Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

CONNECTING KAPORIN'S CONDITION NUMBER AND THE BREGMAN LOG DETERMINANT DIVERGENCE*

ANDREAS A. BOCK† AND MARTIN S. ANDERSEN†

Abstract. This paper presents some theoretical results relating the Bregman log determinant matrix divergence to Kaporin's condition number. These can be viewed as nearness measures between a preconditioner and a given matrix, and we show under which conditions these two functions coincide. We also give examples of constraint sets over which it is equivalent to minimize these two objectives. We focus on preconditioners that are the sum of a positive definite and low-rank matrix, which were developed in a previous work. These were constructed as minimizers of the aforementioned divergence, and we show that they are only a constant scaling from also minimizing Kaporin's condition number. We highlight connections to information geometry and comment on future directions.

Key words. Condition numbers, Bregman divergence, Preconditioner.

AMS subject classifications. 15A12, 15A15, 65F08, 65F55.

1. Introduction.

We study preconditioning of the system

$$Ax = b, \tag{1.1}$$

where A is symmetric positive definite. The purpose of preconditioning is to find a matrix (or operator) P with inverse $H = P^{-1}$ so that the matrix

$$M = H^{\frac{1}{2}} A H^{\frac{1}{2}}$$

is better conditioned, which can accelerate the progress of iterative methods such as the conjugate gradient method (CG) [21]. Throughout, we will use the notation above. We also let I denote the identity matrix, sometimes with a subscript to denote its dimension. The *preconditioned* CG method (PCG) for approximating a solution to $Ax = b$ with a preconditioner P is outlined below:

$$r_0 = b - Ax_0 \tag{1.2a}$$

$$p_0 = Hr_0 \tag{1.2b}$$

for $k = 1, \ldots$

$$\alpha_k = \frac{r_k^* H r_k}{p_k^* A p_k} \tag{1.2c}$$

$$x_{k+1} = x_k + \alpha_k p_k \tag{1.2d}$$

$$r_{k+1} = r_k - \alpha_k A p_k \tag{1.2e}$$

$$\beta_k = \frac{r_{k+1}^* H r_{k+1}}{r_k^* H r_k} \tag{1.2f}$$

$$p_{k+1} = Hr_{k+1} + \beta_k p_k. \tag{1.2g}$$

*Received by the editors on April 22, 2025. Accepted for publication on May 3, 2026. Handling Editor: Heike Fassbender. Corresponding Author: Andreas A. Bock.

†Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark (aasbo@dtu.dk, mskan@dtu.dk).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

Common termination criteria include an upper limit on the number of iterations, or the relative residual reaching some prescribed tolerance.

Recently, the authors studied preconditioners obtained as minimizers of the Bregman log determinant divergence [9, 8]:

$$\mathcal{D}_{\mathrm{LD}}(A,P) = \operatorname{tr}\left(AP^{-1}\right) - \log\det(AP^{-1}) - n,$$

where feasible preconditioners were selected as a sum of a positive definite matrix QQ^* and a low-rank term. We describe this in greater detail in section 4. The results in these papers showed that the divergence is capturing a sense of nearness to A that may be relevant when searching for a preconditioner P for the PCG algorithm, but without any firm theoretical basis connecting $\mathcal{D}_{\mathrm{LD}}(A,P)$ with (1.2). In this note, we establish a relationship between this particular divergence and superlinear convergence of PCG. We highlight the connection between $\mathcal{D}_{\mathrm{LD}}(A,P)$ and Kaporin's condition number [25].

This manuscript is structured as follows. Section 2 begins by discussing different condition numbers and relevant literature. These are generally used to characterize the progress of (1.2), which is covered in subsection 2.1. Next, section 3 contains our main theoretical contributions related to the condition numbers. As mentioned above, we establish a condition on the trace of a preconditioned matrix, which implies that the Bregman log determinant divergence coincides with Kaporin's condition number. Section 4 and subsection 4.1 summarize the low-rank approximation developed in [8]. In subsection 4.2, we show how the preconditioners associated with these approximations can be modified to minimize Kaporin's condition number. Section 5 contains a general result on minimization of the Bregman log determinant divergence and Kaporin's condition number. Section 6 concludes this paper.

2. Spectral and other condition numbers. Let $\sigma(A) = \{\sigma_1(A), \ldots, \sigma_n(A)\}$ denote the similarly ordered singular values of an $n \times n$ matrix A. When A has real eigenvalues, we denote these by $\lambda(A) = \{\lambda_1(A), \ldots, \lambda_n(A)\}$, with $\lambda_1(A) \geq \ldots \geq \lambda_n(A)$.

The condition number of a symmetric positive definite matrix M is defined as

$$\kappa_2(M) = \frac{\sigma_1(M)}{\sigma_n(M)} = \frac{\lambda_1(M)}{\lambda_n(M)}. \tag{2.1}$$

This quantity is sometimes used as a basis for discussion of what is a good preconditioner, when $M = P^{-\frac{1}{2}}AP^{-\frac{1}{2}}$, i.e., a matrix A symmetrically preconditioned by P. An alternative quantity originated with Kaporin in a series of papers [25, 26, 27, 30], with similar approaches dating back to the 80s [13]. Kaporin introduced the following quantity as a function of a symmetric positive definite matrix M:

$$B(M) = \frac{\frac{1}{n}\operatorname{tr}(M)}{\det(M)^{\frac{1}{n}}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\lambda_i(M)}{(\Pi_{i=1}^{n}\lambda_i(M))^{\frac{1}{n}}}, \tag{2.2}$$

and is the arithmetic mean of the eigenvalues of M divided by the geometric mean of the eigenvalues. To the best of our knowledge, (2.2) first appeared in the context of iterative solvers in [25, 27]. See [28] for the extension to asymmetric matrices. We shall refer to (2.2) as *Kaporin's function.* It decreases when eigenvalues of M are clustered and depends only modestly on the smallest eigenvalue if it is well isolated. The latter is a property not shared by $\kappa_2(M)$ in (2.1). Kaporin's functional is also closely related to factored sparse approximate inverses (FSAI) [22, 23, 32, 33, 34, 42]. Equation (2.2) leads to Kaporin's condition number:

$$K(M) = B(M)^n, \tag{2.3}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

which is quasi-convex in M and satisfies the following properties [30], [3, Theorem 13.5]:

$$K(M) \geq 1, \tag{2.4a}$$
$$K(cM) = K(M), \quad \forall c > 0, \tag{2.4b}$$
$$K(X^{-1}MX) = K(M), \quad \forall X \in \mathbb{H}^n_{++}, \tag{2.4c}$$
$$B(M) \leq \kappa_2(M) \leq \left(\kappa_2(M)^{\frac{1}{2}} + \kappa_2(M)^{-\frac{1}{2}}\right)^2 \leq 4K(M). \tag{2.4d}$$

Here, $\mathbb{H}^n_{++}$ is the positive definite cone. Equation (2.4a) follows from the arithmetic-geometric mean inequality, with equality if and only if all the eigenvalues of M are equal. While $\kappa_2(M) = \kappa_2(M^{-1})$, in general,

$$K(M) \neq K(M^{-1}).$$

$B(M)$ has been studied in the optimization literature [11], where it is referred to as an ω condition number. In a recent contribution, [24] minimize this quantity over low rank updates to positive definite matrices.

We also mention that when $n = 2$, then $B(M) = \cos\phi(M)^{-1}$, where $\cos\phi(M)$ is the first antieigenvalue of M in the sense of Gustafson [16, 17]. See [18] for a connection between operator trigonometry and preconditioning.

Finally, we introduce the Bregman log determinant divergence between two matrices $A \in \mathbb{H}^n_{++}$, $P \in \mathbb{H}^n_{++}$ whose eigendecompositions are given by $A = \mathsf{U}\Xi\mathsf{U}^*$ and $P = \mathsf{V}\Omega\mathsf{V}^*$ [35]:

$$\begin{aligned}\mathcal{D}_{\mathrm{LD}}(A, P) &= \operatorname{tr}\left(AP^{-1}\right) - \log\det(AP^{-1}) - n \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathsf{u}_i^*\mathsf{v}_j)^2\left(\frac{\xi_i}{\omega_j} - \ln\left(\frac{\xi_i}{\omega_j}\right) - 1\right).\end{aligned} \tag{2.5}$$

We shall also refer to this as the *Bregman divergence* for brevity, since (2.5) is the only type of Bregman divergence we consider in this paper. See [1] for other divergences. While this quantity is not a condition number, since $\mathcal{D}_{\mathrm{LD}}(A, P)$ is not always greater than 1, we shall use it to express several results related to the convergence rate of PCG.

2.1. Some convergence results. In this section, we present several well-known results related to PCG in terms of the quantities introduced above. The literature on convergence rates of PCG is vast, so we cover only the essentials. Much material here is covered, or expanded upon in great detail, in [3, 15, 37].

Recall the symmetrically preconditioned matrix is denoted

$$M = P^{-\frac{1}{2}}AP^{-\frac{1}{2}}.$$

The arguably most well-known convergence result is in terms of (2.1):

$$\frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq \frac{2}{C^k + C^{-k}} \leq 2C^k, \tag{2.6}$$

where

$$C = \frac{\sqrt{\kappa_2(M)} - 1}{\sqrt{\kappa_2(M)} + 1}.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

Equation (2.6) is, in general, a pessimistic estimate. The associated estimate of PCG iterations $i_{\ell^2}(\epsilon)$ needed for a ϵ-reduction in $\|x - x_0\|_A$, i.e.,

$$\|x - x_{i_{\ell^2}(\epsilon)}\|_A \leq \epsilon \|x - x_0\|_A,$$

where x_0 is an initial guess, is therefore bounded from above by [5, Equation 8]:

$$i_{\ell^2}(\epsilon) = \lceil \frac{1}{2}\sqrt{\kappa_2(M)} \ln\left(\frac{2}{\epsilon}\right)\rceil.$$

It is well known that CG can exhibit superlinear convergence, and several successful attempts, both theoretical and qualitative, at describing this phenomenon have been made since the 1950s [7, 20, 41]. A lesser-known bound can also be expressed in terms of $K(M)$ [30, Theorem 3.1]:

$$\|r_k\|_{P^{-1}} \leq (K(M)^{1/k} - 1)^{k/2} \|r_0\|_{P^{-1}}. \tag{2.7}$$

Equation (2.7) is only useful if

$$\sqrt[n]{K(M)} = B(M) < 2, \tag{2.8}$$

or, equivalently,

$$\log_2 K(M) < n.$$

It is shown in [30, Theorem 3.2] that for any $k \in \mathbb{N}$ and positive β (subject to some mild conditions), there exists a matrix P^{-1} with $B(P^{-\frac{1}{2}} A P^{-\frac{1}{2}}) = \beta$ and an initial residual r_0 such that

$$\|r_k\|_{P^{-1}} = (\beta^{n/k} - 1)^{k/2} \|r_0\|_{P^{-1}}. \tag{2.9}$$

It is, in this sense, an *unimprovable* estimate. The following upper bound on the number of iterations required for an ϵ-reduction in the residual is given in [30, Theorem 4.1] and is valid for any $\sigma \geq 2$:

$$i_K(\epsilon) = \lceil \frac{\sigma \ln(K(M)) + 2\ln(\epsilon^{-1})}{\sigma \ln(\sigma) - (\sigma - 1)\ln(\sigma - 1)} \rceil. \tag{2.10}$$

For $\sigma = 2$, this simplifies to

$$i_K(\epsilon) \leq \lceil n \log_2(B(M)) + \log_2(\epsilon^{-1})\rceil,$$

which is reportedly quite accurate when $n\log_2(B(M)) \gg \log_2(\epsilon^{-1})$, although it is reported that the choice

$$\sigma = 2 + \frac{\log(\epsilon^{-1})}{\ln(K(M))}$$

provides a more precise approximation. As pointed out in [6], when $B(M) \geq 1 + c$, for some $c > 0$, then

$$i_K(\epsilon) \leq nc + \log_2(\epsilon^{-1}),$$

which, in terms of $B(M)$, is a pessimistic a priori estimate on $i_K(\epsilon)$. In [4], several sublinear and superlinear convergence results are reported for a variety of approaches and condition numbers, one of which is

$$\|x_k - x\|_A \leq \left(\frac{3\ln(K(M))}{k}\right)^{k/2} \|x - x_0\|_A, \tag{2.11}$$

which is asymptotically worse than the bound in (2.9). See [5] for more results on iteration and a posteriori error norm estimates and [2] for the study of the rate of convergence for PCG based on different condition numbers.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

3. Kaporin's condition number and the log determinant divergence. We now establish some relations between Kaporin's condition number (2.3) and the Bregman log determinant divergence (2.5).

THEOREM 1.

$$\mathcal{D}_{\mathrm{LD}}(A,P) \geq \ln K(AP^{-1}), \quad \forall A,P \in \mathbb{H}^n_{++}. \tag{3.1}$$

In addition, when

$$\operatorname{tr}\left(AP^{-1}\right) = n, \tag{3.2}$$

we have

$$\mathcal{D}_{\mathrm{LD}}(A,P) = \ln K(P^{-1}A). \tag{3.3}$$

Proof. The inequality Equation (3.1) follows since

$$\frac{1}{n}\operatorname{tr}\left(P^{-1}A\right) - 1 \geq \ln\left(\frac{1}{n}\operatorname{tr}\left(P^{-1}A\right)\right).$$

For equality in Equation (3.1), we see using (3.2), and $M = P^{-\frac{1}{2}}AP^{-\frac{1}{2}}$, that

$$\begin{aligned}
\ln(K(M)) &= n\ln\left(\frac{1}{n}\operatorname{tr}(M)\right) - n\frac{1}{n}\ln(\det(M)) \\
&= n\ln\left(\frac{1}{n}n\right) - \ln(\det(M)) \\
&= -\ln(\det(M)) \\
&= \mathcal{D}_{\mathrm{LD}}(A,P).
\end{aligned}$$

∎

The Jacobi scaling of a matrix $A \in \mathbb{H}^n_{++}$ by the diagonal matrix $\operatorname{diag}(A)$, $\operatorname{diag}(A)_{ii} = A_{ii}$, $i = 1,\ldots,n$, is given by

$$\operatorname{diag}(A)^{-\frac{1}{2}} A \operatorname{diag}(A)^{-\frac{1}{2}}.$$

This scaling is often used to reduce the condition number of a matrix [15]. Note that

$$\operatorname{tr}\left(\operatorname{diag}(A)^{-\frac{1}{2}} A \operatorname{diag}(A)^{-\frac{1}{2}}\right) = n.$$

The importance of this scaling in constructing FSAI preconditioners has also been highlighted [42, Section 2]. One may therefore view the Bregman divergence as penalizing the deviation from the log Kaporin condition number via the term $\operatorname{tr}\left(P^{-1}A\right) - n$.

We can also quantify how close the divergence $\mathcal{D}_{\mathrm{LD}}(A,P)$ is to $\ln(K(P^{-1}A))$, as the following corollary shows.

COROLLARY 2. *Let $A \in \mathbb{H}^n_{++}$, $P = QQ^* \in \mathbb{H}^n_{++}$ so*

$$A = Q(I + \tilde{E})Q^*, \tag{3.4a}$$

$$\tilde{E} = Q^{-1}(A - P)Q^{-*}. \tag{3.4b}$$

Then,

$$\ln(K(P^{-1}A)) = \mathcal{D}_{\mathrm{LD}}(A,P) + O\left(\operatorname{tr}\left(\tilde{E}\right)^2\right).$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

Proof.

$$\begin{aligned}\frac{1}{n}\mathcal{D}_{\mathrm{LD}}(A,P) &= \frac{1}{n}\left(\sum_{i=1}^{n}\lambda_i(\tilde{E}) - \ln(1+\lambda_i(\tilde{E}))\right)\\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i(\tilde{E})\right) - \frac{1}{n}\sum_{i=1}^{n}\ln(1+\lambda_i(\tilde{E}))\\ &\approx \ln\left(1+\frac{1}{n}\sum_{i=1}^{n}\lambda_i(\tilde{E})\right) - \frac{1}{n}\ln\det(I+\tilde{E})\\ &= \ln\left(\frac{1}{n}\sum_{i=1}^{n}1+\lambda_i(\tilde{E})\right) - \frac{1}{n}\ln\det(I+\tilde{E})\\ &= \ln\left(\frac{1}{n}\operatorname{tr}\left(I+\tilde{E}\right)\right) - \frac{1}{n}\ln\det(I+\tilde{E}).\end{aligned}$$

Therefore,

$$\ln(K(M)) = \mathcal{D}_{\mathrm{LD}}(A,P) + O(n^{-1}\operatorname{tr}\left(\tilde{E}\right)^2).$$

□

The approximation used in the proof is of course poor when $\operatorname{tr}\left(\tilde{E}\right) \gg 0$.

Next, we express certain convergence bounds in terms of the Bregman log determinant divergence using Theorem 1.

3.1. The log determinant divergence and convergence of PCG. In light of Theorem 1, we can state some of the convergence results in Subsection 2.1 in terms of the divergence.

Corollary 3. *Let $A \in \mathbb{H}^n_{++}$. For any $k \in \mathbb{N}$, $1 \le k \le n-1$ and $\beta > 1$, there exists a preconditioner $P \in \mathbb{H}^n_{++}$ such that*

$$\begin{aligned}\operatorname{tr}\left(P^{-\frac{1}{2}}AP^{-\frac{1}{2}}\right) &= n,\\ \beta &= e^{\frac{\mathcal{D}_{\mathrm{LD}}(A,P)}{n}},\end{aligned} \tag{3.5}$$

and an initial residual r_0 such that the k^{th} residual, r_k, satisfies

$$\begin{aligned}\|r_k\|_{P^{-1}} &= (\beta^{n/k}-1)^{k/2}\|r_0\|_{P^{-1}}\\ &= (e^{\frac{\mathcal{D}_{\mathrm{LD}}(A,P)}{k}}-1)^{k/2}\|r_0\|_{P^{-1}}.\end{aligned}$$

Proof. Using (3.5), we have

$$K(P^{-1}A)^{1/k} = \left(e^{\ln K(P^{-1}A))}\right)^{1/k} = \left(e^{\mathcal{D}_{\mathrm{LD}}(A,P)}\right)^{1/k} = e^{\frac{\mathcal{D}_{\mathrm{LD}}(A,P)}{k}}.$$

The result follows from (2.4b) and [30, Theorem 3.2]. □

By (2.8), Corollary 3 is nontrivial when

$$\mathcal{D}_{\mathrm{LD}}(A,P) < \ln(2)n.$$

Corollary 3 leads to the following two results, where we note that the trace condition (3.5) is not necessary.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

COROLLARY 4. *Let $A \in \mathbb{H}^n_{++}$ and $P \in \mathbb{H}^n_{++}$ be a preconditioner. The k^{th} residual satisfies*

$$\|r_k\|_{P^{-1}} \leq \left(e^{\frac{\mathcal{D}_{\mathrm{LD}}(A,P)}{k}} - 1\right)^{k/2} \|r_0\|_{P^{-1}}. \tag{3.6}$$

Proof. The result follows from Theorem 1 and [30, Theorem 3.1]. ∎

COROLLARY 5. *Let $A \in \mathbb{H}^n_{++}$ and $P \in \mathbb{H}^n_{++}$ be a preconditioner. For k even such that*

$$3\mathcal{D}_{\mathrm{LD}}(A,P) \leq k < n,$$

we have

$$\|x_k - x\|_A \leq \left(\frac{3\mathcal{D}_{\mathrm{LD}}(A,P)}{k}\right)^{k/2} \|x_0 - x\|_A. \tag{3.7}$$

Proof. Equation (3.7) follows from Theorem 1 and [4, Theorem 4.3]. ∎

We conclude this section with some bounds on the number of iterations required to achieve an ϵ-reduction in the initial residual of PCG.

THEOREM 6. *Let $A \in \mathbb{H}^n_{++}$ and $P \in \mathbb{H}^n_{++}$ be a preconditioner. When $\mathrm{tr}\left(P^{-\frac{1}{2}}AP^{-\frac{1}{2}}\right) = n$, the number of iterations $i_K(\epsilon)$ needed for an ϵ-reduction in the initial $\|r_0\|_{P^{-1}}$ satisfies*

$$i_K(\epsilon) \leq \lceil \ln(2)^{-1} \left(\ln(\epsilon^{-1}) + D(A,P)\right)\rceil.$$

Proof. Using (2.10) for $\sigma = 2$, we have

$$\begin{aligned}
i_K(\epsilon) &\leq \lceil n \log_2(B(P^{-\frac{1}{2}}AP^{-\frac{1}{2}})) + \log_2(\epsilon^{-1})\rceil \\
&= \lceil \log_2(K(P^{-1}A)) + \log_2(\epsilon^{-1})\rceil \\
&= \lceil \log_2\left(\frac{1}{\det(P^{-1}A)}\right) + \log_2(\epsilon^{-1})\rceil \\
&= \lceil \frac{\ln\left(\frac{1}{\det(P^{-1}A)}\right)}{\ln(2)} + \frac{\ln(\epsilon^{-1})}{\ln(2)}\rceil \\
&= \lceil \ln(2)^{-1}\left(\ln(\epsilon^{-1}) - \ln(\det(P^{-1}A))\right)\rceil \\
&= \lceil \ln(2)^{-1}\left(\ln(\epsilon^{-1}) + D(A,P)\right)\rceil.
\end{aligned}$$ ∎

4. Preconditioners based on low-rank approximations. Kaporin also constructs preconditioners using low-rank matrices [26, 27, 29]. In [26], a sparse lower triangular matrix variable G is introduced, and the expression

$$B(GAG^\top)$$

is minimized subject to some structural constraints on G. This is referred to as the first stage. Second, they construct a matrix

$$F = I + CSC^\top \tag{4.1}$$

such that $B(FGAG^\top)$ is minimized. Here, some structure is imposed on $C \in \mathbb{R}^{n\times r}$, and the matrix $S \in \mathbb{R}^{r\times r}$ that minimizes $B(FGAG^\top)$ has the following form:

$$S = \pi_0 C^\top GAG^\top C - (C^\top C)^{-1} \in \mathbb{R}^{r\times r},$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

where

$$\pi_0 = \frac{n - \operatorname{tr}\left((C^\top C)^{-1} C^\top G A G^\top C\right)}{n-r}.$$

[6] adopts a similar two-stage approach. The preconditioner described above can be written as

$$G^{-\top} F^{-1} G^{-1} = G^{-\top}(I + CSC^\top)^{-1} G^{-1}. \tag{4.2}$$

If we let $G^{-\top}G^{-1} \approx A$ be an incomplete Cholesky factorization of A and recall that for some $n \times n$ matrix W with $\operatorname{rank} W = r$,

$$F^{-1} = (I + CSC^\top)^{-1} = I + W,$$

then (4.2) can be written as

$$Q(I+W)Q^\top, \qquad \operatorname{rank} W = r.$$

This is similar to the preconditioners sought in [8, 9]. In (4.1) above, the low-rank term has a specific structure given by the choice of C, which is more restrictive than the general low-rank constraint considered in the previous references. In this report, we focus preconditioners for which $K(M)$ or $\mathcal{D}_{\mathrm{LD}}(A,P)$ is minimized; however, other approaches include κ_2-based minimization, see, e.g. [14, 39] and the references therein.

4.1. Bregman log determinant divergence-based preconditioners. In this section, we describe the preconditioners introduced in [8, 9], which combine approximate factorizations with low-rank approximations sought as minimizers of a Bregman divergence. We introduce the approach taken in these references to generalize the resulting preconditioners in Subsection 4.2 and establish the connection with Kaporin's condition number.

In [9, 8], the authors assume that an approximate factorization QQ^* of the target matrix $A \in \mathbb{H}^n_{++}$ is assumed available (e.g., incomplete Cholesky). Writing A in terms of Q, we obtain

$$A = Q(I + \tilde{E})Q^*, \tag{4.3}$$

where $\tilde{E} = Q^{-1}AQ^{-*} - I$. A low-rank term was introduced to produce the preconditioner

$$P = Q(I + VDV^*)Q^*, \tag{4.4}$$

where $V \in \mathbb{C}^{n\times r}$ and $D \in \mathbb{R}^{r \times r}$ are chosen such that

$$VDV^* \approx \tilde{E}.$$

The approximation above was sought in the sense of the Bregman log determinant divergence introduced in (2.5). This divergence is invariant to congruence transformations. Indeed, using (4.3) and (4.4), we obtain

$$\mathcal{D}_{\mathrm{LD}}(A,P) = \mathcal{D}_{\mathrm{LD}}(I + \tilde{E}, I + VDV^*).$$

Then, the matrices $V \in \mathbb{C}^{n\times r}$ and $D \in \mathbb{R}^{r \times r}$ are found as minimizers of

$$\min_{V \in \mathbb{C}^{n\times r},\, D \in \mathbb{R}^{r\times r}} \quad \mathcal{D}_{\mathrm{LD}}(I + \tilde{E}, I + VDV^*), \tag{4.5a}$$

$$\text{s.t.} \quad I + VDV^* \in \mathbb{H}^n_{++}, \tag{4.5b}$$

$$\operatorname{rank} D \leq r. \tag{4.5c}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

In [9], $\tilde{E}$ was assumed to be positive semidefinite, in which cases V and D were found as the constituents of a truncated singular value decomposition (TSVD) of $\tilde{E}$. [8] generalized the approach and allowed $\tilde{E}$ to be indefinite, and it was shown that a low-rank matrix VDV^* that minimizes (4.5a) can differ from a TSVD of $\tilde{E}$. To see this, let

$$\tilde{E} = U\Theta U^*, \tag{4.6}$$

be an eigendecomposition of $\tilde{E}$, and let $\Pi \in \mathbb{R}^{n\times n}$ be a diagonal matrix with positive entries. The divergence $\mathcal{D}_{\mathrm{LD}}(I + U\Theta U^*, I + U\Pi U^*)$ takes the form of the following simple expression (cf. (2.5)):

$$\mathcal{D}_{\mathrm{LD}}(I + U\Theta U^*, I + U\Pi U^*) = \sum_{i=1}^{n} \left(\frac{1+\theta_i}{1+\pi_i} - \log\left(\frac{1+\theta_i}{1+\pi_i}\right) - 1 \right), \tag{4.7}$$

since eigenspaces of the inputs are aligned. If $\pi_i = \theta_i$ for any $i = 1, \ldots, n$, the i^{th} summand of (4.7) vanishes. If we consider (4.7) as a function of the diagonal matrix Π with the constraint $\operatorname{rank}\Pi \le r$, a minimizer will satisfy $\pi_i = \theta_i$ for some indices i, and $\pi_i = 0$ otherwise (by the rank constraint). The $n - r$ summands of (4.7) that do not vanish are therefore of the form

$$\theta_i - \log(1 + \theta_i).$$

By seeking a low-rank approximation of $\tilde{E}$ in the sense of (4.5), it is therefore of interest to identify the eigenvalues θ_i that take on the largest value under the image of the map

$$\gamma(\lambda) = \lambda - \log(1 + \lambda). \tag{4.8}$$

γ is different from the map $\lambda \mapsto |\lambda|$, which can lead to VDV^* being different from a TSVD of $\tilde{E}$. The map γ induces an order on the eigenvalues of a matrix A, which is captured by the following definition.

Definition 7 (Bregman log determinant truncation [8]). *Let $I + X \in \mathbb{H}^n_{++}$ and let*

$$X = W\Lambda W^*$$

be an eigendecomposition where the diagonal elements of Λ are sorted in algebraically nonincreasing order and W is orthonormal. We introduce a permutation matrix $\mathcal{P}$ and define Z and $\mathfrak{B}$ via

$$X = W\mathcal{P}^\top \mathfrak{B}\mathcal{P}W^* = Z\mathfrak{B}Z^*,$$

where

$$\mathfrak{B} = \begin{bmatrix} \mathfrak{b}_1(X) & & \\ & \ddots & \\ & & \mathfrak{b}_n(X) \end{bmatrix},$$

is the diagonal matrix containing the eigenvalues X given by the order $\succeq$:

$$\mathfrak{b}_i(A) \succeq \mathfrak{b}_j(A) \qquad \text{if} \qquad \gamma(\mathfrak{b}_i(A)) \ge \gamma(\mathfrak{b}_j(A)).$$

Here, γ is the function given in (4.8). We define a BLD truncation of X to order r by

$$\langle\!\langle X \rangle\!\rangle_r = \begin{bmatrix} z_1 | \cdots | z_r \end{bmatrix} \begin{bmatrix} \mathfrak{b}_1(X) & & \\ & \ddots & \\ & & \mathfrak{b}_r(X) \end{bmatrix} \begin{bmatrix} z_1 | \cdots | z_r \end{bmatrix}^*. \tag{4.9}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

In other words, the BLD truncation is found by selecting the rows and columns of an eigendecomposition of a matrix corresponding to the r largest values of the eigenvalues under the map γ.

Definition 7 leads to the following preconditioner.

DEFINITION 8 (Bregman log determinant precondition [8]). *Let $A \in \mathbb{H}^n_{++}$, $QQ^* \in \mathbb{H}^n_{++}$, and $1 \leq r < n$, and assume $\tilde{E}$ has rank greater than r. We call*

$$P = Q(I + \langle\!\langle \tilde{E} \rangle\!\rangle_r)Q^*. \tag{4.10}$$

a Bregman log determinant preconditioner*.*

4.2. Minimization of the Bregman divergence and Kaporin's condition number. In this section, we modify (4.10) by permitting an arbitrary constant scaling of the preconditioner:

$$P_\alpha = Q(\alpha(I - VV^*) + V(I_r + D)V^*)Q^*. \tag{4.11}$$

When $\alpha = 1$, we recover (4.10). We proceed to show that a specific choice of α leads to (4.11) minimizing Kaporin's condition number, as well as the Bregman divergence. Consider the following expression

$$\mathcal{D}_{\mathrm{LD}}(A, P_\alpha) = \mathcal{D}_{\mathrm{LD}}(I + \tilde{E}, \alpha(I - VV^*) + V(I_r + D)V^*). \tag{4.12}$$

Theorem 9 below describes a minimizer of (4.12) as a function of α, letting V and D be fixed:

THEOREM 9. *Suppose $V \in \mathbb{C}^{n\times r}$, $D \in \mathbb{R}^{r\times r}$ have been chosen according to a Bregman truncation, i.e.,*

$$\langle\!\langle \tilde{E} \rangle\!\rangle_r = VDV^*,$$

and set

$$P = Q(I + \langle\!\langle \tilde{E} \rangle\!\rangle_r)Q^*.$$

Then,

$$\alpha^\star = \frac{\operatorname{tr}\left((I + \tilde{E})(I - VV^*)\right)}{n - r} = \frac{\operatorname{tr}\left(P^{-1}A\right) - r}{n - r} = \frac{\sum_{i=r+1}^n 1 + \mathfrak{b}_i(\tilde{E})}{n - r}, \tag{4.13}$$

is the unique minimizer of (4.12) as a function of α. Furthermore,

$$\mathcal{D}_{\mathrm{LD}}(A, P_\alpha) = \ln K(P_\alpha^{-1}A) = -\ln\det(P_\alpha^{-1}A) = -\ln\det(P^{-1}A) + (n - r)\ln(\alpha). \tag{4.14}$$

Proof. Let $\Delta_\alpha = \alpha(I - VV^*) + V(I_r + D)V^*$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathcal{D}_{\mathrm{LD}}(A, P_\alpha) = -\operatorname{tr}\left(\Delta_\alpha^{-1}(I + \tilde{E})\Delta_\alpha^{-1}\frac{\mathrm{d}\Delta_\alpha}{\mathrm{d}\alpha}\right) + \frac{\mathrm{d}}{\mathrm{d}\alpha}\ln\det\Delta_\alpha.$$

Since

$$\frac{\mathrm{d}\Delta_\alpha}{\mathrm{d}\alpha} = I - VV^*,$$

$$-\operatorname{tr}\left(\Delta_\alpha^{-1}(I + \tilde{E})\Delta_\alpha^{-1}\frac{\mathrm{d}\Delta_\alpha}{\mathrm{d}\alpha}\right) = -\alpha^{-2}\operatorname{tr}\left((I + \tilde{E})(I - VV^*)\right).$$

Combining this with

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\ln\det\Delta_\alpha = \alpha^{-1}\operatorname{tr}\left(I - VV^*\right),$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

yields the equation

$$-\alpha^{-2}\operatorname{tr}\left((I+\tilde{E})(I-VV^*)\right)+\alpha^{-1}\operatorname{tr}(I-VV^*)=0.$$

Equation (4.13) follows as a consequence. Uniqueness follows since $\mathcal{D}_{\mathrm{LD}}(A,P_\alpha)$ is convex in $\beta=\alpha^{-1}$.

By direct computation,

$$(I+\tilde{E})(I-VV^*)=\operatorname{tr}\left(P^{-1}A\right)-r.$$

Finally, (4.14) follows since

$$\begin{aligned}\mathcal{D}_{\mathrm{LD}}(A,P_\alpha)&=\frac{\operatorname{tr}\left(P^{-1}A\right)-r}{\frac{\operatorname{tr}(P^{-1}A)-r}{n-r}}-\ln\det(P_\alpha^{-1}A)-(n-r)\\&=-\ln\det(P_\alpha^{-1}A).\end{aligned}$$

∎

α in (4.11) is very similar to the σ variable in [6, Section 3.3], which was used to move the set of smallest eigenvalues of the preconditioned matrix closer to the larger eigenvalues.

While the choice (4.13) minimizes Kaporin's condition number and Bregman divergence as a function of α, the following theorem reveals that many different values of α minimize the condition number, i.e., the map

$$\alpha\mapsto\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}}).$$

THEOREM 10. *Suppose V and D have been chosen according to a Bregman truncation (cf. Definition 8). Then any*

$$\alpha\in[\mathfrak{b}_{r+1}(A),\mathfrak{b}_n(A)],\tag{4.15}$$

will be a minimizer of

$$\alpha\mapsto\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}}).$$

Proof. We have

$$\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}})=\frac{\max\left(1,\frac{\lambda_1((I+\tilde{E})(I-VV^*))}{\alpha}\right)}{\min\left(1,\frac{\lambda_n((I+\tilde{E})(I-VV^*))}{\alpha}\right)}\geq\frac{\lambda_1((I+\tilde{E})(I-VV^*))}{\lambda_n((I+\tilde{E})(I-VV^*))}=:\kappa_2^\star(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}}).\tag{4.16}$$

If $\alpha>\lambda_1((I+\tilde{E})(I-VV^*))$,

$$\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}})=\frac{\alpha}{\lambda_n((I+\tilde{E})(I-VV^*))}>\kappa_2^\star(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}}).$$

Conversely, if $\alpha<\lambda_n((I+\tilde{E})(I-VV^*))$,

$$\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}})=\frac{\lambda_1((I+\tilde{E})(I-VV^*))}{\alpha}>\kappa_2^\star(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}}).$$

The result follows by Definition 7, namely

$$[\mathfrak{b}_{r+1}(A),\mathfrak{b}_n(A)]=[\lambda_1((I+\tilde{E})(I-VV^*)),\lambda_n((I+\tilde{E})(I-VV^*))].$$

∎

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

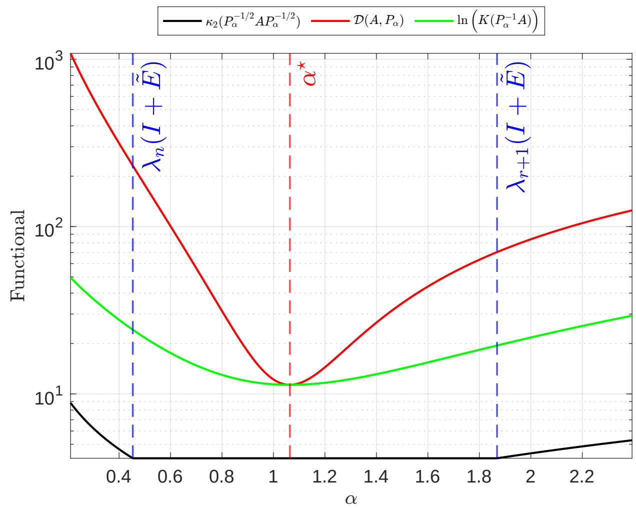

FIG. 1. *Numerical illustration of how α influences the values $\kappa_2(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}})$, $\mathcal{D}_{\mathrm{LD}}(A,P_\alpha)$, $\ln\Big(K(P_\alpha^{-1}A)\Big)$.*

We conclude this section with a numerical illustration of the effect of α on the condition numbers seen above. $A \in \mathbb{H}^n_{++}$ is the matrix `494_bus` from the SuiteSparse Matrix Collection [31] of order $n = 494$ and Q given by a zero fill incomplete Cholesky decomposition of A. P_α is defined in (4.11), with $r = 49$. In Figure 1, $\kappa_2^\star(P_\alpha^{-\frac{1}{2}}AP_\alpha^{-\frac{1}{2}})$, $\mathcal{D}_{\mathrm{LD}}(A,P_\alpha)$, $\ln(K(P_\alpha^{-1}A))$ are plotted as a function of α. The condition number is flat on the interval in (4.15), in agreement with Theorem 10. $\mathcal{D}_{\mathrm{LD}}(A,P_\alpha)$ and $\ln(K(P_\alpha^{-1}A))$ have a minimum at $\alpha^\star$, in agreement with Theorem 9. Indeed, as stated in (4.13), the value of α that minimizes Kaporin's condition number is the *average* of the remaining eigenvalues not selected by the Bregman truncation.

5. A general result on trace scaling. We conclude with a result similar to, but more general than Theorem 9. We recall that PCG converges in a single iteration when the preconditioned matrix is some scalar multiple of the identity, say,

$$P^{-1}A = cI,$$

for $c > 0$. This motivates the study of preconditioners scaled by some constant. In particular, we can choose the constant c such that

$$\operatorname{tr}\left((cP)^{-1}A\right) = c^{-1}\operatorname{tr}\left(P^{-1}A\right) = n \Leftrightarrow c = \frac{\operatorname{tr}\left(P^{-1}A\right)}{n},$$

from which we deduce

$$\mathcal{D}_{\mathrm{LD}}(A,cP) = \ln K((cP)^{-1}A), \tag{5.1}$$

as shown in Theorem 1.

This leads to the following observation, namely that minimization of $\mathcal{D}_{\mathrm{LD}}(A,P)$ is equivalent to minimization of $K(P^{-1}A)$ when considering preconditioners for which the divergence cannot be reduced by scaling P by a positive constant.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

PROPOSITION 11. *Let $A \in \mathbb{H}^n_{++}$ and $C \subseteq \mathbb{H}^n_{++}$ and consider the two following optimization problems:*

$$S_1 = \arg\min_{P\in C} \quad \mathcal{D}_{\mathrm{LD}}(A, P). \tag{5.2}$$

$$S_2 = \arg\min_{P\in C} \quad \ln K(P^{-1}A). \tag{5.3}$$

If C is a cone, then the following statements hold.

(i) $\min_{P\in C} \mathcal{D}_{\mathrm{LD}}(A, P) = \min_{P\in C} \ln K(P^{-1}A)$,
(ii) $S_1 \subset S_2$,
(iii) $\operatorname{cone}(S_1) = S_2$, *where* $\operatorname{cone}(S_1)$ *denotes the conic hull of* S_1.

Proof. Suppose $P^\star$ is a minimizer of (5.3). Since Kaporin's condition number is invariant under any positive scaling of the preconditioner, we can choose $P^\star$ such that $\operatorname{tr}\left((P^\star)^{-1}A\right) = n$. By Theorem 1, $\mathcal{D}_{\mathrm{LD}}(A, P^\star) = K((P^\star)^{-1}A)$, which proves (i). To see (ii), note that any minimizer P of (5.2) must satisfy $\operatorname{tr}\left(P^{-1}A\right) = n$. Indeed, suppose P is a minimizer of (5.2) but $\operatorname{tr}\left(P^{-1}A\right) \neq n$. Define

$$P_c = cP, \quad P \in S_1,$$

where $c = \frac{\operatorname{tr}\left(P^{-1}A\right)}{n}$ such that $\operatorname{tr}\left(P_c^{-1}A\right) = n$. Then,

$$\mathcal{D}_{\mathrm{LD}}(A, P) > \mathcal{D}_{\mathrm{LD}}(A, P_c),$$

since

$$\mathcal{D}_{\mathrm{LD}}(A, P) - \mathcal{D}_{\mathrm{LD}}(A, P_c) = n(c - 1 - \ln(c)) > 0$$

using $x \geq 1 + \ln x$, $x > 0$. As a result, $c = 1$ if $P \in S_1$, since C is a cone. Further, $\mathcal{D}_{\mathrm{LD}}(A, P) = \ln K(P^{-1}A)$. S_2 contains all positive scalings of $P^\star$, which includes $P \in S_1$. Finally, (iii) holds by the scaling invariance of Kaporin's condition number. ∎

We also highlight that $K(P_c^{-1}A) = K(P_{\alpha^\star}^{-1}A)$ for any $c > 0$. In other words, $\ln K(P_c^{-1}A)$ is a flat line in Figure 1, whereas $\ln K(P_\alpha^{-1}A)$ has curvature away from α (recall P_α is given in (4.11)). This is of course obvious from (2.4b), but there are preconditioners prevalent in the literature, such as the randomized Nyström preconditioner [12], where a scaling similar to that controlled by α is used to control the condition number of the preconditioned matrix. Partial scaling of the preconditioner influences the iterates of PCG (as opposed to scaling the entire preconditioner, e.g., $P \mapsto cP$, $c > 0$), so it would be interesting and useful to analyze the consequence of a given choice beyond controlling the condition number.

6. Summary and future directions. We explored a lesser-known condition number, the Kaporin condition number, and established its connection with the Bregman log determinant divergence. The link between Kaporin's condition number and this divergence lies in the trace scaling of the preconditioned matrix M ensuring $\frac{1}{n}\operatorname{tr}(M) = 1$. This led to the error bounds presented in Section 3. In Section 4, we explore this connection in the context of the truncation described in [8]. Subsection 4.2 described this scaling via α (by scaling the orthogonal complement of the truncation) or c (the entire preconditioner). We also highlight a relationship with the condition number. Section 5 contained a result concerning the equivalence of using the divergence and logarithm of the Kaporin condition number as an objective function.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

We comment on some possible future directions. An obvious extension would be to compute approximate factor Q of a matrix A using an FSAI approach [23, 40] and then compensate using a low-rank term for the error $E = A - QQ^\top$ using the low-rank truncation provided in [8]. This could lead to a more principled approach than in [8], where Q is obtained as an incomplete Cholesky factor of A.

Randomized linear algebra is a family of techniques in linear algebra that leverages randomness to approximate large-scale problems that may otherwise be computationally infeasible [19, 36]. Such methods could prove to be useful in estimating a randomized version of the log determinant divergence and therefore providing operational approximate convergence bounds.

The recent paper by Carson et al. [10] contains a wealth of insights on the behavior of CG. The authors point out that not only is clustering of the eigenvalues of a matrix important for CG, but so is their position. As described in [10, Section 2.2], the position of the clusters can have an enormous effect on the residual error in energy norm. The scaling of the preconditioned matrix via α given in Subsection 4.2 can be thought of as controlling the position of the "unpreconditioned" part of the matrix A, and it would be interesting to explore theoretically and quantitatively the impact of α.

The Bregman log determinant divergence has been studied in detail in the field of information geometry [1]. When N is a smooth submanifold of a manifold M, the point dual geodesic projection $P_N^\star$ of P onto N is given by:

$$P_N^\star = \arg\min_{R\in N} \mathcal{D}(P, R),$$

In certain cases, uniqueness of such projection holds, see [1] and [38, Section 3.7] for more details. Candidate preconditioners described by some adequate submanifold $N \subset \mathbb{H}^n_{++}$ could therefore be interesting, and insights or algorithms from information geometry may be relevant to preconditioning or other problems in numerical linear algebra.

Acknowledgments. This work was supported by the Novo Nordisk Foundation under grant number NNF20OC0061894.

Appendix A. Proof of Corollary 2.

Proof.

$$\begin{aligned}
\frac{1}{n}\mathcal{D}_{\mathrm{LD}}(A,P) &= \frac{1}{n}\left(\sum_{i=1}^{n} \lambda_i(\tilde{E}) - \ln(1+\lambda_i(\tilde{E}))\right)\\
&= \left(\frac{1}{n}\sum_{i=1}^{n} \lambda_i(\tilde{E})\right) - \frac{1}{n}\sum_{i=1}^{n}\ln(1+\lambda_i(\tilde{E}))\\
&\approx \ln\left(1+\frac{1}{n}\sum_{i=1}^{n}\lambda_i(\tilde{E})\right) - \frac{1}{n}\ln\det(I+\tilde{E})\\
&= \ln\left(\frac{1}{n}\sum_{i=1}^{n} 1+\lambda_i(\tilde{E})\right) - \frac{1}{n}\ln\det(I+\tilde{E})\\
&= \ln\left(\frac{1}{n}\operatorname{tr}\left(I+\tilde{E}\right)\right) - \frac{1}{n}\ln\det(I+\tilde{E}).
\end{aligned}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

Therefore,

$$\begin{aligned}\mathcal{D}_{\mathrm{LD}}(A,P) &\approx n\ln\left(\frac{1}{n}\operatorname{tr}\left(I+\tilde{E}\right)\right)-\ln\det(I+\tilde{E})\\ &= n\ln\left(\frac{1}{n}\operatorname{tr}\left(I+\tilde{E}\right)\right)-\ln\det(I+\tilde{E})\\ &= \ln(K(M)^n),\end{aligned}$$

and

$$\ln(K(M)^n)=\mathcal{D}_{\mathrm{LD}}(A,P)+O(n^2\operatorname{tr}(\tilde{E})).$$

□

REFERENCES

[1] S.-i. Amari. *Information Geometry and Its Applications*, vol. 194. Springer, Tokyo, 2016.

[2] O. Axelsson. *Condition numbers for the study of the rate of convergence of the conjugate gradient method.* In: Proceedings of the 2nd IMACS International Symposium on Iterative Methods in Linear Algebra, Blagoevgrad, Bulgaria, 17–20 June 1995, 3–33.

[3] O. Axelsson. *Iterative Solution Methods.* Cambridge University Press, New York, 1996.

[4] O. Axelsson and I. Kaporin. On the sublinear and superlinear rate of convergence of conjugate gradient methods. Numer. Algorithms 25:1–22, 2000.

[5] O. Axelsson and I. Kaporin. Error norm estimation and stopping criteria in preconditioned conjugate gradient iterations. *Numer. Linear Algebra Appl.*, 8:265–286, 2001.

[6] O. Axelsson and I. Kaporin. Optimizing two-level preconditionings for the conjugate gradient method. In: *Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3.* Springer, 2001, 3–21.

[7] B. Beckermann and A.B. Kuijlaars. Superlinear convergence of conjugate gradients. *SIAM J. Numer. Anal.*, 39:300–329, 2001.

[8] A.A. Bock and M.S. Andersen. A new matrix truncation method for improving approximate factorisation preconditioners. *arXiv preprint arXiv:2312.06417*, 2023.

[9] A.A. Bock and M.S. Andersen. Preconditioner design via Bregman divergences. *SIAM J. Matrix Anal. Appl.*, 45:1148–1182, 2024.

[10] E. Carson, J. Liesen, and Z. Strakoš. Towards understanding CG and GMRES through examples. *Linear Algebra Appl.*, 692:241–291, 2024.

[11] J.E. Dennis, Jr. and H. Wolkowicz. Sizing and least-change secant methods. *SIAM J. Numer. Anal.*, 30:1291–1314, 1993.

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

[13] V. Gadjokov. Quasi-optimum scaling of symmetric positive-definite matrices. *Comput. Phys. Commun.*, 36:19–24, 1985.

[14] W. Gao, Z. Qu, M. Udell, and Y. Ye. Scalable approximate optimal diagonal preconditioning, *arXiv preprint arXiv:2312.15594*, 2023.

[15] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM, Philadelphia, 1997.

[16] K. Gustafson. A computational trigonometry and related contributions by Russians Kantorovich, Krein, Kaporin. *Comput. Technol.*, 4(3):73–83, 1999d (Novosibirsk, Russia).

[17] K. Gustafson. Operator trigonometry of iterative methods. *Numer. Linear Algebra Appl.*, 4:333–347, 1997.

[18] K. Gustafson. Operator trigonometry of preconditioning, domain decomposition, sparse approximate inverses, successive overrelaxation, minimum residual schemes. *Numer. Linear Algebra Appl.*, 10:291–315, 2003.

[19] N. Halko, P.-G. Martinsson, and J.A. Tropp. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53:217–288, 2011.

[20] R.M. Hayes. Iterative methods of solving linear problems on Hilbert space. In: *Bulletin of the American Mathematical Society*, vol. 58. American Mathematical Society, Providence, RI, 1952, 653–653.

[21] M.R. Hestenes, E. Stiefel, et al. *Methods of Conjugate Gradients for Solving Linear Systems*, vol. 49. NBS Washington, DC, 1952.

[22] T. Huckle. Factorized sparse approximate inverses for preconditioning. *J. Supercomput.*, 25:109–117, 2003.

[23] C. Janna, M. Ferronato, F. Sartoretto, and G. Gambolati. Fsaipack: a software package for high-performance factored sparse approximate inverse preconditioning. *ACM Trans. Math. Softw.*, 41:1–26, 2015.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 42, pp. 430-445, May 2026.

[24] W.L. Jung, D. Torregrosa-Belén, and H. Wolkowicz. The ω-condition number: applications to preconditioning and low rank generalized Jacobian updating. *Comput. Optim. Appl.*, 91:1–48, 2025.

[25] I. Kaporin. An alternative approach to estimating the convergence rate of the CG method. In: Yu. A. Kuznetsov (editor), *Numerical Methods and Software*. Department of Numerical Mathematics, USSR Academy of Sciences, Moscow, 1990, 55–72.

[26] I. Kaporin. Optimization of conjugate gradient algorithms, *Comput. Math. Model.*, 5:139–147, 1994.

[27] I.E. Kaporin. Preconditioning of the method of conjugate gradients for solving discrete analogues of differential problems. *Differ. Uravn.*, 26:1225–1236, 1990.

[28] I.E. Kaporin. Explicitly preconditioned conjugate gradient method for the solution of unsymmetric linear systems. *Int. J. Comput. Math.*, 44:169–187, 1992.

[29] I.E. Kaporin. Two-level explicit preconditionings for the conjugate gradient method. *Diff. Uravn.*, 28:329–339, 1992.

[30] I.E. Kaporin. New convergence results and preconditioning strategies for the conjugate gradient method. *Numer. Linear Algebra Appl.*, 1:179–210, 1994.

[31] S.P. Kolodziej, M. Aznaveh, M. Bullock, J. David, T.A. Davis, M. Henderson, Y. Hu, and R. Sandstrom. The suitesparse matrix collection website interface. *J. Open Sour. Softw.*, 4:1244, 2019.

[32] L.Y. Kolotilina, A.A. Nikishin, and A.Y. Yeremin. Factorized sparse approximate inverse preconditionings. IV: simple approaches to rising efficiency. *Numer. Linear Algebra Appl.*, 6:515–531, 1999.

[33] L.Y. Kolotilina and A.Y. Yeremin. Factorized sparse approximate inverse preconditionings I. Theory. *SIAM J. Matrix Anal. Appl.*, 14:45–58, 1993.

[34] L.Y. Kolotilina and A.Y. Yeremin. Factorized sparse approximate inverse preconditioning II: solution of 3D FE systems on massively parallel computers. *Int. J. High Speed Comput.*, 7:191–215, 1995.

[35] B. Kulis, M.A. Sustik, and I.S. Dhillon, Low-rank kernel learning with Bregman matrix divergences. *J. Mach. Learn. Res.*, 10:341–376, 2009.

[36] P.-G. Martinsson and J.A. Tropp. Randomized numerical linear algebra: foundations and algorithms. *Acta Numer.*, 29:403–572, 2020.

[37] O. Nevanlinna. *Convergence of Iterations for Linear Equations*. Birkhäuser, Basel, 2012.

[38] F. Nielsen. An elementary introduction to information geometry. *Entropy*, 22:1100, 2020.

[39] Z. Qu, W. Gao, O. Hinder, Y. Ye, and Z. Zhou. Optimal diagonal preconditioning. *Oper. Res.*, 73(3):1479–1495, 2024.

[40] F. Schäfer, M. Katzfuss, and H. Owhadi. Sparse Cholesky factorization by Kullback–Leibler minimization. *SIAM J. Sci. Comput.*, 43:A2019–A2046, 2021.

[41] A. Van der Sluis and H.A. van der Vorst. The rate of convergence of conjugate gradients. *Numer. Math.*, 48:543–560, 1986.

[42] A.Y. Yeremin, L.Y. Kolotilina, and A. Nikishin. Factorized sparse approximate inverse preconditionings. III. Iterative construction of preconditioners. *J. Math. Sci.*, 101:3237–3254, 2000.