Research Statement

Chunyang Liao

1 Introduction

My research interests are Mathematical Data Science, Approximation Theory, Optimization, Deep Learning Theory and in particular their intersections. Currently, I am working on Optimal Recovery, which is a subfield of Approximation Theory. The task of optimal recovery is learning a function deterministically from observational data by adopting a worst-case perspective tied to an explicit model assumption made on the functions to be recovered. The goal of my previous work is to construct efficient algorithms to find a best approximant minimizing the worst-case error to the unknown function.

In the optimal recovery scenario, data come as

$$y_i = f(\mathbf{x}_i), \quad i \in [1 : m].$$

It is not assumed that $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are independent realizations of a random variable. I think instead of the inputs as fixed entities whose location might be unfavorable. But we still hunt for optimal ways to recovery f or a quantity $Q(f)$ depending on f from the data y_i's. Without underlying probability distribution, the performance of the recovery process cannot be assessed via generalization error. Instead, it is assessed via the notion of worst-case error, central to the theory of Optimal Recovery [Micchelli and Rivlin, 1977]. The optimal recovery framework is attracting attention in both analytical and computational aspects. It has a variety of applications including parametric PDE [Cohen, Dahmen, DeVore, Fadili, Mula, and Nichols, 2020b], data assimilation [Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk, 2017], system identification [Ettehad and Foucart, 2020a], weather forecasting Foucart et al. [2019] and Information-Based Complexity [Traub, 2003, Plaskota, 1996]. It is related to statistical learning and game theory as well, see Owhadi et al. [2019], Owhadi and Scovel [2019], Li et al. [2007].

The history of Optimal Recovery dates back to 1970s. The first survey paper of optimal recovery gave the problem setting of optimal recovery and a lot of examples that can be explained by this framework, see [Micchelli and Rivlin, 1977]. At that point, the existence of linear optimal recovery map had received much attention, see [Packel, 1988], while there was no numerical receipt for constructing the optimal recovery map. With the development of optimization theory, people are able to characterize the best approximant (in the worst-case sense) as the solution to a convex optimization program and the closed form formula is also available for some specific problem setting, such as Hilbert space setting [Foucart, 2022, Chapter 9]. The best way to approximate a quantity of interest $Q(f)$ in the worst case setting can be found in [DeVore, Foucart, Petrova, and Wojtaszczyk, 2019] and [Ettehad and Foucart, 2020b].

2 Background

Echoing the theory of Optimal Recovery, we consider the function f more abstractly as an element from a normed space $\mathcal{F}$. The output data y_i's, which are evaluations of f at the points $\mathbf{x}_i$'s, can be generalized to linear functionals

ℓ_i's applied to f, so that the data take the form

$$y_i = \ell_i(f), \qquad i \in [1:m]. \tag{1}$$

For convenience, we summarize these data as

$$\mathbf{y} = L(f) = [\ell_1(f); \dots; \ell_m(f)] \in \mathbb{R}^m, \tag{2}$$

where the linear map $L : \mathcal{F} \to \mathbb{R}^m$ is called the observation operator. Relevant situations include the case where $\mathcal{F}$ is the space $\mathcal{C}(\Omega)$ of continuous functions on Ω, which is equipped with the uniform norm, and the case where $\mathcal{F}$ is a Hilbert space $\mathcal{H}$, which is equipped with the norm derived from its inner product.

2.1 Worst-case errors

The worst-case error is central to the optimal recovery problem. I will discuss the worst-case error thoroughly in this subsection. Our goal is to approximate f in full, or a quantity $Q(f)$ such as its integral. Despite the general setting allowing for arbitrary quantities of interest $Q : \mathcal{F} \to Z$, we keep in mind the particular case where $Q = \mathrm{Id}_{\mathcal{F}}$ and where Q is a linear functional. The approximant to $Q(f)$ is obtained by applying a recovery map $R : \mathbb{R}^m \to Z$ to $\mathbf{y} = L(f) \in \mathbb{R}^m$. Given a model set $\mathcal{K}$ and observation map L, we turn to assessing the performance of the recovery maps R. For a single $f \in \mathcal{K}$, the error incurred by approximating $Q(f)$ by $R(\mathbf{y})$, $\mathbf{y} = L(f)$, is evidently $\|Q(f) - R(L(f))\|$. For all $f \in \mathcal{K}$, one adopts a worst-case viewpoint and defines the *local* worst-case error of R over $\mathcal{K}$ as

$$\mathrm{err}_{\mathcal{K}}^{\mathrm{loc}}(L, R(\mathbf{y})) := \sup_{f \in \mathcal{K}, L(f) = \mathbf{y}} \|f - R(\mathbf{y})\|_{\mathcal{F}}. \tag{3}$$

Our objective consists in finding a *local* optimal recovery map $R^{opt} : \mathbb{R}^m \to \mathcal{F}$ sending $\mathbf{y} \in \mathbb{R}^m$ to an element $R^{opt}(\mathbf{y}) \in \mathcal{F}$ that minimizes the local worst-case error $\mathrm{err}_{\mathcal{K}}^{\mathrm{loc}}(L, R(\mathbf{y}))$. Such an element $R^{opt}(\mathbf{y}) \in \mathcal{F}$ can be described, almost tautologically, as a center of a smallest ball containing $\mathcal{K} \cap L^{-1}(\{y\})$, called a Chebyshev center of this set of model- and data-consistent elements. This remark, however, does not come with any practical construction of a Chebyshev center.

The term *local* is used to make a distinction with the more traditional *global* worst-case error of the recovery map $R : \mathbb{R}^m \to \mathcal{F}$, as given by

$$\mathrm{err}_{\mathcal{K}}^{\mathrm{glo}}(L, R) := \sup_{f \in \mathcal{K}} \|f - R(L(f))\|_{\mathcal{F}} = \sup_{y \in L(\mathcal{K})} \sup_{f \in \mathcal{K}, L(f) = \mathbf{y}} \|f - R(\mathbf{y})\|_{\mathcal{F}} = \sup_{y \in L(\mathcal{K})} \mathrm{err}_{\mathcal{K}}^{\mathrm{loc}}(L, R(\mathbf{y})). \tag{4}$$

The minimal value of $\mathrm{err}_{\mathcal{K}}^{\mathrm{glo}}(L, R)$ is called the intrinsic error (of the observation map L over the model set $\mathcal{K}$) and the maps R that achieve this minimal value are called globally optimal recovery maps. Our objective consists in constructing such maps—of course, the map that assigns to $\mathbf{y}$ a Chebyshev center of $\mathcal{K} \cap L^{-1}(\{y\})$ is one of them, but in may be impractical. By contrast, for model sets that are convex and symmetric, the existence of linear maps among the set of globally optimal recovery maps is guaranteed by fundamental results from Optimal Recovery in at least two settings: when F is a Hilbert space and when F is an arbitrary normed space but the full recovery of f gives way to the recovery of a quantity of interest $Q(f)$, Q being a linear functional.

2.2 Model set

My research concentrates on an approximation-based model set that is increasingly scrutinized, see e.g. [Maday, Patera, Penn, and Yano, 2015], [DeVore, Petrova, and Wojtaszczyk, 2017] and [Cohen, Dahmen, Mula, and Nichols, 2020a]. One of the notable advantages of this new model set is that it admits linear optimal recovery maps and offers ways to efficiently compute them.

We first note that the prior information is necessary to say meaningful results about f. For example, one could think of all ways to fit a univariate function through points $(x_1, y_1), \ldots, (x_m, y_m) \in \mathbb{R}^2$ if no restriction is imposed. Thus, a model assumption for the functions of interest is needed. This assumption takes the form $f \in \mathcal{K}$, where the model set $\mathcal{K}$ translates an educated belief about the behavior of realistic functions f.

To put things in context, the prior assumption that f is smooth is often made in Optimal Recovery. The smoothness is typically characterized by a model set $\mathcal{K}$ chosen as the unit ball of Lipschitz, Sobolev, or Besov spaces, see e.g. [Wahba, 1990] and [Melkman and Micchelli, 1979]. Motivated by the rough equivalence between smoothness and approximability, I focus on the model set has form (given approximation parameter $\epsilon > 0$)

$$\mathcal{K} := \{f \in \mathcal{F} : \mathrm{dist}(f, V) \leq \epsilon\}, \tag{5}$$

where V is a linear subspace of $\mathcal{F}$ and $\mathrm{dist}(f, V) := \inf\{\|f - v\|_{\mathcal{F}}, v \in V\}$. In my recent work, I deal with the case where $\mathcal{F} = \mathcal{H}$ is a Hilbert space. Therefore, the model set $\mathcal{K}$ can be written as

$$\mathcal{K} := \{f \in \mathcal{H} : \|f - \mathcal{P}_V(f)\| \leq \epsilon\}, \tag{6}$$

where $\mathcal{P}_V$ is the orthogonal projection onto linear subspace V. It is implicitly assumed that $V \cap \ker(L) = \{0\}$, otherwise the worst-case error goes to infinity. By a dimension argument, the implicit assumption forces $n := \dim(V) \leq m$ i.e., we must place ourselves in an underparametrized regime where there are less model parameters than datapoints. The over-parametrization regime for linear functional Q is studied in [Foucart, 2021]. It's worth to point out that our relevant modeling assumption occurs implicitly in machine learning, namely that the functions of interest are well-approximated by suitable hypothesis classes.

3 Prior work and work in progress

3.1 Full optimal recovery in Hilbert space [Foucart, Liao, Shahrampour, and Wang, 2020]

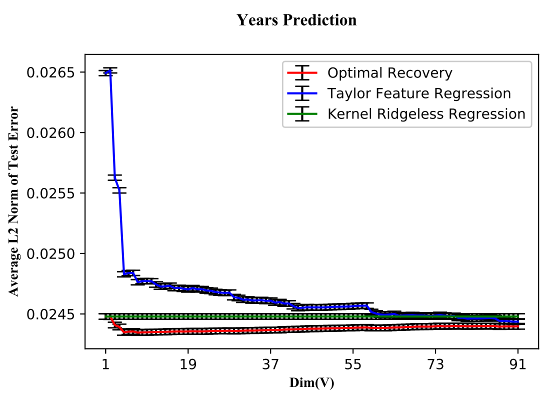
In this paper, we put ourselves into Hilbert space setting and focus on finding the local optimal recovery map. More precisely, we concentrate on the Reproducing Kernel Hilbert Space (RKHS) where the point evaluations at the $\mathbf{x}_i$'s are indeed well-defined and continuous linear functionals on $\mathcal{H}$. Our goal is comparing our framework with traditional kernel method and random feature method in terms of local worst-case error and test mean squared error.

Under the general Hilbert space setting with model set (6), we try to find optimal recovery map R^{opt} such that the local worst-case error is minimized. The solution to the program

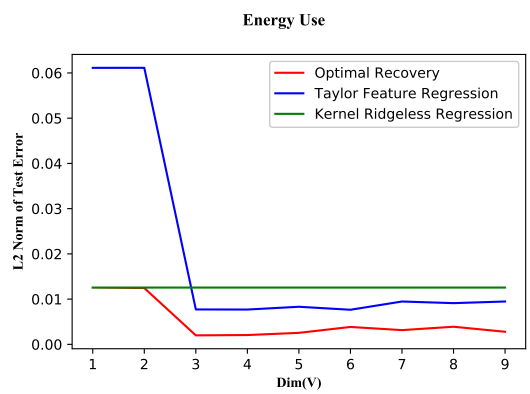
$$\underset{f\in\mathcal{H}}{\text{minimize}}\,\|P_{V^\perp}f\| \quad \text{s.to} \quad Lf=y$$

is proved to be the linear locally optimal recovery map. It also provides an efficient algorithm to compute the Chebyshev center. In addition, the closed form formula for the Chebyshev center is also given, which allows us to deal with infinite-dimensional Hilbert space.

To compare with kernel method and random feature method, we consider the Gaussian kernel and its corresponding Reproducing Kernel Hilbert Space(RKHS). We first point out an interesting relation that he kernel interpolation scheme gives the smallest worst-case error if $V=\text{span}\{u_i, i\in I\}$ for some subset I of $[1:m]$, where u_i are the Riesz representers of the linear functionals $\ell_i\in\mathcal{H}^*$ (notice that $u_i(x)=K_{\mathbf{x}_i}(x)=K(\mathbf{x}_i,x)$ in RKHS). Concerning local worst-case error, we provide a way to compute it for any recovery map thanks to S-lemma (S-procedure) [Pólik and Terlaky, 2007]. Without any doubt, numerical experiments showed that the locally optimal Recovery map (solution to (3.1)) achieves smallest worst-case error. Surprisingly, numerical results (see Figure 1) indicated that our designed locally optimal recovery map has the potential to outperform kernel ridgeless regression and kernel random feature method in the test mean squared error if we choose V as the span of "Taylor features" used in random feature approximation and the data are not independent.



(a) Test error comparison on Years Prediction



(b) Test error comparison on Energy Use

Figure 1: Optimal Recovery and two benchmark regression algorithms on two benchmark datasets. We sort data w.r.t one feature to make data related.

3.2 Optimal recovery from inaccurate data in Hilbert space [Foucart and Liao, 2021]

In this subsection, we consider the case where data comes with observation errors

$$\mathbf{y}=L(f)+e\in\mathbb{R}^m.$$

where the error vector e is not modeled as random noise but through the deterministic ℓ_2-bound $\|e\|_2\le\eta$. We still wish to approximate f by some $\hat{f}\in\mathcal{H}$ (i.e. $Q=\text{Id}_{\mathcal{H}}$). The worst-case errors (3) and (4) need to be adjusted due to bounded observation errors. The *local* worst-case error at $\mathbf{y}$ for $\hat{f}$ becomes

$$\text{lwce}(\mathbf{y}, \hat{f}) := \sup_{\substack{\|f - P_V f\| \leq \epsilon \\ \|L(f) - \mathbf{y}\| \leq \eta}} \|f - \hat{f}\|. \tag{7}$$

As for the *global* worst-case error of $\Delta : \mathbb{R}^m \to \mathcal{H}$, it reads

$$\text{gwce}(\Delta) := \sup_{\substack{\|f - P_V f\| \leq \epsilon \\ \|e\| \leq \eta}} \|f - \Delta(L(f) + e)\|. \tag{8}$$

Earlier works has demonstrated that the following unconstrained regularization program

$$\underset{f \in \mathcal{H}}{\text{minimize}} \quad (1 - \tau)\|f - P_V f\|^2 + \tau\|Lf - \mathbf{y}\|^2, \quad \text{for some } \tau \in [0, 1] \tag{9}$$

provide algorithms that are optimal in both local and global setting. However, the choice of regularization parameter τ is neither clear nor computationally cheap in both local and global setting. The precise choice of regularization parameter τ is the purpose of our work.

[Beck and Eldar, 2007] gives a way to compute the regularization parameter τ. However, it relies on an extension of the S-lemma involving two quadratic constraints. This extension is valid in the complex finite-dimensional setting, but not necessarily in the real setting. Assuming orthonormal observations $L^*L = \text{Id}$, we construct the locally optimal recovery map in real Hilbert spaces by giving the hyperparameter τ implicitly. It's worth to point out the nonlinearity of the local optimal recovery map that sends $\mathbf{y} \in \mathbb{R}^m$ to the best approximant $\hat{f}$, aka Chebyshev center.

The fact that regularization produces globally optimal recovery maps was recognized by [Melkman and Micchelli, 1979, Micchelli, 1993]. However, a recipe for selecting the parameter was not given, except on a specific example. A global optimal recovery map is provided by the linear map sending $\mathbf{y} \in \mathbb{R}^m$ to the minimize of (9) with parameter $\tau = d/(c + d)$, where c, d are solutions to the semidefinite program

$$\underset{c,d \geq 0}{\text{minimize}} \quad \epsilon^2 c + \eta^2 d \qquad \text{s.to} \quad cP_{V^\perp} + dL^*L \succcurlyeq \text{Id}.$$

A surprise arises when we assume orthonormal observation operator L. We proved that regularization maps are globally optimal no matter how the parameter $\tau \in [0, 1]$ is chosen.

4 Future Plans

I am always open to new research areas related to Approximation Theory, mathematical data science and Optimization. I also consider the extension of my current works.

- Local optimal recovery with arbitrary observations. This problem is tackled under a complex finite-dimensional Hilbert space assumption. I plan to consider this problem in any Hilbert spaces. The result can be applied to two space problem mentioned in [Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk, 2017] and [Cohen, Dahmen, DeVore, Fadili, Mula, and Nichols, 2020b].
- Nonlinear approximation space V. I am interested in nonlinear approximated space V that has strong approxi-

mation power, such as the function space generated by Deep Neural Networks with ReLu activation function.

- Over-parametrization regime. The full recovery problem in the over-parametrization regime is more general but still uncovered.

References

Amir Beck and Yonina C Eldar. Regularization in regression with bounded noise: A Chebyshev center approach. *SIAM Journal on Matrix Analysis and Applications*, 29(2):606–625, 2007.

Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1–29, 2017.

A. Cohen, W. Dahmen, O. Mula, and J. Nichols. Nonlinear reduced models for state and parameter estimation. *arXiv preprint arXiv:2009.02687*, 2020a.

Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Jalal Fadili, Olga Mula, and James Nichols. Optimal reduced model algorithms for data-based state estimation. *SIAM Journal on Numerical Analysis*, 58(6):3355–3381, 2020b. doi: 10.1137/19M1255185.

R. DeVore, S. Foucart, G. Petrova, and P. Wojtaszczyk. Computing a quantity of interest from observational data. *Constructive Approximation*, 49:461–508, 2019.

Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Data assimilation and sampling in banach spaces. *Calcolo*, 54(3):963–1007, 2017.

Mahmood Ettehad and S. Foucart. Approximability models and optimal system identification. *Mathematics of Control, Signals, and Systems*, 32:19–41, 2020a.

Mahmood Ettehad and Simon Foucart. Instances of computational optimal recovery: dealing with observation errors. *arXiv preprint arXiv:2004.00192*, 2020b.

S. Foucart. *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press, 2022.

S. Foucart, M. Hielsberg, G. Mullendore, G. Petrova, and P. Wojtaszczyk. Optimal algorithms for computing average temperatures. *Mathematics of Climate and Weather Forecasting*, 5:34–44, 02 2019. doi: 10.1515/mcwf-2019-0003.

Simon Foucart. Instances of computational optimal recovery: refined approximability models. *Journal of Complexity*, 62:101503, 2021.

Simon Foucart and Chunyang Liao. Optimal recovery from inaccurate data in hilbert spaces: Regularize, but what of the parameter? *arXiv preprint arXiv:2111.02601*, 2021.

Simon Foucart, Chunyang Liao, Shahin Shahrampour, and Yinsong Wang. Learning from non-random data in hilbert spaces: An optimal recovery perspective. *arXiv preprint arXiv:2006.03706*, 2020.

Wenye Li, Kin-Hong Lee, and Kwong-Sak Leung. Generalized regularized least-squares learning with predefined features in a Hilbert space. In *Advances in Neural Information Processing Systems*, pages 881–888, 2007.

Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.

Avraham A. Melkman and Charles A. Micchelli. Optimal estimation of linear operators in hilbert spaces from inaccurate data. *SIAM Journal on Numerical Analysis*, 16(1):87–105, 1979.

C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.

Charles A Micchelli. Optimal estimation of linear operators from inaccurate data: a second look. *Numerical Algorithms*, 5(8):375–390, 1993.

H. Owhadi, C. Scovel, and F. Schäfer. Statistical numerical approximation. *Notices of the American Mathematical Society*, 66:1608–1617, 2019.

Houman Owhadi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design.* Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019. doi: 10.1017/9781108594967.

Edward W Packel. Do linear problems have linear optimal algorithms? *SIAM Review*, 30(3):388–403, 1988.

Leszek Plaskota. *Noisy information and computational complexity*, volume 95. Cambridge University Press, 1996.

Imre Pólik and Tamás Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3):371–418, 2007.

Joseph F Traub. *Information-Based Complexity.* John Wiley and Sons Ltd., 2003.

Grace Wahba. *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, 1990. doi: 10.1137/1.9781611970128. URL `https://epubs.siam.org/doi/abs/10.1137/1.9781611970128`.