Regularized Cholesky Decomposition Method for low bit width computing

Zhang Z.
Moscow Institute of Physics and Technology, Dolgoprudny
zhibin@phystech.edu

Digital signal processing (DSP) depends on the accurate inversion of correlation matrices for an array of essential applications. Specifically, within wireless communications, several tasks such as interference whitening in multiantenna receivers , multiantenna channel estimation problem, and the representation of linear systems as $Ax = y$, all rely on the inverse of correlation matrices.

The Cholesky decomposition, favored for its computational efficiency and numerical robustness, occasionally grapples with challenges when confronted with certain ill-conditioned matrices — most notably when faced with a significantly large condition number. In addressing these computational challenges, several algorithms have been proposed, among which the Modified Cholesky Decomposition [1] stands out. However, the diagonal loading method, owing to its straightforward implementation, is gaining popularity in practical applications. Pinpointing an optimal loading value for this method remains a significant research challenge. In the context of low-bit-width computations, rounding errors present a pivotal challenge. Traditional determinate rounding error analysis [2] often provide overestimations, offering limited guidance for practical implementations. In this talk, we introduce probabilistic rounding error analysis [3] as the theoretical foundation. This approach offers a more precise rounding error estimation, allowing for a more accurate assessment of algorithmic stability and reliability, and providing a theoretical basis for the selection of diagonal loading values. Moreover, the potential for rounding errors raises concerns about the non-positive definiteness of correlation matrices, necessitating careful attention in algorithm design and implementation.

In the realm of signal processing, we need to deal with matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{y} + \mathbf{n}$$

where $\mathbf{A} \in \mathbb{C}^{n \times n}$ rank limited matrix, $\mathbf{x}, \mathbf{y}, \mathbf{n} \in \mathbb{C}^{n \times 1}$ correspond to the unknown vector, observation vector, and Gaussian noise vector, respectively. It's noteworthy that the Gaussian additive noise poses challenges as it tends to introduce distortions into correlation matrices. Particularly under conditions of under sampling, such noise can induce a pronounced disparity between the matrix's minimal eigenvalue and its anticipated statistical value[4]. To redress this discrepancy, our research delves intensively into regularization techniques.

This research elucidates the integration strategies of regularization techniques and Cholesky computation protection, particularly under constrained bit-width scenarios, aiming for a comprehensive rectification of the deviation of correlation matrix inverses.

REFERENCES

1. *Cheng, Sheung Hun, and Nicholas J. Higham.* A modified Cholesky algorithm based on a symmetric indefinite factorization. SIAM Journal on Matrix Analysis and Applications 19.4 (1998): 1097-1110.
2. *Higham, Nicholas J.* Accuracy and stability of numerical algorithms. Society for industrial and applied mathematics, 2002.
3. *Higham, Nicholas J., and Theo Mary.* A new approach to probabilistic rounding error analysis. SIAM journal on scientific computing 41.5 (2019): A2815-A2835.
4. *Marchenko, Vladimir Alexandrovich, and Leonid Andreevich Pastur.* Distribution of eigenvalues for some sets of random matrices. Matematicheskii Sbornik 114.4 (1967): 507-536.