

Cholesky分解的单Pass随机算法在秩亏最小二乘中的应用

刘雪翠

上海理工大学理学院, 上海

收稿日期: 2024年4月7日; 录用日期: 2024年6月19日; 发布日期: 2024年6月27日

摘要

通常针对大规模的对称半正定矩阵求解最小二乘问题时, 往往得不到精确解, 而传统的矩阵分解方法在规模增大的情况下大大地增加了求解时间和复杂度, Cholesky分解的单pass随机算法只需对输入矩阵访问一次, 因此本文将Cholesky分解的单pass随机算法与两种常见的求解最小二乘问题的方法结合, 提出了两种新的求解最小二乘问题的算法, 并将两者进行了比较, 最后通过数值实验证明了这两种算法的有效性及可行性。

关键词

Cholesky分解, 矩阵分解, 单Pass随机算法, 最小二乘

The Application of Single-Pass Randomized Algorithm for Cholesky Decomposition in Rank Deficit Least Squares

Xuecui Liu

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 7th, 2024; accepted: Jun. 19th, 2024; published: Jun. 27th, 2024

Abstract

Usually for large-scale symmetric semi-positive definite matrices to solve the least squares problem, often do not get the exact solution, and the traditional matrix factorization method greatly increases the solution time and complexity in the case of increasing scale, the Cholesky decompo-

sition of the single-pass randomized algorithm only needs to access the input matrix once, so this paper combines the single-pass random algorithm of Cholesky decomposition with two common methods to solve the least squares problem, proposes two new algorithms for solving the least squares problem, compares the two, and finally proves the effectiveness and feasibility of the two algorithms through numerical experiments.

Keywords

Cholesky Decomposition, Matrix Factorization, Single-Pass Randomized Algorithm, Least Squares

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 问题的提出

随着计算机技术的不断发展,最小二乘问题的背后是一种优化方法,旨在找到一条直线(或更一般的一个模型),以最小化观测数据点与该直线(或模型)之间的垂直距离的平方和[1]。在实际应用中,最小二乘问题通常涉及到大量数据的分析和模型拟合,以便从观测数据中提取出有用的信息并进行预测[2]。其背后的数学原理和算法在现代科学和工程中扮演着重要的角色[3]。

近年来,随机算法的不断完善,越来越多的人将随机算法应用到最小二乘问题的求解上面,并取得了不错的结果,在文献[4]中,Christos 等人将随机投影的思想运用到求解非负最小二乘问题上去,在文献[5]中,Avron 等人提出了 Blendenpik 算法,是一种基于随机算法求解最小二乘问题的算法。在文献[6]中,Drineas 和 Mahoney 等人提出了两种随机算法求解最小二乘问题,一种是基于随机采样的思想,另一种是基于随机投影的思想。紧接着在文献[7]中,董银丽提出了一种基于随机采样方法求解最小二乘问题的新算法。

针对大规模的对称非正定矩阵分解来求解最小二乘问题,传统的矩阵分解算法至少需要读取输入矩阵两次。而面对存储在内存外部的大规模数据和流型数据时,往往只有一次读取数据的机会。在文献[8]中,提出的 Cholesky 分解的单 pass 随机算法是从随机矩阵投影的角度构造的一种新算法。本文尝试利用这种算法去求解对称半正定矩阵的最小二乘问题。根据求解最小二乘的方法,我们在本文第三章提出了两种新的算法用来求解秩亏最小二乘问题。

2. Cholesky 分解的单 Pass 随机算法

任意给定一个对称半正定矩阵 $A \in R^{n \times n}$, 则我们可以对 A 进行 Cholesky 分解,将 A 表示成一个下三角矩阵 L 与 L 的转置的乘积,即 $A = LL^T$ 。在大数据分析和机器学习中,Cholesky 分解已经成为一种关键的分析工具。随着大数据时代的到来,传统的 LU 和 QR 算法在面对超大规模的对称半正定矩阵分解时,高内存消耗和高计算复杂度已无法满足时代需求,而单 pass 随机 Cholesky 分解只需对输入矩阵访问一次,而且有很高的误差精度,具体内容见算法 1。

3. 用 Cholesky 分解的单 Pass 随机算法求解秩亏最小二乘问题

设矩阵 A 是一个 $n \times n$ 阶矩阵,且 b 是一个 n 元列向量。秩亏最小二乘问题为:求解向量 $x_{n \times 1}$,使得 $\|Ax - b\|$ 达到最小。在本章中我们提出了两种求解最小二乘的方法,具体思想如下。

算法 1. Cholesky 分解的单 pass 随机算法(SPRCH)

输入：半正定矩阵 $A \in \mathbb{R}^{n \times n}$ ，目标值 r ，过采样参数 $p \geq 2$ ， $l = r + p$ ；
 输出：置换矩阵 $P \in \mathbb{R}^{n \times n}$ ，下三角矩阵 $L \in \mathbb{R}^{n \times (r+p)}$ ，使得 $\|PAP^T - LL^T\| \leq \varepsilon$ ， ε 是误差界。
 生成一个 $n \times (r+p)$ 阶的高斯采样矩阵 Ω_1 ；
 计算 $Y_1 = A\Omega_1$ ， $Y_2 = \Omega_1^T Y_1$ ；
 对 Y_1 进行列主元 LU 分解，得到 $PY_1 = L_y U_y$ ；
 令 $B = (\Omega_1^T P^T L_y)^{\dagger} Y_2 (L_y^T P \Omega_1)^{\dagger}$ ；
 对 B 作 Cholesky 分解，得 $B = L_b L_b^T$ ；
 令 $L = L_y L_b$ ；
 返回 L, P 。

第一种：在算法 1 中，我们通过 Cholesky 分解的随机算法得到 $PAP^T \approx LL^T$ ，其中 L 是一个 $n \times l$ 阶列满秩矩阵， U 是一个 $l \times n$ 阶矩阵。所以秩亏最小二乘问题可以转化为：

$$\text{最小化 } \|Ax - b\| \approx \|P^T LL^T Px - b\| = \|LL^T Px - Pb\|。$$

我们可以令 $y = L^T Px$ ， $c = Pb$ 。又因为矩阵 L 列满秩，所以 $y = L^{\dagger} c$ 。接下来再令 $z = Px$ ，矩阵 L^T 不是列满秩矩阵，所以对 L^T 与 z 进行分块：

$$L^T = (L_1, L_2), \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

其中 L_1 是一个 $l \times l$ 阶矩阵， L_2 是一个 $l \times (n-l)$ 阶矩阵。 z_1 是一个 l 元列向量， z_2 是一个 $(n-l)$ 元列向量。由此可知这个方程有无穷个解，为了简便起见，令 $z_2 = 0$ ， $z_1 = L_1^{-1} y$ 。最后通过 $x = P^T z$ 可求出最小二乘解 x 。

第二种：秩亏最小二乘问题还可以转化为：

最小化 $\|Ax - b\| \approx \|P^T LL^T Px - b\| = \|BB^T x - b\|$ ，我们可以令 $B = P^T L$ ， $y = B^T x$ ， $c = Pb$ 。因为矩阵 L 列满秩则 $B = P^T L$ 仍为列满秩矩阵，所以 $y = B^{\dagger} b$ 。接下来直接求解三角方程组 $y = B^T x$ 求出最小二乘解 x 。以上两种算法具体如算法 2 和算法 3：

算法 2. 第一种方法求解秩亏最小二乘问题(SPRCH-LS1)

输入：分解矩阵 $A \in \mathbb{R}^{n \times n}$ ，目标秩 $r < n$ ，随机矩阵的列数 $l = r + p$ ，其中 p 是过采样参数， $b \in \mathbb{R}^{n \times 1}$ ；
 输出：正交置换矩阵 $P \in \mathbb{R}^{n \times n}$ ， $x \in \mathbb{R}^{n \times 1}$ ，使得 $\|Ax - b\| \leq \varepsilon$ ， ε 表示一个比较小的误差界；
 1) 运用算法 1 对 A 做分解得到 L, P ；
 2) 计算 $c = Pb$ ；
 3) 计算 $y = L^{\dagger} c$ ；
 4) 计算 L^T ，并对 L^T 作截断得到 L_1 ；
 5) 计算 $z_1 = L_1^{-1} y$ ，令 $z_2 = 0$ ；
 6) 计算 $x = P^T z$ ；
 7) 返回 x 。

算法 3. 第二种方法求解秩亏最小二乘问题(SPRCH-LS2)

输入：分解矩阵 $A \in \mathbb{R}^{n \times n}$ ，目标秩 $r < n$ ，随机矩阵的列数 $l = r + p$ ，其中 p 是过采样参数， $b \in \mathbb{R}^{n \times 1}$ ；

输出：正交置换矩阵 $P \in \mathbb{R}^{n \times n}$ ， $x \in \mathbb{R}^{n \times 1}$ ，使得 $\|Ax - b\| \leq \varepsilon$ ， ε 表示一个比较小的误差界；

- 1) 运用算法 1 对 A 做分解得到 L, P ;
- 2) 计算 $B = P^T L$;
- 3) 计算 $y = B^+ b$;
- 4) 计算 $x = B^T \setminus y$
- 5) 返回 x 。

4. 数值实验**4.1. 对奇异值慢速衰减的固定输入矩阵**

本节所处理的矩阵是随机生成的奇异值慢速衰减的半正定矩阵。生成方式如下：

- 1) 生成 $\Sigma = \text{diag}(1, \dots, 1, e^{-s}, e^{-2s}, \dots, e^{-(n-r)s})$ ，这里的 r 和 s 都是大于 0 的实数， r 用来控制输入矩阵的秩， s 用来控制衰减速度，令 $s = 190/(n-r-1)$ ，让输入矩阵 A 的奇异值从 1 到 e^{-200} 呈指数型慢速下降；
- 2) 生成一个 $n \times n$ 的随机正交矩阵 U ；
- 3) 令待分解矩阵 $A = U \Sigma U^T$ 。

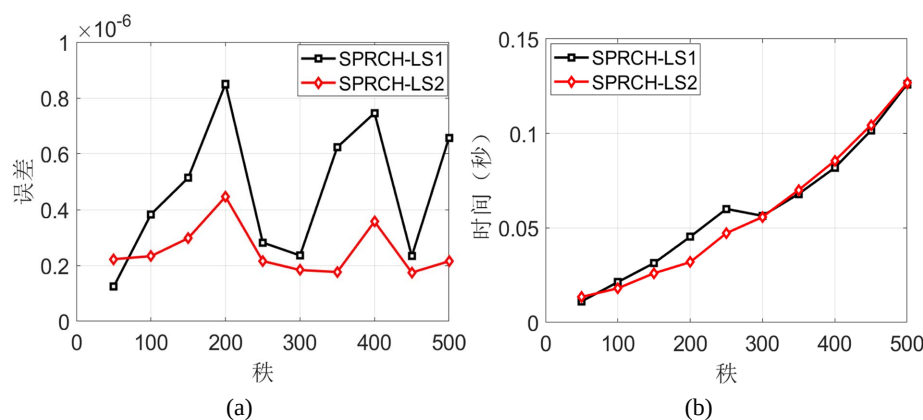


Figure 1. Error and time comparison of Algorithm 2 and Algorithm 3 under a 2000×2000 dimensional matrix with slowly decaying singular values. (a) Error comparison, (b) Time comparison

图 1. 算法 2 和算法 3 在 2000×2000 维且奇异值缓慢衰减的矩阵下的误差和时间比较。
(a) 误差比较, (b) 时间比较

从图 1 中我们可以看出，随着秩的增加，这两种算法的运行时间都在增大，但是误差的变换趋势相同，误差精度都保持在 10^{-6} 左右。

4.2. 对奇异值快速衰减的固定输入矩阵

与 4.1 节类似，这里我们令 $s = 900/(n-r-1)$ ，让输入矩阵 A 的奇异值从 1 到 e^{-1000} 呈指数型快速下降。

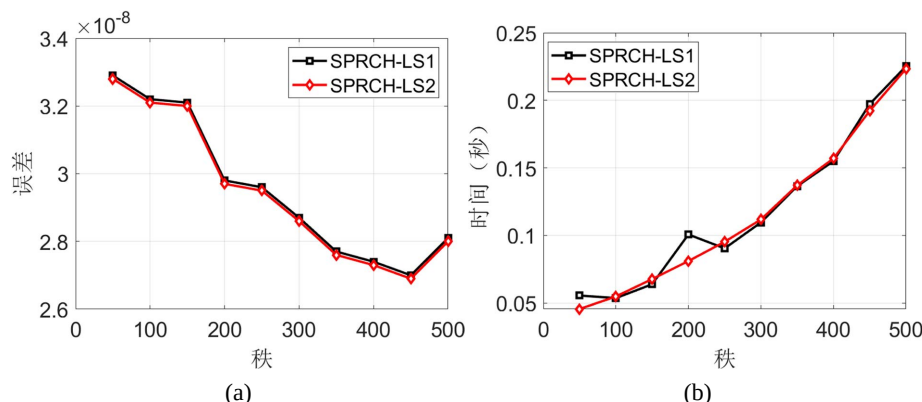


Figure 2. Comparison of error and time between Algorithm 2 and Algorithm 3 under a 2000×2000 dimensional matrix with fast decay of singular values. (a) Error comparison, (b) Time comparison

图 2. 算法 2 和算法 3 在 2000×2000 维且奇异值快速衰减的矩阵下的误差和时间比较。(a) 误差比较, (b) 时间比较

从图 2 可以看出在奇异值快速衰减的条件下, 随着秩的增加, 这两种算法的误差和运行时间基本相似。

5. 总结与展望

5.1. 全文总结

本文所提出的这两种新算法在对处理大规模对称半正定矩阵的最小二乘问题具有很大的优势。这两种算法的误差精度和运行时间成本都很好, 其次我们发现当输入矩阵的奇异值慢速衰减时, 算法 2 略有优势, 而当输入矩阵的奇异值快速衰减时这两种算法的处理结果很接近。

5.2. 本文展望

本文主要提出了两种关于 Cholesky 分解的单 pass 随机算法在秩亏最小二乘中应用的新算法, 这两种新算法为求解大规模对称半正定矩阵的最小二乘问题提供了合适的方案, 但仍存在不足, 比如除了以上两种求解最小二乘的思想, 是否还能将 Cholesky 分解的单 pass 随机算法在运用到其他的求解二乘方法中, 希望在以后的研究中可以给出更多的结果。

参考文献

- [1] Golub, G.H. and Van Loan, C.F. (2012) Matrix Computations. 4th Edition, Johns Hopkins University Press.
- [2] Strang, G. (2006) Introduction to Linear Algebra. 4th Edition, Wellesley-Cambridge Press.
- [3] Luenberger, D.G. and Ye, Y. (2008) Linear and Nonlinear Programming. 3rd Edition, Springer. <https://doi.org/10.1007/978-0-387-74503-9>
- [4] Christos, B. and Petros, D. (2009) Random Projections for the Nonnegative Least-Squares Problem. *Linear Algebra and Its Applications*, **431**, 760-771. <https://doi.org/10.1016/j.laa.2009.03.026>
- [5] Avron, H., Maymounkov, P. and Toledo, S. (2010) Blendenpik: Supercharging LAPACK's Least-Squares Solver. *SIAM Journal on Scientific Computing*, **32**, 1217-1236. <https://doi.org/10.1137/090767911>
- [6] Petros, D., Michael, W., Mahoney, S. and Sarlós, T. (2011) Faster Least Squares Approximation. *Numerische Mathematik*, **117**, 217-249. <https://doi.org/10.1007/s00211-010-0331-6>
- [7] 董银丽, 李鹏程. 新的求解大规模线性最小二乘问题的随机算法[J]. 宝鸡文理学院学报: 自然科学版, 2014, 34(2): 15-19.
- [8] 刘雪翠. Cholesky 分解的单 Pass 随机算法[J]. 理论数学, 2024, 14(1): 87-93. <https://doi.org/10.12677/pm.2024.141010>