AMATH 563 - HOMEWORK 1 (COMPUTATIONAL REPORT)

NATE WHYBRA

Applied Mathematics Department, University of Washington, Seattle, WA
nwhybra@uw.edu

1. Introduction

Given K observations of noisy measurement data, it is often desirable to find a model that accurately depicts the ground truth. In this paper, we investigate least-squares model fitting with l^2 regularization using model functions of the form $f(x) = \sum_{i=0}^{N-1} c_i f_i(x)$ on observations from the $1D$ Levy function. Call the set of functions $\{f_i\}_{i=0}^{N-1}$ a dictionary. In this paper, we test the results of the model-fitting over various dictionaries, different values of N, and different regularization values λ, and compare their accuracies with the mean-squared-error (MSE).

2. Methods

Suppose we have K noisy observations $y_k = f(x_k) + \epsilon_k$ of some unknown function $f : \mathbb{R} \to \mathbb{R}$, where $\epsilon \sim N(0, \sigma^2)$ and for $k \in [K]$. To find a function $\hat{f}$ that approximates f, we introduce a model:

$$g(x) = \sum_{i=0}^{N-1} c_i f_i(x)$$

And try to find the vector $c = [c_0, c_1, ..., c_{N-1}]^T$ that minimizes:

$$h(c) = \frac{1}{2} \sum_{k=0}^{K-1} |g(x_k) - y_k|^2 + \frac{\lambda}{2} \|c\|_2^2$$

$$= \frac{1}{2} \sum_{k=0}^{K-1} \left(\sum_{i=0}^{N-1} c_i f_i(x_k) - y_k \right)^2 + \frac{\lambda}{2} \|c\|_2^2$$

Define $A \in \mathbb{R}^{K \times N}$ to be the matrix where row k is $[f_0(x_k), f_1(x_k), \cdots, f_{N-1}(x_k)]$. Then the above is equivalent to:

$$h(c) = \frac{1}{2} \|Ac - y\|_2^2 + \frac{\lambda}{2} \|c\|^2$$

This is a quadratic function in c which is well know to be convex [1], therefore it has a global minimum which can be found by computing $\frac{dh}{dc}$ and setting it equal to 0. Doing so yields:

$$\frac{dh}{dc} = A^T(Ac - y) + \lambda c := 0$$

$$\implies (A^T A + \lambda I_K)c = A^T y$$

Date: April 16, 2025.

$$\implies \hat{c} = (A^T A + \lambda I_K)^{-1} A^T y$$

Our approximation $\hat{f}$ then takes the form:

$$\hat{f}(x) = \sum_{i=0}^{N-1} \hat{c}_i f_i(x)$$

We proceed by studying the effects of choosing different dictionaries $\{f_i(x)\}_{i=0}^{N-1}$, changing values of N, and values of the regularization parameter λ on fitting noisy data from the $1D$ Levy function.

3. Results

The $1D$ Levy function can be written as:

$$f(x) = \sin(\pi w)^2 + (w-1)^2(1+\sin(2\pi w)^2) \text{ with } w = \frac{10x+3}{4} \text{ and } x \in [-1,1]$$

We are given $K = 50$ observations of f that has noise distributed like $N(0, \sigma^2)$ with $\sigma \approx 0.653$. We considered the following dictionaries:

$$\{f_i\}_{i=0}^{N-1} = \begin{cases} \{\cos(i\pi x)\} & \text{(Cosine)} \\ \{x^i\} & \text{(Monomial)} \\ \{e^{-\frac{(x-z_i)^2}{2l^2}}\} & z_i = -1 + \frac{2i}{N} \quad (RBF) \\ \{x^i\}_{i \leq N-2} \cup \{f\} & \text{(Monomial + Ground Truth)} \end{cases}$$

The last dictionary replaces the highest power monomial with the ground truth function itself. We considered values of $N \in \{10, 50, 100\}$ and values of $\lambda \in \{10^{-10}, 10^{-2}, 1\}$. Our results are summarized in the Tables $1, 2, 3, 4$ and Plots (A), (B), (C), (D), below. Tables 1 through 4 summarize the mean squared error between the fitted models (cosine, monomial, rbf, and monomial + ground truth respectively) over the dictionaries examined and the true Levy function sampled at 500 equidistant points in $[-1, 1]$ for the various values of N and λ chosen. Plots (A), (B), (C), and (D) show the fitted models compared to the true Levy function and observation data.

Table 1. MSE for Cosine Dictionary

	$\lambda = 10^{-10}$	$\lambda = 10^{-2}$	$\lambda = 1$
$N = 10$	2.45	2.45	2.44
$N = 50$	11.30	11.30	11.33
$N = 100$	16.36	16.36	16.38

Table 2. MSE for Monomial Dictionary

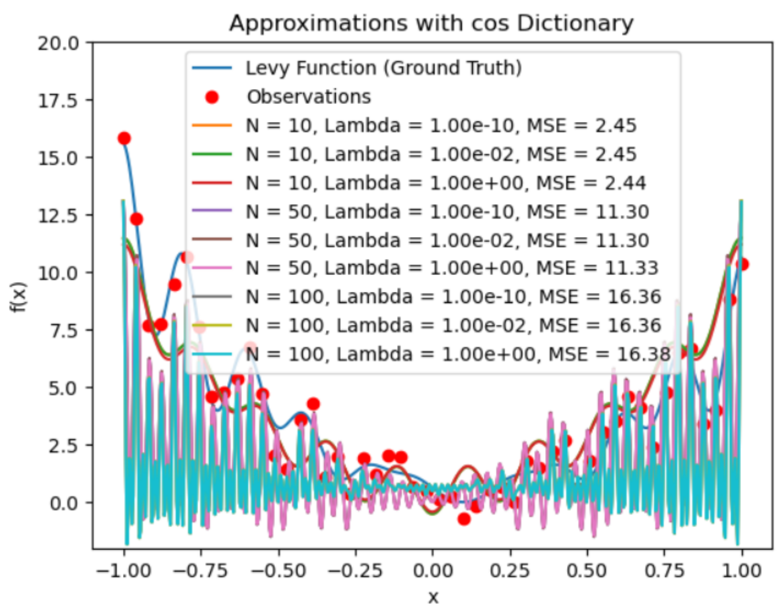
	$\lambda = 10^{-10}$	$\lambda = 10^{-2}$	$\lambda = 1$
$N = 10$	1.07	1.11	1.38
$N = 50$	6.03	0.82	1.28
$N = 100$	29364.77	1.10	1.28

Table 3. MSE for RBF Dictionary

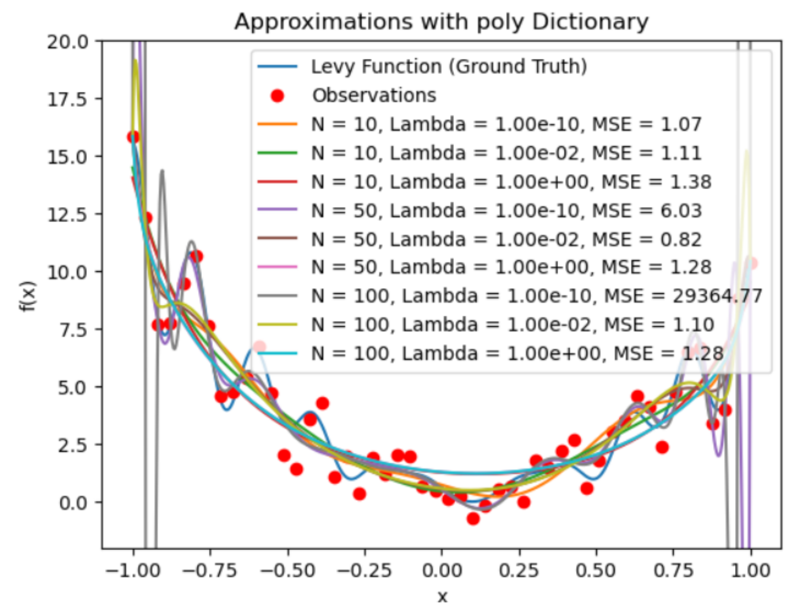
	$\lambda = 10^{-10}$	$\lambda = 10^{-2}$	$\lambda = 1$
$N = 10$	1.61	1.59	2.31
$N = 50$	0.61	1.34	1.72
$N = 100$	0.57	1.30	1.63

Table 4. MSE for Monomial + GT Dictionary

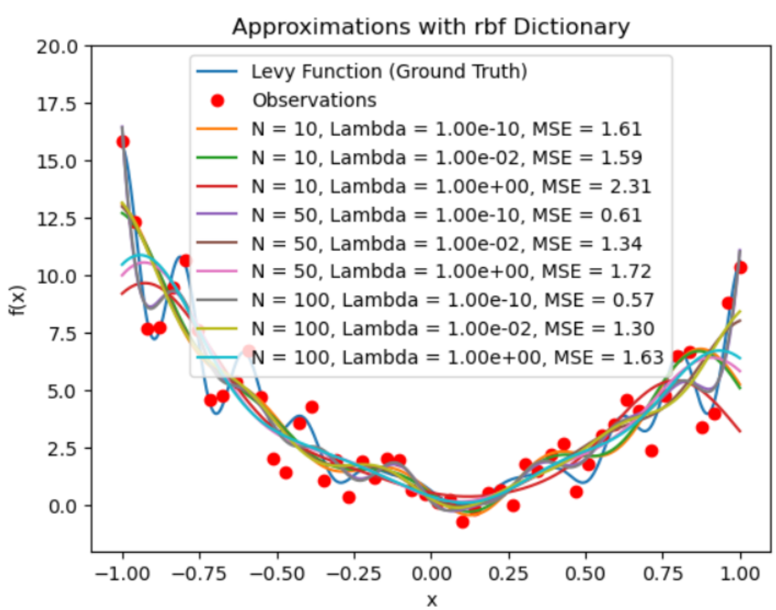
	$\lambda = 10^{-10}$	$\lambda = 10^{-2}$	$\lambda = 1$
$N = 10$	0.05	0.03	0.02
$N = 50$	5.20	0.07	0.01
$N = 100$	23.81	0.11	0.02



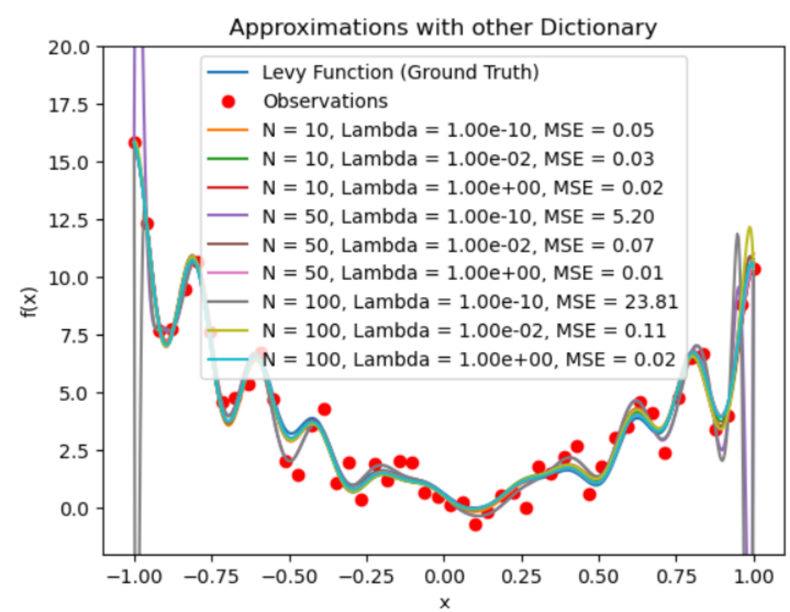
(A) See next page.



(B) See next page.



(C) See next page.



(D) See next page.

4. Summary and Conclusions

Both qualitatively and quantitatively, the approximations from the monomial + ground truth dictionary out perform that of the models from the other dictionaries. With $N = 50$ and $\lambda = 1$, we are able to achieve a small error of 0.01 as opposed to the other dictionaries which show errors larger than 1 for most values of N and λ. The reason for this is that the damping caused by the large regularization parameter causes most of the monomial coefficients to be small and for the true function coefficient to be close to 1. The cosine and monomial dictionaries performed the worst, with clear Runge phenomenon on the edges in the monomial fits, and high frequency spikes in the cosine fits. In contrast, the monomial + ground truth dictionary produced models that both visually resemble the Levy function, and achieve the smallest errors. It makes sense that if your data comes from a function, and you somehow happened to try model fitting with that same function (maybe good luck) you could achieve results like those seen here.

Acknowledgements

The author would like to thank Peter tremendously for helping me double check my numerical results, it was an honor to be assisted by him. Special thanks should be given also to Nash Ward for his advice about plotting, and also to Bamdad Hosseni for help with generating ideas for this report.

References

[1] W. contributors. Regularized least squares. https://en.wikipedia.org/wiki/Regularized%20least%20squares, 2025. [Online; accessed 15-April-2025].

[1]