{k-1} - \beta[\alpha(Ax_k + By_{k-1} - b) + B(y_k - y_{k-1})]\right] + H(y_k - y_{k-1}) \\ &= \partial g(y_k) - B^*\gamma_k + H\Delta y_k. \end{aligned} \tag{3.11}$$

On the other hand, note that (3.7) implies that

$$\gamma_k = \tilde{\gamma}_k + (\gamma_k - \gamma_{k-1}) - (\tilde{\gamma}_k - \gamma_{k-1}) = \tilde{\gamma}_k - \frac{(1-\alpha)}{\alpha}\Delta\gamma_k - \frac{\beta}{\alpha}B\Delta y_k,$$

which in turn, combined with (3.11), gives (3.9). The relation (3.10) follows immediately from (3.3).

Now, the last statement of the lemma follows directly by (3.8)–(3.10) and the definitions of T and M given in (2.11) and (3.4), respectively. ∎

Lemma 3.2.2 *The sequences $\{\Delta y_k\}$ and $\{\Delta \gamma_k\}$ defined in (3.6) satisfy*

$$2\langle B\Delta y_1, \Delta\gamma_1\rangle \geq \|\Delta y_1\|_H^2 - 4d_0, \quad 2\langle B\Delta y_k, \Delta\gamma_k\rangle \geq \|\Delta y_k\|_H^2 - \|\Delta y_{k-1}\|_H^2 \quad \forall k \geq 2, \tag{3.12}$$

where d_0 is as in (3.5).

Proof. Let a point $z^* := (x^*, y^*, \gamma^*)$ be such that $0 \in T(x^*, y^*, \gamma^*)$ (see Assumption 2.3.1) and consider $z_i := (x_i, y_i, \gamma_i)$, $i = 0, 1$. First, note that

$$0 \leq \frac{\beta}{\alpha}\|B\Delta y_1\|^2 + \frac{2}{\alpha}\langle B\Delta y_1, \Delta\gamma_1\rangle + \frac{1}{\alpha\beta}\|\Delta\gamma_1\|^2,$$

where Δy_1 and $\Delta\gamma_1$ are as in (3.6). Hence, by adding $\|\Delta y_1\|_H^2 - 2\langle B\Delta y_1, \Delta\gamma_1\rangle$ to both sides of the above inequality, we obtain

$$\begin{aligned}\|\Delta y_1\|_H^2 - 2\langle B\Delta y_1, \Delta\gamma_1\rangle &\leq \|\Delta y_1\|_H^2 + \frac{\beta}{\alpha}\|B\Delta y_1\|^2 + 2\frac{(1-\alpha)}{\alpha}\langle B\Delta y_1, \Delta\gamma_1\rangle + \frac{1}{\alpha\beta}\|\Delta\gamma_1\|^2 \\ &\leq \|z_1 - z_0\|_M^2 \leq 2\left(\|z^* - z_1\|_M^2 + \|z^* - z_0\|_M^2\right),\end{aligned} \tag{3.13}$$

where M is as in (3.4) and the last inequality is a consequence of the second property in (2.1) with $Q = M$. On the other hand, taking $\tilde{z}_1 = (x_1, y_1, \tilde{\gamma}_1)$, Lemma 3.2.1 implies that $(z_0, z_1, \tilde{z}_1)$ satisfies (2.5a) with T and M as in (2.11) and (3.4), respectively; namely, $M(z_0 - z_1) \in T(\tilde{z}_1)$. Hence, since $0 \in T(z^*)$ and T is monotone, we obtain $\langle M(z_0 - z_1), \tilde{z}_1 - z^*\rangle \geq 0$. Thus, it follows that

$$\begin{aligned}\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 &= \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_1)\|_M^2 - \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_0)\|_M^2 \\ &= \|\tilde{z}_1 - z_1\|_M^2 + 2\langle M(z_0 - z_1), z^* - \tilde{z}_1\rangle - \|\tilde{z}_1 - z_0\|_M^2 \\ &\leq \|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2.\end{aligned} \tag{3.14}$$

Combining (3.6) and (3.7), we have $\tilde{\gamma}_1 - \gamma_1 = [(1-\alpha)\Delta\gamma_1 + \beta B\Delta y_1]/\alpha$. Hence, using the definitions of M, z_1 and $\tilde{z}_1$, we obtain

$$\|\tilde{z}_1 - z_1\|_M^2 = \frac{1}{\alpha\beta}\|\tilde{\gamma}_1 - \gamma_1\|^2 = \frac{\beta}{\alpha^3}\|B\Delta y_1\|^2 + 2\frac{(1-\alpha)}{\alpha^3}\langle B\Delta y_1, \Delta\gamma_1\rangle + \frac{(1-\alpha)^2}{\alpha^3\beta}\|\Delta\gamma_1\|^2$$

and

$$\begin{aligned}\|\tilde{z}_1 - z_0\|_M^2 &\geq \frac{\beta}{\alpha}\|B(y_1 - y_0)\|^2 + \frac{2(1-\alpha)}{\alpha}\langle B(y_1 - y_0), \tilde{\gamma}_1 - \gamma_0\rangle + \frac{1}{\alpha\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 \\ &= \left(\frac{\beta}{\alpha} + 2\frac{(1-\alpha)\beta}{\alpha^2} + \frac{\beta}{\alpha^3}\right)\|B\Delta y_1\|^2 \\ &\quad + 2\left(\frac{(1-\alpha)}{\alpha^2} + \frac{1}{\alpha^3}\right)\langle B\Delta y_1, \Delta\gamma_1\rangle + \frac{1}{\alpha^3\beta}\|\Delta\gamma_1\|^2,\end{aligned}$$

where the last equality is due to (3.6) and (3.7). Hence, it is easy to see that

$$\|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2 \leq \frac{(\alpha-2)}{\alpha^2}\left\|\sqrt{\beta}B\Delta y_1 + \frac{1}{\sqrt{\beta}}\Delta\gamma_1\right\|^2 \leq 0.$$

Thus, it follows from (3.14) that

$$\|z^* - z_1\|_M^2 \leq \|z^* - z_0\|_M^2,$$

which, combined with (3.13), yields

$$\|\Delta y_1\|_H^2 - 2\langle B\Delta y_1, \Delta\gamma_1\rangle \leq 4\|z^* - z_0\|_M^2.$$

Therefore, the first inequality in (3.12) follows from definition of d_0 (see (3.5)) and the fact that $z^* \in T^{-1}(0)$ is arbitrary.

Let us now prove the second inequality in (3.12). First, from the optimality condition of (3.2), and (3.3), we obtain

$$B^*\gamma_j - H(y_j - y_{j-1}) \in \partial g(y_j) \qquad \forall j \geq 1.$$

For every $k \geq 2$, using the previous inclusion for $j = k-1$ and $j = k$, it follows from the monotonicity of the subdifferential of g that

$$\langle B^*(\gamma_k - \gamma_{k-1}) - H(y_k - y_{k-1}) + H(y_{k-1} - y_{k-2}), y_k - y_{k-1}\rangle \geq 0,$$

which, combined with (3.6), yields

$$\langle B\Delta y_k, \Delta\gamma_k\rangle \geq \|\Delta y_k\|_H^2 - \langle H\Delta y_{k-1}, \Delta y_k\rangle \qquad \forall k \geq 2.$$

To conclude the proof, use the first relation in (2.1) with $Q = H$. ■

Let us consider the following quantity:

$$\sigma_\alpha = \frac{1}{1+\alpha(2-\alpha)}. \tag{3.15}$$

Note that $\sigma_2 = 1$, and for any $\alpha \in (0,2)$ we have $\sigma_\alpha \in (0,1)$. The following theorem shows that Algorithm 1 is an instance of the modified HPE framework.

Theorem 3.2.3 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1 and consider $\{(\Delta y_k, \tilde{\gamma}_k)\}$ and σ_α as in (3.6) and (3.15), respectively. Define*

$$z_{k-1} = (x_{k-1}, y_{k-1}, \gamma_{k-1}) \qquad \tilde{z}_k = (x_k, y_k, \tilde{\gamma}_k), \qquad \forall\, k \geq 1, \tag{3.16}$$

and

$$\eta_0 = \frac{4(2-\alpha)\sigma_\alpha}{\alpha}d_0, \qquad \eta_k = \frac{(2-\alpha)\sigma_\alpha}{\alpha}\|\Delta y_k\|_H^2 \qquad \forall\, k \geq 1, \tag{3.17}$$

where d_0 is as in (3.5). Then, the sequence $\{(z_k, \tilde{z}_k, \eta_k)\}$ is an instance of the modified HPE framework, applied for solving (1.7), where $\sigma := \sigma_\alpha$ and M is as in (3.4).

Proof. The inclusion (2.5a) follows from the last statement in Lemma 3.2.1. Let us now show that (2.5b) holds. Using (3.6), (3.7) and (3.16), we obtain

$$\begin{aligned}\|\tilde{z}_k - z_k\|_M^2 &= \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_k\|^2 = \frac{1}{\alpha\beta}\left\|\frac{1}{\alpha}\left[(1-\alpha)\Delta\gamma_k + \beta B\Delta y_k\right]\right\|^2 \\ &= \frac{1}{\alpha^3\beta}\left[(1-\alpha)^2\|\Delta\gamma_k\|^2 + 2(1-\alpha)\beta\langle B\Delta y_k, \Delta\gamma_k\rangle + \beta^2\|B\Delta y_k\|^2\right]. \end{aligned} \tag{3.18}$$

Also, (3.6) and (3.16) imply that

$$\begin{aligned}\|\tilde{z}_k - z_{k-1}\|_M^2 = \|\Delta x_k\|_G^2 + \|\Delta y_k\|_H^2 + \frac{\beta}{\alpha}\|B\Delta y_k\|^2 \\ + 2\frac{(1-\alpha)}{\alpha}\langle B\Delta y_k, \tilde{\gamma}_k - \gamma_{k-1}\rangle + \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2. \end{aligned} \tag{3.19}$$

It follows from (3.7) that

$$\frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{1}{\alpha^3\beta}\left[\|\Delta\gamma_k\|^2 + 2\beta\langle B\Delta y_k, \Delta\gamma_k\rangle + \beta^2\|B\Delta y_k\|^2\right],$$

$$2\frac{(1-\alpha)}{\alpha}\langle B\Delta y_k, \tilde{\gamma}_k - \gamma_{k-1}\rangle = 2\frac{(1-\alpha)}{\alpha^2}\left[\langle B\Delta y_k, \Delta\gamma_k\rangle + \beta\|B\Delta y_k\|^2\right]$$

which, combined with (3.19), yields

$$\begin{aligned}\|\tilde{z}_k - z_{k-1}\|_M^2 = \|\Delta x_k\|_G^2 + \|\Delta y_k\|_H^2 + \left(\frac{\beta}{\alpha} + 2\frac{(1-\alpha)\beta}{\alpha^2} + \frac{\beta}{\alpha^3}\right)\|B\Delta y_k\|^2 \\ + 2\left(\frac{(1-\alpha)}{\alpha^2} + \frac{1}{\alpha^3}\right)\langle B\Delta y_k, \Delta\gamma_k\rangle + \frac{1}{\alpha^3\beta}\|\Delta\gamma_k\|^2. \end{aligned} \tag{3.20}$$

Therefore, combining (3.18) and (3.20), it is easy to verify that

$$\begin{aligned}&\sigma_\alpha\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 \\ &\quad = \sigma_\alpha\|\Delta x_k\|_G^2 + \sigma_\alpha\|\Delta y_k\|_H^2 + 2\frac{(2-\alpha)\sigma_\alpha}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle + \frac{(2-\alpha)^2\sigma_\alpha}{\alpha\beta}\|\Delta\gamma_k\|^2 \\ &\quad \geq 2\frac{(2-\alpha)\sigma_\alpha}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle \geq \eta_k - \eta_{k-1} \qquad \forall\, k \geq 1,\end{aligned}$$

where σ_α is as in (3.15), and the last inequality is due to (3.12) and (3.17). Therefore, (2.5b) holds, and then we conclude that the sequence $\{(z_k, \tilde{z}_k, \eta_k)\}$ is an instance of the modified HPE framework. ∎

3.2.2 Iteration-complexity bounds for the PG-ADMM

In this subsection, we establish pointwise and ergodic iteration-complexity bounds for Algorithm 1. We start by presenting a pointwise bound under the assumption that the relaxation parameter α belongs to $(0,2)$. For this, we first introduce a result which shows that the sequence $\{\|z_k - z_{k-1}\|_M\}$, with $\{z_k\}$ given in (3.16), is monotonically nonincreasing. Then, we consider an auxiliary result which is used to show that the sequence $\{\tilde{\rho}_k\}$, as defined in Theorem 2.2.7 with $\{z_k\}$ and $\{\tilde{z}_k\}$ as in (3.16), is bounded even in the extreme case in which $\alpha = 2$. This latter result is then used to present the ergodic bounds of Algorithm 1 for any $\alpha \in (0,2]$.

Lemma 3.2.4 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1 and consider the sequence $\{z_k\}$ as in (3.16). Then, for every $k \geq 2$,*

$$\|z_k - z_{k-1}\|_M \leq \|z_{k-1} - z_{k-2}\|_M \,,$$

where M is as in (3.4).

Proof. First, note that for any $z \in \mathcal{Z}$, we have

$$\begin{aligned}\|z_{k-1} - z_{k-2}\|_M^2 - \|z_k - z_{k-1}\|_M^2 &= \|z_{k-1} - z + z - z_{k-2}\|_M^2 - \|z_k - z + z - z_{k-1}\|_M^2 \\ &= \|z_{k-2} - z\|_M^2 - \|z_k - z\|_M^2 + 2\left\langle M(z_{k-2} - z_k), z - z_{k-1}\right\rangle.\end{aligned}$$

Letting $z := z_{k-1} + \tilde{z}_{k-1} - \tilde{z}_k$ in the above relations, where $\{\tilde{z}_k\}$ is given in (3.16), it follows that

$$\begin{aligned}&\|z_{k-1} - z_{k-2}\|_M^2 - \|z_k - z_{k-1}\|_M^2 \\ &= \|z_{k-2} - z_{k-1} - \tilde{z}_{k-1} + \tilde{z}_k\|_M^2 - \|z_k - z_{k-1} - \tilde{z}_{k-1} + \tilde{z}_k\|_M^2 + 2\left\langle M(z_{k-2} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle \\ &\geq \|z_{k-2} - z_{k-1} - \tilde{z}_{k-1} + \tilde{z}_k\|_M^2 - \|z_{k-1} - z_k + \tilde{z}_{k-1} - \tilde{z}_k\|_M^2 + 4\left\langle M(z_{k-1} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle \\ &= \|\tilde{z}_k - z_{k-1} - (\tilde{z}_{k-1} - z_{k-2})\|_M^2 - \|z_{k-1} - z_k - (\tilde{z}_{k-1} - \tilde{z}_k)\|_M^2 \,, \end{aligned} \tag{3.21}$$

where the inequality above is due to the monotonicity of the operator T (given in (2.11)), the last part of Lemma 3.2.1 and the following inequality

$$\begin{aligned}\left\langle M(z_{k-2} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle &= \left\langle M(z_{k-2} - z_{k-1}) - M(z_{k-1} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle \\ &\quad + 2\left\langle M(z_{k-1} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle \\ &\geq 2\left\langle M(z_{k-1} - z_k), \tilde{z}_{k-1} - \tilde{z}_k\right\rangle.\end{aligned}$$

Using (3.6), (3.7), and the definitions of z_k and $\tilde{z}_k$ in (3.16), it is easy to see that

$$\begin{aligned}\tilde{z}_k - z_{k-1} - &(\tilde{z}_{k-1} - z_{k-2}) \\ &= (\Delta x_k - \Delta x_{k-1}, \Delta y_k - \Delta y_{k-1}, \tilde{\gamma}_k - \gamma_{k-1} - (\tilde{\gamma}_{k-1} - \gamma_{k-2})) \\ &= \left(\Delta x_k - \Delta x_{k-1}, \Delta y_k - \Delta y_{k-1}, \frac{1}{\alpha}\Big(\Delta\gamma_k - \Delta\gamma_{k-1} + \beta B(\Delta y_k - \Delta y_{k-1})\Big)\right)\end{aligned}$$

and

$$z_{k-1} - z_k - (\tilde{z}_{k-1} - \tilde{z}_k) = \tilde{z}_k - z_k - (\tilde{z}_{k-1} - z_{k-1}) = (0, 0, \tilde{\gamma}_k - \gamma_k - (\tilde{\gamma}_{k-1} - \gamma_{k-1}))$$
$$= \left(0, 0, \frac{1}{\alpha}\Big((1-\alpha)(\Delta\gamma_k - \Delta\gamma_{k-1}) + \beta B(\Delta y_k - \Delta y_{k-1})\Big)\right).$$

Combining the last two relations with (3.21) and the definition of M in (3.4), we obtain

$$\begin{aligned}
\|z_{k-1} - z_{k-2}\|_M^2 - \|z_k - z_{k-1}\|_M^2 \geq\ & \frac{\beta}{\alpha}\|B(\Delta y_k - \Delta y_{k-1})\|^2 \\
& + \frac{2(1-\alpha)}{\alpha^2}\langle B(\Delta y_k - \Delta y_{k-1}), \Delta\gamma_k - \Delta\gamma_{k-1} + \beta B(\Delta y_k - \Delta y_{k-1})\rangle \\
& + \frac{1}{\alpha^3\beta}\|\Delta\gamma_k - \Delta\gamma_{k-1} + \beta B(\Delta y_k - \Delta y_{k-1})\|^2 \\
& - \frac{1}{\alpha^3\beta}\|(1-\alpha)(\Delta\gamma_k - \Delta\gamma_{k-1}) + \beta B(\Delta y_k - \Delta y_{k-1})\|^2.
\end{aligned}$$

By performing some simple algebraic manipulations, the above expression becomes

$$\begin{aligned}
\|z_{k-1} - z_{k-2}\|_M^2 - \|z_k - z_{k-1}\|_M^2 \geq\ & \left(\frac{\beta}{\alpha} + \frac{2(1-\alpha)\beta}{\alpha^2}\right)\|B(\Delta y_k - \Delta y_{k-1})\|^2 \\
& + \left(\frac{2(1-\alpha)}{\alpha^2} + \frac{2}{\alpha^2}\right)\langle B(\Delta y_k - \Delta y_{k-1}), \Delta\gamma_k - \Delta\gamma_{k-1}\rangle + \frac{2\alpha - \alpha^2}{\alpha^3\beta}\|\Delta\gamma_k - \Delta\gamma_{k-1}\|^2 \\
=\ & \frac{(2-\alpha)\beta}{\alpha^2}\left\|B(\Delta y_k - \Delta y_{k-1}) + \frac{1}{\beta}(\Delta\gamma_k - \Delta\gamma_{k-1})\right\|^2 \geq 0,
\end{aligned}$$

where the last inequality follows from the fact that $\alpha \in (0, 2]$. ■

Theorem 3.2.5 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1 with $\alpha \in (0, 2)$ and consider the sequence $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k, \tilde{\gamma}_k)\}$ as in (3.6). Then, for every $k \geq 1$,*

$$0 \in M \begin{pmatrix} \Delta x_k \\ \Delta y_k \\ \Delta\gamma_k \end{pmatrix} + \begin{pmatrix} \partial f(x_k) - A^*\tilde{\gamma}_k \\ \partial g(y_k) - B^*\tilde{\gamma}_k \\ Ax_k + By_k - b \end{pmatrix} \tag{3.22}$$

and

$$\|(\Delta x_k, \Delta y_k, \Delta\gamma_k)\|_M \leq \frac{1}{\sqrt{k}}\sqrt{\frac{2[\alpha(1+\sigma_\alpha) + 8(2-\alpha)\sigma_\alpha]d_0}{\alpha(1-\sigma_\alpha)}},$$

where M, d_0, and σ_α are as (3.4), (3.5) and (3.15), respectively.

Proof. Since $\sigma_\alpha \in (0,1)$ for any $\alpha \in (0,2)$ (see (3.15)), we obtain by combining Theorem 3.2.3, Lemma 3.2.4, and Corollary 2.2.5 that (3.22) holds and

$$\|(\Delta x_k, \Delta y_k, \Delta\gamma_k)\|_M \le \frac{1}{\sqrt{k}}\sqrt{\frac{2(1+\sigma_\alpha)d_0 + 4\eta_0}{1-\sigma_\alpha}}.$$

Hence, to conclude the proof use the definition of η_0 given in (3.17). ■

Remark 3.2.6 For a given tolerance $\bar\rho > 0$, Theorem 3.2.5 implies that in at most $\mathcal{O}(1/\bar\rho^2)$ iterations, Algorithm 1 obtains an approximate solution $(\hat x, \hat y, \hat\gamma)$ and a residual $\hat u$ of (2.11) satisfying

$$M\hat u \in T(\hat x, \hat y, \hat\gamma), \quad \|\hat u\|_M \le \bar\rho, \tag{3.23}$$

where T is as in (2.11). It is worth pointing out that although M may not be invertible, the above complexity result makes sense due to the fact that $\|\hat u\|_M = 0$ yields $M\hat u = 0$, which in turn implies that the triple $(\hat x, \hat y, \hat\gamma)$ is a solution of (1.7). Let λ_M be the largest eigenvalue of M and $(v_{\hat x}, v_{\hat y}, v_{\hat\gamma}) := M\hat u$. For a given tolerance $\rho > 0$, (1.8) follows from (3.23) with $\bar\rho = \rho/\sqrt{\lambda_M}$ and the fact that $\|M(\cdot)\| \le \sqrt{\lambda_M}\|\cdot\|_M$. Hence, Algorithm 1 provides a ρ-approximate solution of (1.7) in at most $\mathcal{O}(1/\rho^2)$ iterations.

We next present an auxiliary result which is essential to obtain ergodic iteration-complexity bounds for Algorithm 1.

Lemma 3.2.7 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1 and consider the sequence $\{(\Delta x_k, \Delta y_k, \Delta\gamma_k, \tilde\gamma_k)\}$ as in (3.6). Then, $\{\tilde\rho_k\}$ given in (2.8) with M and $\{(z_k, \tilde z_k)\}$ as in (3.4) and (3.16), respectively, satisfies*

$$\tilde\rho_k \le \frac{4(1+2\alpha)[\alpha + 4(2-\alpha)\sigma_\alpha]d_0}{\alpha^3} \qquad \forall\, k \ge 1,$$

where d_0 is as in (3.5).

Proof. The same argument used to prove (3.19) and (3.20) yields, for every $k \ge 1$,

$$\|\tilde z_k - z_{k-1}\|_M^2 = \|\Delta x_k\|_G^2 + \|\Delta y_k\|_H^2 + \xi_k, \tag{3.24}$$

where

$$\xi_k := \frac{\beta}{\alpha^3}\|B\Delta y_k\|^2 + \frac{2(1-\alpha)}{\alpha^3}\langle B\Delta y_k, \Delta\gamma_k\rangle + \frac{1}{\alpha^3\beta}\|\Delta\gamma_k\|^2 + \frac{(2-\alpha)}{\alpha}\left[\frac{\beta}{\alpha}\|B\Delta y_k\|^2 + \frac{2}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle\right].$$

Using the definitions of M and z_k given in (3.4) and (3.16), respectively, it follow that

$$
\begin{aligned}
\xi_k &\le \frac{1}{\alpha^2}\|z_k - z_{k-1}\|_M^2 + \frac{(2-\alpha)}{\alpha}\left[\frac{\beta}{\alpha}\|B\Delta y_k\|^2 + \frac{2}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle\right] \\
&= \frac{1}{\alpha^2}\|z_k - z_{k-1}\|_M^2 + \frac{(2-\alpha)}{\alpha}\left[\frac{\beta}{\alpha}\|B\Delta y_k\|^2 + \frac{2(1-\alpha)}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle\right] \\
&+ \frac{2(2-\alpha)}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle \\
&\le \frac{1}{\alpha^2}\|z_k - z_{k-1}\|_M^2 + \frac{(2-\alpha)}{\alpha}\|z_k - z_{k-1}\|_M^2 + \frac{2(1-\alpha)}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle + \frac{2}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle \\
&\le \frac{1+2\alpha-\alpha^2}{\alpha^2}\|z_k - z_{k-1}\|_M^2 + \frac{2(1-\alpha)}{\alpha}\langle B\Delta y_k, \Delta\gamma_k\rangle + \frac{\beta}{\alpha}\|B\Delta y_k\|^2 + \frac{1}{\alpha\beta}\|\Delta\gamma_k\|^2, \quad (3.25)
\end{aligned}
$$

where in the last two inequalities we used the fact that $\alpha \in (0, 2]$ and the first property in (2.1) with $Q = I$, respectively. Combining (3.24), (3.25) and definitions of M and z_k, we obtain, for every $k \ge 1$,

$$
\|\tilde{z}_k - z_{k-1}\|_M^2 \le \frac{1+2\alpha-\alpha^2}{\alpha^2}\|z_k - z_{k-1}\|_M^2 + \|z_k - z_{k-1}\|_M^2 = \frac{1+2\alpha}{\alpha^2}\|z_k - z_{k-1}\|_M^2.
$$

Now, letting $z^* := (x^*, y^*, \gamma^*)$ be an arbitrary solution of (1.7), we obtain from the last inequality and the second relation in (2.1) with $Q = M$ that

$$
\|\tilde{z}_k - z_{k-1}\|_M^2 \le \frac{2(1+2\alpha)}{\alpha^2}\left[\|z^* - z_k\|_M^2 + \|z^* - z_{k-1}\|_M^2\right] \qquad \forall k \ge 1. \tag{3.26}
$$

Since Algorithm 1 is an instance of the modified HPE framework (see Theorem 3.2.3), it follows from (3.26) and Lemma 2.2.3(b) that

$$
\|\tilde{z}_k - z_{k-1}\|_M^2 \le \frac{4(1+2\alpha)}{\alpha^2}\left[\|z^* - z_0\|_M^2 + \eta_0\right] \qquad \forall k \ge 1.
$$

Since z^* is an arbitrary solution of (1.7), the result follows from the definition of $\tilde{\rho}_k$, d_0, and η_0 given in (2.8), (3.5) and (3.17), respectively. ■

Next result presents iteration-complexity bounds for the ergodic sequence associated to Algorithm 1.

Theorem 3.2.8 *Let $\{(x_k, y_k, \gamma_k)\}$ be the sequence generated by Algorithm 1 and consider*

$\{(\Delta x_k, \Delta y_k, \Delta\gamma_k, \tilde{\gamma}_k)\}$ *as in* (3.6). *Define the ergodic sequences as*

$$(x_k^a, y_k^a, \gamma_k^a, \tilde{\gamma}_k^a) = \frac{1}{k}\sum_{i=1}^{k}(x_i, y_i, \gamma_i, \tilde{\gamma}_i), \qquad (r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a) = \frac{1}{k}\sum_{i=1}^{k}(\Delta x_i, \Delta y_i, \Delta\gamma_i), \tag{3.27}$$

$$\varepsilon_{k,x}^a = \frac{1}{k}\sum_{i=1}^{k}\langle G\Delta x_i - A^*\tilde{\gamma}_i, x_k^a - x_i\rangle, \tag{3.28}$$

$$\varepsilon_{k,y}^a = \frac{1}{k}\sum_{i=1}^{k}\left\langle \left(H + \frac{\beta}{\alpha}B^*B\right)\Delta y_i + \frac{(1-\alpha)}{\alpha}B^*\Delta\gamma_i - B^*\tilde{\gamma}_i, y_k^a - y_i\right\rangle. \tag{3.29}$$

Then, for every $k \geq 1$*, there hold* $\varepsilon_{k,x}^a \geq 0$*,* $\varepsilon_{k,y}^a \geq 0$*, and*

$$0 \in M\begin{pmatrix} r_{k,x}^a \\ r_{k,y}^a \\ r_{k,\gamma}^a \end{pmatrix} + \begin{pmatrix} \partial_{\varepsilon_{k,x}^a} f(x_k^a) - A^*\tilde{\gamma}_k^a \\ \partial_{\varepsilon_{k,y}^a} g(y_k^a) - B^*\tilde{\gamma}_k^a \\ Ax_k^a + By_k^a - b \end{pmatrix}, \tag{3.30}$$

$$\|(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a)\|_M \leq \frac{2\sqrt{c_\alpha d_0}}{k}, \quad \varepsilon_{k,x}^a + \varepsilon_{k,y}^a \leq \frac{\tilde{c}_\alpha d_0}{k}, \tag{3.31}$$

where

$$c_\alpha := \frac{\alpha + 4(2-\alpha)\sigma_\alpha}{\alpha}, \quad \tilde{c}_\alpha := \frac{3[3\alpha^2 + 4(1+2\alpha)\sigma_\alpha][\alpha + 4(2-\alpha)\sigma_\alpha]}{2\alpha^3}, \tag{3.32}$$

and M*,* d_0*, and* σ_α *are as in* (3.4), (3.5), *and* (3.15), *respectively.*

Proof. Note that the inclusions (3.8)-(3.9) are equivalent to

$$-(G\Delta x_k - A^*\tilde{\gamma}_k) \in \partial f(x_k), \quad -\left(\left(H + \frac{\beta}{\alpha}B^*B\right)\Delta y_k + \frac{(1-\alpha)}{\alpha}B^*\Delta\gamma_k - B^*\tilde{\gamma}_k\right) \in \partial g(y_k).$$

Hence, by combining Proposition 2.1.1, (3.27) and definition of M, we obtain $\varepsilon_{k,x}^a \geq 0$, $\varepsilon_{k,y}^a \geq 0$, and the first two inclusions of (3.30). The third inclusion of (3.30) holds trivially from (3.10), (3.27) and definition of M. Now, it follows from Theorem 3.2.3 that Algorithm 1 is an instance of the modified HPE where $\{(z_k, \tilde{z}_k)\}$ is given by (3.16). Moreover, it is easy to see that the quantities r_k^a and ε_k^a given in (2.7) satisfy

$$r_k^a = (r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a), \quad \varepsilon_k^a = \frac{1}{k}\sum_{i=1}^{k}\left[\left\langle M\begin{pmatrix} \Delta x_i \\ \Delta y_i \\ \Delta\gamma_i \end{pmatrix}, (x_k^a - x_i, y_k^a - y_i, \tilde{\gamma}_k^a - \tilde{\gamma}_i)\right\rangle\right]. \tag{3.33}$$

Hence, from Theorems 2.2.7 and definition of η_0 in (3.17), we have

$$\|r_k^a\|_M \leq \frac{2\sqrt{(\alpha+4(2-\alpha)\sigma_\alpha)d_0}}{k\sqrt{\alpha}}, \quad \varepsilon_k^a \leq \frac{3[3\alpha^2+4(1+2\alpha)\sigma_\alpha][\alpha+4(2-\alpha)\sigma_\alpha]d_0}{2\alpha^3 k}, \quad (3.34)$$

where in the last inequality we also used Lemma 3.2.7. Now, we claim that $\varepsilon_k^a = \varepsilon_{k,x}^a + \varepsilon_{k,y}^a$. Using this claim, (3.31) follows immediately from (3.32) and (3.34). Hence, to conclude the proof of the theorem, it just remains to prove the above claim. To this end, note that (3.28) and (3.29) yield

$$\begin{aligned}\varepsilon_{k,x}^a + \varepsilon_{k,y}^a = \frac{1}{k}\sum_{i=1}^k &\left[\langle G\Delta x_i, x_k^a - x_i\rangle + \left\langle \left(H + \frac{\beta}{\alpha}B^*B\right)\Delta y_i + \frac{(1-\alpha)}{\alpha}B^*\Delta\gamma_i, y_k^a - y_i\right\rangle\right] \\ &+ \frac{1}{k}\sum_{i=1}^k \langle A\left(x_k^a - x_i\right) + B\left(y_k^a - y_i\right), -\tilde{\gamma}_i\rangle] \,. \end{aligned} \quad (3.35)$$

On the other hand, from (3.27), we obtain

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^k \langle A(x_k^a - x_i) + B(y_k^a - y_i), -\tilde{\gamma}_i\rangle &= \frac{1}{k}\sum_{i=1}^k \langle Ax_k^a + By_k^a - b - (Ax_i + By_i - b), \tilde{\gamma}_k^a - \tilde{\gamma}_i\rangle \\ &= \frac{1}{k}\sum_{i=1}^k \langle -(Ax_i + By_i - b), \tilde{\gamma}_k^a - \tilde{\gamma}_i\rangle \\ &= \frac{1}{k}\sum_{i=1}^k \left\langle \frac{(1-\alpha)}{\alpha}B\Delta y_i + \frac{1}{\alpha\beta}\Delta\gamma_i, \tilde{\gamma}_k^a - \tilde{\gamma}_i\right\rangle \end{aligned}$$

where the last equality is due to (3.10). Hence, the claim follows by combining (3.35), and the definitions of M and ε_k^a in (3.4) and (3.33), respectively. ■

Remark 3.2.9 Using the fact that $\|M(\cdot)\| \leq \sqrt{\lambda_M}\|\cdot\|_M$, where λ_M denotes the largest eigenvalue of M, it follows from the first inequality in (3.31) that

$$\|M(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a)^*\| \leq \frac{2\sqrt{\lambda_M c_\alpha d_0}}{k}.$$

Therefore, for a given tolerance $\rho > 0$, Theorem 3.2.8 implies that in at most $\mathcal{O}(1/\rho)$ iterations of Algorithm 1, we obtain an approximate solution $(\bar{x}, \bar{y}, \bar{\gamma})$ and a residual $(v_{\bar{x}}, v_{\bar{y}}, v_{\bar{\gamma}})$ of (1.7) satisfying

$$v_{\bar{x}} \in \partial_{\varepsilon_{\bar{x}}} f(\bar{x}) - A^*\bar{\gamma}, \qquad v_{\bar{y}} \in \partial_{\varepsilon_{\bar{y}}} g(\bar{y}) - B^*\bar{\gamma}, \qquad v_{\bar{\gamma}} = A\bar{x} + B\bar{y} - b,$$
$$\max\{\|v_{\bar{x}}\|, \|v_{\bar{y}}\|, \|v_{\bar{\gamma}}\|, \varepsilon_{\bar{x}}, \varepsilon_{\bar{y}}\} \leq \rho.$$

Hence, Algorithm 1 provides a relaxed ρ-approximate solution of (1.7) in at most $\mathcal{O}(1/\rho)$ iterations.

Chapter 4

An inexact PG-ADMM and its iteration-complexity analysis

In this chapter, we propose and analyze an inexact proximal generalized ADMM for approximately solving (1.1). This chapter is associated to [1] and is organized as follows. In Section 4.1, we introduce the proposed scheme, whereas Section 4.2 contains its iteration-complexity analysis. Section 4.2 is divided into two subsections. The first one shows that the proposed method falls within the setting of the modified HPE framework of Section 2.2, whereas the last subsection establishes its iteration-complexity bounds to obtain approximate solution of (1.1).

4.1 Inexact PG-ADMM

In this section, we formally state the inexact proximal generalized ADMM for computing approximate solutions of (1.1).

Algorithm 2: Inexact proximal generalized ADMM

0. Let an initial point $(x_0, y_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$, a penalty parameter $\beta > 0$, two error tolerance parameters $\tau_1, \tau_2 \in [0, 1)$, a relaxation factor $\alpha \in (0, 2 - \tau_1)$, and a self-adjoint positive semidefinite linear operator $H \colon \mathcal{Y} \to \mathcal{Y}$ be given, and set $k = 1$.

1. Compute $(\tilde{x}_k, v_k) \in \mathcal{X} \times \mathcal{X}$ such that

$$v_k \in \partial f(\tilde{x}_k) - A^*\tilde{\gamma}_k, \quad \|\tilde{x}_k - x_{k-1} + \beta v_k\|^2 \leq \tau_1 \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \tau_2 \|\tilde{x}_k - x_{k-1}\|^2, \tag{4.1}$$

where

$$\tilde{\gamma}_k = \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b). \tag{4.2}$$

2. Compute an optimal solution $y_k \in \mathcal{Y}$ of the subproblem

$$\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|\alpha(A\tilde{x}_k + By_{k-1} - b) + B(y - y_{k-1})\|^2 \right. \\ \left. + \frac{1}{2}\|y - y_{k-1}\|_H^2 \right\}. \tag{4.3}$$

3. Set

$$x_k = x_{k-1} - \beta v_k, \quad \gamma_k = \gamma_{k-1} - \beta \left[\alpha(A\tilde{x}_k + By_{k-1} - b) + B(y_k - y_{k-1})\right], \tag{4.4}$$

and $k \leftarrow k + 1$, and go to step 1.

Remark 4.1.1 Some comments about Algorithm 2 are in order.

(a) Algorithm 2 is an inexact version of Algorithm 1. It is well-suitable in applications in which subproblem (3.2) is easy to solve whereas (3.1) is not, being necessary therefore to use iterative methods to approximately solve it. The proposed scheme allows inexact solutions of the following inclusion (derived from the first-order optimality condition for (3.1) with $G = \frac{1}{\beta}I$)

$$0 \in \partial f(x) - A^*(\gamma_{k-1} - \beta(Ax + By_{k-1} - b)) + \frac{1}{\beta}(x - x_{k-1}), \tag{4.5}$$

such that a relative error condition is satisfied. The error condition used here is similar to the one studied in [71, 72] in the context of a hybrid proximal extragradient method. It is shown that the new inexact method Algorithm 2 possesses iteration-complexity bounds similar to its exact version Algorithm 1.

(b) If $\tau_1 = \tau_2 = 0$, then the inequality in (4.1), combined with the first relation in (4.4), implies that $\tilde{x}_k = x_k$ and $v_k = (x_{k-1} - x_k)/\beta$. Hence, in view of the definition of $\tilde{\gamma}_k$ in (4.2) and the inclusion in (4.1), we conclude that x_k is a solution of (4.5). Therefore, Algorithm 2 can be seen as a variant of Algorithm 1 in which its first subproblem is approximately solved using a relative error condition. Now, if x_k is a solution of the inclusion in (4.5), then the pair $(\tilde{x}_k, v_k) := (x_k, (x_{k-1} - x_k)/\beta)$ trivially satisfies (4.1).

(c) It is assumed that (4.3) can be easily solved. On the one hand, if the matrix B in (1.1) is not the identity, then subproblem (4.3) with the usual choice $H := \xi I - \beta B^*B$ with $\xi \geq \beta\|B^*B\|$ becomes a prox-subproblem

$$y_k = \arg\min_{y\in\mathcal{Y}} \left\{ g(y) + \frac{\xi}{2}\|y - \hat{y}\|^2 \right\} \tag{4.6}$$

for some $\hat{y} \in \mathcal{Y}$. In many ADMM applications, g is well-structured (e.g., the ℓ_1-norm) and hence the latter problem is easy to solve or even has a closed-form solution. On the other hand, if $B = I$ in (1.1), then $H = 0$ seems to be a natural choice.

Some numerical experiments will be presented in Chapter 6 in order to illustrate the performance of Algorithm 2. In particular, it is verified that the use of the relaxation parameter $\alpha > 1$, specially $\alpha \approx 1.9$, improves considerably its numerical behavior.

4.2 Iteration-complexity of the inexact PG-ADMM

This section analyzes pointwise and ergodic iteration-complexity bounds for Algorithm 2 to obtain an approximate solution of (1.1). It is divided into two subsections. In the first subsection, we show that Algorithm 2 can be regarded as an instance of the modified HPE framework of Section 2.2. The last subsection establishes the iteration-complexity bounds for Algorithm 2.

In order to show that Algorithm 2 falls within the setting of the modified HPE framework, we need to define the elements required by Section 2.2. We consider a linear operator M defined as follows

$$M = \begin{bmatrix} \frac{1}{\beta}I & 0 & 0 \\ 0 & (H + \frac{\beta}{\alpha}B^*B) & \frac{1-\alpha}{\alpha}B^* \\ 0 & \frac{1-\alpha}{\alpha}B & \frac{1}{\alpha\beta}I \end{bmatrix}. \tag{4.7}$$

It can be easily verified that, for every $\beta > 0$ and $\alpha \in (0, 2)$, M is self-adjoint and positive semidefinite. Let us now introduce the constant d_0 given by

$$d_0 = \inf \left\{ \|(x - x_0, y - y_0, \gamma - \gamma_0)\|_M^2 : (x, y, \gamma) \in \Omega^* \right\}, \tag{4.8}$$

where Ω^* is given in Assumption 2.3.1. Note that, if M is positive definite, then d_0 measures the distance in the norm $\|\cdot\|_M$ of the initial point (x_0, y_0, γ_0) to the solution set Ω^*.

Let $\{(x_k, y_k, \gamma_k, \tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 2 and consider the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ defined by

$$z_{k-1} = (x_{k-1}, y_{k-1}, \gamma_{k-1}), \qquad \tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k), \qquad \forall\, k \geq 1. \tag{4.9}$$

It will be shown that, for any given $\rho > 0$, there exists an index k such that $\tilde{z}_k$ is a ρ-approximate solution of (1.7) with residue $r_k := M(z_{k-1} - z_k)$ (see Definition 2.3.2). To this end, we present two technical results. Note first that, from the definitions of $\tilde{\gamma}_k$ and γ_k given in (4.2) and (4.4), respectively, it follows that

$$\tilde{\gamma}_k - \gamma_{k-1} = \frac{\beta}{\alpha} B(y_k - y_{k-1}) + \frac{1}{\alpha}(\gamma_k - \gamma_{k-1}), \quad \forall\, k \geq 1, \tag{4.10}$$

which, in turn, implies that

$$\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{\beta}{\alpha^2} \|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_S^2, \quad \text{where} \quad S = \begin{bmatrix} \beta B^* B & B^* \\ B & \frac{1}{\beta} I \end{bmatrix}. \tag{4.11}$$

For simplicity, we also consider the following linear operators

$$N = \begin{bmatrix} [1 + \alpha(2 - \alpha)]\, \beta B^* B & (1 + \alpha - \alpha^2) B^* \\ (1 + \alpha - \alpha^2) B & \frac{1}{\beta} I \end{bmatrix}, \quad P = \begin{bmatrix} \beta B^* B & (1 - \alpha) B^* \\ (1 - \alpha) B & \frac{(1-\alpha)^2}{\beta} I \end{bmatrix}. \tag{4.12}$$

It is easy to verify that S, N and P are self-adjoint positive semidefinite linear operators for every $\beta > 0$ and $\alpha \in (0, 2)$.

4.2.1 Inexact PG-ADMM in the setting of the modified HPE framework

This subsection is devoted to show that Algorithm 2 can be regarded as an instance of modified HPE framework. In order to show this, we first need to establish some technical lemmas.

Lemma 4.2.1 *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in (4.9). Then, for every $k \geq 1$, the following hold:*

$$\|\tilde{z}_k - z_{k-1}\|_M^2 \geq \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_N^2 \tag{4.13}$$

and

$$\|\tilde{z}_k - z_k\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_k\|^2 + \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_P^2, \tag{4.14}$$

where the matrices M, N and P are as in (4.7) and (4.12).

Proof. Using the fact that $\tilde{z}_k - z_{k-1} = (\tilde{x}_k - x_{k-1}, y_k - y_{k-1}, \tilde{\gamma}_k - \gamma_{k-1})$ and the definition of M in (4.7), we obtain

$$\|\tilde{z}_k - z_{k-1}\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|_H^2 + \frac{\beta}{\alpha}\|B(y_k - y_{k-1})\|^2 \\ + \frac{2(1-\alpha)}{\alpha}\langle B(y_k - y_{k-1}), \tilde{\gamma}_k - \gamma_{k-1}\rangle + \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2.$$

On the other hand, equality (4.10) implies that

$$\langle B(y_k - y_{k-1}), \tilde{\gamma}_k - \gamma_{k-1}\rangle = \frac{\beta}{\alpha}\|B(y_k - y_{k-1})\|^2 + \frac{1}{\alpha}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle,$$

and

$$\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{\beta^2}{\alpha^2}\|B(y_k - y_{k-1})\|^2 + \frac{2\beta}{\alpha^2}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{1}{\alpha^2}\|\gamma_k - \gamma_{k-1}\|^2.$$

Combining the last three equalities, we find

$$\|\tilde{z}_k - z_{k-1}\|_M^2 \geq \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \left(\frac{1}{\alpha} + \frac{2(1-\alpha)}{\alpha^2} + \frac{1}{\alpha^3}\right)\beta\|B(y_k - y_{k-1})\|^2 \\ + \left(\frac{2(1-\alpha)}{\alpha^2} + \frac{2}{\alpha^3}\right)\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{1}{\alpha^3\beta}\|\gamma_k - \gamma_{k-1}\|^2.$$

Thus, (4.13) follows from the last equality and the definition of N in (4.12).

Let us now prove (4.14). Using $\tilde{z}_k - z_k = (\tilde{x}_k - x_k, 0, \tilde{\gamma}_k - \gamma_k)$ (see (4.9)) and the definition of M in (4.7), we have

$$\|\tilde{z}_k - z_k\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_k\|^2 + \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_k\|^2.$$

It follows from (4.10) and some algebraic manipulations that

$$\|\tilde{\gamma}_k - \gamma_k\|^2 = \frac{\beta^2}{\alpha^2}\|B(y_k - y_{k-1})\|^2 + \frac{2(1-\alpha)\beta}{\alpha^2}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle \\ + \frac{(1-\alpha)^2}{\alpha^2}\|\gamma_k - \gamma_{k-1}\|^2.$$

Therefore, the desired equality now follows by combining the last two equalities and the definition of P in (4.12). ■

Lemma 4.2.2 *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in (4.9). Then, for every $k \geq 1$,*

$$M\,(z_{k-1} - z_k) \in T(\tilde{z}_k),$$

where T and M are as in (2.11) and (4.7), respectively.

Proof. This result follows directly from Lemma 3.2.1 with $G = \frac{1}{\beta}I$ and $\tilde{z}_k$ replaced by $(\tilde{x}_k, y_k, \tilde{\gamma}_k)$. ■

The proof of the next lemma is similar to the one of Lemma 3.2.2. We present it for the sake of completeness.

Lemma 4.2.3 *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 2. Then, the following hold:*

(a) $2\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \geq \|y_1 - y_0\|_H^2 - 4d_0$, *where d_0 is as in (4.8);*

(b) $2\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle \geq \|y_k - y_{k-1}\|_H^2 - \|y_{k-1} - y_{k-2}\|_H^2$, *for every $k \geq 2$.*

Proof. (a) Consider z_0, z_1 and $\tilde{z}_1$ as in (4.9), and let an arbitrary $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$ (see Assumpiton 2.3.1). Note that, in view of the definition of d_0 in (4.8), in order to establish (a), it is sufficient to prove that

$$\Theta := \|y_1 - y_0\|_H^2 - 2\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \leq 4\|z^* - z_0\|_M^2, \tag{4.15}$$

where M is as in (4.7). Let us then show (4.15). From the definitions of M and $\{z_k\}$, we have

$$\begin{aligned}\|z_1 - z_0\|_M^2 &= \frac{1}{\beta}\|x_1 - x_0\|^2 + \|y_1 - y_0\|_H^2 + \frac{\beta}{\alpha}\|B(y_1 - y_0)\|^2 \\ &\quad + \frac{2(1-\alpha)}{\alpha}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle + \frac{1}{\alpha\beta}\|\gamma_1 - \gamma_0\|^2 \\ &= \frac{1}{\beta}\|x_1 - x_0\|^2 + \Theta + \left\|\frac{\sqrt{\beta}}{\sqrt{\alpha}}B(y_1 - y_0) + \frac{1}{\sqrt{\alpha\beta}}(\gamma_1 - \gamma_0)\right\|^2.\end{aligned}$$

Hence, we obtain

$$\Theta \leq \|z_1 - z_0\|_M^2 \leq 2\left(\|z^* - z_1\|_M^2 + \|z^* - z_0\|_M^2\right), \tag{4.16}$$

where the last inequality is due to the second property in (2.1). We will now prove that

$$\|z^* - z_1\|_M^2 \leq \|z^* - z_0\|_M^2. \tag{4.17}$$

From Lemma 4.2.2, we have $M(z_0 - z_1) \in T(\tilde{z}_1)$ where T and M are as in (2.11) and (4.7) respectively. Thus, using the fact that $0 \in T(z^*)$ and T is monotone, we obtain $\langle M(z_0 - z_1), z^* - \tilde{z}_1\rangle \le 0$. Hence,

$$\begin{aligned}\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 &= \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_1)\|_M^2 - \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_0)\|_M^2 \\ &= \|\tilde{z}_1 - z_1\|_M^2 + 2\langle M(z_0 - z_1), z^* - \tilde{z}_1\rangle - \|\tilde{z}_1 - z_0\|_M^2 \\ &\le \|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2.\end{aligned}$$

Using (4.14), the inequality in (4.1), and the first equality in (4.4) (all with $k = 1$), we have

$$\|\tilde{z}_1 - z_1\|_M^2 \le \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{\tau_2}{\beta}\|\tilde{x}_1 - x_0\|^2 + \frac{1}{\alpha^3}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_P^2,$$

where P is as in (4.12). Now, (4.13) with $k = 1$ becomes

$$\|\tilde{z}_1 - z_0\|_M^2 \ge \frac{1}{\beta}\|\tilde{x}_1 - x_0\|^2 + \frac{1}{\alpha^3}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_N^2$$

where N is as in (4.12). Combining the last three inequalities and the fact that $\tau_2 < 1$ (see Algorithm 2), we find

$$\begin{aligned}\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 &\le \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{1}{\alpha^3}\left(\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_P^2 - \|(y_1 - y_0, \gamma_1 - \gamma_0)\|_N^2\right) \\ &= \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 - \frac{2-\alpha}{\alpha^2}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_S^2, \qquad (4.18)\end{aligned}$$

where the last equality is due to the fact that $P - N = -\alpha(2-\alpha)S$, with S given in (4.11). The last inequality, (4.11) with $k = 1$ and the fact that $\alpha \in (0, 2 - \tau_1)$ yield

$$\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 \le \frac{\alpha + \tau_1 - 2}{\alpha^2}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_S^2 \le 0,$$

which implies that (4.17) holds. Therefore, (a) now follows by combining (4.16) and (4.17).

(b) The proof of this statement is the same as the last part of the Lemma 3.2.2. ■

Now we are ready to show that Algorithm 2 is an instance of the modified HPE framework. We consider the following quantities

$$\sigma := \max\left\{\frac{1 + \alpha\tau_1}{1 + \alpha(2-\alpha)}, \tau_2\right\} \quad \text{and} \quad \xi := \frac{1}{\alpha^3}[\sigma(1 + \alpha - \alpha^2) + (1 - \tau_1)\alpha - 1]. \qquad (4.19)$$

Note that, in view of the assumptions on α, τ_1 and τ_2 in Algorithm 2, we trivially have $\sigma \in (0,1)$ and $\xi > 0$. Furthermore, if $\tau_1 = \tau_2 = 0$, we have $\sigma = \sigma_\alpha$, where σ_α is as in (3.15).

Theorem 4.2.4 *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in (4.9). Consider $\{\eta_k\}$ defined by*

$$\eta_0 = 4\xi d_0, \qquad \eta_k = \xi \|y_k - y_{k-1}\|_H^2, \qquad \forall\, k \geq 1, \tag{4.20}$$

where d_0 and ξ are as in (4.8) and (4.19), respectively. Then, for every $k \geq 1$,

$$M\,(z_{k-1} - z_k) \in T(\tilde{z}_k), \qquad \|z_k - \tilde{z}_k\|_M^2 + \eta_k \leq \sigma \|z_{k-1} - \tilde{z}_k\|_M^2 + \eta_{k-1}, \tag{4.21}$$

where T, M and σ are as in (2.11), (4.7) and (4.19), respectively. As a consequence, Algorithm 2 is an instance of the modified HPE framework with $\sigma < 1$.

Proof. The inclusion in (4.21) follows from Lemma 4.2.2. Let us now show the inequality in (4.21). Using (4.14) and the first relation in (4.4), we have

$$\begin{aligned}\|\tilde{z}_k - z_k\|_M^2 &= \frac{1}{\beta}\left\|\tilde{x}_k - x_{k-1} + \beta v_k\right\|^2 + \frac{1}{\alpha^3}\left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_P^2 \\ &\leq \frac{\tau_1}{\beta}\left\|\tilde{\gamma}_k - \gamma_{k-1}\right\|^2 + \frac{\tau_2}{\beta}\left\|\tilde{x}_k - x_{k-1}\right\|^2 + \frac{1}{\alpha^3}\left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_P^2,\end{aligned}$$

where the inequality is due to the second condition in (4.1). It follows from the last inequality, (4.13) and the fact that $\sigma \geq \tau_2$ (see (4.19)) that

$$\sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 \geq a_k \tag{4.22}$$

where

$$a_k := -\frac{\tau_1}{\beta}\left\|\tilde{\gamma}_k - \gamma_{k-1}\right\|^2 + \frac{1}{\alpha^3}\left(\sigma \left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_N^2 - \left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_P^2\right).$$

We will show that $a_k \geq \eta_k - \eta_{k-1}$, where the sequence $\{\eta_k\}$ is defined in (4.20). From (4.11), we find

$$\frac{\tau_1}{\beta}\left\|\tilde{\gamma}_k - \gamma_{k-1}\right\|^2 = \frac{1}{\alpha^3}\left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_{\alpha\tau_1 S}^2,$$

which, combined with definition of a_k, yields

$$a_k = \frac{1}{\alpha^3}\left\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\right\|_{\sigma N - \alpha\tau_1 S - P}^2.$$

Hence, using the definitions of N, S and P in (4.11) and (4.12), we obtain

$$a_k = \frac{1}{\alpha^3}\left(\hat{\xi}\beta\|B(y_k - y_{k-1})\|^2 + 2\xi\,\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{\tilde{\xi}}{\beta}\|\gamma_k - \gamma_{k-1}\|^2\right), \tag{4.23}$$

where

$$\hat{\xi} = \sigma(1+\alpha(2-\alpha)) - \alpha\tau_1 - 1, \quad \xi = \sigma(1+\alpha-\alpha^2) + (1-\tau_1)\alpha - 1, \quad \tilde{\xi} = \sigma - \alpha\tau_1 - (1-\alpha)^2. \tag{4.24}$$

Now, from the definition of σ given in (4.19), we obtain $\sigma \geq (1+\alpha\tau_1)/(1+\alpha(2-\alpha))$. Hence, $\hat{\xi} \geq 0$ and

$$\tilde{\xi} \geq \frac{1+\alpha\tau_1}{1+\alpha(2-\alpha)} - \alpha\tau_1 - (1-\alpha)^2 = \frac{\alpha^2(2-\tau_1-\alpha)(2-\alpha)}{1+\alpha(2-\alpha)} > 0,$$

where the last inequality is due to the fact that $\alpha \in (0, 2-\tau_1)$. Moreover, since $\sigma \in (0,1)$ (see (4.19)), we find

$$\xi = \sigma(1+\alpha-\alpha^2) + \alpha - \tau_1\alpha - 1 > \sigma(1+\alpha(2-\alpha)) - \alpha\tau_1 - 1 = \hat{\xi}.$$

Thus, $\xi > \hat{\xi} \geq 0$, and $\tilde{\xi} \geq 0$. Hence, from (4.23) and Lemma 4.2.3, it follows that

$$a_k \geq \frac{2\xi}{\alpha^3}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle \geq \begin{cases} \dfrac{1}{\alpha^3}\left(\xi\,\|y_1 - y_0\|_H^2 - 4\xi d_0\right), & k=1, \\ \dfrac{1}{\alpha^3}\left(\xi\,\|y_k - y_{k-1}\|_H^2 - \xi\,\|y_{k-1} - y_{k-2}\|_H^2\right), & k \geq 2, \end{cases}$$

which, combined with the definitions of $\{\eta_k\}$ in (4.20), yields $a_k \geq \eta_k - \eta_{k-1}$ for every $k \geq 1$. Hence, the desired inequality now follows from (4.22).

■

4.2.2 Iteration-complexity bounds for the inexact PG-ADMM

We next establish the iteration-complexity for Algorithm 2 in order to compute an approximate solution of (1.1). First, we present a pointwise iteration-complexity bound and subsequently we derive an ergodic iteration-complexity bound to obtain a relaxed approximate solution of (1.7) in the sense of (1.9). We mention that the pointwise iteration-complexity bound presented in Theorem 4.2.5 can also be derived from Theorem 4.2.4 combined with Theorem 2.2.4. However, we decide to present a direct and easy to follow proof, for completeness and convenience of the reader.

Theorem 4.2.5 *For a given tolerance $\rho > 0$, Algorithm 2 generates a ρ-approximate solution $(\tilde{x}_i, y_i, \tilde{\gamma}_i)$ of (1.7) with an associated residue $r_i = M(z_{i-1} - z_i)$ in at most $\mathcal{O}(d_0/\rho^2)$ iterations, where $\{z_i\}$ and d_0 are as in (4.9) and (4.8), respectively.*

Proof. First note that, in view of the inclusion in (4.21), we have $r_k := M(z_{k-1} - z_k)$ is a residue to the inclusion in (2.12) associated to $\tilde{z}_k$, for every $k \geq 1$. Let λ_M be the largest eigenvalue of M in (4.7). Hence, combining the definition of r_k, the inequality in (4.21) and simple algebra, we obtain

$$\begin{aligned} \|r_k\|^2 &\leq \lambda_M \|z_{k-1} - z_k\|_M^2 \leq 2\lambda_M\left[\|z_{k-1} - \tilde{z}_k\|_M^2 + \|\tilde{z}_k - z_k\|_M^2\right] \\ &\leq 2\lambda_M\left[(\sigma+1)\|z_{k-1} - \tilde{z}_k\|_M^2 + \eta_{k-1} - \eta_k\right], \end{aligned} \tag{4.25}$$

for every $k \geq 1$. From Proposition 4.2.4, Algorithm 2 is an instance of the modified HPE framework with $\{(z_k, \tilde{z}_k)\}$ and $\{\eta_k\}$ given in (4.9) and (4.20), respectively. Then, it follows from Lemma 2.2.3(b) and (4.25) that, for every $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$,

$$\begin{aligned}\sum_{k=1}^{i} \|r_k\|^2 &\leq \frac{2\lambda_M}{1-\sigma} \sum_{k=1}^{i} \left[(\sigma+1)\left(\|z_{k-1} - z^*\|_M^2 - \|z_k - z^*\|_M^2\right) + 2(\eta_{k-1} - \eta_k)\right] \\ &\leq \frac{2\lambda_M}{1-\sigma}\left((\sigma+1)\|z_0 - z^*\|_M^2 + 2\eta_0\right),\end{aligned}$$

which in turn, in view of the definitions of d_0 and η_0 given in (4.8) and (4.20), implies that there exists a scalar $c > 0$ such that

$$\sum_{k=1}^{i} \|r_k\|^2 \leq c d_0. \tag{4.26}$$

In particular, the latter inequality implies that $\{r_k\}$ converges to zero. Hence, let i be the first index in which $\|r_i\| \leq \rho$ (which is equivalent to say that $\tilde{z}_i$ is a ρ-approximate solution with residue r_i). Note that if $i = 1$, then the statement of the theorem trivially follows. Now assume that $i > 1$. It follows from (4.26) that

$$(i-1)\rho^2 < \sum_{k=1}^{i-1} \|r_k\|^2 \leq c d_0$$

and hence $i = \mathcal{O}(d_0/\rho^2)$, concluding the proof of the theorem. ∎

The next theorem presents the ergodic iteration-complexity bound for Algorithm 2.

Theorem 4.2.6 *Let $\{(x_k, y_k, \gamma_k, \tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 2 and consider the sequences $\{(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a)\}$ and $\{r_k^a\}$ defined by*

$$(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a) = \frac{1}{k}\sum_{i=1}^{k} (x_i, y_i, \gamma_i, \tilde{x}_i, \tilde{\gamma}_i), \qquad r_k^a = \frac{1}{k}\sum_{i=1}^{k} (z_{i-1} - z_i), \tag{4.27}$$

where $\{z_i\}$ is as in (4.9). Then, for every $k \geq 1$, there exist $\varepsilon_{k,x}^a, \varepsilon_{k,y}^a \geq 0$ such that the following relations hold

$$M r_k^a \in \left(\partial_{\varepsilon_{k,x}^a} f(\tilde{x}_k^a) - A^*\tilde{\gamma}_k^a,\ \partial_{\varepsilon_{k,y}^a} g(y_k^a) - B^*\tilde{\gamma}_k^a,\ A\tilde{x}_k^a + B y_k^a - b\right) \tag{4.28}$$

$$\|M r_k^a\| \leq \frac{\sqrt{\vartheta d_0}}{k}, \qquad \max\{\varepsilon_{k,x}^a, \varepsilon_{k,y}^a\} \leq \frac{\vartheta d_0}{k}, \tag{4.29}$$

where M and d_0 are as in (4.7) and (4.8), respectively, and ϑ is a positive scalar depending on (α, τ_1, τ_2) and the largest eigenvalue of M.

Proof. First of all, define $(v_i, u_i, w_i) = M(z_{i-1} - z_i)$ for every $i \geq 1$. Hence, it follows from Proposition 4.2.4, (2.11), and (4.7) that

$$v_i + A^*\tilde{\gamma}_i \in \partial f(\tilde{x}_i), \qquad u_i + B^*\tilde{\gamma}_i \in \partial g(y_i), \qquad w_i = A\tilde{x}_i + By_i - b. \tag{4.30}$$

On the one hand, from the above equality and (4.27), we have

$$w_k^a := \frac{1}{k}\sum_{i=1}^{k} w_i = A\tilde{x}_k^a + By_k^a - b. \tag{4.31}$$

Now, in view of the inclusions in (4.30), it follows from (4.27) and Proposition 2.1.1 that the sequences $\{\varepsilon_{k,x}^a\}$ and $\{\varepsilon_{k,y}^a\}$ defined by

$$\varepsilon_{k,x}^a := \frac{1}{k}\sum_{i=1}^{k} \langle v_i + A^*\tilde{\gamma}_i, \tilde{x}_i - \tilde{x}_k^a \rangle, \qquad \varepsilon_{k,y}^a := \frac{1}{k}\sum_{i=1}^{k} \langle u_i + B^*\tilde{\gamma}_i, y_i - y_k^a \rangle, \tag{4.32}$$

are nonnegative and

$$\frac{1}{k}\sum_{i=1}^{k} v_i \in \partial_{\varepsilon_{k,x}^a} f(\tilde{x}_k^a) - A^*\tilde{\gamma}_k^a, \quad \frac{1}{k}\sum_{i=1}^{k} u_i \in \partial_{\varepsilon_{k,y}^a} g(y_k^a) - B^*\tilde{\gamma}_k^a. \tag{4.33}$$

The inclusion in (4.28) follows from (4.31) and (4.33) and the fact that $\sum_{i=1}^{k}(v_i, u_i, w_i) = M(z_0 - z_k)$. Therefore, the proof of the existence of the elements $\varepsilon_{k,x}^a, \varepsilon_{k,y}^a \geq 0$ such that (4.28) holds is completed.

Let us now prove that (4.29) holds for $\varepsilon_{k,x}^a$ and $\varepsilon_{k,y}^a$ as defined above. Since Algorithm 2 is an instance of the modified HPE framework with $\sigma < 1$ (see Proposition 4.2.4), using Theorem 2.2.7, we have

$$\|r_k^a\|_M \leq \frac{2\sqrt{d_0 + \eta_0}}{k}, \qquad \varepsilon_k^a \leq \frac{3(3 - 2\sigma)(d_0 + \eta_0)}{2(1 - \sigma)k}, \tag{4.34}$$

where

$$\varepsilon_k^a = \frac{1}{k}\sum_{i=1}^{k} \langle M(z_{i-1} - z_i), \tilde{z}_i - \tilde{z}_k^a \rangle, \tag{4.35}$$

with $\{\tilde{z}_i\}$ given in (4.9) and $\tilde{z}_k^a := (\tilde{x}_k^a, y_k^a, \tilde{\gamma}_k^a)$. It is well-known that $\|Mr_k^a\|^2 \leq \lambda_M \|r_k^a\|_M^2$, where λ_M is the largest eigenvalue of M. Hence, using the first inequality in (4.34) and the definition of η_0 in (4.20), we conclude that the bound on $\|Mr_k^a\|$ in (4.29) holds with $\vartheta = \vartheta_1 := 4\lambda_M(1 + 4\xi)$. It remains to show the second estimate in (4.29). Using (4.32), we have

$$\begin{aligned}\varepsilon^a_{k,x}+\varepsilon^a_{k,y} &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle v_i,\tilde{x}_i-\tilde{x}^a_k\rangle+\langle u_i,y_i-y^a_k\rangle+\langle\tilde{\gamma}_i,A\tilde{x}_i-A\tilde{x}^a_k+By_i-By^a_k\rangle\Big)\\ &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle v_i,\tilde{x}_i-\tilde{x}^a_k\rangle+\langle u_i,y_i-y^a_k\rangle+\langle\tilde{\gamma}_i,w_i-w^a_k\rangle\Big),\end{aligned}$$

where the last equality is due to the definitions of w_i and w^a_k in (4.30) and (4.31), respectively. Additionally, the definitions of w_i, w^a_k and $\tilde{\gamma}^a_k$ imply that

$$\frac{1}{k}\sum_{i=1}^{k}\langle\tilde{\gamma}_i,w_i-w^a_k\rangle=\frac{1}{k}\sum_{i=1}^{k}\langle\tilde{\gamma}_i-\tilde{\gamma}^a_k,w_i-w^a_k\rangle=\frac{1}{k}\sum_{i=1}^{k}\langle w_i,\tilde{\gamma}_i-\tilde{\gamma}^a_k\rangle.$$

Therefore, since $z_i=(x_i,y_i,\gamma_i)$ and $M(z_{i-1}-z_i)=(v_i,u_i,w_i)$, it follows that $\varepsilon^a_{k,x}+\varepsilon^a_{k,y}=\varepsilon^a_k$, where ε^a_k is given in (4.35). Hence, using the estimate on ε^a_k given in (4.34) and the definition of η_0 in (4.20), we conclude that the second inequality in (4.29) holds with $\vartheta=\vartheta_2:=3(3-2\sigma)(1+4\xi)/2(1-\sigma)$. Therefore, the estimations in (4.29) trivially follow by defining $\vartheta=\max\{\vartheta_1,\vartheta_2\}$. ■

Remark 4.2.7 It follows from Theorem 4.2.6 that, for a given tolerance $\rho>0$, in at most $k=\mathcal{O}(\max\{\sqrt{d_0},d_0\}/\rho)$ iterations, the triple $(\tilde{x}^a_k,y^a_k,\tilde{\gamma}^a_k)$, together with r^a_k, satisfies the inclusion in (4.28) with $\varepsilon^a_{k,x},\varepsilon^a_{k,y}\geq 0$ and $\max\left\{\|Mr^a_k\|,\varepsilon^a_{k,x},\varepsilon^a_{k,y}\right\}\leq\rho$. Hence, the triple $(\tilde{x}^a_k,y^a_k,\tilde{\gamma}^a_k)$ can be seen as a relaxed ρ-approximate solution of (1.7) with residue $(v_{\bar{x}},v_{\bar{y}},v_{\bar{\gamma}}):=Mr^a_k$ in the sense that the inclusions in (1.7) are relaxed by using the ε-subdifferential operator instead of the subdifferential (see (1.9)). Therefore, Algorithm 2 provides a relaxed ρ-approximate solution of (1.7) in at most $\mathcal{O}(1/\rho)$ iterations. It should be mentioned that the quantities $\varepsilon^a_{k,x}$ and $\varepsilon^a_{k,y}$ can be explicitly computed (see (4.32)). Their expressions are not explicitly stated in order to simplify the statement of the theorem.

Chapter 5

An inexact proximal ADMM and its iteration-complexity analysis

In this chapter, we propose and analyze an inexact proximal ADMM for computing approximate solutions of (1.1). This chapter is related to [3] and is organized as follows. In Section 5.1, we introduce the proposed method and discuss its relationship with other ADMM variants. Section 5.2 is devoted to the iteration-complexity analysis of the proposed scheme. This section is divided into two subsections. The first one shows that our scheme falls within the setting of the modified HPE framework of Section 2.2, whereas in the last subsection, we establish the iteration-complexity bound for the proposed scheme in order to obtain an approximate solution of (1.1).

5.1 An inexact proximal ADMM (P-ADMM)

The inexact proximal ADMM proposed here is described as follows.

Algorithm 3: Inexact proximal ADMM

0. Let an initial point $(x_0, y_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$, a penalty parameter $\beta > 0$, two error tolerance parameters $\tau_1, \tau_2 \in [0, 1)$, and a self-adjoint positive semidefinite linear operator $H \colon \mathcal{Y} \to \mathcal{Y}$ be given. Choose a stepsize parameter

$$\theta \in \left(0, \frac{1 - 2\tau_1 + \sqrt{(1-2\tau_1)^2 + 4(1-\tau_1)}}{2(1-\tau_1)}\right), \tag{5.1}$$

and set $k = 1$.

1. Compute $(v_k, \tilde{x}_k) \in \mathcal{X} \times \mathcal{X}$ such that

$$v_k \in \partial f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \qquad \|\tilde{x}_k - x_{k-1} + \beta v_k\|^2 \le \tau_1 \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \tau_2 \|\tilde{x}_k - x_{k-1}\|^2, \tag{5.2}$$

where

$$\tilde{\gamma}_k = \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b), \tag{5.3}$$

and compute an optimal solution $y_k \in \mathcal{Y}$ of the subproblem

$$\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|A\tilde{x}_k + By - b\|^2 + \frac{1}{2} \|y - y_{k-1}\|_H^2 \right\}. \tag{5.4}$$

2. Set

$$x_k = x_{k-1} - \beta v_k, \qquad \gamma_k = \gamma_{k-1} - \theta\beta\,(A\tilde{x}_k + By_k - b) \tag{5.5}$$

and $k \leftarrow k + 1$, and go to step 1.

Remark 5.1.1 Some remaks about Algorithm 3 are in order:

(a) If $\tau_1 = \tau_2 = 0$, then $\tilde{x}_k = x_k$ due to the inequality in (5.2) and the first relation in (5.5). Hence, since $v_k = (x_{k-1} - x_k)/\beta$, the first subproblem of Step 1 is equivalent to compute an exact solution $x_k \in \mathcal{X}$ of the following subproblem

$$\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 + \frac{1}{2\beta} \|x - x_{k-1}\|^2 \right\}, \tag{5.6}$$

and then Algorithm 3 becomes the proximal ADMM (1.5) with stepsize parameter $\theta \in (0, (1+\sqrt{5})/2)$ and proximal terms given by $(1/\beta)I$ and H. Therefore, the proposed method can be seen as an extension of the proximal ADMM (1.5) in which subproblem (5.6) is solved inexactly using a relative approximate criterion.

(b) Subproblem (5.4) contains a proximal term defined by a self-adjoint positive semidefinite linear operator H which, appropriately chosen, makes the subproblem easier to solve or even to have closed-form solution. For instance, if $H = sI - \beta B^*B$ with $s > \beta\|B\|^2$, subproblem (5.4) is equivalent to

$$\min_{y\in\mathcal{Y}} \left\{ g(y) + \frac{s}{2}\|y - \bar{y}\|^2 \right\},$$

for some $\bar{y} \in \mathcal{Y}$, which has a closed-form solution in many applications. For example, if $g(\cdot) = \|\cdot\|_1$, then to solve the above problem corresponds to evaluating the well-known (explicitly computed) thresholding operator, see (6.5); we refer the reader to [6,62] for other examples in which the solution of the above proximal subproblem can be explicitly computed.

(c) The use of a relative approximate criterion in (5.4) requires, as far as we know, the stepsize parameter $\theta \in (0, 1]$. However, since, in many applications, the second subproblem (5.4) is solved exactly and a stepsize parameter $\theta > 1$ accelerates the method, here only the first subproblem is assumed to be solved inexactly.

(d) The inexact proximal ADMM is close related to [29, Algorithm 2]. Indeed, the latter method corresponds to the former one with $H = 0$, $\theta = 1$ and the following condition

$$2\beta|\langle \tilde{x}_k - x_{k-1}, v_k\rangle| + \beta^2\|v_k\|^2 \le \tau_1\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 \tag{5.7}$$

instead of the inequality in (5.2). Numerical comparisons between the inexact proximal ADMM and [29, Algorithm 2] will be provided in Chapter 6.

Some preliminary numerical experiments to illustrate the advantages of Algorithm 3 are reported in Chapter 6.

5.2 Iteration-complexity of the inexact P-ADMM

In this section, we present an iteration-complexity analysis for the inexact proximal ADMM in order to obtain approximate solution of (1.1). As previously mentioned, our analysis is done by showing that it is an instance of the modified HPE framework for computing approximate solutions of the Lagrangian system (1.7). Thus, we need to introduce the elements required by the setting of Section 2.2. Namely, consider the self-adjoint positive semidefinite linear operator

$$M = \begin{bmatrix} I/\beta & 0 & 0 \\ 0 & (H + \beta B^*B) & 0 \\ 0 & 0 & I/(\theta\beta) \end{bmatrix}. \tag{5.8}$$

In this setting, the quantity d_0 defined in (2.6) becomes

$$d_0 = \inf\left\{\|(x - x_0, y - y_0, \gamma - \gamma_0)\|_M^2 : (x, y, \gamma) \in T^{-1}(0)\right\}, \tag{5.9}$$

where T is as in (2.11).

5.2.1 Inexact P-ADMM in the setting of the modified HPE framework

Our main goal in this subsection is to show that Algorithm 3 falls within the setting of the modified HPE framework. We start by presenting a preliminary technical result, which basically shows that a certain sequence generated by Algorithm 3 satisfies the inclusion in (2.5b) with T and M as in (2.11) and (5.8), respectively.

Lemma 5.2.1 *Consider (x_k, y_k, γ_k) and $(\tilde{x}_k, \tilde{\gamma}_k)$ generated at the k-iteration of Algorithm 3. Then,*

$$\frac{1}{\beta}(x_{k-1} - x_k) \in \partial f(\tilde{x}_k) - A^*\tilde{\gamma}_k, \tag{5.10}$$

$$(H + \beta B^*B)(y_{k-1} - y_k) \in \partial g(y_k) - B^*\tilde{\gamma}_k, \tag{5.11}$$

$$\frac{1}{\theta\beta}(\gamma_{k-1} - \gamma_k) = A\tilde{x}_k + By_k - b. \tag{5.12}$$

As a consequence, $z_k = (x_k, y_k, \gamma_k)$ and $\tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k)$ satisfy inclusion (2.5a) with T and M as in (2.11) and (5.8), respectively.

Proof. Inclusion (5.10) follows trivially from the inclusion in (5.2) and the first relation in (5.5). Now, from the optimality condition of (5.4) and the definition of $\tilde{\gamma}_k$ in (5.3), we obtain

$$\begin{aligned}
0 &\in \partial g(y_k) - B^*\gamma_{k-1} + \beta B^*(A\tilde{x}_k + By_k - b) + H(y_k - y_{k-1}) \\
&= \partial g(y_k) - B^*[\gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b)] + \beta B^*B(y_k - y_{k-1}) + H(y_k - y_{k-1}) \\
&= \partial g(y_k) - B^*\tilde{\gamma}_k + \beta B^*B(y_k - y_{k-1}) + H(y_k - y_{k-1}).
\end{aligned}$$

which proves to (5.11). The relation (5.12) follows immediately from the second relation in (5.5). To end the proof, note that the last statement of the lemma follows directly by (5.10)–(5.12) and definitions of T and M in (2.11) and (5.8), respectively. ∎

The following result presents some relations satisfied by the sequences generated by the inexact proximal ADMM. These relations are essential to show that the latter method is an instance of the modified HPE framework.

Lemma 5.2.2 *Let $\{(x_k, y_k, \gamma_k)\}$ and $\{(\tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 3. Then, the following hold:*

(a) *for any $k \geq 1$, we have*

$$\tilde{\gamma}_k - \gamma_{k-1} = \frac{1}{\theta}(\gamma_k - \gamma_{k-1}) + \beta B(y_k - y_{k-1}), \quad \tilde{\gamma}_k - \gamma_k = \frac{1-\theta}{\theta}(\gamma_k - \gamma_{k-1}) + \beta B(y_k - y_{k-1});$$

(b) *we have*

$$\frac{1}{2}\|y_1 - y_0\|_H^2 - \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \leq 2\max\left\{1, \frac{\theta}{2-\theta}\right\} d_0,$$

where d_0 is as in (5.9);

(c) *for every $k \geq 2$, we have*

$$\begin{aligned}\frac{1}{\theta}\left\langle \gamma_k - \gamma_{k-1}, B(y_k - y_{k-1})\right\rangle \geq \frac{1-\theta}{\theta}\left\langle \gamma_{k-1} - \gamma_{k-2}, B(y_k - y_{k-1})\right\rangle \\ + \frac{1}{2}\|y_k - y_{k-1}\|_H^2 - \frac{1}{2}\|y_{k-1} - y_{k-2}\|_H^2.\end{aligned}$$

Proof. (a) The first relation follows by noting that the definitions of $\tilde{\gamma}_k$ and γ_k in (5.3) and (5.5), respectively, yield

$$\tilde{\gamma}_k - \gamma_{k-1} = -\beta(A\tilde{x}_k + By_{k-1} - b) = \frac{1}{\theta}(\gamma_k - \gamma_{k-1}) + \beta B(y_k - y_{k-1}).$$

The second relation in (a) follows trivially from the first one.

(b) First, note that

$$\begin{aligned}0 &\leq \frac{1}{2\beta}\left\|\frac{1}{\sqrt{\theta}}(\gamma_1 - \gamma_0) + \beta B(y_1 - y_0)\right\|^2 \\ &= \frac{1}{2\theta\beta}\|\gamma_1 - \gamma_0\|^2 + \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle + \frac{\beta}{2}\|B(y_1 - y_0)\|^2,\end{aligned}$$

which, for every $z^* = (x^*, y^*, \gamma^*) \in \Omega^*$, yields

$$\begin{aligned}&\frac{1}{2}\|y_1 - y_0\|_H^2 - \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \\ &\qquad \leq \frac{1}{2}\left(\|y_1 - y_0\|_H^2 + \frac{1}{\theta\beta}\|\gamma_1 - \gamma_0\|^2 + \beta\|B(y_1 - y_0)\|^2\right) \\ &\qquad \leq \|y_1 - y^*\|_H^2 + \|y_0 - y^*\|_H^2 + \frac{1}{\theta\beta}\|\gamma_1 - \gamma^*\|^2 + \frac{1}{\theta\beta}\|\gamma_0 - \gamma^*\|^2 \\ &\qquad + \beta\|B(y_1 - y^*)\|^2 + \beta\|B(y_0 - y^*)\|^2,\end{aligned}$$

where the last inequality is due to the second property in (2.1). Hence, using (5.8), we obtain

$$\frac{1}{2}\|y_1 - y_0\|_H^2 - \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \le \|z_1 - z^*\|_M^2 + \|z_0 - z^*\|_M^2, \tag{5.13}$$

where $z_0 = (x_0, y_0, \gamma_0)$ and $z_1 = (x_1, y_1, \gamma_1)$. On the other hand, from Lemma 5.2.1 with $k = 1$, we have $M(z_0 - z_1) \in T(\tilde{z}_1)$, where $\tilde{z}_1 = (\tilde{x}_1, y_1, \tilde{\gamma}_1)$ and T is as in (2.11). Using this fact and the monotonicity of T, we obtain $\langle \tilde{z}_1 - z^*, M(z_0 - z_1)\rangle \ge 0$ for all $z^* = (x^*, y^*, z^*) \in \Omega^*$. Hence,

$$\begin{aligned}\|z^* - z_0\|_M^2 - \|z^* - z_1\|_M^2 &= \|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 + 2\langle \tilde{z}_1 - z^*, M(z_0 - z_1)\rangle \\ &\ge \|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2.\end{aligned} \tag{5.14}$$

It follows from (5.8), item (a), and some direct calculations that

$$\begin{aligned}\|\tilde{z}_1 - z_1\|_M^2 &= \frac{1}{\beta}\|\tilde{x}_1 - x_1\|^2 + \frac{1}{\theta\beta}\|\tilde{\gamma}_1 - \gamma_1\|^2 \\ &= \frac{1}{\beta}\|\tilde{x}_1 - x_1\|^2 + \frac{1}{\theta\beta}\left\|\frac{1-\theta}{\theta}(\gamma_1 - \gamma_0) + \beta B(y_1 - y_0)\right\|^2 \\ &= \frac{1}{\beta}\|\tilde{x}_1 - x_1\|^2 + \frac{(1-\theta)^2}{\beta\theta^3}\|\gamma_1 - \gamma_0\|^2 + \frac{\beta}{\theta}\|B(y_1 - y_0)\|^2 \\ &+ \frac{2(1-\theta)}{\theta^2}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle.\end{aligned} \tag{5.15}$$

Moreover, (5.8) and item (a) also yield

$$\begin{aligned}&\|\tilde{z}_1 - z_0\|_M^2 = \frac{1}{\beta}\|\tilde{x}_1 - x_0\|^2 + \|y_1 - y_0\|_{(\beta B^*B + H)}^2 + \frac{1}{\theta\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 \\ &\ge \frac{1}{\beta}\|\tilde{x}_1 - x_0\|^2 + \beta\|B(y_1 - y_0)\|^2 + \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{1-\tau_1\theta}{\theta\beta}\left\|\frac{1}{\theta}(\gamma_1 - \gamma_0) + \beta B(y_1 - y_0)\right\|^2 \\ &= \frac{1}{\beta}\|\tilde{x}_1 - x_0\|^2 + \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{[1 + (1-\tau_1)\theta]\beta}{\theta}\|B(y_1 - y_0)\|^2 + \frac{1-\tau_1\theta}{\beta\theta^3}\|\gamma_1 - \gamma_0\|^2 \\ &+ \frac{2(1-\tau_1\theta)}{\theta^2}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle.\end{aligned} \tag{5.16}$$

Combining the above two conclusions, we obtain

$$\begin{aligned}&\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \ge \frac{1}{\beta}\left(\|\tilde{x}_1 - x_0\|^2 - \|\tilde{x}_1 - x_1\|^2 + \tau_1\|\tilde{\gamma}_1 - \gamma_0\|^2\right) \\ &+ (1-\tau_1)\beta\|B(y_1 - y_0)\|^2 + \frac{2-\theta-\tau_1}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 + \frac{2(1-\tau_1)}{\theta}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle.\end{aligned} \tag{5.17}$$

Now, note that the inequality in (5.2) with $k = 1$ and the definition of x_1 in (5.5) imply that

$$0 \le \tau_2\|\tilde{x}_1 - x_0\|^2 - \|\tilde{x}_1 - x_1\|^2 + \tau_1\|\tilde{\gamma}_1 - \gamma_0\|^2$$

which, combined with (5.17) and $\tau_2 \in [0,1)$, yields

$$
\begin{aligned}
&\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \\
&\quad \geq (1-\tau_1)\beta\|B(y_1 - y_0)\|^2 + \frac{2-\theta-\tau_1}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 + \frac{2(1-\tau_1)}{\theta}\langle B(y_1-y_0), \gamma_1 - \gamma_0\rangle \\
&\quad = \frac{1-\theta}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 + (1-\tau_1)\left\|\sqrt{\beta}B(y_1 - y_0) + \frac{1}{\theta\sqrt{\beta}}(\gamma_1 - \gamma_0)\right\|^2 \geq \frac{1-\theta}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2.
\end{aligned}
$$

Hence, if $\theta \in (0,1]$, then we have

$$
\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \geq 0,
$$

which, combined with (5.14), yields

$$
\|z_1 - z^*\|_M^2 \leq \|z_0 - z^*\|_M^2. \tag{5.18}
$$

Now, if $\theta > 1$, then we have

$$
\begin{aligned}
\|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2 &\leq \frac{\theta - 1}{\beta\theta^2}\|\gamma_1 - \gamma_0\|^2 \\
&\leq \frac{2(\theta-1)}{\theta}\left(\frac{1}{\beta\theta}\|\gamma_1 - \gamma^*\|^2 + \frac{1}{\beta\theta}\|\gamma_0 - \gamma^*\|^2\right) \\
&\leq \frac{2(\theta-1)}{\theta}\left[\|z_0 - z^*\|_M^2 + \|z_1 - z^*\|_M^2\right]
\end{aligned} \tag{5.19}
$$

where the second inequality is due to the second property in (2.1), and the last inequality is due to (5.8) and definitions of z_0, z_1 and z^*. It follows from (5.1) that $\theta < (1+\sqrt{5})/2$, in particular, $\theta < 2$. Hence, adding (5.14) and (5.19), we obtain

$$
\|z_1 - z^*\|_M^2 \leq \frac{3\theta - 2}{2-\theta}\|z_0 - z^*\|_M^2.
$$

Thus, it follows from (5.18) and the last inequality that

$$
\|z_1 - z^*\|_M^2 \leq \max\left\{1, \frac{3\theta-2}{2-\theta}\right\}\|z_0 - z^*\|_M^2. \tag{5.20}
$$

Therefore, the desired inequality follows from (5.13), (5.20) and the definition of d_0 in (5.9).

(c) From the optimality condition for (5.4), the definition of $\tilde{\gamma}_k$ in (5.3) and item (a), we have, for every $k \geq 1$,

$$
\partial g(y_k) \ni B^*(\tilde{\gamma}_k - \beta B(y_k - y_{k-1})) - H(y_k - y_{k-1}) = \frac{1}{\theta}B^*(\gamma_k - (1-\theta)\gamma_{k-1}) - H(y_k - y_{k-1}).
$$

For any $k \geq 2$, using the above inclusion with $k \leftarrow k$ and $k \leftarrow k-1$ and the monotonicity of ∂g , we obtain

$$
\begin{aligned}
&\frac{1}{\theta}\langle B^*(\gamma_k - \gamma_{k-1}) - (1-\theta)B^*(\gamma_{k-1} - \gamma_{k-2}), y_k - y_{k-1}\rangle \\
&\geq \langle H(y_k - y_{k-1}), y_k - y_{k-1}\rangle - \langle H(y_{k-1} - y_{k-2}), y_k - y_{k-1}\rangle \\
&\geq \frac{1}{2}\|y_k - y_{k-1}\|_H^2 - \frac{1}{2}\|y_{k-1} - y_{k-2}\|_H^2,
\end{aligned}
$$

where the last inequality is due to the first property in (2.1), and so the proof of the lemma follows. ■

We next consider a technical result.

Lemma 5.2.3 *Let scalars* τ_1, τ_2 *and* θ *be as in step 0 of Algorithm 3. Then, there exists a scalar* $\sigma \in [\tau_2, 1)$ *such that the matrix*

$$
L = \begin{bmatrix} \sigma - 1 + (\sigma - \tau_1)\theta & -|(1-\theta)[\sigma - 1 + (1-\tau_1)\theta]| \\ -|(1-\theta)[\sigma - 1 + (1-\tau_1)\theta]| & \sigma - 1 + (2 - \theta - \tau_1)\theta \end{bmatrix} \tag{5.21}
$$

is positive definite.

Proof. Since τ_1 and θ are fixed scalars given in step 0 of Algorithm 3, the determinant and trace of L are polynomial functions of σ denoted here by $\Phi(\sigma)$ and $\tilde{\Phi}(\sigma)$, respectively. It is easy to see that

$$
\Phi(1) = \theta^2(1-\tau_1)\left[-(1-\tau_1)\theta^2 + (1-2\tau_1)\theta + 1\right], \quad \tilde{\Phi}(1) = [3 - 2\tau_1 - \theta]\theta.
$$

Note that the upper bound on θ given in (5.1), namely,

$$
\hat{\theta} := \frac{1 - 2\tau_1 + \sqrt{(1-2\tau_1)^2 + 4(1-\tau_1)}}{2(1-\tau_1)}
$$

corresponds to the positive root of the quadratic $q(\theta) = -(1-\tau_1)\theta^2 + (1-2\tau_1)\theta + 1$, which appears in the expression of $\Phi(1)$. Hence, since $\tau_1 \in [0,1)$ and $\theta \in (0, \hat{\theta})$, we can conclude that $\Phi(1) > 0$. Now, by using $\tau_1 \in [0,1)$ and some simple algebraic manipulations, it can be verified that $\hat{\theta} < 3 - 2\tau_1$, which, combined with the fact that $\theta \in (0, \hat{\theta})$, yields $\tilde{\Phi}(1) > 0$. Therefore, there exists $\hat{\sigma} \in [0,1)$ such that $\Phi(\sigma) > 0$ and $\tilde{\Phi}(\sigma) > 0$ for all $\sigma \in [\hat{\sigma}, 1)$, which in turn implies that $L := L(\sigma)$ is positive definite for all $\sigma \in [\hat{\sigma}, 1)$. The statement of the lemma follows now by choosing $\sigma = \max\{\tau_2, \hat{\sigma}\}$. ■

In the following, we show that the inexact proximal ADMM can be regarded as an instance of the modified HPE framework.

Theorem 5.2.4 *Let $\{(x_k, y_k, \gamma_k)\}$ and $\{(\tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 3. Let also T, M and d_0 be as in (2.11), (5.8) and (5.9), respectively. Define*

$$z_0 = (x_0, y_0, \gamma_0), \quad \mu = \frac{4[\sigma - 1 + (1-\tau_1)\theta]}{\theta^{3/2}} \max\left\{1, \frac{\theta}{2-\theta}\right\}, \quad \eta_0 = \mu d_0 \tag{5.22}$$

and, for all $k \geq 1$,

$$z_k = (x_k, y_k, \gamma_k), \qquad \tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k), \tag{5.23}$$

$$\eta_k = \frac{[\sigma - 1 + (2-\theta-\tau_1)\theta]}{\beta\theta^3}\|\gamma_k - \gamma_{k-1}\|^2 + \frac{[\sigma - 1 + (1-\tau_1)\theta]}{\theta}\|y_k - y_{k-1}\|_H^2, \tag{5.24}$$

where $\sigma \in [\tau_2, 1)$ is given by Lemma 5.2.3. Then, $(z_k, \tilde{z}_k, \eta_k)$ satisfies the error condition in (2.5b) for every $k \geq 1$. As a consequence, the inexact proximal ADMM is an instance of the modified HPE framework with $\sigma < 1$.

Proof. First of all, since $\sigma < 1$ and the matrix L in (5.21) is positive definite (in particular, l_{11} is positive), we have

$$[\sigma - 1 + (1-\tau_1)\theta] \geq [\sigma - 1 + (\sigma - \tau_1)\theta] = l_{11} > 0. \tag{5.25}$$

Now, using (5.8) and definitions of $\{z_k\}$ and $\{\tilde{z}_k\}$ in (5.23), we obtain

$$\|\tilde{z}_k - z_{k-1}\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|_H^2 + \beta\|B(y_k - y_{k-1})\|^2 + \frac{1}{\beta\theta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2,$$

$$\|\tilde{z}_k - z_k\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_k\|^2 + \frac{1}{\beta\theta}\|\tilde{\gamma}_k - \gamma_k\|^2.$$

Hence,

$$\begin{aligned}\sigma\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 &= \frac{1}{\beta}\left(\sigma\|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 + \tau_1\|\tilde{\gamma}_k - \gamma_{k-1}\|^2\right) \\ &+ \sigma\|y_k - y_{k-1}\|_H^2 + \sigma\beta\|B(y_k - y_{k-1})\|^2 + \frac{\sigma - \tau_1\theta}{\beta\theta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 - \frac{1}{\beta\theta}\|\tilde{\gamma}_k - \gamma_k\|^2.\end{aligned} \tag{5.26}$$

Note that the inequality in (5.2) and definition of x_k in (5.4) imply that

$$0 \leq \tau_2\|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 + \tau_1\|\tilde{\gamma}_k - \gamma_{k-1}\|^2$$

which, combined with (5.26) and the fact that $\sigma \geq \tau_2$, yields

$$\begin{aligned}\sigma\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 \geq \sigma\|y_k - y_{k-1}\|_H^2 + \sigma\beta\|B(y_k - y_{k-1})\|^2 \\ + \frac{\sigma - \tau_1\theta}{\beta\theta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 - \frac{1}{\beta\theta}\|\tilde{\gamma}_k - \gamma_k\|^2.\end{aligned} \tag{5.27}$$

On the other hand, it follows from Lemma 5.2.2(a) that

$$
\begin{aligned}
&\frac{\sigma-\tau_1\theta}{\beta\theta}\|\tilde{\gamma}_k-\gamma_{k-1}\|^2-\frac{1}{\beta\theta}\|\tilde{\gamma}_k-\gamma_k\|^2\\
&=\frac{\sigma-\tau_1\theta}{\beta\theta}\left\|\frac{1}{\theta}(\gamma_k-\gamma_{k-1})+\beta B(y_k-y_{k-1})\right\|^2-\frac{1}{\beta\theta}\left\|\frac{1-\theta}{\theta}(\gamma_k-\gamma_{k-1})+\beta B(y_k-y_{k-1})\right\|^2\\
&=\frac{\sigma-1+(2-\theta-\tau_1)\theta}{\beta\theta^3}\|\gamma_k-\gamma_{k-1}\|^2+\frac{(\sigma-1-\tau_1\theta)\beta}{\theta}\|B(y_k-y_{k-1})\|^2\\
&+\frac{2[\sigma-1+(1-\tau_1)\theta]}{\theta^2}\langle\gamma_k-\gamma_{k-1},B(y_k-y_{k-1})\rangle.
\end{aligned}
$$

Hence, combining the last equality and (5.27), we obtain

$$
\begin{aligned}
\sigma\|\tilde{z}_k-z_{k-1}\|_M^2-\|\tilde{z}_k-z_k\|_M^2\geq{}&\sigma\|y_k-y_{k-1}\|_H^2\\
&+\frac{[\sigma-1+(2-\theta-\tau_1)\theta]}{\beta\theta^3}\|\gamma_k-\gamma_{k-1}\|^2+\frac{[\sigma-1+(\sigma-\tau_1)\theta]\beta}{\theta}\|B(y_k-y_{k-1})\|^2\\
&+\frac{2[\sigma-1+(1-\tau_1)\theta]}{\theta^2}\langle\gamma_k-\gamma_{k-1},B(y_k-y_{k-1})\rangle. \qquad (5.28)
\end{aligned}
$$

We will now consider two cases: $k=1$ and $k>1$.

Case 1 ($k=1$): Since $[\sigma-1+(1-\tau_1)\theta]>0$ (see (5.25)), it follows from (5.28) with $k=1$ and Lemma 5.2.2(b) that

$$
\begin{aligned}
\sigma\|\tilde{z}_1-z_0\|_M^2-\|\tilde{z}_1-z_1\|_M^2\geq{}&\left[\sigma+\frac{[\sigma-1+(1-\tau_1)\theta]}{\theta^{3/2}}\right]\|y_1-y_0\|_H^2\\
&+\frac{[\sigma-1+(2-\theta-\tau_1)\theta]}{\beta\theta^3}\|\gamma_1-\gamma_0\|^2+\frac{[\sigma-1+(\sigma-\tau_1)\theta]\beta}{\theta}\|B(y_1-y_0)\|^2\\
&-\frac{4[\sigma-1+(1-\tau_1)\theta]}{\theta^{3/2}}\max\left\{1,\frac{\theta}{2-\theta}\right\}d_0
\end{aligned}
$$

which, combined with definitions of η_0 and η_1 in (5.22) and (5.24), respectively, yields

$$
\begin{aligned}
\sigma\|\tilde{z}_1-z_0\|_M^2-\|\tilde{z}_1-z_1\|_M^2+\eta_0-\eta_1\geq{}&\frac{[\sigma-1+(\sigma-\tau_1)\theta]\beta}{\theta}\|B(y_1-y_0)\|^2\\
&+\left[\sigma+\frac{[\sigma-1+(1-\tau_1)\theta]}{\theta^{3/2}}-\frac{[\sigma-1+(1-\tau_1)\theta]}{\theta}\right]\|y_1-y_0\|_H^2.
\end{aligned}
$$

From the last inequality and some algebraic manipulations, we obtain

$$
\begin{aligned}
\sigma\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 + \eta_0 - \eta_1 \\
&\geq \frac{[\sigma - 1 + (\sigma - \tau_1)\theta]}{\theta}\left(\beta\|B(y_1 - y_0)\|^2 + \frac{1}{\sqrt{\theta}}\|y_1 - y_0\|_H^2\right) \\
&+ \left[\sigma + \frac{1-\sigma}{\sqrt{\theta}} - \frac{[\sigma - 1 + (1-\tau_1)\theta]}{\theta}\right]\|y_1 - y_0\|_H^2 \\
&= \frac{[\sigma - 1 + (\sigma - \tau_1)\theta]}{\theta}\left(\beta\|B(y_1 - y_0)\|^2 + \frac{1}{\sqrt{\theta}}\|y_1 - y_0\|_H^2\right) \\
&+ \frac{[(1-\sigma)(1+\sqrt{\theta} - \theta) + \tau_1\theta]}{\theta}\|y_1 - y_0\|_H^2. \qquad (5.29)
\end{aligned}
$$

Using (5.1), we have $\theta \in]0, (1+\sqrt{5})/2[$ which in turn implies that $(1 + \sqrt{\theta} - \theta) \geq 0$. Hence, inequality (2.5b) with $k = 1$ follows from (5.25), (5.29) and the fact that $\sigma < 1$.

Case 2 ($k > 1$): Since $[\sigma - 1 + (1 - \tau_1)\theta] > 0$ (see (5.25)), it follows from (5.28) and Lemma 5.2.2(c) that

$$
\begin{aligned}
\sigma\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 &\geq \frac{[\sigma - 1 + (1-\tau_1)\theta]}{\theta}\left(\|y_k - y_{k-1}\|_H^2 - \|y_{k-1} - y_{k-2}\|_H^2\right) \\
&+ \frac{[\sigma - 1 + (2 - \theta - \tau_1)\theta]}{\beta\theta^3}\|\gamma_k - \gamma_{k-1}\|^2 + \frac{[\sigma - 1 + (\sigma - \tau_1)\theta]\beta}{\theta}\|B(y_k - y_{k-1})\|^2 \\
&+ \frac{2(1-\theta)[\sigma - 1 + (1-\tau_1)\theta]}{\theta^2}\langle \gamma_{k-1} - \gamma_{k-2}, B(y_k - y_{k-1})\rangle
\end{aligned}
$$

which, combined with definition of $\{\eta_k\}$ in (5.24) and the Cauchy-Schwarz inequality, yields

$$
\begin{aligned}
\sigma\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 + \eta_{k-1} - \eta_k \\
&\geq \frac{[\sigma - 1 + (2 - \theta - \tau_1)\theta]}{\beta\theta^3}\|\gamma_{k-1} - \gamma_{k-2}\|^2 + \frac{[\sigma - 1 + (\sigma - \tau_1)\theta]\beta}{\theta}\|B(y_k - y_{k-1})\|^2 \\
&- \frac{2|(1-\theta)[\sigma - 1 + (1-\tau_1)\theta]|}{\theta^2}\|\gamma_{k-1} - \gamma_{k-2}\|\|B(y_k - y_{k-1})\| \\
&= \frac{1}{\theta}\left\langle L\begin{bmatrix}\sqrt{\beta}\|B(y_k - y_{k-1})\| \\ \|\gamma_{k-1} - \gamma_{k-2}\|/\theta\sqrt{\beta}\end{bmatrix}, \begin{bmatrix}\sqrt{\beta}\|B(y_k - y_{k-1})\| \\ \|\gamma_{k-1} - \gamma_{k-2}\|/\theta\sqrt{\beta}\end{bmatrix}\right\rangle
\end{aligned}
$$

where L is as in (5.21). Therefore, since L is positive definite (see Lemma 5.2.3(b)), we conclude that inequality (2.5b) also holds for $k > 1$.

To end the proof, note that the last statement of the proposition follows trivially from the first one and Lemma 5.2.1. ■

5.2.2 Iteration-complexity bounds for the inexact P-ADMM

We are now ready to establish pointwise and ergodic iteration-complexity bounds for the inexact proximal ADMM in order to obtain an approximate solution of problem (1.1).

Theorem 5.2.5 *Consider the sequences $\{(x_k, y_k, \gamma_k)\}$ and $\{(\tilde{x}_k, \tilde{\gamma}_k)\}$ generated by Algorithm 3. Then, for every $k \geq 1$,*

$$\begin{pmatrix} \frac{1}{\beta}(x_{k-1} - x_k) \\ (H + \beta B^* B)(y_{k-1} - y_k) \\ \frac{1}{\beta\theta}(\gamma_{k-1} - \gamma_k) \end{pmatrix} \in \begin{bmatrix} \partial f(\tilde{x}_k) - A^*\tilde{\gamma}_k \\ \partial g(y_k) - B^*\tilde{\gamma}_k \\ A\tilde{x}_k + By_k - b \end{bmatrix} \tag{5.30}$$

and there exist $\sigma \in (0,1)$ and $i \leq k$ such that

$$\left(\frac{1}{\beta}\|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2_{(H+\beta B^* B)} + \frac{1}{\beta\theta}\|\gamma_i - \gamma_{i-1}\|^2\right)^{1/2} \leq \frac{\sqrt{d_0}}{\sqrt{k}}\sqrt{\frac{2(1+\sigma)+4\mu}{1-\sigma}},$$

where d_0 and μ are as in (5.9) and (5.22), respectively.

Proof. This result follows by combining Theorem 5.2.4 and Theorem 2.2.4. ■

Remark 5.2.6 For a given tolerance $\bar{\rho} > 0$, Theorem 5.2.5 ensures that in at most $\mathcal{O}(1/\bar{\rho}^2)$ iterations, Algorithm 3 provides an approximate solution $(\hat{x}, \hat{y}, \hat{\gamma})$ of the Lagrangian system (1.7) together with a residual $r := (r_x, r_y, r_\gamma)$ in the sense that

$$\frac{1}{\beta}r_x \in \partial f(\hat{x}) - A^*\hat{\gamma}, \qquad (H + \beta B^* B)r_y \in \partial g(\hat{y}) - B^*\hat{\gamma}, \qquad \frac{1}{\beta\theta}r_\gamma = A\hat{x} + B\hat{y} - b,$$

and $\|(r_x, r_y, r_\gamma)\|_M \leq \bar{\rho}$, where M is as in (5.8). Note that, for a given tolerance $\rho > 0$, the above relations are equivalent to (1.8) with $(v_{\hat{x}}, v_{\hat{y}}, v_{\hat{\gamma}}) := Mr$, $\bar{\rho} := \rho/\sqrt{\lambda_M}$, where λ_M is the largest eigenvalue of M, and the fact that $\|M(\cdot)\| \leq \sqrt{\lambda_M}\|\cdot\|_M$. Therefore, Algorithm 3 provides a ρ-approximate solution of (1.7) in at most $\mathcal{O}(1/\rho^2)$ iterations.

Theorem 5.2.7 *Let the sequences $\{(x_k, y_k, \gamma_k)\}$ and $\{(\tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 3. Consider the ergodic sequences $\{(x_k^a, y_k^a, \gamma_k^a)\}$, $\{(\tilde{x}_k^a, \tilde{\gamma}_k^a)\}$, $\{(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a)\}$ and $\{(\varepsilon_{k,x}^a, \varepsilon_{k,y}^a)\}$ defined by*

$$(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a) = \frac{1}{k}\sum_{i=1}^{k}(x_i, y_i, \gamma_i, \tilde{x}_i, \tilde{\gamma}_i), \quad (r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a) = \frac{1}{k}\sum_{i=1}^{k}(r_{i,x}, r_{i,y}, r_{i,\gamma}), \tag{5.31}$$

$$(\varepsilon_{k,x}^a, \varepsilon_{k,y}^a) = \frac{1}{k}\sum_{i=1}^{k}\left(\langle r_{i,x}/\beta + A^*\tilde{\gamma}_i, \tilde{x}_i - \tilde{x}_k^a\rangle, \langle (H + \beta B^* B)\, r_{i,y} + B^*\tilde{\gamma}_i, y_i - y_k^a\rangle\right), \tag{5.32}$$

where

$$(r_{i,x}, r_{i,y}, r_{i,\gamma}) = (x_{i-1} - x_i, y_{i-1} - y_i, \gamma_{i-1} - \gamma_i)\,. \tag{5.33}$$

Then, for every $k \geq 1$, we have $\varepsilon^a_{k,x}, \varepsilon^a_{k,y} \geq 0$,

$$\begin{pmatrix} \frac{1}{\beta} r^a_{k,x} \\ (H + \beta B^* B) r^a_{k,y} \\ \frac{1}{\beta\theta} r^a_{k,\gamma} \end{pmatrix} \in \begin{bmatrix} \partial_{\varepsilon^a_{k,x}} f(\tilde{x}^a_k) - A^* \tilde{\gamma}^a_k \\ \partial_{\varepsilon^a_{k,y}} g(y^a_k) - B^* \tilde{\gamma}^a_k \\ A\tilde{x}^a_k + B y^a_k - b, \end{bmatrix}, \tag{5.34}$$

and there exists $\sigma \in (0,1)$ such that

$$\left(\frac{1}{\beta} \|r^a_{k,x}\|^2 + \|r^a_{k,y}\|^2_{(H+\beta B^* B)} + \frac{1}{\beta\theta} \|r^a_{k,\gamma}\|^2 \right)^{1/2} \leq \frac{2\sqrt{(1+\mu) d_0}}{k} \tag{5.35}$$

and

$$\varepsilon^a_{k,x} + \varepsilon^a_{k,y} \leq \frac{3(3-2\sigma)(1+\mu) d_0}{2(1-\sigma)k}, \tag{5.36}$$

where d_0 and μ are as in (5.9) and (5.22), respectively.

Proof. By combining Theorem 5.2.4, the definition of η_0 in (5.22), and Theorem 2.2.7, we conclude that inequality (5.35) holds, and

$$\varepsilon^a_k \leq \frac{3(3-2\sigma)(1+\mu) d_0}{2(1-\sigma)k}, \tag{5.37}$$

where

$$\varepsilon^a_k = \frac{1}{k} \left[\sum_{i=1}^{k} \Big(\langle r_{i,x}/\beta, \tilde{x}_i - \tilde{x}^a_k \rangle + \langle (H + \beta B^* B)\, r_{i,y}, y_i - y^a_k \rangle + \langle r_{i,\gamma}/(\theta\beta), \tilde{\gamma}_i - \tilde{\gamma}^a_k \rangle \Big) \right] \tag{5.38}$$

On the other hand, (5.12), (5.31) and (5.33) yield

$$A\tilde{x}_k + B y_k = \frac{1}{\theta\beta} r_{k,\gamma} + b, \quad A\tilde{x}^a_k + B y^a_k = \frac{1}{\theta\beta} r^a_{k,\gamma} + b.$$

Additionally, it follows from definitions of $r_{i,\gamma}$ and $r^a_{k,\gamma}$ that

$$\frac{1}{k} \sum_{i=1}^{k} \langle \tilde{\gamma}_i, r_{i,\gamma} - r^a_{k,\gamma} \rangle = \frac{1}{k} \sum_{i=1}^{k} \langle \tilde{\gamma}_i - \tilde{\gamma}^a_k, r_{i,\gamma} - r^a_{k,\gamma} \rangle = \frac{1}{k} \sum_{i=1}^{k} \langle \tilde{\gamma}_i - \tilde{\gamma}^a_k, r_{i,\gamma} \rangle.$$

Hence, combining the identity in (5.38) with the last two equations, we have

$$\begin{aligned}\varepsilon_k^a &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle r_{i,x}/\beta, \tilde{x}_i - \tilde{x}_k^a\rangle + \langle (H+\beta B^*B)\, r_{i,y}, y_i - y_k^a\rangle\Big) + \frac{1}{k}\sum_{i=1}^{k}\big\langle \tilde{\gamma}_i, \big(r_{i,\gamma} - r_{k,\gamma}^a\big)/(\theta\beta)\big\rangle \\ &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle r_{i,x}/\beta, \tilde{x}_i - \tilde{x}_k^a\rangle + \langle (H+\beta B^*B)\, r_{i,y}, y_i - y_k^a\rangle + \langle \tilde{\gamma}_i, A\tilde{x}_i - A\tilde{x}_k^a + By_i - By_k^a\rangle\Big) \\ &= \frac{1}{k}\sum_{i=1}^{k}\langle r_{i,x}/\beta + A^*\tilde{\gamma}_i, \tilde{x}_i - \tilde{x}_k^a\rangle + \frac{1}{k}\sum_{i=1}^{k}\langle (H+\beta B^*B)\, r_{i,y} + B^*\tilde{\gamma}_i, y_i - y_k^a\rangle = \varepsilon_{k,x}^a + \varepsilon_{k,y}^a,\end{aligned}$$

where the last equality is due to the definitions of $\varepsilon_{k,x}^a$ and $\varepsilon_{k,y}^a$ in (5.32). Therefore, the inequality in (5.36) follows trivially from the last equality and (5.37).

To finish the proof of the theorem, note that direct use of Proposition 2.1.1(b) (for f and g), (5.30)–(5.33) give $\varepsilon_{k,x}^a$, $\varepsilon_{k,y}^a \geq 0$ and the inclusion in (5.34). ■

Remark 5.2.8 For a given tolerance $\bar{\rho} > 0$, Theorem 5.2.7 ensures that in at most $\mathcal{O}(1/\bar{\rho})$ iterations, Algorithm 3 provides, in the ergodic sense, an approximate solution $(\bar{x}, \bar{y}, \bar{\gamma})$ of the Lagrangian system (1.7) together with residues $\bar{r} := (r_{\bar{x}}, r_{\bar{y}}, r_{\bar{\gamma}})$ and $(\varepsilon_{\bar{x}}, \varepsilon_{\bar{y}})$ such that

$$\frac{1}{\beta} r_{\bar{x}} \in \partial_{\varepsilon_{\bar{x}}} f(\bar{x}) - A^*\bar{\gamma}, \quad (H+\beta B^*B) r_{\bar{y}} \in \partial_{\varepsilon_{\bar{y}}} g(\bar{y}) - B^*\bar{\gamma}, \quad \frac{1}{\beta\theta} r_{\bar{\gamma}} = A\bar{x} + B\bar{y} - b,$$

and $\max\{\|(r_{\bar{x}}, r_{\bar{y}}, r_{\bar{\gamma}})\|_M, \varepsilon_{\bar{x}}, \varepsilon_{\bar{y}}\} \leq \bar{\rho}$, where M is as in (5.8). For a given tolerance $\rho > 0$, the above relations are equivalent to (1.9) with $(v_{\bar{x}}, v_{\bar{y}}, v_{\bar{\gamma}}) := M\bar{r}$, $\bar{\rho} := \rho/\sqrt{\lambda_M}$, where λ_M is the largest eigenvalue of M, and the fact that $\|M(\cdot)\| \leq \sqrt{\lambda_M}\|\cdot\|_M$. Hence, Algorithm 3 provides a relaxed ρ-approximate solution of (1.7) in at most $\mathcal{O}(1/\rho)$ iterations. The above ergodic complexity bound is better than the pointwise one by a factor of $\mathcal{O}(1/\rho)$; however, the above inclusion is, in general, weaker than that of the pointwise case due to the ε-subdifferentials of f and g instead of subdifferentials.

Chapter 6

Numerical experiments

In this chapter we report some numerical experiments to illustrate the performance of the ADMM variants analyzed in Chapters 3, 4, and 5. All experiments were performed on MATLAB R2015a using an Intel(R) Core i7 2.4GHz computer with 8GB of RAM.

We considered two classes of problems, namely, LASSO and ℓ_1-regularized logistic regression. We are more interested in showing the efficiency of the proposed inexact ADMM variants. For this, we considered some randomly generated problems and we also collected non-simulated data sets, namely, six biomedical data sets from the Elvira biomedical repository [16] representing different types of cancer and one artificial "Madelon" data set from the ICU Machine Learning Repository [22]. Each one of them is associated with a matrix $D \in \Re^{m\times n}$ and a vector $d \in \Re^m$ and are listed in more detail in Table 6.1 below.

Table 6.1: List of non-simulated data sets

Data sets	m	n
Colon tumor gene expression [4]	62	2000
Central nervous system (CNS) [63]	60	7129
Leukemia cancer-ALLMLL [38]	38	7129
Lung cancer-Michigan [8]	96	7129
Lymphoma-Harvard [69]	77	7129
Prostate cancer [70]	102	12600
Madelon [44]	2000	500

6.1 Strategies

In this section, we define the initial parameters and the strategies used to present some comparisons among the considered ADMM variants. Initially, it is important to note that in all our implementations and for all algorithms, we set the initial point $(x_0, y_0, \gamma_0) = (0, 0, 0)$, and the penalty parameter $\beta = 1$. In the following, we specify some details regarding the implementation of each tested algorithms:

Algorithm 1: In our implementation of Algorithm 1, we chose different values of α, namely $\alpha \in \{1.0, 1.3, 1.5, 1.7, 1.9\}$. We set $(G, H) = (0, 0)$, and used the following condition as a stopping criterion

$$\|M(z_k - z_{k-1})\|_\infty \leq 10^{-4}, \tag{6.1}$$

where $z_k := (x_k, y_k, \gamma_k)$ is the sequence generated by Algorithm 1 and M is as in (3.4).

Algorithm 2: We report the numerical performance of Algorithm 2 to solve the two classes of problems, LASSO and ℓ_1-regularized logistic regression.

Different values of the relaxation parameter α were considered in order to illustrate its effect and show that, similarly to the exact generalized ADMM, the performance of the algorithm improves considerably when $\alpha > 1$, specially $\alpha \approx 1.9$. Algorithm 2 was compared with its "exact" version, namely, the generalized ADMM considered in Chapter 3. The latter method corresponds to Algorithm 1 with $(G, H) = (0, 0)$ and x_k being such that there exists a residue v_k satisfying

$$v_k \in \partial f(x_k) - A^* \left[\gamma_{k-1} + \beta \left(Ax_k + By_{k-1} - b\right)\right], \qquad \|v_k\| \leq 10^{-8}.$$

Note that the above inclusion with $v_k = 0$ is the one derived from the first-order optimality condition for (3.1) with $G = 0$. It should be mentioned that the applications considered here are such that the solution of the second subproblem of the three analyzed algorithms can be explicitly computed.

For the first test problem, the algorithms were tested using six non-simulated data sets reported in Table 6.1. In addition, for the second class of problems, we select all data sets from Table 6.1.

For all tests, we used the same overall termination condition (6.1), with M and z_k given in (4.7) and (4.9), respectively. In Algorithm 2, the remaining initialization data were $\tau_1 = 0.99(2 - \alpha)$, $\tau_2 = 1 - 10^{-8}$ and $H = 0$, and a hybrid inner stopping criterion was used; specifically, the inner-loop terminates when v_k satisfies either the inequality in (4.1) or $\|v_k\| \leq 10^{-8}$. The latter strategy was also used in [29, 30, 80] and it is motivated by the fact that, close to a solution, the former condition seems to be more restrictive than the latter.

Algorithm 3: We also report some numerical tests to illustrate the performance of Algorithm 3 in the two classes of problems, LASSO and ℓ_1-regularized logistic regression. Our main goal is to show that, in some applications, the method performs better with a stepsize parameter $\theta > 1$ instead of the choice $\theta = 1$ as considered in the related literature. Similarly to the strategy use in Algorithm 2, we also used a hybrid inner stopping criterion for Algorithm 3, i.e., the inner-loop terminates when v_k satisfies either the inequality in (5.2) or $\|v_k\| \leq 10^{-8}$. We set $\tau_1 = 0.99(1+\theta-\theta^2)/(\theta(2-\theta))$, $\tau_2 = 1 - 10^{-8}$ and $H = 0$. For a comparison purpose, we also run [29, Algorithm 2], denoted here by relerr-ADMM; see Remark 5.1.1(d) for more details on the relationship between Algorithm 3 and the relerr-ADMM. As suggested in [29], the error tolerance parameter τ_1 in (5.7) was taken equal to 0.99. For all tests, both algorithms stopped when the condition (6.1) was satisfied, where M is as in (5.8) and $z_k := (x_k, y_k, \gamma_k)$ is the sequence generated by the respective algorithms.

6.2 LASSO problem

We consider the following LASSO problem [77, 78]

$$\min_{x \in \Re^n} \frac{1}{2}\|Dx - d\|^2 + \mu\|x\|_1, \tag{6.2}$$

where $D \in \Re^{m\times n}$, $d \in \Re^m$, $\mu > 0$ is a regularization parameter, and $\|\cdot\|_1$ denotes the ℓ_1-norm. In our experiment, we scaled d and the columns of D in order to have unit ℓ_2-norm. The regularization parameter μ was set equal to $0.1\|D^*d\|_\infty$, where $\|\cdot\|_\infty$ denotes the maximum norm. By introducing a new variable, the above problem is usually rewritten as

$$\min\left\{\frac{1}{2}\|Dx - d\|^2 + \mu\|y\|_1 : \; y - x = 0, \; x \in \Re^n, y \in \Re^n\right\}. \tag{6.3}$$

Obviously, (6.3) is an instance of (1.1) with

$$f(x) = \frac{1}{2}\|Dx - d\|^2, \quad g(y) = \mu\|y\|_1, \quad A = -I, \quad B = I \quad \text{and} \quad b = 0. \tag{6.4}$$

First, we verify the performance of Algorithm 1, for solving problem (6.2). Note that, with the specifications in (6.4), the subproblems (3.1) and (3.2) have closed-form solutions

$$x_k = (D^*D + \beta I)^{-1}(D^*d + \beta y_{k-1} - \gamma_{k-1}), \quad y_k = \mathcal{S}_{\frac{\mu}{\beta}}\left(\alpha x_k + (1-\alpha)y_{k-1} + \frac{1}{\beta}\gamma_{k-1}\right),$$

where, for a scalar $\kappa > 0$, $\mathcal{S}_\kappa : \Re^n \to \Re^n$ is the shrinkage operator [7] defined as

$$\mathcal{S}_\kappa^i(w) = \text{sign}(w^i)\max(0, |w^i| - \kappa) \quad i = 1, 2, \ldots, n, \tag{6.5}$$

with sign$(\cdot)$ denotes the sign function. In our experiments of Algorithm 1, the matrix D was randomly generated and the vector $d \in \Re^m$ was chosen as $d = Dx + \sqrt{0.001}y$, where the $(100/n)$-sparse vector $x \in \Re^n$ and the noisy vector $y \in \Re^m$ were also randomly generated.

Table 6.2: Performance of Algorithm 1 to solve three randomly generated LASSO problems

Dim. of D	$\alpha = 1.0$		$\alpha = 1.3$		$\alpha = 1.5$		$\alpha = 1.7$		$\alpha = 1.9$	
$m \times n$	Iter	Time	Iter	Time	Iter	Time	Iter	Time	Iter	Time
900×3000	27	7.2	21	5.6	19	5.0	19	5.0	47	12.6
1200×4000	26	14.8	23	13.1	21	12.5	20	12.4	49	32.1
1500×5000	26	26.6	21	24.5	20	27.1	20	24.0	46	58.1

The performance of Algorithm 1 to solve the three randomly generated LASSO problem instances is reported in Table 6.2, in which "Iter" and "Time" denote the number of iterations and the CPU time in seconds, respectively. From this table, we can see that, in all considered instances of (6.3), Algorithm 1 with $\alpha \in \{1.3, 1.5, 1.7\}$ performed better than Algorithm 1 with $\alpha \in \{1, 1.9\}$. Moreover, Algorithm 1 with $\alpha = 1.7$ presented the best performance. Therefore, we can conclude that Algorithm 1 with a suitable relaxation factor $\alpha > 1$ outperformed the standard ADMM (which corresponds to Algorithm 1 with $\alpha = 1$) in our numerical experiments.

We also tested Algorithm 2 for the problem (6.2). In view of (6.4), the pair $(\tilde{x}_k, v_k)$ in (4.1) can be obtained by computing an approximate solution $\tilde{x}_k$ with a residual v_k of the following linear system

$$(D^*D + \beta I)x = (D^*d + \beta y_{k-1} - \gamma_{k-1}). \tag{6.6}$$

For approximately solving the above linear system, we used the conjugate gradient method [60] with starting point $D^*d + \beta y_{k-1} - \gamma_{k-1}$. Similarly to the previous case, the subproblem (4.3) has a closed-form solution

$$y_k = \mathcal{S}_{\frac{\mu}{\beta}}\left(\alpha\tilde{x}_k + (1-\alpha)y_{k-1} + \frac{1}{\beta}\gamma_{k-1}\right),$$

where $\mathcal{S}$ is as in (6.5).

Table 6.3: Performance of Algorithms 1 and 2 for six instances of the LASSO problem

	$\alpha = 1.0$		$\alpha = 1.3$		$\alpha = 1.5$		$\alpha = 1.7$		$\alpha = 1.9$	
Data set	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2
	Number of outer iterations									
Colon	114	116	89	88	77	78	69	69	63	63
CNS	321	319	249	248	217	217	194	194	182	182
Leukemia	600	600	431	431	370	370	330	329	320	320
Lung	535	535	412	412	357	357	315	315	282	282
Lymphoma	331	331	255	255	222	222	196	196	176	176
Prostate	430	431	331	331	287	287	254	254	227	227
	Total number of inner iterations									
Colon	4656	2136	3639	1607	3149	1450	2822	1308	2576	1216
CNS	16064	10060	12466	7818	10862	6871	9712	6203	9108	6024
Leukemia	17365	11351	12478	8033	10715	6909	9556	6196	9263	6195
Lung	22836	12516	17588	9622	15240	8373	13451	7475	12048	6881
Lymphoma	15182	8619	11703	6522	10180	5850	8998	5208	8072	4796
Prostate	35002	19562	26944	15083	23374	13088	20700	11906	18478	11003
	CPU time in seconds									
Colon	23.3	16.4	18.2	12.3	17.0	10.9	14.4	9.7	13.1	9.2
CNS	944.4	754.4	743.4	584.6	643.1	515.6	576.7	472.9	538.7	449.0
Leukemia	1290.4	1119.2	927.8	789.0	797.0	679.4	710.5	606.1	689.4	600.4
Lung	1470.9	1114.7	1159.5	872.3	998.5	762.5	880.5	670.9	788.8	607.6
Lymphoma	931.0	769.7	728.1	601.8	634.6	489.0	564.1	433.3	504.1	393.1
Prostate	5926.5	4325.1	4494.2	3509.2	3900.2	3083.7	3438.1	2664.4	3103.0	2343.7

Table 6.3 displays the numerical results obtained. In order to compare the algorithms, we consider the number of outer iterations, the total number of accumulated inner iterations and the CPU time in seconds. In Figure 6.1, we plot the arithmetic mean of the latter three comparisons criteria for each algorithm for solving the six LASSO problem instances. From these results, one can see that the number of outer iterations of Algorithm 2 and Algorithm 1 are basically the same for every considered relaxation parameter α. In particular,

the numerical advantage of using $\alpha > 1$, specially $\alpha \approx 1.9$, is also verified for Algorithm 2. Algorithm 2 performed at least 33% less inner iterations than Algorithm 1, reaching, in some instances, 50% less inner iterations. Note that this performance improvement also reflected favorably in terms of CPU time.

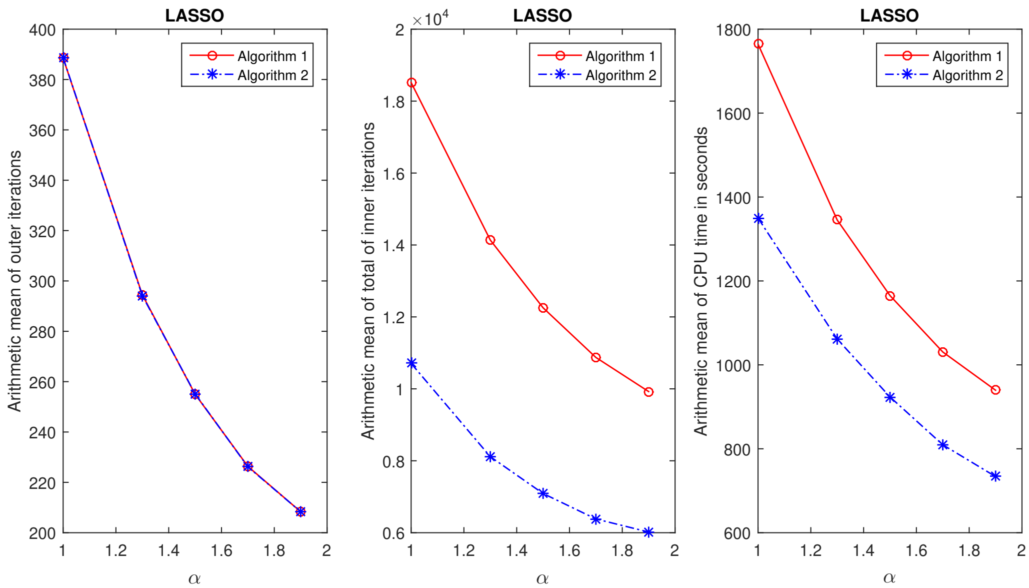

Figure 6.1: Arithmetic mean of the LASSO problem results given in Table 6.3

Now, let us discuss the performance of Algorithm 3 for approximately solving problem (6.2). In this case, the pair $(\tilde{x}_k, v_k)$ in (5.2) was obtained using the same strategy as in Algorithm 2, i.e., we applied the conjugate gradient method [60] with starting point $D^*d + \beta y_{k-1} - \gamma_{k-1}$ in order to obtain an approximate solution $\tilde{x}_k$ with residual v_k of the linear system (6.6). Note that subproblem (5.4) also has a closed-form solution

$$y_k = \mathcal{S}_{\frac{\mu}{\beta}}\left(\tilde{x}_k + \frac{1}{\beta}\gamma_{k-1}\right),$$

where $\mathcal{S}$ is the shrinkage operator defined in (6.5).

We tested the relerr-ADMM and Algorithm 3 for solving 3 randomly generated LASSO problem instances. For a given dimension $m \times n$, we generated a random matrix D and choose vector $d \in \Re^m$ as $d = Dx + \sqrt{0.001}y$, where the $(100/n)-$sparse vector $x \in \Re^n$ and the noisy vector $y \in \Re^m$ were also generated randomly. We also tested the relerr-ADMM and Algorithm 3 on six standard cancer data sets given in Table 6.1. Their performances are listed in Tables 6.4 and 6.5, in which "Out" and "Inner" denote the number of iterations and the total number of inner iterations of the methods, respectively, whereas "Time" is the

CPU time in seconds. From these tables, we see that the relerr-ADMM and Algorithm 3 with $\theta = 1$ presented similar performances. However, Algorithm 3 with $\theta = 1.3$ and $\theta = 1.6$ clearly outperformed the relerr-ADMM.

Table 6.4: Performance of the relerr-ADMM and Algorithm 3 to solve three randomly generated LASSO problems

Dim. of D	relerr-ADMM			Alg. 3 ($\theta = 1$)			Alg. 3 ($\theta = 1.3$)			Alg. 3 ($\theta = 1.6$)		
$m \times n$	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time
900×3000	27	206	12.3	27	206	11.9	23	183	10.4	21	202	9.6
1200×4000	27	207	26.2	27	207	25.6	24	191	22.2	21	197	19.9
1500×5000	25	186	42.2	25	186	42.2	22	169	39.1	20	190	35.8

Table 6.5: Performance of the relerr-ADMM and Algorithm 3 for six instances of the LASSO problem

Data set	relerr-ADMM			Alg. 3 ($\theta = 1$)			Alg. 3 ($\theta = 1.3$)			Alg. 3 ($\theta = 1.6$)		
	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time
Colon	116	2298	18.3	116	2136	17.4	107	1977	16.0	99	1990	15.3
CNS	319	10077	823.5	319	10060	793.4	315	10292	817.1	312	11029	831.1
Leukemia	600	11390	1216.5	600	11351	1172.6	427	7948	845.5	362	7068	741.3
Lung	535	12499	1321.4	535	12516	1218.4	404	9332	924.8	338	8426	777.6
Lymphoma	331	8737	769.2	331	8619	765.0	264	6901	610.3	216	6038	521.4
Prostate	430	19400	4559.3	431	19562	4303.1	358	16465	3592.9	328	16989	3536.1

Figures 6.2, 6.3, and 6.4 summarize the results presented in Tables 6.3 and 6.5 for the following inexact versions: Algorithm 2 with $\alpha = 1.3, 1.5, 1.7, 1.9$, relerr-ADMM and Algorithm 3 with $\theta = 1.3, 1.6$. We omit the results related to Algorithm 2 with $\alpha = 1.0$ and Algorithm 3 with $\theta = 1.0$, because they are identical and, basically, the same as those of the relerr-ADMM. In these figures we can easily verify the superiority of Algorithm 2, especially with $\alpha = 1.9$.

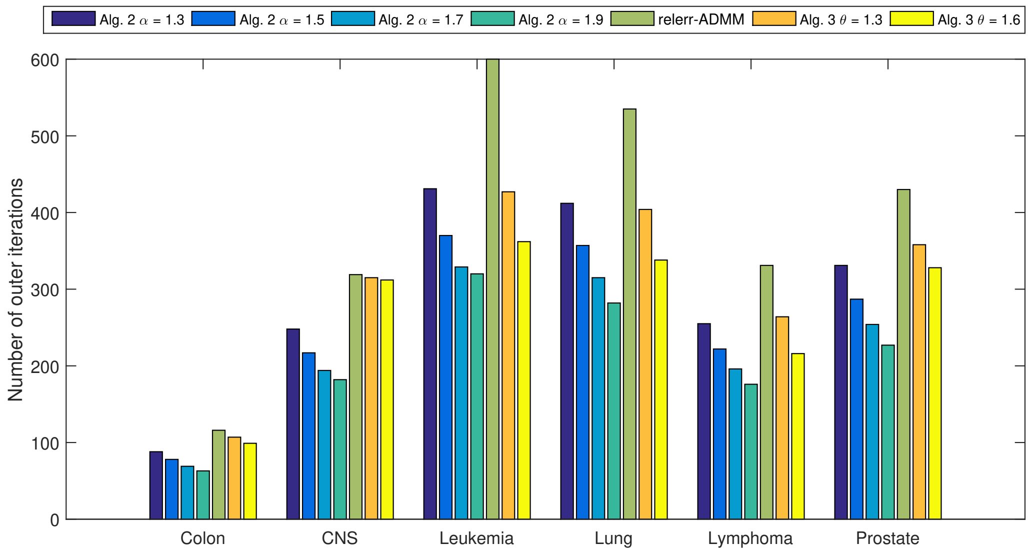

Figure 6.2: LASSO problem: number of outer iterations

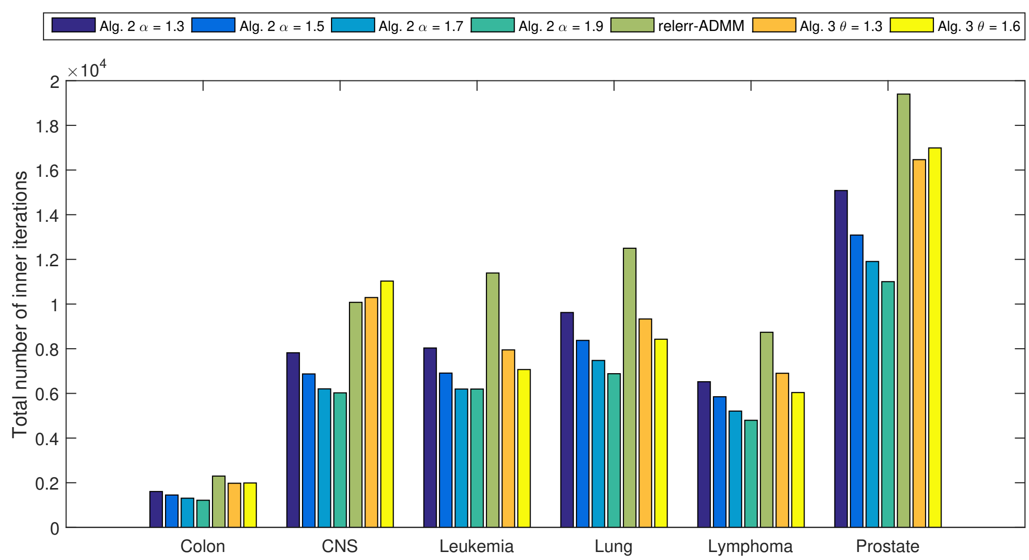

Figure 6.3: LASSO problem: total number of inner iterations

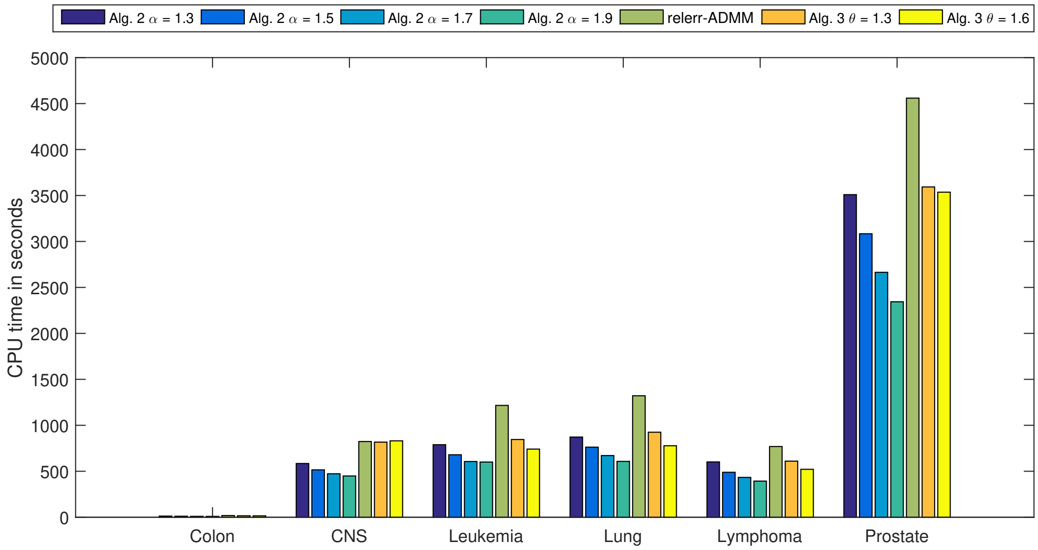

Figure 6.4: LASSO problem: CPU time in seconds

6.3 ℓ_1-Regularized logistic regression problem

Consider the ℓ_1-regularized logistic regression problem [51]

$$\min_{t\in\Re, u\in\Re^n} \frac{1}{m}\sum_{i=1}^{m} \log\Big(1+\exp\big(-d^i\left(\langle D_i,u\rangle + t\right)\big)\Big) + \mu\,\|u\|_1\,, \tag{6.7}$$

where $D_i \in \Re^n$ are the rows of a matrix $D \in \Re^{m\times n}$, $d^i \in \{-1,+1\}$ are the coordinates of a vector $d \in \Re^m$ and $\mu > 0$ is a regularization parameter. In our experiment, the matrix D and the vector d were chosen as described in the beginning of this chapter (see Table 6.1). We scaled the columns of D in order to have unit ℓ_2-norm and set $\mu = 0.5\lambda_{\max}$, where $\lambda_{\max}$ is as defined in [51, Subsection 2.1].

By defining $z^{i:j} := (z^i,\ldots,z^j) \in \Re^{j-i+1}$ for $j \geq i$, problem (6.7) can be rewritten as an instance of (1.1) in which

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \log\Big(1+\exp\big(-d^i\left(\left\langle D_i, x^{2:n+1}\right\rangle + x^1\right)\big)\Big), \qquad g(y) = \mu\left\|y^{2:n+1}\right\|_1, \tag{6.8}$$
$$A = -I, \qquad B = I, \qquad \text{and} \qquad b = 0.$$

First we apply Algorithm 2 to solve problem (6.7). In order to compute a pair $(\tilde{x}_k, v_k)$ as in (4.1), we implemented the limited-memory BFGS method [60, Algorithm 7.5] with starting point equal to $(0,\ldots,0)$. The subproblem (4.3) has a closed-form solution $y_k := (y_k^1, y_k^{2:n+1})$ given by

$$y_k^1 = \alpha\tilde{x}_k^1 + (1-\alpha)y_{k-1}^1 + \frac{1}{\beta}\gamma_{k-1}^1, \quad y_k^{2:n+1} = \mathcal{S}_{\frac{\mu}{\beta}}\left(\alpha\tilde{x}_k^{2:n+1} + (1-\alpha)y_{k-1}^{2:n+1} + \frac{1}{\beta}\gamma_{k-1}^{2:n+1}\right),$$

where $\mathcal{S}$ is the shrinkage operator as defined in (6.5).

Table 6.6 displays the numerical results obtained. As in Subsection 6.2, the methods were compared in terms of the number of outer iterations, the total number of inner iterations and the CPU time in seconds. In Figure 6.5, we plot the arithmetic mean of the latter three comparison criteria for each method for solving the seven ℓ_1-regularized logistic regression problem instances. By analyzing Table 6.6 and Figure 6.5, one can see that Algorithm 2 performed, basically, the same number of outer iterations than Algorithm 1. Regarding the total number of inner iterations, Algorithm 2 performed at least 41% less than Algorithm 1, reaching, in some instances, 60% less inner iterations. Note that the saving with respect to CPU times was very expressive. Specifically, Algorithm 2 was at least 48% faster than Algorithm 1. The reason lies in the difficulty to solve (3.1) for the ℓ_1-regularized logistic regression problem.

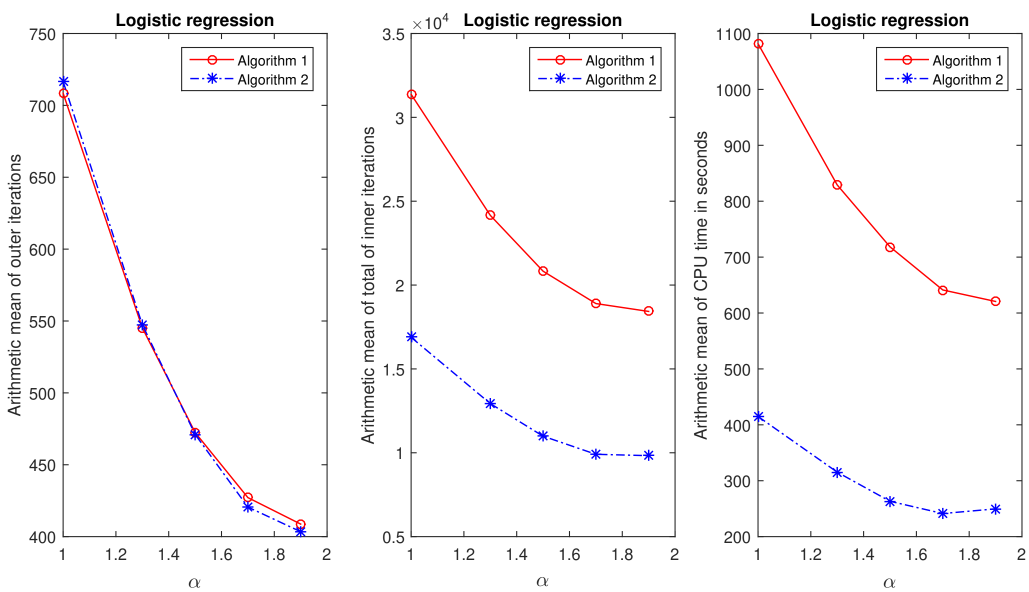

Figure 6.5: Arithmetic mean of the ℓ_1-regularized logistic regression problem results given in Table 6.6

We also tested Algorithm 3 applied for solving seven ℓ_1-regularized logistic regression problem (6.7) using the data sets given in Table 6.1. The pair $(\tilde{x}_k, v_k)$ in (5.2) also was

Table 6.6: Performance of Algorithms 1 and 2 for seven instances of the ℓ_1-regularized logistic regression problem

	$\alpha = 1.0$		$\alpha = 1.3$		$\alpha = 1.5$		$\alpha = 1.7$		$\alpha = 1.9$	
Data set	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2
	Number of outer iterations									
Colon	337	370	259	253	224	216	197	196	176	175
CNS	278	278	213	216	185	186	163	163	145	144
Leukemia	624	625	480	481	416	416	367	367	328	328
Lung	513	551	400	435	347	375	380	378	528	548
Lymphoma	375	375	287	289	248	251	219	223	195	197
Prostate	879	882	676	678	585	585	516	512	462	457
Madelon	1953	1935	1502	1480	1302	1269	1148	1105	1027	975
	Total number of inner iterations									
Colon	18645	9912	14460	7033	12334	5784	10949	5515	9688	4883
CNS	15515	8758	11881	6781	10259	5969	9068	5086	8077	4528
Leukemia	27859	15402	21486	11763	18560	10354	16271	8951	14538	7925
Lung	28487	15744	22329	13005	18813	10642	20320	10559	28931	16208
Lymphoma	21638	11191	16485	8666	14248	7443	12590	6546	11228	5826
Prostate	68770	37327	52865	28419	45705	24842	40480	22902	36160	21267
Madelon	38698	19857	29584	14859	25871	11898	22601	9806	20371	8159
	CPU time in seconds									
Colon	48.3	21.8	37.4	13.5	31.8	10.3	28.3	9.8	24.7	8.7
CNS	302.0	107.9	232.2	88.9	199.0	79.0	177.4	68.9	159.2	61.5
Leukemia	417.1	168.6	337.4	131.8	279.7	110.1	243.0	93.4	215.8	91.4
Lung	844.1	352.2	638.4	292.5	539.1	239.8	572.5	242.9	822.8	363.72
Lymphoma	527.5	190.0	402.5	156.3	351.9	134.3	308.7	121.8	276.0	108.1
Prostate	3844.6	1246.5	2950.4	918.9	2562.0	807.8	2271.1	761.8	2036.8	782.1
Madelon	1589.2	817.6	1205.2	605.6	1065.1	461.3	887.6	390.6	809.4	332.2

obtained with the aid of the limited-memory BFGS method [60, Algorithm 7.5], being the starting point the origin. Again, the subproblem (5.4) has a closed-form solution $y_k =$

$(y_k^1, y_k^{2:n+1})$ given by

$$y_k^1 = \tilde{x}_k^1 + \frac{1}{\beta}\gamma_{k-1}^1, \qquad y_k^{2:n+1} = \mathcal{S}_{\frac{\mu}{\beta}}\left(\tilde{x}_k^{2:n+1} + \frac{1}{\beta}\gamma_{k-1}^{2:n+1}\right),$$

where $\mathcal{S}$ is the shrinkage operator given in (6.5).

Tables 6.7 reports the performances of the relerr-ADMM and Algorithm 3 for solving the aforementioned seven instances of the problem (6.7). In Table 6.7, "Out" and "Inner" are the number of iterations and the total of inner iterations of the methods, respectively, whereas "Time" is the CPU time in seconds. Similarly to the numerical results of Section 6.2, we observe that the relerr-ADMM and Algorithm 3 with $\theta = 1$ had similar performances, whereas Algorithm 3 with $\theta = 1.3$ and $\theta = 1.6$ outperformed the relerr-ADMM. Therefore, the efficiency of the inexact proximal ADMM for solving real-life applications is illustrated.

Table 6.7: Performance of the relerr-ADMM and Algorithm 3 for seven instances of the ℓ_1-regularized logistic regression problem

Data set	relerr-ADMM			Alg. 3 ($\theta = 1$)			Alg. 3 ($\theta = 1.3$)			Alg. 3 ($\theta = 1.6$)		
	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time	Out	Inner	Time
Colon	335	11621	26.0	370	9912	22.7	276	7694	15.6	234	6903	13.9
CNS	278	10116	172.5	278	8758	151.4	245	7836	135.2	229	7286	123.8
Leukemia	624	17788	237.5	625	15402	221.7	601	14825	211.5	592	14987	201.7
Lung	519	19715	568.1	551	15744	428.8	539	16235	482.5	547	15948	442.1
Lymphoma	374	14358	324.8	375	11191	226.6	356	10773	228.4	353	10811	237.5
Prostate	879	41145	1720.1	882	37327	1463.9	688	29367	1183.7	560	28239	1384.3
Madelon	1957	22830	890.7	1935	19857	923.8	1938	19790	929.8	1961	26553	1131.3

Figures 6.6, 6.7, and 6.8 were constructed with the numerical values contained in Tables 6.6 and 6.7 of the following inexact methods: Algorithm 2 with $\alpha = 1.3, 1.5, 1.7, 1.9$, relerr-ADMM and Algorithm 3 with $\theta = 1.3, 1.6$. It can be easily seen that, in most tests, Algorithm 2, especially with $\alpha = 1.9$, obtained the best numerical performance.

We end this section by making some remarks. First, Algorithm 3 was tested with other values of θ different from the ones presented in tables 6.4, 6.5 and 6.7, and we observed the following: (i) if $\theta \in [0.1, 1.6]$, then the performance of Algorithm 3 improved as θ was increased; (ii) if $\theta \in (1.6, (\sqrt{5}+1)/2)$, then Algorithm 3 performed similarly to its exact version, since the relative error condition (5.2) became stringent. Second, the classical proximal gradient method and its accelerated versions such as FISTA can also be applied to solve LASSO and ℓ_1-regularized logistic regression problems. Numerical comparisons showing that the relerr-ADMM is competitive with FISTA for solving the aforementioned problems were reported in [29]. Therefore, since Algorithm 3 performed better than the relerr-ADMM for these applications, we can conclude that Algorithm 3 is also competitive with FISTA.

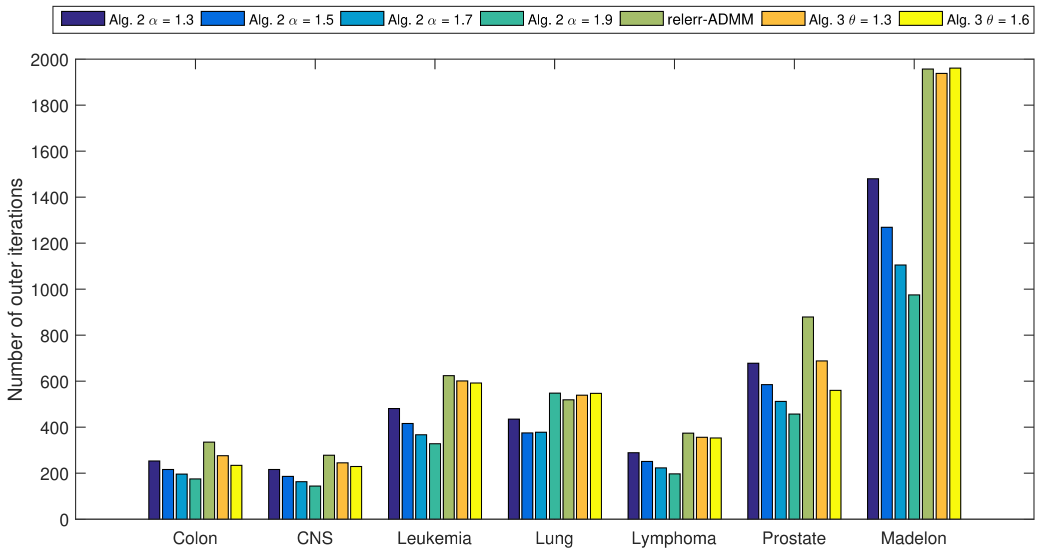

Figure 6.6: ℓ_1-Regularized logistic regression problem: number of outer iterations

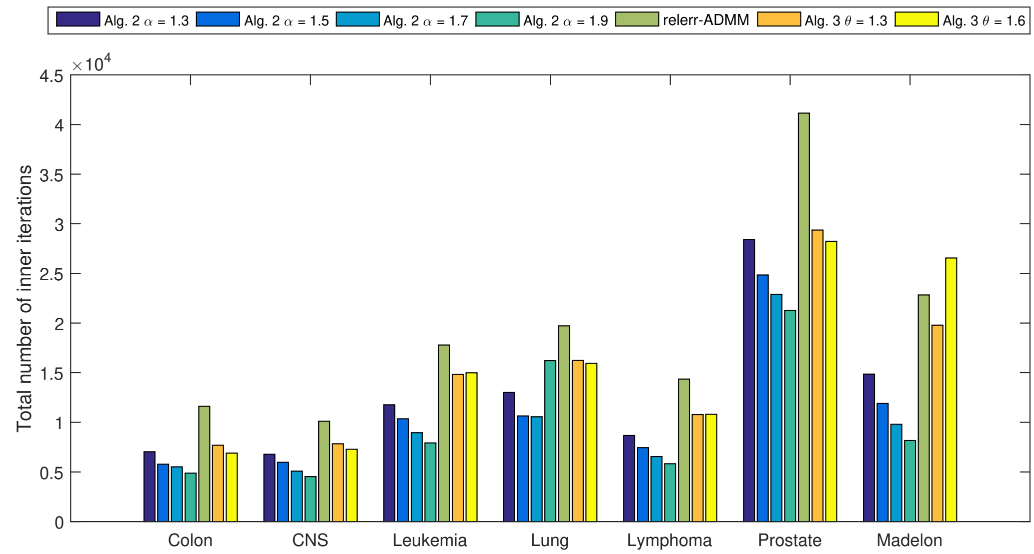

Figure 6.7: ℓ_1-Regularized logistic regression problem: total number of inner iterations

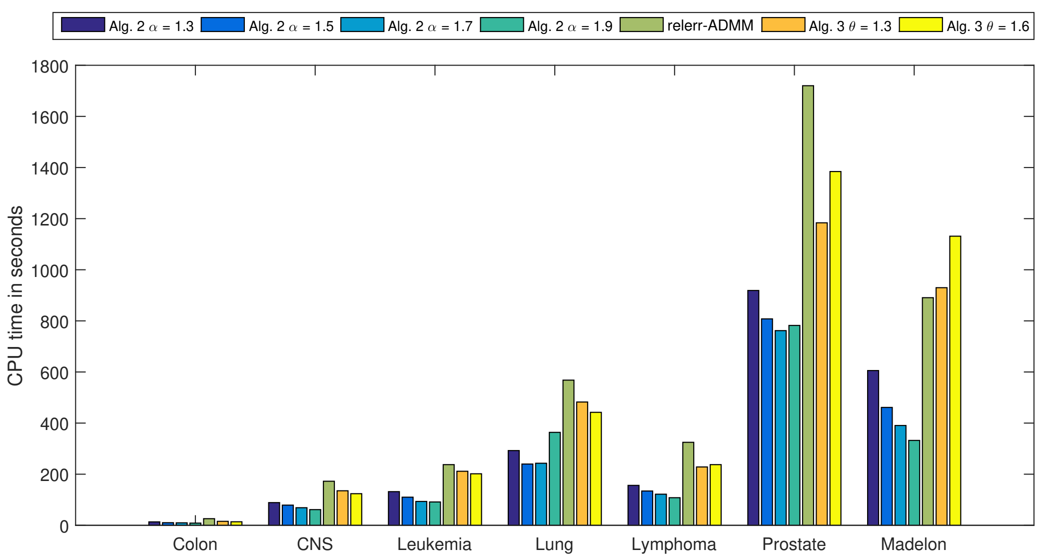

Figure 6.8: ℓ_1-Regularized logistic regression problem: CPU time in seconds

Chapter 7

Final remarks

In this thesis, we proposed and analyzed some variants of the alternating direction method of multipliers (ADMM) for computing approximate solutions of linearly constrained convex optimization problems. Initially, we studied iteration-complexity results for a proximal generalized ADMM. Specifically, for a given tolerance $\rho > 0$, we established $\mathcal{O}(1/\rho^2)$ pointwise and $\mathcal{O}(1/\rho)$ ergodic iteration-complexity bounds for the proximal generalized ADMM to obtain an approximate solution of the Lagrangian system associated to the aforementioned optimization problem. We also proposed and analyzed two inexact variants of the (generalized) proximal ADMM. These variants are such that their first partial subproblems are approximately solved using relative error conditions based on the works of Solodov and Svaiter [71–74]. It was shown that from a theoretical view point, the proposed inexact schemes have pointwise and ergodic iteration-complexity bounds similar to their exact versions, whereas from a computational viewpoint the proposed schemes are relatively cheaper and more efficient. Our analysis is essentially based on showing that these considered schemes can be seen as special instances of a hybrid proximal extragradient framework for solving monotone inclusion problems. Some numerical experiments were carried out in order to illustrate the numerical behavior of the methods. They confirm that appropriately chosen parameters can improve the performance of the methods and indicate that the proposed inexact versions represents an useful tool for solving some real-life applications that can be formulated as linearly constrained convex optimization problems. Finally, a possible direction for future research would be to analyze inexact variants of the regularized ADMMs due to their improved iteration-complexity bounds. This would be interesting also to improve the applicability of these methods. Another direction, would be to explore the proximal terms of the inexact proximal ADMM in order to enlarge the region in which one can choose the relaxation parameter included in the Lagrange multipliers update rule.

Bibliography

[1] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. An inexact proximal generalized alternating direction method of multipliers. *Submitted to Comput. Optim. Appl.*, 2018.

[2] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. Iteration-complexity analysis of a generalized alternating direction method of multipliers. *J. Glob. Optim.*, 73(2):331–348, 2019.

[3] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. A partially inexact proximal alternating direction method of multipliers and its iteration-complexity analysis. *J. Optim. Theory Appl. doi:10.1007/s10957-019-01525-8*, 2019.

[4] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.*, 96(12):6745–6750, 1999.

[5] H. Attouch and M. Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. *Pac. J. Optim.*, 5(1):17–37, 2008.

[6] A. Beck. *First-Order Methods in Optimization.* SIAM, Philadelphia, 2017.

[7] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[8] D. G. Beer, S. L. R. Kardia, C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.*, 8(8):816, 2002.

[9] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York, 1982.

[10] S. Bitterlich, R. I. Boţ, E. R. Csetnek, and G. Wanka. The Proximal Alternating Minimization Algorithm for two-block separable convex optimization problems with linear constraints. *arXiv preprint arXiv:1806.00260*, 2018.

[11] R. I. Boţ and E. R. Csetnek. ADMM for monotone operators: convergence analysis and rates. *Adv. Computat. Math.*, 2018.

[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[13] K. Bredies and H. Sun. A proximal point analysis of the preconditioned alternating direction method of multipliers. *J. Optim. Theory Appl.*, 173(3):878–907, 2017.

[14] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[15] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter. ϵ-enlargements of maximal monotone operators: theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 25–43. Kluwer Acad. Publ., Dordrecht, 1999.

[16] A. Cano, A. Masegosa, and S. Moral. ELVIRA biomedical data set repository, 2005.

[17] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, 2011.

[18] C. Chen, B. He, and X. Yuan. Matrix completion via an alternating direction method. *IMA J. of Numer. Anal.*, 32(1):227–245, 06 2011.

[19] E. Corman and X. Yuan. A generalized proximal point algorithm and its convergence rate. *SIAM J. Optim.*, 24(4):1614–1638, 2014.

[20] Y. Cui, X. Li, D. Sun, and K. C. Toh. On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. *J. Optim. Theory Appl.*, 169(3):1013–1041, 2016.

[21] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, 66(3):889–916, 2016.

[22] D. Dheeru and E. K. Taniskidou. UCI machine learning repository, 2018.

[23] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

[24] J. Eckstein. Parallel alternating direction multiplier decomposition of convex programs. *J. Optim. Theory Appl.*, 80(1):39–62, 1994.

[25] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optim. Method Softw.*, 4(1):75–83, 1994.

[26] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3, Ser. A):293–318, 1992.

[27] J. Eckstein and P. J. S. Silva. A practical relative error criterion for augmented Lagrangians. *Math. Program.*, 141(1):319–348, 2013.

[28] J. Eckstein and W. Yao. Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. *Pacific J. Optim.*, 11(4):619–644, 2015.

[29] J. Eckstein and W. Yao. Approximate ADMM algorithms derived from Lagrangian splitting. *Comput. Optim. Appl.*, 68(2):363–405, 2017.

[30] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas–Rachford splitting and special cases of the ADMM. *Math. Program.*, 170(2):417–444, 2018.

[31] E. X. Fang, H. Bingsheng, H. Liu, and Y. Xiaoming. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.*, 7(2):149–187, 2015.

[32] M. Fortin and R. Glowinski. On decomposition-coordination methods using an augmented lagrangian. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and Its Applications*, pages 97 – 146. Elsevier, 1983.

[33] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and Its Applications*, pages 299 – 331. Elsevier, Amsterdam, 1983.

[34] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2(1):17–40, 1976.

[35] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer Series in Computational Physics. Springer-Verlag, Houston, 1984.

[36] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Society for Industrial and Applied Mathematics, Philadelphia, 1989.

[37] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *R.A.I.R.O. Anal. Numér.*, 9(2):41–76, 1975.

[38] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

[39] M. L. N. Gonçalves. On the pointwise iteration-complexity of a dynamic regularized ADMM with over-relaxation stepsize. *Appl. Math. Comput.*, 336:315–325, 2018.

[40] M. L. N. Gonçalves, M. M. Alves, and J. G. Melo. Pointwise and ergodic convergence rates of a variable metric proximal alternating direction method of multipliers. *J. Optim. Theory Appl.*, 177(2):448–478, 2018.

[41] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Extending the ergodic convergence rate of the proximal ADMM. *arXiv preprint arXiv:1611.02903*, 2016.

[42] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. *SIAM J. Optim.*, 27(1):379–407, 2017.

[43] Y. Gu, B. Jiang, and H. Deren. A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. *arXiv preprint arXiv:1506.02221*, 2015.

[44] I. Guyon, S. Gunn, A. B. Hur, and G. Dror. Result analysis of the NIPS 2003 feature selection challenge. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural Inf. Process. Syst.*, pages 545–552. MIT Press, 2005.

[45] W. W. Hager, M. Yashtini, and H. Zhang. An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. *SIAM J. Numer. Anal.*, 54(3):1535–1556, 2016.

[46] B. He, L. Z. Liao, D. Han, and H. Yang. A new inexact alternating directions method for monotone variational inequalities. *Math. Program.*, 92(1):103–118, 2002.

[47] B. He, H. Liu, Z. Wang, and X. Yuan. A strictly contractive peaceman–rachford splitting method for convex programming. *SIAM J. Optim.*, 24(3):1011–1040, 2014.

[48] B. He and X. Yuan. On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.*, 50(2):700–709, 2012.

[49] B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.*, 130(3):567–577, 2015.

[50] Y. He and R. D. C. Monteiro. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM J. Optim.*, 26(1):29–56, 2016.

[51] K. Koh, S. J. Kim, and S. Boyd. An interior-point method for large-scale l_1-regularized logistic regression. *J. Mach. Learn. Res.*, 8:1519–1555, 2007.

[52] T. Lin, S. Ma, and S. Zhang. An extragradient-based alternating direction method for convex minimization. *Found. Comput. Math.*, pages 1–25, 2015.

[53] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[54] M. Marques Alves, R. D. C. Monteiro, and B. F. Svaiter. Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. *SIAM J. Optim.*, 26(4):2730–2743, 2016.

[55] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(R3):154–158, 1970.

[56] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.

[57] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[58] M. Ng, F. Wang, and X. Yuan. Inexact alternating direction methods for image recovery. *SIAM J. Sci. Comput.*, 33(4):1643–1668, 2011.

[59] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan. A general analysis of the convergence of ADMM. *arXiv preprint arXiv:1502.02009*, 2015.

[60] J. Nocedal and S. J. Wright. *Numerical Optimization 2nd*. Springer, New York, 2006.

[61] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao, Jr. An accelerated linearized alternating direction method of multipliers. *SIAM J. Imaging Sci.*, 8(1):644–681, 2015.

[62] N. Parikh and S. Boyd. Proximal algorithms. *Found. Trends Optim.*, 1(3):127–239, 2014.

[63] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415(6870):436–442, 2002.

[64] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[65] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[66] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

[67] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259 – 268, 1992.

[68] R. Shefi and M. Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. Optim.*, 24(1):269–297, 2014.

[69] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.*, 8(1):68–74, 2002.

[70] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2):203–209, 2002.

[71] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[72] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[73] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.

[74] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[75] H. Sun. Analysis of fully preconditioned ADMM with relaxation in Hilbert spaces. *arXiv preprint arXiv:1611.04801*, 2016.

[76] M. Tao and X. Yuan. On the optimal linear convergence rate of a generalized proximal point algorithm. *J. Sci. Comput.*, 74(2):826–850, 2018.

[77] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, 58(1):267–288, 1996.

[78] R. J. Tibshirani. The lasso problem and uniqueness. *Electron. J. Stat.*, 7:1456–1490, 2013.

[79] X. Wang and X. Yuan. The linearized alternating direction method of multipliers for dantzig selector. *SIAM J. Sci. Comput.*, 34(5):2792–2811, 2012.

[80] J. Xie, A. Liao, and X. Yang. An inexact alternating direction method of multipliers with relative error criteria. *Optim. Lett.*, 11(3):583–596, 2017.

[81] M. H. Xu. Proximal alternating directions method for structured variational inequalities. *J. Optim. Theory Appl.*, 134(1):107–117, Jul 2007.

[82] Z. Xu, M. Figueiredo, X. Yuan, C. Studer, and T. Goldstein. Adaptive relaxed ADMM: Convergence theory and practical implementation. *arXiv preprint arXiv:1704.02712*, 2017.

[83] J. Yang and X. Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comput.*, 82(281):301–329, 2013.