The Journal of Open Source Software

cxreg: An R package for complex-valued lasso and graphical lasso

Younghoon Kim[1*], **Navonil Deb**[2*], **and Sumanta Basu**[2¶]

1 Center for Data Science for Enterprise and Society, Cornell University, United States **2** Department of Statistics and Data Science, Cornell University, United States ¶ Corresponding author * These authors contributed equally.

DOI: 10.21105/joss.09698

Software

- Review
- Repository
- Archive

Editor: George K. Thiruvathukal

Reviewers:

- @adallak
- @tambstat

Submitted: 25 July 2025
Published: 03 August 2026

Summary

Pathwise coordinate descent (Friedman et al., 2007) is a widely used iterative optimization method for solving penalized linear regression (*lasso*; Tibshirani, 1996) and penalized Gaussian log-likelihood (*graphical lasso*; e.g., Banerjee et al., 2008) problems. For a given sequence of tuning parameters λ in the penalty term, the algorithm updates one coefficient at a time by computing partial residuals.

The coordinate descent algorithm for complex-valued lasso and Gaussian graphical lasso, implemented in the cxreg R package, extends the standard pathwise coordinate descent method to complex-valued data (Deb et al., 2024). The algorithm operates by *realifying* complex numbers through a ring isomorphism (Herstein, 1991). Specifically, for a complex number $z \in \mathbb{C}$, we define the map:

$$\varphi(z) = \begin{pmatrix} \mathrm{Re}(z) & -\mathrm{Im}(z) \\ \mathrm{Im}(z) & \mathrm{Re}(z) \end{pmatrix}.$$

It is known that φ is a field isomorphism between $\mathbb{C}$ and the set $\mathcal{M}^{2\times 2}$ whose entries consist of $a, b \in \mathbb{R}$ to form 2 by 2 matrices (Artin, 2018),

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Under this mapping, algebraic operations in the complex domain correspond directly to those in the real domain. We extend these operations to the p-dimensional case $(p > 1)$; see Section 3.1 of Deb et al. (2024) for details.

Using this correspondence, we transform complex-valued linear regression problems into equivalent real-valued formulations. The same idea applies to the Gaussian log-likelihood for complex-valued variables. For the ℓ_1 penalty used in lasso and graphical lasso, we note that the modulus of a complex coefficient equals the entrywise ℓ_2 norm of its real and imaginary parts. Consequently, the lasso for p-dimensional complex variables is equivalent to a $2p$-dimensional group lasso (Yuan & Lin, 2006), where each group contains the real and imaginary components of one complex variable.

We also incorporate standard computational enhancements, including warm starts, using solutions from previous λ values, and active set selection to limit updates to relevant predictors (see Hastie et al., 2015, Chapter 5.4). The R package `cxreg` adopts a familiar interface similar to `glmnet` (Friedman et al., 2010), and a detailed description is available in the package vignette.

Statement of need

Solving complex-valued penalized Gaussian likelihood problems (and, as a component, penalized linear regression) is central to high-dimensional time series analysis. In the frequency domain, examining partial spectral coherence is analogous to studying partial correlations in Gaussian graphical models (e.g., Priestley, 1988). With the development of the local Whittle likelihood approximation (Whittle, 1951), the inverse spectral density matrix, known as the spectral precision matrix, provides a way to characterize associations among variables in high-dimensional time series data.

These associations, represented by the spectral precision matrix, are evaluated at fixed frequencies. However, because the spectral density matrix estimator is inconsistent at a single frequency (see Brockwell & Davis, 1991, Chapter 10), it is common to use an averaged, smoothed periodogram, obtained by aggregating periodogram matrices across neighboring frequencies, as a consistent sample analog. Consequently, spectral precision matrices must be computed across multiple neighboring frequencies around each frequency of interest. Such analyses involve repeated high-dimensional optimization, as required for statistical inference on spectral precision matrices (Krampe & Paparoditis, 2025). Therefore, developing efficient algorithms for complex-valued penalized Gaussian likelihood estimation is crucial for scalable and accurate frequency-domain analysis.

State of the field

Although coordinate descent is computationally efficient and widely used to solve the lasso (Friedman et al., 2007) and graphical lasso (Friedman et al., 2008), particularly under sparsity assumptions, its extension to complex-valued settings remains limited. This gap largely arises from the lack of methods for updating complex-valued partial residuals within the coordinate descent framework.

Several alternative approaches have been proposed for complex-valued graphical modeling, but many do not explicitly exploit sparsity within a coordinate descent framework, which can limit scalability in high-dimensional settings. Recent software, such as the R package tsglasso (Dallakyan et al., 2022) and the Python package FreDom (Dallakyan, 2024), provide related implementations based on closed-form updates derived via Wirtinger calculus. Another line of work employs the alternating direction method of multipliers (ADMM; Baek et al., 2023).

In contrast, the algorithm implemented in this package is specifically designed to leverage sparsity within a coordinate descent framework. The method follows Deb et al. (2024), where comparisons with benchmark approaches, including the complex-valued extension of the nodewise regression CIPE algorithm (Fiecas et al., 2019) and its pathwise variants, demonstrate improvements in both estimation accuracy and computational efficiency.

Acknowledgements

ND and SB acknowledge partial support from NIH award R21NS120227. In addition, SB acknowledges partial support from NSF awards DMS-1812128, DMS-2210675, DMS-2239102 and NIH award R01GM135926. The authors thank the two referees for their helpful comments and suggestions, which improved the clarity and presentation of the manuscript.

References

Artin, M. (2018). *Algebra*. Pearson.

Baek, C., Düker, M.-C., & Pipiras, V. (2023). Local Whittle estimation of high-dimensional long-run variance and precision matrices. *The Annals of Statistics*, *51*(6), 2386–2414. https://doi.org/10.1214/23-AOS2330

Banerjee, O., El Ghaoui, L., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research*, *9*, 485–516.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods*. Springer Science & Business Media.

Dallakyan, A. (2024). On learning time series DAGs: A frequency domain approach. *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2024.06.003

Dallakyan, A., Kim, R., & Pourahmadi, M. (2022). Time series graphical lasso and sparse VAR estimation. *Computational Statistics & Data Analysis*, *176*, 107557. https://doi.org/10.1016/j.csda.2022.107557

Deb, N., Kuceyeski, A., & Basu, S. (2024). Regularized estimation of sparse spectral precision matrices. *arXiv Preprint arXiv:2401.11128*. https://arxiv.org/abs/2401.11128

Fiecas, M., Leng, C., Liu, W., & Yu, Y. (2019). Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, *13*(2), 4079–4101. https://doi.org/10.1214/19-EJS1621

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*(2), 302–332. https://doi.org/10.1214/07-AOAS131

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22. https://doi.org/10.18637/jss.v033.i01

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations* (Vol. 143). CRC Press. https://doi.org/10.1201/b18401

Herstein, I. N. (1991). *Topics in algebra*. John Wiley & Sons.

Krampe, J., & Paparoditis, E. (2025). Frequency domain statistical inference for high-dimensional time series. *Journal of the American Statistical Association*, *120*(551), 1580–1592. https://doi.org/10.1080/01621459.2025.2479244

Priestley, M. B. (1988). *The spectral analysis of time series*. Oxford University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Whittle, P. (1951). *Hypothesis testing in time series analysis*. Almqvist & Wiksells boktr.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *68*(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x