Machine Learning for Risk Management: Kernels with Shape Constraints (Summary)

Kernel methods and the associated rich function class called reproducing kernel Hilbert space (RKHS) are among the most flexible tools in machine learning (ML) and statistics. In our project we designed novel kernel-based techniques and investigated their practical efficiency. The constructed methods are

1. capable of ensuring hard shape constraints in RKHSs in a principled way with theoretical guarantees,
2. applicable in portfolio optimization with kernel-based divergence and goodness-of-fit measures,
3. able to solve infinitely many related tasks by exploiting their dependencies (such as conditional quantile estimation for all quantiles simultaneously) using vector-valued RKHSs.

Our methods were developed in line with finance and economics, and were tested both in these areas and beyond to explore their robustness. Below a brief summary of our results is provided.

- **Hard shape constraints in RKHSs:** Shape constraints (such as non-negativity, monotonicity, convexity or sub-modularity) form the basis of numerous successful applications in ML, finance and economics. However enforcing these shape requirements in a *hard* way (for instance at all points of an interval) is a rather challenging problem. We proved that with second-order cone programming it is possible to encode affine shape constraints on function values and on function derivatives in a hard fashion in RKHSs. The solution is capable of handling (i) both real-valued and matrix-valued kernels[1], (ii) SDP (semidefinite programming) constraints[2], (iii) generalized coverings expressible in terms of balls and halfspaces. We demonstrated the efficiency of the proposed technique in (i) joint quantile regression with applications to economics and the analysis of aircraft trajectories, (ii) estimation of production functions, (iii) safety-critical control, (iv) vehicle trajectory reconstruction from noisy location observations, and (v) shape optimization (catenary problem). Our contributions with real-valued kernels were published at NeurIPS (the most prestigious ML conference) [5], the extended results which allow matrix-valued kernels and adaptive coverings have been submitted to JMLR (the largest ML journal) [4]. In addition, we disseminated our results at Eurecom [8], SMAI-MODE [9], SPIGL [10], EUROPT [11], Aix-Marseille University [12], TAMU [13], MMS [14] and IFAC WC [15].

[1]These choices correspond to performing approximation with real-valued and vector-valued functions, respectively.

[2]SDP constraints include for instance convexity on a subset of the variables.

- **Kernel minimum divergence portfolios:** Portfolio optimization is a fundamental problem in finance, with a recently proposed exciting approach called distribution targeting. The idea of distribution targeting is to optimize the portfolio to match the investor's preference expressed via a target distribution of returns. Unfortunately, existing methods in this direction (based on Kullback-Leibler and f-divergences) suffer from computational and modeling bottlenecks. To mitigate these severe shortcomings, we proposed to use alternative divergence measures based on kernels, maximum mean discrepancy (MMD) and kernel Stein discrepancy (KSD). On the
  - theoretical front: We computed analytically the mean embedding (the underlying representation of probability measures used in MMD) for various new target distribution - kernel pairs. We proved that such analytical knowledge leads to better concentration properties of MMD estimators, extended the result to the case of unbounded kernels, and derived minimax lower bounds.
  - practical side: We performed a wide range of numerical experiments to demonstrate the practical efficiency of the proposed semi-analytic MMD and KSD estimators on both synthetic and real-world examples. Our numerical results show that the improved concentration properties of the semi-analytic estimators translate to less noisy and more pronounced U-shape of the objective functions, and hence easier solution when used in portfolio optimization. According to our real-world experiments on portfolio benchmarks from the Kenneth French's library the proposed technique (i) performs favorably compared to classical portfolio optimization schemes on balanced portfolios in normal economic conditions, (ii) MMD and KSD with light-tailed targets often improve the out-of-sample skewness and kurtosis of the portfolio. In addition, the adapted cross-entropy optimization technique allows handling various constraints (such as budget or budget with non-negativity) naturally on the portfolio weights.

  Our work has been submitted to the SIAM Journal on Financial Mathematics [2].

- **Solving infinitely many related tasks:** We showed how to tackle the learning problem of infinitely many tasks where the output is a real-valued function over the hyperparameter space in a nonparametric fashion using operator-valued kernels; examples include quantile regression, cost-sensitive classification, and density level set estimation. We extended the optimization methodology (both in the primal and the dual domain) for the vector-valued case and for learning with integral losses. We demonstrated the efficiency of the approach in style transfer at NeurIPS (CtrlGen21, [3]). In addition, we disseminated our results at LIKE [6] and at CMStatistics [7].

  We provided (submitted to AISTATS-2022, [1]) a flexible optimization framework dedicated to robustness and sparsity when the target variable is function-valued. Particularly, we extended existing results in the literature to a wider class of loss functions, by considering specific p-norms in the output space $L^2[\Theta, \mu]$ where Θ denotes the domain of hyperparameters and the probability measure μ captures the importance of the different hyperparameter values. To handle these losses, we proposed novel optimization algorithms based on a linear spline representation of the dual variables in case of $p \neq 2$, and revisited the case $p = 2$ of available techniques by proposing a different scheme for approximating the integral operator. The latter scheme widens significantly the class of applicable kernels by not requiring analytical access to the eigensystem of the integral operator associated to the kernel. We contrasted and demonstrated the practical efficiency of the designed new methods for functional output regression over the classical setting of squared loss.

References

[1] Alex Lambert, Dimitri Bouche, Zoltán Szabó, and Florence d'Alché-Buc. Functional output regression with infimal convolution: Exploring the Huber and ϵ-insensitive losses. In *Artificial Intelligence and Statistics Conference (AISTATS)*, 2022. (submitted).

[2] Linda Chamakh and Zoltán Szabó. Kernel minimum divergence portfolios. Technical report, 2021. (https://arxiv.org/abs/2110.09516; submitted to SIAM Journal on Financial Mathematics).

[3] Alex Lambert[†], Sanjeel Parekh[†], Zoltán Szabó, and Florence d'Alché-Buc. Emotion transfer using vector-valued infinite task learning. Technical report, 2021. ([†]contributed equally; https://arxiv.org/abs/2102.05075).

[4] Pierre-Cyril Aubin-Frankowski and Zoltán Szabó. Handling hard affine SDP shape constraints in RKHSs. Technical report, 2021. (http://arxiv.org/abs/2101.01519; submitted to Journal of Machine Learning Research).

[5] Pierre-Cyril Aubin-Frankowski and Zoltán Szabó. Hard shape-constrained kernel machines. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 384–395, Vancouver, Canada, 8-10 December 2020. (poster presentation; 20.1% acceptance rate).

Additional Dissemination

[6] Continuous emotion transfer using RKHSs. Lifting Inference with Kernel Embeddings winter school and workshop (LIKE22; Bern, Switzerland), presented by Zoltán Szabó, 12 January 2022.

[7] Vector-valued infinite task learning in style transfer. CMStatistics, Advanced Statistical Methods for High Dimensional Data session, presented by Zoltán Szabó, 19 December 2021.

[8] Shape constraints meet kernel machines. Data Science Seminar at Eurecom, presented by Zoltán Szabó, 4 November 2021.

[9] Hard shape-constrained kernel regression. SMAI-MODE, presented by Pierre-Cyril Aubin-Frankowski, 7-9 September 2020.

[10] Hard shape-constrained kernel regression. Joint Structures and Common Foundations of Statistical Physics, Information Geometry and Inference for Learning (SPIGL), presented by Pierre-Cyril Aubin-Frankowski, 27-31 July 2020.

[11] Kernel regression with hard shape constraints. Workshop on Advances in Convex Optimization (EUROPT), presented by Zoltán Szabó, 8 July 2021.

[12] Vector-valued prediction with RKHSs and hard shape constraints. Computer Science and Systems Laboratory, Aix-Marseille University, presented by Zoltán Szabó, 20 May 2021.

[13] Shape-constrained kernel machines. Department of Statistics, Texas A&M University, presented by Zoltán Szabó, 12 February 2021.

[14] Kernel machines with hard shape constraints. Meeting on Mathematical Statistics (MMS): Robustness and Computational Efficiency of Algorithms in Statistical Learning, presented by Zoltán Szabó, 15 December 2020.

[15] Kernel regression for vehicle trajectory reconstruction under speed and inter-vehicular distance constraints. IFAC World Congress (IFAC WC), presented by Pierre-Cyril Aubin-Frankowski, 11-17 July 2020.