**Les Cahiers du GERAD** ISSN: 0711–2440

# A spectral preconditioner for the conjugate gradient method with iteration budget

Y. Diouane, S. Gürol, O. Mouthal, D. Orban

G–2026–19

September 2026



La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2026
– Bibliothèque et Archives Canada, 2026

Legal deposit – Bibliothèque et Archives nationales du Québec, 2026
– Library and Archives Canada, 2026

**GERAD** HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

**Tél. : 514 340-6053**
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

# A spectral preconditioner for the conjugate gradient method with iteration budget

**Youssef Diouane** [a]
**Selime Gürol** [b]
**Ossama Mouthal** [b]
**Dominique Orban** [a]

[a] *GERAD & Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Qc), Canada, H3T 1J4*

[b] *CERFACS / CECI CNRS UMR 5318, Toulouse, France*

youssef.diouane@polymtl.ca
gurol@cerfacs.fr
mouhtal@cerfacs.fr
dominique.orban@polymtl.ca

**September 2026**
**Les Cahiers du GERAD**
**G–2026–19**

**Abstract :** We study the solution of large symmetric positive-definite linear systems in a matrix-free setting with a limited iteration budget. We focus on the preconditioned conjugate gradient (PCG) method with spectral preconditioning. Spectral preconditioners map a subset of eigenvalues to a positive cluster via a scaling parameter, and leave the remainder of the spectrum unchanged, in hopes to reduce the number of iterations to convergence. We formulate the design of the spectral preconditioners as a constrained optimization problem. The optimal cluster placement is defined to minimize the error in energy norm at a fixed iteration. This optimality criterion provides new insight into the design of efficient spectral preconditioners when PCG is stopped short of convergence. We propose practical strategies for selecting the scaling parameter, hence the cluster position, that incur negligible computational cost. Numerical experiments highlight the importance of cluster placement and demonstrate significant improvements in terms of error in energy norm, particularly during the initial iterations.

**Keywords :** Linear systems; matrix-free; conjugate gradient method; deflated CG; spectral preconditioner; error in energy norm

**Acknowledgements:** This work was funded by French National Programme LEFE/INSU.

# 1 Introduction

We consider large symmetric positive-definite (SPD) linear systems, where the matrix is implicitly represented as an *operator*, using the conjugate gradient (CG) method [19]. A standard approach to accelerate convergence is to introduce a preconditioner. There is no universal strategy for constructing effective preconditioners; we refer the reader to the surveys [3, 23, 30] for an overview.

One important class of preconditioners adapted to matrix-free settings is that of limited memory preconditioners (LMPs) [14]. This approach is closely related to deflation [10, 24], augmentation methods [6, 11], balancing strategies [21], and low-rank updates [4]. The main idea is to modify the eigenspectrum of the preconditioned matrix such that a small set of eigenvalues of the linear system matrix are either mapped to a positive scalar value [14] or are removed [10] in hopes to reduce the number of iterations to converge.

In this work, we focus on *spectral preconditioners* [5, 12, 14], which are constructed using extreme eigenvalues and corresponding eigenvectors. These preconditioners come in two different types. The first type maps the extreme eigenvalues to a positive scalar, and is called a *deflating preconditioner* in [12] and a *spectral LMP* in [14, 27]. The second type shifts the extreme eigenvalues by a common positive value, and is called a coarse grid preconditioner [5]. In practice, the scalar value is often set to one [12, 14, 27].

Both approaches are considered in [12], where the authors analyze the eigenvalue distribution of the preconditioned systems to first-order approximation when the eigenspectrum is approximately known. For instance, in large-scale weather prediction [8], the spectral LMP is constructed using Ritz pairs extracted from PCG recurrences [25]. Alternatively, parallel randomized algorithms [17] can be used to approximate extreme eigenvalues and their corresponding eigenvectors [9].

Our main objective is to design a spectral preconditioner that minimizes the error after $\ell$ steps, where $\ell$ is fixed, motivated by applications in which the solver must be stopped after a prescribed number of iterations. The resulting preconditioner is closely related to the deflating preconditioner [12, 14] and is also connected to deflation techniques [24].

Formulating the requirements of the preconditioner as an optimization problem reveals its structure: the preconditioner maps a selected subset of eigenvalues to positive values while leaving the rest unchanged. Since explicitly computing this set of eigenvalues is impractical, we propose a strategy to identify the relevant part of the spectrum. We show that the preconditioner should be constructed from the extreme eigenvalues, and we propose several strategies to define the positioning of the cluster. These strategies are also shown to be linked to the energy norm of the error, which is precisely the quantity minimized by CG.

The paper is organized as follows. In Section 2, we review the CG method and its convergence properties, and then discuss the characteristics of an efficient preconditioner. Our main contributions are presented in Sections 3 and 4, where we define the *scaled* spectral preconditioner and analyze its properties. We also outline four strategies for selecting the scaling parameter, which determines the placement of the eigenvalue cluster, to *reduce the total number of iterations and improve convergence in the early stages*. In Section 5, we report numerical experiments on matrices with extreme eigenvalues. Conclusions and perspectives appear in Section 6.

## Notation

Matrix $A \in \mathbb{R}^{n\times n}$ is always SPD. Its spectral decomposition is $A = S\Lambda S^\top$ with $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \dots \geq \lambda_n > 0$, and $S = \begin{bmatrix} s_1 & \cdots & s_n \end{bmatrix}$ being orthogonal. The notation $\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ denotes the diagonal matrix whose diagonal entries are $\lambda_1, \dots, \lambda_n$. The $A$-norm, or *energy norm*, of $x \in \mathbb{R}^n$ is $\|x\|_A = \sqrt{x^\top Ax}$. The identity matrix of size $n$ is $I_n$. The set of eigenvalues of $X = X^T \in \mathbb{R}^{n\times n}$ is $\mathrm{spec}(X)$. The cardinality of a subset $\pi \subseteq \{1, \dots, n\}$ is $|\pi|$.

In what follows, all results are derived under the assumption of exact arithmetic.

# 2 Background

## 2.1 Conjugate Gradient

The CG [19] is a method for $Ax = b$ with $A \in \mathbb{R}^{n\times n}$ SPD and $b \in \mathbb{R}^n$. If $x_0 \in \mathbb{R}^n$ is an initial guess and $r_0 = b - Ax_0$ is the initial residual, then at every step $\ell = 1, 2, \dots, \mu$, CG produces a unique approximation [25]

$$x_\ell \in x_0 + \mathcal{K}_\ell(A, r_0) \quad \text{such that} \quad r_\ell \perp \mathcal{K}_\ell(A, r_0), \tag{1}$$

where $\mathcal{K}_\ell(A, r_0) := \operatorname{span}\{r_0, Ar_0, \dots, A^{\ell-1}r_0\}$ is the $\ell$-th Krylov subspace generated by $A$ and $r_0$ and $\mu$ is the grade of $r_0$ with respect to $A$, i.e., the maximum dimension of the Krylov subspace generated by $A$ and $r_0$ [25]. In exact arithmetic $x_\mu = x^*$, where $x^*$ is the exact solution. The most popular and efficient implementation of (1) is the original formulation of Hestenes and Stiefel [19], which recursively updates coupled 2-term recurrences for $x_{\ell+1}$, the residual $r_{\ell+1} := b - Ax_{\ell+1}$, and the search direction $p_{\ell+1}$. Algorithm 1 states the complete procedure. A common stopping criterion is based on sufficient decrease of the relative residual norm $\|r_\ell\|_2/\|r_0\|_2$. However, in certain practical implementations, such as data assimilation, a fixed number of iterations is used as a stopping criterion due to budget constraints. The preconditioned CG (PCG) variant is also presented, together with its companion formulation, Algorithm 2, which will be detailed in Section 2.3.

**Algorithm 1** CG

$r_0 = b - Ax_0$
$\rho_0 = r_0^\top r_0$
$p_0 = r_0$
**for** $\ell = 0, 1, \dots$ **do**
  $q_\ell = Ap_\ell$
  $\alpha_\ell = \rho_\ell/(q_\ell^\top p_\ell)$
  $x_{\ell+1} = x_\ell + \alpha_\ell p_\ell$
  $r_{\ell+1} = r_\ell - \alpha_\ell q_\ell$
  $\rho_{\ell+1} = r_{\ell+1}^\top r_{\ell+1}$
  $\beta_{\ell+1} = \rho_{\ell+1}/\rho_\ell$
  $p_{\ell+1} = r_{\ell+1} + \beta_{\ell+1} p_\ell$
**end for**

**Algorithm 2** PCG

$\hat{r}_0 = b - A\hat{x}_0$ // $\hat{x}_0 = x_0$
$z_0 = F\hat{r}_0$
$\hat{\rho}_0 = \hat{r}_0^\top z_0$
$\hat{p}_0 = z_0$
**for** $\ell = 0, 1, \dots$ **do**
  $\hat{q}_\ell = A\hat{p}_\ell$
  $\hat{\alpha}_\ell = \hat{\rho}_\ell/(\hat{q}_\ell^\top \hat{p}_\ell)$
  $\hat{x}_{\ell+1} = \hat{x}_\ell + \hat{\alpha}_\ell \hat{p}_\ell$
  $\hat{r}_{\ell+1} = \hat{r}_\ell - \hat{\alpha}_\ell \hat{q}_\ell$
  $z_{\ell+1} = F\hat{r}_{\ell+1}$
  $\hat{\rho}_{\ell+1} = \hat{r}_{\ell+1}^\top z_{\ell+1}$
  $\hat{\beta}_{\ell+1} = \hat{\rho}_{\ell+1}/\hat{\rho}_\ell$
  $\hat{p}_{\ell+1} = z_{\ell+1} + \hat{\beta}_{\ell+1}\hat{p}_\ell$
**end for**

## 2.2 Convergence properties of CG

The approximation (1) minimizes the error in energy norm:

$$\|x^* - x_\ell\|_A^2 = \min_{p\in\mathbb{P}_\ell(0)} \|p(A)(x^* - x_0)\|_A^2 = \min_{p\in\mathbb{P}_\ell(0)} \sum_{i=1}^n p(\lambda_i)^2 \frac{\eta_i^2}{\lambda_i}, \tag{2}$$

where $\eta_i = s_i^\top r_0$ and $\mathbb{P}_\ell(0)$ is the set of polynomials of degree at most $\ell$ with value 1 at zero [25]. Thus, at each iteration, before reaching the solution, CG solves a certain weighted polynomial approximation problem over the discrete set $\{\lambda_1, \dots, \lambda_n\}$. Moreover, if $z_1^{(\ell)}, \dots, z_\ell^{(\ell)}$ are the $\ell$ roots of the unique solution $p_\ell$ to (2) [15, 28],

$$\|x^* - x_\ell\|_A^2 = \sum_{i=1}^n p_\ell(\lambda_i)^2 \frac{\eta_i^2}{\lambda_i} = \sum_{i=1}^n \prod_{j=1}^\ell \left(1 - \frac{\lambda_i}{z_j^{(\ell)}}\right)^2 \frac{\eta_i^2}{\lambda_i}. \tag{3}$$

The $z_j^{(\ell)}$ are the *Ritz values* [28]. From (3), if $z_j^{(\ell)}$ is close to a $\lambda_i$, we expect a significant reduction in the error in energy norm. Based on the above, Van der Sluis and Van der Vorst [28] explain the rate of convergence of CG in terms of the convergence of the Ritz values to eigenvalues of $A$.

The widely stated convergence bound for CG

$$\frac{\|x^* - x_\ell\|_A}{\|x^* - x_0\|_A} \leq 2 \left( \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^\ell, \tag{4}$$

where $\kappa(A) := \lambda_1/\lambda_n$ is the condition number of $A$. While (4) provide the worst-case behavior of CG, the convergence properties may vary significantly from the worst case for a specific right hand side [2].

### 2.3 Properties of a good preconditioner

In many practical applications, a preconditioner is essential for accelerating the convergence of CG [3, 30]. Assume that a preconditioner $F = UU^\top \in \mathbb{R}^{n\times n}$ is available in factored form, where $U$ is non singular, and consider the system with split preconditioner

$$U^\top AUy = U^\top b, \tag{5}$$

whose matrix is also SPD. System (5) can then be solved with CG. The latter updates estimate $y_\ell$ that can be used to recover $\hat{x}_\ell := Uy_\ell$. Algorithm 2, the PCG method, is equivalent to the procedure just described, but only involves products with $F$ and does not assume knowledge of $U$ [13, p.532]. PCG updates $\hat{x}_\ell$ directly as an approximate solution in $\hat{x}_0 + U\mathcal{K}_\ell(U^\top AU, U^\top r_0)$, with $\hat{x}_0 = x_0$, as in (2), $\hat{x}_\ell$ minimizes the energy norm

$$\|x^* - \hat{x}_\ell\|_A^2 = \min_{p\in\mathbb{P}_\ell(0)} \|Up(U^\top AU)U^{-1}(x^* - x_0)\|_A^2. \tag{6}$$

Although there is no general method for building a good preconditioner [3, 30], leveraging the convergence properties of CG on (6) often leads to the following criteria: (i) $F$ should approximate the inverse of $A$, (ii) $F$ should be cheap to apply, (iii) $\kappa(U^\top AU)$ should be smaller than $\kappa(A)$, and (iv) $U^\top AU$ should have a more favorable distribution of eigenvalues than $A$.

## 3 A scaled spectral preconditioner

We focus on the spectral preconditioners [4, 12, 14] derived from the spectrum of $A$ that can be expressed as

$$F = I_n + SDS^\top, \tag{7}$$

where $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is a diagonal matrix of rank $k$, $1 < k < n$, such that $I_n + D$ is SPD. Preconditioner $F$ can be expressed in factorized form $F = UU^\top$, with

$$U = S(I_n + D)^{1/2}S^\top, \tag{8}$$

In what follows, we denote by $\hat{x}_\ell(F)$ the $\ell$-th PCG iterate computed with the preconditioner $F$ defined in (7). We also let $\mathbf{PCG}(F)$ denote the PCG algorithm when using $F$ as a preconditioner. The next theorem provides an explicit expression for the energy norm of the error produced by $\mathbf{PCG}(F)$.

**Theorem 1.** *Consider $\mathbf{PCG}(F)$ with $F$ defined in (7). Assume that, at iteration $\ell$, the $\mathbf{PCG}(F)$ has not yet reached the solution. Then for any $j \leq \ell$,*

$$\|x^* - \hat{x}_j(F)\|_A^2 = \sum_{i=1}^n \frac{\eta_i^2}{\lambda_i} \hat{p}_j((d_i + 1)\lambda_i)^2, \tag{9}$$

*where* $\{\hat{p}_j\} = \arg\min_{p\in\mathbb{P}_j(0)} \sum_{i=1}^n \frac{\eta_i^2}{\lambda_i} p((d_i + 1)\lambda_i)^2$.

See proof on page 12.

In the spectral LMP [27] for data assimilation problems, the coefficients $\{d_i\}$ are chosen so that the $k$ largest eigenvalues $\{\lambda_i\}$ are mapped into an existing cluster, typically located at 1, while the remaining eigenvalues remain unchanged. Here, we ask whether a better choice for $d_i$ can be made.

### 3.1 An optimal choice for $D$

We wish to find $D^* = \operatorname{diag}(d_1^*, \dots, d_n^*)$ as a solution of

$$\begin{aligned} \min_{(d_1,\dots,d_n)\in\mathbb{R}^n} \quad & \tfrac{1}{2}\|x^* - \hat{x}_\ell(F)\|_A^2 \\ \text{s.t.} \quad & F = I_n + SDS^\top \quad \text{with } D = \operatorname{diag}(d_1,\dots,d_n), \\ & I_n + D \text{ is positive-definite}, \\ & \operatorname{rank}(D) \le k. \end{aligned} \tag{10}$$

Problem (10) defines the best low-rank update of the identity matrix in the directions of the eigenvectors to minimize the error in energy norm at iteration $\ell$.

In Theorem 2, we will show that the entries of a particular class of solutions can be constructed using $k$ eigenvalues of $A$ and a scaling factor $\theta^*$, which is a root of a specific polynomial defined later. The choice of eigenvalues used in this construction depends on the properties of $A$, $b$, and $\ell$. To identify the appropriate $k$ eigenvalues, we consider the set

$$\Pi_k = \{\pi_k \subset \{1, 2, \dots, n\} \mid |\pi_k| = k\}.$$

For any $\pi_k \in \Pi_k$, the indices in $\pi_k$ determine the $k$ nonzero diagonal elements of $D$, and thus define a preconditioner $F$. The complementary set of $\pi_k$ will be denoted $\bar{\pi}_k$.

To prove Theorem 2, we first present a lemma on the existence and unicity of the polynomial minimizing,

$$\min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2, \tag{11}$$

which is important for constructing the solution in Theorem 2.

**Lemma 1.** *Consider* $\mathbf{PCG}(F)$ *with* $F$ *feasible for* (10). *Assume that, at iteration* $\ell$, *the* $\mathbf{PCG}(F)$ *has not yet reached the solution. Then, for any* $\pi_k \in \Pi_k$, *there exists a unique solution to* (11) *which has only positive roots.*

See proof on page 12.

In the rest of the paper, we assume that $\ell \ge 1$ and that at iteration $\ell$, $\mathbf{PCG}(F)$ has not yet reached the solution with $F$ feasible for (10).

**Theorem 2.** *Let* $\pi_k^* \in \arg\min_{\pi_k\in\Pi_k} \left\{ \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \right\}$, *and* $\hat{p}_\ell \in \mathbb{P}_\ell(0)$, *with a positive root* $\theta^* > 0$ *and such that* $\{\hat{p}_\ell\} = \arg\min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$.

*Then,* $d^* := (d_1^*, \dots, d_n^*)$, *where* $d_i^* := \begin{cases} \frac{\theta^*}{\lambda_i} - 1, & \text{if } i \in \pi_k^*, \\ 0, & \text{if } i \in \bar{\pi}_k^*, \end{cases}$, *solves* (10).

*In addition,*

$$\|x^* - \hat{x}_\ell(F^*)\|_A^2 = \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\lambda_i)^2, \tag{12}$$

*where* $F^* := I_n + SD^*S^\top$ *with* $D^* = \operatorname{diag}(d_1^*, \dots, d_n^*)$.

See proof on page 13.

The following corollary provides an explicit expression for $F^*$ together with a square root decomposition.

**Corollary 1.** *Solution of (10) can be expressed as*

$$F^* := F(\theta^*, \pi_k^*) = I_n + \sum_{i \in \pi_k^*} \left( \frac{\theta^*}{\lambda_i} - 1 \right) s_i s_i^\top = U^* {U^*}^\top, \tag{13}$$

*where* $U^* = {U^*}^\top := I_n + \sum_{i \in \pi_k^*} \left( \sqrt{\frac{\theta^*}{\lambda_i}} - 1 \right) s_i s_i^\top$, *and* $\pi_k^*$ *and* $\theta^*$ *are defined in Theorem 2.*

Note that $F^*$ maps eigenvalues $\{\lambda_i\}_{i \in \pi_k^*}$ at $\theta^*$, while leaving the rest of the spectrum unchanged. Specifically,

$$F^* A = \theta^* \sum_{i \in \pi_k^*} s_i s_i^\top + \sum_{i \in \bar{\pi}_k^*} \lambda_i s_i s_i^\top.$$

Hence, the spectrum is $\{\theta^* \text{ (with multiplicity } k)\} \cup \{\lambda_i \mid i \in \bar{\pi}_k^*\}$. The following corollary states that $\mathbf{PCG}(F^*)$ guarantees a reduction in the energy norm of the error compared to the unpreconditioned case. In fact, since $D = 0$ is feasible for (10), the $\mathbf{PCG}(I_n)$ iterates coincide with those of CG. Hence, by the optimality of $F^*$ for (10), the result follows directly.

**Corollary 2.** $\mathbf{PCG}(F^*)$ *leads to an improved reduction in the energy norm of the error compared to the unpreconditioned case:*

$$\|x^* - \hat{x}_\ell(F^*)\|_A^2 \leq \|x^* - x_\ell\|_A^2. \tag{14}$$

Inequality (14) is particularly relevant for applications in which the reduction of the error in the energy norm is a primary objective and PCG is terminated prior to convergence.

The optimal set $\pi_k^*$ and the value of $\theta^*$ depend on the distribution of the eigenvalues $\{\lambda_i\}$, the components of the initial residual $\{\eta_i\}$, and $\ell$. Thus, computing $\pi_k^*$ and $\theta^*$ can be computationally infeasible in practical applications. In the next section, we propose a practical choice of $\pi_k^*$ obtained by minimizing the condition number; a closely related idea appears in [1] in the context of hierarchical matrices. Practical strategies for selecting the scaling parameter, as alternatives to $\theta^*$, are also proposed.

## 3.2 A practical choice for $\pi_k^*$

We first derive an upper bound on $\|x^* - \hat{x}_\ell(F^*)\|_A$ based on the condition number.

**Lemma 2.** *Let* $\kappa_{\bar{\pi}_k} := \max_{i \in \bar{\pi}_k} \lambda_i / \min_{i \in \bar{\pi}_k} \lambda_i$ *be defined for any* $\pi_k \in \Pi_k$. *Then,*

$$\|x^* - \hat{x}_\ell(F^*)\|_A \leq 2 \left( \frac{\sqrt{\kappa_{\bar{\pi}_k^a}} - 1}{\sqrt{\kappa_{\bar{\pi}_k^a}} + 1} \right)^\ell \|x^* - x_0\|_A,$$

*where* $\pi_k^a \in \arg \min_{\pi_k \in \Pi_k} \kappa_{\bar{\pi}_k}$.

See proof on page 13.

The next proposition gives a characterization of $\pi_k^a$.

**Proposition 1.** *Let* $\mathcal{J} = \arg \min_{1 \leq j \leq k+1} \lambda_j / \lambda_{n-k+j-1}$, $j_0 \in \mathcal{J}$, *and* $\pi_k \in \Pi_k$ *such that* $\bar{\pi}_k = \{j_0, j_0 + 1, \ldots, j_0 + n - k - 1\}$. *Then,* $\pi_k \in \arg \min_{\pi_k' \in \Pi_k} \kappa_{\bar{\pi}_k'}$.

See proof on page 14.

Proposition 1 shows that once an index $j_0 \in \mathcal{J}$ has been identified, an optimal $\pi_k^a$ can be constructed simply by ensuring that the smallest and largest values of $\bar{\pi}_k^a$ correspond to $j_0$ and $n-k-1+j_0$, respectively. As a result, the preconditioner can be constructed by using $\pi_k^a$ as follows

$$F_\theta := F(\theta, \pi_k^a) \tag{15}$$

where $\theta > 0$, and strategies for selecting it will be presented later.

Let $j_0 \in \mathcal{J}$ such that $\bar{\pi}_k^a = \{j_0, j_0+1, \ldots, n-k+j_0-1\}$. The value of $j_0$ determines which part of the eigenspectrum is used in constructing the preconditioner. There are three possible cases:

*Case 1.* Using the largest $k$ eigenvalues: $j_0 = k+1$ and $\pi_k^a = \{1, 2, \ldots, k\}$.

*Case 2.* Using the smallest $k$ ones: $j_0 = 1$ and $\pi_k^a = \{n-k+1, n-k+2, \ldots, n\}$.

*Case 3.* Using a mixture of smaller and larger ones: $1 < j_0 < k+1$ and $\pi_k^a = \{1, \ldots, j_0-1\} \cup \{n-k+j_0, \ldots, n\}$.

Now that $\pi_k^a$ has been set according to the application context (*Cases 1, 2*, or *3*), we provide a practical estimate of $\theta$ in the next section. In particular, we present four strategies for selecting $\theta$.

# 4 On the choice of the scaling parameter $\theta$

## 4.1 $\theta$ as the mid-range between $\lambda_{j_0}$ and $\lambda_{n-k+j_0-1}$

Our goal is to select $\theta$ such that the resulting $\mathbf{PCG}(F_\theta)$ iterates yield an error comparable to (11) with $\pi_k = \pi_k^a$. We begin by examining the connection with deflation methods [20, 24]. The deflation method, when used with the deflation subspace constructed from the eigenvectors $(s_i)_{i\in\pi_k^a}$ corresponding to the eigenvalues $(\lambda_i)_{i\in\pi_k^a}$, generates iterates

$$x_\ell^{\mathrm{D}} = \sum_{i\in\pi_k^a} \frac{1}{\lambda_i} s_i s_i^\top b + P z_\ell, \tag{16}$$

where $P = I_n - \sum_{i\in\pi_k^a} s_i s_i^\top$ is the orthogonal projection onto $\mathrm{span}\{s_i \mid i \in \bar{\pi}_k^a\}$ and $z_\ell$ designates the iterate generated by CG when solving the projected system (see Proposition 2)

$$PA\, z = Pb \tag{17}$$

starting from $z_0 = x_0$.

The following theorem provides the polynomial expression for $\left\|x^* - x_\ell^{\mathrm{D}}\right\|_A^2$.

**Theorem 3.** *The energy norm of the error for $x_\ell^D$ defined in (16) is given by*

$$\left\|x^* - x_\ell^D\right\|_A^2 = \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2. \tag{18}$$

See proof on page 14.

Let $\{p_\ell^{\mathrm{D}}\} := \arg\min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$ denote the polynomial that attains (18). The following theorem presents the main result of this section.

**Theorem 4.** *Let $F_\theta$ be defined as (15) with $\theta > 0$. Let $x_\ell^D$ be given by (16). Then,*

$$\left\|x^* - x_\ell^D\right\|_A \leq \|x^* - \hat{x}_\ell(F_\theta)\|_A \leq \frac{\alpha(\theta)}{\theta} \left\|x^* - x_{\ell-1}^D\right\|_A, \tag{19}$$

*with* $\alpha(\theta) = \max\left(|\lambda_{j_0} - \theta|, |\theta - \lambda_{n-k+j_0-1}|\right)$.

See proof on page 14.

Note that choosing $\theta > 0$ such that $\alpha(\theta)/\theta \geq 1$ in (19) would give a pessimistic upper bound. For a better bound, we select $\theta > 0$ such that $\alpha(\theta)/\theta < 1$, which is equivalent to imposing $\theta > \lambda_{j_0}/2$. The value of $\theta$ that minimizes $\alpha(\theta)/\theta$ is $\theta_{\mathrm{m}} = (\lambda_{j_0} + \lambda_{n-k+j_0-1})/2$, for which $\alpha(\theta_{\mathrm{m}})/\theta_{\mathrm{m}} = (\lambda_{j_0} - \lambda_{n-k+j_0-1})/(\lambda_{j_0} + \lambda_{n-k+j_0-1}) < 1$.

In practice, we do not have access to $\lambda_{j_0}$ and $\lambda_{n-k+j_0-1}$, so we propose practical choices of $\theta$ such that $\alpha(\theta)/\theta < 1$, depending on which eigenvalues are used to build the preconditioner:

*Case 1.* Using the largest eigenvalues, i.e., $j_0 = k+1$. If the smallest eigenvalue can be estimated. Then, practical choices are $\theta = \big(\lambda_k + \lambda_n\big)/2$ or $\theta = \lambda_k$.

*Case 2.* Using the smallest eigenvalues, i.e., $j_0 = 1$. If the largest eigenvalue can be estimated. Then, practical choices are $\theta = \big(\lambda_1 + \lambda_{n-k+1}\big)/2$ or $\theta = \lambda_1$.

*Case 3.* Using a mixture of largest and smallest eigenvalues, i.e., $1 < j_0 < k+1$. Then, two practical choices are $\theta = \big(\lambda_{j_0-1} + \lambda_{n-k+j_0}\big)/2$ or $\theta = \lambda_{j_0-1}$.

### 4.2 $\theta$ with respect to the initial iteration

In this section, we investigate the choice of $\theta$ as the root of the polynomial $p_1^D$ given by

$$\{p_1^D\} = \arg \min_{p \in \mathbb{P}_1(0)} \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2. \tag{20}$$

**Theorem 5.** *Let $\theta_1$ be the root of $p_1^D$. Then, $\|x^* - \hat{x}_1(F_{\theta_1})\|_A^2 = \|x^* - x_1^D\|_A^2$.*

See proof on page 15.

An explicit form of $\theta_1$ is provided in the following theorem.

**Theorem 6.** $\theta_1 = \dfrac{r_0^\top A r_0 - \sum_{i\in\pi_k^a} \lambda_i (s_i^\top r_0)^2}{r_0^\top r_0 - \sum_{i\in\pi_k^a} (s_i^\top r_0)^2}$ *is the root of $p_1^D$.*

See proof on page 15.

In the next subsection, we investigate a choice of $\theta$ that yields a smaller error than the unpreconditioned iterate at every $\mathbf{PCG}(F_\theta)$ iterate.

### 4.3 $\theta$ providing lower error in energy norm

In this section, we focus on the first case where $j_0 = k+1$, i.e., the preconditioner is constructed using the largest k eigenvalues. We present the analysis only for this case. The other cases can be treated in a similar manner.

We now characterize the values of $\theta$ for which the preconditioned iterates achieve a smaller error than the unpreconditioned ones. Specifically, we focus on the interval $\theta \in [\lambda_{k+1}, \lambda_k]$ and show that, for such choices, there exists a polynomial that promotes favorable $\mathbf{PCG}(F_\theta)$ convergence.

**Lemma 3.** *For any $\theta \in [\lambda_{k+1}, \lambda_k]$, and any polynomial $p$ of degree $\ell$ such that $p(0) = 1$ and whose roots all lie in $[\lambda_n, \lambda_1]$, there exists a polynomial $\hat{p}$ of degree $\ell$ such that $\hat{p}(0) = 1$ and*

$$\begin{aligned} |\hat{p}(\theta)| &\leq |p(\lambda_i)|, \quad i = 1, \dots, k \\ |\hat{p}(\lambda_i)| &\leq |p(\lambda_i)|, \quad i = k+1, \dots, n. \end{aligned}$$

See proof on page 15.

Now, we can present a result that enables comparing the error in energy norm between the preconditioned and the unpreconditioned iterates.

**Theorem 7.** *Let $\theta \in [\lambda_{k+1}, \lambda_k]$. Then, $\|x^* - \hat{x}_\ell(F_\theta)\|_A \leq \|x^* - x_\ell\|_A$.*

See proof on page 16.

Theorem 7 offers a range of choices for $\theta$. Let us remind that to construct $F_\theta$, we are given $k$ eigenpairs. As a result, one practical choice is $\theta = \lambda_k$.

In the next section, we present and analyze the choice of setting $\theta = \lambda_n$.

### 4.4 $\theta$ as the smallest eigenvalue

For matrices of the form $A = \rho I_n + X$ (with $X \in \mathbb{R}^{n\times n}$ symmetric positive semidefinite and rank deficient), it is common to choose $\theta = \lambda_n$ [14]. Such a structure of $A$ arises naturally in regularized least-squares problems, where $\lambda_n = 1$ [27]. Motivated by this strategy of choosing $\theta$ as the smallest eigenvalue, we analyze in the next theorem how the error in the energy norm along the iterates of $\mathbf{PCG}(F_{\lambda_n})$ can be compared to that of deflated CG.

**Theorem 8.** *Let $\varepsilon \geq 0$ be a fixed tolerance and assume that $s_n^\top r_0 = \eta_n \neq 0$. There exists an iterate $\ell_0$ of deflated CG such that $|\lambda_n - v_{\ell_0}^{(\ell_0)}| \leq \varepsilon$, where $v_{\ell_0}^{(\ell_0)}$ is the smallest root of the polynomial $\{p_{\ell_0}^D\} = \arg\min_{p\in\mathbb{P}_{\ell_0}(0)} \sum_{i\in\bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$. Let $x_\ell^D$ be generated as (16). Then, the following bounds hold:*

$$\begin{cases} \|x^* - \hat{x}_\ell(F_{\lambda_n})\|_A^2 \leq \|x^* - x_\ell^D\|_A^2 + \sum\limits_{i\in\pi_k^a} \frac{\eta_i^2}{\lambda_i}, & \text{if } \ell < \ell_0, \\ \|x^* - \hat{x}_\ell(F_{\lambda_n})\|_A^2 \leq \|x^* - x_\ell^D\|_A^2 + \frac{\varepsilon^2}{\lambda_n^2} \sum\limits_{i\in\pi_k^a} \frac{\eta_i^2}{\lambda_i}, & \text{if } \ell \geq \ell_0. \end{cases} \tag{21}$$

See proof on page 16.

Theorem 8 shows two-phase convergence behavior. It states that if $\ell < \ell_0$, the terms $\sum_{i\in\pi_k^a} \frac{\eta_i^2}{\lambda_i}$ may slow down convergence and even lead to worse results compared to the unpreconditioned case. However, if $\ell \geq \ell_0$, the behavior is expected to resemble that of deflated CG.

## 5 Numerical experiments

### 5.1 Approximating largest and smallest eigenvalues

In practice, the largest eigenvalues can be approximated when solving sequences of linear least-squares problems arising in nonlinear least-squares problems [27]. The Lanczos process underlying CG can be exploited on each linear problem to extract approximate spectral information [25]. This spectral information is used to design a preconditioner for the subsequent linear least-squares problem [27]. Another approach to approximate the largest eigenvalues is randomized eigenvalue decomposition [17], which has been shown to be efficient for constructing preconditioner to solve SPD linear systems [9].

For the smallest eigenvalues, in the context of solving linear systems with multiple right-hand sides, Stathopoulos and Orginos [26] proposed a CG-based approach to approximate the smallest eigenvalues and their corresponding eigenvectors relying on Rayleigh–Ritz projections and to reuse them as a deflation subspace for later linear systems. There is another way to approximate the smallest eigenvalues and their corresponding eigenvectors by using harmonic projection techniques [22, 24]. A comparative study between Rayleigh–Ritz and harmonic projections in the context of approximating the smallest eigenpairs was presented in [29].

### 5.2 Computational complexity

The computational cost of calculating the eigenpairs depends on the strategy employed and on which eigenpairs are targeted. For instance, approximating eigenspectrum from CG coefficients at iteration

$k$ requires $\mathcal{O}(nk^2)$ flops, which accounts for the eigendecomposition of a $k \times k$ tridiagonal matrix as well as the matrix-vector products with the $n$-dimensional Lanczos vectors (normalized residual vectors produced by CG) [8]. When randomized algorithms are used to approximate the eigenpairs, the overall cost is dominated by at least one matrix–vector product with $A$ [17]. A key advantage of randomized methods is that they are parallelizable and can provide spectral information in advance for PCG.

Applying $F_\theta$ requires storing the $k$ selected eigenvectors $\{s_i\}_{i\in\pi_k^a}$, the associated eigenvalues $\{\lambda_i\}_{i\in\pi_k^a}$, and the scalar parameter $\theta$. This results in a memory cost of $\mathcal{O}(kn)$. The total computational cost of performing matrix-vector products with $F_\theta$ is $\mathcal{O}(kn)$ flops [14]. There is an extra cost associated with constructing $\theta_1$, but it is dominated by a single multiplication with the matrix $A$.

## 5.3 Experimental setup

We restrict our attention to $Ax = b$, where $A \in \mathbb{R}^{n\times n}$ is a diagonal matrix with $n = 10^6$. In this setting, the eigenvalues used to build the preconditioner are exact, and the preconditioner $F_\theta$ is therefore diagonal. We consider the eigenvalue distribution [16]:

$$\lambda_i = \lambda_n + \left(\frac{n-i}{n-1}\right)(\lambda_1 - \lambda_n)\rho^{i-1}, \quad \text{for} \quad i = 1, \dots, n, \tag{22}$$

with $\lambda_1 = 10^6, \quad \lambda_n = 1$ and $\rho = 0.75$. The preconditioner $F_\theta$ is constructed using $k \in \{30, 40, 50\}$ largest eigenvalues. We denote $\theta_{\mathrm{m}} = \big(\lambda_k + \lambda_n\big)/2$ and $\theta_{\mathrm{r}} = \lambda_k$.

The choices for $\theta_{\mathrm{m}}$ and $\theta_{\mathrm{r}}$ are motivated from Section 4.1. In addition, further theoretical results are provided for $\theta_{\mathrm{r}}$ in Section 4.3.

We use $b = [1, \dots, 1]^\top/\sqrt{n}$, and $x_0 = 0$. We compare the performance of the methods of Table 1 in terms of the relative error in energy norm at each iteration.

**Table 1: Description of methods used in the numerical experiments.**

| Method | Description |
|---|---|
| CG | Algorithm 1 applied to $Ax = b$ |
| $\mathbf{PCG}(F_{\theta_{\mathrm{r}}})$ | Algorithm 2 applied to $Ax = b$ using $F = F_{\theta_{\mathrm{r}}}$ |
| $\mathbf{PCG}(F_{\theta_1})$ | Algorithm 2 applied to $Ax = b$ using $F = F_{\theta_1}$ |
| $\mathbf{PCG}(F_{\theta_{\mathrm{m}}})$ | Algorithm 2 applied to $Ax = b$ using $F = F_{\theta_{\mathrm{m}}}$ |
| **DefCG** | Algorithm 1 applied to the projected system (17) |

In all plots below, the eigenvalues are indexed by $u = (i - \frac{1}{2})/N$ for $i = 1, \dots, N$, so that $u \in (0, 1)$ represents the relative position of the $i$-th eigenvalue in the ordered spectrum. The abscissa is displayed on a logit scale, $\operatorname{logit}(u) = \log(u/(1-u))$, which expands both ends of the interval $(0, 1)$. This transformation enhances the visibility of extreme eigenvalues (corresponding to very small or very large indices) by spreading them apart, while keeping the bulk of the spectrum compressed in the center.

In the supplementary material [7], we present experiments for other eigenvalue distributions, allowing us to recover the different cases described in Section 3.2.

## 5.4 Numerical results

Figure 1 shows eigenvalue distributions of the preconditioned matrix $FA$ for different choices for $k$ and $F$. It also shows the convergence of $\mathbf{PCG}(F)$ in terms of the error in the relative energy norm. Note that the CG does not depend on $k$; for this reason, and for clarity of presentation, the corresponding CG results are shown only in the first column of Figure 1 (i.e., $k = 30$).

As shown in Figure 1, all methods require fewer iterations than the unpreconditioned case to achieve an approximation of the solution for which $\|x^* - x_\ell\|_A \approx 10^{-8}\|x^*\|_A$. $\mathbf{PCG}(F_{\theta_{\mathrm{r}}})$ achieves

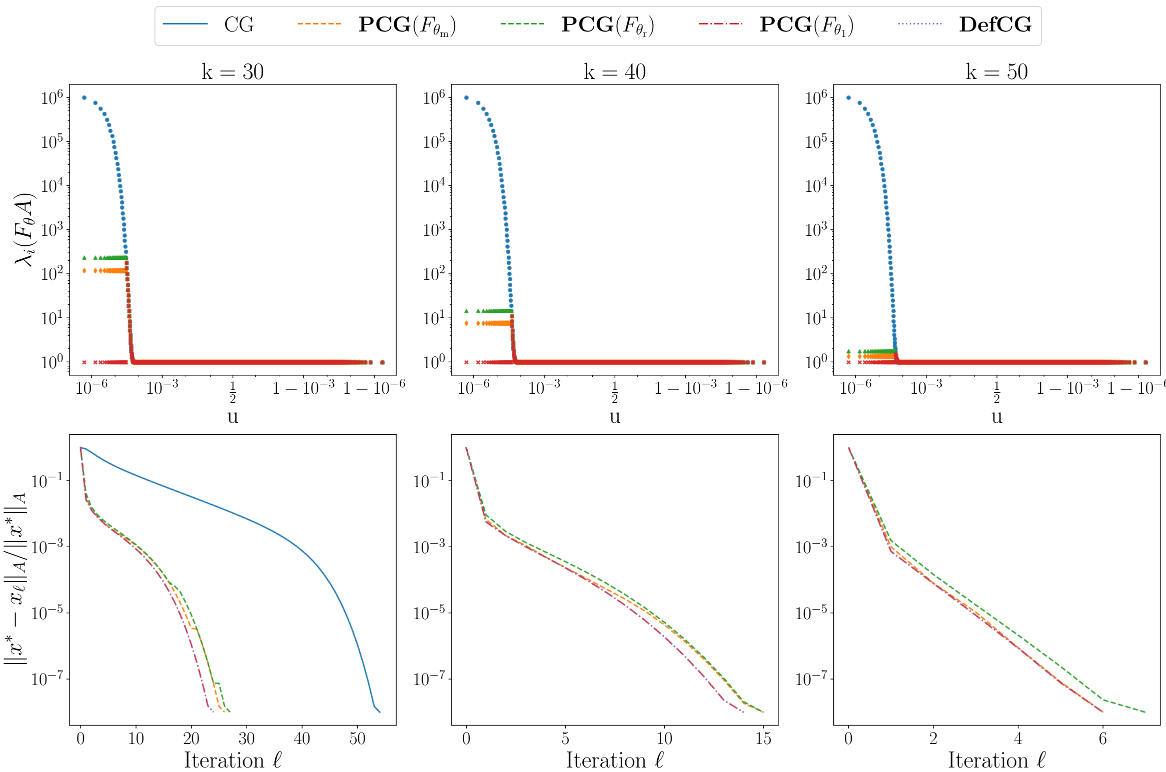


**Figure 1: Top: Eigenvalue distributions of $F_\theta A$ for $k = 30$, $40$, and $50$ (from left to right). Bottom: Relative errors in energy norm versus iteration $\ell$, including DefCG.**

better convergence than CG at all iterates, as guaranteed by Theorem 7. **PCG**($F_{\theta_1}$) achieves better convergence than CG and also outperforms both **PCG**($F_{\theta_r}$) and **PCG**($F_{\theta_m}$). Its convergence remains close to that of **DefCG**. **PCG**($F_{\theta_m}$) produces iterates closer to those of **DefCG**, as this behaviour is explicitly motivated for the choice of $\theta_m$ as explained in Section 4.1. In addition, **PCG**($F_{\theta_m}$) outperforms slightly **PCG**($F_{\theta_r}$), which can be explained by $\frac{\alpha(\theta_m)}{\theta_m} \leq \frac{\alpha(\theta_r)}{\theta_r}$ (See Theorem 4).

## 5.5 Discussion

Both the theoretical analysis and the numerical experiments suggest that deflated CG gives the best result. However, in practice, since only approximate eigenpairs are typically available, deflated CG can become computationally demanding for large-scale problems. Constructing the deflation projector [20] requires storing additional vectors and involves extra multiplications with $A$, as well as additional vector operations. This results in non-negligible overhead, both in memory usage and in the number of matrix-vector products.

One motivation for the scaled spectral preconditioner is to mimic the favorable convergence behavior of deflated CG while significantly reducing this overhead. In follow-up work, we will provide theoretical and numerical result showing that the scaled spectral preconditioner remains robust when approximate eigenpairs are used and can achieve an approximation close to deflated CG at low cost.

## 5.6 Numerical experiments with the choice of $\theta = \lambda_n$

To show the behavior in Section 4.4, we consider the same distribution as (22) with $n = 100, \rho = 0.75, \lambda_1 = 10^4$ and $\lambda_n = 1$. The preconditioner uses the $k = 10$ largest eigenvalues.

We consider two different choices for the right-hand side vector $b$, based on the distribution of $\zeta_i := \frac{\eta_i^2}{\lambda_i}$, defined as follows: (i) **Fast Decay of $\zeta_i$:** a fast decay in the values of $\zeta_i$ is given by $\zeta_i = \zeta_n + (\frac{i-1}{n-1})(\zeta_1 - \zeta_n)0.9^{n-i}$. We set $\zeta_N = 1$ and $\zeta_1 = 10^3$. The right-hand side vector is then defined as $b = [\sqrt{\zeta_1 \lambda_1}, \sqrt{\zeta_2 \lambda_2}, \ldots, \sqrt{\zeta_n \lambda_n}]$. (ii) **Fast Growth of $\zeta_i$:** we reverse the order of $\zeta_i$ defined in the previous case to obtain a fast-growing distribution: $b = [\sqrt{\zeta_n \lambda_1}, \sqrt{\zeta_{n-1} \lambda_2}, \ldots, \sqrt{\zeta_1 \lambda_n}]$.

As shown in Figure 2, **PCG**$(F_{\lambda_n})$ reduces the number of iterations to converge, but it is not necessarily better than CG, especially for $\ell < 15$. We observe that the convergence improves for $\ell \geq 15$ and becomes closer to **DefCG** for $\ell \geq 20$, this can be interpreted as a consequence of the convergence of the smallest Ritz value toward $\lambda_n$ in **DefCG**, in accordance with the results of Theorem 8 (see the supplementary material [7, Fig. SM3]). Similarly, the slow convergence of **PCG**$(F_{\lambda_n})$ in the early iterations can be explained by Theorem 8 where $\sum_{i=1}^{k} \zeta_i$ is large enough to slow down the convergence. For the case with a fast growth distribution of $\zeta_i$, **PCG**$(F_{\lambda_n})$ exhibits faster convergence in the early iterations, similar to **DefCG**, because the terms $\sum_{i=1}^{k} \zeta_i$ in (21) remain very small.

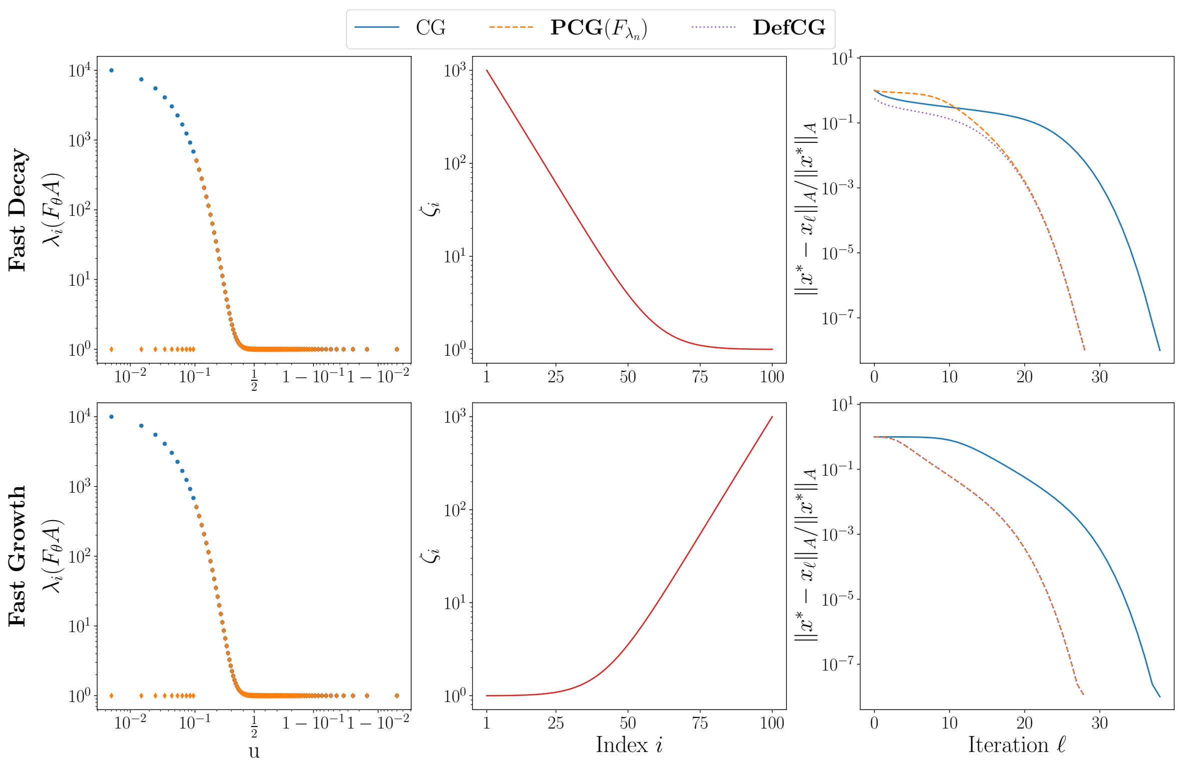


**Figure 2: Eigenvalues of the preconditioned matrix, the values of $\zeta_i$, and the energy norm of the relative error along iterations for fast decay and fast growth of $\zeta_i$.**

# 6 Conclusion and perspectives

We have introduced and analyzed a class of scaled spectral preconditioners for the conjugate gradient method applied to large symmetric positive-definite linear systems with extreme eigenvalues, particularly when the CG number of iterations is limited. Starting from an optimization viewpoint, we derived a preconditioner that minimizes the error after a prescribed number of iterations. A key outcome of our analysis is that the preconditioner should be built from the extreme eigenvalues, and that the position of the resulting eigenvalue cluster can be chosen according to several principled criteria. The numerical experiments on matrices with extreme eigenvalues confirm that the scaled spectral preconditioners can significantly accelerate the rate of convergence of PCG.

In follow-up work, we will provide a detailed theoretical analysis of the preconditioner when it is used for a sequence of SPD matrices, or equivalently, when it is constructed from approximate eigenvalues.

# A Proofs

**Proof of Theorem 1.** Let $U$ be defined as (8). As in (6),

$$\begin{aligned}\|x^* - \hat{x}_j(F)\|_A^2 &= \min_{p\in\mathbb{P}_j(0)} \|Up(U^\top AU)U^{-1}(x^* - x_0)\|_A^2 \\ &= \min_{p\in\mathbb{P}_j(0)} \|p(U^\top AU)(x^* - x_0)\|_A^2 = \min_{p\in\mathbb{P}_j(0)} \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} p((d_i+1)\lambda_i)^2,\end{aligned}$$

where we use the identity $Up(U^\top AU)U^{-1} = UU^{-1}p(U^\top AU)$, since both $A$ and $U$ are diagonalizable in the same eigenbasis. □

**Proof of Lemma 1.** Let us define the SPD matrix $\Sigma := \operatorname{diag}((\lambda_i)_{i\in\bar{\pi}_k})$ and $\eta := [\eta_i]_{i\in\bar{\pi}_k}^\top$. Applying CG to the linear system $\Sigma\, y = \eta$ , with the initial guess $y_0 = 0$, implicitly solves

$$\arg \min_{p\in\mathbb{P}_j(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \tag{23}$$

at the $j$-th iteration, with $j \le \widetilde{\ell}$ and $\widetilde{\ell}$ being the grade of $\eta$ with respect to $\Sigma$. Thus, to prove the existence and uniqueness of the polynomial minimizing (11), it suffices to show that $\widetilde{\ell} > \ell$, since CG produces the unique polynomial minimizing (23).

By contradiction, assume that $\widetilde{\ell} \le \ell$. We first consider the case $\widetilde{\ell} \ge 1$; the particular case $\widetilde{\ell} = 0$ will be handled separately. Let $\{p_{\widetilde{\ell}}\} = \arg\min_{p\in\mathbb{P}_{\widetilde{\ell}}(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$, and $\sigma$ be one of its positive roots (a Ritz value at iteration $\widetilde{\ell}$ when solving $\Sigma\, y = \eta$). Define a preconditioner $\widetilde{F}$ with $d_i = \sigma/\lambda_i - 1$ for $i \in \pi_k$, and $d_i = 0$ for $i \in \bar{\pi}_k$. Then,

$$\begin{aligned}\min_{p\in\mathbb{P}_{\widetilde{\ell}}(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 &\le \min_{p\in\mathbb{P}_{\widetilde{\ell}}(0)} \left[\sum_{i\in\pi_k} \frac{\eta_i^2}{\lambda_i} p(\sigma)^2 + \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2\right] \\ &= \|x^* - \hat{x}_{\widetilde{\ell}}(\widetilde{F})\|_A^2 \\ &\le \sum_{i\in\pi_k} \frac{\eta_i^2}{\lambda_i} p_{\widetilde{\ell}}(\sigma)^2 + \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p_{\widetilde{\ell}}(\lambda_i)^2 = \min_{p\in\mathbb{P}_{\widetilde{\ell}}(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2.\end{aligned}$$

As a result, $\|x^* - \hat{x}_{\widetilde{\ell}}(\widetilde{F})\|_A^2 = \min_{p\in\mathbb{P}_{\widetilde{\ell}}(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 = 0$, since CG applied to the linear system $\Sigma\, y = \eta$ terminates at step $\widetilde{\ell}$. Since $\widetilde{\ell} \le \ell$, the last equality contradicts the assumption that $\hat{x}_\ell(\widetilde{F}) \ne x^*$.

We now consider the case $\widetilde{\ell} = 0$, which occurs only if $\eta = 0$. Define the preconditioner $\widetilde{F}$ with $d_i = \sigma/\lambda_i - 1$ for $i \in \pi_k$, and $d_i = 0$ for $i \in \bar{\pi}_k$ with $\sigma > 0$. From Theorem 1, and since $\eta_i = 0$ for $i \in \bar{\pi}_k$,

$$\|x^* - \hat{x}_1(\widetilde{F})\|_A^2 = \min_{p\in\mathbb{P}_1(0)} \sum_{i\in\pi_k} \frac{\eta_i^2}{\lambda_i} p(\sigma)^2 + \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \le \sum_{i\in\pi_k} \frac{\eta_i^2}{\lambda_i} \hat{p}_1(\sigma) = 0,$$

where $\hat{p}_1(\lambda) = 1 - \frac{\lambda}{\sigma}$. This again contradicts the assumption that $\hat{x}_1(\widetilde{F}) \ne x^*$. Finally, we conclude that $\widetilde{\ell} > \ell$.

Since the roots of the polynomial minimizing (11) correspond to Ritz values generated by CG applied to $\Sigma\, y = \eta$, all the roots are positive. □

**Proof of Theorem 2.** First, the existence of such a $\pi_k^*$ follows from the fact that the set $\Pi_k$ is finite and that, for each $\pi_k$, the inner minimization over $\mathbb{P}_\ell(0)$ is well defined. Let $F = I_n + SDS^\top \in \mathbb{R}^{n\times n}$ with $D = \mathrm{diag}(d_1,\dots,d_n)$ where $(d_1,\dots,d_n)$ is feasible for (10). Let $\mathcal{Z} = \{i \in \{1,\dots,n\} \mid d_i = 0\}$ and $\mathcal{N} = \{i \in \{1,\dots,n\} \mid d_i \neq 0\}$. Then, using Theorem 1,

$$\begin{aligned}\|x^* - \hat{x}_\ell(F)\|_A^2 &= \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\mathcal{N}} \frac{\eta_i^2}{\lambda_i} p((d_i+1)\lambda_i)^2 + \sum_{i\in\mathcal{Z}} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \\ &\geq \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\mathcal{Z}} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2.\end{aligned} \tag{24}$$

Since $D$ has rank at most $k$, $|\mathcal{Z}| \geq n-k$.

Let $\bar{\pi}_k \subseteq \mathcal{Z}$ such that $|\bar{\pi}_k| = n-k$. Then, using (24), the optimality of set $\pi_k^*$, and the fact that $\hat{p}_\ell(\theta^*) = 0$, we get

$$\begin{aligned}\|x^* - \hat{x}_\ell(F)\|_A^2 \geq \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \quad &\geq \quad \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\lambda_i)^2 \\ &= \quad \sum_{i\in\pi_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\theta^*)^2 + \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\lambda_i)^2 \\ &\geq \quad \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\pi_k^*} \frac{\eta_i^2}{\lambda_i} p(\theta^*)^2 + \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \\ &= \quad \min_{p\in\mathbb{P}_\ell(0)} \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} p((d_i^*+1)\lambda_i)^2 \\ &= \quad \|x^* - \hat{x}_\ell(F^*)\|_A^2. \qquad \text{(by Theorem 1)}\end{aligned}$$

Hence,

$$\|x^* - \hat{x}_\ell(F)\|_A^2 \geq \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\lambda_i)^2 \geq \|x^* - \hat{x}_\ell(F^*)\|_A^2. \tag{25}$$

Since (25) holds for any feasible $F$, $d^*$ solves (10). Moreover, as (25) also holds for $F^*$, $\|x^* - \hat{x}_\ell(F^*)\|_A^2 = \sum_{i\in\bar{\pi}_k^*} \frac{\eta_i^2}{\lambda_i} \hat{p}_\ell(\lambda_i)^2$. □

**Proof of Lemma 2.** Let $\pi_k \in \Pi_k$, then by Theorem 2, we get

$$\begin{aligned}\|x^* - \hat{x}_\ell(F^*)\|_A^2 &\leq \min_{p\in\mathbb{P}_\ell(0)} \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \\ &\leq \min_{p\in\mathbb{P}_\ell(0)} \max_{i\in\bar{\pi}_k} p(\lambda_i)^2 \sum_{i\in\bar{\pi}_k} \frac{\eta_i^2}{\lambda_i} \\ &\leq \min_{p\in\mathbb{P}_\ell(0)} \max_{i\in\bar{\pi}_k} p(\lambda_i)^2 \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} \\ &\leq 4\left(\frac{\sqrt{\kappa_{\bar{\pi}_k}}-1}{\sqrt{\kappa_{\bar{\pi}_k}}+1}\right)^{2\ell} \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} = 4\left(\frac{\sqrt{\kappa_{\bar{\pi}_k}}-1}{\sqrt{\kappa_{\bar{\pi}_k}}+1}\right)^{2\ell} \|x^* - x_0\|_A^2,\end{aligned}$$

where $\kappa_{\bar{\pi}_k} := \max_{i\in\bar{\pi}_k} \lambda_i / \min_{i\in\bar{\pi}_k} \lambda_i$ defines the condition number associated with the set of eigenvalues, $(\lambda_i)_{i\in\bar{\pi}_k}$, which are not used in constructing the preconditioner.

Since, $\kappa \mapsto \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2\ell}$ is increasing on $[1, +\infty[$, thus

$$\|x^* - \hat{x}_\ell(F^*)\|_A^2 \leq 4\left(\frac{\sqrt{\kappa_{\bar{\pi}_k^a}}-1}{\sqrt{\kappa_{\bar{\pi}_k^a}}+1}\right)^{2\ell} \|x^* - x_0\|_A^2,$$

where $\pi_k^a \in \arg\min_{\pi_k \in \Pi_k} \kappa_{\bar{\pi}_k}$. □

**Proof of Proposition 1.** Let $\pi'_k \in \Pi_k$ and set $j'_0 = \min_{i \in \bar{\pi}'_k} \{i\}$ and $\bar{j}'_0 = \max_{i \in \bar{\pi}'_k} \{i\}$. Using the fact that $\bar{\pi}'_k \subset \{1, \dots, n\}$ with elements ranging between $j'_0$ and $\bar{j}'_0$, we deduce that $\bar{\pi}'_k \subseteq \{j'_0, j'_0 + 1, \dots, \bar{j}'_0\}$. Since $|\bar{\pi}'_k| = n - k$, we get

$$\bar{j}'_0 \geq j'_0 + n - k - 1 \quad \text{and} \quad j'_0 \leq k + 1.$$

Since $(\lambda_i)_{1 \leq i \leq n}$ are given in decreasing order, $\kappa_{\bar{\pi}'_k} = \frac{\max_{i \in \bar{\pi}'_k} \lambda_i}{\min_{i \in \bar{\pi}'_k} \lambda_i} = \frac{\lambda_{j'_0}}{\lambda_{\bar{j}'_0}} \geq \frac{\lambda_{j'_0}}{\lambda_{j'_0+n-k-1}}$.

Consider $\pi_k \in \Pi_k$ such that $\bar{\pi}_k = \{j_0, j_0 + 1, \dots, j_0 + n - k - 1\}$. Then, for all $\pi'_k \in \Pi_k$, we get

$$\kappa_{\bar{\pi}'_k} \geq \frac{\lambda_{j'_0}}{\lambda_{j'_0+n-k-1}} \geq \frac{\lambda_{j_0}}{\lambda_{j_0+n-k-1}} = \kappa_{\bar{\pi}_k},$$

i.e., $\pi_k \in \arg\min_{\pi'_k \in \Pi_k} \kappa_{\bar{\pi}'_k}$. □

**Proof of Theorem 3.** The exact solution of $A\,x = b$ can be written as $x^* = \sum_{i=1}^{n} \frac{1}{\lambda_i} s_i s_i^\top b$. From (16)

$$x^* - x_\ell^{\mathrm{D}} = \sum_{i \in \bar{\pi}_k^a} \frac{1}{\lambda_i} s_i s_i^\top b - P z_\ell. \tag{26}$$

By Proposition 3, $z_\ell = \sum_{i \in \pi_k^a} s_i s_i^\top x_0 + S_{\bar{\pi}_k^a} \hat{y}_\ell$. Since $P s_i = 0$ for $i \in \pi_k^a$ and $P s_i = s_i$ for $i \in \bar{\pi}_k^a$, we obtain $P z_\ell = S_{\bar{\pi}_k^a}\, \hat{y}_\ell$. Inserting into (26),

$$x^* - x_\ell^{\mathrm{D}} = S_{\bar{\pi}_k^a} \big(\Lambda_{\bar{\pi}_k^a}^{-1} S_{\bar{\pi}_k^a}^\top b - \hat{y}_\ell\big).$$

Therefore, using $S_{\bar{\pi}_k^a}^\top A S_{\bar{\pi}_k^a} = \Lambda_{\bar{\pi}_k^a}$, we obtain, $\|x^* - x_\ell^{\mathrm{D}}\|_A^2 = \|\Lambda_{\bar{\pi}_k^a}^{-1} S_{\bar{\pi}_k^a}^\top b - \hat{y}_\ell\|_{\Lambda_{\bar{\pi}_k^a}}^2$.

CG applied to the reduced system, $\Lambda_{\bar{\pi}_k^a} \hat{y} = S_{\bar{\pi}_k^a}^\top b$, with starting vector $\hat{y}_0 = S_{\bar{\pi}_k^a}^\top x_0$ satisfies

$$\begin{aligned}
\|\Lambda_{\bar{\pi}_k^a}^{-1} S_{\bar{\pi}_k}^\top b - \hat{y}_\ell\|_{\Lambda_{\bar{\pi}_k^a}}^2 &= \min_{p \in \mathbb{P}_\ell(0)} \big\|p(\Lambda_{\bar{\pi}_k^a})(\Lambda_{\bar{\pi}_k^a}^{-1} S_{\bar{\pi}_k^a}^\top b - \hat{y}_0)\big\|_{\Lambda_{\bar{\pi}_k^a}}^2 \\
&= \min_{p \in \mathbb{P}_\ell(0)} \big\|p(\Lambda_{\bar{\pi}_k^a}) \Lambda_{\bar{\pi}_k^a}^{-1} S_{\bar{\pi}_k^a}^\top (b - S_{\bar{\pi}_k^a} \Lambda_{\bar{\pi}_k^a} S_{\bar{\pi}_k^a}^\top x_0)\big\|_{\Lambda_{\bar{\pi}_k^a}}^2 \\
&= \min_{p \in \mathbb{P}_\ell(0)} \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2.
\end{aligned}$$

□

**Proof of Theorem 4.** Let us start by proving the first inequality. Using expression (9) together with the definition of $F_\theta$, we obtain

$$\begin{aligned}
\|x^* - \hat{x}_\ell(F_\theta)\|_A^2 &= \min_{p \in \mathbb{P}_\ell(0)} \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i} p(\theta)^2 + \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \\
&\geq \min_{p \in \mathbb{P}_\ell(0)} \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 = \big\|x^* - x_\ell^{\mathrm{D}}\big\|_A^2,
\end{aligned}$$

where the last equality follows from Theorem 3. To prove the second equality, we define $\widetilde{p}(\lambda) = \big(1 - \frac{\lambda}{\theta}\big)\, p_{\ell-1}^{\mathrm{D}}(\lambda) \in \mathbb{P}_\ell(0)$, where $\{p_{\ell-1}^{\mathrm{D}}\} = \arg\min_{p \in \mathbb{P}_{\ell-1}(0)} \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$. By construction, $\widetilde{p}(\theta) = 0$. Then,

$$\|x^* - \hat{x}_\ell(F_\theta)\|_A^2 = \min_{p \in \mathbb{P}_\ell(0)} \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i} p(\theta)^2 + \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2$$

$$\le \sum_{i\in\pi_k^a} \frac{\eta_i^2}{\lambda_i}\widetilde{p}(\theta)^2 + \sum_{i\in\bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i}\widetilde{p}(\lambda_i)^2 = \sum_{i\in\bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p_{\ell-1}^{\mathrm{D}}(\lambda_i)^2 \left(1-\frac{\lambda_i}{\theta}\right)^2$$

$$\le \max_{i\in\bar{\pi}_k^a}\left(1-\frac{\lambda_i}{\theta}\right)^2 \left\|x^*-x_{\ell-1}^{\mathrm{D}}\right\|_A^2 = \frac{\alpha(\theta)^2}{\theta^2}\left\|x^*-x_{\ell-1}^{\mathrm{D}}\right\|_A^2.$$

Here we have used (18), as well as the identity $\alpha(\theta)/\theta = \max_{i\in\bar{\pi}_k^a}(1-\lambda_i/\theta)$, which follows from the ordering of the eigenvalues. □

**Proof of Theorem 5.** By using the first inequality from Theorem 4 for the particular choice $\ell = 1$,

$$\begin{aligned}\|x^*-x_1^D\|_A^2 &\le \|x^*-\hat{x}_1(F_{\theta_1})\|_A^2\\ &= \min_{p\in\mathbb{P}_1(0)} \sum_{i\in\pi_k^a}\frac{\eta_i^2}{\lambda_i}p(\theta_1)^2 + \sum_{i\in\bar{\pi}_k^a}\frac{\eta_i^2}{\lambda_i}p(\lambda_i)^2\\ &\le \sum_{i\in\pi_k^a}\frac{\eta_i^2}{\lambda_i}p_1^D(\theta_1)^2 + \sum_{i\in\bar{\pi}_k^a}\frac{\eta_i^2}{\lambda_i}p_1^D(\lambda_i)^2 = \|x^*-x_1^D\|_A^2.\end{aligned}$$

□

**Proof of Theorem 6.** From Theorem 5,

$$\|x^*-\hat{x}_1(F_{\theta_1})\|_A^2 = \min_{p\in\mathbb{P}_1(0)} \sum_{i\in\bar{\pi}_k^a}\frac{\eta_i^2}{\lambda_i}p(\lambda_i)^2.$$

This minimization problem is equivalent to performing one iteration of CG on $\Sigma\, y = \eta$, where the SPD matrix is given by $\Sigma = \mathrm{diag}((\lambda_i)_{i\in\bar{\pi}_k^a})$ and $\eta = [\eta_i]_{i\in\bar{\pi}_k^a}^\top$, using the initial guess $y_0 = 0$. The corresponding Ritz value after this first iteration is [25, p.194],

$$\theta_1 = \frac{\eta^\top\Sigma\eta}{\eta^\top\eta} = \frac{\sum_{i\in\bar{\pi}_k^a}\eta_i^2\,\lambda_i}{\sum_{i\in\bar{\pi}_k^a}\eta_i^2}.$$

$\theta_1$ can also be written in terms of $\pi_k^a$, i.e.

$$\theta_1 = \frac{\sum\limits_{i=1}^n \lambda_i\,(s_i^\top r_0)^2 - \sum\limits_{i\in\pi_k^a}\lambda_i\,(s_i^\top r_0)^2}{\sum\limits_{i=1}^n (s_i^\top r_0)^2 - \sum\limits_{i\in\pi_k^a}(s_i^\top r_0)^2} = \frac{r_0^\top A r_0 - \sum\limits_{i\in\pi_k^a}\lambda_i\,(s_i^\top r_0)^2}{r_0^\top r_0 - \sum\limits_{i\in\pi_k^a}(s_i^\top r_0)^2}.$$

□

**Proof of Lemma 3.** Let us denote $(\mu_j)_{1\le j\le\ell}$ the roots of the polynomial $p$ given in decreasing order, so $p(\lambda) = \prod_{i=1}^\ell\left(1-\frac{\lambda}{\mu_i}\right)$ for any $\lambda \ge 0$. Three cases may occur:

Case 1: For all $j \in \{1,\dots,\ell\}$, $\mu_j < \theta$. We choose $\hat{p}(\lambda) = p(\lambda)$. Then for $i \in \{k+1,\dots,n\}$, we have $|\hat{p}(\lambda_i)| = |p(\lambda_i)|$. For $i \in \{1,\dots,k\}$, using the property that $\mu_j < \theta \le \lambda_i$, we obtain

$$1-\frac{\lambda_i}{\mu_j} \le 1-\frac{\theta}{\mu_j} \le 0.$$

Thus, we have $|1-\frac{\theta}{\mu_j}| \le |1-\frac{\lambda_i}{\mu_j}|$, and consequently $|\hat{p}(\theta)| \le |p(\lambda_i)|$.

Case 2: For all $j \in \{1, \ldots, \ell\}$, $\theta \leq \mu_j$. We choose $\hat{p}(\lambda) = \prod_{j=1}^{\ell} \left(1 - \frac{\lambda}{\theta}\right) = \left(1 - \frac{\lambda}{\theta}\right)^{\ell}$. Then simply for $i \in \{1, \ldots, k\}$, $|\hat{p}(\theta)| = 0 \leq |p(\lambda_i)|$. For $i \in \{k+1, \ldots, n\}$, using the property $\lambda_{k+1} \leq \theta \leq \mu_j$, we obtain

$$0 \leq 1 - \frac{\lambda_i}{\lambda_{k+1}} \leq 1 - \frac{\lambda_i}{\theta} \leq 1 - \frac{\lambda_i}{\mu_j}.$$

Therefore, for $i = k+1, \ldots, n$, $|\hat{p}(\lambda_i)| \leq |p(\lambda_i)|$.

Case 3: Let $s \in \{1, \ldots, \ell - 1\}$ such that for $j = 1, \ldots, s$, $\theta \leq \mu_j \leq \lambda_1$, and for $j = s+1, \ldots, \ell$, $\lambda_n \leq \mu_j < \theta$. We choose

$$\hat{p}(\lambda) = \prod_{j=1}^{s} \left(1 - \frac{\lambda}{\theta}\right) \prod_{j=s+1}^{\ell} \left(1 - \frac{\lambda}{\mu_j}\right) = \left(1 - \frac{\lambda}{\theta}\right)^{s} \prod_{j=s+1}^{\ell} \left(1 - \frac{\lambda}{\mu_j}\right).$$

We have $\hat{p}(\theta) = 0$, so $|\hat{p}(\theta)| \leq |p(\lambda_i)|$ for $i \in \{1, \ldots, k\}$. For $i \in \{k+1, \ldots, n\}$ and $j \in \{1, \ldots, s\}$, we have

$$0 \leq 1 - \frac{\lambda_i}{\lambda_{k+1}} \leq 1 - \frac{\lambda_i}{\theta} \leq 1 - \frac{\lambda_i}{\mu_j},$$

because $\lambda_{k+1} \leq \theta \leq \mu_j$. Therefore, for $i = k+1, \ldots, n$, $|\hat{p}(\lambda_i)| \leq |p(\lambda_i)|$. □

**Proof of Theorem 7.** From (3),

$$\|x^* - x_\ell\|_A^2 = \min_{p \in \mathbb{P}_\ell(0)} \|p\,(A)\,(x^* - x_0)\|_A^2 = \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} p_\ell(\lambda_i)^2. \tag{27}$$

From Lemma 3, there exists a polynomial $\hat{p}$ of degree $\ell$ with $\hat{p}(0) = 1$ such that

$$|\hat{p}(\theta)| \leq |p_\ell(\lambda_i)|, \quad i \in \{1, \ldots, k\}$$
$$|\hat{p}(\lambda_i)| \leq |p_\ell(\lambda_i)|, \quad i \in \{k+1, \ldots, n\}.$$

Applying these inequalities to (27) yields

$$\begin{aligned} \|x^* - x_\ell\|_A^2 &= \sum_{i=1}^{n} \frac{\eta_i^2}{\lambda_i} p_\ell(\lambda_i)^2 \geq \sum_{i=1}^{k} \frac{\eta_i^2}{\lambda_i} \hat{p}(\theta)^2 + \sum_{i=k+1}^{n} \frac{\eta_i^2}{\lambda_i} \hat{p}(\lambda_i)^2 \\ &\geq \min_{p \in \mathbb{P}_\ell(0)} \sum_{i=1}^{k} \frac{\eta_i^2}{\lambda_i} p(\theta)^2 + \sum_{i=k+1}^{n} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 = \|x^* - \hat{x}_\ell(F_\theta)\|_A^2. \end{aligned}$$

□

**Proof of Theorem 8.** Let us first show the existence of $\ell_0$. Let $(v_j^{(\ell)})_{1 \leq j \leq \ell}$ denote the roots of $p_\ell^D$ given in decreasing order. Since $\eta_n \neq 0$, the smallest eigenvalue $\lambda_n$ will be reached in at most $n - k$ iterations of deflated CG. Therefore, the set $\{\ell \in \{1, \ldots, n-k\} \mid |\lambda_n - v_\ell^{(\ell)}| \leq \varepsilon\}$ is non-empty. Let us define $\ell_0 = \min\{\ell \in \{1, \ldots, n-k\} \mid |\lambda_n - v_\ell^{(\ell)}| \leq \varepsilon\}$. At the $\ell$-th iterate, $\mathbf{PCG}(F_{\lambda_n})$ satisfies

$$\begin{aligned} \|x^* - \hat{x}_\ell(F_{\lambda_n})\|_A^2 &= \min_{p \in \mathbb{P}_\ell(0)} \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_n)^2 + \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p(\lambda_i)^2 \\ &\leq \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i} p_\ell^D(\lambda_n)^2 + \sum_{i \in \bar{\pi}_k^a} \frac{\eta_i^2}{\lambda_i} p_\ell^D(\lambda_i)^2 \\ &= \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i} p_\ell^D(\lambda_n)^2 + \|x^* - x_\ell^D\|_A^2. \end{aligned} \tag{28}$$

For $\ell < \ell_0$, since $p_\ell^D(\lambda_n)^2 = \prod_{j=1}^{\ell} \left(1 - \frac{\lambda_n}{v_j^{(\ell)}}\right)^2 \leq 1$, we obtain from (28)

$$\|x^* - \hat{x}_\ell(F_{\lambda_n})\|_A^2 \leq \|x^* - x_\ell^D\|_A^2 + \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i}.$$

For $\ell \geq \ell_0$, by using the interlacing property,

$$p_\ell^D(\lambda_n)^2 \leq \left(1 - \frac{\lambda_n}{v_\ell^{(\ell)}}\right)^2 \leq \left(1 - \frac{\lambda_n}{v_{\ell_0}^{(\ell_0)}}\right)^2 \leq \frac{\varepsilon^2}{\lambda_n^2}.$$

As a result, from (28), $\|x^* - \hat{x}_\ell(F_{\lambda_n})\|_A^2 \leq \|x^* - x_\ell^D\|_A^2 + \frac{\varepsilon^2}{\lambda_n^2} \sum_{i \in \pi_k^a} \frac{\eta_i^2}{\lambda_i}$. □

# B Properties of the projected system (17)

**Proposition 2.** *The matrix $PA$ defined in (17) is symmetric positive semi-definite. In addition,*

- $PA = AP = PAP$.
- *The system $PAz = Pb$ is consistent.*

**Proof.** See [6]. Note that in [6], $P$ is defined for a general deflation subspace $W$. □

**Proposition 3.** *The iterate $z_\ell$ generated by CG when solving (17) starting with $z_0 = x_0$ satisfy*

$$z_\ell = \sum_{i \in \pi_k^a} s_i s_i^\top x_0 + S_{\bar{\pi}_k^a} \hat{y}_\ell,$$

*where $S_{\bar{\pi}_k^a} = [s_i]_{i \in \bar{\pi}_k^a}$. Here, $\hat{y}_\ell$ generated with CG when solving $\Lambda_{\bar{\pi}_k^a} \hat{y} = S_{\bar{\pi}_k^a}^\top b$ starting with $\hat{y}_0 = S_{\bar{\pi}_k^a}^\top x_0$ where $\Lambda_{\bar{\pi}_k^a} = \operatorname{diag}((\lambda_i)_{i \in \bar{\pi}_k^a})$.*

**Proof.** From the expression of $PA$, it is easy to see that $PA = \sum_{i \in \bar{\pi}_k^a} \lambda_i s_i s_i^\top$. In addition $P\ b = (I - \sum_{i \in \pi_k^a} s_i s_i^\top) b = \sum_{i \in \bar{\pi}_k^a} s_i s_i^\top b = S_{\bar{\pi}_k^a}^\top S_{\bar{\pi}_k^a} b$. The proof follows directly from [18] by decomposing CG onto the range and null space of $PA$. □

# References


[1] E. Agullo, E. Darve, L. Giraud, and Y. Harness. Low-rank factorizations in data sparse hierarchical algorithms for preconditioning symmetric positive definite matrices. SIAM J. Matrix Anal. Appl., 39(4): 1701–1725, 2018.

[2] B. Beckermann and Arno B. J. Kuijlaars. Superlinear CG convergence for special right-hand sides. Electronic Transactions on Numerical Analysis, 14:1–19, 2002.

[3] M. Benzi. Preconditioning techniques for large linear systems: A survey. J. Comput. Phys., 182(2): 418–477, 2002.

[4] L. Bergamaschi. A survey of low-rank updates of preconditioners for sequences of symmetric linear systems. Algorithms, 13(4), 2020.

[5] B. Carpentieri, I. S. Duff, and L. Giraud. A class of spectral two-level preconditioners. SIAM J. Sci. Comput., 25(2):749–765, 2003.

[6] O. Coulaud, L. Giraud, P. Ramet, and X. Vasseur. Deflation and augmentation techniques in Krylov linear solvers. Dev. Parallel-Distrib.-Grid Cloud Comput. Eng., pages 249–275, 2013.

[7] Y. Diouane, S. Gürol, O. Mouhtal, and D. Orban. Supplementary material for “A spectral preconditioner for the conjugate gradient method with iteration budget”, 2026.

[8] M. Fisher, J. Nocedal, Y. Trémolet, and S. J. Wright. Data assimilation in weather forecasting: a case study in PDE-constrained optimization. Optim. Eng., 10(3):409–426, 2009.

[9] Z. Frangella, J. A. Tropp, and M. Udell. Randomized Nyström preconditioning. SIAM J. Matrix Anal. Appl., 44(2):718–752, 2023.

[10] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. SIAM J. Sci. Comput., 23 (2):442–462, 2001.

[11] A. Gaul, M. H. Gutknecht, J. Liesen, and R. Nabben. A framework for deflated and augmented Krylov subspace methods. SIAM J. Matrix Anal. Appl., 34(2):495–518, 2013.

[12] L. Giraud and S. Gratton. On the sensitivity of some spectral preconditioners. SIAM J. Matrix Anal. Appl., 27(4):1089–1105, 2006.

[13] G. H. Golub and C. F. Van Loan. Matrix Computations. The Johns Hopkins University Press, Baltimore, 4th edition, 2013.

[14] S. Gratton, A. Sartenaer, and J. Tshimanga. On a class of limited memory preconditioners for large scale linear systems with multiple right-hand sides. SIAM J. Optim., 21(3):912–935, 2011.

[15] S. Gratton, D. Titley-Peloquin, Ph. L. Toint, and J. T. Ilunga. Differentiating the method of conjugate gradients. SIAM J. Matrix Anal. Appl., 35(1):110–126, 2014.

[16] A. Greenbaum and Z. Strakoš. Predicting the behavior of finite precision lanczos and conjugate gradient computations. SIAM J. Matrix Anal. Appl., 13(1):121–137, 1992.

[17] N. Halko, PG. Martinsson, and J.A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review, vol. 53, no. 2, pp. 217–288., 2011.

[18] K. Hayami. Convergence of the conjugate gradient method on singular systems. arXiv preprint arXiv:1809.00793, 2018.

[19] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bur. Stand., 49:409–435, 1952.

[20] K. Kahl and H. Rittich. The deflated conjugate gradient method: Convergence, perturbation and accuracy. Linear Algebra Appl., 515:111–129, 2017.

[21] R. Nabben and C. Vuik. A comparison of deflation and the balancing preconditioner. SIAM J. Sci. Comput., 27(5):1742–1759, 2006.

[22] M. L. Parks, E. De Sturler, G. Mackey, D. D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. SIAM J. Sci. Comput., 28(5):1651–1674, 2006.

[23] J. W. Pearson and J. Pestana. Preconditioners for Krylov subspace methods: An overview. GAMM Mitt., 43(4):e202000015, 2020.

[24] Y. Saad, M. Yeung, J. Erhel, and F. Guyomarc'h. A deflated version of the conjugate gradient algorithm. SIAM J. Sci. Comput., 21(5):1909–1926, 2000.

[25] Yousef Saad. Iterative methods for sparse linear systems. SIAM, 2003.

[26] A. Stathopoulos and K. Orginos. Computing and deflating eigenvalues while solving multiple right-hand side linear systems with an application to quantum chromodynamics. SIAM J. Sci. Comput., 32(1): 439–462, 2010.

[27] J. Tshimanga, S. Gratton, A. T. Weaver, and A. Sartenaer. Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. Q. J. R. Meteorol. Soc., 134 (632):751–769, 2008.

[28] A. Van der Sluis and H. A. Van der Vorst. The rate of convergence of conjugate gradients. Numer. Math., 48(5):543–560, 1986.

[29] N. Venkovic, P. Mycek, L. Giraud, and O. Le Maître. Comparative study of harmonic and Rayleigh-Ritz procedures with applications to deflated conjugate gradients. PhD thesis, CERFACS, 2020.

[30] A. J. Wathen. Preconditioning. Acta Numer., 24:329–376, 2015.