Menemui Matematik (Discovering Mathematics) 46(3) (2024) 13-19

Menemui Matematik
(Discovering Mathematics)

journal homepage: https://myjms.mohe.gov.my/index.php/dismath/

Quasi-Newton Method for Sparse Matrix Factorization with Frobenius norm Regularization

Muhammad Yusril Firdaus Faris[1], Fong Peng Lim[2], Nor Aliza Ab Rahmin[3], Wah June Leong[4*]

[1, 2, 3, 4]*Department of Mathematics and Statistics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor.*

[1]208322@student.upm.edu.my, [2]fongpeng@upm.edu.my, [3]aliza@upm.edu.my, [4]leongwj@upm.edu.my
*Corresponding author

Received: 20 September 2024
Accepted: 24 November 2024

ABSTRACT

This paper considers quasi-Newton method for sparse matrix factorization (SMF) that incorporates Frobenius norm regularization to control overfitting and enhance generalization in data-driven applications. Sparse matrix factorization seeks to decompose a matrix into 2 matrices while promoting sparsity in one (or both) of the resulting factors, which is particularly useful in applications such as recommendation systems, signal processing, and dimensionality reduction. The Frobenius norm regularization is employed to penalize large parameter values, ensuring sparser factorization. The proposed quasi-Newton method, leveraging approximate second-order information, efficiently optimizes the objective function with significantly reduced computational overhead compared to full Newton methods. Experimental results on an example demonstrate the efficacy of the method in achieving high-quality sparse factorizations under different regularization parameter.

Keywords: Sparse matrix factorization, quasi-Newton method, Frobenius norm regularization, BFGS update.

INTRODUCTION

Matrix factorization techniques have become fundamental tools in numerous fields, including machine learning (Sim et al., 2022), data mining (Al-Hakeem et al. (2023), Sun et al. (2023)), signal processing, bioinformatics (Woo et al., 2023) and disease modelling (Kon and Labadin, 2019). By decomposing a large data matrix into smaller, more interpretable components, matrix factorization methods provide a compact representation of data, often revealing underlying patterns and relationships. However, when working with high-dimensional data, such as in recommendation systems, image processing, and text mining, achieving both accuracy and interpretability is critical. To this end, *Sparse Matrix Factorization* (SMF) has emerged as a popular technique, where sparsity constraints are introduced to yield factors with a significant proportion of zero entries. Sparse factors not only improve interpretability but also reduce computational complexity, making SMF particularly valuable for handling large-scale datasets.

Despite its advantages, SMF poses unique challenges in terms of optimization. The introduction of sparsity constraints makes the optimization problem more complex, often requiring specialized algorithms that balance the need for sparse solutions with accurate matrix approximations. Traditional optimization methods, such as gradient descent, can struggle to handle this trade-off efficiently, leading to slow convergence and suboptimal solutions, especially in large-scale applications. Consequently, there is a growing interest in developing more sophisticated optimization techniques to address these challenges.

In this paper, we propose to employ quasi-Newton method SMF problem to tackle the optimization difficulties inherent in sparse factorization problems. Quasi-Newton methods are well-known for their ability to approximate second-order derivative information without the computational expense of full Newton methods. This makes them particularly well-suited for large-scale optimization problems like SMF.

The remainder of this paper is organized as follows: Section 2 details the formulation of our sparse matrix factorization model and the quasi-Newton method. Section 3 presents experimental results, demonstrating the performance of our method. Finally, Section 4 concludes the paper with a discussion of future research directions.

SPARSE MATRIX FACTORIZATION AND QUASI-NEWTON METHODS

Sparse matrix factorization aims to decompose a given matrix $X \in R^{m\times n}$ into the product of two matrices $U \in R^{m\times p}$ and $V \in R^{p\times n}$, such that $X \approx UV$. A critical goal of this factorization is to enforce sparsity in the resulting matrices U and/or V, meaning that many elements of these matrices should be zero, making the factorization easier to interpret and reducing the computational cost for large-scale data. To achieve this, sparsity constraints are typically introduced through regularization techniques.

The objective function for sparse matrix factorization can be formulated using the Frobenius norm, which measures the reconstruction error between the original matrix X and the factorized product UV. The associated optimization problem is defined as:

$$\min_{U,V} \|X - UV\|_F^2,$$

where $\|A\|_F = \sqrt{trace(AA^T)}$ denotes the Frobenius norm of a matrix A. This formulation minimizes the sum of squared differences between the entries of X and UV, aiming for a close approximation of the original matrix.

To promote sparsity in the factor matrices, regularization terms are added to the objective function. The most common types of regularization are based on the l_2 −norm (which is also the Frobenius norm in matrix setting), which encourages sparsity by penalizing the sum of squared values, controlling for large values in the matrices. The regularized objective function becomes:

$$\min_{U,V} [f(U,V) = \|X - UV\|_F^2 + \beta_1 \|U\|_F^2 + \beta_2 \|V\|_F^2], \tag{1}$$

where β_1 and β_2 are regularization parameters controlling the level of sparsity for matrices U and V, respectively.

Optimization using the BFGS Method

The optimization problem defined in (1) is and difficult to solve efficiently, especially for large-scale matrices. To address this, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method, a popular quasi-Newton optimization algorithm is utilized in this paper. BFGS is designed to approximate the second-order derivative (Hessian matrix) of the objective function without the need for explicit computation of the Hessian, making it computationally efficient for large-scale problems.

The BFGS method iteratively updates the factor matrices U and V by solving the following optimization problem:

$$(U_{k+1}, V_{k+1}) = (U_k, V_k) - \alpha_k H_k \nabla f(U_k, V_k), \tag{2}$$

where α_k is the step size, H_k is the approximate inverse Hessian at iteration (U_k, V_k), and $\nabla f(U_k, V_k)$, is the gradient of the objective function with respect to U and V at (U_k, V_k). The (inverse) BFGS formula (Dennis and Moré, 1977) for iteration (2) is given as below:

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}\ ,\ k \geq 0, \text{with } H_0 = I, \tag{3}$$

where:

- $s_k = (U_{k+1}, V_{k+1}) - (U_k, V_k)$ is the step in the variable space,
- $y_k = \nabla f(U_{k+1}, V_{k+1}) - \nabla f(U_k, V_k)$ is the difference in the gradients,
- H_k is the current approximation of the inverse Hessian matrix,
- I is the identity matrix.

Note that (3) is updated iteratively using gradient information, avoiding the need for explicit computation of second derivatives, which is particularly beneficial for large matrices. The BFGS method enjoys rapid convergence properties due to its use of second-order information, making it a suitable choice for sparse matrix factorization, where the optimization landscape is typically complex and involves non-convexities. In addition to its computational efficiency, the BFGS method can handle regularization terms naturally, allowing it to incorporate the sparsity-inducing regularization penalties, such as the Frobenius norm regularization in the factorization process.

Global Convergence with Armijo Line Search

The BFGS quasi-Newton method, when paired with the Armijo line search (also called a backtracking line search), has convergence guarantees under certain assumptions. The Armijo condition for a step size α_k, given as follow

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T p_k, \tag{4}$$

where $0 < c_1 < 1$ is a constant (often chosen as $c_1 = 10^{-4}$), and p_k is the search direction, ensures sufficient descent in the objective function, which helps control the step size during optimization.

For global convergence to a stationary point (i.e., a point where $\nabla f(x^*) = 0$), the BFGS method combined with an Armijo line search (4) has been shown to work under the following conditions (Nocedal and Wright (2001), Sim et al. (2018, 2019, 2022, 2023)):

Assumption 1.

i. The objective function $f(x)$ is twice continuously differentiable,
ii. The Hessian of the objective function, $\nabla^2 f(x)$, is Lipschitz continuous near the solution, meaning there exists a positive constant M such that:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq M\,\|x - y\|,\ \forall\, x, y.$$

iii. The initial point x_0 is sufficiently close to the local minimizer x^*, and the inverse Hessian approximation is sufficiently accurate.

Theorem 1 (Yuan and Sun (1991))**.**
Under Assumption 1, if the Armijo condition (4) is satisfied at each iteration, the sequence of iterates $\{x_k\}$ generated by the BFGS algorithm obeys

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \qquad \square$$

NUMERICAL EXPERIMENTS AND COMPARISONS

In this section, we report results of some numerical experiments with BFGS method for SMF. For the purpose of illustration, we consider the case where sparsity is required only for U, namely

$$\min_{U,V}[f(U,V) = \|X - UV\|_F^2 + \beta\|U\|_F^2].$$

We give the following example where X is given as below with 3 different values of β are considered, i.e. $\beta = 0.2, 0.6, 1.0$,

$$X = \begin{bmatrix} 0.8507 & 0.8154 & 0.6126 & 0.2278 & 0.7386 \\ 0.5606 & 0.8790 & 0.9900 & 0.4981 & 0.5860 \\ 0.9296 & 0.9889 & 0.5277 & 0.9009 & 0.2467 \\ 0.6967 & 0.0005 & 0.4795 & 0.5747 & 0.6664 \\ 0.5828 & 0.8654 & 0.8013 & 0.8452 & 0.0835 \end{bmatrix}.$$

Dimension of U and V is set as 5×5. The algorithm was coded in MATLAB. We stopped the iteration when the condition $\|\nabla f(U_k, V_k)\|_F \leq 10^{-4}$.

Table 1: BFGS for SMF

	$\|\nabla f(U_k, V_k)\|_F$		
k	$\beta = 0.2$	$\beta = 0.6$	$\beta = 1.0$
1	0.42000000	1.39000000	1.07000000
5	0.09360000	0.36100000	0.48100000
10	0.00804000	0.02940000	0.03920000
15	0.00198000	0.00664000	0.00886000
20	0.00064300	0.00223000	0.00297000
25	0.00007250	0.00016000	0.00021000

Table 2: Sparsity of U

	No.nonzero component	
$\beta = 0.2$	$\beta = 0.6$	$\beta = 1.0$
17	15	11

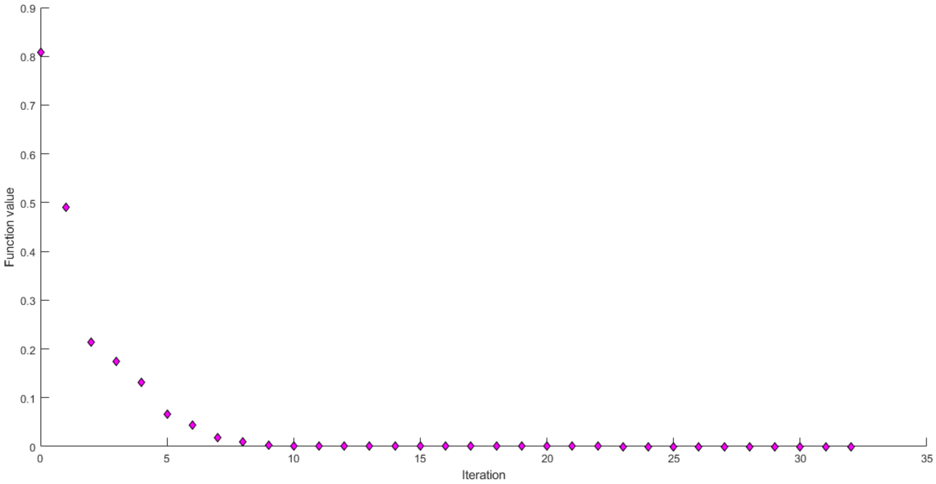

Figure 1: Value of f: $\beta = 0.2$

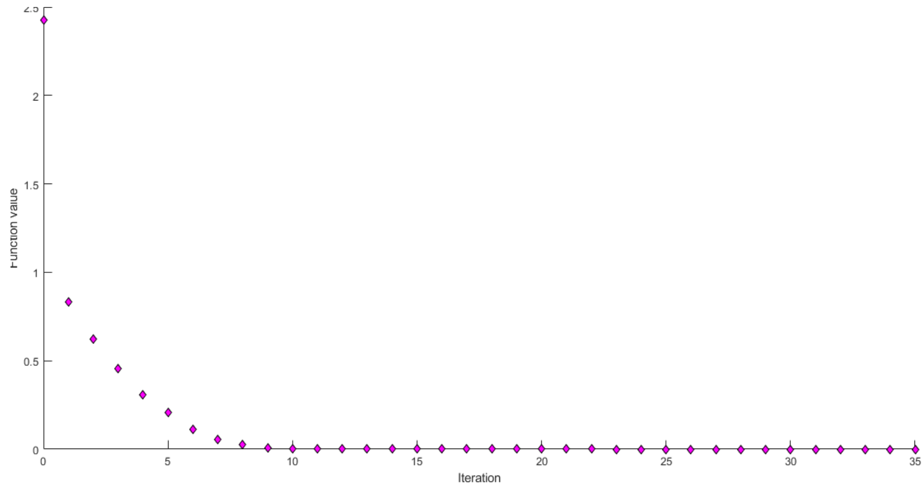

Figure 2: Value of f: $\beta = 0.6$

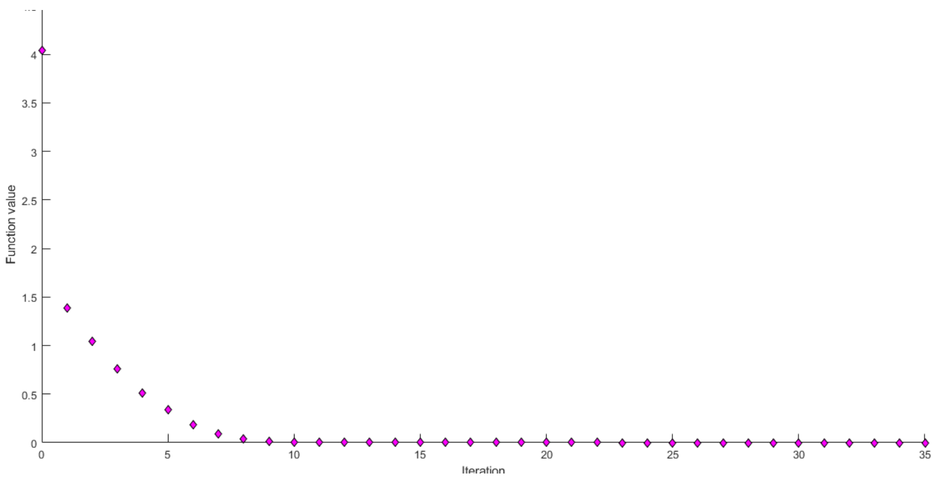

Figure 3: Value of f: $\beta = 1.0$

The numerical results obtained from our experiments provide evidence that increasing the regularization parameter significantly enhances the sparsity of the resulting matrix factors in sparse SMF. Specifically, as we varied the regularization parameter, a clear trend emerged: larger values of the parameter consistently led to sparser factor matrices. This behavior aligns with our theoretical expectations regarding Frobenius norm regularization, which penalizes larger parameter values and thus encourages the model to favor solutions with reduced complexity. In practical terms, higher sparsity not only reduces the memory footprint of the factor matrices but also enhances the computational efficiency during inference, making the system more scalable.

Furthermore, the results indicate that the BFGS method is an efficient optimization tool in achieving these sparse factorizations. The BFGS method's use of approximate second-order information allows it to converge more rapidly than first-order methods while maintaining a manageable computational overhead.

CONCLUSION

In conclusion, this paper presents a quasi-Newton method for SMF that effectively integrates Frobenius norm regularization to mitigate overfitting and improve generalization in data-driven contexts. By focusing on decomposing matrices while promoting sparsity in the factors, the method addresses key challenges in applications such as recommendation systems, signal processing, and dimensionality reduction. The incorporation of Frobenius norm regularization not only ensures sparser factorization but also stabilizes the optimization process by penalizing excessive parameter values. The proposed approach, which utilizes approximate second-order information, demonstrates a significant reduction in computational costs compared to traditional full Newton methods, making it more feasible for large-scale applications. Experimental results validate the method's effectiveness, showcasing its ability to produce high-quality sparse factorizations across varying penalty parameters.

REFERENCES

Dennis, J. E., and Moré, J. J. (1977), Quasi-Newton Methods, Motivation, and Theory. *SIAM Rev.*, **19(1)**: 46 – 89.

Al-Hakeem, H. A., Arasan, J., Mustafa, M. S. B., and Lim, F. P. (2023), Generalized Exponential Distribution with Interval-Censored Data and Time Dependent Covariate. *Commun. Stat. Simul. Comput.*, **52(12)**: 6149 – 6159.

Kon, C. M. L., and Labadin, J. (2019), Simulating Dengue: Comparison of Observed and Predicted Cases from Generic Reaction-Diffusion Model for Transmission of Mosquito-Borne Diseases. *Matematika*, **35(3)**, 309 – 330.

Nocedal, J., and Wright, S. J. (2006), Numerical Optimization (2nd ed.). New York: Springer.

Sim, H. S., Leong, W. J., Chen, C. Y., and Ibrahim, S. N. I. (2018), Multi-Step Spectral Gradient Methods with Modified Weak Secant Relation for Large-Scale Unconstrained Optimization. *Numer. Algebra Control Optim.*, **8(3)**: 377 – 387.

Sim, H. S., Leong, W. J., and Chen, C. Y. (2019), Gradient Method with Multiple Damping for Large-Scale Unconstrained Optimization. *Optim. Lett.*, **13(1)**: 617 – 632.

Sim, H. S., Chen, C. Y., Leong, W. J., and Li, J. (2022), Nonmonotone Spectral Gradient Method based on Memoryless Symmetric Rank-One Update for Large-Scale Unconstrained Optimization. *J. Ind. Manag. Optim.*, **18(6)**.

Sim, H. S., Ling, W. S. Y., Leong, W. J., and Chen, C. Y. (2023), Proximal Linearized Method for Sparse Equity Portfolio Optimization with Minimum Transaction Cost. *J. Inequal. Appl.*, **2023(1)**: 152.

Sun, W., Chee, C. S., Kok, L. Y., Lim, F. P., and Samsudin, S. (2023), Evaluation of Differences in the Performance Strategies of Top and Bottom Basketball Teams Utilizing Rank-Sum Ratio Comprehensive. *Front. Sports Act. Living*, **4**: 1052909.

Woo, G. Y. H., Sim, H. S., Goh, Y. K., and Leong, W. J. (2023), Proximal Variable Metric Method with Spectral Diagonal Update for Large-Scale Sparse Optimization. *J. Franklin Inst.*, **360(7)**: 4640 – 4660.

Yuan, Y., and Sun, W. (1991), Global Convergence of Quasi-Newton Methods with Armijo Line Search. *J. Comput. Math.*, **9(2)**: 125 – 135.