Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A general scheme for log-determinant computation of matrices via stochastic polynomial approximation

Wei Peng, Hongxia Wang *

Department Mathematics and System Science, National University of Defense Technology, Changsha, Hunan, 410073, China

ARTICLE INFO

Article history:
Received 31 March 2016
Received in revised form 21 July 2017
Accepted 1 November 2017
Available online xxxx

Keywords:
Log-determinant
Orthogonal polynomial expansions
Stochastic algorithm
Counting spanning trees

ABSTRACT

We study the approximation of determinant for large scale matrices with low computational complexity. This paper develops a generalized stochastic polynomial approximation frame as well as a stochastic Legendre approximation algorithm to calculate log-determinants of large-scale positive definite matrices based on the prior eigenvalue distributions. The generalized frame is implemented by weighted L_2 orthogonal polynomial expansions with an efficient recursion formula and matrix–vector multiplications. So the proposed scheme is efficient both in computational complexity and data storage. Respective error bounds are given in theory which guarantee the convergence of the proposed algorithms. We illustrate the effectiveness of our method by numerical experiments on both synthetic matrices and counting spanning trees.

1. Introduction

The problem of large-scale log-determinant computation is widely studied, which arises from machine learning and data mining applications such as Gaussian process [1,2], sparse inverse covariance estimation [3,4] and counting spanning trees in large-scale networks [5,6]. In this paper, we consider the stochastic estimation of log-determinant of a positive symmetric matrix $\Sigma \in \mathbb{R}^{n\times n}$.

A typical method to solve the problem is Cholesky decomposition. But it is not always affordable for large scale log-determinant computation. First, for Σ is often sparse, which could be stored efficiently, the operation of Cholesky decomposition tends to fill in the zero entries, devastating the sparsity and costing storage even with reordering techniques such like AMD [7]. Second, Cholesky decomposition costs $\mathcal{O}(n^3)$ flops in dense cases, which is expensive when n is quite large.

Previous studies have developed several randomized methods to approximate log-determinant of a symmetric positive matrix. In general, existing stochastic methods for the problem are based on the trace estimation via Monte-Carlo methods [8]. Barry et al. [9] and Boutsidis et al. [10] use randomized Taylor expansions in a variety of settings. However, from the viewpoint of polynomial approximation, the Taylor series is hardly to be the optimal. Later, the stochastic Chebyshev expansion method [11] is also presented, which accelerates the Taylor approximation significantly. However, this method requires the minimal and maximal eigenvalue of a matrix to dilate the objective function in proofs and algorithms, which costs extra computation. In recent, the log-determinant algorithm is extended to a general case of estimating the trace of $f(A)$ via the similar frame [12].

Besides polynomial approximations, the subspace iteration based estimator for *Trace*(A) or $\log\det(I+A)$ is introduced by Saibaba et al. [13], where the Hermitian positive definite matrix A has k dominant eigenvalues. The essence of the

* Corresponding author.
E-mail addresses: weipeng0098@126.com (W. Peng), wanghongxia@nudt.edu.cn (H. Wang).

https://doi.org/10.1016/j.camwa.2017.11.001
0898-1221/© 2017 Elsevier Ltd. All rights reserved.

subspace method is to maintain the larger eigenvalues while drop the smaller ones to reduce the scale of the objective matrix. Essentially, this method consider the prior that the matrix A has k dominant eigenvalues.

As a popular viewpoint in probabilistic numerics is to regard epistemic uncertainty in numerical objects as random variables, a prior knowledge about the uncertainty might enhance existing deterministic methods. In this paper, we extend previous research to a general case based on the eigenvalue distribution prior. First, a generalized polynomial approximating frame for log-determinant computation is proposed in the weighted L_2 space. A stochastic orthogonal polynomial expansion approximating method is given as well as a deterministic error bound, only an upper bound of eigenvalues is required which could be obtained by the Gershgorin circle theorem or the power method. The proofs of the deterministic part are standard adaptations of known techniques in Aune et al. [14], Boutsidis et al. [10] and Han et al. [11]. We also notice that a practical error bound could hardly be given without the prior distribution of eigenvalues due to the singularity of $\log x$ at 0. Thus a probabilistic error bound is presented through eigenvalues regarded as random variables. Through theoretical and numerical results, we note the similarity between the weight function and the eigenvalue distribution might be related to the convergence rate. Therefore, orthogonal polynomials with certain weight functions could be selected to accelerate the convergence based on the prior eigenvalue distribution. Numerical experiments involving graph theory are also conducted to demonstrate the strategy.

During the review period of the manuscript, we noticed the very recent SLQ method proposed by Ubaru et al. [15]. The SLQ algorithm is primarily based on the theory of Lanczos Quadrature [16], which is applied to estimate $\mathbf{v}^T f(A)\mathbf{v}$ through Gaussian quadrature with an implicit weight function to approximate the eigenvalue distribution instead of a given prior. The numerical comparison with SLQ is given in the experiment part. Another appealing recent work is using Bayesian quadrature proposed by Fitzsimons et al. [17], which updates the eigenvalue distribution via posterior estimation.

The paper is organized as follows. Weighted L_2 space and orthogonal polynomial approximation are introduced in Section 2 as preliminaries. Section 3 proposes a general algorithm including the Legendre approximation as a special case. Deterministic and probabilistic bounds are derived in Section 4. In Section 5, we conduct several numerical experiments and Section 6 concludes the paper.

2. $L_2(\omega)$ and orthogonal polynomial approximation

In this section, the weighted L_2 space $L_2(\omega)$ is introduced. The best orthogonal polynomial approximation respect to the weight function $\omega(x)$ is then given as preparation for our general frame.

Let the weight function $\omega : (-1, 1) \to \mathbb{R}$ be continuous and positive almost everywhere, satisfying $\int_{-1}^{1} \omega(x)dx = 1$ and $0 < \int_{-1}^{1} x^{2n}\omega(x)dx < +\infty$, $n \geq 0$. The weighted L_2 space with a weight function $\omega(x)$ is defined as

$$L_2(\omega) = \left\{ f \,|\, \langle f, f\rangle_\omega = \int_{-1}^{1} |f(x)|^2 \omega(x)dx < +\infty \right\},$$

where the inner product $\langle f, g\rangle_\omega = \int_{-1}^{1} f(x)g(x)\omega(x)dx$ which induces a norm $\|f\|_\omega = \left(\int_{-1}^{1} f(x)^2\omega(x)dx\right)^{\frac{1}{2}}$.

A family of orthogonal polynomials $\{Q_i(x) | i = 0, 1, 2, \ldots, N\}$ respect to $\omega(x)$ could be obtained by

$$Q_{i+1}(x) = \frac{x - \langle xQ_i, Q_i\rangle_\omega}{k_i} Q_i(x) - \frac{\langle xQ_i, Q_{i-1}\rangle_\omega}{k_i} Q_{i-1}(x), i = 1, 2, \ldots \quad (1)$$

where $Q_0(x)$ and $Q_1(x)$ are given satisfying $\langle Q_0, Q_1\rangle_\omega = 0$, $k_i \in \mathbb{R}$ is chosen as a positive real number so that $\|Q_{i+1}(x)\|_\omega = 1$. It follows $deg(Q_i(x)) = i$ and $\langle Q_i, Q_j\rangle_\omega = \delta_{ij}$ immediately. Denote $S_N = span\{1, x, x^2, \ldots, x^N\}$, which is the collection of all polynomials with degrees no more than N. Then we have $S_N = span\{Q_i(x) | i = 0, 1, 2, \ldots, N\}$.

Assume $p_{0,\omega}(x) = 1$ and $p_{1,\omega}(x) = q_1x - q_0$ satisfying $\langle p_{0,\omega}, p_{1,\omega}\rangle_\omega = 0$ with given $\omega(x)$. Then by (1),

$$p_{i+1,\omega}(x) = \frac{x - b_i}{a_i} p_{i,\omega}(x) - d_i p_{i-1,\omega}(x), i = 1, 2, \ldots, N.$$

Particularly, let $p_{1,\omega}(x)$ be the second term of Legendre polynomials, Chebyshev polynomials of the first kind (Chebyshev I) and Chebyshev polynomials of the second kind (Chebyshev II) respectively, and some specified $\omega(x)$ is chosen accordingly, we have

$$(q_1, q_0, a_i, b_i, d_i) = \begin{cases} (1, 0, \frac{i+1}{2i+1}, 0, \frac{i}{i+1}) & , \ \omega(x) = \frac{1}{2} \text{ (Legendre)} \\ (1, 0, \frac{1}{2}, 0, 1) & , \ \omega(x) = \frac{1}{\sqrt{1-x^2}} \text{(Chebyshev I)} \\ (2, 0, \frac{1}{2}, 0, 1) & , \ \omega(x) = \sqrt{1-x^2} \text{(Chebyshev II).} \end{cases} \quad (2)$$