# pysr3: A Python Package for Sparse Relaxed Regularized Regression

Aleksei Sholokhov [1], **Peng Zheng** [2], **and Aleksandr Aravkin** [1,2]

1 Department of Applied Mathematics, University of Washington **2** Department of Health Metrics Sciences, University of Washington

Summary

Datasets increasingly contain more and more potential covariates that may be related to different research questions. The presence of irrelevant or weakly relevant features can be counterproductive to modeling, as over-parametrized models may lead to unstable estimates, invalid inference, and low prediction accuracy. Reliable feature selection is a requirement for a wide range of regression settings.

Feature selection methods are a hot research topic (Buscemi & Plaia, 2020),(Miao & Niu, 2016),(Li et al., 2020), with a plethora of numerical approaches and corresponding implementations. However, most of the current state-of-the-art tools custom-tailor their implementation to their mathematical approach (e.g. use an optimization scheme that works only for a particular regularizer). In practice, it forces practitioners to re-implement their workflow for each method that they want to use or try, even when the difference between methods is minor. The absence of universal open source implementations effectively blocks practitioners from comparing all available methods, with the effect of slowing uptake of research results in the field.

We fill this gap by implementing recently developed universal solvers (Zheng et al., 2018), (Sholokhov et al., 2022) that (1) work with most popular regularized regression techniques, and (2) improve selection accuracy of any regularized regression approach using new relaxation reformulations. To date, the library supports linear models (classic regression) and linear mixed effects models. Because of full compatibility with sklearn, all pysr3 models can be used in pipeline with classic modelling blocks such as data pre-processors, randomized grid search, cross-validation, and quality metrics.

Statement of Need

Reliable automated feature selection requires easy-to-use libraries, so that practitioners can test and compare multiple regularization approaches using their data, choose the most effective method, and apply the analysis at scale. These libraries should be flexible and modular enough to accommodate future developments, such as newly proposed regularizers, without forcing the practitioner to implement new solvers for each approach. At the same time, the libraries must be efficient and robust enough to handle common challenges such as ill-conditioned problems that arise in datasets with correlated predictors. The PySR3 library is designed to easily include new loss functions, constraints, information criteria, and regularization strategies. All PySR3 models fully adhere the standards and interface requirements of sklearn (Buitinck et al., 2013), providing a familiar interface to users.

Currently, PySR3 models can be used to automatically select features in both linear regression models and linear mixed effects (LME) models, which extend linear regression to clustered data settings. LME models commonly arise in longitudinal data analysis and meta-analysis.

Feature selection for linear mixed-effects models is harder than for linear regression models due to non-linearity of LMEs and within-group correlations. To the best of our knowledge, there are no standalone Python packages for mixed-effect feature selection, while the alternatives implemented in R, such as lmmlasso (Schelldorfer et al., 2011) and glmmlasso (Schelldorfer et al., 2014), take more time to converge and yield worse selection accuracy, as discussed in (Sholokhov et al., 2022).

Core idea and structure of `pysr3`

The baseline optimization method of PySR3 is proximal gradient descent (PGD). PGD exploits the fact that most feature selection methods minimize the sum of a smooth loss that captures the negative log likelihood of the data and a non-smooth sparsity promoting regularizer. PGD works well as long as the regularizer has an implementable proximal operator. Many widely-used regularizers have proximal operators that are well known and either have efficient numerical routines or closed form solutions (Zheng et al., 2018). Examples include the zero-norm (L0), least absolute shrinkage and selection operator (LASSO), adaptive LASSO (A-LASSO), and smoothly clipped absolute deviation (SCAD) regularizers.

Each regularizer included in PySR3 can also be used in its relaxed SR3 form (Zheng et al., 2018). SR3 preconditions the likelihood, improving the performance of feature selection methods. PGD on the SR3-transformed problem takes fewer iterations to converge, and the features selected are more accurate and have lower false positive rates across simulated examples for both linear regression (Zheng et al., 2018) and LME (Sholokhov et al., 2022) models, as illustrated on Figure 1 (adapted from Figure 1 from (Sholokhov et al., 2022)).
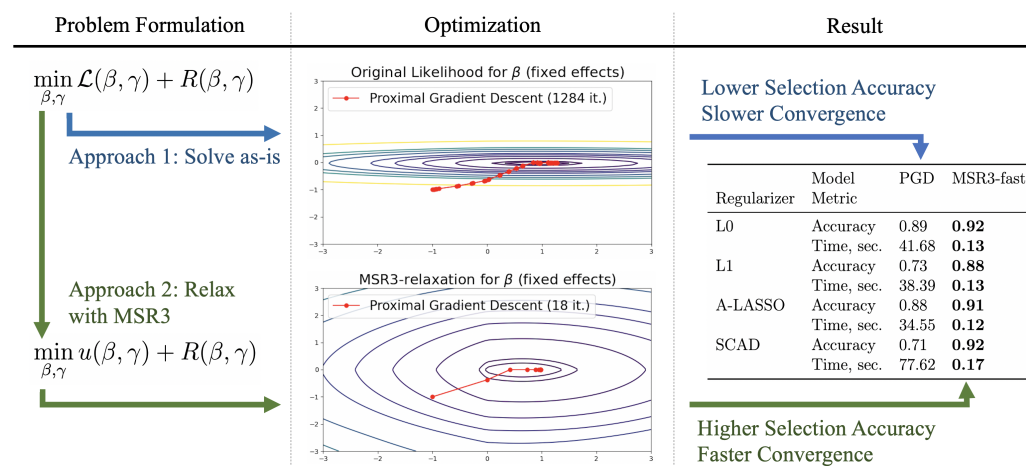

Figure 1: SR3 relaxation makes a problem more well-conditioned which accelerates the optimization.

More information about the structure of the library can be found in documentation, while the mathematical contributions are extensively discussed in (Zheng et al., 2018) for linear regression and in (Sholokhov et al., 2022),(Aravkin et al., 2022) for linear mixed effects models.

Ongoing Research and Dissemination

The manuscripts "A Relaxation Approach to Feature Selection for Linear Mixed Effects Models" and "Analysis of Relaxation Methods for Feature Selection in Mixed Effects Models" are undergoing simultaneous peer-review. Since its introduction in (Zheng et al., 2018), SR3 has been cited 92 times, and used in model discovery (Mendible et al., 2020), optimal dose management (Levin et al., 2019) and inverse problems (Baraldi et al., 2019). The

LME extensions of PySR3 was developed for variable selection in meta-analysis, which is a fundamental problem in risk factor analysis for the Global Burden of Disease study (Murray et al., 2020).

References

Aravkin, A., Burke, J., Sholokhov, A., & Zheng, P. (2022). Analysis of relaxation methods for feature selection in mixed effects models. *arXiv Preprint arXiv:2209.10575*. https://doi.org/10.48550/arXiv.2209.10575

Baraldi, R., Kumar, R., & Aravkin, A. (2019). Basis pursuit denoise with nonsmooth constraints. *IEEE Transactions on Signal Processing*, *67*(22), 5811–5823. https://doi.org/10.1109/tsp.2019.2946029

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: Experiences from the scikit-learn project. *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108–122. https://doi.org/10.48550/arXiv.1309.0238

Buscemi, S., & Plaia, A. (2020). Model selection in linear mixed-effect models. *AStA Advances in Statistical Analysis*, *104*(4), 529–575. https://doi.org/10.1007/s10182-019-00359-z

Levin, R., Aravkin, A. Y., & Kim, M. (2019). A proof of principle: Multi-modality radiotherapy optimization. *arXiv Preprint arXiv:1911.05182*. https://doi.org/10.48550/arXiv.1911.05182

Li, X., Wang, Y., & Ruiz, R. (2020). A survey on sparse learning models for feature selection. *IEEE Transactions on Cybernetics*. https://doi.org/10.1109/TCYB.2020.2982445

Mendible, A., Brunton, S. L., Aravkin, A. Y., Lowrie, W., & Kutz, J. N. (2020). Dimensionality reduction and reduced-order modeling for traveling wave physics. *Theoretical and Computational Fluid Dynamics*, *34*(4), 385–400. https://doi.org/10.1007/s00162-020-00529-9

Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, *91*, 919–926. https://doi.org/10.1016/j.procs.2016.07.111

Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., & others. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, *396*(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Schelldorfer, J., Bühlmann, P., & DE GEER, S. V. (2011). Estimation for high-dimensional linear mixed-effects models using l1-penalization. *Scandinavian Journal of Statistics*, *38*(2), 197–214. https://doi.org/10.1111/j.1467-9469.2011.00740.x

Schelldorfer, J., Meier, L., & Bühlmann, P. (2014). Glmmlasso: An algorithm for high-dimensional generalized linear mixed models using L1-penalization. *Journal of Computational and Graphical Statistics*, *23*(2), 460–477. https://doi.org/10.1080/10618600.2013.773239

Sholokhov, A., Burke, J. V., Santomauro, D. F., Zheng, P., & Aravkin, A. (2022). A relaxation approach to feature selection for linear mixed effects models. *arXiv Preprint arXiv:2205.06925*. https://doi.org/10.48550/arXiv.2205.06925

Zheng, P., Askham, T., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2018). A unified framework for sparse relaxed regularized regression: SR3. *IEEE Access*, *7*, 1404–1423. https://doi.org/10.1109/ACCESS.2018.2886528