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, whose main task is to approximate a function deterministically from observations by adopting a worst-case perspective tied to an explicit model assumption made on the function to be recovered. A function f is acquired through point evaluations

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

and these data should be used to recover f. Importantly, the evaluation points are considered fixed entities in Optimal Recovery: they cannot be chosen in a favorable way, as in Information-Based Complexity [Traub, 2003], nor do they occur as independent realizations of a random variable, as in Statistical Learning Theory [Hastie, Tibshirani, and Friedman, 2009]. In particular, without an underlying probability distribution, the performance of the recovery process cannot be assessed via generalization error. Instead, it is assessed via a notion of worst-case error, central to the theory of Optimal Recovery.

The theory of Optimal Recovery was developed in the 70's-80's as a subfield of Approximation Theory (see the surveys [Micchelli and Rivlin, 1977, 1985]). At that time, researchers focused on the purely analytic approaches, but did not put the emphasize on computational issues. Due to the lack of algorithmic solutions and numerical advances, the developments of Optimal Recovery slowed down and reached their limitations. Recently, researchers accept a more computational perspective by exploiting the full power of modern Optimization Theory, then some new advances occur. For example, what is called the spline algorithm in Optimal Recovery has recently made a reappearance in Machine Learning circles as minimum-norm interpolation [de Boor, 1977, Wahba, 1990], of course with a different motivation. Moreover, Optimal Recovery in Reproducing Kernel Hilbert Spaces is the same as kernel ridgeless regression in Machine Learning [Liang and Rakhlin, 2020]. It also relates to statistical numerical approximation and Gaussian Process [Owhadi et al., 2019]. Other applications of Optimal Recovery include numerical/parametric PDE [Chen et al., 2021, Cohen et al., 2020], system identification [Ettehad and Foucart, 2020], weather forecasting [Foucart et al., 2019] and so on.

2 Optimal Recovery Perspective

In Optimal Recovery, the function f is considered 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 map**.

Data itself is not enough to produce a meaningful reconstruction of f. Therefore, a model assumption is necessary, which is an educated belief about properties of realistic functions f. My research concentrates on a new approximation-based model, which is motivated by the rough equivalence between smoothness and approximability. The **model set** $\mathcal{K}$ takes form (given approximation parameter $\epsilon > 0$)

$$\mathcal{K} := \{f \in \mathcal{F} : \mathrm{dist}(f, V) \leq \epsilon\}, \tag{3}$$

where V is a linear subspace of $\mathcal{F}$ and $\mathrm{dist}(f, V) = \inf_{v \in V} \|f - v\|_{\mathcal{F}}$. This kind of model set was scrutinized, see e.g. [Maday, Patera, Penn, and Yano, 2015] and [DeVore, Petrova, and Wojtaszczyk, 2017]. I consider this kind of model set because it not only characterizes the smoothness of the function but also introduces computational advantages.

A **recovery map** $\Delta : \mathbb{R}^m \to Z$ produces an approximant $\Delta(\mathbf{y})$ of f. The *local* worst-case error of Δ at any given $\mathbf{y} \in \mathbb{R}^m$ is defined as

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

The term *local* is used above to make a distinction with the *global* worst-case error, defined as

$$\mathrm{err}^{\mathrm{glo}}_{\mathcal{K}}(L, \Delta) := \sup_{y \in L(\mathcal{K})} \mathrm{err}^{\mathrm{loc}}_{\mathcal{K}}(L, \Delta(\mathbf{y})) = \sup_{f \in \mathcal{K}} \|f - \Delta(L(f))\|_{\mathcal{F}}. \tag{5}$$

The map Δ is called locally/globally optimal recovery map if it minimizes local/global worst-case error, respectively. One of my goal is to design tractable algorithms to find both locally and globally optimal recovery maps.

3 Prior work

My work focused on the case where $\mathcal{F} = \mathcal{H}$ is a Hilbert space. Therefore, the model set $\mathcal{K}$ can be written as

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

where $\mathcal{P}_V$ is the orthogonal projection onto linear subspace V. To get bounded worst-case error, it is implicitly assumed that $V \cap \ker(L) = \{0\}$. By a dimension argument, the implicit assumption forces $n := \dim(V) \le m$. In other words, we must place ourselves in an underparametrized regime.

3.1 Optimal Recovery in Hilbert space from exact Observation

This subsection summarizes my work accepted by Sampling Theory, Signal Processing, and Data Analysis [Foucart, Liao, Shahrampour, and Wang, 2022]. We constructed the locally optimal recovery map (hence globally optimal recovery map) and provided an explicit formula, which allows us to deal with infinite-dimensional Hilbert space. In addition, We provided a way to compute the worst-case error in finite-dimensional Hilbert space as well. To the best of my knowledge, it is the first place to give an exact computation of the worst-case error. Finally, we shed light on the relationship between Optimal Recovery and Kernel Method.

Under the general Hilbert space setting with model set (6), the locally optimal recovery map $R^{\text{opt}}(\mathbf{y})$ is the solution to the program

$$\underset{f \in \mathcal{H}}{\text{minimize}} \, \|P_{V^\perp} f\| \quad \text{s.to} \quad Lf = \mathbf{y}. \tag{7}$$

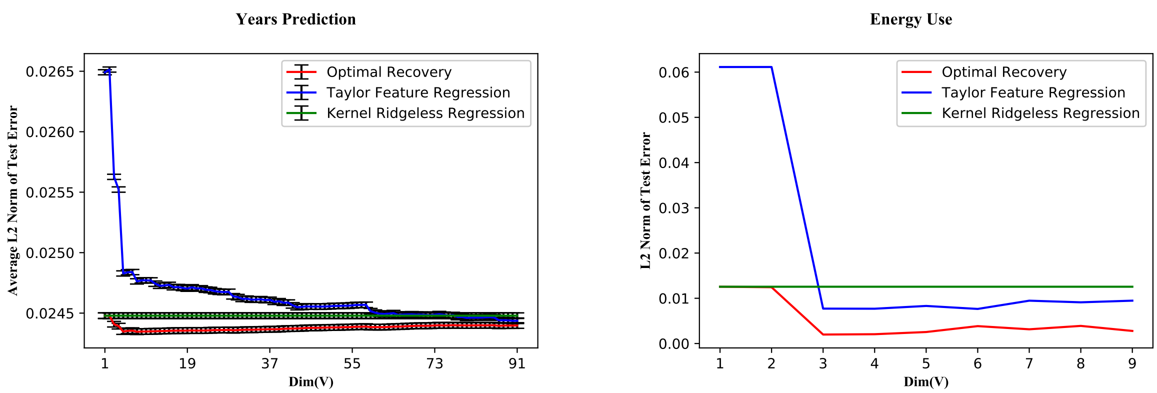
Since we are in Hilbert spaces and ℓ_i's are linear functionals on $\mathcal{H}$, by Riesz Representers Theorem, there are $u_i \in \mathcal{H}$ such that $\ell_i(f) = \langle u_i, f\rangle$ for all $i \in [1:m]$. Let $\{v_1, \dots, v_n\}$ be a basis of linear space V, we define matrix $\mathbf{G} \in \mathbb{R}^{m\times m}$ such that $\mathbf{G}_{i,i'} = \langle u_i, u_{i'}\rangle$ and matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ such that $\mathbf{C}_{i,j} = \langle u_i, v_j\rangle$. The closed form of $R^{\text{opt}}(\mathbf{y})$ is

$$R^{\text{opt}}(\mathbf{y}) = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j, \tag{8}$$

where coefficient vectors $\mathbf{b} = (\mathbf{C}^\top \mathbf{G}^{-1}\mathbf{C})^{-1}\mathbf{C}^\top \mathbf{G}^{-1}\mathbf{y}$ and $\mathbf{a} = \mathbf{G}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b})$. Apparently, the closed form formula indicates that the locally optimal recovery map is linear.

To compute the worst-case error, we first note that the optimization problem (4) is a Quadratic Optimization with one quadratic constraints when $\mathcal{H}$ is a finite dimensional Hilbert space. Then we use S-lemma [Pólik and Terlaky, 2007] to write it as a semidefinite program.

There is an interesting relation between kernel interpolation and optimal recovery. Specifically, if we are in a Reproducing Kernel Hilbert space with $V = \text{span}\{u_i, i \in I\}$ for some subset I of $[1:m]$ (notice that $u_i(x) = K_{\mathbf{x}_i}(x) = K(\mathbf{x}_i, x)$ in RKHS), the program (7) is indeed kernel ridgeless regression (kernel interpolation) scheme. 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 Spaces

In this subsection, I try to generalize the Optimal Recovery framework by integrating ℓ_2 bounded observation error. This work has been accepted by Constructive Approximation [Foucart and Liao, 2021].

We consider the case where data comes with observation errors, i.e. $\mathbf{y} = L(f) + \mathbf{e} \in \mathbb{R}^m$. The error vector $\mathbf{e}$ is not modeled as random noise but through the deterministic ℓ_2-bound, i.e. $\|\mathbf{e}\|_2 \leq \eta$. The worst-case errors (4) and (5) need to be adjusted to

$$\text{lwce}(\mathbf{y}, \Delta) := \sup_{\substack{\|f - P_V f\| \leq \epsilon \\ \|L(f) - \mathbf{y}\| \leq \eta}} \|f - \Delta(\mathbf{y})\|, \tag{9}$$

$$\text{gwce}(\Delta) := \sup_{\substack{\|f - P_V f\| \leq \epsilon \\ \|\mathbf{e}\| \leq \eta}} \|f - \Delta(L(f) + \mathbf{e})\|. \tag{10}$$

Earlier works has demonstrated that the following unconstrained regularization program

$$\underset{f\in\mathcal{H}}{\text{minimize}} \quad (1-\tau)\|f-P_V f\|^2 + \tau\|Lf-\mathbf{y}\|^2, \quad \text{for some } \tau\in[0,1] \tag{11}$$

provide algorithms that are optimal in both local and global setting. In the local setting, Beck and Eldar [2007] constructed the locally optimal recovery map by applying an extension of the S-lemma involving two quadratic constraints. However, this extension is valid only in the complex finite-dimensional Hilbert space. In the global setting, regularization produces globally optimal recovery maps was recognized by [Melkman and Micchelli, 1979, Micchelli, 1993]. Nevertheless, a recipe for selecting the parameter was not given, except on a specific example.

We give a full picture of selecting the regularization parameter in both local setting and global setting. In the local setting, we constructed the locally optimal recovery map in real Hilbert spaces by assuming orthonormal observations L, i.e. $L^*L = \mathrm{Id}$ (L^* is the adjoint operator of L). The optimal regularization parameter $\tau_\sharp$ is obtained by solving the following equation,

$$\lambda_{\min}((1-\tau)P_{V^\perp}+\tau L^*L) - \frac{(1-\tau)^2\epsilon^2-\tau^2\eta^2}{(1-\tau)\epsilon^2-\tau\eta^2+(1-\tau)\tau(1-2\tau)\delta^2} = 0, \tag{12}$$

where $\lambda_{\min}$ means the smallest eigenvalue and δ is precomputed as $\delta = \min\{\|P_{V^\perp}f\| : Lf = \mathbf{y}\} = \min\{\|Lf-\mathbf{y}\| : f\in V\}$. This equation can be solved easily by using the bisection method or Newton/secant method. It is worth to point out that the local optimal recovery map is not linear since the computation of $\tau_\sharp$ does depend on observation vector $\mathbf{y}$.

In contrast, the globally optimal recovery map is linear. It is one of several globally optimal recovery maps, since the locally optimal one (which is nonlinear) is also globally optimal. We proved that the optimal regularization parameter $\tau = d/(c+d)$, where c, d are solutions to the semidefinite program

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

where the constraint $cP_{V^\perp} + dL^*L \succcurlyeq \mathrm{Id}$ reads as the matrix $cP_{V^\perp} + dL^*L - \mathrm{Id}$ is a positive semidefinite matrix. Since the optimal regularization parameter does not depend on $\mathbf{y}$, then we can say that the globally optimal recovery map is linear. In addition, this semidefinite program can be solve in an offline stage, which is another computational advantage.

Finally, a surprise arises when we assume orthonormal observation operator L in the global setting. The regularization parameter does not need to be chosen with care after all, since regularization maps are globally optimal no matter how the parameter $\tau\in[0,1]$ is chosen.

4 Ongoing Projects and Future Plans

I am trying to extend my current works in the following directions:

- **Full recovery from other types of error**. It's still unclear how to design optimal recovery algorithm if other types of observation error are considered.
- **Quantity of Interest** $Q(f)$. We can also recovery a quantity $Q(f)$ related to f, such as the integral of f over domain. When Q is a linear functional of f, approximation of $Q(f)$ from exact observational data and noisy data were addressed in [DeVore et al., 2019] and [Ettehad and Foucart, 2021], respectively. But a more general linear quantity of interest Q was not considered.
- **Multispace Problem**. Besides a single linear approximation space V, we can take a nested sequence of approximation space $V_0 \subset V_1 \subset \cdots \subset V_k$ and define model set $\mathcal{K}$ as

$$\mathcal{K} = \bigcap_{i=0}^{k} \mathcal{K}_k = \bigcap_{i=0}^{k} \{f \in \mathcal{F} : \mathrm{dist}(f, V_k) \leq \epsilon_k\},$$

  where approximation parameters satisfy $\epsilon_0 \geq \epsilon_1 \geq \cdots \geq \epsilon_k$. This new model set is motivated from reduced modeling in parametric partial differential equation and can be used to incorporate over-parametrization regime, see Foucart [2021] for details. I plan to consider the full recovery problem under this new model set. Numerical results indicate that regularization program produces a globally optimal recovery map but the choice of regularization parameter is still not clear.
- **Non-linear Approximation Space** V. The linear space V is selected in (3) because it brings computational advantages. However, it does not have strong approximation capability. Therefore, a non-linear space V, such as the function space generated by deep-neural-network, is more favorable.
- **Applications of Optimal Recovery framework**. The optimal recovery framework has been applied in many areas, such as signal processing Shenoy and Parks [1992], image processing Muresan and Parks [2001], robust estimation Tempo [1987] and time series prediction Vicino et al. [1987]. Due to the lack of computational methods, only limited cases can be computed efficiently. With the development of the computational aspect, I believe that we can explore more applications of Optimal Recovery.

I am also open to new research topics related to Mathematical Data Science, Approximation Theory and Deep Learning. Some old projects I did include the approximation power of Deep Neural Network, semi-definite programming in approximation theory and communication-efficient distributed gradient clipping algorithm.

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.

Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear pdes with gaussian processes. *J. Comput. Phys.*, 447:110668, 2021.

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, 2020.

Carl de Boor. Computational aspects of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 69–91. Springer, 1977.

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.

M. Ettehad and S. Foucart. Instances of computational optimal recovery: dealing with observation errors. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1438–1456, 2021.

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

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

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

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.

S. Foucart, C. Liao, S. Shahrampour, and Y. Wang. Learning from non-random data in hilbert spaces: An optimal recovery perspective. *Sampling Theory, Signal Processing, and Data Analysis*, 20, 2022.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Springer, second edition, 2009.

Tengyuan Liang and Alexander Rakhlin. Just interpolate: Kernel "Ridgeless" regression can generalize. *The Annals of Statistics*, 48(3):1329 – 1347, 2020.

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.

C. A. Micchelli and T. J. Rivlin. Lectures on optimal recovery. In *Numerical Analysis Lancaster 1984*, pages 21–93. Springer, 1985.

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

D.D. Muresan and T.W. Parks. Optimal recovery approach to image interpolation. In *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, volume 3, pages 848–851 vol.3, 2001. doi: 10.1109/ICIP.2001.958253.

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

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

R.G. Shenoy and T.W. Parks. An optimal recovery approach to interpolation. *IEEE Transactions on Signal Processing*, 40(8):1987–1996, 1992. doi: 10.1109/78.150000.

Roberto Tempo. Robust estimation and filtering in the presence of bounded noise. In *26th IEEE Conference on Decision and Control*, volume 26, pages 1614–1615, 1987. doi: 10.1109/CDC.1987.272715.

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

A. Vicino, R. Tempo, R. Genesio, and M. Milanese. Optimal error and gmdh predictors: A comparison with some statistical techniques. *International Journal of Forecasting*, 3(2):313–328, 1987. ISSN 0169-2070. doi: https://doi.org/10.1016/0169-2070(87)90012-4. URL `https://www.sciencedirect.com/science/article/pii/0169207087900124`.

Grace Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.