Please cite the Published Version

Optimizing ADMM and Over-Relaxed ADMM Parameters for Linear Quadratic Problems

Jintao Song[1,2], **Wenqi Lu**[3,4], **Yunwen Lei**[5], **Yuchao Tang**[6], **Zhenkuan Pan**[2], **Jinming Duan**[1†]

[1] School of Computer Science, University of Birmingham, UK
[2] College of Computer Science and Technology, Qingdao University, China
[3] Department of Computing and Mathematics, Manchester Metropolitan University, UK
[4] Centre for Computational Science and Mathematical Modelling, Coventry University, UK
[5] Department of Mathematics, University of Hong Kong, HK
[6] School of Mathematics and Information Science, Guangzhou University, China

Abstract

The Alternating Direction Method of Multipliers (ADMM) has gained significant attention across a broad spectrum of machine learning applications. Incorporating the over-relaxation technique shows potential for enhancing the convergence rate of ADMM. However, determining optimal algorithmic parameters, including both the associated penalty and relaxation parameters, often relies on empirical approaches tailored to specific problem domains and contextual scenarios. Incorrect parameter selection can significantly hinder ADMM's convergence rate. To address this challenge, in this paper we first propose a general approach to optimize the value of penalty parameter, followed by a novel closed-form formula to compute the optimal relaxation parameter in the context of linear quadratic problems (LQPs). We then experimentally validate our parameter selection methods through random instantiations and diverse imaging applications, encompassing diffeomorphic image registration, image deblurring, and MRI reconstruction.

1 Introduction

ADMM is a versatile algorithm with applications spanning various domains, including compressed sensing (Hou, Li, and Zhang 2022; Liu et al. 2023), image processing (Chan, Wang, and Elgendy 2016; Yazaki, Tanaka, and Chan 2019), and machine learning (Li et al. 2022; Zhou and Li 2023). Although introduced in the 1970s for optimization, its roots can be traced back to the 1950s as a method to solve elliptic and parabolic partial difference equations (Boyd et al. 2011). ADMM leverages the convergence strengths of the method of multipliers and the decomposability property of dual ascent. It is particularly useful in addressing convex optimization of considerable scale, beyond the capacity of conventional solvers. The ongoing research and outstanding algorithmic performance have significantly contributed to its widespread adoption, highlighting the growing importance of exploring its theoretical properties, particularly regarding parameter selection (Ghadimi et al. 2014; Wang et al. 2019).

ADMM, from a technical viewpoint, decomposes complex optimization problems into manageable sub-problems, often solvable using point-wise, closed-form solvers (Candès et al. 2011; Lu et al. 2016; Thorley et al. 2021; Jia et al. 2021; Duan et al. 2023). It proceeds by iteratively updating these sub-problems alternately until a solution meeting the original problem's objectives and constraints is attained. Within ADMM, the augmented Lagrange function incorporates penalty terms associated with the constraints. The penalty parameters determine the strength of these penalty terms. As highlighted in (Deng and Yin 2016), the convergence rate of ADMM is directly impacted by these penalty parameters. The optimal selection of such parameters can significantly enhance the algorithm's convergence rate. However, the lack of a universal method to compute these parameters optimally remains a challenge.

The convergence rate of ADMM can be further accelerated by leveraging information from prior iterations during the computation of subsequent iterations. Such a technique is known as over-relaxation and often used in conjunction with ADMM (De Pierro and Iusem 1986; Zhang et al. 2020). Numerous research endeavors have been devoted to defining appropriate values for the resultant relaxation parameter. Notably, in the study conducted by (Eckstein 1994), the authors proposed a widely acknowledged empirical range of values, typically falling within $[1.5, 1.8]$, which however is not always the case according to our findings in this paper. Despite a multitude of papers presenting specific guidelines for selecting this parameter, many real-world application papers (Stellato et al. 2020; Duan et al. 2023) still resort to empirically determined values. This reliance on empirical choices is due to the absence of a straightforward and efficient method that can promptly and optimally determine this relaxation parameter.

The objective of this paper is to introduce novel methods for the selection of optimal parameters within both ADMM and over-relaxed ADMM. As an example, we focus on linear quadratic problems (LQPs), particularly with applications tailored to image processing. The theories developed in this paper could offer valuable insights for addressing other non-quadratic problems, such as non-smooth L_1 optimization. More specifically, we have identified four key contributions of this paper, summarized as follows:

- We perform a comprehensive convergence analysis of

†Corresponding author: j.duan@bham.ac.uk

the ADMM algorithm as applied to LQPs, effectively demonstrating its unconditional convergence within the context of LQPs. This is achieved by initially converting the ADMM iterations into its fixed-point iterations, which facilitates the derivation of the iteration matrix. Subsequently, we theoretically show that the spectral radius of the iteration matrix is bounded by 1, regardless of the value of the penalty parameter.

- We propose a general optimization method for the selection of the optimal penalty parameter in ADMM. We achieve this by utilizing numerical gradient descent to minimize the spectral radius of the iteration matrix. Moreover, in specific scenarios like image deblurring and MRI reconstruction, we show the existence of an closed-form solution for accurately determining the optimal penalty parameter within ADMM.

- We establish, for the first time, the existence of an closed-form solution for determining the relaxation parameter in over-relaxed ADMM. We find that for any arbitrary value of the penalty parameter, there exists a corresponding relaxation parameter, computed from the closed-form solution, that minimizes the spectral radius of the iteration matrix. Consequently, we can transform the original joint optimization problem, with respect to both penalty and relaxation parameters, into a single-variable optimization problem focused only on the penalty parameter.

- We verify our proposed parameter selection methods through random instantiations and practical real-world imaging applications, encompassing diffeomorphic image registration, image deblurring, MRI reconstruction. This approach sets us apart from previous methods, e.g., (Ghadimi et al. 2014), that only depend on simulated data for validation purpose.

2   Related Works

(Boley 2013) studied the convergence rate of ADMM for both quadratic and linear programs via the spectral analysis based on a novel matrix recurrence. While acknowledging that the penalty parameters of ADMM can influence its convergence rate, they did not offer guidance on how to select these parameters. To address this issue, (Ghadimi et al. 2014) reformulated ADMM into a fixed-point iteration system to analyze the impact of parameters on the convergence rate of ADMM and over-relaxed ADMM. By minimizing the spectral radius of the iteration matrix, they successfully derived optimal penalty and relaxation parameters for quadratic programming. (Teixeira et al. 2015) extended the applicability of Ghadimi's theory by transforming the distributed quadratic programming into an equivalent constrained quadratic programming. (França and Bento 2016) introduced a method that determines the relaxation parameter for semi-definite programming through the analysis of the problem's condition number.

(Boyd et al. 2011) suggested an empirical parameter update strategy for ADMM's penalty parameters. The idea is to maintain a proportional relationship between the norms of primal and dual residuals, ensuring their convergence to zero within a specified factor. (Xu, Figueiredo, and Goldstein 2017) proposed an adaptive ADMM approach by applying the Barzilai-Borwein spectral method to the original ADMM algorithm. Their method allows to dynamically update penalty parameters in each iteration based on primal and dual residuals. Inspired by this work, (Mavromatis, Foti, and Vavalis 2020) introduced a weighted penalty parameter ADMM algorithm for solving optimal power flow problems. Their approach involves the computation of absolute values from the admittance matrix and the Hessian matrix in each ADMM iteration. These values are then used to recalibrate the penalty parameters, aiming to refine the accuracy of parameter estimation.

However, certain limitations exist in the current research landscape. Firstly, many methods (Boyd et al. 2011; Xu, Figueiredo, and Goldstein 2017; Wohlberg 2017; Mhanna, Verbič, and Chapman 2018) rely on primal and dual residuals for estimating optimal parameters during iterations, but there often lack closed-form or explicit pre-iteration parameter selection approaches. Secondly, existing parameter selection techniques, based on the spectral analysis of the iteration matrix (Ghadimi et al. 2014; França and Bento 2016), predominantly focus on specific problem types (e.g., standard quadratic problem with L being an identity matrix). These methods requiring the spectral radius of the iteration matrix to be computable in an explicit form, which restricts their applicability and generalization ability (Stellato et al. 2020). In this paper, we will propose effective methods to address these two challenges.

3   Methodology

This section starts with the introduction of essential notations utilized in the subsequent formulations. We proceed by presenting the concept of fixed-point iterations, which serves as a foundational element for both the convergence analysis and parameter selection processes. Following this, we proceed to apply both ADMM and its over-relaxed variant to address LQPs. In the final stages, we propose novel methods for selecting the penalty and relaxation parameters. This is accomplished through the conversion of ADMM and over-relaxed ADMM into the form of fixed-point iterations, followed by the utilization of spectral radius analysis.

3.1   Notations and Fixed-Point Iterations

Let $\mathbb{R}$ and $\mathbb{C}$ denote respectively the set of real and complex numbers, $\mathbb{R}_{++}$ denote the set of positive numbers, $\mathcal{S}^{n \times n}$ denote the set of $n \times n$ matrices, and I^n (or I) be the $n \times n$ identity matrix. For the square matrix T and its corresponding eigenvalues $\lambda's$, we define the nth smallest eigenvalue of T as $\lambda_n(T)$, and the spectral radius of T as $\rho(T)$.

Fixed-point iterations involve the iterative process below

$$u^{k+1} = Tu^k + c,$$

where $T \in \mathcal{S}^{n \times n}$ is known as the iteration matrix, $u \in \mathbb{R}^n$, and $c \in \mathbb{R}^n$. It was shown in (Ghadimi et al. 2014) that the convergence factor ζ of this fixed-point iteration system is equal to $\rho(T)$. Here, the convergence factor ζ is defined as

$$\zeta \triangleq \sup_{k:u^k \neq u^*} \frac{\|u^{k+1} - u^*\|}{\|u^k - u^*\|},$$

where $\| \cdot \|$ represents the L_2 norm, and u^* denotes the optimal solution (i.e., so-called ground truth). The sequence $\{u_k\}$ is Q-sublinear if $\zeta = 1$, Q-linear if $\zeta < 1$, and Q-superlinear if $\zeta = 0$. Throughout this paper, the letter Q has been omitted when referring to the convergence rate. For linearly convergence sequences with $\zeta \in (0, 1)$, if we define t_ϵ as the smallest iteration count to ensure $\|u^{k+1} - u^*\| < \varepsilon$ for all $k > t_\varepsilon$, then t_ε can be calculated by $(\log(\varepsilon) - \log(\sigma)) / \log(\zeta)$, where σ denotes the worst case distance between u^0 and u^*, i.e., $\|u^0 - u^*\| < \sigma$. This suggests that by reducing the value of the the convergence factor ζ, the iteration count can be decreased, leading to a faster convergence rate.

3.2 ADMM for LQPs

The LQPs for image processing we study in this paper have the following structure

$$\min_u \frac{\mu}{2}\|Au - f\|^2 + \frac{1}{2}\|Lu\|^2, \quad (1)$$

where $\mu \in \mathbb{R}_{++}$ is the regularization parameter; $A \in \mathbb{R}^{m \times n}$ or $\mathbb{C}^{m \times n}$ ($m \leq n$) is an encoding matrix; $u \in \mathbb{R}^n$ or $\mathbb{C}^n$ is the unknown vector; $f \in \mathbb{R}^m$ or $\mathbb{C}^m$ is the input vector; and $L \in \mathbb{R}^{n \times n}$ is a regularization matrix. The value of μ determines the output quality, whereas smaller values of μ tend to yield smoother results. By differentiating (1) with respect to u and setting the respective derivative to zero, we have the following linear system

$$\left(\mu A^T A + L^T L\right) u = \mu A^T f. \quad (2)$$

When addressing the solution of Equation (2), two primary challenges arise: Firstly, in certain scenarios like our MRI reconstruction and diffeomorphic image registration, where $\left(\mu A^T A + L^T L\right)$ may be positive semi-definite, the process of inverting such a matrix becomes unfeasible. Secondly, in the context of higher-dimensional cases like 3D medical image registration (Thorley et al. 2021), even if the matrix $\left(\mu A^T A + L^T L\right)$ remains positive definite, the process of matrix inversion becomes computationally expensive. To address these two issues, we propose to use ADMM to handle the original problem (1), as an alternative to using the normal equation to solve (2).

To apply ADMM, we introduce an auxiliary variable $w \in \mathbb{R}^n$, a Lagrangian multiplier $b \in \mathbb{R}^n$, and a penalty parameter $\theta \in \mathbb{R}_{++}$, transforming (1) into the following augmented Lagrange function

$$\mathcal{L}(u, w; b) = \frac{\mu}{2}\|Au - f\|^2 + \frac{1}{2}\|Lw\|^2 + \frac{\theta}{2}\|w - u - b\|^2. \quad (3)$$

To optimize (3) with ADMM, we need to decompose it into two sub-problems with respect to u and w and then update the Lagrangian multipliers b until the process converges. The following Algorithm 1 outlines the optimization process using ADMM. In Algorithm 1, we have $w^{k+1} = (L^T L + \theta I)^{-1}(\theta u^k + \theta b^k)$ and $u^{k+1} = (\mu A^T A + \theta I)^{-1}(\theta w^{k+1} - \theta b^k + \mu A^T f)$. It is worth noting that while matrix inversion is applied to both variables w^{k+1} and u^{k+1}, fast solvers exist in specific cases due to the distinctive structure of $A^T A$ and $L^T L$. For instance, in diffeomorphic image registration,

---

Algorithm 1: ADMM for LQPs

Input: matrices A and L; parameter μ and θ
Initialize: u^0 and b^0
Repeat:

$$w^{k+1} = \arg\min_w \frac{1}{2}\|Lw\|^2 + \frac{\theta}{2}\|w - u^k - b^k\|^2$$
$$u^{k+1} = \arg\min_u \frac{\mu}{2}\|Au - f\|^2 + \frac{\theta}{2}\|w^{k+1} - u - b^k\|^2$$
$$b^{k+1} = b^k + u^{k+1} - w^{k+1}$$

until some stopping criterion is met

$A^T A$ takes on a rank-1 form, allowing efficient inversion through the Morris-Sherman equation (Bartlett 1951; Thorley et al. 2021). Similarly, in MRI reconstruction and diffeomorphic image registration, $L^T L$ can be effectively diagonalized using the discrete Fourier transformation basis functions (Goldstein and Osher 2009; Duan et al. 2023). Consequently, the application of ADMM to solve LQPs offers distinct advantages.

Theorem 1. *In order to determine the optimal penalty parameter θ^* in ADMM automatically, we need to transform the ADMM iterations in Algorithm 1 into the following fixed-point iteration system, solely with respect to the variable u*

$$u^{k+1} = (I + Q) u^k - \left(\mu A^T A + \theta I\right)^{-1}\left(\theta \mu A^T f\right), \quad (4)$$

where $I + Q$ is the iteration matrix with Q defined as

$$Q = \theta(\mu A^T A + \theta I)^{-1}((L^T L + \theta I)^{-1}(\theta I - \mu A^T A) - I). \quad (5)$$

Next, given a value of μ, we can prove

$$\rho (I + Q) \leq 1, \quad (6)$$

regardless of the value of θ. As per Section 3.1, we know that the convergence factor ζ of Algorithm 1 is equal to the spectral radius of the iteration matrix. As such, ζ is bounded by 1, meaning Algorithm 1 or (4) is unconditionally convergent.

Proof. Detailed derivations proving the equivalence between Algorithm 1 and the fixed-point iteration system (4), as well as Inequality (6), have been provided in Appendix 1 of the arXiv version of this paper. $\square$

Next, we search the optimal parameter θ^* that minimizes the convergence rate of Algorithm 1. Since ζ is dependent on the penalty parameter θ, the objective is to identify a value for θ that minimizes the convergence factor ζ. For this, we define the following minimization problem

$$\min_\theta \zeta(\theta), \quad (7)$$

where $\zeta(\theta) = \rho(I + Q(\theta))$. From Inequality (6) in Theorem 1 we have $\lambda_i(Q(\theta)) \in [-1, 0]$, and we can also easily derive $\lambda_i(I + Q(\theta)) = 1 + \lambda_i(Q(\theta))$. As such, we have $\rho(I + Q(\theta)) = 1 + \lambda_n(Q(\theta))$, with which the minimization problem (7) can be converted to

$$\min_\theta \lambda_n(Q(\theta)). \quad (8)$$

Though the minimization problem (8) is a one-dimensional optimization problem with respect to only θ,

Algorithm 2: Over-relaxed ADMM for LQPs
Input: matrices A and L; parameter μ, θ and α
Initialize: u^0 and b^0
Repeat:
$\quad w^{k+1} = \arg\min_{w} \frac{1}{2}\|Lw\|^2 + \frac{\theta}{2}\|w - u^k - b^k\|^2$
$\quad u^{k+1} = \arg\min_{u} \frac{\mu}{2}\|Au - f\|^2 + \frac{\theta}{2}\|\alpha w^{k+1} - \alpha u^k - b^k\|^2$
$\quad b^{k+1} = b^k + u^{k+1} - \alpha w^{k+1} - (1-\alpha)u^k$
until some stopping criterion is met

computing θ^* directly is however not trivial. This is the reason why a general applicable method for optimizing θ is still lacking. Previous works (Ghadimi et al. 2014; Teixeira et al. 2015) were based on the assumption that $\lambda_n(Q)$ can be explicitly written for spectral analysis. However, in practical applications such as diffeomorphic registration in Section 4.2, this is a significant limitation. To address this challenge, we propose to use numerical gradient descent to optimize θ

$$\theta^{k+1} = \theta^k - t\nabla\lambda_n\left(Q(\theta^k)\right), \tag{9}$$

where t denotes the step size. In this study, we employed the central finite difference scheme to compute gradients. Compared to the one-sided finite difference method, this scheme offers better numerical stability. It also provides more accurate estimation of gradients. It is important to note that this gradient descent method is general, as it does not need to know the explicit form of the eigenvalues of matrix Q. The definition for the central finite difference is given by

$$\nabla\lambda_n\left(Q(\theta^k)\right) \approx \frac{\lambda_n\left(Q(\theta^k+\eta)\right) - \lambda_n\left(Q(\theta^k-\eta)\right)}{2\eta},$$

where η represents a small value. In our experiments, we set this value within the range of 10^{-5} to 10^{-3}, which led to a satisfactory convergence of the gradient descent (9).

3.3 Over-Relaxed ADMM

Over-relaxation technique can be used in the ADMM algorithm and further accelerate the convergence rate of ADMM. This method is achieved by introducing an additional relaxation parameter α and replacing w^{k+1} in Algorithm 1 with $\alpha w^{k+1} + (1-\alpha)u^k$. Algorithm 2 outlines the optimization process of the augmented Lagrange function (3) using over-relaxed ADMM.

To investigate the influence of relaxation parameter α on convergence, we transform Algorithm 2 into its fixed-point iteration system. Such a conversion approach is in line with Proof of Theorem 1 in Appendix. The resulting fixed-point iteration system is given as follows

$$u^{k+1} = (I + \alpha Q)u^k - \alpha\left(\mu A^T A + \theta I\right)^{-1}\left(\theta\mu A^T f\right).$$

After obtaining the iteration matrix $I + \alpha Q$, we can analyze the spectral radius of this matrix to determine the optimal relaxation parameter α^*.

Theorem 2. *The optimal α^* can be directly calculated using the following closed-form formula*

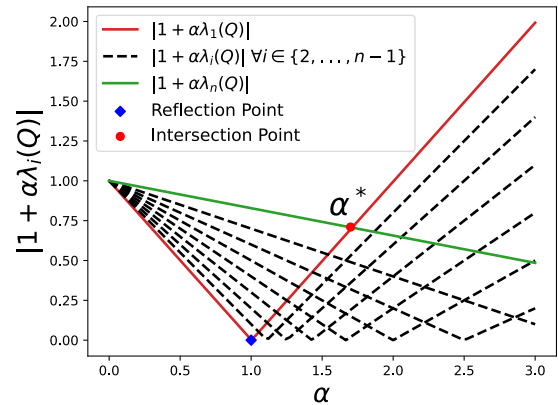$$\alpha^* = -\frac{2}{\lambda_1\left(Q(\theta)\right) + \lambda_n\left(Q(\theta)\right)}, \tag{10}$$

Figure 1: Relationship between $|1 + \alpha\lambda_i(Q)|$ and the value of α. The slope of each line before reflection is $\lambda_i(Q)$. The spectral radius before the intersection point is governed by the green line, while after the reflection, it is determined by the reflected red line. The intersection point corresponds to the optimal α^* as well as the minimum spectral radius of the iteration matrix $I + \alpha Q$.

where $Q(\theta)$ is a matrix whose entries reply on the value of θ. As per Equation (10), we can compute the optimal relaxation parameter α^* as long as a value of θ is given.

Proof. To prove Theorem 2, we begin with the following two-dimensional joint optimization problem

$$\min_{\theta,\alpha}\zeta(\theta,\alpha), \tag{11}$$

where $\zeta(\theta,\alpha) = \rho(I + \alpha Q(\theta))$. In order to express the spectral radius in terms of the eigenvalue structure, we first derive the equality $\lambda_i(I + \alpha Q(\theta)) = 1 + \alpha\lambda_i(Q(\theta))$, and the spectral radius $\rho(I + \alpha Q(\theta))$ is then defined as

$$\max_i|1 + \alpha\lambda_i(Q(\theta))|, \ \forall i \in \{1, ..., n\}. \tag{12}$$

From Inequality (6) in Theorem 1, we know $\lambda_i(Q(\theta)) \in [-1, 0]$. Based on this and (12), we plot Figure 1 to demonstrate the correlation between the absolute eigenvalue of the iteration matrix and the relaxation parameter. From this figure, it is straightforward to express the spectral radius as the following piecewise function

$$\rho = \begin{cases} 1 + \alpha\lambda_n(Q), & \text{if } -1 - \alpha\lambda_1(Q) \leq 1 + \alpha\lambda_n(Q) \\ -1 - \alpha\lambda_1(Q), & \text{if } -1 - \alpha\lambda_1(Q) > 1 + \alpha\lambda_n(Q) \end{cases}, \tag{13}$$

where with a slight abuse of notation, we use ρ to represent $\rho(I + \alpha Q(\theta))$.

Once we have (13), our objective is to minimize it in order to enhance the convergence rate. From Figure 1 again, it becomes evident that the smallest spectral radius is located at the intersection point where the following equality holds

$$-1 - \alpha\lambda_1(Q) = 1 + \alpha\lambda_n(Q),$$

from which α can be computed using a closed-form solution as follows

$$\alpha = -\frac{2}{\lambda_1(Q(\theta)) + \lambda_n(Q(\theta))},$$

which exactly verifies the validity of Equation (10). $\qquad\square$

If now we plug the optimal α^* into (13), we can convert the joint minimization problem (11) into the following minimization problem

$$\min_{\theta,\alpha} \zeta(\theta,\alpha) \Rightarrow \min_{\theta} \frac{\lambda_1(Q(\theta)) - \lambda_n(Q(\theta))}{\lambda_1(Q(\theta)) + \lambda_n(Q(\theta))}, \quad (14)$$

which is a single-variable optimization problem with respect to only θ. This problem can be minimized with a numerical gradient descent method similar to Equation (9). Once θ^* is found, α^* can be computed using the closed-form solution (10). It is worth noting that even if θ is not optimal, α computed via (10) can still accelerate convergence.

4 Experiments

In this section, we will first test the generalization ability of our proposed parameter selection method through random instantiations. Following that, we will apply the proposed parameter selection methods to diffeomorphic image registration, image deblurring, and MRI reconstruction. We will compare our optimal ADMM algorithm and over-relaxed variant (oADMM) with gradient descent (GD), gradient descent with Nesterov's acceleration (GD-N) (Nesterov 1983; Bartlett and Duan 2021), gradient descent with Nesterov's acceleration and restart (GD-NR) (O'donoghue and Candes 2015; Bartlett and Duan 2021), as well as conjugate gradient (CG). In all of our experiments, we chose the step size in gradient-based methods using the Lipschitz constant of the corresponding problem. It is worth noting that optimal values for penalty parameters can be determined analytically for image deblurring and MRI reconstruction problems. However, for image registration numerical gradient descent is required to compute these parameters.

4.1 Generalization Ability

Emphasizing that our approach is model-driven, the selection of parameters[1] relies on the matrices A and L in the minimization problem (1). As such, the measure of generalization ability lies in how effectively our method performs as A and L undergo variations, which is in contrast to data-driven methods, where the generalization ability is often examined using multiple different datasets.

We presented Figure 2 to demonstrate the generalization ability of our approach, where the analysis is based on 50 random instantiations of $A \in \mathbb{R}^{200 \times 50}$ and $L \in \mathbb{R}^{200 \times 50}$ while keeping f and μ fixed. For ADMM, we employed numerical gradient descent to minimize (8) with respect to θ. For oADMM, we utilized numerical gradient descent to minimize (36) with regard to θ, whilst the optimal value of α was calculated using (10) once θ^* was found. We note that the optimal values of θ for both ADMM and oADMM are similar and that the optimal values of α are not within $[1.5, 1.8]$ as suggested in (Eckstein 1994). As evident from Figure 2, the calculated optimal values consistently result in faster convergence rates for both ADMM and oADMM, reaffirming the generalization ability of our proposed parameter selection methods.

[1]The parameter selection also relies on the regularization parameter μ which we fix as a constant in this paper.
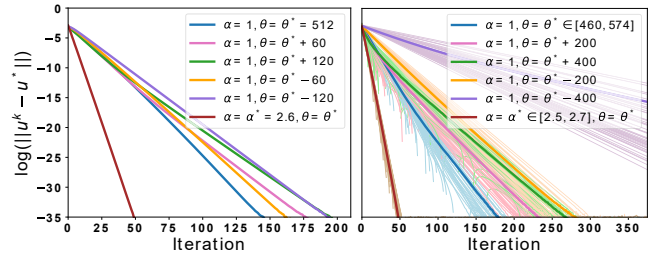

Figure 2: Left: Convergence rates of different methods and parameter values based on 1 random instantiation of A and L. Right: Convergence rates based on 50 random instantiations of A and L. The solid lines represent the average over 50 instantiations. The algorithm is ADMM when $\alpha = 1$, and oADMM when $\alpha = \alpha^*$.

4.2 Diffeomorphic Image Registration

Computing a diffeomorphic deformation can be treated as modelling a dynamical system (Beg et al. 2005), given by an ordinary differential equation (ODE): $\partial\phi/\partial t = \mathbf{v}_t(\phi_t)$, where $\phi_0 = \mathrm{Id}$ is the identity transformation and $\mathbf{v}_t$ indicates the velocity field at time $t (\in [0,1])$. The ODE can be solved by Euler integration, in which the deformation field ϕ is calculated as the compositions of a series of small deformations, defined as $\phi = (\mathrm{Id} + \frac{\mathbf{v}_{t_{N-1}}}{N}) \circ \cdots \circ (\mathrm{Id} + \frac{\mathbf{v}_{t_1}}{N}) \circ (\mathrm{Id} + \frac{\mathbf{v}_{t_0}}{N})$. If the velocity fields $\mathbf{v}_{t_i}$ are sufficiently small whilst satisfying some smoothness constraints, the resulting composition is a diffeomorphic deformation.

To compute the velocity fields whilst satisfying these diffeomorphic constraints, we minimize the following linear quadratic problem (Thorley et al. 2021)

$$\min_{v_x,v_y} \frac{\mu}{2}\|\langle I_x, v_x\rangle + \langle I_y, v_y\rangle + I_t\|^2 + \frac{1}{2}\|\nabla v_x\|^2 + \frac{1}{2}\|\nabla v_y\|^2, \quad (15)$$

where $I_x, I_y \in \mathbb{R}^n$ denote the spatial derivatives of the image; $I_t \in \mathbb{R}^n$ represents the temporal derivative of the image; and $v_x, v_y \in \mathbb{R}^n$ denote the velocity field in x and y directions. In this case, by setting

$$A = \begin{pmatrix} \mathrm{diag}(\langle I_x, I_x\rangle) & \mathrm{diag}(\langle I_x, I_y\rangle) \\ \mathrm{diag}(\langle I_y, I_x\rangle) & \mathrm{diag}(\langle I_y, I_y\rangle) \end{pmatrix} \in \mathbb{R}^{2n \times 2n}$$

and

$$L = \begin{pmatrix} \nabla^T\nabla & 0 \\ 0 & \nabla^T\nabla \end{pmatrix} \in \mathbb{R}^{2n \times 2n},$$

we can use numerical gradient descent to compute optimal parameters for both ADMM and oADMM.

In Figure 3, we show results obtained through the introduced diffeomorphic registration technique. We examine the impact of the penalty parameter θ in both ADMM and oADMM, and then evaluate the convergence efficiency of different algorithms. Given a pair of images (depicted as source and target in the figure), we can compute a deformation (shown in the bottom left panel) that ensures a positive Jacobian determinant (shown in the bottom middle panel) for all pixel positions. In the top right panel, we show the correlation between the spectral radius of the iteration matrix and θ in both ADMM and oADMM. As can be seen,
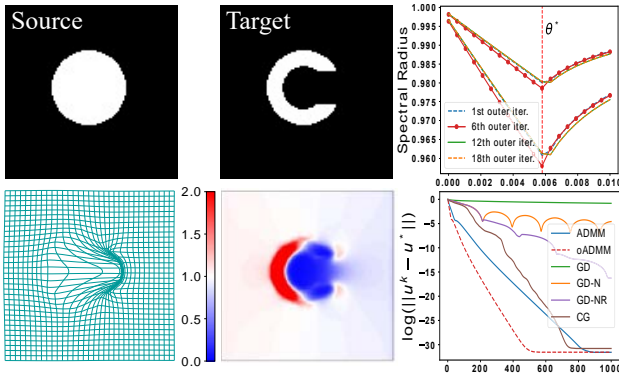
Figure 3: Illustration of diffeomorphic image registration results, visualization of the correlation between spectral radius and θ, and comparison of convergence rates of algorithms. The x-axes of the two plots in the third column represent the values of θ and iteration numbers, respectively.

there exists an unique optimal value where the spectral radius is minimized. As such, when using numerical gradient descent, it is possible to find the optimal value of θ that can considerably reduce iteration counts. This panel also illustrates that θ^*, producing the smallest spectral radius for oADMM, closely aligns with that of ADMM. Furthermore, due to the two-loop[2] iterative nature of diffeomorphic image registration, the data term of (15) undergoes slight changes at each iteration of the outer loop. These changes however do not significantly influence the value of θ^*, as evident from the top right panel. Therefore, given a specific value of μ, it is sufficient to use gradient descent to search θ^* for each outer iteration. Finally, in the bottom right panel, convergence rates among different algorithms are compared. As is evident, the parameter-optimized oADMM algorithm remains the fastest in terms of convergence rate.

4.3 Image Deblurring

In this application, we look at a phantom test image. The image went through a Gaussian blur of size 7×7 and standard deviation 2, followed by an additive zero-mean white Gaussian noise with standard deviation 10^{-4}. The top left and middle panels of Figure 4 depict the original and blurred images, respectively. To deblur the image we minimize the following problem

$$\min_u \frac{\mu}{2}\|Ku - f\|^2 + \frac{1}{2}\|u\|^2, \quad (16)$$

where $K \in \mathcal{S}^{n \times n}$ is the matrix representing the blur operator, $u \in \mathbb{R}^n$ is the vectorized unknown clean image, and $f \in \mathbb{R}^n$ is the vectorized input image. By setting $A = K$ and $L = I$, the matrix Q in (5) for this application has the form of

$$Q = \theta(\mu K^T K + \theta I)^{-1}((I + \theta I)^{-1}(\theta I - \mu K^T K) - I).$$

Since K is a convolution matrix derived from the Gaussian kernel function, the eigenvalues of $K^T K$ can be calculated using the two-dimensional discrete Fourier transform

---

[2]We did not use the pyramid implementation as in (Thorley et al. 2021), so we ended up with a two-loop algorithm comprising inner ADMM/oADMM iterations and outer warping iterations.
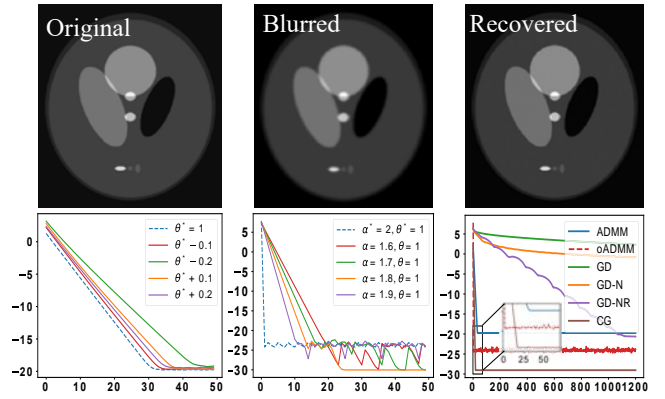

Figure 4: Demonstration of image deblurring effects and convergence rates of different algorithms. The x-axis and y-axis of each plot in the second row represent iteration numbers and $\log(\|u^k - u^*\|)$, respectively.

(Capus and Brown 2003). With $\lambda\left(K^T K\right)$, we can derive the maximum eigenvalues of Q as

$$\lambda_n(Q) = -\frac{\theta + \theta\mu\lambda_i(K^T K)}{\theta^2 + \mu\lambda_i(K^T K) + \theta + \theta\mu\lambda_i(K^T K)}, \quad (17)$$

where i is either 1 or n. Since $\lambda_n(Q)$ in this case can be explicitly written, we can derive closed-form solutions for the parameters in ADMM and over-relaxed ADMM. In Theorem 3, we give their optimal parameters.

Theorem 3. *Firstly, to tackle the optimization problem* (16) *using ADMM, given a regularization parameter $\mu \in \mathbb{R}_{++}$, the optimal value of the penalty parameter in ADMM can be expressed in closed form*

$$\theta^* = \begin{cases} \sqrt{\mu}, & \text{if } \mu \leq 1 \\ 1, & \text{otherwise} \end{cases},$$

which was derived by minimizing the value of $\lambda_n(Q)$ in (17).

If over-relaxed ADMM is used to tackle the optimization problem (16), *the optimal penalty and relaxation parameters are given by*

$$\theta^* = 1 \text{ and } \alpha^* = 2,$$

among which θ^* was determined by minimizing the problem (36) *with $\lambda_1(Q)$ and $\lambda_n(Q)$ defined in* (17), *and then α^* was computed using* (10) *with θ^*.*

Proof. Detailed derivations have been given in Appendix 2 of the arXiv version of this paper. □

In Figure 4, the top right panel displays the deblurred image from oADMM (comparable results were achieved with ADMM), which closely resembles the original image. Note that we set the regularization parameter μ to 10^3 for this experiment. The bottom left and middle panels demonstrate that with the optimal θ^* and α^* there is a clear enhancement over the model's convergence, and that the optimal values of θ for both ADMM and oADMM are the same in this case. The bottom right panel shows that ADMM, CG, and oADMM exhibit superior performance than GD, GD-N, and GD-NR. Upon a detailed examination from the zoomed-in

window, CG addresses this quadratic problem very well, albeit still needing multiple iterations to attain convergence. In contrast, oADMM achieves convergence in a single step[3], outperforming all compared algorithms.

4.4 MRI Reconstruction

To reconstruct MR images we minimize the problem

$$\min_u \frac{\mu}{2}\|\mathcal{D}\mathcal{F}u - f\|^2 + \frac{1}{2}\|\nabla u\|^2, \quad (18)$$

where $\mathcal{D} \in \mathbb{R}^{m \times n}$ $(m < n)$ is the sampling matrix; $\mathcal{F} \in \mathbb{C}^{n \times n}$ is the Fourier transform matrix; $u \in \mathbb{C}^n$ is a complex-valued MR image stacked as a column vector; $f \in \mathbb{C}^m$ is the undersampled k-space data; ∇ denotes the first-order gradient operator. By setting $A = \mathcal{D}\mathcal{F}$ and $L = \nabla$, the matrix Q in (5) for this application has the form of

$$Q = \theta(\mu M_1 + \theta I)^{-1}((\nabla^T\nabla + \theta I)^{-1}(\theta I - \mu M_1) - I), \quad (19)$$

where $M_1 = \mathcal{F}^T\mathcal{D}^T\mathcal{D}\mathcal{F}$. Due to the use of periodic boundary conditions, $\nabla^T\nabla$ can be efficiently diagonalized in the form of $\mathcal{F}^T G \mathcal{F}$, where G is a diagonal matrix. Equation (19) can be simplified to $\mathcal{F}^T M_2 \mathcal{F}$, where M_2 is given as

$$M_2 = \theta(\mu\mathcal{D}^T\mathcal{D} + \theta I)^{-1}((G + \theta I)^{-1}(\theta I - \mu\mathcal{D}^T\mathcal{D}) - I),$$

which is a diagonal matrix. The eigenvalues of $\mathcal{F}^T M_2 \mathcal{F}$ are simply the values along the diagonal of M_2. If we define $\lambda_i(G)$ as the ith smallest eigenvalue of G, and d_i as the diagonal value of $\mathcal{D}^T\mathcal{D}$ at the position where $\lambda_i(G)$ is indexed from G, the maximum eigenvalues of Q can be derived as

$$\lambda_n(Q) = -\frac{\theta\lambda_i(G) + \theta\mu d_i}{\theta^2 + \mu d_i\lambda_i(G) + \theta\lambda_i(G) + \theta\mu d_i},$$

where $i \in \{1, ..., n\}$. Since $\lambda_n(Q)$ can be written explicitly, we can derive the closed-form solution for θ in ADMM. In Theorem 3, we present the optimal value for this parameter. If over-relaxed ADMM is used to solve (18), a closed-form solution still exists for θ. It is however too cumbersome to derive them in this case. As such, the penalty parameter θ in over-relaxed ADMM was searched by gradient descent, and once θ^* was found the optimal relaxation parameter α^* can be directly obtained using Equation (10) with θ^*.

Theorem 4. *To tackle the optimization problem* (18) *using ADMM, given a regularization parameter $\mu \in \mathbb{R}_{++}$, the optimal value of the penalty parameter in ADMM can be expressed in closed form*

$$\theta^* = \begin{cases} \sqrt{\mu a}, & \text{if} \quad \mu \leq 2b - a \\ \sqrt{\frac{\mu ab}{\mu + a - b}}, & \text{if} \quad 2b - a < \mu \leq a \\ \mu, & \text{if} \quad a < \mu \leq b \\ \sqrt{\frac{\mu cb}{\mu + c - b}}, & \text{otherwise} \end{cases},$$

where a, b and c are defined as follows

$$a = \lambda_1((\mathcal{D}^T\mathcal{D}) \odot G)$$
$$b = \lambda_1((1 - \mathcal{D}^T\mathcal{D}) \odot G),$$
$$c = \lambda_x((\mathcal{D}^T\mathcal{D}) \odot G)$$

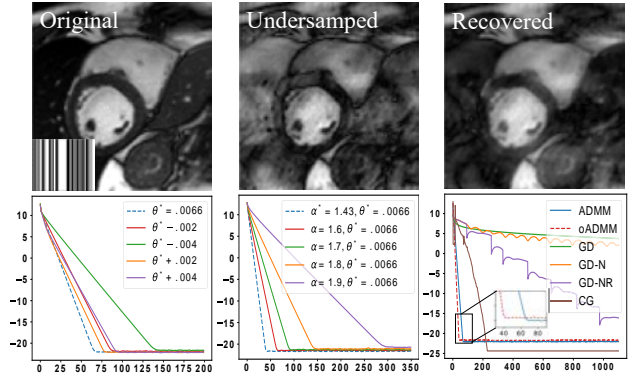[3]The spectral radius in this case is close to zero, leading to a superlinear convergence rate.

Figure 5: Demonstration of MRI reconstruction results and comparison of convergence rates among algorithms. The x-axis and y-axis of each plot in the second row represent iteration numbers and $\log(\|u^k - u^*\|)$, respectively.

where $\odot$ is the hadamard product; $\lambda_1((\mathcal{D}^T\mathcal{D}) \odot G)$ denotes the smallest eigenvalue of $(\mathcal{D}^T\mathcal{D}) \odot G$, excluding the eigenvalues corresponding to zero entries on the diagonal of $\mathcal{D}^T\mathcal{D}$; $\lambda_1((1 - \mathcal{D}^T\mathcal{D}) \odot G)$ denotes the smallest eigenvalue of $(1 - \mathcal{D}^T\mathcal{D}) \odot G$, excluding the eigenvalues corresponding to zero entries on the diagonal $1 - \mathcal{D}^T\mathcal{D}$; and $\lambda_x((\mathcal{D}^T\mathcal{D}) \odot G)$ represents the largest eigenvalue of $(\mathcal{D}^T\mathcal{D}) \odot G$, excluding the eigenvalues corresponding to zero entries along the diagonal of $\mathcal{D}^T\mathcal{D}$.

Proof. Detailed derivations have been given in Appendix 3 of the arXiv version of this paper. □

In Figure 5, we reconstruct a cardiac MR image from k-space. The original image (displayed in the top left panel) was first transformed into k-space using the Fourier transformation. Then 50% of the data there was taken using a cartesian sampling mask, displayed in the original image. This undersampled data was then corrupted by an additive zero-mean white Gaussian noise with standard deviation 1 to form f in (18). The reconstruction (top right), despite some slight blurring due to the smooth regularization, clearly enhances image quality compared to that displayed in the top middle panel, which is a direct reconstruction of f using the inverse Fourier transformation. The bottom left and middle panels of this figure illustrate that the choice of θ and α has a significant impact on the convergence rate and that our proposed methods result in faster convergence. Thanks to the utilization of these optimal parameters, we observe a clear superiority of our ADMM and oADMM over GD, its accelerated variants, and CG in terms of convergence efficiency.

5 Conclusion

In this paper, we presented automated techniques for selecting optimal penalty and relaxation parameters within the framework of ADMM and over-relaxed ADMM for linear quadratic problems. Our approaches involve a numerical gradient descent method for estimating the penalty parameter and a novel closed-form solution for determining the optimal relaxation parameter. We verified the generalizability and efficacy of these approaches through random instantiations and real-world imaging applications.

Appendix 1

In Algorithm 1 of ADMM, the u-update, w-update, and b-update are respectively provided as follows

$$w^{k+1} = \left(L^T L + \theta I\right)^{-1} \left(\theta u^k + \theta b^k\right), \qquad (20)$$

$$u^{k+1} = \left(\mu A^T A + \theta I\right)^{-1} \left(\theta w^{k+1} - \theta b^k + \mu A^T f\right), \quad (21)$$

$$b^{k+1} = b^k + u^{k+1} - w^{k+1}. \qquad (22)$$

From Equation (21) we obtain

$$b^k = w^{k+1} - u^{k+1} - \frac{\mu A^T A u^{k+1}}{\theta} + \frac{\mu A^T f}{\theta},$$

with which Equation (22) can be re-written as follows

$$b^{k+1} = -\frac{\mu A^T A}{\theta} u^{k+1} + \frac{\mu A^T f}{\theta},$$

which implies

$$b^k = -\frac{\mu A^T A}{\theta} u^k + \frac{\mu A^T f}{\theta}. \qquad (23)$$

By replacing b^k in (20) with (23), w-update in (20) can be re-written as

$$w^{k+1} = (L^T L + \theta I)^{-1}((\theta I - \mu A^T A)u^k - \mu A^T f). \qquad (24)$$

Substituting (24), along with (23), into (21) yields the fixed iteration system below, solely with respect to the variable u

$$u^{k+1} = (I + Q) u^k - \left(\mu A^T A + \theta I\right)^{-1} \left(\theta \mu A^T f\right), \quad (25)$$

where $I + Q$ is the iteration matrix[4] with Q defined as

$$Q = \theta(\mu A^T A + \theta I)^{-1}((L^T L + \theta I)^{-1}(\theta I - \mu A^T A) - I),$$

or equivalently

$$Q = -(\mu A^T A + \theta I)^{-1}(L^T L + \theta I)^{-1}(\theta \mu A^T A + \theta L^T L). \qquad (26)$$

The iteration matrix $I + Q$ can be derived as

$$I + Q = (\mu A^T A + \theta I)^{-1}(L^T L + \theta I)^{-1}(\theta^2 I + \mu L^T L A^T A). \qquad (27)$$

So far, we have successfully transformed the ADMM iterations in Algorithm 1 into the fixed-point iterations (25). Next, we need to analyze the spectral radius of the iteration matrix to establish the convergence of ADMM. Recall that for the matrix Q and its corresponding eigenvalues $\lambda's$, we define $\lambda_i(Q)$ as the ith smallest eigenvalue of Q, $\rho(Q)$ as the spectral radius of Q, and $\|\cdot\|$ as the spectral norm.

First, by adding (26) and (27), we obtain

$$I + 2Q = (\mu A^T A + \theta I)^{-1}(L^T L + \theta I)^{-1}$$
$$(\theta^2 I + \mu L^T L A^T A - \theta L^T L - \theta \mu A^T A),$$

which is equivalent to

$$I + 2Q = (\mu A^T A + \theta I)^{-1}(L^T L + \theta I)^{-1}$$
$$(\theta I - L^T L)(\theta I - \mu A^T A).$$

---

[4]The iteration matrix may not be a symmetric matrix, so in theory its eigenvalues could be complex. However, in our applications we did not observe such an effect.

In theory, altering the sequence of matrix multiplication will yield a matrix that has the same eigenvalues to the original one. In this case, we define the matrix S below

$$S = (L^T L + \theta I)^{-1}(\theta I - L^T L)(\theta I - \mu A^T A)(\mu A^T A + \theta I)^{-1},$$

which has the same eigenvalues as $I + 2Q$. Now, if we analyze the spectral norm of S, we can have the following two inequalities

$$\|S_1\| = \|(L^T L + \theta I)^{-1}(\theta I - L^T L)\|$$
$$= \max_i \left|\frac{\theta - \lambda_i(L^T L)}{\theta + \lambda_i(L^T L)}\right| \leq 1, \forall i \in \{1, ..., n\}.$$

$$\|S_2\| = \|(\theta I - \mu A^T A)(\mu A^T A + \theta I)^{-1}\|$$
$$= \max_i \left|\frac{\theta - \lambda_i(\mu A^T A)}{\theta + \lambda_i(\mu A^T A)}\right| \leq 1, \forall i \in \{1, ..., n\}.$$

Combining such two inequalities, we can draw the conclusion that

$$\rho(I + 2Q) = \rho(S) \leq \|S\| \leq \|S_1\|\|S_2\| = 1.$$

As such, the following inequality holds

$$-1 \leq 1 + 2\lambda_i(Q) \leq 1, \ \forall i \in \{1, ..., n\},$$

which means

$$\lambda_i(Q) \in [-1, 0], \ \forall i \in \{1, ..., n\},$$

and

$$\rho(I + Q) \leq 1.$$

This essentially proves that the spectral radius of the iteration matrix is bounded by 1, thereby allowing us to conclude that the ADMM or Algorithm 1 for linear quadratic problems converges unconditionally.

Appendix 2

Given a value of the regularization parameter μ, the maximum eigenvalue of Q is of the following form

$$\lambda_n(Q(\theta)) = -\frac{\theta + \theta\mu\lambda_i(K^T K)}{\theta^2 + \mu\lambda_i(K^T K) + \theta + \theta\mu\lambda_i(K^T K)}.$$

First, we derive θ^* in ADMM. It is evident that $\lambda_n(Q)$ increases monotonically as i in $\lambda_i(K^T K)$ increases when $\theta \geq 1$, but decreases monotonically as i in $\lambda_i(K^T K)$ increases when $\theta < 1$. Hence, we only need to discuss these two scenarios separately. Due to the fact that $\lambda_1(K^T K)$ close to 0 and $\lambda_n(K^T K) = 1$, $\lambda_n(Q)$ can be re-written in the piecewise function as follows

$$\lambda_n(Q(\theta)) = \begin{cases} -\dfrac{\theta + \theta\mu}{\theta^2 + \mu + \theta + \mu\theta} & \text{if } \theta < 1 \\ -\dfrac{1}{\theta + 1} & \text{if } \theta \geq 1 \end{cases}.$$

Now we need to minimize $\lambda_n(Q(\theta))$ with respect to θ. When $\theta \geq 1$, $\lambda_n(Q(\theta))$ increases monotonically when θ increases. The optimal penalty parameter in this case that minimizes $\lambda_n(Q)$ is given by

$$\theta^* = 1, \ \lambda_n(Q(\theta^*)) \approx -\frac{1}{2}.$$

When $\theta < 1$, it is straightforward to know

$$\theta^* = \sqrt{\mu}, \ \lambda_n\left(Q(\theta^*)\right) = -\frac{1+\mu}{2\sqrt{\mu}+1+\mu}.$$

So far, we have computed the optimal values of θ^* for the two cases. The next step is to determine which of the two cases has a smaller function value (e.g., smaller spectral radius). As such, we have

$$\theta^* = \begin{cases} \sqrt{\mu}, & \text{if } -\frac{1+\mu}{2\sqrt{\mu}+1+\mu} \leq -\frac{1}{2} \\ 1, & \text{otherwise} \end{cases},$$

which is equivalent to the following form

$$\theta^* = \begin{cases} \sqrt{\mu}, & \text{if } \mu \leq 1 \\ 1, & \text{otherwise} \end{cases}.$$

Next, we optimize the parameters in over-relaxed ADMM. When $\theta \geq 1$, $\lambda_1(Q(\theta))$ and $\lambda_n(Q(\theta))$ are respectively defined as

$$\lambda_1(Q(\theta)) = -\frac{\theta+\theta\mu}{\theta^2+\mu+\theta+\theta\mu}, \ \lambda_n(Q(\theta)) = -\frac{1}{\theta+1}.$$

In this case, the minimization problem (14) in the main paper can be converted to

$$\min_{\theta} \frac{\lambda_1(Q(\theta)) - \lambda_n(Q(\theta))}{\lambda_1(Q(\theta)) + \lambda_n(Q(\theta))} \Rightarrow \min_{\theta} \frac{(\theta-1)\mu}{2\theta+\mu+\theta\mu}.$$

When $\theta \geq 1$, the objective function value of the converted problem is greater or equal to zero. In such a case, the minimum of the objective function is achieved when $\theta^* = 1$. The optimal value of α^* is computed by plugging $\theta^* = 1$ into Equation (10) in the main paper, resulting in

$$\alpha^* = -\frac{2}{\lambda_1\left(Q(\theta^*)\right) + \lambda_n\left(Q(\theta^*)\right)} = 2.$$

Similarly, when $\theta \leq 1$, $\lambda_1(Q)$ and $\lambda_n(Q)$ are respectively defined as follows

$$\lambda_1(Q(\theta)) = -\frac{1}{\theta+1}, \ \lambda_n(Q(\theta)) = -\frac{\theta+\theta\mu}{\theta^2+\mu+\theta+\theta\mu}.$$

In this case, the minimization problem (14) in the main paper can be converted to

$$\min_{\theta} \frac{\lambda_1(Q(\theta)) - \lambda_n(Q(\theta))}{\lambda_1(Q(\theta)) + \lambda_n(Q(\theta))} \Rightarrow \min_{\theta} \frac{(1-\theta)\mu}{2\theta+\mu+\theta\mu}.$$

When $\theta \leq 1$, the objective function value of the converted problem is also greater or equal to zero. This implies that $\theta^* = 1$ and the corresponding optimal over-relaxation parameters $\alpha^* = 2$. As such, for both cases we have the optimal penalty and over-relaxation parameters given as follows

$$\theta^* = 1 \text{ and } \alpha^* = 2,$$

which verifies the validity of the closed-form solutions in Theorem 3.

Appendix 3

We start from the following equation

$$\lambda_n(Q) = \lambda_n(\mathcal{F}^T M_2 \mathcal{F}) = \lambda_n(M_2), \qquad (28)$$

where M_2 is a diagonal matrix defined as

$$M_2 = \theta(\mu\mathcal{D}^T\mathcal{D} + \theta I)^{-1}((G + \theta I)^{-1}(\theta I - \mu\mathcal{D}^T\mathcal{D}) - I).$$

If we define a vector $v \in \mathbb{R}^n$, each element of which corresponds to a diagonal entry of M_2, we have

$$v_i = \theta\frac{1}{\mu(\mathcal{D}^T\mathcal{D})_{ii} + \theta}\left(\frac{\theta - \mu(\mathcal{D}^T\mathcal{D})_{ii}}{G_{ii} + \theta} - 1\right)$$

$$= -\frac{\theta G_{ii} + \theta\mu(\mathcal{D}^T\mathcal{D})_{ii}}{\theta^2 + \mu(\mathcal{D}^T\mathcal{D})_{ii}G_{ii} + \theta G_{ii} + \theta\mu(\mathcal{D}^T\mathcal{D})_{ii}}$$

$$= \begin{cases} -\frac{\theta G_{ii}}{\theta^2 + \theta G_{ii}} & \text{if } (\mathcal{D}^T\mathcal{D})_{ii} = 0 \\ -\frac{\theta G_{ii} + \theta\mu}{\theta^2 + \mu G_{ii} + \theta G_{ii} + \theta\mu} & \text{if } (\mathcal{D}^T\mathcal{D})_{ii} = 1 \end{cases}.$$

According to Equation (28), the spectral radius of Q is equivalent to the maximum value found on the diagonal of M_2, so we have

$$\lambda_n(Q) = \lambda_n(M_2) = \max_i v_i =$$

$$\max_i\left\{-\frac{\theta G_{ii}}{\theta^2 + \theta G_{ii}}, -\frac{\theta G_{ii} + \theta\mu}{\theta^2 + \mu G_{ii} + \theta G_{ii} + \theta\mu}\right\}, \quad (29)$$

subject to the constraints with respect to the binary matrix $\mathcal{D}^T\mathcal{D}$. We need to divide the maximization problem (29) into two separate maximization problems, as follows

$$\max_i\left\{-\frac{G_{ii}}{\theta + G_{ii}}\right\} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 0, \qquad (30)$$

and

$$\max_i\left\{-\frac{\theta G_{ii} + \theta\mu}{\theta^2 + \mu G_{ii} + \theta G_{ii} + \theta\mu}\right\} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 1. \qquad (31)$$

If we now introduce four new variables

$$a = \arg\min_i G_{ii} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 1,$$

$$b = \arg\min_i G_{ii} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 0,$$

$$c = \arg\max_i G_{ii} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 1,$$

and

$$d = \arg\max_i G_{ii} \ s.t. \ (\mathcal{D}^T\mathcal{D})_{ii} = 0,$$

then (30) and (31) can be converted into the following two equivalent maximization problems

$$\max\left\{-\frac{b}{\theta + b}, -\frac{d}{\theta + d}\right\}, \qquad (32)$$

and

$$\max\left\{-\frac{\theta a + \theta\mu}{\theta^2 + \mu a + \theta a + \theta\mu}, -\frac{\theta c + \theta\mu}{\theta^2 + \mu c + \theta c + \theta\mu}\right\}. \qquad (33)$$

If we denote the solution for the maximization (32) as x, then we have

$$x = \max\left\{-\frac{b}{\theta+b}, -\frac{d}{\theta+d}\right\} = -\frac{b}{\theta+b}.$$

If we define the solution for the maximization (33) as y, then we have

$$y = \begin{cases} -\dfrac{\theta a + \mu\theta}{\theta^2 + \mu a + \theta a + \mu\theta} & \text{if } \theta > \mu \\ -\dfrac{\theta c + \mu\theta}{\theta^2 + \mu c + \theta c + \mu\theta} & \text{if } \theta < \mu \\ -1/2 & \text{if } \theta = \mu \end{cases}.$$

Next, we must compare x and y to determine identify the larger one as $\lambda_n(Q)$, and subsequently minimize it with respect to θ. This results in the following discussions.

First, we consider the case when $\theta > \mu$ and $\mu < a$. In this case, we have

$$F \triangleq x - y = \frac{(a + \mu - b)\theta^2 - \mu ab}{(\theta^2 + \mu a + \theta a + \mu\theta)(\theta + b)}.$$

The denominator of F is always greater than zero, so we only need to consider the sign of the numerator. In the case when $\mu \leq b - a$, $F < 0$, then we have $y > x$, leading to $\theta^* = \sqrt{\mu a}$. In the case when $\mu > b - a$, $F = 0$ holds only when θ satisfies the following equation

$$\theta = \sqrt{\frac{\mu ab}{\mu + a - b}}.$$

If $\theta \geq e_1$ where e_1 denotes $\sqrt{\mu ab/(\mu + a - b)}$, we have $x \geq y$, then $\theta^* = e_1$. If $\theta < e_1$, then $x < y$. In this case, we should compute θ^* from y. If $e_1 \geq \sqrt{\mu a}$, which is equivalent to $\mu \leq 2b - a$, then $\theta^* = \sqrt{\mu a}$. If $e_1 < \sqrt{\mu a}$, $\theta^* = e_1$. Consequently, we can conclude

$$\theta^* = \begin{cases} \sqrt{\mu a}, & \text{if } \mu \leq 2b - a \\ e_1, & \text{if } 2b - a < \mu \leq a \end{cases}. \quad (34)$$

Next, we consider the case when $\theta < \mu$ and $\mu > c$. In this case, we have

$$F = x - y = \frac{(c + \mu - b)\theta^2 - \mu cb}{(\theta^2 + \mu c + \theta c + \mu\theta)(\theta + b)},$$

where $c \geqslant b$. F in this case increases monotonically in θ. Moreover, $F = 0$ holds only when θ satisfies the following equation

$$\theta = \sqrt{\frac{\mu cb}{\mu + c - b}}.$$

As $c \geqslant b$, it is easy to know $e_2 < \sqrt{\mu c}$ in which e_2 denotes $\sqrt{\mu cb/(\mu + c - b)}$. In this case, when $\theta \leq e_2$, y monotonically decreases in θ. As such, y achieves its minimum when $\theta^* = e_2$. This implies that when $\mu > c$, the optimal value x and y are exactly the same, which reads

$$\theta^* = e_2, \text{ if } \mu > c. \quad (35)$$

Lastly, we consider the case when $\theta = \mu$ and $a \leq \mu \leq c$. In this case, we have

$$F \triangleq x - y = \frac{\theta - b}{2(\theta + b)}.$$

It is easy to know if $\theta > b$, then $x > y$. If $\theta \leq b$, then y is the largest eigenvalue. In this case, the optimal θ^* is given as follows

$$\theta^* = \begin{cases} \mu, & \text{if } a < \mu \leq b \\ e_2, & \text{if } b < \mu \leq c \end{cases}. \quad (36)$$

By combining Equations (34), (35), and (36), we have addressed all scenarios where $\mu < a$, $a \leq \mu \leq c$, and $\mu > c$. As such, we derive the optimal value of θ^* for ADMM as follows

$$\theta^* = \begin{cases} \sqrt{\mu a}, & \text{if } \mu \leq 2b - a \\ \sqrt{\dfrac{\mu ab}{\mu + a - b}}, & \text{if } 2b - a < \mu \leq a \\ \mu, & \text{if } a < \mu \leq b \\ \sqrt{\dfrac{\mu cb}{\mu + c - b}}, & \text{if } b < \mu \end{cases},$$

which verifies the validity of the closed-form solution in Theorem 4.

References

Bartlett, J.; and Duan, J. 2021. Accelerated first order methods for variational imaging. *arXiv preprint arXiv:2110.02813*.

Bartlett, M. S. 1951. An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 22(1): 107–111.

Beg, M. F.; Miller, M. I.; Trouvé, A.; and Younes, L. 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61: 139–157.

Boley, D. 2013. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM Journal on Optimization*, 23(4): 2183–2207.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1): 1–122.

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3): 1–37.

Capus, C.; and Brown, K. 2003. Fractional Fourier transform of the Gaussian and fractional domain signal support. *IEE Proceedings-Vision, Image and Signal Processing*, 150(2): 99–106.

Chan, S. H.; Wang, X.; and Elgendy, O. A. 2016. Plug-and-play ADMM for image restoration: fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1): 84–98.

De Pierro, A. R.; and Iusem, A. 1986. A relaxed version of Bregman's method for convex programming. *Journal of Optimization Theory and Applications*, 51: 421–440.

Deng, W.; and Yin, W. 2016. On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66: 889–916.

Duan, J.; Jia, X.; Bartlett, J.; Lu, W.; and Qiu, Z. 2023. Arbitrary order total variation for deformable image registration. *Pattern Recognition*, 109318.

Eckstein, J. 1994. Parallel alternating direction multiplier decomposition of convex programs. *Journal of Optimization Theory and Applications*, 80(1): 39–62.

França, G.; and Bento, J. 2016. An explicit rate bound for over-relaxed ADMM. In *2016 IEEE International Symposium on Information Theory (ISIT)*, 2104–2108. IEEE.

Ghadimi, E.; Teixeira, A.; Shames, I.; and Johansson, M. 2014. Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems. *IEEE Transactions on Automatic Control*, 60(3): 644–658.

Goldstein, T.; and Osher, S. 2009. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2): 323–343.

Hou, R.; Li, F.; and Zhang, G. 2022. Truncated residual based plug-and-play ADMM algorithm for MRI reconstruction. *IEEE Transactions on Computational Imaging*, 8: 96–108.

Jia, X.; Thorley, A.; Chen, W.; Qiu, H.; Shen, L.; Styles, I. B.; Chang, H. J.; Leonardis, A.; De Marvao, A.; O'Regan, D. P.; et al. 2021. Learning a model-driven variational network for deformable image registration. *IEEE Transactions on Medical Imaging*, 41(1): 199–212.

Li, Q.; Kailkhura, B.; Goldhahn, R.; Ray, P.; and Varshney, P. K. 2022. Robust decentralized learning using ADMM with unreliable agents. *IEEE Transactions on Signal Processing*, 70: 2743–2757.

Liu, Y.; Huang, K.; Yang, C.; and Wang, Z. 2023. Distributed network reconstruction based on binary compressed sensing via ADMM. *IEEE Transactions on Network Science and Engineering*.

Lu, W.; Duan, J.; Qiu, Z.; Pan, Z.; Liu, R. W.; and Bai, L. 2016. Implementation of high-order variational models made easy for image processing. *Mathematical Methods in the Applied Sciences*, 39(14): 4208–4233.

Mavromatis, C.; Foti, M.; and Vavalis, M. 2020. Auto-tuned weighted-penalty parameter ADMM for distributed optimal power flow. *IEEE Transactions on Power Systems*, 36(2): 970–978.

Mhanna, S.; Verbič, G.; and Chapman, A. C. 2018. Adaptive ADMM for distributed AC optimal power flow. *IEEE Transactions on Power Systems*, 34(3): 2025–2035.

Nesterov, Y. E. 1983. A method of solving a convex programming problem with convergence rate $O(\frac{1}{k^2})$. In *Doklady Akademii Nauk*, volume 269, 543–547. Russian Academy of Sciences.

O'donoghue, B.; and Candes, E. 2015. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15: 715–732.

Stellato, B.; Banjac, G.; Goulart, P.; Bemporad, A.; and Boyd, S. 2020. OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12(4): 637–672.

Teixeira, A.; Ghadimi, E.; Shames, I.; Sandberg, H.; and Johansson, M. 2015. The ADMM algorithm for distributed quadratic problems: parameter selection and constraint preconditioning. *IEEE Transactions on Signal Processing*, 64(2): 290–305.

Thorley, A.; Jia, X.; Chang, H. J.; Liu, B.; Bunting, K.; Stoll, V.; de Marvao, A.; O'Regan, D. P.; Gkoutos, G.; Kotecha, D.; et al. 2021. Nesterov accelerated ADMM for fast diffeomorphic image registration. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, 150–160.

Wang, J.; Yu, F.; Chen, X.; and Zhao, L. 2019. Admm for efficient deep learning with global convergence. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 111–119.

Wohlberg, B. 2017. ADMM penalty parameter selection by residual balancing. *arXiv preprint arXiv:1704.06209*.

Xu, Z.; Figueiredo, M.; and Goldstein, T. 2017. Adaptive ADMM with spectral penalty parameter selection. In *Artificial Intelligence and Statistics*, 718–727. PMLR.

Yazaki, Y.; Tanaka, Y.; and Chan, S. H. 2019. Interpolation and denoising of graph signals using plug-and-play ADMM. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5431–5435.

Zhang, R.; Yan, K.; Li, G.; Jiang, T.; Li, X.; and Chen, H. 2020. Privacy-preserving decentralized power system economic dispatch considering carbon capture power plants and carbon emission trading scheme via over-relaxed ADMM. *International Journal of Electrical Power & Energy Systems*, 121: 106094.

Zhou, S.; and Li, G. Y. 2023. Federated learning via inexact ADMM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.