CLOSED-FORM PARAMETER SELECTION FOR ADMM AND OVER-RELAXED ADMM VIA SPECTRAL UPPER-ENVELOPE ANALYSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

The Alternating Direction Method of Multipliers (ADMM) and its over-relaxed variant (oADMM) are among the most effective optimization solvers in modern machine learning. However, their convergence performance depends critically on penalty and relaxation parameters that are challenging to tune properly. In practice, these parameters are typically selected through manual tuning, iterative procedures such as gradient-based search, or adaptive schemes such as residual balancing. Such approaches incur substantial computational overhead while offering limited theoretical guidance. To address this issue, we introduce a spectral upper-envelope framework, supported by strong theoretical guarantees, that provides analytical parameter-selection rules for ADMM and oADMM over a broad class of linear quadratic problems (LQPs). By formulating these algorithms as fixed-point iterations, we relate their convergence behavior to the spectral radius of the corresponding iteration matrix. This viewpoint leads to a tight and tractable upper bound, from which closed-form parameter choices can be derived, thereby eliminating the need for manual or iterative tuning. Extensive stochastic experiments show that the proposed parameters closely match optimal values obtained via gradient-descent search. In addition, experiments on two application domains, namely heterogeneous parabolic partial differential equations (PDEs) and diffeomorphic image registration, demonstrate that the proposed framework enables ADMM and oADMM to achieve faster convergence than state-of-the-art iterative solvers. Overall, our spectral upper-envelope framework provides a practical and theoretically grounded approach to parameter selection for ADMM and oADMM.

1 INTRODUCTION

In this paper, we consider the following regularized linear quadratic problem:

$$\min_{u} \ \frac{\mu}{2}\|Au - f\|_2^2 + \frac{1}{2}\|Lu\|_2^2, \tag{1}$$

where $\mu \in \mathbb{R}_{++}$ is the fidelity parameter, $A \in \mathbb{R}^{m\times n}$ (or $\mathbb{C}^{m\times n}$, with $m \leq n$) is the forward encoding operator, $u \in \mathbb{R}^n$ (or $\mathbb{C}^n$) is the unknown variable, $f \in \mathbb{R}^m$ denotes the observed data, and $L \in \mathbb{R}^{p\times n}$ (or $\mathbb{C}^{p\times n}$) is a regularization operator. Problems of this form arise ubiquitously in inverse problems (Yazaki et al., 2019), compressed sensing (Liu et al., 2023), regularized estimation (Chan et al., 2016), and can be interpreted as Gaussian maximum-a-posteriori (MAP) estimation (Chung et al., 2024). They are also closely connected to discretized parabolic partial differential equations (PDEs), widely used in engineering science, quantum mechanics, and financial mathematics (de Souza Santos & Ferreira, 2024).

The first-order optimality condition of equation 1 yields the normal equation: $(\mu A^\top A + L^\top L)u = \mu A^\top f$, whose direct inversion is often computationally undesirable. The system matrix $\mu A^\top A + L^\top L$ may fail to be positive definite, as in applications such as MRI reconstruction (Song et al., 2024), and can therefore be singular, so that its inverse does not exist. Even when the system is positive definite, direct solution methods become prohibitively expensive in large-scale settings, particularly in high-dimensional problems such as 3D image registration (Thorley et al., 2021).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

While iterative methods such as the conjugate gradient algorithm provide a practical alternative, they are generic Krylov subspace solvers and do not explicitly exploit the structure of $A^\top A$ and $L^\top L$. More recently, Song et al. (2024) proposed efficient ADMM-based algorithms that leverage this structure more directly. In particular, ADMM reformulations lead to subproblems involving shifted operators of the form $\mu A^\top A + \theta I$ and $L^\top L + \theta I$, which are positive definite for $\theta > 0$ and hence more amenable to efficient inversion. These subproblems can often be solved significantly faster by exploiting problem-specific structure, including multigrid methods (Nash, 2000), FFT- or DCT-based diagonalization (Lu et al., 2016; Duan et al., 2023), or closed-form updates via Sherman–Morrison–Woodbury identities (Vercauteren et al., 2009; Thorley et al., 2021). These solvers are typically pointwise in nature and can be efficiently parallelised using modern GPU-based frameworks such as PyTorch (Thorley et al., 2021; Duan et al., 2023).

Despite these advantages, the performance of ADMM and its over-relaxed variant (oADMM) (Eckstein & Bertsekas, 1992) remains highly sensitive to the choice of the penalty and relaxation parameters. In particular, these parameters can strongly affect the spectral radius of the underlying fixed-point iteration and thus the practical convergence speed. However, their effective and efficient selection remains nontrivial. In this work, we address this issue by developing an analytical parameter-selection framework for ADMM and oADMM tailored to linear quadratic problems (LQPs). The main contributions of this paper are summarized as follows:

- We propose a novel spectral upper-envelope framework for selecting algorithmic parameters in ADMM and oADMM to solve the linear quadratic problem equation 1 efficiently in an iterative manner. Optimal parameter selection can be formulated as minimizing the spectral radius of the iteration matrix. Instead, our framework minimizes an upper bound of the spectral radius, yielding closed-form solutions and avoiding iterative search such as gradient descent (Song et al., 2024). As a result, the method is highly efficient while maintaining comparable accuracy.
- Within the spectral upper-envelope framework, a key theoretical contribution is a disk-based geometric characterization of the spectrum in the complex plane, which localizes the nonreal eigenvalues of the ADMM iteration matrix and extends to those of oADMM via an affine mapping. This geometric perspective enables a tight upper bound on the spectral radius.
- We apply ADMM and oADMM, together with the proposed spectral upper-envelope analysis, to two real-world applications. The first considers a general parabolic PDE, which we reformulate as a linear quadratic problem. This however leads to the inversion of $I + \mu L^T A^T A L$, precluding the direct use of ADMM/oADMM and the proposed framework. We therefore derive the dual formulation, yielding the inversion of $\mu A^T A + L^T L$ and thus enabling the application of ADMM/oADMM. The second application is diffeomorphic image registration, where we compare the proposed methods with several state-of-the-art iterative algorithms, demonstrating their superiority.

2 Related Work

Existing works on ADMM and oADMM parameter selection can be broadly grouped into two categories. The first focuses on convergence analysis and parameter tuning via spectral or fixed-point perspectives. Boley (2013) studied the local linear convergence of ADMM for quadratic and linear programs using matrix recurrences, highlighting the strong impact of penalty parameters on convergence behavior. Building on this viewpoint, Ghadimi et al. (2014); Song et al. (2024) reformulated ADMM as a fixed-point iteration and selected parameters by minimizing the spectral radius of the iteration matrix, establishing a foundation for spectral parameter tuning in quadratic settings. However, these approaches typically rely on numerical search or problem-specific design (Franca & Bento, 2016), and no closed-form rules are available for the regularized linear quadratic problem considered here, limiting both generality and practicality.

The second category focuses on adaptive or empirical update strategies. Boyd et al. (2011) proposed a residual-balancing rule for updating the ADMM penalty parameter by maintaining a prescribed relation between primal and dual residual norm (Goldstein et al., 2015), while Xu et al. (2017a) introduced an adaptive ADMM scheme based on Barzilai–Borwein spectral ideas. Related problem-dependent updates have also been studied by Mavromatis et al. (2020); Mhanna et al. (2018). More recently, McCann & Wohlberg (2024) developed a robust and simple penalty-selection strategy for ADMM, and Lozenski et al. (2026) proposed an adaptive multiparameter selection method for mul-

Algorithm 1: ADMM for LQPs	**Algorithm 2:** Over-relaxed ADMM for LQPs
Input: matrices A and L; parameters μ and θ	**Input:** matrices A and L; parameters μ, θ and β
Initialize: u^0 and b^0	**Initialize:** u^0 and b^0
Repeat:	**Repeat:**
$w^{k+1} = \arg\min_w \frac{1}{2}\|Lw\|^2 + \frac{\theta}{2}\|w - u^k - b^k\|^2$	$w^{k+1} = \arg\min_w \frac{1}{2}\|Lw\|^2 + \frac{\theta}{2}\|w - u^k - b^k\|^2$
$u^{k+1} = \arg\min_u \frac{\mu}{2}\|Au - f\|^2 + \frac{\theta}{2}\|w^{k+1} - u - b^k\|^2$	$\tilde{w}^{k+1} = \beta w^{k+1} + (1-\beta)u^k$
$b^{k+1} = b^k + u^{k+1} - w^{k+1}$	$u^{k+1} = \arg\min_u \frac{\mu}{2}\|Au - f\|^2 + \frac{\theta}{2}\|\tilde{w}^{k+1} - u - b^k\|^2$
	$b^{k+1} = b^k + u^{k+1} - \tilde{w}^{k+1}$
Until some stopping criterion is met.	**Until** some stopping criterion is met.

ticonstraint and multiblock ADMM. These works illustrate the continuing practical importance of reliable parameter tuning, but they typically rely on residual- or curvature-related quantities during the iterations and leave the underlying optimality mechanism largely implicit.

3 METHODOLOGY

In this section, we develop a spectral framework for one-shot parameter selection in ADMM and oADMM. The framework proceeds in several steps. First, we reformulate the ADMM/oADMM iterations as fixed-point systems, which facilitates the characterization of their convergence behavior in terms of the spectral radius of the corresponding iteration matrix. We then introduce two spectral envelope functions that determine an upper bound on the spectral radius. Finally, we describe how this spectral upper-envelope framework can be used to guide parameter selection in ADMM and oADMM.

3.1 FIXED-POINT ITERATION FORMULATION FOR ADMM AND OADMM

To apply ADMM or oADMM, we introduce an auxiliary variable $w \in \mathbb{R}^n$ and a dual variable $b \in \mathbb{R}^n$, and rewrite equation 1 in the form of the following scaled augmented Lagrangian:

$$\mathcal{L}_\theta(w, u, b) = \frac{\mu}{2}\|Au - f\|_2^2 + \frac{1}{2}\|Lw\|_2^2 + \frac{\theta}{2}\|w - u - b\|_2^2,$$

where $\theta \in \mathbb{R}_{++}$ denotes the penalty parameter. Algorithms 1 and 2 summarize the ADMM and oADMM optimization procedures for the augmented Lagrangian above. In Algorithm 1, the ADMM iterations can be equivalently expressed as a fixed-point iteration solely with respect to the primal variable u:

$$u^{k+1} = M_1 u^k + c,$$

where $c \in \mathbb{R}^n$ is a constant vector that does not affect the convergence rate, and the iteration matrix M_1 is given by

$$M_1 = I + Q, \text{ where } Q = \theta(\mu A^\top A + \theta I)^{-1}\Big((L^\top L + \theta I)^{-1}(\theta I - \mu A^\top A) - I\Big). \tag{2}$$

Similarly, the oADMM iterations in Algorithm 2 also admit a fixed-point representation, but with a different iteration matrix given by

$$M_2 = I + \beta Q = (1 - \beta)I + \beta M_1, \tag{3}$$

where $\beta \in \mathbb{R}_{++}$ is a relaxation parameter. From the standard fixed-point iteration theory (Ghadimi et al., 2014; Song et al., 2024), the asymptotic convergence of both ADMM and oADMM is governed by the spectral radii of their iteration matrices (i.e., M_1 and M_2). Consequently, optimal parameter selection can be formulated as a spectral-radius minimization problem over M_1 and M_2.

3.2 SPECTRAL UPPER- AND LOWER-ENVELOPE FUNCTIONS

To establish explicit parameter selection rules, we analyze the spectra of the iteration matrices M_1 and M_2 via the matrices $Q_A := \mu A^\top A$ and $Q_L := L^\top L$, which are assumed to be symmetric posi-

tive definite.[1] Let $\alpha \in [\alpha_1, \alpha_n]$ and $\ell \in [\ell_1, \ell_n]$ denote the eigenvalues of Q_A and Q_L, respectively. We then define the rectangular spectral domain:

$$\Omega := [\alpha_1, \alpha_n] \times [\ell_1, \ell_n].$$

For $\theta > 0$ and $(\alpha, \ell) \in \Omega$, we define

$$h_{\alpha,\ell}(\theta) := \frac{\alpha\ell + \theta^2}{(\alpha + \theta)(\ell + \theta)}.$$

Based on the family $\{h_{\alpha,\ell}(\theta)\}_{(\alpha,\ell)\in\Omega}$, we introduce the spectral upper- and lower-envelope functions as

$$S_{\max}(\theta) := \max_{(\alpha,\ell)\in\Omega} h_{\alpha,\ell}(\theta), \quad S_{\min}(\theta) := \min_{(\alpha,\ell)\in\Omega} h_{\alpha,\ell}(\theta).$$

These functions respectively represent the pointwise maximum and minimum of the family over the spectral domain $(\alpha, \ell) \in \Omega$. The next proposition shows that these extrema are attained at four corners of Ω.

Proposition 1. *For every $\theta > 0$, let $\mathcal{V}$ denote the set of four corner vertices of the spectral domain Ω,*

$$\mathcal{V} := \{(\alpha_1, \ell_1), (\alpha_1, \ell_n), (\alpha_n, \ell_1), (\alpha_n, \ell_n)\}.$$

Then, it holds that

$$S_{\max}(\theta) = \max_{(\alpha,\ell)\in\mathcal{V}} h_{\alpha,\ell}(\theta), \quad S_{\min}(\theta) = \min_{(\alpha,\ell)\in\mathcal{V}} h_{\alpha,\ell}(\theta). \tag{4}$$

Proof. The proof of Proposition 1 is deferred to Appendix A.1. □

The two envelope functions defined in equation 4 play a central role in determining the upper-bound functions used in the spectral analysis of ADMM and oADMM in the next sections.

3.3 Parameter Selection for ADMM

We first consider how to choose the penalty parameter θ for ADMM. This can be determined by solving the following optimization problem:

$$\theta^\star_{\text{ADMM}} = \arg\min_{\theta>0} \rho(M_1(\theta)), \tag{5}$$

where $\rho(M_1(\theta))$ denotes the spectral radius of the iteration matrix $M_1(\theta)$, which is given by

$$\rho(M_1(\theta)) = \max_{\lambda\in\sigma(M_1(\theta))} |\lambda|,$$

where $\sigma(M_1(\theta))$ denotes the spectrum (i.e., the set of all eigenvalues) of the iteration matrix $M_1(\theta)$. Since $M_1(\theta)$ is not necessarily symmetric, its eigenvalues may be complex. It is therefore important to understand the geometric distribution of $\sigma(M_1(\theta))$ in the complex plane. To this end, we establish the following lemma.

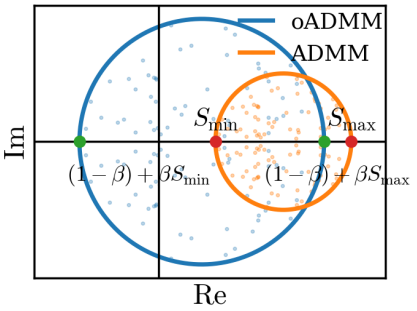


Figure 1: Geometric illustration of the spectra of $M_1(\theta)$ and $M_2(\theta)$ for a given θ. The orange and blue dots represent the complex eigenvalues of $M_1(\theta)$ and $M_2(\theta)$, respectively.

Lemma 2. *For every $\theta > 0$, the spectrum of the iteration matrix $M_1(\theta)$ is contained in a closed disk in the complex plane, i.e.,*

$$\sigma(M_1(\theta)) \subset \mathcal{D}_\theta,$$

where

$$\mathcal{D}_\theta := \left\{ z \in \mathbb{C} : \left| z - \frac{S_{\min}(\theta) + S_{\max}(\theta)}{2} \right| \le \frac{S_{\max}(\theta) - S_{\min}(\theta)}{2} \right\}. \tag{6}$$

In particular, the disk $\mathcal{D}_\theta$ is centered at the midpoint of $S_{\min}(\theta)$ and $S_{\max}(\theta)$, with radius equal to half of their difference.

[1] In the semidefinite case, the same analysis can be carried out on the positive spectral support by taking α_1 and ℓ_1 as the smallest positive eigenvalues of Q_A and Q_L, respectively. This excludes the zero-eigenvalue modes and yields a cleaner characterization of the nontrivial contraction behavior. For clarity, we focus here on the positive definite setting.

Proof. The proof is deferred to Appendix A.2. □

This lemma essentially states that all eigenvalues of $M_1(\theta)$ lie in a disk centered on the real axis (see Fig. 1, yellow disk). The fact that the disk is centered on the real axis is crucial, because the maximum modulus over this disk is attained at its intersection with the real axis, namely $S_{max}(\theta)$ and $S_{min}(\theta)$. The Lemma immediately yields the following upper-bound function for $\rho(M_1(\theta))$.

Theorem 3. *For every $\theta > 0$, the spectral radius of $M_1(\theta)$ satisfies*

$$\rho(M_1(\theta)) \leq S_{\max}(\theta).$$

Proof. By Lemma 2, the spectrum $\sigma(M_1(\theta))$ is contained in the closed disk equation 6. Since $S_{\min}(\theta) > 0$, this disk lies entirely in the right half-plane, and the point farthest from the origin is its rightmost endpoint $S_{\max}(\theta)$. The result follows. □

By Theorem 3, the parameter selection problem equation 5 for ADMM over a complex spectrum can be reduced to a scalar surrogate by minimizing $S_{\max}(\theta)$ with respect to θ:

$$\hat{\theta}^{\star}_{\mathrm{ADMM}} = \arg\min_{\theta>0} S_{\max}(\theta). \tag{7}$$

$\hat{\theta}^{\star}_{\mathrm{ADMM}}$ is an approximation of $\theta^{\star}_{\mathrm{ADMM}}$ in equation 5. Next, the theorem below shows that the optimal penalty parameter θ for equation 7 has a closed-form expression.

Theorem 4. *The minimizer of equation 7 is given by*

$$\hat{\theta}^{\star}_{\mathrm{ADMM}} = \begin{cases} \sqrt{\alpha_n \ell_1} & \alpha_n < \ell_1 \\ \sqrt{\alpha_1 \ell_n} & \ell_n < \alpha_1 \\ \theta_1 & \textit{otherwise}, \end{cases} \quad \text{where } \theta_1 = \sqrt{\frac{\alpha_n\alpha_1(\ell_n-\ell_1)+\ell_n\ell_1(\alpha_n-\alpha_1)}{(\alpha_n-\alpha_1)+(\ell_n-\ell_1)}}. \tag{8}$$

Proof. By Proposition 1, it suffices to compare the values of $h_{\alpha,\ell}(\theta)$ at the four corner vertices $\mathcal{V}$. The proof includes two cases, depending on whether the spectral intervals $[\alpha_1, \alpha_n]$ and $[\ell_1, \ell_n]$ are disjoint or overlapping. The full proof is deferred to Appendix A.3. □

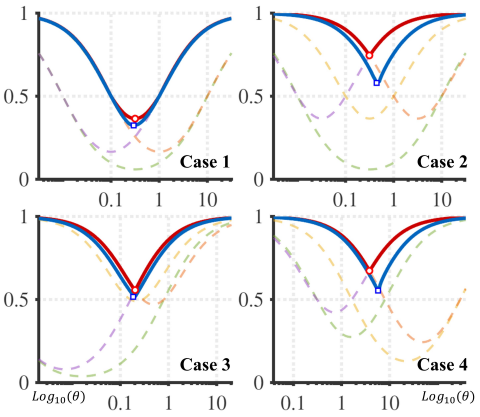


Figure 2: The spectral upper-envelope analysis characterizes the relationship between the true spectral radius and its upper bound for ADMM. In each subfigure, the solid red curve represents the upper-envelope function $S_{\max}(\theta)$, which bounds the true spectral radius, shown by the solid blue curve. The dashed curves in each subfigure correspond to $h_{\alpha,\ell}(\theta)$ evaluated at $\mathcal{V}$.

To illustrate Theorem 4, we present Fig. 2, where one can see that the minimizers of the proposed upper bounds closely approximate those of the true spectral radius functions.

3.4 Parameter Selection for oADMM

We next consider how to choose the parameters θ and β for oADMM. They can be found by tackling the following optimization problem:

$$(\theta^{\star}_{\mathrm{oADMM}}, \beta^{\star}_{\mathrm{oADMM}}) = \arg\min_{\theta>0,\ \beta>1} \rho(M_2(\theta,\beta)), \tag{9}$$

where $\rho(M_2(\theta,\beta))$ is the spectral radius of the iteration matrix $M_2(\theta,\beta)$. The relationship between the spectral radii of the iteration matrices $M_1(\theta)$ and $M_2(\theta,\beta)$ is affine, as shown in equation 3. Based on this, $\rho(M_2(\theta,\beta))$ can be expressed as:

$$\rho(M_2(\theta,\beta)) = \max_{\lambda\in\sigma(M_1(\theta))} |(1-\beta)+\beta\lambda|,$$

All eigenvalues of $M_2(\theta,\beta)$ lie in the blue disk in Fig. 1, the disk undergoes an affine transformation consisting of a scaling by β and a translation by $1-\beta$. As a result, the maximum module over the transformed disk can also be attained at its intersection with the real axis (see the green points in Fig. 1). This observation yields the following upper-bound function for $\rho(M_2(\theta,\beta))$.

Theorem 5. *For every $\theta > 0$ and $\beta > 1$,*

$$\rho(M_2(\theta,\beta)) \le J(\theta,\beta) := \max_{r\in\{S_{\min}(\theta),\, S_{\max}(\theta)\}} \left|(1-\beta)+\beta r\right|. \tag{10}$$

Proof. The proof follows directly from Lemma 2 by applying the affine transformation. □

By Theorem 5, the parameter selection problem for oADMM over a complex spectrum can be reduced to a scalar surrogate by minimizing $J(\theta,\beta)$ with respect to θ and β:

$$(\hat{\theta}^{\star}_{\text{oADMM}}, \hat{\beta}^{\star}_{\text{oADMM}}) = \arg\min_{\theta>0,\ \beta>0} J(\theta,\beta). \tag{11}$$

$(\hat{\theta}^{\star}_{\text{oADMM}}, \hat{\beta}^{\star}_{\text{oADMM}})$ are an approximation of $(\theta^{\star}_{\text{oADMM}}, \beta^{\star}_{\text{oADMM}})$ in equation 9. Next, the theorem below shows that the optimal relaxation parameter β for equation 11 has a closed-form expression for a given θ.

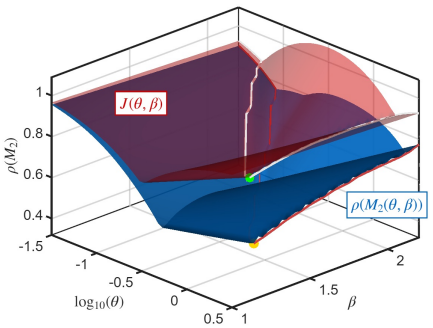


Figure 3: The relationship between the true spectral radius function $\rho(M_2(\theta,\beta))$ and its upper-bound function $J(\theta,\beta)$ in the context of oADMM. On the white curve superimposed on $J(\theta,\beta)$, each point is given by $\big(\theta, \hat{\beta}^*(\theta), S(\theta)\big)$, where $\hat{\beta}^*(\theta)$ is computed from equation 12 and $S(\theta)$ is evaluated from equation 14 for a given θ. On the red curve superimposed on $\rho(M_2(\theta,\beta))$, each point is given by $\big(\theta, \beta^*(\theta), \rho(M_2(\theta,\beta^*(\theta)))\big)$, where $\beta^*(\theta) = \arg\min_\beta \rho\big(M_2(\theta,\beta)\big)$ is obtained via gradient-descent search for the given θ, and $\rho(M_2(\theta,\beta^*(\theta)))$ is evaluated at the corresponding θ and $\beta^*(\theta)$.

Theorem 6. *For a given $\theta > 0$, the minimizer of $J(\theta,\beta)$ with respect to β is given by*

$$\hat{\beta}^{\star}_{\text{oADMM}}(\theta) = \frac{2}{2-\big(S_{\min}(\theta)+S_{\max}(\theta)\big)}. \tag{12}$$

Proof. Equation 12 was first established in (Song et al., 2024), but only for the real-valued case. In this paper, we extend the analysis to the complex-valued setting, yielding a more general result. □

By substituting equation 12 into $J(\theta,\beta)$ in equation 10, the joint optimization problem equation 11 can be reformulated as:

$$\hat{\theta}^{\star}_{\text{oADMM}} = \arg\min_{\theta>0} S(\theta), \tag{13}$$

where $S(\theta)$ is given by

$$S(\theta) = \frac{S_{\max}(\theta) - S_{\min}(\theta)}{2-\big(S_{\max}(\theta)+S_{\min}(\theta)\big)}. \tag{14}$$

Technically speaking, the joint optimization problem equation 11 over θ and β is equivalent to the optimization problem equation 13 over θ alone, with the optimal β given by equation 12. Therefore, one only needs to focus on finding a solution to equation 13, which we show next.

For a given θ, $S_{\min}(\theta)$ and $S_{\max}(\theta)$ can be obtained at the four corner vertices of the spectral domain Ω (i.e., $\mathcal{V}$). Consequently, $S(\theta)$ defined in equation 14 can be also expressed entirely using these corner vertices. For any two distinct vertices $(\alpha_i,\ell_j) \neq (\alpha_k,\ell_q) \in \mathcal{V}$, we define the following function:

$$\phi_{(i,j),(k,q)}(\theta) := \frac{\left|h_{\alpha_i,\ell_j}(\theta) - h_{\alpha_k,\ell_q}(\theta)\right|}{2-\big(h_{\alpha_i,\ell_j}(\theta)+h_{\alpha_k,\ell_q}(\theta)\big)}.$$

Since $S_{\max} - S_{\min} > 0$, we add the absolute value in the numerator to account for the ordering of the two vertices. As a result, $\phi_{(i,j),(k,q)} = \phi_{(k,q),(i,j)}$. Therefore, there are $\binom{4}{2} = 6$ distinct cases for ϕ.

Proposition 7. *For every $\theta > 0$, the upper-envelope function $S(\theta)$ in equation 14, restricted to the four corner vertices, is given as follows:*

$$S(\theta) = \max_{\substack{(i,j)\in\{1,n\}^2 \\ (k,q)\in\{1,n\}^2}} \phi_{(i,j),(k,q)}(\theta),$$

which implies that the upper-bound function of the spectral radius can be expressed as the upper envelope of the six distinct functions of ϕ.

Proof. The proof is straightforward. Since $S(\theta)$ is monotonically increasing in $S_{\max}(\theta)$ and monotonically decreasing in $S_{\min}(\theta)$, it follows that $S(\theta)$ is attained at the upper envelope of the six functions of ϕ. □

Based on Proposition 7, the theorem below restricts the minimizer of $S(\theta)$ to a finite candidate set.

Theorem 8. *The global minimizer of equation 13 lies in the candidate set, given as follows:*

$$\mathcal{C} = \Big\{ \min(\alpha_n, \ell_n),\ \theta_1,\ \max(\alpha_1, \ell_1),\ \theta_2 \Big\}$$

where θ_1 is characterized by $\phi_{(1,1),(n,n)}(\theta) = 0$, and given already in equation 8. θ_2 is the unique positive stationary point of $\phi_{(1,n),(n,1)}(\theta)$ whenever it exists, and therefore admits a closed-form expression obtained by solving a one-variable quadratic equation. Using this set, the optimal penalty parameter can be selected via:

$$\hat{\theta}^{\star}_{\text{oADMM}} = \arg\min_{\theta \in \mathcal{C}} S(\theta).$$

Proof. The proof is deferred to Appendix A.4. □

Theorem 8 establishes the result for oADMM. When combined with Theorem 6, it reduces the original tuning problem over (θ, β) to finite evaluatic function $S(\theta)$. Consequently, the penalty and relaxatio

Theorem 8 establishes the result for oADMM. When combined with Theorem 6, it reduces the original tuning problem over (θ, β) to finite evaluations over the candidate set $\mathcal{C}$ in closed-form function $S(\theta)$. Consequently, the penalty and relaxation parameters can be determined efficiently.

4 Numerical Experiments

In this section, we evaluate the effectiveness and generalization ability of the proposed spectral upper-envelope methods under several scenarios. We begin by assessing the accuracy and robustness of the methods using random instantiations of A and L. We then apply the methods to real-world problems, spanning heterogeneous parabolic PDEs and diffeomorphic image registration.

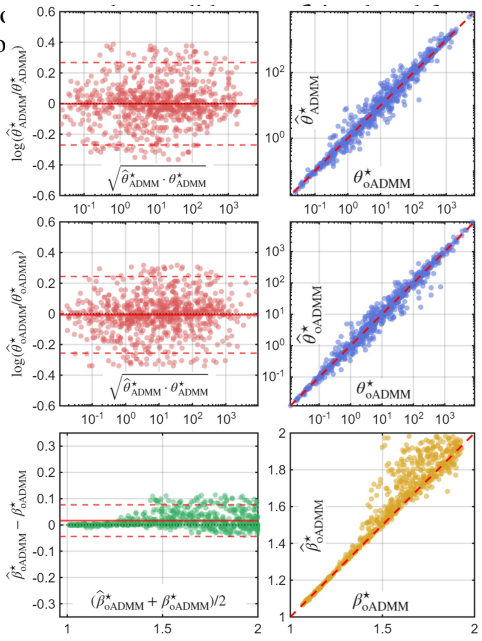


Figure 4: Bland-Altman and parity plots for ADMM and oADMM parameters.

Stochastic Instantiations: To evaluate the accuracy of the proposed parameter selection rules, we conduct 1000 stochastic comparisons against gradient-descent-based search methods (Song et al., 2024). In each trial, matrices $A \in \mathbb{R}^{100\times 50}$ and $L \in \mathbb{R}^{100\times 50}$ are randomly instantiated, and the resulting parameters from both approaches are compared. Figure 4 presents the corresponding Bland–Altman and parity plots, with quantitative results summarized in Table 1. Both ADMM and oADMM exhibit consistent behavior: the Bland–Altman plots show small bias with moderate dispersion around zero, while the parity plots indicate close agreement with the diagonal across several orders of magnitude. For oADMM, the relaxation parameter β also remains closely aligned with the diagonal, with only a slight positive bias observed. Overall, the discrepancies remain small relative to the parameter range across 1000 trials, indicating that the proposed analytical rules are robust and closely approximate search-based optimal parameters. At the same time, they avoid the computational cost of iterative tuning, which requires repeated evaluation of the spectral radius of the iteration matrix.

Linear Heterogeneous Parabolic PDEs: We consider a class of diffusion-type parabolic PDEs and show that implicit time discretization leads naturally to linear quadratic optimization problems of the form studied in this work. This demonstrates that the proposed framework applies beyond synthetic

Table 1: Statistical comparison of proposed and search-based parameter selection methods over 1,000 random tests, using log-scale metrics for θ and linear-scale metrics for β.

	Bland–Altman analysis				Parity analysis	
Parameter	Bias	Std. dev.	95% LoA-L	95% LoA-U	Corr.	RMSE
ADMM θ	-0.001	0.137	-0.270	0.268	0.985	0.176
oADMM θ	-0.007	0.128	-0.257	0.243	0.990	0.163
oADMM β	0.016	0.031	-0.044	0.077	0.953	0.091

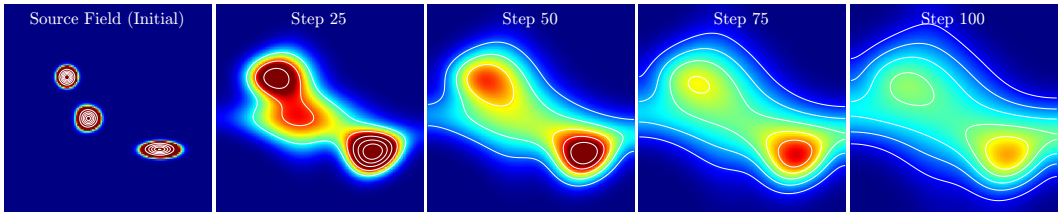


Figure 5: A representative diffusion example illustrating the spread of three instantaneously released Gaussian sources in a heterogeneous medium $k(x)$. The source term f is active only at the initial time step and vanishes for all $t > 0$, representing an instantaneous pulse-like release.

matrix settings to PDE-constrained problems arising in applications such as heat conduction, porous media flow, and subsurface transport (Su et al., 2022; Zhang et al., 2023; Cavalcante et al., 2024).

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain, and let $u : \Omega \times [0, T] \to \mathbb{R}$ and $f : \Omega \times [0, T] \to \mathbb{R}$. We consider

$$\frac{\partial u(x,t)}{\partial t} = \nabla \cdot \big(k(x)\nabla u(x,t)\big) + f(x,t), \tag{15}$$

where $k(x) > 0$. Applying a backward Euler discretization with step size $\tau > 0$ to the PDE yields an implicit update for u^{k+1}, which we discretize below.

Assume a discretized 2D regular grid with $n = N_x N_y$ nodes, and let $u \in \mathbb{R}^n$ denote the discrete state and $f^{k+1} \in \mathbb{R}^n$ the discretized forcing term. Next, we define the discrete gradient operator $D = [D_x\ D_y]^\top \in \mathbb{R}^{2n\times n}$, where $D_x, D_y \in \mathbb{R}^{n\times n}$ are spatial finite difference operators. The corresponding discrete divergence operator is given by $-D^\top$. Next, let $K \in \mathbb{R}^{2n\times 2n}$ be a diagonal matrix with strictly positive entries corresponding to the discretized diffusion coefficient $k(x)$ applied componentwise to the gradient field. The backward Euler step leads to the linear system

$$\big(I + \tau D^\top K D\big)u^{k+1} = u^k + \tau f^{k+1}, \tag{16}$$

which is equivalently characterized as the optimality condition of a quadratic problem (see Appendix A.5), consisting of a temporal fidelity term, a diffusion regularizer $\|K^{1/2}Du\|_2^2$, and a linear forcing term. This primal formulation requires solving linear systems involving the inversion of $I + \tau D^\top K D$, which has the structure $\mu I + A^\top L^\top L A$. Such a form precludes the direct application of ADMM/oADMM and the proposed framework.

For this, we derive the dual formulation. Introducing the auxiliary variable $v \in \mathbb{R}^{2n}$ with $v = Du$, and the dual variable $p \in \mathbb{R}^{2n}$, we derive the dual problem (see Appendix A.5) as

$$\min_{p\in\mathbb{R}^{2n}} \ \frac{\tau}{2}\|D^\top p + f^{k+1}\|_2^2 + \frac{1}{2}\|K^{-1/2}p\|_2^2 + \langle D^\top p + f^{k+1},\, u^k\rangle. \tag{17}$$

In contrast to the primal, the dual formulation involves systems of the form $\tau DD^\top + K^{-1}$, corresponding to $\mu A^\top A + L^\top L$. This structure is compatible with ADMM/oADMM and enables efficient splitting within the proposed framework.[2]

In the numerical experiment shown in Fig. 5, we compare ADMM and oADMM, equipped with our proposed parameter-selection rules, against standard iterative solvers: gradient descent (GD), Nesterov-accelerated gradient descent (GD-N) (Nesterov, 1983; Bartlett & Duan, 2021), Nesterov-accelerated gradient descent with restart (GD-NR) (O'donoghue & Candes, 2015; Bartlett & Duan,

[2]Notably, both resulting subproblems in ADMM/oADMM admit efficient solvers: one reduces to diagonal matrix inversion, while the other can be solved efficiently using the discrete cosine transform (DCT) under the Neumann boundary conditions adopted in this work.

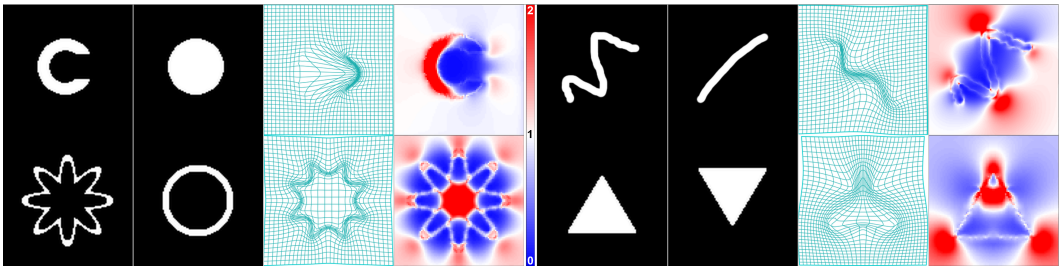


Figure 6: Visualization of large deformations on four image pairs. For each case the columns show the source image, target image, mesh grid, and Jacobian determinant, respectively.

2021), conjugate gradient (CG), and preconditioned CG (PCG) with incomplete Cholesky factorization. Step sizes for gradient-based methods are set using the Lipschitz constant of the objective. As shown in Fig. 7 (left), both ADMM and oADMM achieve competitive convergence, with oADMM among the top-performing methods.

Diffeomorphic Image Registration: Computing a diffeomorphic deformation can be treated as modelling a dynamical system (Beg et al., 2005), described by the ordinary differential equation (ODE) $\partial\boldsymbol{\phi}_t/\partial t = \boldsymbol{v}_t(\boldsymbol{\phi}_t)$, where $\boldsymbol{\phi}_0 = \mathrm{Id}$ is the identity transformation and $\boldsymbol{v}_t$ denotes the velocity field at time $t \in [0,1]$. The ODE can be solved via Euler integration, where the deformation field $\boldsymbol{\phi}$ is obtained as a composition of small deformations, namely $\boldsymbol{\phi} = (\mathrm{Id} + \frac{1}{N}\boldsymbol{v}_{t_{N-1}}) \circ \cdots \circ (\mathrm{Id} + \frac{1}{N}\boldsymbol{v}_{t_1}) \circ (\mathrm{Id} + \frac{1}{N}\boldsymbol{v}_{t_0})$. If the velocity fields $\boldsymbol{v}_{t_i}$ are sufficiently small and satisfy suitable smoothness constraints, the resulting composition is diffeomorphic. To estimate the velocity fields under these constraints, we consider the following linear quadratic problem (Thorley et al., 2021):

$$\min_{v_x, v_y} \ \frac{\mu}{2}\|I_x \odot v_x + I_y \odot v_y + I_t\|_2^2 + \frac{1}{2}\|Dv_x\|_2^2 + \frac{1}{2}\|Dv_y\|_2^2, \tag{18}$$

where $I_x, I_y \in \mathbb{R}^n$ are the discrete spatial image derivatives, $I_t \in \mathbb{R}^n$ is the temporal derivative, $v_x, v_y \in \mathbb{R}^n$ are the velocity components, $\odot$ denotes pointwise multiplication, and $D \in \mathbb{R}^{2n\times n}$ is the discrete gradient on the 2D grid.

With $v = (v_x, v_y) \in \mathbb{R}^{2n}$, $A = [\mathrm{diag}(I_x)\ \mathrm{diag}(I_y)] \in \mathbb{R}^{n\times 2n}$, and $L = \mathrm{diag}(D^\top D, D^\top D)$, the objective becomes $\frac{\mu}{2}\|Av + I_t\|_2^2 + \frac{1}{2}\|L^{1/2}v\|_2^2$ and therefore fits the proposed quadratic framework.[3]

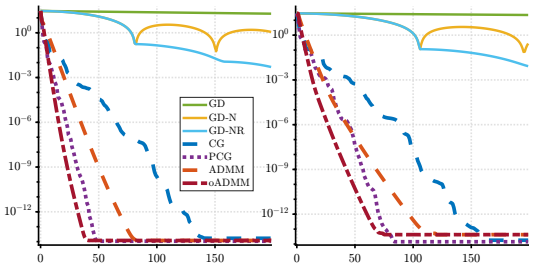


Figure 7: Convergence rates of different methods for the PDE (left) and the image registration problem (right). x-axis: iteration count; y-axis: $\log L^2$ error between iterates and the ground truth.

Against the same iterative solvers, oADMM achieves the fastest convergence in the image-registration experiment shown in Fig. 7 (right). Figure 6 presents four image pairs to illustrate the computation of large diffeomorphic deformation fields, together with the corresponding Jacobian determinants, which remain nonnegative.

5 Conclusion

In this paper, we developed an analytical parameter-selection framework for ADMM and oADMM applied to linear quadratic problems. Based on a spectral upper-envelope analysis, the method yields a closed-form piecewise characterization of the ADMM penalty parameter and a finite parameter candidate set for oADMM. We validated the proposed rules through extensive stochastic experiments and representative applications, showing that the analytically predicted parameters closely match spectral-radius-optimal choices across several settings. In particular, applications to inhomogeneous parabolic PDEs and diffeomorphic image registration demonstrate both the practicality of the framework and its ability to accelerate convergence compared to existing iterative methods.

[3]The two ADMM/oADMM subproblems use a rank-one inversion and a DFT solver under periodic boundary conditions.

AI USE STATEMENT

Generative AI tools were used to assist with language editing. All AI-assisted suggestions were manually checked by the authors. The authors take full responsibility for all mathematical claims, proofs, and the final content of the paper.

REFERENCES

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

M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.

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

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

T. M. Cavalcante et al. Simulation of two-phase flow in 3D fractured reservoirs using a projection-based embedded discrete fracture model on unstructured tetrahedral grids (pedfm-u). *Advances in Water Resources*, 188:104703, 2024.

Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1): 84–98, 2016.

Julianne Chung, Jiahua Jiang, Scot M Miller, and Arvind K Saibaba. Hybrid projection methods for solution decomposition in large-scale bayesian inverse problems. *SIAM Journal on Scientific Computing*, 46(2):S97–S119, 2024.

Daniel de Souza Santos and Tiago Alessandro Espinola Ferreira. Neural network learning of Black-Scholes equation for option pricing, 2024. URL `https://arxiv.org/abs/2405.05780`.

Jinming Duan, Xi Jia, Joseph Bartlett, Wenqi Lu, and Zhaowen Qiu. Arbitrary order total variation for deformable image registration. *Pattern Recognition*, 137:109318, 2023.

Jonathan Eckstein and Dimitri P Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1): 293–318, 1992.

Guilherme Franca and José Bento. An explicit rate bound for over-relaxed ADMM. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2104–2108. IEEE, 2016. doi: 10.1109/ISIT.2016.7541670.

Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson. Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems. *IEEE Transactions on Automatic Control*, 60(3):644–658, 2014.

Tom Goldstein, Min Li, and Xiaoming Yuan. Adaptive primal-dual splitting methods for statistical learning and image processing. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2089–2097. Curran Associates, Inc., 2015.

Yishun Liu, Keke Huang, Chunhua Yang, and Zhen Wang. Distributed network reconstruction based on binary compressed sensing via ADMM. *IEEE Transactions on Network Science and Engineering*, 10(4):2141–2153, 2023.

Luke Lozenski, Michael T McCann, and Brendt Wohlberg. An adaptive multiparameter penalty selection method for multiconstraint and multiblock ADMM. *IEEE Open Journal of Signal Processing*, 2026.

Wenqi Lu, Jinming Duan, Zhaowen Qiu, Zhenkuan Pan, Ryan Wen Liu, and Li Bai. Implementation of high-order variational models made easy for image processing. *Mathematical Methods in the Applied Sciences*, 39(14):4208–4233, 2016.

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

Michael T McCann and Brendt Wohlberg. Robust and simple ADMM penalty parameter selection. *IEEE Open Journal of Signal Processing*, 5:402–420, 2024.

Sleiman Mhanna, Gregor Verbič, and Archie C Chapman. Adaptive ADMM for distributed ac optimal power flow. *IEEE Transactions on Power Systems*, 34(3):2025–2035, 2018.

Stephen G Nash. A multigrid approach to discretized optimization problems. *Optimization Methods and Software*, 14(1-2):99–116, 2000.

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

Brendan O'donoghue and Emmanuel Candes. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15:715–732, 2015.

Jintao Song, Wenqi Lu, Yunwen Lei, Yuchao Tang, Zhenkuan Pan, and Jinming Duan. Optimizing ADMM and over-relaxed ADMM parameters for linear quadratic problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8117–8125, 2024.

Danyang Su, Mingliang Xie, Klaus Ulrich Mayer, and Kerry T. B. MacQuarrie. Simulation of diffusive solute transport in heterogeneous porous media with dipping anisotropy. *Frontiers in Water*, 4:974145, 2022.

Alexander Thorley, Xi Jia, Hyung Jin Chang, Boyang Liu, Karina Bunting, Victoria Stoll, Antonio de Marvao, Declan P O'Regan, Georgios Gkoutos, Dipak Kotecha, et al. Nesterov accelerated ADMM for fast diffeomorphic image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 150–160. Springer, 2021.

Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.

Brendt Wohlberg. Admm penalty parameter selection by residual balancing. *arXiv preprint arXiv:1704.06209*, 2017.

Zheng Xu, Mario Figueiredo, and Tom Goldstein. Adaptive ADMM with spectral penalty parameter selection. In *Artificial Intelligence and Statistics*, pp. 718–727. PMLR, 2017a.

Zheng Xu, Mario AT Figueiredo, Xiaoming Yuan, Christoph Studer, and Tom Goldstein. Adaptive relaxed admm: Convergence theory and practical implementation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7234–7243. IEEE, 2017b.

Yoshinao Yazaki, Yuichi Tanaka, and Stanley H Chan. 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)*, pp. 5431–5435. IEEE, 2019.

Shi-Ting Zhang, Yang Hu, Qiang He, and Qian-Ping Li. A diffuse interface–lattice Boltzmann model for conjugate heat transfer with imperfect interface. *Computers & Mathematics with Applications*, 154:174–190, 2023.

A Appendix

A.1 Proof of Proposition 1

We consider the scalar spectral function

$$h_{\alpha,\ell}(\theta) \;=\; h(\alpha,\ell,\theta) \;:=\; \frac{\alpha\ell+\theta^2}{(\alpha+\theta)(\ell+\theta)}, \quad (\alpha,\ell)\in\Omega, \quad \theta>0,$$

where the spectral domain is defined by the rectangle

$$\Omega=[\alpha_1,\alpha_n]\times[\ell_1,\ell_n].$$

The associated upper and lower spectral envelopes are given by

$$S_{\max}(\theta) \;:=\; \max_{(\alpha,\ell)\in\Omega} h(\alpha,\ell,\theta), \quad S_{\min}(\theta) \;:=\; \min_{(\alpha,\ell)\in\Omega} h(\alpha,\ell,\theta).$$

For subsequent analysis, we record the partial derivatives of h:

$$\partial_\alpha h(\alpha,\ell,\theta)=\frac{\theta(\ell-\theta)}{(\alpha+\theta)^2(\ell+\theta)}, \quad \partial_\ell h(\alpha,\ell,\theta)=\frac{\theta(\alpha-\theta)}{(\ell+\theta)^2(\alpha+\theta)},$$

and

$$\partial_\theta h(\alpha,\ell,\theta)=\frac{(\alpha+\ell)(\theta^2-\alpha\ell)}{(\alpha+\theta)^2(\ell+\theta)^2}.$$

In particular, the signs of these derivatives admit the simple characterizations

$$\operatorname{sign}\big(\partial_\alpha h(\alpha,\ell,\theta)\big)=\operatorname{sign}(\ell-\theta), \quad \operatorname{sign}\big(\partial_\ell h(\alpha,\ell,\theta)\big)=\operatorname{sign}(\alpha-\theta), \tag{19}$$

and

$$\operatorname{sign}\big(\partial_\theta h(\alpha,\ell,\theta)\big)=\operatorname{sign}(\theta^2-\alpha\ell).$$

These relations show that, for any fixed $\theta>0$, the function $h(\alpha,\ell,\theta)$ is monotone in each variable separately. As a consequence, its extrema over the rectangular domain Ω are attained at the boundary, and in fact at the four corner points of Ω. These derivative identities and monotonicity properties will be used repeatedly in the subsequent analysis.

A.2 Proof of Lemma 2

Recall that the ADMM iteration matrix

$$M_1=I+Q, \quad Q=\theta(Q_A+\theta I)^{-1}\big((Q_L+\theta I)^{-1}(\theta I-Q_A)-I\big),$$

where $Q_L=L^\top L$ and $Q_A=A^\top A$. We construct

$$I+2Q=(Q_A+\theta I)^{-1}(Q_L+\theta I)^{-1}(\theta I-Q_L)(\theta I-Q_A). \tag{20}$$

We then define two real symmetric Cayley-type matrices

$$R_A:=(\theta I-Q_A)(Q_A+\theta I)^{-1}, \quad R_L:=(Q_L+\theta I)^{-1}(\theta I-Q_L). \tag{21}$$

By the invariance of the spectrum under cyclic permutations of matrix products, the matrix in equation 20 is isospectral to R_AR_L, and hence

$$\sigma(I+2Q)=\sigma(R_AR_L).$$

Since $M_1=I+Q$ and $M_2=(1-\beta)I+\beta M_1$, it follows that

$$\sigma(M_1)=\sigma\big(\tfrac{1}{2}(I+R_AR_L)\big), \;\; \sigma(M_2)=\sigma((1-\beta)I+\beta M_1) \tag{22}$$

The problem reduces to analyzing the spectrum of R_AR_L. Since R_A and R_L are symmetric, their eigenvalues are real, but the spectrum of their product need not be real. Moreover, $\sigma(R_AR_L)=\sigma(R_LR_A)$, so the two factors can be interchanged without changing the spectrum.

If either factor is definite, the product is similar to a symmetric matrix, and the desired spectral inclusion follows from standard symmetric factorization and Rayleigh quotient bounds. When neither factor is definite, we first consider the case where both are nonsingular and hence indefinite, treating real and nonreal eigenvalues separately. We then handle singular cases by a nullspace reduction and a continuity argument. The following theorem establishes the spectral inclusion for arbitrary symmetric factors, from which Lemma 2 follows.

Theorem 9. *Let $R_A, R_L \in \mathbb{R}^{n\times n}$ be symmetric, with ordered eigenvalues $a_1 \le \cdots \le a_n$ and $b_1 \le \cdots \le b_n$, respectively. Define*

$$\sigma_{\min} := \min_{1\le i,j\le n} a_i b_j = \min\{a_1b_1, a_1b_n, a_nb_1, a_nb_n\},$$
$$\sigma_{\max} := \max_{1\le i,j\le n} a_i b_j = \max\{a_1b_1, a_1b_n, a_nb_1, a_nb_n\}.$$

No restriction is imposed on the inertia. Then

$$\sigma(R_AR_L) \subseteq \mathcal{D}[\sigma_{\min}, \sigma_{\max}],$$

where, for $s_- \le s_+$,

$$\mathcal{D}[s_-, s_+] := \left\{ z \in \mathbb{C} : \left| z - \frac{s_- + s_+}{2} \right| \le \frac{s_+ - s_-}{2} \right\}$$

is the closed disk with diameter $[s_-, s_+]$.

Proof. Suppose first that one factor is definite. If $R_A \succ 0$, then

$$R_AR_L \sim R_A^{1/2} R_L R_A^{1/2}.$$

If $R_A \prec 0$, writing $H = -R_A \succ 0$ gives

$$R_AR_L \sim -H^{1/2} R_L H^{1/2}.$$

Thus the product is similar to a symmetric matrix, and Rayleigh–Ritz gives the claimed interval. The case of definite R_L is analogous.

Next, we consider the case in which neither R_A nor R_L is definite. There are two scenarios according to singularity: 1) both R_A and R_L are nonsingular; 2) at least one of them is singular. If both R_A and R_L are nonsingular, then, since neither matrix is definite, both must be indefinite. Cases 1 and 2 handle their real and nonreal eigenvalues, respectively. Case 3 handles the second scenario.

First suppose both factors are nonsingular and hence indefinite. In an orthogonal eigenbasis of R_A, write

$$R_A = \begin{pmatrix} -D_{00} & 0 \\ 0 & D_{11} \end{pmatrix}, \qquad R_L = \begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix},$$

where $D_{00}, D_{11} \succ 0$. Since R_L is symmetric, $B_{10} = B_{01}^\top$. Let

$$p = \lambda_{\max}(D_{00}) = -a_1, \qquad q = \lambda_{\max}(D_{11}) = a_n, \qquad \beta_1 = b_1 < 0 < b_n = \beta_n.$$

Case 1: Real eigenvalues.

For real eigenvalue $\lambda \in \mathbb{R}$,

$$R_AR_L - \lambda I = R_A\left(R_L - \lambda R_A^{-1}\right) = R_AK_\lambda,$$

so $\lambda \in \sigma(R_AR_L)$ if and only if

$$K_\lambda = \begin{pmatrix} B_{00} + \lambda D_{00}^{-1} & B_{01} \\ B_{10} & B_{11} - \lambda D_{11}^{-1} \end{pmatrix}$$

is singular. We have

$$\beta_1 I \preceq B_{00} \preceq \beta_n I, \quad \beta_1 I \preceq B_{11} \preceq \beta_n I, \quad D_{00}^{-1} \succeq p^{-1} I, \quad D_{11}^{-1} \succeq q^{-1} I,$$

and

$$\sigma_{\max} = \max\{-p\beta_1, q\beta_n\}, \qquad \sigma_{\min} = \min\{-p\beta_n, q\beta_1\}.$$

If $\lambda > \sigma_{\max}$, define

$$A_\lambda := B_{00} + \lambda D_{00}^{-1} \succeq (\beta_1 + \lambda/p) I \succ 0,$$

and

$$D_\lambda := B_{11} - \lambda D_{11}^{-1} \preceq (\beta_n - \lambda/q) I \prec 0.$$

Since A_λ is the positive upper-left block of K_λ, its Schur complement is

$$S_\lambda = D_\lambda - B_{10}A_\lambda^{-1}B_{01} \preceq D_\lambda \prec 0.$$

Consequently,

$$\det K_\lambda = \det(A_\lambda)\det(S_\lambda) \neq 0,$$

so K_λ is nonsingular.

If $\lambda < \sigma_{\min}$, then

$$A_\lambda := B_{00} + \lambda D_{00}^{-1} \preceq (\beta_n + \lambda/p)I \prec 0,$$

and

$$D_\lambda := B_{11} - \lambda D_{11}^{-1} \succeq (\beta_1 - \lambda/q)I \succ 0.$$

Since D_λ is the positive lower-right block of K_λ, its Schur complement is

$$\widetilde{S}_\lambda = A_\lambda - B_{01}D_\lambda^{-1}B_{10} \preceq A_\lambda \prec 0.$$

Thus $\det K_\lambda = \det(D_\lambda)\det(\widetilde{S}_\lambda) \neq 0$, and K_λ is again nonsingular.

Therefore every real eigenvalue must lie in $[\sigma_{\min}, \sigma_{\max}]$.

Case 2: Nonreal eigenvalues.

Let $\lambda \in \sigma(R_A R_L) \setminus \mathbb{R}$ and let $v \in \mathbb{C}^n \setminus \{0\}$ be a corresponding eigenvector. Then

$$R_L v = \lambda R_A^{-1} v.$$

Left-multiplying by v^* gives

$$v^* R_L v = \lambda v^* R_A^{-1} v.$$

Since both quadratic forms are real while λ is nonreal, this implies

$$v^* R_L v = v^* R_A^{-1} v = 0.$$

The spectral inequalities

$$(R_L - \beta_1 I)(R_L - \beta_n I) \preceq 0, \qquad (R_A^{-1} + p^{-1}I)(R_A^{-1} - q^{-1}I) \succeq 0$$

hold. Expanding these inequalities and using $v^* R_L v = v^* R_A^{-1} v = 0$ yields

$$v^* R_L^2 v \leq -\beta_1\beta_n \|v\|_2^2, \qquad v^* R_A^{-2} v \geq \frac{\|v\|_2^2}{pq}.$$

Since $\|R_L v\|_2^2 = |\lambda|^2 \|R_A^{-1} v\|_2^2$, we obtain

$$|\lambda|^2 \leq pq(-\beta_1\beta_n).$$

Moreover,

$$\sigma_{\max} = \max\{-p\beta_1, q\beta_n\} \geq \sqrt{pq(-\beta_1\beta_n)},$$

and

$$-\sigma_{\min} = \max\{p\beta_n, -q\beta_1\} \geq \sqrt{pq(-\beta_1\beta_n)}.$$

Hence

$$|\lambda| \leq r := \min\{\sigma_{\max}, -\sigma_{\min}\}.$$

The triangle inequality yields

$$\left|\lambda - \frac{\sigma_{\min} + \sigma_{\max}}{2}\right| \leq |\lambda| + \left|\frac{\sigma_{\min} + \sigma_{\max}}{2}\right| \leq r + \left|\frac{\sigma_{\min} + \sigma_{\max}}{2}\right| = \frac{\sigma_{\max} - \sigma_{\min}}{2}.$$

Therefore,

$$\lambda \in \mathcal{D}[\sigma_{\min}, \sigma_{\max}].$$

Case 3: Singularity.

Assume that R_A is singular. The simplest case is that R_A is a zero matrix. In this case, $R_A R_L$ is also a zero matrix, so all eigenvalues are zero, implying

$$\sigma(R_A R_L) \subseteq \mathcal{D}[\sigma_{\min}, \sigma_{\max}].$$

Otherwise, since orthogonal similarity preserves the spectrum of the product, we may work without loss of generality in an orthonormal basis and write

$$R_A = \begin{pmatrix} 0 & 0 \\ 0 & \widehat{R}_A \end{pmatrix}, \qquad R_L = \begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix},$$

where $\widehat{R}_A$ is symmetric and nonsingular. Then

$$R_A R_L = \begin{pmatrix} 0 & 0 \\ \widehat{R}_A B_{10} & \widehat{R}_A B_{11} \end{pmatrix},$$

so

$$\sigma(R_A R_L) = \{0\} \cup \sigma(\widehat{R}_A B_{11}).$$

Since B_{11} is a symmetric compression with $\sigma(B_{11}) \subseteq [b_1, b_n]$ and $\sigma(\widehat{R}_A) \subseteq [a_1, a_n]$, the reduced disk is contained in $\mathcal{D}[\sigma_{\min}, \sigma_{\max}]$.

If B_{11} is nonsingular, the preceding result applies directly. Otherwise, choose $\varepsilon_j \to 0$ so that $B_{11} + \varepsilon_j I$ is nonsingular, apply the result to $\widehat{R}_A(B_{11} + \varepsilon_j I)$, and pass to the limit by continuity of the characteristic roots and disk endpoints.

Since $0 \in [a_1, a_n]$, we have $\sigma_{\min} \leq 0 \leq \sigma_{\max}$, so the zero eigenvalues are also included. If only R_L is singular, apply the same argument to $R_L R_A$, noting that $\sigma(R_A R_L) = \sigma(R_L R_A)$. this completes the proof of Theorem 9. □

We now apply Theorem 9 to the matrix M_1 defined in Appendix A.2. Recall that R_A and R_L are defined in equation 21 by:

$$R_A := (\theta I - Q_A)(Q_A + \theta I)^{-1}, \quad R_L := (Q_L + \theta I)^{-1}(\theta I - Q_L).$$

Let the eigenvalues of Q_A and Q_L be $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\ell_1, \ldots, \ell_n\}$, respectively. For a given $\theta > 0$, define the smallest and largest pairwise products of the eigenvalues of R_A and R_L by:

$$\sigma_{\min}(\theta) := \min_{i,j} \left(\frac{\theta - \alpha_i}{\alpha_i + \theta} \cdot \frac{\theta - \ell_j}{\ell_j + \theta} \right), \quad \sigma_{\max}(\theta) := \max_{i,j} \left(\frac{\theta - \alpha_i}{\alpha_i + \theta} \cdot \frac{\theta - \ell_j}{\ell_j + \theta} \right).$$

By Theorem 9, the spectrum of $R_A R_L$ is contained in the disk

$$\sigma(R_A R_L) \subseteq \left\{ z \in \mathbb{C} : \left| z - \frac{\sigma_{\min}(\theta) + \sigma_{\max}(\theta)}{2} \right| \leq \frac{\sigma_{\max}(\theta) - \sigma_{\min}(\theta)}{2} \right\}.$$

Recall that

$$S_{\min}(\theta) := \min_{i,j} \frac{\alpha_i \ell_j + \theta^2}{(\alpha_i + \theta)(\ell_j + \theta)}, \quad S_{\max}(\theta) := \max_{i,j} \frac{\alpha_i \ell_j + \theta^2}{(\alpha_i + \theta)(\ell_j + \theta)}.$$

Then the following identities hold:

$$S_{\min}(\theta) = \frac{1 + \sigma_{\min}(\theta)}{2}, \quad S_{\max}(\theta) = \frac{1 + \sigma_{\max}(\theta)}{2}.$$

Consequently, since the spectrum of $M_1 = \frac{1}{2}(I + R_A R_L)$ can be obtained from $\sigma(R_A R_L)$ via the affine transformation $z \mapsto \frac{1}{2}(1 + z)$, we have

$$\sigma(M_1) = \sigma\big(\tfrac{1}{2}(I + R_A R_L)\big) \subseteq \left\{ z \in \mathbb{C} : \left| z - \frac{S_{\min}(\theta) + S_{\max}(\theta)}{2} \right| \leq \frac{S_{\max}(\theta) - S_{\min}(\theta)}{2} \right\}.$$

This completes the proof for Lemma 2 □

A.3 Proof of Theorem 4

To prove the theorem, we distinguish between two main cases: the disjoint case, where either $\alpha_n < \ell_1$ or $\ell_n < \alpha_1$, and the overlapping case, where $[\alpha_1, \alpha_n] \cap [\ell_1, \ell_n] \neq \emptyset$. These correspond to three possible configurations, each analyzed below.

Configuration 1: $\alpha_n < \ell_1$**.**

We first show that any minimizer $\hat{\theta}^\star_{\mathrm{ADMM}}$ must lie in the interval $[\alpha_n, \ell_1]$. For this, we first consider $\theta < \alpha_n$. From equation 19, for all $\ell \in [\ell_1, \ell_n]$,

$$\partial_\alpha h(\alpha, \ell, \theta) > 0,$$

so the maximum over α is attained at $\alpha = \alpha_n$. Evaluating at this point,

$$\partial_\ell h(\alpha_n, \ell, \theta) > 0,$$

which implies that the maximum over ℓ is attained at $\ell = \ell_n$. Hence,

$$S_{\max}(\theta) = h(\alpha_n, \ell_n, \theta).$$

Since $\theta < \alpha_n < \ell_n$, we have

$$\partial_\theta h(\alpha_n, \ell_n, \theta) < 0,$$

so $S_{\max}(\theta)$ is strictly decreasing on $(0, \alpha_n)$, and no minimizer lies in this region.

Next, we consider $\theta > \ell_1$. By an analogous argument,

$$S_{\max}(\theta) = h(\alpha_1, \ell_1, \theta),$$

and $S_{\max}(\theta)$ is strictly increasing on (ℓ_1, ∞). Hence, no minimizer lies in this region either.

Therefore, any minimizer must lie in $[\alpha_n, \ell_1]$. Restricting $\theta \in [\alpha_n, \ell_1]$, we have

$$\partial_\alpha h(\alpha, \ell, \theta) > 0, \quad \partial_\ell h(\alpha, \ell, \theta) < 0,$$

so

$$S_{\max}(\theta) = h(\alpha_n, \ell_1, \theta).$$

The minimizer is obtained from the first-order optimality condition

$$\partial_\theta h(\alpha_n, \ell_1, \theta) = 0,$$

which yields the closed-form solution

$$\hat{\theta}^\star_{\mathrm{ADMM}} = \sqrt{\alpha_n \ell_1}. \tag{23}$$

Configuration 2: $\ell_n < \alpha_1$**.**

The case is symmetric to Configuration 1. The corresponding closed-form solution is

$$\hat{\theta}^\star_{\mathrm{ADMM}} = \sqrt{\ell_n \alpha_1}. \tag{24}$$

Configuration 3: $[\alpha_1, \alpha_n] \cap [\ell_1, \ell_n] \neq \emptyset$**.**

We first define

$$l := \max(\alpha_1, \ell_1), \quad r := \min(\alpha_n, \ell_n),$$

so that $l \leq r$. The proof again proceeds as follows.

We claim that for all $\theta > 0$,

$$S_{\max}(\theta) = \max\{h(\alpha_n, \ell_n, \theta),\ h(\alpha_1, \ell_1, \theta)\}. \tag{25}$$

If $\theta \leq r$, then

$$\partial_\alpha h(\alpha, \ell_n, \theta) \geq 0, \quad \partial_\ell h(\alpha_n, \ell, \theta) \geq 0,$$

which implies

$$h(\alpha_n, \ell_n, \theta) \geq \max\{h(\alpha_n, \ell_1, \theta),\ h(\alpha_1, \ell_n, \theta)\}.$$

If $\theta \geq l$, then similarly

$$h(\alpha_1, \ell_1, \theta) \geq \max\{h(\alpha_n, \ell_1, \theta),\ h(\alpha_1, \ell_n, \theta)\}.$$

Since for any $\theta > 0$, either $\theta \leq r$ or $\theta \geq l$, it follows that the corner points (α_1, ℓ_n) and (α_n, ℓ_1) are dominated. This establishes equation 25.

Using equation 25, the minimizer is obtained by solving

$$h(\alpha_n, \ell_n, \theta) = h(\alpha_1, \ell_1, \theta).$$

Solving for θ yields

$$\hat{\theta}^{\star}_{\text{ADMM}} = \sqrt{\frac{\alpha_n\alpha_1(\ell_n - \ell_1) + \ell_n\ell_1(\alpha_n - \alpha_1)}{(\alpha_n - \alpha_1) + (\ell_n - \ell_1)}}. \tag{26}$$

Since this reduces to a linear equation in θ^2, there exists a unique positive solution. Moreover, this solution corresponds to the unique minimizer of $S_{\max}(\theta)$.

Combining the results from Configurations 1–3 completes the proof of Theorem 4. □

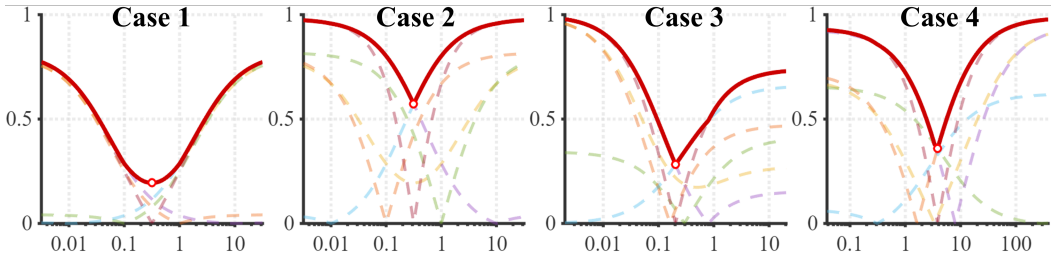


Figure 8: The relationship between the upper-envelope function $S(\theta)$ (shown in red) and its six envelope branches of ϕ (dotted lines). The x-axis denotes the penalty parameter θ, and the y-axis denotes the spectral radius. In each case, the minimizer of $S(\theta)$ is attained at the turning point of the corresponding red branch.

A.4 Proof of Theorem 8

The derivation of this theorem is nontrivial. We first provide a proof sketch, followed by a complete proof for each case of parameter selection.

A.4.1 Proof Sketch

Our goal is to derive the global minimizer of the following spectral upper-envelope function $S(\theta)$ consisting of six branches. To better understand this function, we plot Fig. 8. The proof in this appendix aims to find the minimizer of the red solid line under different conditions.

$$S(\theta) = \max_{\substack{(i,j)\in\{1,n\}^2 \\ (k,q)\in\{1,n\}^2}} \phi_{(i,j),(k,q)}(\theta). \tag{27}$$

Recall that the function $h(\alpha, \ell, \theta)$ in Appendix A.1 is symmetric in terms of α and ℓ. For this, we just need to consider the minimization of equation 27 over three representative configurations in terms of the four corner points:

Configuration 1: $\alpha_1 < \alpha_n < \ell_1 < \ell_n$,

Configuration 2: $\alpha_1 < \ell_1 < \ell_n < \alpha_n$,

Configuration 3: $\alpha_1 < \ell_1 < \alpha_n < \ell_n$.

As established in Lemma 10 and 11, all five branches $\phi_{\ell_1}(\theta)$, $\phi_{\alpha_1}(\theta)$, $\phi_{\alpha_n}(\theta)$, $\phi_{\ell_n}(\theta)$, and $\phi_{\theta_1}(\theta)$ exhibit a v-shaped structure, decreasing to zero and then increasing to 1. We first characterize the upper envelope of these five branches before introducing the sixth branch.

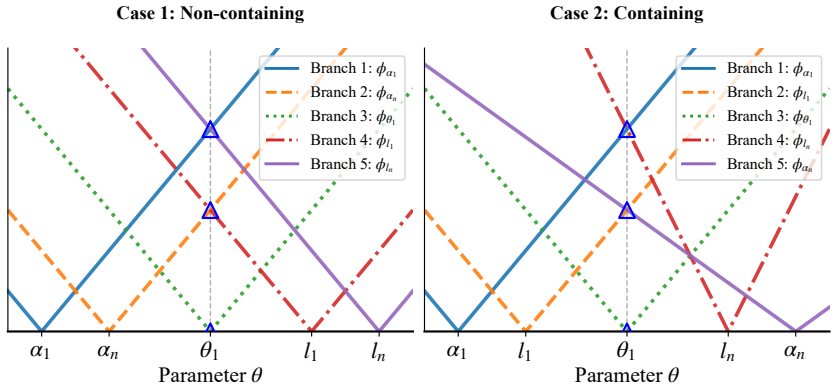


Figure 9: Illustration of the minimization over five branches. For both containing and non-containing cases, the unique minimizer of the upper envelope is always attained at $\theta = \theta_1$. This figure is a simplified illustration of Fig. 8.

The minimization over these five branches yields two further cases. The first is the **non-containing** case (i.e., Configuration 1 and 3). As illustrated in Fig. 9 (left), the branch pairs $(\phi_{\alpha_1}, \phi_{\ell_n})$ and $(\phi_{\alpha_n}, \phi_{\ell_1})$ intersect precisely at $\theta = \theta_1$, rendering θ_1 the unique candidate point for the minimizer. The second is the **containing** case (i.e., Configuration 2), where the upper envelope (i.e., $\mathcal{E}(\theta)$ defined below) is still minimized at $\theta = \theta_1$, making it the only candidate point for the minimizer, as shown in Fig. 9 (right).

Let $\mathcal{E}(\theta) := \max\{\phi_{\alpha_1}(\theta), \phi_{\alpha_n}(\theta), \phi_{\ell_1}(\theta), \phi_{\ell_n}(\theta), \phi_{\theta_1}(\theta)\}$ denote the upper envelope of the five branches. We now introduce the sixth branch $\phi_{(1,n),(n,1)}(\theta)$ and analyze it under two scenarios:

- **Disjoint scenario:** For $\alpha_1 < \alpha_n < \ell_1 < \ell_n$, the branch $\phi_{(1,n),(n,1)}(\theta)$ intersects $\mathcal{E}(\theta)$ at two critical points: the inner-left point α_n and the inner-right point ℓ_1. Within the interval (α_n, ℓ_1), we have $\phi_{(1,n),(n,1)}(\theta) > \mathcal{E}(\theta)$, as shown in Case 1 of Fig 8 where the yellow dotted line coincides with the red solid line. Consequently, there is no longer a guarantee that the minimizer occurs at $\theta = \theta_1$. In this case, the minimum must instead be sought over the set $\{\alpha_n, \ell_1, \theta_2\}$.
- **Overlapping scenario:** In other parameter orderings, $\phi_{(1,n),(n,1)}(\theta) < \mathcal{E}(\theta)$ holds over the relevant interval, corresponding to Cases 2–4 in Fig. 8, where in each case the yellow dotted line lies strictly below the red solid line. In these cases, the sixth branch does not affect the minimum of the upper envelope function $\mathcal{E}(\theta)$, and the minimizer still occurs at $\theta = \theta_1$.

By combining the cases above, the global minimizer of the upper envelope over all six branches reduces to evaluating the upper envelope only on a finite candidate set:

$$\mathcal{C} = \Big\{\min(\alpha_n, \ell_n),\ \theta_1,\ \max(\alpha_1, \ell_1),\ \theta_2\Big\}.$$

A.4.2 Full Proof

Recall that the function ϕ is characterized by six distinct branches:

$$\phi_{(1,1),(1,n)},\quad \phi_{(1,1),(n,1)},\quad \phi_{(1,n),(n,n)},\quad \phi_{(n,1),(n,n)},\quad \phi_{(1,1),(n,n)},\quad \phi_{(1,n),(n,1)}.$$

These branches correspond to evaluating the function h at the four corner points:

$$h_{\alpha_1,\ell_1},\quad h_{\alpha_1,\ell_n},\quad h_{\alpha_n,\ell_1},\quad h_{\alpha_n,\ell_n}.$$

Recall that the upper-envelope function $S(\theta)$ of all six branches is defined as equation 27. The global minimizer of $S(\theta)$ must arise in one of the following two scenarios:

- It is a stationary point of one of the six branches;
- It is an intersection point at which the 1st-order derivatives of the two branches have opposite signs. For simplicity, we refer to these points as turning points.

These two scenarios divide the discussion into the three parts presented below.

Part 1: Derivation of Stationary Points. We begin by analysing the stationary points of the six branches defined above and establish the following lemma.

Lemma 10. *Given two distinct vertices $(\alpha_i, \ell_j) \neq (\alpha_k, \ell_q)$, any stationary point of*

$$\phi_{(i,j),(k,q)}(\theta) := \frac{\left|h_{\alpha_k,\ell_q}(\theta) - h_{\alpha_i,\ell_j}(\theta)\right|}{2 - \big(h_{\alpha_k,\ell_q}(\theta) + h_{\alpha_i,\ell_j}(\theta)\big)} \tag{28}$$

must satisfy the quadratic equation

$$c_2\theta^2 + c_1\theta + c_0 = 0, \tag{29}$$

where

$$c_2 = (a_k + \ell_q) - (a_i + \ell_j), \quad c_1 = 2(a_k\ell_q - a_i\ell_j), \quad c_0 = a_k\ell_q(a_i + \ell_j) - a_i\ell_j(a_k + \ell_q).$$

Proof. The absolute value in equation 28 does not affect the stationary points and can therefore be safely removed. Without loss of generality, we assume that

$$h_{\alpha_k,\ell_q}(\theta) \geq h_{\alpha_i,\ell_j}(\theta).$$

For notational convenience, define

$$m(\theta) := h_{\alpha_k,\ell_q}(\theta), \quad n(\theta) := h_{\alpha_i,\ell_j}(\theta), \quad u(\theta) := m(\theta) - n(\theta), \quad v(\theta) := 2 - \big(m(\theta) + n(\theta)\big).$$

Then

$$\phi_{(i,j),(k,q)}(\theta) = \frac{u(\theta)}{v(\theta)}.$$

Differentiating this expression with respect to θ, we obtain

$$\phi'_{(i,j),(k,q)}(\theta) = \frac{u'(\theta)v(\theta) - u(\theta)v'(\theta)}{v(\theta)^2}.$$

Since $v(\theta)^2 > 0$, any stationary point must satisfy

$$m'(\theta)\big(1 - n(\theta)\big) = n'(\theta)\big(1 - m(\theta)\big). \tag{30}$$

Substituting the expressions for $m(\theta)$ and $n(\theta)$ into equation 30 yields

$$\frac{(a_k + \ell_q)(\theta^2 - a_k\ell_q)}{(\theta + a_k)^2(\theta + \ell_q)^2} \cdot \frac{\theta(a_i + \ell_j)}{(\theta + a_i)(\theta + \ell_j)} = \frac{(a_i + \ell_j)(\theta^2 - a_i\ell_j)}{(\theta + a_i)^2(\theta + \ell_j)^2} \cdot \frac{\theta(a_k + \ell_q)}{(\theta + a_k)(\theta + \ell_q)}.$$

Rearranging the above expression, we obtain

$$c_2\theta^2 + c_1\theta + c_0 = 0,$$

where c_2, c_1, and c_0 are defined as in Lemma 10. This completes the proof. □

Based on Lemma 10, we can explicitly compute the stationary point of each of the six branches. For the following four branches, the stationary points can be computed from the first-order optimality condition $\phi'(\theta) = 0$, all of which yielding negative solutions:

$$\phi'_{(1,1),(1,n)}(\theta) = 0 \;\Rightarrow\; \theta = -\alpha_1, \qquad \phi'_{(n,1),(n,n)}(\theta) = 0 \;\Rightarrow\; \theta = -\alpha_n,$$

$$\phi'_{(1,1),(n,1)}(\theta) = 0 \;\Rightarrow\; \theta = -\ell_1, \qquad \phi'_{(1,n),(n,n)}(\theta) = 0 \;\Rightarrow\; \theta = -\ell_n.$$

Hence, none of these branches admits a minimizer of $S(\theta)$.

For the branch $\phi_{(1,1),(n,n)}$, any stationary point must satisfy equation 29. Since all coefficients in equation 29 have the same sign, the quadratic admits no positive solution. Consequently, this branch has no real-valued stationary point. As such the only branch that may admit a minimizer of $S(\theta)$ is $\phi_{(1,n),(n,1)}$, for which $\theta_2 \in \mathbb{R}_{++}$ can be obtained by solving

$$\phi'_{(1,n),(n,1)}(\theta_2) = 0. \tag{31}$$

Lemma 11. *Among the six branches, five are v-shaped and are defined on $\mathbb{R}_{++}$, the set of strictly positive real numbers:*

$$\phi_{(1,1),(n,1)},\ \ \phi_{(1,1),(1,n)},\ \ \phi_{(n,n),(n,1)},\ \ \phi_{(n,n),(1,n)},\ \ \phi_{(1,1),(n,n)}.$$

Each of these branches has a v-shaped profile and admits a unique zero, which occur at

$$\ell_1,\ \ \alpha_1,\ \ \alpha_n,\ \ \ell_n,\ \ \theta_1,$$

respectively.

Proof. To prove the lemma, it is sufficient to consider a representative branch. The remaining cases follow by the same argument. For this, let us consider

$$\phi_{(1,1),(n,1)}(\theta),$$

for which at $\theta = \ell_1$ we have

$$h_{\alpha_1,\ell_1}(\ell_1) = h_{\alpha_n,\ell_1}(\ell_1).$$

As such

$$\phi_{(1,1),(n,1)}(\ell_1) = 0.$$

By Lemma 10, we can derive

$$\phi'_{(1,1),(n,1)}(\theta) = \begin{cases} < 0, & \theta < \ell_1 \\ > 0, & \theta > \ell_1, \end{cases}$$

which means that $\phi_{(1,1),(n,1)}$ is strictly decreasing on $(0, \ell_1)$ and strictly increasing on (ℓ_1, ∞), and therefore attains its minimum at $\theta = \ell_1$. Note that the function $\phi_{(1,1),(n,1)}$ is non-differentiable at the point $\theta = \ell_1$, so $\phi_{(1,1),(n,1)}$ is v-shaped with a unique minimizer at the point. The same reasoning above applies to all other cases. This completes the proof. □

Next, for convenience, we relabel the five v-shaped branches according to the locations of their unique minimizers:

$$\begin{aligned} \phi_{\ell_1}(\theta) &:= \phi_{(1,1),(n,1)}(\theta), \\ \phi_{\alpha_1}(\theta) &:= \phi_{(1,1),(1,n)}(\theta), \\ \phi_{\alpha_n}(\theta) &:= \phi_{(n,n),(n,1)}(\theta), \\ \phi_{\ell_n}(\theta) &:= \phi_{(n,n),(1,n)}(\theta), \\ \phi_{\theta_1}(\theta) &:= \phi_{(1,1),(n,n)}(\theta). \end{aligned} \tag{32}$$

Each function above $\phi_\epsilon(\theta)$ is a v-shaped curve that is not differentiable at its unique minimum, attained at $\theta = \epsilon$.

Part 2: Derivation of the Minimizer of the Upper Envelope of the Five Branches. We now analyze the upper envelope of the five v-shaped branches and prove that its minimum is attained at $\theta = \theta_1$ where θ_1 is given in equation 8.

Lemma 12. *Let $\mathcal{E}(\theta)$ be the upper-envelope function of the five branches:*

$$\mathcal{E}(\theta) := \max\Big\{\phi_{\alpha_1}(\theta),\, \phi_{\alpha_n}(\theta),\, \phi_{\ell_1}(\theta),\, \phi_{\ell_n}(\theta),\, \phi_{\theta_1}(\theta)\Big\}, \quad \theta > 0.$$

Then $\mathcal{E}(\theta)$ attains its minimum at $\theta = \theta_1$.

Proof. By Lemma 11, each branch is v-shaped with a unique minimizer occurring at

$$\alpha_1, \quad \alpha_n, \quad \ell_1, \quad \ell_n, \quad \theta_1,$$

respectively. Hence, any minimizer of $\mathcal{E}(\theta)$ must occur at a non-differentiable point, which is a turning point (see Fig 9 for illustration of such a point) consisting of a decreasing left half-branch and an increasing right half-branch. We classify all possible turning points according to their relative position to the corner points. By symmetry, it it sufficient to consider only the following three representative configurations.

Configuration 1: $\alpha_1 < \alpha_n < \ell_1 < \ell_n$.

The derivation under this configuration can be illustrated in Fig. 9 left. No minimizer can lie in $(0, \alpha_1]$ or $[\ell_n, \infty)$, since in these intervals all five branches share the same monotonicity (i.e., they are either all increasing or all decreasing). Thus the only interval that can contain a minimizer is (α_1, ℓ_n). We can further exclude the following two intervals due to the fact that the corresponding turning points in these intervals are below that at $\theta = \theta_1$.

$$(\alpha_1, \alpha_n], \quad [\ell_1, \ell_n).$$

Hence, only the interval (α_n, ℓ_1) remains. On this interval, ϕ_{θ_1} (the green branch in Fig. 9 left) also attains its minimum at θ_1 and does not introduce any additional minimizers. The only remaining turning point occurs between ϕ_{α_n} and ϕ_{ℓ_1} (the orange and red branches in Fig. 9 left), whose unique intersection is also at θ_1. Thus, the final minimizer is attained at $\theta = \theta_1$.

Configuration 2: $\alpha_1 < \ell_1 < \alpha_n < \ell_n$.

The analysis is the same as in Configuration 1.

Configuration 3: $\alpha_1 < \ell_1 < \ell_n < \alpha_n$.

The derivation for this configuration is illustrated in Fig. 9 (right). The turning point between ϕ_{α_1} and ϕ_{ℓ_n} occurs at $\theta_1 > \ell_1$, which excludes the interval $(\alpha_1, \ell_1]$. Similarly, the turning point between ϕ_{ℓ_1} and ϕ_{α_n} occurs at $\theta_1 < \ell_n$, excluding the interval $[\ell_n, \alpha_n)$. Therefore, only the interval (ℓ_1, ℓ_n) remains for consideration. Within this interval, all corresponding turning points lie below the turning point at θ_1, and therefore cannot be minimizers. Consequently, the final minimizer is attained at $\theta = \theta_1$.

Combining these three configurations, we conclude that the minimizer of the upper-envelope function $\mathcal{E}(\theta)$ can be only attained at $\theta = \theta_1$. The same conclusion remains valid in cases where certain corner points coincide. □

Part 3: Analysis of the Sixth Branch. We now analyze the sixth branch $\phi_{(1,n),(n,1)}(\theta)$ to determine whether it introduces additional candidate points beyond θ_1 identified in Lemma 12. The zero point θ_0 of this branch, at which $\phi_{(1,n),(n,1)}(\theta_0) = 0$, is characterized by

$$h_{\alpha_1,\ell_n}(\theta_0) = h_{\alpha_n,\ell_1}(\theta_0),$$

which yields

$$\theta_0^2 = \frac{\ell_1\ell_n(\alpha_n - \alpha_1) - \alpha_1\alpha_n(\ell_n - \ell_1)}{(\alpha_n - \alpha_1) - (\ell_n - \ell_1)}.$$

The stationary point of this branch is obtained from

$$\phi'_{(1,n),(n,1)}(\theta_2) = 0,$$

which produces another candidate point, denoted by θ_2. We can further prove that these two candidate points (i.e., θ_0 and θ_2) are mutually exclusive: either θ_0 or θ_2 lies in $\mathbb{R}_{++}$. Thus, the admissible point is the unique turning point of this branch, implying that $\phi_{(1,n),(n,1)}$ also forms a v-shaped curve on $\mathbb{R}_{++}$.

Next, we examine how this branch modifies the overall envelope structure. For this, we consider the following two scenarios.

Disjoint Scenario: $\alpha_1 < \alpha_n < \ell_1 < \ell_n$.

We compare the sixth branch $\phi_{(1,n),(n,1)}$ with the upper-envelope function $\mathcal{E}(\theta)$ of the five branches. In this configuration, $\mathcal{E}(\theta)$ is determined by $\phi_{\alpha 1}(\theta)$ and $\phi_{\ell_n}(\theta)$.

First, we let

$$\phi_{\alpha_1}(\theta) = \phi_{(1,n),(n,1)}(\theta)$$

to calculate the interaction point. This leads to

$$\frac{h_{\alpha_1,\ell_1} - h_{\alpha_1,\ell_n}}{2 - \left(h_{\alpha_1,\ell_1} + h_{\alpha_1,\ell_n}\right)} = \frac{h_{\alpha_n,\ell_1} - h_{\alpha_1,\ell_n}}{2 - \left(h_{\alpha_n,\ell_1} + h_{\alpha_1,\ell_n}\right)}.$$

Observing the structure of the above equation, we define the following function:

$$g_y(t) := \frac{t-y}{2-(t+y)}.$$

$g_y(t)$ is strictly increasing in t, implying

$$h_{\alpha_1,\ell_1} = h_{\alpha_n,\ell_1} \quad \Rightarrow \quad \theta = \ell_1.$$

Next, we let

$$\phi_{\ell_n}(\theta) = \phi_{(1,n),(n,1)}(\theta).$$

With the similar derivations above, we know

$$\theta = \alpha_n$$

In conclusion, the sixth branch intersects the other five branches only at α_n and ℓ_1. As such, on the interval $[\alpha_n, \ell_1]$, we have

$$\phi_{(1,n),(n,1)}(\theta) \geq \mathcal{E}(\theta).$$

In this case, θ_1 is no longer a candidate point. Finally, these new candidate points form the valid set as follows:

$$\{\alpha_n,\ \ell_1,\ \theta_2\},$$

provided that $\theta_2 \in (\alpha_n, \ell_1)$.

Overlapping Scenario: $\alpha_1 < \ell_1 < \ell_n < \alpha_n$ or $\alpha_1 < \ell_1 < \alpha_n < \ell_n$. For this configuration, the same reasoning as in Configuration 1 shows that

$$\phi_{(1,n),(n,1)}(\theta) \leq \mathcal{E}(\theta),$$

corresponding to Cases 2–4 in Fig. 8, where in each case the yellow dotted line represents the sixth branch, which always lies below the corresponding upper envelope in red. Hence, the sixth branch does not alter the structure of the envelope, and the unique candidate point remains

$$\theta_1.$$

Combining the results from Parts 1–3, the minimization of the upper-envelope function $S(\theta)$ of the six branches is reduced to an optimization problem over the following set:

$$\mathcal{C} = \Big\{\min(\alpha_n, \ell_n),\ \theta_1,\ \max(\alpha_1, \ell_1),\ \theta_2\Big\},$$

which completes the proof of Theorem 8 □

A.5 Derivation of the Dual Formulation

The PDE linear system equation 16 is equivalent to minimizing the following quadratic optimization problem:

$$\min_{u\in\mathbb{R}^n} \ \frac{1}{2\tau}\|u-u^k\|_2^2 + \frac{1}{2}\|K^{1/2}Du\|_2^2 - \langle f^{k+1}, u\rangle. \tag{33}$$

Introducing the auxiliary variable $v = Du$, we rewrite equation 33 as the constrained problem

$$\min_{u\in\mathbb{R}^n,\ v\in\mathbb{R}^{2n}} \ \frac{1}{2\tau}\|u-u^k\|_2^2 + \frac{1}{2}\|K^{1/2}v\|_2^2 - \langle f^{k+1}, u\rangle \quad \text{s.t. } v = Du. \tag{34}$$

The associated Lagrangian, with dual variable $p \in \mathbb{R}^{2n}$, is

$$\mathcal{L}(u,v,p) = \frac{1}{2\tau}\|u-u^k\|_2^2 + \frac{1}{2}\|K^{1/2}v\|_2^2 - \langle f^{k+1}, u\rangle + \langle p, v-Du\rangle. \tag{35}$$

We derive the dual function by minimizing $\mathcal{L}$ with respect to u and v. The terms depending on v are

$$\frac{1}{2}\|K^{1/2}v\|_2^2 + \langle p, v\rangle.$$

Setting the gradient to zero yields

$$Kv + p = 0, \quad \Rightarrow \quad v^* = -K^{-1}p.$$

Substituting back gives

$$\inf_v \left(\frac{1}{2}\|K^{1/2}v\|_2^2 + \langle p, v \rangle \right) = -\frac{1}{2}\|K^{-1/2}p\|_2^2. \tag{36}$$

The terms depending on u are

$$\frac{1}{2\tau}\|u - u^k\|_2^2 - \langle f^{k+1}, u \rangle - \langle p, Du \rangle.$$

Using $\langle p, Du \rangle = \langle D^\top p, u \rangle$, this becomes

$$\frac{1}{2\tau}\|u - u^k\|_2^2 - \langle f^{k+1} + D^\top p, u \rangle.$$

The first-order optimality condition gives

$$\frac{1}{\tau}(u - u^k) - (f^{k+1} + D^\top p) = 0,$$

so that

$$u^* = u^k + \tau\big(f^{k+1} + D^\top p\big). \tag{37}$$

Substituting u^* back, we obtain

$$\inf_u \left(\frac{1}{2\tau}\|u - u^k\|_2^2 - \langle f^{k+1} + D^\top p, u \rangle \right) = -\frac{\tau}{2}\|f^{k+1} + D^\top p\|_2^2 - \langle f^{k+1} + D^\top p, u^k \rangle. \tag{38}$$

Combining equation 36 and equation 38, the dual function is

$$\mathcal{G}(p) = -\frac{\tau}{2}\|D^\top p + f^{k+1}\|_2^2 - \frac{1}{2}\|K^{-1/2}p\|_2^2 - \langle D^\top p + f^{k+1},\, u^k \rangle.$$

Maximizing $\mathcal{G}(p)$ is equivalent to minimizing its negative, leading to the dual problem

$$\min_{p \in \mathbb{R}^{2n}} \ \frac{\tau}{2}\|D^\top p + f^{k+1}\|_2^2 + \frac{1}{2}\|K^{-1/2}p\|_2^2 + \langle D^\top p + f^{k+1},\, u^k \rangle, \tag{39}$$

which coincides with equation 17. □

B Additional Experiments

This section provides additional comparisons designed to isolate the practical cost of parameter selection. We first compare the proposed analytical rules with the iterative spectral-radius search used by Song et al. (2024), and then compare our one-shot selection with adaptive ADMM and oADMM methods. We also report an end-to-end wall-clock comparison and additional robustness tests for larger and rank-deficient problems.

B.1 Analytical Tuning vs. Iterative Spectral-Radius Search

As a representative search-based baseline, we follow the numerical spectral-radius optimization strategy of Song et al. (2024). For each candidate penalty parameter θ, the spectral radius $\rho(M(\theta))$ is computed numerically, and θ is refined by gradient descent. The gradient is approximated using centered finite differences, so each update requires repeated spectral-radius evaluations; in our implementation, these evaluations use an implicitly restarted Arnoldi eigensolver.

In contrast, our ADMM rule uses only the extremal eigenvalues $\alpha_1, \alpha_n, \ell_1$, and ℓ_n, after which the parameter follows directly from Theorem 4. For oADMM, the continuous search over (θ, β) is reduced to evaluations of the finite candidate set in Theorem 8. Neither rule requires iterative minimization of the spectral radius.

Table 2 reports the parameter-selection time. The advantage of analytical tuning increases with dimension: at 1024×1024, the proposed rule requires 0.01303 seconds, compared with 259.07 seconds for iterative search, giving a speedup of approximately 1.99×10^4.

Table 2: Parameter-selection cost of the proposed analytical rule and the iterative spectral-radius search of Song et al. (2024).

Matrix size	Ours (s)	Numerical search (s)	Speedup
256×256	0.00085	4.449	5.2×10^3
512×512	0.00303	22.901	7.6×10^3
1024×1024	0.01303	259.07	1.99×10^4

B.2 One-Shot vs. Adaptive Parameter Selection

We next compare the proposed one-shot rules with three representative adaptive methods: residual-balancing ADMM (RB-ADMM) (Wohlberg, 2017), spectral adaptive ADMM (SA-ADMM) (Xu et al., 2017a), and adaptive relaxed ADMM (AR-ADMM) (Xu et al., 2017b). RB-ADMM and SA-ADMM adapt the ADMM penalty parameter, whereas AR-ADMM jointly adapts the penalty and relaxation parameters of oADMM.

For the linear-quadratic problems studied here, fixing the model and algorithmic parameters fixes the iteration matrix and hence its asymptotic convergence factor. Our method therefore selects the parameters once, before the iterations begin. Adaptive schemes instead update their parameters along the optimization trajectory and can depend strongly on their initial values. To assess this sensitivity, we initialize every adaptive method from five values relative to our analytically selected penalty $\widehat{\theta}$:

$$\frac{\theta_0}{\widehat{\theta}} \in \left\{10^{-2}, 10^{-1}, 1, 10, 10^2\right\}. \tag{40}$$

All methods use identical initial primal and dual variables and terminate when the relative L_2 error reaches 10^{-6}. We use the image-registration problem as a structured optimization benchmark at resolutions 128^2, 256^2, and 512^2.

Table 3 reports wall-clock time in seconds. For each adaptive method, the first value is the best runtime over the five initializations in equation 40, and the value in parentheses is the worst. The proposed oADMM rule is faster than AR-ADMM at all three resolutions, including AR-ADMM's best initialization. For ADMM, SA-ADMM is competitive under favorable initialization, but its runtime is substantially more variable. At 512^2, SA-ADMM ranges from 21.90 to 131.63 seconds, while analytically tuned ADMM takes 19.03 seconds. RB-ADMM is more sensitive still: its best run takes 335.85 seconds, and its worst run exceeds the 3000-second time limit.

Table 3: Wall-clock comparison with adaptive parameter-selection methods. For each adaptive method, the best runtime across five initializations is reported first and the worst is shown in parentheses. All values are in seconds.

Method	128^2	256^2	512^2
Ours (oADMM)	**0.11**	**1.68**	**9.10**
AR-ADMM	0.17 (0.92)	2.50 (7.95)	12.80 (40.72)
Ours (ADMM)	**0.37**	3.08	**19.03**
SA-ADMM	0.45 (2.70)	**2.65** (15.93)	21.90 (131.63)
RB-ADMM	2.35 (11.70)	15.44 (93.20)	335.85 ($>$ 3000)

These results highlight the difference between one-shot and online parameter selection. Adaptive methods may reach effective parameter regimes, but their transient behavior depends on initialization and on information accumulated during the iterations. The proposed rules determine the parameters directly from spectral information and avoid this adaptation phase.

B.3 End-to-End Wall-Clock Efficiency

Iteration counts alone do not capture the different per-iteration and startup costs of ADMM, oADMM, CG, and PCG. We therefore perform a single-GPU wall-clock comparison on the same 1024×1024 image-registration instance. For PCG, we use an incomplete Cholesky preconditioner

with drop tolerance 10^{-2}. For the proposed methods, startup time includes the extraction of the spectral endpoints and the one-shot parameter selection; for ICT-PCG, it includes construction of the incomplete Cholesky factorization.

Table 4: End-to-end wall-clock comparison on a 1024×1024 image-registration instance.

Method	Iterations	Time/iter. (ms)	Startup (s)	Total (s)	Final rel. residual
Ours (oADMM)	539	7.9	0.0130	**4.27**	9.38×10^{-7}
Ours (ADMM)	1051	7.7	0.0131	8.10	9.70×10^{-7}
CG	6367	7.9	0	50.30	9.27×10^{-7}
ICT-PCG (10^{-2})	596	143.4	0.387	85.85	9.93×10^{-7}

ADMM and oADMM have nearly identical per-iteration costs, while over-relaxation reduces the total runtime by 47.4%, corresponding to a $1.90\times$ speedup. Although ICT-PCG requires a comparable number of iterations, its sparse triangular solves are substantially more expensive on the GPU: its per-iteration cost is approximately $18.2\times$ that of oADMM and its total runtime, including startup, is approximately $20.1\times$ larger.

B.4 Scalability and Rank-Deficient Problems

Finally, we assess whether the analytical approximation remains accurate as the matrix dimension grows and when the forward operator is underdetermined. We use the spectral-radius gap

$$\Delta\rho := \rho\big(M(\widehat{\theta})\big) - \rho\big(M(\theta^\star)\big) \geq 0, \tag{41}$$

where $\widehat{\theta}$ is selected by the proposed rule and $\theta^\star$ is obtained through an independent numerical search. The same definition is used for oADMM with its jointly selected penalty and relaxation parameters.

Table 5 (left) shows that the gap remains small for random matrices up to dimension 2048×2048 and does not deteriorate with dimension. To test an underdetermined and rank-deficient setting, we also sample 200 instances with $A \in \mathbb{R}^{50\times 100}$, for which $A^\top A$ is necessarily singular. As shown in Table 5 (right), both ADMM and oADMM retain small gaps relative to an independent numerical search.

Table 5: Additional robustness experiments measured by the spectral-radius gap $\Delta\rho$. Left: increasing matrix dimension. Right: 200 underdetermined instances with $A \in \mathbb{R}^{50\times 100}$.

Dimension	Median $\Delta\rho$	90th percentile
256×256	8.74×10^{-4}	8.81×10^{-3}
512×512	6.72×10^{-4}	5.45×10^{-3}
1024×1024	7.49×10^{-4}	4.39×10^{-3}
2048×2048	4.96×10^{-4}	2.98×10^{-3}

Method	Median $\Delta\rho$	90th percentile
ADMM	9.77×10^{-4}	9.41×10^{-3}
oADMM	1.83×10^{-3}	2.40×10^{-3}