ESTIMATES OF THE TRACE OF THE INVERSE OF A SYMMETRIC MATRIX USING THE MODIFIED CHEBYSHEV ALGORITHM

GÉRARD MEURANT*

In memory of Gene H. Golub

Abstract. In this paper we study how to compute an estimate of the trace of the inverse of a symmetric matrix by using Gauss quadrature and the modified Chebyshev algorithm. As auxiliary polynomials we use the shifted Chebyshev polynomials. Since this can be too costly in computer storage for large matrices we also propose to compute the modified moments with a stochastic approach due to Hutchinson [9].

1. Introduction. In [3] Bai and Golub studied how to bound the trace of the inverse $\mathrm{tr}(A^{-1})$ and the determinant $\det(A)$ of symmetric positive definite matrices, see also [4] or [7]. The bounds of the trace of the inverse are based on the computation of the three first moments which are the traces of A^i, $i = 0, 1, 2$. Unfortunately as we will see there are examples for which these bounds are far from being sharp. In this paper we study how to improve these results by computing more moments. A lower bound of the trace of the inverse is computed by using Gauss quadrature and the modified Chebyshev algorithm [11]. This works well but can be expensive in terms of computer storage. Therefore we also describe a stochastic approach to compute the modified moments which are needed by our algorithm.

There are many applications in physics for which it is important to compute the trace of the inverse, particularly in lattice Quantum Chromodynamics (QCD).

The contents of the paper are the following. We first review the results of Bai and Golub [3] to motivate our study. Then, we give a short description of the modified Chebyshev algorithm and describe an example for which the Bai and Golub bounds are far from being sharp and the Chebyshev algorithm breaks down. Using the shifted Chebyshev polynomials as auxiliary polynomials to compute modified moments we show that tight bounds can be obtained with the modified Chebyshev algorithm. In this algorithm the traces of auxiliary matrices which are denser and denser as the algorithm proceed have to be computed. To avoid this, we propose to compute these traces using a stochastic approach due to Hutchinson [9]. Even though the numerical results are not as good as with the genuine modified moments, it is much cheaper in terms of storage.

2. The results of Bai and Golub. In [3] Bai and Golub obtained bounds using Gauss quadrature analytically. This can be done by writing the trace of the inverse as a Riemann–Stieltjes integral. More generally let

$$\mu_r = \mathrm{tr}(A^r) = \sum_{i=1}^{n} \lambda_i^r = \int_a^b \lambda^r \, d\alpha, \tag{2.1}$$

be the moments related to α, a piecewise constant measure (that we do not know explicitly) with jumps of heights 1 at the eigenvalues of A, a and b are respectively lower and upper bounds of the smallest and largest eigenvalues of A. The trace of the inverse is obtained with $r = -1$. The first three moments ($r = 0, 1, 2$) are easily

*(gerard.meurant@gmail.com)

computed

$$\mu_0 = n, \quad \mu_1 = \mathrm{tr}(A) = \sum_{i=1}^{n} a_{i,i}, \quad \mu_2 = \mathrm{tr}(A^2) = \sum_{i,j=1}^{n} a_{i,j}^2 = \|A\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm. A simple 2–point Gauss-Radau quadrature rule (see for instance [5] or [6]) for the integral in equation (2.1) is written as

$$\mu_r = \int_a^b \lambda^r \, d\alpha = \bar{\mu}_r + R_r.$$

The approximation of the integral is

$$\bar{\mu}_r = w_0 t_0^r + w_1 t_1^r, \tag{2.2}$$

where the weights w_0, w_1 and the node t_1 are to be determined. The other node t_0 is prescribed to be a or b, the ends of the integration interval. From Gauss quadrature theory (see Golub and Welsch [8] or Gautschi [5]) we know that t_0 and t_1 are the eigenvalues of a 2×2 matrix. Hence, they are the roots of the characteristic polynomial and solutions of a quadratic equation that we write as $c\xi^2 + d\xi - 1 = 0$. Because of equation (2.2), this implies that

$$c\bar{\mu}_r + d\bar{\mu}_{r-1} - \bar{\mu}_{r-2} = 0. \tag{2.3}$$

For $r = 0, 1, 2$ the quadrature rule is exact, $\bar{\mu}_r = \mu_r$ and we know that t_0 is a root of the quadratic equation. This gives two equations for c and d,

$$\begin{aligned} c\mu_2 + d\mu_1 - \mu_0 &= 0, \\ ct_0^2 + dt_0 - 1 &= 0. \end{aligned}$$

By solving this linear system we obtain the values of c and d,

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \mu_2 & \mu_1 \\ t_0^2 & t_0 \end{pmatrix}^{-1} \begin{pmatrix} \mu_0 \\ 1 \end{pmatrix}.$$

The unknown root t_1 of the quadratic equation is obtained by using the product of the roots, $t_1 = -1/(t_0 c)$. When the nodes are known, the weights are found by solving

$$\begin{aligned} w_0 t_0 + w_1 t_1 &= \mu_1, \\ w_0 t_0^2 + w_1 t_1^2 &= \mu_2. \end{aligned}$$

This gives

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} t_0 & t_1 \\ t_0^2 & t_1^2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

To bound $\mathrm{tr}(A^{-1})$ Bai and Golub used equation (2.3) with $r = 1$,

$$c\bar{\mu}_1 + d\bar{\mu}_0 - \bar{\mu}_{-1} = 0.$$

But $\bar{\mu}_0 = \mu_0$ and $\bar{\mu}_1 = \mu_1$. Hence,

$$\bar{\mu}_{-1} = (\,\mu_1 \quad \mu_0\,) \begin{pmatrix} c \\ d \end{pmatrix},$$

which gives

$$\bar{\mu}_{-1} = (\,\mu_1 \quad \mu_0\,) \begin{pmatrix} \mu_2 & \mu_1 \\ t_0^2 & t_0 \end{pmatrix}^{-1} \begin{pmatrix} \mu_0 \\ 1 \end{pmatrix}.$$

Then,

$$\mu_{-1} = \bar{\mu}_{-1} + R_{-1}(\lambda),$$

and the remainder is

$$R_{-1}(\lambda) = -\frac{1}{\eta^4} \int_a^b (\lambda - t_0)(\lambda - t_1)^2 \, d\alpha,$$

for some $a < \eta < b$. If the prescribed node is $t_0 = a$ the remainder is negative and $\bar{\mu}_{-1}$ is an upper bound of μ_{-1}. It is a lower bound if $t_0 = b$. This leads to the following result.

THEOREM 2.1 (Bai and Golub). *Let A be a symmetric positive definite matrix of order n, $\mu_1 = \mathrm{tr}(A)$, $\mu_2 = \|A\|_F^2$, the spectrum of A being contained in $[a, b]$, then*

$$(2.4)\quad (\,\mu_1 \quad n\,) \begin{pmatrix} \mu_2 & \mu_1 \\ b^2 & b \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} \le \mathrm{tr}(A^{-1}) \le (\,\mu_1 \quad n\,) \begin{pmatrix} \mu_2 & \mu_1 \\ a^2 & a \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix}.$$

3. The modified Chebyshev algorithm. The modified Chebyshev algorithm was developed by J. Wheeler in 1974 [11], from an algorithm due to P. Chebyshev in 1859. Its aim is to compute the coefficients of the three–term recurrence satisfied by orthogonal polynomials π_k using the moments associated to a given (and perhaps unknown) measure. However, the map giving the coefficients as functions of the moments is ill–conditioned (see Gautschi [5]) and this algorithm may suffer from numerical problems. To avoid these problems the modified algorithm applies the Chebyshev algorithm to modified moments instead of ordinary moments, see also Sack and Donovan [10].

The modified moments (using known orthogonal polynomials p_k) are defined as

$$(3.1)\qquad m_k = \int_a^b p_k(\lambda)\, d\alpha, \;\; k \ge 0.$$

The mixed moments related to p_l and α are

$$\sigma_{k,l} = \int_a^b \pi_k(\lambda) p_l(\lambda)\, d\alpha(\lambda), \;\; k \ge 0, l \ge 0.$$

If the three–term recurrence relation for the unknown orthogonal polynomials π_k related to the measure α is

$$(3.2)\quad \gamma_{k+1}\pi_{k+1}(\lambda) = (\lambda - \alpha_{k+1})\pi_k(\lambda) - \eta_k \pi_{k-1}(\lambda), \quad \pi_{-1}(\lambda) \equiv 0,\, \pi_0(\lambda) \equiv \pi_0,$$

and the known orthogonal polynomials satisfy

$$b_{k+1}p_{k+1}(\lambda) = (\lambda - a_{k+1})p_k(\lambda) - c_k p_{k-1}(\lambda), \quad p_{-1}(\lambda) \equiv 0,\ p_0(\lambda) \equiv p_0. \tag{3.3}$$

The modified moments are

$$m_l = \frac{\sigma_{0,l}}{\pi_0} = \int_a^b p_l(\lambda)\, d\alpha.$$

The value π_0 can be chosen arbitrarily. Then, we can obtain formulas for the coefficients α_k and η_k given the mixed moments as well as recurrences to compute the mixed moments. Assume we know $2m$ modified moments. The modified Chebyshev algorithm is the following for computing the coefficients $\alpha_k, k = 1, \dots, m$ and $\eta_k, k = 1, \dots, m-1$,

$$\sigma_{-1,l} = 0,\ l = 1, \dots, 2m-2, \quad \sigma_{0,l} = m_l \pi_0,\ l = 0, 1, \dots, 2m-1$$

$$\alpha_1 = a_1 + b_1 \frac{m_1}{m_0},$$

and for $k = 1, \dots, m-1$, we choose the normalization parameter $\gamma_k > 0$ and
for $l = k, \dots, 2m-k-1$,

$$\sigma_{k,l} = \frac{1}{\gamma_k}[b_{l+1}\sigma_{k-1,l+1} + (a_{l+1} - \alpha_k)\sigma_{k-1,l} + c_l \sigma_{k-1,l-1} - \eta_{k-1}\sigma_{k-2,l}],$$

from which the coefficients can be computed as

$$\alpha_{k+1} = a_{k+1} + b_{k+1}\frac{\sigma_{k,k+1}}{\sigma_{k,k}} - b_k \frac{\sigma_{k-1,k}}{\sigma_{k-1,k-1}},$$

$$\eta_k = b_k \frac{\sigma_{k,k}}{\sigma_{k-1,k-1}}.$$

For obtaining orthonormal polynomials the coefficients γ_k are chosen to have a norm equal to 1. If they are chosen equal to 1 we obtain monic polynomials. The Chebyshev algorithm is obtained by choosing $p_k(\lambda) = \lambda^k$.

4. Examples using the Chebyshev algorithm. The example is the matrix arising from the 5–point finite difference approximation of the Poisson equation in a unit square with an $m \times m$ mesh. This gives a matrix A of order $n = m^2$, with

$$A = \begin{pmatrix} T & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T \end{pmatrix}$$

each block being of order m and

$$T = \begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{pmatrix}.$$

We first use a small matrix with $n = 36$. The trace of the inverse is 13.7571. The Bai and Golub lower and upper bounds obtained using the first three moments are 10.2830 and 24.3776. However, if we consider a larger problem with $n = 900$ for which the trace of the inverse is 512.6442, the bounds computed from three moments are 261.0030 and 8751.76; the upper bound is a large overestimate.

As we have seen one can compute more moments $\mathrm{tr}(A^i)$, $i > 2$ and from the moments recover (with the Chebyshev algorithm) the Jacobi matrix of the coefficients α_k and η_k whose eigenvalues and eigenvectors allows to compute an approximation of the trace of the inverse, which is the moment of order -1, using the Gauss quadrature rule which gives a lower bound for the integral. Results are given for $n = 36$ in table 4.1. They seem fine after $k = 4$ which corresponds to the computation of 8 moments. However, the moment matrices (whose elements are μ_{i+j-2}) are ill–conditioned and if we continue the computations after $k = 10$ they are not positive definite anymore. Table 4.2 gives the results for $n = 900$. The ill–conditioning of the moment matrices do not allow to go further. Hence, this method is not feasible for large matrices.

TABLE 4.1
$n = 36$, *Chebyshev*, $tr(A^{-1}) = 13.7571$

k	est.
1	9.0000
2	11.3684
3	12.5714
4	13.1581
5	13.4773
6	13.6363
7	13.7139
8	13.7452
9	13.7550
10	13.7568

TABLE 4.2
$n = 900$, *Chebyshev*, $tr(A^{-1}) = 512.6442$

k	est.
1	225.0000
2	296.7033
3	344.6869
4	375.8398
5	400.0648
6	418.2138
7	433.1216
8	444.9913
9	455.0122
10	463.2337

5. Numerical examples using the modified Chebyshev algorithm. Since the moment matrices are too ill–conditioned, we can try to use modified moments to solve this problem. We consider the shifted Chebyshev polynomials of the first kind as the auxiliary orthogonal polynomials p_k. The drawback is that we need to have estimates of the smallest and largest eigenvalues of A. On the interval $[\lambda_{min}, \lambda_{max}]$

these polynomials satisfy the following three-term recurrence

$$C_0(\lambda) \equiv 1, \quad \left(\frac{\lambda_{max} - \lambda_{min}}{2}\right) C_1(\lambda) = \lambda - \left(\frac{\lambda_{max} + \lambda_{min}}{2}\right),$$

$$\left(\frac{\lambda_{max} - \lambda_{min}}{4}\right) C_{k+1}(\lambda) = \left(\lambda - \frac{\lambda_{max} + \lambda_{min}}{2}\right) C_k(\lambda) - \left(\frac{\lambda_{max} - \lambda_{min}}{4}\right) C_{k-1}(\lambda).$$

From these relations we can compute the trace of the matrices $C_i(A)$, $i = 0, \dots, 2m$ which are the modified moments. The modified Chebyshev algorithm generates the coefficients of monic polynomials corresponding to the measure related to the problem. We symmetrize this Jacobi matrix and obtain the nodes and weights of the Gauss quadrature rule from the Golub and Welsch algorithm [8]. The nodes are the eigenvalues of the Jacobi matrix and the weights are the squares of the first components of the normalized eigenvectors. The fonction to consider is $f(x) = 1/x$. Results are displayed in tables 5.1 for $n = 36$ and 5.2 for $n = 900$. Note that upper bounds can be obtained by using the Gauss–Radau quadrature rule instead of the Gauss rule. Using the modified moments there are no breakdowns in the computations and we obtain quite good results for the trace of the inverse. This example illustrates the benefits of using modified moments.

Table 5.1
$n = 36$, modified moments, $tr(A^{-1}) = 13.7571$

k	est.
1	9.0000
2	11.3684
3	12.5714
4	13.1581
5	13.4773
6	13.6363
7	13.7139
8	13.7452
9	13.7550
10	13.7568
11	13.7571

Table 5.2
$n = 900$, modified moments, $tr(A^{-1}) = 512.6442$

k	est.
5	400.0648
10	463.2560
15	489.5383
20	502.0008
25	508.0799
30	510.9301
35	512.1385
40	512.5469

6. Monte–Carlo estimates. Another possibility to compute the trace of the inverse is to consider the diagonal elements of A^{-1}. From Golub and Meurant [6] we know how to estimate $(e^i)^T A^{-1} e^i$. However, this approach requires computing n such estimates. This might be too costly if n is large. In Bai, Fahey, Golub, Menon and Richter [2] it was proposed to use a Monte Carlo technique based on the following proposition, see Hutchinson [9] and also Bai, Fahey and Golub [1].

PROPOSITION 6.1. *Let B be a symmetric matrix of order n with $\mathrm{tr}(B) \neq 0$. Let $\mathcal{Z}$ be a discrete random variable with values 1 and -1 with equal probability 0.5 and let z be a vector of n independent samples from $\mathcal{Z}$. Then $z^T B z$ is an unbiased estimator of $\mathrm{tr}(B)$,*

$$E(z^T B z) = \mathrm{tr}(B),$$

$$\mathrm{var}(z^T B z) = 2 \sum_{i \neq j} b_{i,j}^2,$$

where $E(\cdot)$ denotes the expected value and var *denotes the variance.*

The method proposed in [2] is to first generate p sample vectors z^k, $k = 1, \ldots, p \ll n$ and then to estimate $(z^k)^T A^{-1} z^k$ by running the Lanczos algorithm, see [6]. This gives p estimates σ_k from which an unbiased estimate of the trace is derived as

$$\mathrm{tr}(A^{-1}) \approx \frac{1}{p} \sum_{k=1}^{p} \sigma_k.$$

If we have p lower bounds σ_k^L and p upper bounds σ_k^U, we obtain lower and upper bounds by computing the means

$$\frac{1}{p} \sum_{k=1}^{p} \sigma_k^L \leq \frac{1}{p} \sum_{k=1} (z^k)^T A^{-1} z^k \leq \frac{1}{p} \sum_{k=1}^{p} \sigma_k^U.$$

The quality of such an estimation was assessed in Bai, Fahey and Golub [1].

Table 6.1 gives the results for our example with $n = 36$ and doing 5 iterations of the Lanczos algorithm to compute the bounds of $(z^k)^T A^{-1} z^k$ using the Gauss and Gauss–Radau quadrature rules. Results for $=$ 900 with 30 iterations are given in table 6.2. The results are good even though we do not always obtain lower and upper bounds, but not as good as with the modified Chebyshev algorithm.

TABLE 6.1
$n = 36$, *Monte Carlo, 5 it.,* $tr(A^{-1}) = 13.7571$

p	G	G–R b_L	G–R b_U
1	12.8274	12.8749	12.9087
2	14.7464	14.8440	14.9300
3	14.8973	14.9681	15.0277
4	13.6203	13.6777	13.7226
5	13.9216	13.9918	14.0495

7. Mixing the modified Chebyshev algorithm and the Monte–Carlo estimates. Since the matrices that have to be computed when using the modified moments in the modified Chebyshev algorithm are denser and denser when k increases, it can be costly to compute and to store them for large matrices. Therefore it is tempting to combine the modified Chebyshev algorithm and the Monte Carlo estimates of the trace of a matrix to compute approximate modified moments. Instead

Table 6.2
$n = 900$, Monte Carlo, 30 it., tr$(A^{-1}) = 512.6442$

p	G	G–R b_L	G–R b_U
1	478.1734	478.3272	478.4955
2	466.4618	466.5600	466.6667
3	458.1058	458.1850	458.2703
4	466.5929	466.6975	466.8028
5	511.1780	511.2772	511.3732

of computing the matrices $C_i(A)$ and their traces, we can choose p random vectors z^j, $j = 1, \ldots, p$, compute $C_i(A)z^j$ by three–term vector recurrences and obtain a Monte–Carlo estimate of the trace of $C_i(A)$ by averaging the values $(z^j)^T C_i(A) z^j$. The results for $n = 36$ and $k = 5$ are given in table 7.1 and for $n = 900$ and $k = 30$ in table 7.2. Of course, the results are not as good as when using the exact traces of the matrices $C_i(A)$. They are of the same order of accuracy as those obtained with the Monte Carlo method on A^{-1}. The best result is given by $p = 5$. The results for $p = 5$ and increasing values of k are displayed in table 7.3. We obtain good bounds when k is large. The advantages of this algorithm is that we do not need to store the matrices C_k and we do not have to run Lanczos iterations as when estimating $(z^k)^T A^{-1} z^k$. However, it can be difficult to find a good value of p. It is better to increase k rather than p.

Table 7.1
$n = 36$, $k = 5$, Modified moments + Monte Carlo, tr$(A^{-1}) = 13.7571$

p	G
1	12.8274
2	14.7289
3	14.8535
4	13.5780
5	13.8215
6	14.1153
7	14.1134
8	14.5652
9	14.9474
10	14.7008

Table 7.2
$n = 900$, $k = 30$, Modified moments + Monte Carlo, tr$(A^{-1}) = 512.6442$

p	G
1	478.1734
2	466.3618
3	457.9893
4	466.4000
5	510.8105
6	499.5213
7	494.8707
8	488.1057
9	530.4409
10	526.7022

8. Conclusions. In this paper we have shown that the trace of the inverse of a symmetric matrix can be estimated using Gauss quadrature and the modified

TABLE 7.3
$n = 900$, $p = 5$, Modified moments + Monte Carlo, $tr(A^{-1}) = 512.6442$

k	G
5	396.4725
10	449.9707
15	480.7157
20	499.4022
25	508.0922
30	510.8105
35	511.7008
40	511.9745
45	512.0306
50	512.0414

Chebyshev algorithm. Stochastic estimates allows to compute the modified moments without using too much storage. Choosing the number of points in the quadrature rule and the number of random vectors is still an open problem. Estimates for the determinant of A can be obtained using the techniques developed in this paper and the identity $\ln(\det(A)) = \mathrm{tr}(\ln(A))$.

REFERENCES

[1] Z. Bai, M. Fahey and G.H. Golub, *Some large scale matrix computation problems*, J. Comp. Appl. Math., v 74, (1996), pp 71–89.

[2] Z. Bai, M. Fahey, G.H. Golub, M. Menon and E. Richter, *Computing partial eigenvalue sum in electronic structure calculation*, Report SCCM 98-03 Stanford University, (1998).

[3] Z. Bai and G.H. Golub, *Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices*, Annals Numer. Math., v 4, (1997), pp 29–38.

[4] Z. Bai and G.H. Golub, *Some unusual matrix eigenvalue problems*, in Proceedings of Vecpar'98 - Third International Conference for Vector and Parallel Processing, J. Palma, J. Dongarra and V. Hernandez Eds., Springer, (1999), pp 4–19.

[5] W. Gautschi, *Orthogonal polynomials: computation and approximation*, Oxford University Press, (2004).

[6] G.H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, D.F. Griffiths and G.A. Watson eds., Pitman Research Notes in Mathematics, v 303, (1994), pp 105–156.

[7] G.H. Golub and G. Meurant, *Matrices, moments and quadrature with applications*, to be published by Princeton University Press, (2009).

[8] G.H. Golub and J.H. Welsch, *Calculation of Gauss quadrature rules*, Math. Comp., v 23, (1969), pp 221–230.

[9] M. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Comm. Statist. Simul., v 18, (1989), pp 1059–1076.

[10] R.A. Sack and A. Donovan, *An algorithm for Gaussian quadrature given modified moments*, Numer. Math., v 18 n 5, (1972), pp 465–478.

[11] J.C. Wheeler, *Modified moments and Gaussian quadrature*, in Proceedings of the international conference on Padé approximants, continued fractions and related topics, Univ. Colorado, Boulder, Rocky Mtn. J. Math., v 4 n 2, (1974), pp 287–296.