Optimal Transport and Distributional Robustness

Silvana M. Pesenti (University of Toronto),
Brendan Pass (University of Alberta),
Marc Henry (Pennsylvania State University)
Yanqin Fan (University of Washington)

Workshop no: 24w5163
March 24 – March 29, 2024

1 Overview

Optimal transport (OT) and distributional robust optimisation (DRO) provide essential technical tools to all disciplines where stochastic optimisation and robustness are important. In insurance and risk management, computing model-free measures of aggregate risks such as VaR and Expected Short-Fall is of utmost importance and routinely done. With potentially large numbers of different risks, knowledge of the dependence structure is often lacking. As a result, researchers and policy makers rely on the worst risk measures defined as the maximum value of the risk measures over all the joint measures of the individual risks with fixed marginal measures.

In causal inference in economics and biostatistics, distributional treatment effects such as the variance and the proportion of participants who benefit from the treatment depend on the joint distribution of the potential outcomes. Even with ideal randomised experiments such as double-blind clinical trials, such joint distribution is unavailable and as a result, only the lower and upper bounds on them are identified from the sample information. Although seemingly different, these problems share the same mathematical structure, that of optimising the expected value of a functional of multiple variables with fixed marginals, i.e., that of an optimal transport problem. When the marginals are univariate, such optimisation problems have traditionally been handled by using copulas and have explicit solutions for important classes of functionals. However, when the marginals are multivariate, direct computation of the value function is extremely challenging and no significant progress has been made until very recently thanks to the recent breakthroughs in compu-

tational and statistical estimation of the optimal transport problem with possibly multiple marginals.

The workshop brought together researchers at various stages of their career who are involved in recent breakthroughs in the theory, computation and statistical estimation of optimal transport, including multi-marginal and weak optimal transport, as well as Wasserstein distance based distributional robustness, together with researchers involved in recent breakthroughs in causal inference, inequality measurement, partial identification, option pricing, risk and uncertainty management, and robust decision making, that are rendered possible by the new optimal transport and distributional robustness tools.

2 Presentation Highlights

2.1 Keynotes: Notions of robustness and their relation to optimal transport

- In the first keynote of the conference, Jose Blanchet presented an optimal transport (OT) integrated formulation of distributionally robust optimisation (DRO) and statistical robustness. He made a distinction between statistical robustness, concerned with data contamination and small fractions of outliers, and distributional robustness, which is concerned with quantifying model errors under potential changes in the environment and global shifts. He argued that statistically robust procedures are significantly different from distributionally robust ones. In robust statistics, the data is drawn from a contaminated version $\bar{\mathbb{P}}$ of the true original environment $\mathbb{P}$, but the empirical distribution $\bar{\mathbb{P}}_n$ converges to $\bar{\mathbb{P}}$ and the deployment environment $\mathbb{P}^*$ is the same as the original environment $\mathbb{P}$. In DRO, $\bar{\mathbb{P}} = \mathbb{P}$ (no contamination), but the deployment environment $\mathbb{P}^*$ is different from the original environment $\mathbb{P}$, or, alternatively, the empirical distribution $\bar{\mathbb{P}}$ does not converge to the original environment DGP $\mathbb{P}$. In an adversarial approach to computing optimal procedures, robust statistical procedures are solutions of a non convex non smooth optimisation problem (harder game to play) $\max_{\bar{\mathbb{P}} \in \mathcal{B}(\mathbb{P})} \min_\theta \mathcal{R}(\theta, \bar{\mathbb{P}}_n)$; whereas DRO procedures are solutions to a convex optimisation problem (easier game to play) in a game where nature moves last (harder game to win) $\min_{\theta \in \Psi} \max_{\bar{\mathbb{P}} \in \mathcal{B}(\bar{\mathbb{P}}_n)} \mathcal{R}(\theta, \bar{\mathbb{P}})$. In both cases, $\mathcal{B}$ is a Wasserstein ball, and $\mathcal{R}(\theta, \mathbb{P})$ is a risk function of θ indexing the model for the deployment environment, and of the actual DGP $\mathbb{P}$. (Blanchet et al., 2021)

- Xiaohong Chen gave an excellent review of sieve estimation and inference in nonparametric and semiparametric models including likelihood models, moment condition models, and minimum distance models. Applications of optimal transport in economic and statistical models lead to semiparametric models with optimal transport map as a nonparametric function. The methodology reviewed by Xiaohong is potentially useful in developing estimation and inference in models involving the optimal transport map. (Chen, 2006)

- Lars Hansen presented distributionally robust optimisation as a trade-off between "best guesses" and potentially "bad outcomes". He quoted George Box: "Since all

models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about safety from mice when there are tigers abroad." He emphasised the need for an ability to identify the sources of uncertainty that are big and relevant to the decision problem at hand. he classified three types of uncertainty: risk, modelled with a prior, misspecification of the prior (also called ambiguity), and misspecification of the likelihood. He presented an approach to decision making that involves distributionally robust optimisation with respect to a set of prior specifications (around a reference prior) and a set of possible likelihood specifications (around a reference model). He then Discusses applications of this approach to climate change mitigation policy. (Cerreia-Vioglio et al., 2024)

- The keynote of Ludger Rüschendorf was an extensive review from the early works of optimal couplings, to risk bounds under dependence uncertainty, application to worst-case portfolio allocation, and ordering results for risk models. It started with Fréchet-Hoeffding bounds and the early works of bounds on portfolio sums of the Value-at-Risk, to its connection to mass transportation via rearrangement. Next, an overview of risk measure theory, their worst-case and best-case bounds, and optimal couplings. Finally, recent works on improving risk bounds by adding additional structure such as dependence information. (Rüschendorf, 2013)

2.2 Theoretical advances in optimal transport

- Hang Cheung presented work on Hamilton-Jacobi-Bellman equations in Wasserstein space, focusing on establishing well-posedness of viscosity solutions. (Cheung et al., 2023)

- Joshua Hiew discussed a dynamic approach to optimal transport. Beginning with product measure, the solutions to the entropically regularised OT problems were characterised by an ordinary differential equation, where the variable is one over the usual regularisation parameter. A similar approach applies to linearly constrained variants, and resulting numerical simulations were presented. (Joshua Zoen-Git Hiew et al., 2024)

- The presentation of Zhaonan Qu linked the matrix balancing arising in Sinkhorn's algorithm with choice modelling. He showed how results from the choice modelling literature can be used to understand convergence rates of the Sinkhorn algorithm for sparse matrices. (Qu et al., 2023)

- Cale Rankin's presentation featured new results on a Jordan-Kinderlehrer-Otto type schemes for gradient flows with respect to general optimal transport costs. The motivating example for this work was the Bregman divergence, and further results on the geometry induced on the space of probability measures by optimal transport with a Bregman divergence cost were also presented. (Rankin and Wong, 2024)

- Ruodu Wang proposed a refinement of Strassen's Theorem on the almost sure characterisation of the convex order: $X \leq_{cv} Y$ if and only if there is a probability space

and $X' =_d X$ and $Y' =_d Y$ such that $X' = \mathbb{E}[Y'|X']$. The refinement allows the construction to be on the same probability space. Namely $X \leq_{cv} Y$ if and only if $X =_d \mathcal{E}[Y|\mathcal{G}]$ for some σ-field $\mathcal{G}$. This is achieved by constructing a backward Monge map for the corresponding martingale optimal transport problem. (Nutz et al., 2022)

- Leonard Wong presented new work on the adapted Wasserstein distance between Gaussian measures. The adapted Wasserstein distance is a variant of the Wasserstein distance designed for the situation when the marginal measures are temporally separated (so one cannot look into the future). Explicit solutions to the OT problem corresponding to the traditional Wasserstein distance between Gaussian's are well known; Wong and collaborators have obtained explicit solutions between Gaussian's in the adapted case. (Gunasingam and Wong, 2024)

2.3 Applications to portfolio choice and risk management

- Jonathan Ansari's talk presented new rearrangement-based inequality and applications to portfolio allocation. The work assumes that the marginal distributions are given but the dependence structure is only partially known, e.g., via conditionally independent factor models. Of interest are inequalities of risk measures of portfolio sums under different stochastic orderings of the underlying risk factors. (Ansari and Rüschendorf, 2024)

- Carole Bernard presented recent work on cost-efficient portfolio strategies (the cheapest portfolio strategy which results in a given return distribution) under ambiguity leading to robust portfolio choices. Specifically, the interest was on uncertainty sets induced by stochastic integral orders and in particular characterising the "least favourable distribution", that then results in a robust cost efficient portfolio. (Bernard et al., 2022)

- Rui Gao's talk discusses distributional robustness in a dynamic setting, where perturbations from a nominal stochastic process are characterised by causal transport mappings. A recursive reformulation is presented to evaluate the worst-case risk of a random sequence. Furthermore, dynamic programming reformulations are developed to identify the optimal robust decision rule. (Yang et al., 2022)

- The presentation of Tiantian Mao dived into distributional robust model aggregation problem by studying the most "conservative" distribution of an uncertainty set. The most conservative distribution defined as the smallest distribution dominating all distributions in an uncertainty set. This approach results in model aggregation risk measures which are more prudent then worst case risk measures. (Mao et al., 2022)

- Silvana Pesenti's talk argues that in financial and risk management applications, where gains and losses are treated differently, quantification between distributions should not be symmetric. Thus, the work proposes asymmetric divergences based on OT, where the cost function are proper scoring functions from the statistical literature. (Pesenti and Vanduffel, 2023)

- Steven Vanduffel considered the problem of worst-case risk measures when the underlying distribution is only partially known. Examples include the Value-at-Risk with first two moment constraints and distortion risk measures, where the uncertainty sets is described by the 2-Wasserstein distance and fixed first and second moment. Steven further presented applications to multivariate robust portfolio allocation. (Bernard et al., 2023)

- Yao Xie's presentation was on the computation of Wasserstein distributional robust optimisation. Leveraging the dual formulation, the idea is to sample from the least favourable distribution. In particular using flow-based methods and generative models, the flow-DRO method decouples the data generation process and learning the least favourable distribution via the JKO-flow. (Xu et al., 2024)

2.4 Applications to econometrics and statistics

- Stéphane Bonhomme presented work in the framework of econometric models that take the form $\mathbb{E}[h(Y, X)|X] = m(X; \theta)$ with fixed θ. He argued that some components of $\theta = (\eta, \beta)$, say β, may be heterogeneous, and proposed inference on the homogeneous model, with adjustment to the standard errors to account for the fact that β could be random in a Wasserstein ball around the Dirac. (Bonhomme et al., 2024)

- Onil Boussim presented his work on the identification of potential outcome distributions under endogenous treatment and selection.

- Yanqin Fan's talk focused on the identification of a general class of moment equality models when not all the variables in the model are observed in the same data sets. Examples include identification and estimation of measures of fairness when the sensitive variables are not observed in the main data set containing observations on the other variables. Using optimal transport theory, her work establishes identified set for the parameter of interest. (Fan et al., 2024a)

- Mario Ghossoub's presentation considered an optimal allocation problem with initial endowments X_i for each individual i with utility U_i. If the U_i are all monotonic, concave and strictly Schur concave, then there exist Pareto optimal allocations, and they are all comonotonic. The work goes further to give a characterisation of the Pareto optimal (hence comonotonic) allocations if in addition the U_i are all translation invariant and positively homogeneous. (Ghossoub and Zhu, 2024)

- Alice Qi presented her work on targeted policy learning. In contrast to the popular utilitarian policy learning which maximises the expected value of the post-treatment outcome, targeted policy learning focuses on the lower tail of the distribution of the post-treatment outcome and thus maximises the outcome of the disadvantaged. The targeted welfare function in her work is a distributionally robust version of the utilitarian welfare function with a specific uncertainty set for the distribution of the target population. (Fan et al., 2024b)

- Samory Kpotufe presented recent results on the understanding of core problems in "AI" such as domain adaptation, transfer, multitask, representation, or lifelong learning. They all concern learning in heterogeneous and ever-changing environments. Results presented include recent measures of discrepancy between distributions, fine-tuned to given estimation problems (classification, bandits, etc) that offer a more optimistic picture than existing probability metrics (e.g. Wasserstein, TV) or divergences (KL, Renyi, etc) in terms of achievable rates; and non adaptively achievable minimax oracle rates in some seemingly simple extensions of basic settings, e.g., extensions to multiple choices of source datasets (as in multitask or multi source learning), or extensions to multiple prediction models to transfer (i.e., model selection under distribution shift). (Kalan and Kpotufe, 2023)

- Xiaoting Sun's presentation examined a problem of matching of students to schools. Student preferences over school are identified from a single realisation of the matching under the assumption that the matching is stable (no pair would prefer to be matched to their current assignment) and the observation of variables that shift student preferences but not school priorities, and vice-versa. (He et al., 2021)

- Johannes Wiesel presented his research on the out of sample prediction error of the $\sqrt{LASSO}$ and related estimators. The paper establishes a new bound on the out-of-sample prediction error of such estimators and shows that linear predictors based on these estimators solve a distributionally robust optimisation problem: they minimise the worst-case prediction error over distributions that are close to each other in a type of max-sliced Wasserstein metric. They also provide an oracle recommendation for the choice of regularisation parameter that guarantees good out-of-sample prediction error. (Olea et al., 2024)

- Gaoqian Xu presented his research on Wasserstein distributionally robust optimisation when the ambiguity set restricts the marginal distributions to respective Wasserstein balls of the reference marginal distributions. This work is motivated by empirical applications in which multiple data sets must be combined. A notable application is the robust evaluation of distributional causal effects of a binary treatment. (Fan et al., 2023)

- Wendao Xue presented her job market paper using optimal transport to address differential measurement error issue prevalent in causal inference. She adopts the method of sieves reviewed in Xiaohong Chen's talk to estimating both the optimal transport map and the average treatment effect of a binary treatment in the presence of differential measurement error. (Xue, 2023)

- Kelvin Zhang presented a new approach to the monopolist's problem. A new dual formulation to the problem was presented, as well as a new regularity theory. This approach sheds new light on the celebrated Rochet-Chone solution to the problem on the two dimensional square. (McCann and Zhang, 2023)

Figure 1: Conference picture of in-person participants at the Meeting

3 Outcome of the Meeting

The workshop, the first of this kind, brought together researchers from computer science, economics, finance, mathematics, statistics, and operations research, whose work develops optimal transport and distributional robust optimisation tools and applies those methodologies to causal inference, inequality measurement, partial identification, option pricing, risk and uncertainty management, robust decision making.

The workshop facilitated interactions and research collaborations among researchers from these different fields, diverse backgrounds and different career stages and profiles. This was achieved with a flexible program that included round tables, mentoring activities, informal discussions, and introductory talks on a few main themes or topics, that let people see how their own research fits with an area they were not previously familiar with.

All participants were very enthusiastic about the scientific content of this workshop and found it thought-provoking. It is the belief of the organisers that this workshop is the first of a long series of forthcoming scientific events which may have a significant impact on the development of OT and DR methods in the aforementioned fields.

4 Meeting Participants

The following researchers participated in-person or virtually in the meeting:

Pesenti, Silvana (University of Toronto); Fan, Yanqin (University of Washington); Henry, Marc (The Pennsylvania State University); Ansari, Jonathan (University of Salzburg); Bernard, Carole (Grenoble Ecole de Management); Blanchet, Jose (Stanford); Bonhomme,

Stephane Olivier (University of Chicago); Boussim, Onil (Pennsylvania State University); Cheung, Hang (University of Calgary); Duchi, John (Stanford); Gao, Rui (University of Texas at Austin); Ghossoub, Mario (University of Waterloo); Hiew, Joshua Zoen-Git (University of Alberta); Kpotufe, Samory (Columbia University); Lin, Liyuan (University of Waterloo); Mao, Tiantian (University of Science and Technololgy of China); Nguyen, Viet Anh (Chinese University of Hong Kong); Pass, Brendan (University of Alberta); Qi, Alice (University of Washington); Qu, Zhaonan (Stanford University); Rankin, Cale (University of Toronto); Ruschendorf, Ludger (University Freiburg); Vanduffel, Steven (Solvay Business School - Vrije Universiteit Brussel); Wang, Ruodu (University of Waterloo); Wiesel, Johannes (Carnegie Mellon University); Wong, Ting Kam Leonard (University of Toronto); Xie, Yao (Georgia Institute of Technology); Xu, Gaoqian (University of Washington); Xue, Wendao (University of Washington); Zhang, Kelvin Shuangjian (Fudan University); Acciaio, Beatrice (ETH Zurich); Charpentier, Arthur (UQAM); Chen, Xiaohong (Yale University); Cheng, Ziteng (University of Toronto); Coache, Anthony (University of Toronto); Deb, Nabarun (Booth School of Business, University of Chicago); Eckstein, Stephan (ETH Zurich); Galichon, Alfred (New York University); Gunsilius, Florian (University of Michigan); Hallin, Marc (Université libre de Bruxelles); Hansen, Lars (University of Chicago); Kroell, Emma (University of Toronto); Miao, Kathleen (University of Toronto); Moresco, Marlon (Concordia University); Raghunathan, Aditi (Carnegie Mellon University); Sun, Xiaoting (Simon Fraser University); Tam, Brandon (University of Toronto); Tomiyama, Hideyuki (Pennsylvania State University); Wiesel, Johannes (Carnegie Mellon University); Yang, Yang (University of Calgary); Zhao, Yidi (Pennsylvania State University); Zhao, Jiwei (University of Wisconsin-Madison); Zozoungbo, Christelle (Pennsylvania State University); d'Haultfoeuille, Xavier (CREST-ENSAE)

References

Ansari, J. and Rüschendorf, L. (2024). Supermodular and directionally convex comparison results for general factor models. *Journal of Multivariate Analysis*, 201:105264.

Bernard, C., Junike, G., Lux, T., and Vanduffel, S. (2022). Cost-efficient payoffs under model ambiguity. *arXiv preprint arXiv:2207.02948*.

Bernard, C., Pesenti, S. M., and Vanduffel, S. (2023). Robust distortion risk measures. *Mathematical Finance*. (forthcoming).

Blanchet, J., Murthy, K., and Zhang, F. (2021). Optimal transport based distributionally robust optimization: Structural properties and iterative schemes. *arXiv preprint arXiv:1810.02403*.

Bonhomme, S., Christensen, T., and Weidner, M. (2024). Heterogeneity robust estimation and inference. *in progress*.

Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F., and Marinacci, M. (2024). Making decisions under model misspecification. *arXiv preprint arXiv:2008.01071*.

Chen, X. (2006). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*.

Cheung, H., Qiu, J., and Badescu, A. (2023). A Viscosity Solution Theory of Stochastic Hamilton-Jacobi-Bellman Equations in the Wasserstein Space. *arxiv preprint arXiv:2310.14446*.

Fan, Y., Park, H., and Xu, G. (2023). Quantifying distributional model risk in marginal problems via optimal transport. *Working Paper*.

Fan, Y., Pass, B., and Shi, X. (2024a). Identification and inference in moment models under data combination via optimal transport. *Working Paper*.

Fan, Y., Qi, A., and Xu, G. (2024b). Targeted policy learning. *Working Paper*.

Ghossoub, M. and Zhu, M. (2024). Pareto efficiency in pure exchange economies with monotone concave Schur concave utilities. *in progress*.

Gunasingam, M. and Wong, T.-K. L. (2024). Adapted optimal transport between Gaussian processes in discrete time. *arXiv preprint arXiv:2404.06625*.

He, Y., Sinha, S., and Sun, X. (2021). Identification and estimation in many-to-one two-sided matching without transfers. *arXiv preprint arXiv:2104.02009*.

Joshua Zoen-Git Hiew, J. Z.-G., Nenna, L., and Pass, B. (2024). An ordinary differential equation for entropic optimal transport and its linearly constrained variants. *arxiv preprint arXiv:2403.20238*.

Kalan, M. and Kpotufe, S. (2023). Tight rates in supervised outlier transfer learning. *arXiv preprint arXiv:2310.04686*.

Mao, T., Wang, R., and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *arXiv preprint arXiv:2201.06370*.

McCann, R. J. and Zhang, K. S. (2023). A duality and free boundary approach to adverse selection. *arXiv preprint arXiv:2301.07660*.

Nutz, M., Wang, R., and Zhang, Z. (2022). Martingale transport and Monge maps. *arXiv preprint arXiv:2209.07660*.

Olea, J. L. M., Rush, C., Velez, A., and Wiesel, J. (2024). The out-of-sample prediction error of the $\sqrt{LASSO}$ and related estimators. *arXiv preprint arXiv:2211.07608*.

Pesenti, S. M. and Vanduffel, S. (2023). Optimal transport divergences induced by scoring functions. *arXiv preprint arXiv:2311.12183*.

Qu, Z., Galichon, A., and Ugander, j. (2023). On Sinkhorn's Algorithm and Choice Modeling. *arXiv preprint arXiv:2310.00260*.

Rankin, C. and Wong, T.-K. L. (2024). JKO schemes with general transport costs. *arXiv preprint arXiv:2402.17681*.

Rüschendorf, L. (2013). Mathematical risk analysis. *Springer Ser. Oper. Res. Financ. Eng. Springer, Heidelberg*.

Xu, C., Lee, J., Cheng, X., and Xie, Y. (2024). Flow-based distributionally robust optimization. *IEEE Journal on Selected Areas in Information Theory*.

Xue, W. (2023). Correcting strategic misreporting behavior on outcomes in estimating treatment effect. *Working Paper*.

Yang, J., Zhang, L., Chen, N., Gao, R., and Hu, M. (2022). Decision-making with side information: A causal transport robust approach. *Optimization Online*.