

MatrixLM: A Julia package to obtain closed-form least squares estimates for matrix linear models.

Gregory Farage^{1*}, Jane W Liang^{2*}, Chenhao Zhao¹, and Šaunak Sen¹

¹ Division of Biostatistics, Department of Preventive Medicine, University of Tennessee Health Science Center, United States ² Division of Research, Kaiser Permanente Northern California, United States * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [↗](#)

Submitted: 05 February 2026
Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

MatrixLM is an open-source Julia package for fitting matrix linear models, which extend classical linear regression to a bilinear framework for matrix-valued responses. It is designed for analyzing high-throughput assays in which both rows and columns of the data matrix have associated covariates, such as in metabolomics, proteomics, or chemical genetic screens.

In a matrix linear model, the entries of a response matrix are modeled as a joint function of sample-level covariates (e.g., treatment group, demographic factors) and feature-level covariates (e.g., molecular or anatomical annotations, biological groups, pathways). MatrixLM implements efficient estimation and inference for this class of models using efficient matrix operations whenever possible allowing users to fit large numbers of models while retaining an explicit linear model interpretation. The inputs include a response matrix and two design matrices encoding the row and column covariates, and the main outputs include estimated coefficients, standard errors, and test statistics for user-specified contrasts.

Compared with workflows built from many separate univariate models, MatrixLM provides a unified interface for specifying, fitting, and summarizing matrix linear models. This reduces code duplication, improves reproducibility, and makes it easier to express hypotheses that naturally involve both sample- and feature-level information (for example, testing for differential effects across feature groups or experimental conditions). By providing an efficient implementation in Julia, MatrixLM enables researchers to perform interpretable analyses of large structured matrix-valued data.

Statement of need

High-throughput studies in biology and medicine often produce matrix-shaped data where each row corresponds to a sample (e.g., a patient, mutant strain, or experimental unit) and each column represents a molecular measurement (e.g., metabolite, gene, or phenotype). In many applications, both the samples and the measured features have associated metadata that should be incorporated into the analysis. However, existing tools either ignore these annotations or handle them in a fragmented, two-step fashion.

Standard approaches often involve fitting a separate model to each feature (e.g., using t-tests or linear models) and then performing a second-stage enrichment or grouping analysis. This approach is limited in two key ways: (1) it does not handle overlapping or quantitative annotations well, and (2) it fails to exploit shared structure among features or among samples. Dimension-reduction and machine learning methods address some of these issues, but often sacrifice interpretability and do not provide familiar statistical outputs like effect sizes or

81 This optimization problem has a closed-form solution:

$$\hat{B} = (X^T X)^{-1} X^T Y Z (Z^T Z)^{-1},$$

82 when $X^T X$ and $Z^T Z$ are invertible. From \hat{B} , the software can construct fitted values
83 $\hat{Y} = X \hat{B} Z^T$ and standard errors and test statistics for entries of \hat{B} or for user-specified linear
84 contrasts, directly analogous to classical linear models.

85 When there is only a single feature (one column of Y) and no feature-level design matrix Z ,
86 this framework reduces to ordinary linear regression with design matrix X . The MatrixLM
87 package therefore generalizes familiar linear modeling ideas to matrix-valued outcomes with
88 structured annotations on both rows and columns.

89 Figures

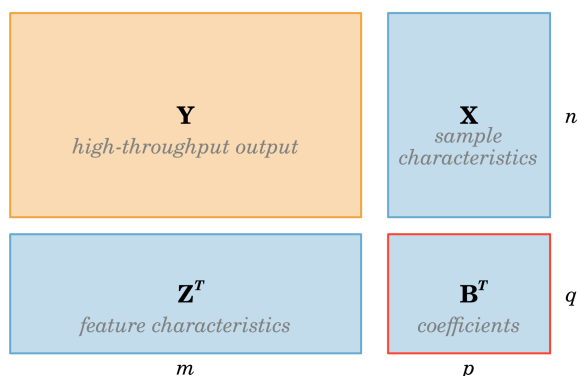


Figure 1: A visualization of the response ($Y : n \times m$), sample covariates ($X : n \times p$), feature covariates ($Z : m \times q$), and coefficients ($B : p \times q$) matrices for a matrix linear model. The dimensions in the model correspond to n samples/individuals, m features/measurements, p sample covariates, and q feature covariates; the matrix B is to be estimated.

90 Conclusion and future directions

91 MatrixLM provides a practical implementation of matrix linear models for encoding relationships
92 and groupings high-throughput, matrix-shaped data with annotations on both samples and
93 features. By combining a flexible formula interface with fast, closed-form least-squares
94 estimation, the package makes it straightforward for applied researchers to encode biological
95 or experimental structure directly into their models.

96 In ongoing work, we are extending this framework to penalized matrix linear models for high-
97 dimensional settings: a companion Julia package, MatrixLMnet, that implements elastic-net
98 and related penalties on the coefficient matrix to enable variable selection and regularization
99 in matrix linear models.

100 AI usage disclosure

101 The authors used GitHub Copilot to assist in coding testing functions. All Copilot-suggested
102 code was reviewed, tested, and validated by the human authors to ensure correctness. The
103 authors take full responsibility for the content of this manuscript.

104 **Acknowledgements**

105 This work started when JWL was a summer intern at UCSF, and continued when she was a
106 scientific programmer at the University of Tennessee Health Science Center (UTHSC). We
107 thank both UCSF and UTHSC for funding, and a supportive environment. This work was
108 partly supported by National Institutes of Health grants GM-070683 (SS), GM-078338 (SS),
109 GM-123489 (SS, GF), ES-022841 (SS), and DA-044223 (SS).

110 **References**

- 111 Farage, G., Zhao, C., Choi, H. Y., Garrett, T. J., Elam, M. B., Kechris, K., & Sen, Ś. (2025).
112 Matrix linear models for connecting metabolite composition to individual characteristics.
113 *Metabolites*, 15(2). <https://doi.org/10.3390/metabo15020140>
- 114 Liang, J. W., Nichols, R. J., & Sen, Ś. (2019). Matrix Linear Models for High-Throughput
115 Chemical Genetic Screens. *Genetics*, 4(212), 1063–1073. [https://doi.org/10.1534/genetics.](https://doi.org/10.1534/genetics.119.302299)
116 [119.302299](https://doi.org/10.1534/genetics.119.302299)

DRAFT