Electronic Transactions on Numerical Analysis.
Volume 43, pp. 70-89, 2014.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ESTIMATES FOR THE BILINEAR FORM $\mathbf{x}^T\mathbf{A}^{-1}\mathbf{y}$ WITH APPLICATIONS TO LINEAR ALGEBRA PROBLEMS*

PARASKEVI FIKA†, MARILENA MITROULI†, AND PARASKEVI ROUPA†

Abstract. Let $A \in \mathbb{R}^{p\times p}$ be a nonsingular matrix and x, y vectors in $\mathbb{R}^p$. The task of this paper is to develop efficient estimation methods for the bilinear form $x^T A^{-1} y$ based on the extrapolation of moments of the matrix A at the point -1. The extrapolation method and estimates for the trace of A^{-1} presented in Brezinski et al. [Numer. Linear Algebra Appl., 19 (2012), pp. 937–953] are extended, and families of estimates efficiently approximating the bilinear form requiring only few matrix vector products are derived. Numerical approximations of the entries and the trace of the inverse of any real nonsingular matrix are presented and several numerical results, discussions, and comparisons are given.

Key words. extrapolation, matrix moments, matrix inverse, trace

AMS subject classifications. 65F15, 65F30, 65B05, 65C05, 65J10, 15A18, 15A45

1. Introduction and motivation for the problem. Let A be a real nonsingular matrix of order p, and let x, y be real vectors of length p. The subject of this work is to estimate the bilinear form $x^T A^{-1} y$. The evaluation of this form arises in several applications including network analysis, signal processing, nuclear physics, quantum mechanics, and computational fluid dynamics [15, 19, 20]. The motivation for this problem stems from the fact that for a specific selection of vectors x and y, by estimating the bilinear form $x^T A^{-1} y$, we can approximate several useful quantities arising frequently in many linear algebra problems. Let us name some of them.

Elements of the matrix A^{-1}. By choosing as vectors x and y appropriate columns of the identity matrix of order p, notated as I_p, estimates for the diagonal and off-diagonal entries of A^{-1} can be computed. In network analysis, it is important to calculate the resolvent subgraph centrality $((I_p - \alpha A)^{-1})_{ii}$ and the resolvent subgraph communicability $((I_p - \alpha A)^{-1})_{ij}$ of nodes i and j, where α is an appropriate parameter [5, 12]. The computation of the diagonal of the inverse of a matrix appears also in graph theory, machine learning, electronic structure calculations, and portfolio analysis, and various methods are proposed in the literature for this task [4, 21]. Uncertainty quantification in risk analysis requires the diagonal entries of the inverse covariance matrices for evaluating the degree of confidence in the quality of the data [21].

The trace of the matrix A^{-1}. For appropriately chosen random vectors x, the evaluation of $x^T A^q x$ can lead to the estimation of the trace of matrix powers, denoted as $\mathrm{Tr}(A^q)$, for any $q \in \mathbb{Q}$ as described in [8]. The specific case of $q = -1$ was presented in [7]. This case has attracted a lot of attention in view of the computation of the trace of the inverse $\mathrm{Tr}(A^{-1})$ of symmetric matrices [2, 7, 14]. The trace of the inverse of a matrix is required in several applications arising from the fields of statistics, fractals, lattice quantum chromodynamics, crystals, network analysis, and graph theory [7, 8, 14]. Also, in network analysis, the resolvent Estrada index is calculated from the trace $\mathrm{Tr}((I_p - \alpha A)^{-1})$ [5, 12].

Given a nonsingular matrix $A \in \mathbb{R}^{p\times p}$ and vectors $x, y \in \mathbb{R}^p$, the bilinear form $x^T A^{-1} y$ can be computed explicitly by using dense matrix computational methods. In general, these methods require $\mathcal{O}(p^3)$ floating point operations, and therefore, if A is sufficiently large, a direct method is not an option. In case that A is a symmetric matrix, the bilinear form can be

* Received October 31, 2013. Accepted June 15, 2014. Published online on September 9, 2014. Recommended by L. Reichel.

†University of Athens, Department of Mathematics, Panepistimiopolis, 15784 Athens, Greece ({pfika, mmitroul, parask_roupa}@math.uoa.gr).

ETNA
Kent State University
http://etna.math.kent.edu

transformed to a Riemann-Stieltjes integral and can be approximated by applying quadrature rules using orthogonal polynomial theory and the Lanczos method [1, 14]. These methods require a complexity of order $\mathcal{O}(jp^2)$, where j is the number of (Lanczos) iterations. For any nonsingular complex square matrix A and x, y complex vectors, approaches for approximating $x^* A^{-1} y$ are given in [20]. These approaches are based on non-Hermitian generalizations of Vorobyev moment problems.

The goal of this paper is to develop a different approach which employs extrapolation techniques. For any nonsingular square matrix A and vectors x, y, we obtain estimates for $x^T A^{-1} y$ by extrapolation of the moments of A at the point -1. The derived families of estimates $e_\nu, \nu \in \mathbb{R}$, require only some inner products and few matrix-vector products and can provide an accurate estimate for an appropriate selection of ν. However, the choice of this "good" value of ν is not yet solved and remains an important open problem.

The paper is organized as follows. The extrapolation procedure is developed and approximations of the bilinear form $x^T A^{-1} y$ are presented in Section 2. Applications to the estimation of the elements of A^{-1} and the trace $\mathrm{Tr}(A^{-1})$ are given in Sections 3–4. The extrapolation method is illustrated by numerical experiments in Section 5, where details on the method, other considerations, and comparisons are discussed. Concluding remarks end the paper.

Throughout the paper, $\|x\|$ denotes the Euclidean norm of the vector x, $\|A\|$ is the spectral norm of the matrix A, and $(\cdot, \cdot)$ the Euclidean inner product.

2. Estimation of $\mathbf{x^T A^{-1} y}$ via an extrapolation procedure.

Claude Brezinski in 1999 introduced the extrapolation of the moments of a matrix for estimating the norm of the error when solving a linear system [6]. Extensions of this approach are presented in [9, 10], where applications of the error estimates in least squares problems and regularization are developed. Generalization of the extrapolation approach to Hilbert spaces for compact, linear operators are described in [7] for the estimation of the trace of the inverse of an invertible linear operator and in [8] for the estimation of the trace of a power of a positive self-adjoint linear operator.

Next, we will extend the extrapolation procedure to the derivation of estimates for the bilinear form $x^T A^{-1} y$ for any nonsingular matrix A. We first consider the quadratic case where $x = y$.

2.1. Estimates for $\mathbf{x^T A^{-1} x}$.

Let us recall the singular value decomposition (*SVD*) of a given matrix $A \in \mathbb{R}^{p \times p}$

$$A = U \Sigma V^T = \sum_{k=1}^{p} \sigma_k u_k v_k^T,$$

where $U = [u_1, \ldots, u_p], V = [v_1, \ldots, v_p]$ are orthogonal matrices and the singular values in the diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ are ordered according to $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$.

For a real vector $x \in \mathbb{R}^p$ it holds that

$$Ax = \sum_{k=1}^{p} \sigma_k (v_k, x) u_k, \quad A^T x = \sum_{k=1}^{p} \sigma_k (u_k, x) v_k, \quad \text{and} \quad A^{-1} x = \sum_{k=1}^{p} \sigma_k^{-1} (u_k, x) v_k.$$

Let us define the moments of A as follows:

$$\begin{aligned} c_{2n}(x) &= (x, (A^T A)^n x), & c_{2n+1}(x) &= (x, A(A^T A)^n x), & n &\geq 0, \\ c_{2n}(x) &= (x, (AA^T)^n x), & c_{2n+1}(x) &= (x, A^T (AA^T)^n x), & n &\leq 0. \end{aligned}$$

Using some moments for $n \geq 0$ as interpolation conditions, we obtain estimates by extrapolation of these moments at $n = -1$,

$$c_{-1}(x) = (x, A^T(AA^T)^{-1}x) = (x, A^{-1}x).$$

By defining the moments of the nonsymmetric matrix A in this way, we can now express them as summations derived from the SVD of A :

$$c_{2n}(x) = \sum_k \sigma_k^{2n}(x, v_k)^2 = \sum_k \sigma_k^{2n} a_k^2, \qquad n \geq 0,$$

$$c_{2n}(x) = \sum_k \sigma_k^{2n}(x, u_k)^2 = \sum_k \sigma_k^{2n} b_k^2, \qquad n \leq 0,$$

$$c_{2n+1}(x) = \sum_k \sigma_k^{2n+1}(x, v_k)(x, u_k) = \sum_k \sigma_k^{2n+1} a_k b_k,$$

where $a_k = (x, v_k)$ and $b_k = (x, u_k)$.

We notice that the moment $c_{-1}(x)$ can be estimated by keeping one or two terms in its expansion

$$(x, A^{-1}x) = \sum_k \sigma_k^{-1}(x, v_k)(x, u_k) = \sum_k \sigma_k^{-1} a_k b_k. \tag{2.1}$$

Usually we neither know the singular values nor a_k and b_k, but we are able to compute the moments $c_n(x)$, for $n \geq 0$, by considering appropriate interpolation conditions. In the sequel, the moments $c_n(x)$ will be denoted as c_n, and all the denominators of the estimates are assumed to be different from zero.

One-term estimates. Approximations of c_{-1} can be obtained by keeping only one term in the summation (2.1), that is,

$$c_{-1} = (x, A^{-1}x) \simeq s^{-1}\alpha\beta,$$

where the unknowns s, α, and β are determined by the following interpolation conditions:

$$c_0 = (x, x) = (V^T x, V^T x) = \sum_k (x, v_k)^2 = \alpha^2,$$

$$c_0 = (x, x) = (U^T x, U^T x) = \sum_k (x, u_k)^2 = \beta^2,$$

$$c_1 = (x, Ax) = \sum_k \sigma_k(x, v_k)(x, u_k) = s\alpha\beta,$$

$$c_2 = (x, A^T Ax) = (Ax, Ax) = \sum_k \sigma_k^2(x, v_k)^2 = s^2\alpha^2.$$

This is a nonlinear system of four equations with three unknowns. Since a unique solution does not exist, we obtain the following values of s.

LEMMA 2.1.

$$|s| = |c_0^{-\nu/2-1} c_1^{\nu+1} c_2^{-\nu/2}|, \quad \nu \in \mathbb{R}.$$

Proof. Solving the system, we have the following expressions for s, which are gathered into the compact formula

$$|s| = |c_0^{-i/2-1} c_1^{i+1} c_2^{-i/2}|, \quad i = 0, -1, -2.$$

We notice also that the above formula can be extended to any real number ν. Indeed,

$$|c_0^{-\nu/2-1} c_1^{\nu+1} c_2^{-\nu/2}| = |(\beta^2)^{-\nu/2-1} (s\alpha\beta)^{\nu+1} (s^2\alpha^2)^{-\nu/2}| = |s|,$$

assuming that $\alpha^2 = \beta^2 = c_0$. ∎

Replacing $|s|$ from Lemma 2.1 in the formula $c_{-1} \simeq s^{-2} c_1$, we obtain the following family of estimates for the moment c_{-1}.

$$e_\nu = c_0^{\nu+2} c_1^{-2\nu-1} c_2^{\nu}, \quad \nu \in \mathbb{R}. \tag{2.2}$$

In case that $c_1 = 0$, by choosing $\nu = -1/2$, we avoid division by zero. For small values of c_1, formula (2.2) yields estimates of c_{-1} for appropriate values of $\nu \in \mathbb{R}$.

REMARK 2.2. For $\nu = 0$, formula (2.2) gives $e_0 = c_0^2/c_1$, which is the one-term estimate stated in [7].

PROPOSITION 2.3. *The family of estimates (2.2) satisfy the relations*

$$e_\nu = \rho^\nu e_0, \quad e_\nu = \rho\, e_{\nu-1}, \qquad \textit{where } \rho = c_0 c_2/c_1^2,\ \nu \in \mathbb{R} \tag{2.3}$$

and e_ν is a nondecreasing function of $\nu \in \mathbb{R}$ for $c_1 > 0$ and nonincreasing for $c_1 < 0$.

Proof. We have

$$e_\nu = c_0^{\nu+2} c_1^{-2\nu-1} c_2^{\nu} = \left(\frac{c_0 c_2}{c_1^2}\right)^\nu \frac{c_0^2}{c_1} = \rho^\nu e_0,$$

where $\rho = c_0 c_2/c_1^2$. We also have $e_\nu = \rho^\nu e_0 = \rho(\rho^{\nu-1} e_0) = \rho e_{\nu-1}$.

It holds that $(x, Ax)^2 \le (x, x)(Ax, Ax)$ by the Cauchy-Schwarz inequality [16]. This implies that $c_1^2 \le c_0 c_2$ and thus $\rho = c_0 c_2/c_1^2 \ge 1$. Therefore, if $c_1 > 0$, that is $e_0 > 0$, then $e_\nu = \rho e_{\nu-1} \ge e_{\nu-1}$ for any $\nu \in \mathbb{R}$, whereas if $c_1 < 0$, then $e_\nu = \rho e_{\nu-1} \le e_{\nu-1}$ since e_ν is negative for any $\nu \in \mathbb{R}$. ∎

Next, we see that there exists a ν_0 such that e_{ν_0} gives the exact value of c_{-1}.

LEMMA 2.4. *Let $A \in \mathbb{R}^{p\times p}$ be a positive real matrix, i.e., $(x, Ax) > 0, \forall x \ne 0$. There exists a value ν_0 given by*

$$\nu_0 = \frac{\log(c_{-1}/e_0)}{\log(\rho)}, \quad \rho = c_0 c_2/c_1^2$$

such that $e_{\nu_0} = c_{-1}$.

Proof. Since $\nu \in \mathbb{R}$ and $\rho \ge 1$, it holds that $\lim_{\nu\to+\infty} e_\nu = +\infty$ and $\lim_{\nu\to-\infty} e_\nu = 0$ and thus the domain of the function e_ν is $\mathbb{R}$, and its range is the set of positive real numbers, i.e., $e_\nu : \mathbb{R} \to (0, +\infty)$.

If $c_{-1} > 0$, there exists a value $\nu_0 = \frac{\log(c_{-1}/e_0)}{\log(\rho)} \in \mathbb{R}$ satisfying $e_{\nu_0} = \rho^{\nu_0} e_0 = c_{-1}$. On the other hand, if $c_{-1} < 0$, there exists a value $\nu_0 = \frac{\log(|c_{-1}|/e_0)}{\log(\rho)} + i(\frac{\pi}{\log(\rho)}) = \gamma + \delta i \in \mathbb{C}$ satisfying $e_{\nu_0} = -\rho^\gamma e_0 = -|c_{-1}| = c_{-1}$ since $\delta \log(\rho) = \pi$. ∎

REMARK 2.5. A similar result can be proved if $A \in \mathbb{R}^{p\times p}$ is a negative real matrix, i.e., if $(x, Ax) < 0, \forall x \ne 0$.

Nevertheless, in practice, it is not possible to compute this ideal value ν_0 as it requires a priori knowledge of the exact value of c_{-1}. However, we can find an upper bound for ν_0 by the following result.

ETNA
Kent State University
http://etna.math.kent.edu

COROLLARY 2.6. *Let $A \in \mathbb{R}^{p\times p}$ be a positive real matrix. It holds that*

$$\nu_0 \leq \frac{\log(c_1/(c_0\sigma_p))}{\log(\rho)},$$

where σ_p is the smallest singular value of the matrix $A \in \mathbb{R}^{p\times p}$.

Proof. By the Cauchy-Schwarz inequality, we have

$$c_{-1} = (x, A^{-1}x) \leq \|A^{-1}x\|\|x\| \leq \|A^{-1}\|\|x\|^2 = \frac{c_0}{\sigma_p}.$$

Thus, $\log(\frac{c_{-1}}{e_0}) \leq \log(\frac{c_0}{e_0\sigma_p})$ implies $\nu_0 \leq \frac{\log(c_1/(c_0\sigma_p))}{\log(\rho)}$. □

More one-term estimates. More estimates can be obtained if we consider the following interpolation condition given by the moment $\tilde{c}_2$ instead of c_2,

$$\tilde{c}_2 = (A^T x, A^T x) = \sum_k \sigma_k^2 (x, u_k)^2 = s^2\beta^2.$$

Then, we get the following family of estimates for the moment c_{-1},

$$\tilde{e}_\nu = c_0^{\nu+2} c_1^{-2\nu-1} \tilde{c}_2^{\nu}, \quad \nu \in \mathbb{R}. \tag{2.4}$$

Indeed, $\tilde{e}_\nu = c_0^{\nu+2} c_1^{-2\nu-1} \tilde{c}_2^{\nu} = (\alpha^2)^{\nu+2}(s\alpha\beta)^{-2\nu-1}(s^2\beta^2)^\nu = s^{-1}\alpha\beta$ since $\alpha^2 = \beta^2 = c_0$.

The estimates in (2.4) satisfy the relations

$$\tilde{e}_\nu = \tilde{\rho}^\nu e_0 \qquad \text{for } \tilde{\rho} = c_0 \tilde{c}_2 / c_1^2 \tag{2.5}$$

and have a similar monotonic behavior as those in (2.3).

Another family of estimates for the moment c_{-1}^2 for a symmetric positive definite matrix is given in [9]. More formulae yielding families of one-term estimates valid for any $\nu \in \mathbb{R}$ can be derived. All these formulae, for an appropriate selection of $\nu \in \mathbb{R}$, produce the same one-term estimates.

Two-term estimates. By keeping two terms in relation (2.1), the moment c_{-1} can be approximated as follows.

$$c_{-1} = (x, A^{-1}x) \simeq \hat{e}_\nu = s_1^{-1}\alpha_1\beta_1 + s_2^{-1}\alpha_2\beta_2.$$

Let us consider the following interpolation conditions:

$$\begin{aligned}
c_0 &= (x,x) = (V^T x, V^T x) = \sum_k (x, v_k)^2 = \alpha_1^2 + \alpha_2^2,\\
c_0 &= (x,x) = (U^T x, U^T x) = \sum_k (x, u_k)^2 = \beta_1^2 + \beta_2^2,\\
c_{2j} &= (x, (A^T A)^j x) = \sum_k \sigma_k^{2j} (x, v_k)^2 = s_1^{2j}\alpha_1^2 + s_2^{2j}\alpha_2^2,\\
\tilde{c}_{2j} &= (x, (AA^T)^j x) = \sum_k \sigma_k^{2j} (x, u_k)^2 = s_1^{2j}\beta_1^2 + s_2^{2j}\beta_2^2,\\
c_{2j+1} &= (x, A(A^T A)^j x) = \sum_k \sigma_k^{2j+1} (x, v_k)(x, u_k) = s_1^{2j+1}\alpha_1\beta_1 + s_2^{2j+1}\alpha_2\beta_2,\\
\tilde{c}_{2j+1} &= (x, A^T(AA^T)^j x) = \sum_k \sigma_k^{2j+1} (x, v_k)(x, u_k) = s_1^{2j+1}\alpha_1\beta_1 + s_2^{2j+1}\alpha_2\beta_2,
\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

for different values of $j \geq 0$. The following family of two-term estimates $\hat{e}_\nu$ can be derived,

$$\hat{e}_\nu = e_0 + \frac{c_0 c_2 - c_1^2}{c_1} \frac{c_0 \tilde{c}_{\nu+2} - c_1 c_{\nu+1}}{c_1 c_{\nu+3} - c_2 \tilde{c}_{\nu+2}}, \qquad \nu \in \mathbb{N}, \tag{2.6}$$

where e_0 is the one-term estimate of (2.2) for $\nu = 0$. Indeed, replacing in (2.6) the interpolation conditions given by the moments $c_{\nu+1}$, $c_{\nu+2}$, $\tilde{c}_{\nu+2}$, $c_{\nu+3}$, ($\tilde{c}_\nu = c_\nu$, if ν is odd), we get $\hat{e}_\nu = s_1^{-1}\alpha_1\beta_1 + s_2^{-1}\alpha_2\beta_2$.

In case that $c_1 = 0$, equation (2.6) can be rewritten in the form

$$\hat{e}_\nu = \frac{c_0^2 c_{\nu+3} - c_0 c_2 c_{\nu+1} - c_0 c_1 \tilde{c}_{\nu+2} + c_1^2 c_{\nu+1}}{c_1 c_{\nu+3} - c_2 \tilde{c}_{\nu+2}},$$

and thus division by zero is avoided.

As a consequence of the Cauchy-Schwarz inequality, the denominator $c_1 c_{\nu+3} - c_2 \tilde{c}_{\nu+2}$ is always positive for a symmetric positive definite matrix. When the Cauchy-Schwarz inequality holds as an equality, i.e., the vector x coincides with an eigenvector of the matrix, formula (2.6) cannot be used since the denominator is equal to zero.

REMARK 2.7. The moments of formula (2.6) are indexed by ν, and thus it is required that $\nu \in \mathbb{N}$. On the contrary, the moments of (2.2) are raised to powers of ν which can be any real number.

2.2. Estimates for $\mathbf{x^T A^{-1} y}$. For $x \neq y$, we define the bilinear moment

$$c_{-1}(x, y) = (x, A^{-1}y).$$

It holds that $x^T A^{-1} y = x^T (A^T A)^{-1} u$, where $u = A^T y$. Then we can use the polarization identity $x^T(A^T A)^{-1}u = \frac{1}{4}(w^T(A^T A)^{-1}w - z^T(A^T A)^{-1}z)$, where $w = x + u$ and $z = x - u$.

We set the moments $g_n(x) = (x, (A^T A)^n x)$, $n \in \mathbb{Z}$. Then,

$$c_{-1}(x, y) = \frac{1}{4}(g_{-1}(w) - g_{-1}(z)). \tag{2.7}$$

Estimates for $c_{-1}(x, y)$ can be obtained by considering one- or two-term estimates for the moments $g_{-1}(w)$ and $g_{-1}(z)$ from formulae (2.3), (2.5), and (2.6), respectively.

2.3. Estimates for symmetric matrices. If A is a symmetric matrix, we can prove that e_0 in (2.3) and $\hat{e}_0$ in (2.6) are lower bounds for c_{-1}.

LEMMA 2.8. *Let $A \in \mathbb{R}^{p \times p}$ be a symmetric matrix. The one-term estimate e_0 coincides with the lower bound for $(x, A^{-1}x)$ obtained by using the Gauss quadrature rule and one Lanczos iteration of the approach presented in* [14].

The two-term estimate $\hat{e}_0$ coincides with the lower bound for $(x, A^{-1}x)$ obtained by using the Gauss quadrature rule and two Lanczos iterations [14].

Proof. In [14] it is proved that the estimate (lower bound using the Gauss quadrature rule) of $x^T A^{-1} x$ obtained in the kth iteration is given by the (1,1) element of the inverse of the $k \times k$ Jacobi matrix J_k, multiplied by $\|x\|^2$. For one Lanczos iteration ($k = 1$), the Jacobi matrix becomes $J_1 = [u^T A u]$, where $u = x/\|x\|$. Therefore,

$$x^T A^{-1} x \simeq \|x\|^2 (u^T A u)^{-1} = \|x\|^2 (x^T A x / \|x\|^2)^{-1} = c_0 (c_1/c_0)^{-1} = c_0^2/c_1 = e_0.$$

For $k = 2$ Lanczos iterations, applying the algorithm of [1, 14], we have the Jacobi matrix

$$J_2 = \begin{bmatrix} c_1/c_0 & \|r\| \\ \|r\| & \mu \end{bmatrix}, \quad \text{where} \quad r = A\frac{x}{\|x\|} - \frac{c_1}{c_0}\frac{x}{\|x\|}, \quad \mu = \frac{r^T A r}{\|r\|^2} = \frac{r^T A r}{r^T r}.$$

Since A is symmetric, we have $r^T r = \frac{c_2}{c_0} - \frac{c_1^2}{c_0^2}$ and $r^T A r = \frac{c_3 c_0^2 + c_1^3 - 2c_1 c_2 c_0}{c_0^3}$. Thus, $\mu = \frac{r^T A r}{r^T r} = \frac{c_3 c_0^2 + c_1^3 - 2c_1 c_2 c_0}{c_0^2 c_2 - c_0 c_1^2}$. It holds that

$$J_2^{-1} = \frac{1}{\frac{c_1}{c_0}\mu - \|r\|^2} \begin{bmatrix} \mu & -\|r\| \\ -\|r\| & c_1/c_0 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} x^T A^{-1} x \simeq \|x\|^2 J_2^{-1}(1,1) &= c_0 \frac{\mu}{\frac{c_1}{c_0}\mu - \|r\|^2} \\ &= \frac{c_1^3 + c_0^2 c_3 - 2c_0 c_1 c_2}{c_1 c_3 - c_2^2} = \frac{c_0^2}{c_1} + \frac{(c_0 c_2 - c_1^2)^2}{c_1^2 c_3 - c_1 c_2^2}, \end{aligned}$$

which is $\hat{e}_0$ in (2.6) since $\tilde{c}_2 = c_2$. $\square$

If the matrix A is also positive definite, additional bounds for ν_0 can be derived. Corollary 2.6 yields that $\nu_0 \leq \frac{\log(c_1/(c_0 \lambda_{min}))}{\log(\rho)}$, where λ_{min} is the smallest eigenvalue of the matrix A. In addition, we obtain an interval in which ν_0 lies.

PROPOSITION 2.9. *Let $A \in \mathbb{R}^{p \times p}$ be a symmetric positive definite matrix. It holds that*

$$0 \leq \nu_0 \leq \frac{\log(m)}{\log(\rho)},$$

where $m = \frac{(1+\kappa(A))^2}{4\kappa(A)}$ and $\kappa(A)$ is the spectral condition number of A.

Proof. For a symmetric positive definite matrix A and for any vector x, it holds that [7]

$$\frac{c_0^2}{c_1} \leq x^T A^{-1} x \leq m \frac{c_0^2}{c_1}.$$

Thus, $0 < e_0 \leq c_{-1} \leq m e_0$, and therefore

$$e_0 \leq c_{-1} \quad \Rightarrow \quad \frac{\log(1)}{\log(\rho)} \leq \frac{\log(c_{-1}/e_0)}{\log(\rho)} \quad \Rightarrow \quad \nu_0 \geq 0$$

since $\log(\rho) > 0$ as $\rho \geq 1$. Respectively, we get

$$c_{-1} \leq m e_0 \quad \Rightarrow \quad \log(\frac{c_{-1}}{e_0}) \leq \log(m) \quad \Rightarrow \quad \nu_0 \leq \frac{\log(m)}{\log(\rho)}. \qquad \square$$

REMARK 2.10. The double inequality of Proposition 2.9 shows that, if A is orthogonal, then $\kappa(A) = 1$, and it follows that $\nu_0 = 0$, which shows that $e_0 = c_{-1}$.

Since A is symmetric, we can use the polarization identity

$$x^T A^{-1} y = \frac{1}{4}(w^T A^{-1} w - z^T A^{-1} z)$$

for the evaluation of the bilinear form $x^T A^{-1} y$. Then, the bilinear moment $c_{-1}(x, y)$ can be expressed as

$$c_{-1}(x, y) = \frac{1}{4}(c_{-1}(w) - c_{-1}(z)), \tag{2.8}$$

where $w = x + y$ and $z = x - y$.

Estimates for $c_{-1}(x, y)$ can be obtained by considering one- or two-term estimates for the moments $c_{-1}(w)$ and $c_{-1}(z)$ given by formulae (2.3), (2.5), and (2.6), respectively.

ETNA
Kent State University
http://etna.math.kent.edu

3. Estimates for the elements of the matrix $\mathbf{A^{-1}}$. Let $A = [a_{ij}] \in \mathbb{R}^{p\times p}$, for $i, j = 1, \ldots, p$, and let δ_i be the ith column of the identity matrix. Then $c_{-1}(\delta_i) = (A^{-1})_{ii}$.

PROPOSITION 3.1. *The families of one-term estimates for the diagonal elements of the matrix A^{-1} are*

$$(A^{-1})_{ii} \simeq \rho^\nu \frac{1}{a_{ii}}, \quad \rho = \frac{s_i}{a_{ii}^2} \qquad \textit{or} \qquad (A^{-1})_{ii} \simeq \tilde{\rho}^\nu \frac{1}{a_{ii}}, \quad \tilde{\rho} = \frac{\tilde{s}_i}{a_{ii}^2}, \qquad \nu \in \mathbb{R},$$

where $s_i = \sum_{k=1}^p a_{ki}^2$ and $\tilde{s}_i = \sum_{k=1}^p a_{ik}^2$.

Proof. We have

$$\begin{aligned}
c_0(\delta_i) &= \delta_i^T \delta_i = 1,\\
c_1(\delta_i) &= \delta_i^T A \delta_i = a_{ii},\\
c_2(\delta_i) &= (A\delta_i)^T A\delta_i = \delta_i^T A^T A \delta_i = \sum_{k=1}^p a_{ki}^2 = s_i,\\
\tilde{c}_2(\delta_i) &= (A^T\delta_i)^T A^T\delta_i = \delta_i^T A A^T \delta_i = \sum_{k=1}^p a_{ik}^2 = \tilde{s}_i.
\end{aligned}$$

Replacing the above quantities in formulae (2.3) and (2.5), we obtain the result. □

PROPOSITION 3.2. *The one-term estimates for the elements of the matrix A^{-1} using the one-term estimates e_0 of (2.2) are*

$$\begin{aligned}
(A^{-1})_{ii} &\simeq \frac{1}{a_{ii}},\\
(A^{-1})_{ij} &\simeq \frac{1}{4}\left(\frac{(\tilde{s}_j + 2a_{ji} + 1)^2}{\sum_{t=1}^p (s_{tj} + a_{ti})^2} - \frac{(\tilde{s}_j - 2a_{ji} + 1)^2}{\sum_{t=1}^p (-s_{tj} + a_{ti})^2} \right),
\end{aligned}$$

where $s_{tj} = \sum_{k=1}^p a_{tk} a_{jk}$ and $\tilde{s}_j$ is as in Proposition 3.1.

Proof. The diagonal entries of the matrix A^{-1} can be estimated using the one-term estimate e_0 of (2.2), i.e., $c_{-1}(\delta_i) \simeq e_0(\delta_i) = c_0^2(\delta_i)/c_1(\delta_i)$. It holds that

$$c_{-1}(\delta_i) = \delta_i^T A^{-1} \delta_i = (A^{-1})_{ii}, \quad c_0(\delta_i) = \delta_i^T \delta_i = 1, \quad \text{and} \quad c_1(\delta_i) = \delta_i^T A \delta_i = a_{ii},$$

and thus

$$(A^{-1})_{ii} \simeq e_0(\delta_i) = a_{ii}^{-1}.$$

By setting $w = \delta_i + A^T\delta_j$ and $z = \delta_i - A^T\delta_j$ in (2.7), we obtain estimates for the off-diagonal elements of A^{-1}. Replacing the moments c_i by the moments g_i in the one-term estimates e_0 for (2.2), we obtain estimates for the moments $x^T(A^TA)^{-1}x$. It holds that

$$\begin{aligned}
g_0(w) = w^T w &= \sum_{k=1, k\neq i}^p a_{jk}^2 + (1 + a_{ji})^2 = \sum_{k=1}^p a_{jk}^2 + 1 + 2a_{ji},\\
g_0(z) = z^T z &= \sum_{k=1, k\neq i}^p a_{jk}^2 + (1 - a_{ji})^2 = \sum_{k=1}^p a_{jk}^2 + 1 - 2a_{ji},
\end{aligned}$$

$$
\begin{aligned}
g_1(w) = w^T(A^TA)w = (Aw)^T(Aw) &= \sum_{t=1}^{p}(a_{ti}(1+a_{ji}) + \sum_{k=1,k\neq i}^{p} a_{tk}a_{jk})^2 \\
&= \sum_{t=1}^{p}(\sum_{k=1}^{p} a_{tk}a_{jk} + a_{ti})^2, \\
g_1(z) = z^T(A^TA)z = (Az)^T(Az) &= \sum_{t=1}^{p}(a_{ti}(1-a_{ji}) - \sum_{k=1,k\neq i}^{p} a_{tk}a_{jk})^2 \\
&= \sum_{t=1}^{p}(-\sum_{k=1}^{p} a_{tk}a_{jk} + a_{ti})^2.
\end{aligned}
$$

Replacing in $c_{-1}(x,y) \simeq \frac{1}{4}(g_0^2(w)/g_1(w) - g_0^2(z)/g_1(z))$ the moments $g_0(w)$, $g_1(w)$, $g_0(z)$, and $g_1(z)$, we obtain the result. □

3.1. Estimates for the elements of symmetric matrices. If A is a symmetric matrix, then further estimates for its off-diagonal entries can be derived.

PROPOSITION 3.3. *The one-term estimate for the off-diagonal elements of A^{-1} using the one-term estimates e_0 of (2.2) is*

$$
(A^{-1})_{ij} \simeq \frac{-4a_{ij}}{(a_{ii}+a_{jj})^2 - 4a_{ij}^2}, \qquad i \neq j.
$$

Proof. By setting $x = \delta_i$ and $y = \delta_j$ in the bilinear form $x^TA^{-1}y$ and using the polarization identity (2.8), we obtain estimates for the off-diagonal elements of A^{-1}. It holds that $\delta_i^TA^{-1}\delta_j = \frac{1}{4}(w^TA^{-1}w - z^TA^{-1}z)$, where $w = \delta_i + \delta_j$ and $z = \delta_i - \delta_j$. Using the one-term formula e_0 for the moments $w^TA^{-1}w$ and $z^TA^{-1}z$ and considering that $\delta_i^TA^{-1}\delta_j = (A^{-1})_{ij}$, we get $(A^{-1})_{ij} \simeq \frac{1}{4}(c_0^2(w)/c_1(w) - c_0^2(z)/c_1(z))$. Since, for $i \neq j$ it holds that

$$
\begin{aligned}
&c_0(w) = 2 = c_0(z), \quad c_1(w) = w^TAw = a_{ii} + a_{jj} + 2a_{ij}, \quad \text{and} \\
&c_1(z) = z^TAz = a_{ii} + a_{jj} - 2a_{ij},
\end{aligned}
$$

the above formula yields

$$
(A^{-1})_{ij} \simeq \frac{1}{4}\left[\left(\frac{4}{a_{ii}+a_{jj}+2a_{ij}}\right) - \left(\frac{4}{a_{ii}+a_{jj}-2a_{ij}}\right)\right] = \frac{-4a_{ij}}{(a_{ii}+a_{jj})^2 - 4a_{ij}^2}. \qquad \square
$$

Next, we can prove that e_0 in (2.2) and $\hat{e}_0$ in (2.6) for $x = \delta_i$ are lower bounds for the diagonal elements of A^{-1}.

LEMMA 3.4. *The one-term estimate for the entry $(A^{-1})_{ii}$, $e_0(\delta_i) = 1/a_{ii}$, coincides with the lower bound of $(A^{-1})_{ii}$ given in [14] using the Gauss quadrature rule and one Lanczos iteration.*

Proof. Following Lemma 2.8, the estimate (lower bound using the Gauss rule) for $x^TA^{-1}x$ obtained in the first iteration is

$$
x^TA^{-1}x \simeq (x^TAx/\|x\|^2)^{-1}\|x\|^2 = (c_1/c_0)^{-1}c_0 = c_0^2/c_1,
$$

which implies that

$$
(A^{-1})_{ii} = \delta_i^TA^{-1}\delta_i \simeq c_0^2(\delta_i)/c_1(\delta_i) = 1/a_{ii}. \qquad \square
$$

ETNA
Kent State University
http://etna.math.kent.edu

LEMMA 3.5. *The two-term estimate for the entry $(A^{-1})_{ii}$, $\hat{e}_0$ in (2.6) for $x = \delta_i$, coincides with the lower bound of $(A^{-1})_{ii}$ given in [14] using the Gauss quadrature rule and two Lanczos iterations.*

Proof. In the same way as for $k = 1$, we obtain lower bounds for any $k \in \mathbb{N}$. For $k = 2$, the following formula obtained by Gauss quadrature is given in [14, Theorem 11.1].

$$\frac{s_{ii}}{a_{ii}s_{ii} - (\sum_{k\neq i} a_{ki}^2)^2} \leq (A^{-1})_{ii}, \qquad i = 1, \dots, p, \tag{3.1}$$

where $s_{ii} = \sum_{t\neq i}\sum_{k\neq i} a_{ti}a_{kt}a_{ki}$.

The estimate $\hat{e}_0$ in (2.6) for $x = \delta_i$ gives

$$(A^{-1})_{ii} \simeq \frac{c_1^3(\delta_i) + c_3(\delta_i) - 2c_1(\delta_i)c_2(\delta_i)}{c_1(\delta_i)c_3(\delta_i) - c_2^2(\delta_i)}$$

since $c_0(\delta_i) = 1$.

For a symmetric matrix A, it holds that

$$c_1(\delta_i) = a_{ii}, \qquad c_2(\delta_i) = \sum_{k=1}^{p} a_{ki}^2 = \sum_{k\neq i} a_{ki}^2 + a_{ii}^2,$$

and

$$\begin{aligned}
c_3(\delta_i) &= \sum_{t=1}^{p}\sum_{k=1}^{p} a_{it}a_{tk}a_{ki} = \sum_{t=1}^{p}\left(\sum_{k\neq i} a_{it}a_{tk}a_{ki} + a_{it}a_{ti}a_{ii}\right)\\
&= \sum_{t=1}^{p}\sum_{k\neq i} a_{it}a_{tk}a_{ki} + \sum_{t=1}^{p} a_{it}a_{ti}a_{ii}\\
&= \sum_{t\neq i}\sum_{k\neq i} a_{it}a_{tk}a_{ki} + \sum_{k\neq i} a_{it}a_{ti}a_{ii} + \sum_{t=1}^{p} a_{it}a_{ti}a_{ii}\\
&= s_{ii} + c_1(\delta_i)\sum_{k\neq i} a_{ki}^2 + c_1(\delta_i)\sum_{t=1}^{p} a_{it}a_{ti}\\
&= s_{ii} + c_1(\delta_i)(c_2(\delta_i) - c_1^2(\delta_i)) + c_1(\delta_i)c_2(\delta_i) = s_{ii} + 2c_1(\delta_i)c_2(\delta_i) - c_1^3(\delta_i).
\end{aligned}$$

Thus, $\sum_{k\neq i} a_{ki}^2 = c_2(\delta_i) - c_1^2(\delta_i)$, and $s_{ii} = c_3(\delta_i) - 2c_1(\delta_i)c_2(\delta_i) + c_1^3(\delta_i)$. Inserting these identities into (3.1), we have

$$\frac{s_{ii}}{a_{ii}s_{ii} - (\sum_{k\neq i} a_{ki}^2)^2} = \frac{c_1^3(\delta_i) + c_3(\delta_i) - 2c_1(\delta_i)c_2(\delta_i)}{c_1(\delta_i)c_3(\delta_i) - c_2^2(\delta_i)}. \qquad \square$$

4. A family of estimates for the trace of the matrix $\mathbf{A^{-1}}$. For a symmetric matrix A, the trace of its inverse, $\mathrm{Tr}(A^{-1})$, can be related to the moment c_{-1} due to a stochastic result proved by Hutchinson in [17]. Let X be a discrete random variable taking the values 1 and -1 with equal probability 0.5, and let x be a vector of p independent samples from X (for simplicity, we write in this case $x \in X^p$). It holds that $E(c_{-1}(x)) = \mathrm{Tr}(A^{-1})$, where $E(\cdot)$ denotes the expected value.

Let $A \in \mathbb{R}^{p\times p}$ be any nonsingular matrix. In order to apply Hutchinson's result for the estimation of $\mathrm{Tr}(A^{-1})$, we define the matrix [1, 7]

$$M = \frac{1}{2}(A^{-1} + A^{-T}) = \frac{1}{2}((A^TA)^{-1}A^T + A(A^TA)^{-1}).$$

ETNA
Kent State University
http://etna.math.kent.edu

The matrix M is symmetric and Tr$(M) = $ Tr(A^{-1}). We define the moments

$$d_n(x) = (x, (((A^TA)^nA^T + A(A^TA)^n)/2)x), \quad n = -1, 0, 1, \dots.$$

In the sequel, d_n denotes the moment $d_n(x)$. We have

$$d_{-1} = (z, Mz) = c_{-1}, \qquad d_0 = (x, ((A^T + A)/2)x) = (x, Ax) = c_1,$$

and, more generally, for all $n = -1, 0, 1, \dots$,

$$d_n = \sum_{k=1}^{p} \sigma_k^{2n+1} a_k b_k, \tag{4.1}$$

where a_k and b_k are defined in Section 2.1. Applying the extrapolation procedure, we obtain one- and two-term estimates for the moment d_{-1}. By keeping one term in the summation (4.1) and imposing that $d_n = s^{2n+1}\alpha\beta$, the following expressions of s can be derived.

LEMMA 4.1.

$$s^2 = d_0^{-\nu/2-1} d_1^{\nu+1} d_2^{-\nu/2}, \qquad \nu \in \mathbb{R}.$$

Since $d_{-1} \simeq s^{-4}d_1$, we get the following family of estimates for the moment d_{-1}.

$$t_\nu = d_0^{\nu+2} d_1^{-2\nu-1} d_2^{\nu} \simeq d_{-1}, \qquad \nu \in \mathbb{R}. \tag{4.2}$$

Then, $E(t_\nu)$, for $x \in X^p$, is an estimate for Tr(A^{-1}) for any matrix A.

By keeping two terms in the summation (4.1), along the same lines as in Section 2, we obtain a family of estimates $\hat{t}_\nu$ which are the same as $\hat{e}_\nu$ in (2.6) with the moments d_i in place of c_i. The expected values of these estimates $E(\hat{t}_\nu)$, for $x \in X^p$, are estimates for Tr(A^{-1}).

5. Implementation and numerical examples.

5.1. Computational complexity of the estimates. The one- and two-term estimates require some inner products and few matrix-vector products (mvps). It is worth pointing out that the evaluation of $(A^TA)^n$, $n = 1, 2$, required for the initial moments c_2, c_3, c_4, and d_1, d_2, d_3, d_4, is never carried out by explicitly forming the products $(A^TA)x$, $(A^TA)^2x$, $(A^TA)^kA^Tx$, and $A(A^TA)^kx$, $k = 1, 2, \dots$, but these expressions are computed by successive matrix-vector products.

In particular, by computing the initial matrix-vector product (mvp) $w_1 = Ax$, the moments $c_1 = x^Tw_1$ and $c_2 = w_1^Tw_1$ are derived by only one additional inner product. In this way, for symmetric A, by computing $w_2 = Aw_1$, we obtain the moments $c_3 = w_1^Tw_2$ and $c_4 = w_2^Tw_2$ with two more inner products. If A is a nonsymmetric matrix, the additional mvps $\tilde{w}_1 = A^Tx$, $\tilde{w}_2 = A^Tw_1$, and $\tilde{w}_3 = A\tilde{w}_1$ are required for the moments $c_3 = \tilde{w}_1^T\tilde{w}_2$, $c_4 = \tilde{w}_2^T\tilde{w}_2$, $\tilde{c}_2 = \tilde{w}_1^T\tilde{w}_1$, and $\tilde{c}_2 = \tilde{w}_3^T\tilde{w}_3$.

Table 5.1 displays the number of arithmetic operations required for the computation of the estimates e_ν, $\tilde{e}_\nu$, and $\hat{e}_\nu$ for dense and banded matrices with bandwidth q. In Table 5.2, we can observe the number of arithmetic operations required for the estimation of the bilinear form $x^TA^{-1}y$ using (2.7) for any matrix A. Formula (2.8) requires twice the number of operations reported in Table 5.1 for symmetric matrices. We notice that the computation of each moment $g_n(x) = (x, (A^TA)^nx)$, $n = 0, 1, \dots$, requires n mvps.

The elements of A^{-1} can be approximated by the one-term estimate e_0 performing only few scalar operations. Any other e_ν requires arithmetic operations of order $\mathcal{O}(p)$, whereas the two-term estimates require operations of order $\mathcal{O}(p^2)$. Table 5.3 gives the number of arithmetic operations required for the computation of the estimates t_ν in (4.2).

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.1
Arithmetic operations for the estimation of the moment $x^T A^{-1} x$.

Matrix A	e_ν	$\tilde{e}_\nu$	$\hat{e}_\nu$, ν even	$\hat{e}_\nu$, ν odd
dense	$\mathcal{O}(p^2)$	$\mathcal{O}(2p^2)$	$\mathcal{O}((\nu+3)p^2)$	$\mathcal{O}((\nu+2)p^2)$
symmetric dense	$\mathcal{O}(p^2)$	$\mathcal{O}(p^2)$	$\mathcal{O}((\nu/2+2)p^2)$	$\mathcal{O}((\nu/2+3/2)p^2)$
banded	$\mathcal{O}(qp)$	$\mathcal{O}(2qp)$	$\mathcal{O}((\nu+3)qp)$	$\mathcal{O}((\nu+2)qp)$
symmetric banded	$\mathcal{O}(qp)$	$\mathcal{O}(qp)$	$\mathcal{O}((\nu/2+2)qp)$	$\mathcal{O}((\nu/2+3/2)qp)$

TABLE 5.2
Arithmetic operations for the estimation of the bilinear form $x^T A^{-1} y$.

Matrix A	e_0	e_ν	$\tilde{e}_\nu$	$\hat{e}_\nu$
nonsymmetric	$\mathcal{O}(3p^2)$	$\mathcal{O}(5p^2)$	$\mathcal{O}(9p^2)$	$\mathcal{O}((2\nu+7)p^2)$

TABLE 5.3
Arithmetic operations for the estimation of the moment d_{-1}.

t_0	t_ν	$\hat{t}_\nu$
$\mathcal{O}(6p^2)$	$\mathcal{O}(10p^2)$	$\mathcal{O}((14+4\nu)p^2)$

5.2. Numerical examples. This section presents extensive numerical experiments validating the behavior of the one- and two-term estimates. All computations were performed in vectorized form in MATLAB (R2009b) 64-bit on an Intel Core i7 computer with 8 Gb RAM. The so–called *exact* values reported in this section were obtained by using the *inv* function in MATLAB.

EXAMPLE 5.1 (Monotonic behavior of the one-term estimates). We test the monotony of the family of one-term estimates $e_\nu = \rho^\nu e_0$ in (2.3). We consider the *parter* matrix of order 3000 obtained by using the *gallery* function in MATLAB. *Parter* is a well-conditioned ($\kappa(A) = 4.6694$) Cauchy- and Toeplitz matrix with elements $a_{ij} = 1/(i-j+0.5)$. In Table 5.4 we estimate the element $A^{-1}_{1500,1500} = 2.0271\text{e-}1$. The best value of ν is $\nu_0 = -9.9978\text{e-}1$. We notice that the one-term estimates e_ν increase as ν increases since $c_1 = 2 > 0$.

We also consider the *orsreg1* matrix of order 2205 obtained from the *University of Florida Sparse Matrix Collection* [11]. This matrix is sparse and ill-conditioned ($\kappa(A) = 1.5394\text{e}4$). In Table 5.5 we estimate the element $A^{-1}_{1490,1490} = -5.7741\text{e-}3$. The best value of ν is $\nu_0 = 5.0027$. We notice that e_ν is a decreasing function of ν since $c_1 = -1.2640\text{e}4$.

In Figure 5.1 we illustrate the quality of approximating a part of the diagonal of A^{-1}. We depict the exact value and the one-term estimates of (2.3) for different values of ν for the first 50 diagonal elements of the *grcar* matrix of order 4000. The *grcar* matrix, which is Toeplitz and well-conditioned ($\kappa(A) = 3.6277$), is obtained by using the *gallery* function in MATLAB. We observe that for most of the elements, $e_{-0.75}$ is a good estimate. We notice that e_ν increases as ν increases.

TABLE 5.4
Increasing family of estimates for $A^{-1}_{1500,1500} = 2.0271\text{e-}1$ *for the parter matrix of order* 3000.

ν	e_ν	Relative error
-1	2.0267e-1	1.9821e-4
ν_0	2.0271e-1	Exact value
-0.9	2.2182e-1	9.4289e-2
-0.8	2.4279e-1	1.9771e-1
-0.7	2.6573e-1	3.1090e-1
-0.6	2.9084e-1	4.3478e-1

TABLE 5.5
Decreasing family of estimates for $A^{-1}_{1490,1490} = -5.7741\text{e-}3$ *for the orsreg1 matrix of order 2205.*

ν	e_ν	Relative error
2	-4.3969e-4	9.2385e-1
2.5	-6.7510e-4	8.8308e-1
3	-1.0365e-3	8.2048e-1
3.5	-1.5915e-3	7.2437e-1
4	-2.4436e-3	5.7680e-1
4.5	-3.7519e-3	3.5022e-1
5	-5.7606e-3	2.3254e-3
ν_0	-5.7741e-3	Exact value
5.5	-8.8448e-3	5.3183e-1
6	-1.3580e-2	1.3520e0
6.5	-2.0851e-2	2.6112e0
7	-3.2015e-2	4.5446e0

5.3. Estimates for matrix entries. If we consider symmetric matrices, we can compare the estimates derived from Gauss quadrature methods with the estimates produced by extrapolation.

EXAMPLE 5.2 (Comparison with the Gauss quadrature method). In Table 5.6, we report the results for an example also given in [14, Table 11.6] for the *Poisson* matrix of order 900 obtained by using the *gallery* function in MATLAB. This matrix is symmetric, block tridiagonal (sparse), and ill-conditioned ($\kappa(A) = 5.6492\text{e}2$). We estimate the element $A^{-1}_{150,150} = 0.3602$. We observe that e_0, which is also the bound obtained by Gauss quadrature in one iteration ($k = 1$), is not a good approximation. However, for $\nu = 2.12$, the value of $e_\nu = 0.3599$ is a fair estimation attained by one mvp. The best value of ν is $\nu_0 = 2.1250$.

Using Gauss or Gauss-Radau quadrature rules, we obtain the same value 0.3599 after $k = 20$ iterations, whereas a very good approximation of the exact value is achieved after $k = 40$ iterations.

In Table 5.7 we report results for an example also given in [1, Table 1] for the *Heat flow* matrix of order 900. This matrix is symmetric, block tridiagonal (sparse), and well-conditioned ($\kappa(A) = 2.6$). We estimate the element $A^{-1}_{1,1} = 0.5702$. The best value of ν is $\nu_0 = 1.0668$. We notice that the relative error of the one-term estimate e_ν for $\nu = 1$ (which is very close to ν_0) is of order $\mathcal{O}(10^{-3})$. The two-term estimates $\hat{e}_0$ and $\hat{e}_1$ do not reduce the order of the relative error. However, a relative error of order $\mathcal{O}(10^{-6})$ can be attained by Gauss quadrature in only $k = 4$ iterations.

ETNA
Kent State University
http://etna.math.kent.edu

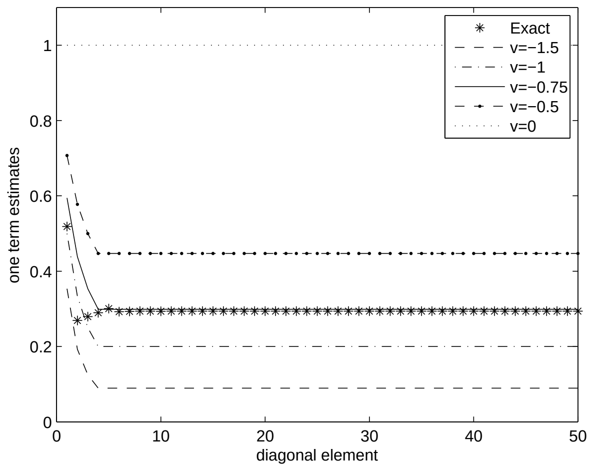

FIG. 5.1. *Estimating the diagonal of the inverse of the grcar matrix of order 4000.*

TABLE 5.6
Estimates for $A^{-1}_{150,150} = 0.3602$ *for the Poisson matrix of order 900.*

	Relative error	Estimates
$e_0 =$ Gauss $(k = 1)$	3.0593e-1	0.2500
e_2	2.1251e-2	0.3525
$e_{2.1}$	4.2858e-3	0.3586
$e_{2.12}$	8.5768e-4	0.3599
$\hat{e}_0 =$ Gauss $(k = 2)$	1.4576e-1	0.3077
$\hat{e}_1$	1.6555e-1	0.3006
Gauss $(k = 20)$	8.2489e-4	0.3599
Gauss $(k = 40)$	2.9294e-5	0.3602

EXAMPLE 5.3 (Estimating the diagonal elements of the inverse of covariance matrices). *Covariance* matrices are symmetric positive definite of the form $A = XX^T$, where X is the data matrix. We tested *covariance matrices* (α, β) with entries $a_{ii} = 1 + i^{\alpha}$ and $a_{ij} = 1/|i-j|^{\beta}$, for $i \neq j$, where $\alpha, \beta \in \mathbb{R}$ [3].

The mean relative error of the diagonal entries of a matrix is defined as the value $\left(\sum_{i=1}^{p} |a_{ii} - e(\delta_i)|/|a_{ii}|\right)/p$. Table 5.8 presents the mean relative errors of the diagonal elements of inverse covariance matrices of order $p = 4000$ for various α and β using the one-term estimates e_ν for $\nu = 0, 1/4, 1/2, 3/4, 1$. We notice that even the one-term estimate $e_0 = 1/a_{ii}$, which requires only one division, gives a good result. In the last row we report the execution time in second required for the estimation of the whole diagonal.

5.4. Estimation of Tr$(\mathbf{A}^{-1})$. Estimates for Tr(A^{-1}) can be obtained by realizing N experiments and then computing the mean value of the quantities $t(x_i)$, for $x_i \in X^p$, where $t(x_i)$ denote any of the one-term or two-term estimates for the moment $d_{-1}(x_i)$, i.e., Tr$(A^{-1}) \simeq \tau = \frac{1}{N}\sum_{i=1}^{N} t(x_i)$. Actually, the computation of the one-term or two-term trace estimates τ require N times the arithmetic operations reported in Tables 5.1 and 5.3. More details about the implementation of the trace computation can be found in [7]. For the esti-

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.7
Estimates for $A^{-1}_{1,1} = 0.5702$ for the Heat flow matrix of order 900.

	Relative error	Estimates
$e_0 =$ Gauss $(k = 1)$	2.5686e-2	0.5556
e_1	1.6284e-3	0.5693
$\hat{e}_0 =$ Gauss $(k = 2)$	1.0194e-3	0.5696
$\hat{e}_1$	1.4790e-3	0.5694
Gauss $(k = 4)$	2.2083e-6	0.5702

TABLE 5.8
Mean relative error of the diagonal entries of the inverse of covariance matrices.

	(α, β)			
	$(1, 2)$	$(2, 1/2)$	$(1/2, 4)$	$(1, 1)$
$\kappa(A)$	2.9956e3	9.6207e6	7.6118e1	2.9109e3
e_0	2.4416e-4	8.0099e-5	3.0162e-3	2.6710e-4
$e_{1/4}$	1.8553e-4	6.2590e-5	2.3172e-3	1.8500e-4
$e_{1/2}$	1.2510e-4	1.5996e-4	1.6111e-3	9.9504e-5
$e_{3/4}$	6.2785e-5	3.2393e-4	8.9787e-4	4.4659e-5
e_1	3.3206e-5	5.3747e-4	1.8367e-4	8.2616e-5
time	4.3279e-2	3.8132e-2	3.5464e-2	3.6578e-2

mation of Tr(A^{-1}), we use samples of size $N = 50$ in order to ensure a better convergence to a standard normal distribution.

EXAMPLE 5.4 (A nonsymmetric matrix). We consider the *parter* matrix P of order 1000 obtained by using the *gallery* function in MATLAB. In Tables 5.9 and 5.10 we report the relative errors arising in the computation of Tr(P^{-1}), also tested in [7, Example 4]. We notice that the selected values of ν do not much influence the values of the estimates since their evaluation is based on a statistical result. The best value ν_0 of ν is computed as the mean value of the best values obtained for each sample. However, the relative error of the one-term estimate is considerably better for $\nu = 3.6$ than for $\nu = 0$, the value which was presented in [7].

EXAMPLE 5.5 (A covariance matrix). In Table 5.11 we test a covariance matrix A of order $p = 4000$, with $\alpha = 1/2$ and $\beta = 2$, also used in [18, Table 1]. The exact value of Tr(A^{-1}) is 1.1944e2 and $\kappa(A) = 6.7094$e1. The methods employed in [18] for the estimation of Tr(A^{-1}) have relative errors of order $\mathcal{O}(10^{-3})$ for different sample sizes. We notice that the one-term estimate can attain, for appropriate value of ν, a relative error of order $\mathcal{O}(10^{-5})$.

EXAMPLE 5.6 (A diagonal dominant matrix). We consider a tridiagonal matrix $S(\gamma, \delta)$. The off-diagonal elements of S are random numbers between 0 and 1, whereas its diagonal entries lie in the interval (γ, δ). This matrix is diagonal dominant for an appropriate choice of (γ, δ). We have tested the matrix S for various values of (γ, δ). In Table 5.12 we notice that, as the values of the diagonal entries increase, better approximations of Tr(S^{-1}) can be obtained.

5.5. Networks. In network analysis, it is important to extract numerical quantities that describe characteristic features of the graph of a given network. Some of these properties, such as the importance of a node, the ease of traveling from one node to another, etc., can be

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.9
Relative errors in the estimation of $\mathrm{Tr}(P^{-1}) = 2.0358\mathrm{e}2$ *for the parter matrix of order 1000 using one-term estimates.*

Estimates t_ν	Relative error
t_0	9.9366e-4
$t_{0.8}$	7.7408e-4
$t_{1.6}$	5.5436e-4
$t_{2.4}$	3.3451e-4
$t_{3.2}$	1.1451e-4
$t_{3.4}$	5.9486e-5
$t_{3.6}$	4.4556e-6

TABLE 5.10
Relative errors in the estimation of $\mathrm{Tr}(P^{-1}) = 2.0358\mathrm{e}2$ *for the parter matrix of order 1000 using two-term estimates.*

Estimates $\hat{t}_\nu$	Relative error
$\hat{t}_0$	1.5508e-3
$\hat{t}_1$	8.4604e-4
$\hat{t}_2$	7.8382e-4
$\hat{t}_3$	7.8603e-4

obtained by the diagonal elements of a matrix function f applied to the adjacency matrix A of the network.

The ease of traveling between nodes i and j with $i \neq j$ can be defined by the so-called f-subgraph communicability $(f(A))_{ij}$. Also, the importance of a node i can be defined by the f-subgraph centrality $(f(A))_{ii}$. Particularly, the most important node in a given network can be thought of as the node with the largest f-subgraph centrality [5, 13].

A matrix function which calculates subgraph centrality is the matrix resolvent

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \cdots = \sum_{k=0}^{\infty} \alpha^k A^k,$$

where $0 < \alpha < \dfrac{1}{\rho(A)}$ with $\rho(A)$ the spectral radius of A. Bounds imposed on α ensure that $I - \alpha A$ is nonsingular and that the geometric series converges to its inverse.

The resolvent Estrada index is defined as $\mathrm{Tr}((I_p - \alpha A)^{-1})$, the resolvent subgraph centrality of a node i is the diagonal element $((I_p - \alpha A)^{-1})_{ii}$, and the resolvent subgraph communicability of nodes i and j is the element $((I_p - \alpha A)^{-1})_{ij}$. In the following numerical results, we use the parameter $\alpha = 0.85/\lambda_{max}$, where λ_{max} is the largest eigenvalue of A [5, 12].

We consider the Erdös networks from the Pajek group of the *University of Florida Sparse Matrix Collection* [11]. They represent various subnetworks of the Erdös collaboration network. *Erdös 982* is a singular symmetric matrix of order $p = 5822$. However, for $A =$*Erdös 982*, the matrix resolvent $B = (I_p - \alpha A)$ is nonsingular with $\kappa(B) = 1.1300\mathrm{e}2$.

We also examine the matrix *ca-GrQc* from the Snap group of the *University of Florida Sparse Matrix Collection*, which represents a collaboration network for the arXiv General Relativity. The *ca-GrQc* matrix is a singular symmetric matrix of order $p = 5242$, whereas $B = (I_p - \alpha A)$, for $A =$*ca-GrQc*, is nonsingular with $\kappa(B) = 2.2378\mathrm{e}1$.

TABLE 5.11
Estimates for $\mathrm{Tr}(A^{-1}) = 1.1944\mathrm{e}2$ *for a covariance matrix* A *of order* $p = 4000$.

	Relative error	Estimates
e_0	2.2442e-1	9.2636e1
e_1	1.3125e-1	1.0376e2
e_2	2.6886e-2	1.1623e2
$e_{2.2}$	4.5554e-3	1.1890e2
$e_{2.24}$	2.8150e-5	1.1944e2
$\hat{e}_0$	8.8750e-2	1.0884e2

TABLE 5.12
Relative errors in estimating $\mathrm{Tr}(S^{-1})$ *for the matrix* S *of order 1000 using two-term estimates.*

	(γ, δ)				
	$(10, 20)$	$(20, 30)$	$(30, 40)$	$(40, 50)$	$(50, 60)$
$\hat{t}_0$	1.8784e-2	2.7315e-3	5.9190e-4	2.2566e-4	6.5519e-5
$\hat{t}_1$	2.5495e-2	3.7943e-3	8.9577e-4	3.6824e-4	1.1241e-4
$\hat{t}_2$	3.0673e-2	4.7538e-3	1.1833e-3	5.0573e-4	1.5829e-4
$\hat{t}_3$	3.4638e-2	5.6071e-3	1.4523e-3	6.3722e-4	2.0296e-4

Finally, we use the adjacency matrix *pref* of order $p = 4000$. This matrix represents connected simple graphs and is taken from the toolbox CONTEST of MATLAB [22]. The condition number of the matrix $B = (I_p - \alpha A)$, for $A =$*pref*, is $\kappa(B) = 1.2043\mathrm{e}1$.

We compare the estimated resolvent Estrada index of the *Erdös 982* matrix with its exact value $5.8833\mathrm{e}3$ (Table 5.13), the estimated resolvent subgraph centrality of node 5 of the *ca-GrQc* matrix with its exact value 1.0003 (Table 5.14), and the estimated resolvent subgraph communicability of nodes 1 and 17 of the *pref* matrix with its exact value $2.0418\mathrm{e}{-1}$ (Table 5.15).

In the Tables 5.13, 5.14, and 5.15 the satisfactory relative error that appeared in these computations is reported. Specifically, in Table 5.13, we notice that the one-term estimate e_ν for $\nu = 1.25$ attains a relative error of order $\mathcal{O}(10^{-5})$, whereas the relative errors of the two-term estimates are of order $\mathcal{O}(10^{-3})$. In Table 5.14, we notice that for various values of ν, we achieve a satisfactory relative error using one- or two-term estimates. In particular, the relative error of the one-term estimate e_ν for $\nu = 1$ is of order $\mathcal{O}(10^{-7})$, but the relative errors of the two-term estimates are smaller. Finally, in Table 5.15, it is also noted that the relative error of the one-term estimate e_ν for $\nu = 2$ is of order $\mathcal{O}(10^{-5})$. On the other hand, the relative errors of the two-term estimates are worse.

In Figures 5.2 and 5.3, we display the exact value and the one-term estimates of (2.3) for different values of ν estimating the resolvent subgraph centrality of the first 50 nodes of the *smallw* matrix of order 3000 obtained by the toolbox CONTEST of MATLAB [22] and of the *minnesota* matrix of order 2642 from the *University of Florida Sparse Matrix Collection* [11]. We notice that for appropriate values of ν, we obtain very satisfactory approximations.

6. Concluding remarks. In this paper, we extended the extrapolation techniques developed in [6, 7, 8, 9] and proposed families of estimates for the bilinear moment $c_{-1}(x, y) = x^T A^{-1} y$ for any nonsingular matrix $A \in \mathbb{R}^{p\times p}$. Approximations of the elements and the trace of A^{-1} were derived and implemented in several numerical experiments.

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.13
Relative errors of the resolvent Estrada index for the Erdös 982 matrix of order 5822.

	Relative error
e_0	1.0399e-2
e_1	2.1272e-3
$e_{1.25}$	4.8475e-5
$e_{1.5}$	2.0346e-3
$\hat{e}_0$	1.7295e-3
$\hat{e}_1$	2.2086e-3
$\hat{e}_2$	2.5830e-3
$\hat{e}_3$	2.8838e-3

TABLE 5.14
Relative errors of the resolvent subgraph centrality of node 5 *for the ca-GrQc matrix of order 5242.*

	Relative error
e_0	3.4721e-4
$e_{0.9}$	3.4835e-5
e_1	1.2055e-7
$e_{1.1}$	3.4595e-5
$e_{1.5}$	1.7347e-4
$\hat{e}_0$	2.2197e-16
$\hat{e}_1$	0
$\hat{e}_2$	0
$\hat{e}_3$	0

TABLE 5.15
Relative errors of the resolvent subgraph communicability of nodes 1 *and* 17 *for the pref matrix of order 4000.*

	Relative error
e_0	7.2989e-1
e_1	4.4738e-1
e_2	5.9324e-5
$\hat{e}_0$	3.2726e-1
$\hat{e}_1$	4.2900e-1
$\hat{e}_2$	4.8159e-1
$\hat{e}_3$	5.1084e-1

In case that A is a symmetric matrix, some values of the one- and two-term estimates can be interpreted as Gauss quadrature formulae, and thus they are lower bounds of c_{-1}. Nevertheless, there are differences between Gauss quadrature and extrapolation techniques. Mainly, by using the extrapolation method, families of estimates can be derived which are valid for any nonsingular matrix and thus can be used for nonsymmetric problems. In the performed numerical tests, the efficiency of the approximation using estimates derived either from Gauss quadrature or extrapolation methods is subject to the choice of the symmetric matrix.

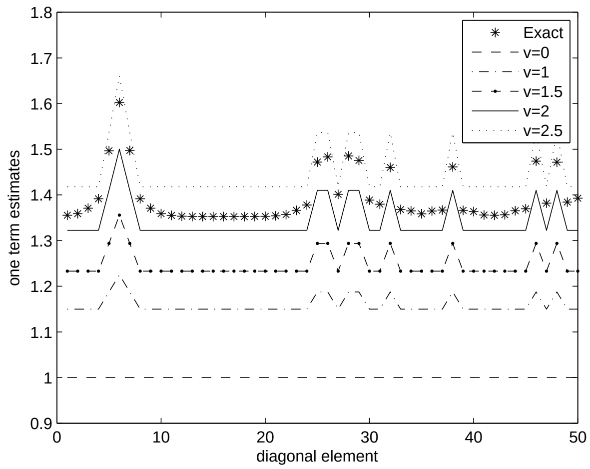

FIG. 5.2. *Estimating the resolvent subgraph centrality of the* smallw *matrix.*

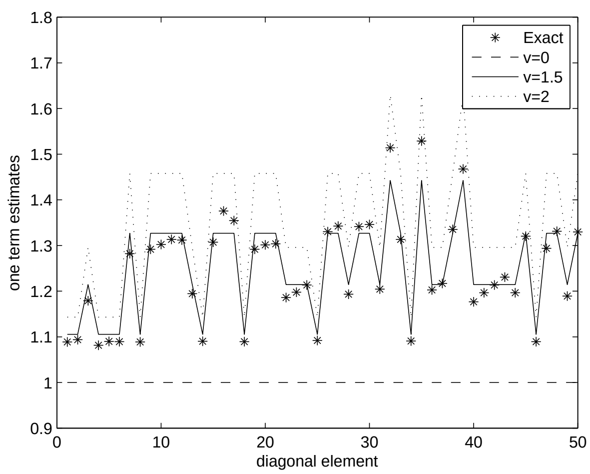

FIG. 5.3. *Estimating the resolvent subgraph centrality of the* minnesota *matrix.*

The presented numerical results show the convincing behavior of the derived estimates and indicate that they can be used in the approximation of useful quantities arising in a variety of linear algebra problems. According to the numerical tests we performed, it seems that the one- and two-term estimates are not very sensitive to perturbations of the initial matrix.

Extrapolation methods can provide a very good estimate of one matrix-vector product, but the problem is that we do not know the best value of ν a priori. The specification of this value remains an important open problem. For symmetric positive definite matrices, a range of values in which this best value lies is specified. However, a thorough study is needed to obtain sharper intervals. The estimation of $x^T f(A)y$ for an appropriate smooth function is also under consideration.

ETNA
Kent State University
http://etna.math.kent.edu

Acknowledgments. The authors would like to thank Professor Claude Brezinski for inspiring discussions and advices. The authors are grateful to the reviewers of this paper whose valuable remarks significantly improved this work. The first author (P. F.) acknowledge financial support from the Foundation for Education and European Culture (IPEP).

REFERENCES

[1] Z. BAI, M. FAHEY, AND G. GOLUB, *Some large scale computation problems*, J. Comput. Appl. Math., 74 (1996), pp. 71–89.

[2] Z. BAI AND G. H. GOLUB, *Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices*, Annals Numer. Math., 4 (1997), pp. 29–38.

[3] C. BEKAS, A. CURIONI, AND I. FEDULOVA, *Low-cost data uncertainty quantification*, Concurrency Comput.: Pract. Exper., 24 (2012), pp. 908–920.

[4] C. BEKAS, E. KOKIOPOULOU, AND Y. SAAD, *An estimator for the diagonal of a matrix*, Appl. Numer. Math., 57 (2007), pp. 1214–1229.

[5] M. BENZI AND C. KLYMKO, *Total communicability as a centrality measure*, J. Complex Networks, 1 (2013), pp. 124–149.

[6] C. BREZINSKI, *Error estimates for the solution of linear systems*, SIAM J. Sci. Comput., 21 (1999), 764–781.

[7] C. BREZINSKI, P. FIKA, AND M. MITROULI, *Moments of a linear operator on a Hilbert space, with applications to the trace of the inverse of matrices and the solution of equations*, Numer. Linear Algebra Appl., 19 (2012), pp. 937–953.

[8] ———, *Estimations of the trace of powers of positive self-adjoint operators by extrapolation of the moments*, Electron. Trans. Numer. Anal., 39 (2012), pp. 144–155.
http://etna.mcs.kent.edu/vol.39.2012/pp144-155.dir

[9] C. BREZINSKI, G. RODRIGUEZ, AND S. SEATZU, *Error estimates for linear systems with applications to regularization*, Numer. Algorithms, 49 (2008), pp. 85–104.

[10] ———, *Error estimates for the regularization of least squares problems*, Numer. Algorithms, 51 (2009), pp. 61–76.

[11] T. A. DAVIS AND Y. HU, *The University of Florida Sparse Matrix Collection*, ACM Trans. Math. Software, 38 (2011), 1 (25 pages).
Available at http://www.cise.ufl.edu/research/sparse/matrices/.

[12] E. ESTRADA AND D. J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.

[13] C. FENU, D. MARTIN, L. REICHEL, AND G. RODRIGUEZ, *Network analysis via Partial Spectral Factorization and Gauss Quadrature*, SIAM J. Sci. Comput., 35 (2013), pp. A2046–A2068.

[14] G. H. GOLUB AND G. MEURANT, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, 2010.

[15] G. H. GOLUB, M. STOLL, AND A. WATHEN, *Approximation of the scattering amplitude and linear systems*, Electron. Trans. Numer. Anal., 31 (2008), pp. 178–203.
http://etna.mcs.kent.edu/vol.31.2008/pp178-203.dir

[16] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge 1985.

[17] M. HUTCHINSON, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Comm. Statist. Simulation Comput., 18 (1989), pp. 1059–1076.

[18] V. KALANTZIS, C. BEKAS, A. CURIONI, AND E. GALLOPOULOS, *Accelerating data uncertainty quantification by solving linear systems with multiple right-hand sides*, Numer. Algorithms, 62 (2013), pp. 637–653.

[19] P. E. SAYLOR AND D. C. SMOLARSKI, *Why Gaussian quadrature in the complex plane?*, Numer. Algorithms, 26 (2001), pp. 251–280.

[20] Z. STRAKOŠ AND P. TICHÝ, *On efficient numerical approximation of the bilinear form $c^* A^{-1} b$*, SIAM J. Sci. Comput., 33 (2011), pp. 565–587.

[21] J. TANG AND Y. SAAD, *A probing method for computing the diagonal of a matrix inverse*, Numer. Linear Algebra Appl., 19 (2012), pp. 485–501.

[22] A. TAYLOR AND D. J. HIGHAM, *CONTEST: Toolbox files and documentation*. Available at www.mathstat. strath.ac.uk/research/groups/numerical_analysis/contest/toolbox