Hybrid Regularization for Random Feature Models

Tony Maccagnano [1], Sarah Skaggs [2], Madeline Straus [3], Advay Vyas [4]

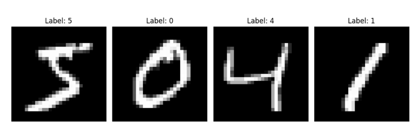
[1] Michigan Tech, [2] Virginia Tech, [3] Tufts University, [4] University of Texas at Austin

Abstract

Hybrid regularization methods have been well-documented as effective solvers for ill-posed inverse problems. These methods have been recently extended into the machine learning domain for training random feature models (RFMs) and have applications in bioinformatics, 3D reconstruction, and image processing. Hybrid regularization methods, specifically hybrid Krylov subspace methods, have shown to be competitive with classical regularization techniques for training RFMs. While these methods have shown success, they are often computationally expensive. In the context of inverse problems, new hybrid Krylov subspace methods have been developed that are inner product free, making them more efficient. In this project, we consider the application of these hybrid Krylov subspace solvers for the training of RFMs.

Background

Image classification involves mapping a set of input images to a corresponding set of output labels. For this project, we use the widely studied MNIST dataset.



MNIST Dataset Sample

A common and effective approach in image classification is to represent it as a **large scale linear inverse problem** of the form

$$\mathbf{B} \approx \mathbf{A}\mathbf{X}$$

where $\mathbf{B}$ is the observed data (labels), $\mathbf{A}$ represents the forward model, and $\mathbf{X}$ is the unknown coefficient matrix we want to approximate.

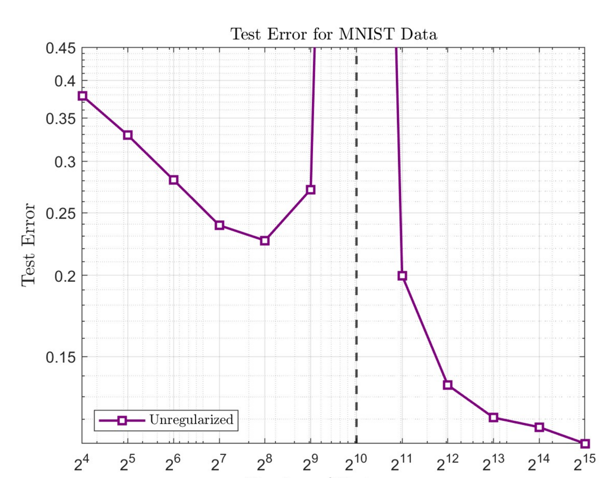
Random Feature Models

We can use **Random Feature Models (RFMs)** to transform the image data such that a linear model can be applied.

Challenges

- Ill-posedness
- **Double descent phenomenon**
- Poor generalization

These challenges can be mitigated with various regularization techniques.



Regularization

Iterative regularization solves the least squares problem in smaller, increasing spaces. A popular approach is using **Krylov subspaces**, where the order-k Krylov subspace of a matrix $\mathbf{A}$ with respect to a vector $\mathbf{b}$ is defined by

$$\mathcal{K}_k(\mathbf{A},\mathbf{b}) = \operatorname{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}\}$$

Issue: relatively unstable semi-convergence.

Tikhonov regularization solves the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda^2\|\mathbf{x}\|_2^2$$

where $\lambda > 0$ is the **penalty term** and $\|\mathbf{x}\|_2^2$ is the regularization term.

Issue: choice of λ heavily influences solution quality.

Hybrid regularization is a combination of iterative and variational regularization.

- Uses iterative methods to build problem subspaces (e.g. Krylov)
- Applies variational regularization (e.g. Tikhonov) inside those subspaces

Approach

Random Feature Model Process

Our goal is to apply a random, nonlinear transformation $\mathbf{A}: \mathbb{R}^{n_i} \to \mathbb{R}^m$ to each row in input matrix of images $\mathbf{Y} \in \mathbb{R}^{n\times n_i}$ in order to identify patterns in the data. We choose to define $\mathbf{A}$ as:

$$\mathbf{A}(\mathbf{Y}) = \left[a(\mathbf{Y}\mathbf{K} + 1_n\mathbf{c}^T)\ 1_n\right] \in \mathbb{R}^{n\times m}$$

- $\mathbf{K} \in \mathbb{R}^{n_i\times(m-1)}$ and $\mathbf{c} \in \mathbb{R}^{m-1}$ are random values from a unit hypersphere.
- a is the ReLU activation function, $a(x) = \max(x, 0)$.
- The dimension m controls the expressiveness of the RFM.

The RFM training consists of finding $\mathbf{X} \in \mathbb{R}^{m\times n_o}$ such that $\mathbf{A}(\mathbf{Y})\cdot\mathbf{X} \approx \mathbf{B}$. This corresponds to solving the least squares problem

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \frac{1}{2n}\|\mathbf{A}(\mathbf{Y})\mathbf{X} - \mathbf{B}\|_F^2.$$

For this problem, we need to use a solver that can run efficiently and be stable enough to train the RFM thoroughly.

Hybrid Regularization Methods

Previously, in [3], the **Hybrid LSQR** method (available in MATLAB's IR Tools package [2]) was successfully applied to train RFMs. Notably, the method:

- Avoided the double descent phenomenon
- Performed competitively against existing methods
- However, this method uses computationally expensive inner products

Our proposal is to instead utilize a **Hybrid LSLU** method, introduced in [1], which improves on Hybrid LSQR by using a **Hessenberg process** that

- Does not require inner products
- Requires less storage and work

Through these improvements, this unique approach is advantageous for both mixed precision and high-performance computing scenarios.

Future Work

In the future, we want to investigate the regularization parameter calculations as well as the early stopping criterion to counteract the double descent phenomenon that occurs in later iterations. We also would like to explore the randomized sketch-and-solve version of Hybrid LSLU which has the potential to address some of the issues we currently see. Finally, we are looking into additional datasets to use LSLU with RFMs on and ways to optimize and refine our algorithms to increase efficiency.

Acknowledgments

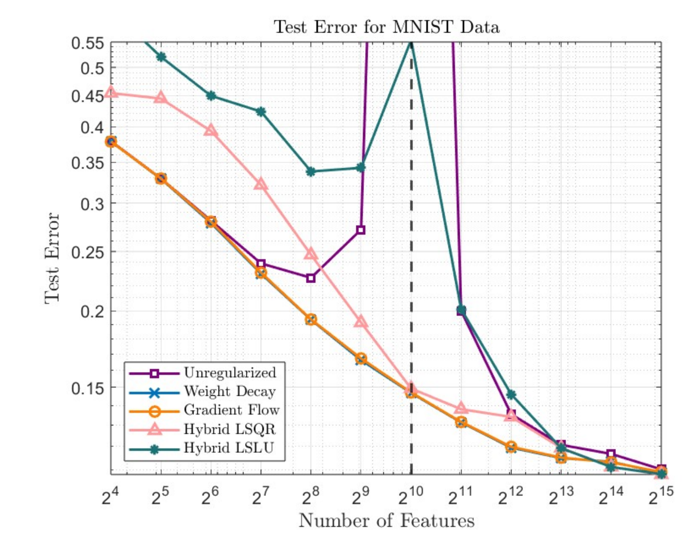
This work is supported by the US NSF award DMS-2349534. We also would like to thank our amazing project mentor Dr. Chelsea Drum and the rest of the Emory Computational Mathematics for Data Science REU team for their invaluable help and resources.

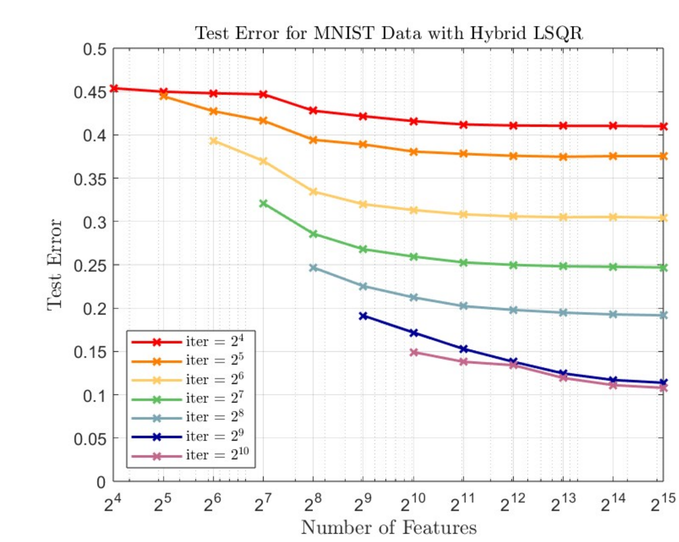
Numerical Results

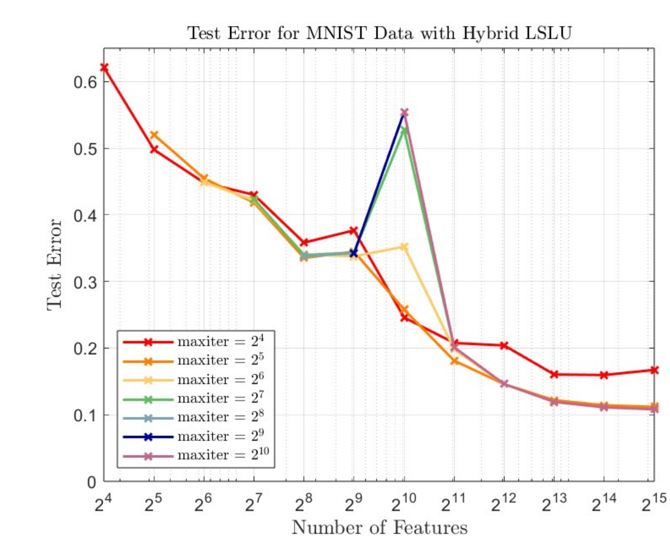
For our experiments, we tested the Hybrid LSLU method with the MNIST dataset, then compared the results to Hybrid LSQR. These graphs compare error, iterations, and features for both hybrid and standard regularization methods.

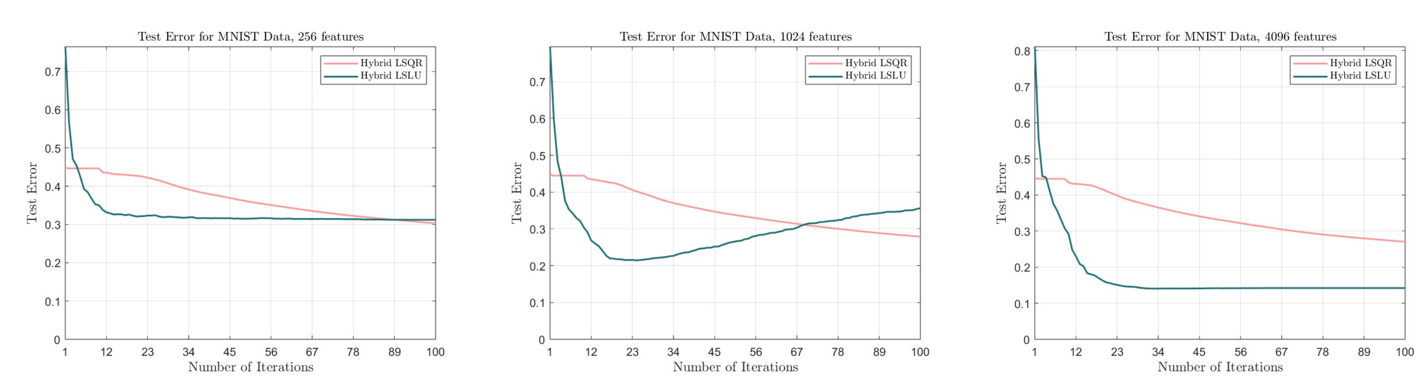
Methods and Parameters

- Unregularized
- Weight Decay & Gradient Flow (optimal parameter)
- Hybrid LSQR (GCV)
- Hybrid LSLU (wGCV)









Comparison of LSQR (GCV) and LSLU (adaptive wGCV) over 100 iterations

References

[1] Ariana N Brown et al. "Inner Product Free Krylov Methods for Large-Scale Inverse Problems". In: *arXiv preprint arXiv:2409.05239* (2024).

[2] Silvia Gazzola, Per Christian Hansen, and James G Nagy. "IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems". In: *Numerical Algorithms* 81 (2019), pp. 773–811.

[3] Kelvin Kan, James Nagy, and Lars Ruthotto. "Hybrid Regularization Methods Achieve Near-Optimal Regularization in Random Feature Models". In: *Submitted to Transactions on Machine Learning Research* (2024).

[4] Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems*. Vol. 20. 2007, pp. 1177–1184.