



SPS Newsletter

The Newsletter of the Stochastic Programming Society

Volume 2 Number 1

May 2021

Contents

<i>Chair's Column</i>	
Güzin Bayraksan	1
<i>SPS virtual seminar series: decision making in an uncertain world</i>	
Güzin Bayraksan and Francesca Maggioni	3
<i>INFORMS Frederick W. Lanchester Prize 2020:</i>	
<i>Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric</i>	
Peyman Mohajerin Esfahani and Daniel Kuhn	5
<i>Young Researcher Highlight:</i>	
<i>Dupačová-Prékopa Best Student Paper Prize Finalists</i>	
Junyi Liu	8
Rui Peng Liu	10
<i>Advice for Applying SP in the Real World</i>	
Stein W. Wallace	11
<i>Real World Impact of Stochastic Programming for</i>	
<i>The Electricity Sector Case</i>	
Alexandre Street and Davi Valladão	14
<i>COVID-19</i>	
Claudia Sagastizábal	17
<i>SP Events in 2021</i>	20
<i>Date Change of ICSP XVI</i>	
David Woodruff	21

Send your comments to the Editor:
 Francesca Maggioni
francesca.maggioni@unibg.it

Chair's Column

Güzin Bayraksan
 Ohio State University (USA)
bayraksan.1@osu.edu

Dear Colleagues,

What a year it has been! I hope everyone is doing well and keeping healthy.

Over the course of its first year, the new Committee on Stochastic Programming (COSP) has created the Stochastic Programming Society (SPS) Twitter and LinkedIn accounts. Our aim was to increase communication between our members. Little did we know that these would be invaluable during the pandemic. In addition, last summer, we organized a Virtual Seminar Series and, with the permission of speakers, put the videos of the talks on our newly established YouTube channel. I am happy to report that, in addition to hundreds of live attendees, as of this writing, our virtual seminar series have been collectively viewed more than 5500 times on our society's YouTube channel.

Despite a challenging year, it has been a great year of recognition for our community. I would like to highlight several prestigious awards¹ won by our esteemed colleagues. These awards recognize contributions in almost every aspect in the field, spanning **theory**, **methodology**, and **real-world impact**. (The below is listed by time of award first, and then alphabetically by last name).

¹Not an exhaustive list. Apologies for any missed awards.

- **Peyman Mohajerin Esfahani** and **Daniel Kuhn** were awarded the **2020 Frederick W. Lanchester Prize** by INFORMS for their influential work on distributionally robust optimization using the Wasserstein metric. The Lanchester prize is awarded for the best contribution to operations research and the management sciences published in English in the past five years. I would like to highlight that the prize is not awarded every year. You can find a summary of their work in this newsletter.
- **Jong-Shi Pang** has been elected a member of the **National Academy of Engineering** (NAE) for the development of methods to advance the theory and applications of optimization and operations research. Election to the NAE is among the highest professional distinctions accorded to an engineer, and academy membership honors those who have made outstanding contributions to their fields.
- **Mario Veiga Ferraz Pereira**, of the famed SDDP method, has been elected an international member of the **National Academy of Engineering** for his contributions to the methodology and implementation of multistage stochastic optimization in hydroelectric scheduling, energy planning, and policy.
- **Nilay Noyan** has been part of a team at Amazon that has been awarded the **2021 INFORMS Prize**. The 2021 INFORMS prize was awarded to Amazon, notably Amazon Transportation Services, for its long-lasting institutional achievement in integrating operations research into organizational decision making.
- Three of the 2021 **NSF CAREER Award** winners this year, **Yongjia Song**, **Phebe Vayanos**, and **Weijun Xie** are stochastic programmers!

Please join me in congratulating the amazing accomplishments of our colleagues and the impact that they are making! And please let me know about any of our colleagues who have been missed and deserve special recognition on our Twitter and LinkedIn feeds.

Together with last year's NAE and National Academy of Sciences inductees, the preeminent researchers **Alexander Shapiro**, **Arkadi Nemirovski** and **Jorge Nocedal**, these awards evidence the prominent role our community is playing and the theoretical and practical impact stochastic optimization is making.

I would like to use this opportunity to announce that the **next International Conference on Stochastic Programming (ICSP) has been postponed one year to 2023**. Given the domino effect the pandemic caused on the conference cycles—especially in light of ISMP-2021 being postponed to 2022—a decision has been reached by COSP to postpone ICSP for one year as well. Please mark your calendars for July 24–28, 2023, where we hope to see you in person in Davis, California, USA.

I hope you enjoy our second newsletter. As before, we are highlighting young researchers, the Dupačová-Prékopa best student paper prize finalists **Junyi Liu** and **Rui Peng Liu**, and real-world impact of stochastic programming in the electricity sector (**Alexandre Street** and **Davi Valladão**) and for the Covid-19 pandemic (**Claudia Sagastizábal**). Don't miss the perspective article by **Stein Wallace** on how to apply stochastic programming in the real world!

Last but not least, I am very grateful to all members of the Committee on Stochastic Programming for their many contributions.

I wish all of you healthy, happy, and creative days, and I look forward to seeing you at the next meeting held in person.



<https://www.stoprog.org/>



<https://twitter.com/stoprogsociety>



<https://www.linkedin.com/groups/13799735/>



https://www.youtube.com/channel/UCMSOCh_B-o0tW6FsgQCUrHg

Stochastic Programming Society Virtual Seminar Series:

Decision Making in an Uncertain World

Güzin Bayraksan* and Francesca Maggioni**

*Ohio State University (USA)

**University of Bergamo (ITALY)

bayraksan.1@osu.edu and francesca.maggioni@unibg.it

In summer of 2020, one of the most extraordinary summers in recent history, when many conferences were cancelled or postponed due to the Covid-19 pandemic, the Stochastic Programming Society (SPS) [1] held a virtual seminar series aptly titled “Decision Making in an Uncertain World”. The series gave the SPS community a way to learn about new research in the area and connect with colleagues across continents and different time zones via their shared interest in decision making under uncertainty.

The Virtual Seminar series were designed by the governing board of the Stochastic Programming Society, also known as the “Committee on Stochastic Programming” (COSP) [2], thanks to the suggestion of a member of SPS, Vincent Leclère. The committee decided on the name of the seminar series, the format of the seminars, and names of the speakers, taking into account their provenance from all around the World and research topics. The webinars took place biweekly on Fridays 7-8 AM Pacific Time in order to allow people at many different time zones to attend the events, starting on May 29th, 2020 using Zoom platform, kindly supported by the Integrated Systems Engineering Department of the Ohio State University. The series consisted of seven talks with presenters from various parts of the world: two from Europe (Steffen Rebennack and Daniel Kuhn), one from South America (Alejandro Jofré) and four from the United States (Cynthia Rudin, Alexander Shapiro, David Morton and Katya Scheinberg). The seminar series covered a variety of topics including theory, computations, and applications. It also included topics that are relevant to the society but that

go beyond the traditional topics of interest to the society such as machine learning. Topics covered by the series included the following areas of research:

- *Real-world Stochastic Programming Applications* such as safe COVID-19 Reopening (David Morton’s talk) and pricing in electricity markets and extension to markets with massive entry of renewable energies and distributed generation (Alejandro Jofré’s talk), Brazilian interconnected power system problem (Alexander Shapiro’s talk);
- *Theory and Computations of Stochastic Programming* such as cut-sharing in Stochastic Dual Dynamic Programming (Steffen Rebennack’s talk) and computational and theoretical aspects of solving multistage stochastic programs (Alexander Shapiro’s talk);
- *Distributionally Robust Optimization* such as moderate deviations theory and distributionally robust optimization, which aim to learn from correlated data (Daniel Kuhn’s talk);
- *Related Fields* such as interpretability versus explainability in *Machine Learning* (Cynthia Rudin’s talk), and convergence analysis of *Stochastic Algorithms* (Katya Scheinberg’s talk).

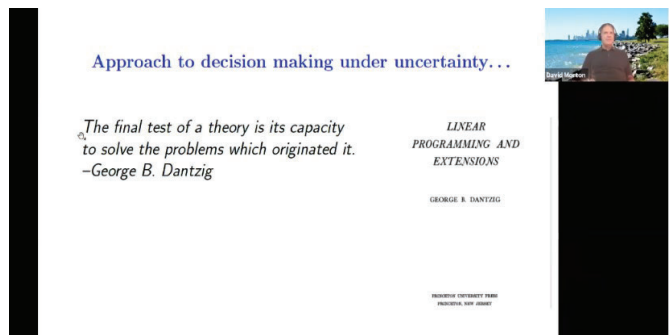


Figure 1: David Morton’s answer to “What is your approach to decision making under uncertainty?” during his talk on Covid-19 Reopening.

Because the seminars covered a wide variety of topics, COSP decided to begin each talk by asking the speakers what their approach to decision making under uncertainty was. Answers to this question by the

speakers (in alphabetical order of their last names) provided stimulating “food for thought” to the community:

- Alejandro Jofré advocated “combining Stochastic Optimization with interaction of different agents” in a system where all agents are affected by uncertainty. He emphasized taking into account the interaction of different agents, who may have their own objectives.
- Daniel Kuhn, instead of supporting one approach, “praised many different approaches depending on the problem at hand” and inspired the community to “find new approaches” that have not been used before.
- David Morton quoted one of the pioneers of Stochastic Programming, George B. Dantzig, “The final test of a theory is its capacity to solve the problems which originated it” (see Figure 1), encouraging the SPS community to apply its models and methods to solve real-world problems.
- Steffen Rebennack said that his approach to decision making under uncertainty is to “incorporate all available information on the uncertainty in an optimization model.”
- Cynthia Rudin promoted “decision making in a human-aware way” (see Figure 2). In this decision-making paradigm, machine learning tools are decision aids to human decision makers, rather than providing the decisions themselves. This necessitates the machine learning models to be interpretable by humans.
- Katya Scheinberg advised “putting [most of the] eggs in a solid basket,” updating the common advice “don’t put all your eggs in one basket.” She related this updated advice to both real-life decisions under uncertainty and the success of the algorithms she has been analyzing.
- Alexander Shapiro mentioned that historically many different communities worked on decision making problems under uncertainty, including Markov Decision Processes, Stochastic Optimal Control, and Stochastic Programming. Even though commonalities exist, these communities

primarily worked independently with different approaches. He encouraged the communities to come together and learn from each other. He also urged the creation of a library of decision-making problems under uncertainty.

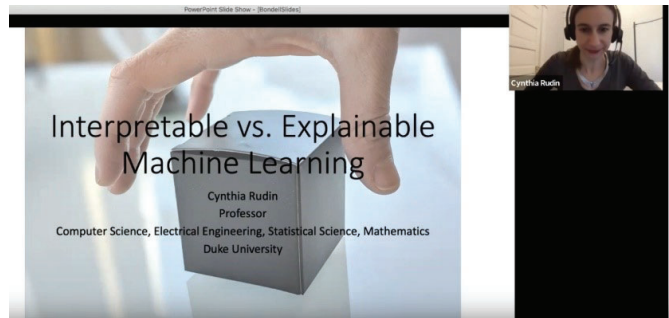


Figure 2: Cynthia Rudin’s talk on Interpretable Machine Learning.

With the permission of the presenters, COSP posted recordings of the webinars on the new Stochastic Programming Society YouTube channel [3]. Since June 15, 2020, the talks collectively generated more than 5000 total views. This was a great opportunity to learn about new research in the optimization under uncertainty area and connect with colleagues from all over the world.

Acknowledgements. A version of this article appeared in Newsletter 34 of the European Women in Mathematics, devoted to the ongoing pandemic and its consequences. We thank for the permission to reprint it.

REFERENCES

- [1] <https://stoprog.org/>
- [2] <https://stoprog.org/cosp-members>
- [3] https://www.youtube.com/channel/UCMSOCh_B-o0tW6FsgQCurHg/videos

INFORMS Frederick W. Lanchester Prize 2020:

Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric

Peyman M. Esfahani* and Daniel Kuhn**

*Delft University of Technology (THE NETHERLANDS)

**EPFL, Lausanne (SWITZERLAND)

P.MohajerinEsfahani@tudelft.nl and

daniel.kuhn@epfl.ch

The 2020 INFORMS Frederick W. Lanchester Prize for the best contribution to operations research and management science in the past five years was awarded to Peyman Mohajerin Esfahani and Daniel Kuhn at 2020 INFORMS Annual Meeting for the paper “Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations” published in Mathematical Programming 171(1-2), 115–166, 2018. Below a summary of this paper.

Consider the stochastic optimization problem

$$J^* := \inf_{x \in \mathbb{X}} \left\{ \mathbb{E}^{\mathbb{P}} [h(x, \xi)] = \int_{\Xi} h(x, \xi) \mathbb{P}(d\xi) \right\} \quad (1)$$

where $\mathbb{X} \subseteq \mathbb{R}^n$, $\Xi \subseteq \mathbb{R}^m$ and the loss function $h(x, \xi)$ depends both on the decision vector x and the random vector ξ , which is governed by a probability distribution \mathbb{P} supported on Ξ . In practice the distribution \mathbb{P} may only be indirectly observable through independent training samples $\hat{\xi}_1, \dots, \hat{\xi}_N$. A *data-driven solution* for problem (1) is a feasible decision \hat{x}_N constructed from the training data. We also aim to construct a *data-driven certificate* \hat{J}_N , that is, a safe estimate of the out-of-sample performance $\mathbb{E}^{\mathbb{P}} [h(\hat{x}_N, \xi)]$. Specifically, we hope to ensure that the inequality $\mathbb{E}^{\mathbb{P}} [h(\hat{x}_N, \xi)] \leq \hat{J}_N$ holds with a high probability, and we refer to this probability as the *reliability*. Our ideal goal is to find a data-driven solution \hat{x}_N with the lowest possible out-of-sample performance. This is impossible, however, because \mathbb{P} is unknown, and the out-of-sample performance cannot

be computed. We thus pursue the more modest but achievable goal to find a data-driven solution with a low certificate and a high reliability.

A natural approach to generate data-driven solutions \hat{x}_N is to approximate \mathbb{P} with the discrete empirical probability distribution $\hat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_i}$. This amounts to approximating the original stochastic program (1) with the *sample-average approximation* (SAA) problem

$$\hat{J}_{\text{SAA}} := \inf_{x \in \mathbb{X}} \mathbb{E}^{\hat{\mathbb{P}}_N} [h(x, \xi)] \quad (2)$$

If the feasible set \mathbb{X} is compact and the loss function is uniformly continuous in x across all $\xi \in \Xi$, then the optimal value and the optimal solutions of the SAA problem (2) converge almost surely to their counterparts in the true problem (1) as N tends to infinity [2, Theorem 5.3]. The SAA problem has been conceived primarily for situations where the distribution \mathbb{P} is known and additional samples can be acquired cheaply via random number generation. However, the optimal solutions of the SAA problem tend to display a poor out-of-sample performance in situations where N is small and where the acquisition of additional samples would be costly or impossible.

This prompts us to propose an alternative approach that explicitly accounts for our ignorance of the true data-generating distribution \mathbb{P} and that offers attractive performance guarantees even when the acquisition of additional samples from \mathbb{P} is impossible or expensive. Specifically, we design an ambiguity set $\mathbb{B}_\varepsilon(\hat{\mathbb{P}}_N)$ containing all distributions that could have generated the available training samples with high confidence. This ambiguity set enables us to define the data-driven decision \hat{x}_N and the certificate \hat{J}_N as the optimal value and an optimal solution of a distributionally robust optimization (DRO) problem of the form

$$\hat{J}_N := \inf_{x \in \mathbb{X}} \sup_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}} [h(x, \xi)]. \quad (3)$$

Following the pioneering work [3], we construct $\mathbb{B}_\varepsilon(\hat{\mathbb{P}}_N)$ as a ball of radius ε around the empirical distribution $\hat{\mathbb{P}}_N$ with respect to the (first) Wasserstein distance. We will demonstrate that the optimal value \hat{J}_N as well as any optimal solution \hat{x}_N (if it exists) of the DRO problem (3) offer rigorous finite sample and asymptotic consistency guar-

antees. Moreover, for many loss functions of practical interest, the DRO problem (3) is computationally tractable and admits a reformulation reminiscent of the SAA problem (2).

1. Statistical guarantees

The (first) Wasserstein distance between two probability distributions \mathbb{Q}_1 and \mathbb{Q}_2 on \mathbb{R}^m is defined as

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \int_{\mathbb{R}^{2m}} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) : \right. \\ \left. \begin{array}{l} \Pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2, \text{ respectively} \end{array} \right\},$$

where $\|\cdot\|$ represents an arbitrary norm on \mathbb{R}^m . In addition, we define the Wasserstein ambiguity set with radius $\varepsilon \geq 0$ as

$$\mathbb{B}_\varepsilon(\widehat{\mathbb{P}}_N) := \left\{ \mathbb{Q} \in \mathcal{M}(\Xi) : d_W(\widehat{\mathbb{P}}_N, \mathbb{Q}) \leq \varepsilon \right\}, \quad (4)$$

where $\mathcal{M}(\Xi)$ is the set of all probability distributions on Ξ .

One may use [4, Theorem 2] to derive an a priori estimate of the probability that the unknown data-generating distribution \mathbb{P} falls within the Wasserstein ball $\mathbb{B}_\varepsilon(\widehat{\mathbb{P}}_N)$. Specifically, for any fixed $\beta \in (0, 1)$ the Wasserstein ball contains \mathbb{P} with confidence $1 - \beta$ if its radius ε exceeds

$$\varepsilon_N(\beta) := \begin{cases} \left(\frac{\log(c_1 \beta^{-1})}{c_2 N} \right)^{1/\max\{m, 2\}} & \text{if } N \geq \frac{\log(c_1 \beta^{-1})}{c_2}, \\ \left(\frac{\log(c_1 \beta^{-1})}{c_2 N} \right)^{1/a} & \text{if } N < \frac{\log(c_1 \beta^{-1})}{c_2}. \end{cases}$$

Note that the Wasserstein ball with radius $\varepsilon_N(\beta)$ can thus be viewed as a confidence region for \mathbb{P} . This insight forms the basis for several statistical guarantees for problem (3).

Theorem 1 (Finite sample guarantee). *Suppose that $\mathbb{E}^{\mathbb{P}}[\exp(\|\xi\|^a)] \leq A$ for some $a > 1$ and $A > 0$ and that $\beta \in (0, 1)$. If \widehat{J}_N and \widehat{x}_N represent the optimal value and an optimizer of the DRO problem (3) with Wasserstein ambiguity set $\mathbb{B}_{\varepsilon_N(\beta)}(\widehat{\mathbb{P}}_N)$, then we have for all $N \geq 1$ that*

$$\mathbb{P}^N \left\{ \mathbb{E}^{\mathbb{P}}[h(\widehat{x}_N, \xi)] \leq \widehat{J}_N \right\} \geq 1 - \beta. \quad (5)$$

Note that (5) guarantees that the certificate \widehat{J}_N provides a $(1 - \beta)$ -upper confidence bound on the

out-of-sample performance of the data-driven decision \widehat{x}_N for all $N \geq 1$. While Theorems 1 and 2 provide strong theoretical justification for using Wasserstein ambiguity sets, in practice, the radius $\varepsilon_N(\beta)$ is rather conservative, and it is prudent to calibrate the Wasserstein radius via data-driven techniques from statistics. Examples include the hold-out method, k -fold cross-validation or bootstrapping.

Theorem 2 (Asymptotic consistency). *Suppose that $\mathbb{E}^{\mathbb{P}}[\exp(\|\xi\|^a)] \leq A$ for some $a > 1$ and $A > 0$ and that $\beta_N \in (0, 1)$ satisfies $\sum_{N=1}^{\infty} \beta_N < \infty$ and $\lim_{N \rightarrow \infty} \varepsilon_N(\beta_N) = 0$.¹ Assume also that \widehat{J}_N and \widehat{x}_N represent the optimal value and an optimizer of the DRO problem (3) with Wasserstein ambiguity set $\mathbb{B}_{\varepsilon_N(\beta_N)}(\widehat{\mathbb{P}}_N)$.*

- (i) *If $h(x, \xi)$ is upper semicontinuous in ξ and there exists $L \geq 0$ with $|h(x, \xi)| \leq L(1 + \|\xi\|)$ for all $x \in \mathbb{X}$ and $\xi \in \Xi$, then \mathbb{P}^∞ -almost surely we have $\widehat{J}_N \downarrow J^*$ (converges from above) as $N \rightarrow \infty$, where J^* is the optimal value of (1).*
- (ii) *If the assumptions of assertion (i) hold, \mathbb{X} is closed, and $h(x, \xi)$ is lower semicontinuous in x for every $\xi \in \Xi$, then any accumulation point of $\{\widehat{x}_N\}_{N \in \mathbb{N}}$ is \mathbb{P}^∞ -almost surely an optimal solution for (1).*

One can show that all assumptions of Theorem 2 are necessary as well as sufficient, that is, relaxing any of these conditions can invalidate the asymptotic consistency result.

2. Tractable reformulation

We now prove that the inner worst-case expectation problem in (3) over the Wasserstein ball (4) can be reformulated as a finite convex program for many loss functions $h(x, \xi)$ of practical interest. For ease of notation, in this section we suppress the dependence on the decision variable x . Thus, we examine the worst-case expectation problem

$$\sup_{\mathbb{Q} \in \mathbb{B}_\varepsilon(\widehat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[\ell(\xi)], \quad \ell(\xi) = \max_{k \leq K} \ell_k(\xi), \quad (6)$$

involving a loss function that is defined as the pointwise maximum of measurable component functions

¹A possible choice is $\beta_N = \exp(-\sqrt{N})$.

$\ell_k(\xi)$, $k \leq K$. The focus on loss functions representable as pointwise maxima is non-restrictive unless we impose some structure on the component functions ℓ_k . Our key tractability results are predicated on the following convexity assumption.

Assumption 3 (Convexity). *The set $\Xi \subseteq \mathbb{R}^m$ is convex and closed, and the negative component functions $-\ell_k$ are proper, convex, and lower semicontinuous for all $k \leq K$. In addition, ℓ_k is not identically equal to $-\infty$ on Ξ , $k \leq K$.*

Assumption 3 essentially stipulates that $\ell(\xi)$ can be written as a maximum of concave functions. The worst-case expectation problem (6) constitutes an infinite-dimensional optimization problem over probability distributions and thus appears to be intractable. However, one can demonstrate that (6) can be re-expressed as a finite-dimensional convex program by leveraging tools from robust optimization.

Theorem 4 (Exact convex reduction). *If Assumption 3 holds, then for any $\varepsilon \geq 0$ the worst-case expectation (6) equals the optimal value of the finite convex program*

$$\begin{aligned} \min \quad & \lambda\varepsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t.} \quad & \lambda, s_i \in \mathbb{R}, z_{ik}, \nu_{ik} \in \mathbb{R}^m, \quad \forall i \leq N, \forall k \leq K \\ & [-\ell_k]^*(z_{ik} - \nu_{ik}) + \sigma_{\Xi}(\nu_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \\ & \|z_{ik}\|_* \leq \lambda, \end{aligned}$$

where $[-\ell_k]^*$ denotes the conjugate of $-\ell_k$, σ_{Ξ} is the support function of Ξ and $\|\cdot\|_*$ is the norm dual to $\|\cdot\|$.

Stress test experiments are instrumental to assess the quality of candidate decisions in stochastic optimization. Meaningful stress tests require a good understanding of the extremal distributions from within the Wasserstein ball that achieve the worst-case expectation (6) for various loss functions. We now argue that such extremal distributions can be constructed systematically from the solution of a convex program dual to the one described in Theorem 4.

Theorem 5 (Worst-case distributions). *If Assumption 3 holds, then for any $\varepsilon \geq 0$ the worst-case expectation (6) coincides with the optimal value of the*

finite convex program

$$\begin{aligned} \max \quad & \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \alpha_{ik} \ell_k \left(\hat{\xi}_i - \frac{q_{ik}}{\alpha_{ik}} \right) \\ \text{s.t.} \quad & \alpha_{ik} \in \mathbb{R}_+, \quad q_{ik} \in \mathbb{R}^m \quad \forall i \leq N, \quad \forall k \leq K \\ & \hat{\xi}_i - \frac{q_{ik}}{\alpha_{ik}} \in \Xi \quad \forall i \leq N, \quad \forall k \leq K \\ & \sum_{k=1}^K \alpha_{ik} = 1 \quad \forall i \leq N \\ & \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \|q_{ik}\| \leq \varepsilon. \end{aligned}$$

If $\{\alpha_{ik}^*, q_{ik}^*\}_{r \in \mathbb{N}}$ is an optimal solution of the above problem, then the discrete probability distribution

$$\mathbb{Q}^* := \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \alpha_{ik}^* \delta_{\xi_{ik}^*} \quad \text{with} \quad \xi_{ik}^* = \hat{\xi}_i - \frac{q_{ik}^*}{\alpha_{ik}^*}$$

belongs to $\mathbb{B}_{\varepsilon}(\hat{\mathbb{P}}_N)$ and attains the supremum of (6).

3. Applications and MOSEK software package

The class of loss functions (6) fulfilling Assumptions 3 is rich and encompasses several interesting special cases for which the convex reformulation in Theorem 4 reduces to an explicit linear program. This is the case when the 1-norm or the ∞ -norm is used in the definition of the Wasserstein metric and if $\ell(\xi)$ belongs to any of the following function classes: (i) a pointwise maximum or minimum of affine functions; (ii) the indicator function of a closed polytope or the indicator function of the complement of an open polytope; (iii) the optimal value of a parametric linear program whose cost or right-hand side coefficients depend linearly on ξ . Tractable reformulations of (3) are also available when (iv) the random vector ξ can be viewed as a stochastic process and the loss function is additively separable, and (v) the loss function is convex in ξ and may therefore not be representable as a pointwise maximum of finitely many concave functions as postulated by Assumption 3.

Recently, MOSEK 9.2 introduced parameterization in their Fusion API (available for Python, C++, Java and .NET). As it is particularly well-suited for (re-)solving large-scale optimization problems, MOSEK have dedicated a jupyter notebook for the Wasserstein DRO problems reformulated as in Theorem 4, using their Fusion API for Python [5]. The

notebook also covers various data-driven techniques to calibrate the Wasserstein radius including cross-validation and bootstrapping. These methods are exemplified in the context of mean-risk portfolio optimization.

REFERENCES

- [1] P. Mohajerin Esfahani, and D. Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations." *Mathematical Programming* 171(1-2), 115–166, 2018.
- [2] A. Shapiro, D. Dentcheva, and A. Ruszczyński. "Lectures on stochastic programming: modeling and theory". *Society for Industrial and Applied Mathematics*, 2014.
- [3] G. Pflug, and D. Wozabal. "Ambiguity in portfolio selection." *Quantitative Finance* 7(4), 435–442, 2007.
- [4] N. Fournier, and A. Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure." *Probability Theory and Related Fields* 162 (3), 707–738, 2015.
- [5] MOSEK ApS, "Fusion API for Python" *MOSEK Optimization Suite 9.2.38*, 2021, Available at https://nbviewer.jupyter.org/github/MOSEK/Tutorials/blob/master/dist-robust-portfolio/Data-driven_distributionally_robust_portfolio.ipynb.

The ICSP XV Dupačová-Prékopa Student Paper Prize Finalists

Two-stage stochastic programming with linearly bi-parameterized quadratic recourse

Junyi Liu

University of Southern California (USA)

junyiliu@usc.edu

A finalist of the 2019 Dupačová-Prékopa Best Student Paper Prize was Junyi Liu at ICSP XV for the paper "Two-stage stochastic programming with linearly bi-parameterized quadratic recourse", coauthored with Ying Cui, Jong-Shi Pang, and her adviser Suvrajeet Sen, published in SIAM Journal on Optimization, 30(3):2530–2558, 2020. Below a summary of this paper.

The two-stage stochastic program (SP) is one of the standard SP models in which the first-stage decision is made prior to observing the uncertainty, and the second-stage recourse decision is undertaken so as to adapt to the observation in an optimal manner. To date, this class of problems has an overwhelming feature in the literature [1], that is the cost vector in the objective of the second stage program is independent of the first-stage decision. Our paper focuses on an extended two-stage SP with linearly bi-parameterized recourse as follows:

$$\underset{x \in X \subseteq \mathbb{R}^{n_1}}{\text{minimize}} \quad \zeta(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\xi}}[\psi(x, \tilde{\xi})], \quad (1)$$

where the recourse function $\psi(x, \xi)$ is the optimal objective value of the quadratic program:

$$\begin{aligned} \psi(x, \xi) \triangleq \min_y \quad & [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Q y \\ \text{s.t.} \quad & A(\xi)x + Dy \geq b(\xi), y \in \mathbb{R}^{n_2}. \end{aligned} \quad (2)$$

In this setting, $\tilde{\xi}$ is a random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and ξ without the tilde refers to a realization of the random variable. Then

$f(\xi)$, $G(\xi)$, $A(\xi)$, and $b(\xi)$ are respectively the realizations of these functions composed with the random variable $\tilde{\xi}$. We have the blanket assumption that $\varphi(\bullet)$ is a convex function on a compact convex set X ; Q is a symmetric positive semidefinite matrix; and the recourse function $\psi(x, \xi)$ satisfies the relatively complete recourse property. This linearly bi-parameterized recourse has numerous applications, such as two-stage shipment planning with pricing [2], two-stage power systems planning with renewable energy, and two-stage SP with linear complementarity constraint.

In general, the linearly bi-parameterized recourse $\psi(x, \xi)$ (2) is a nonconvex function. The loss of convexity is possibly the main reason that the state-of-the-art of computational two-stage SP remained largely under the restricted setting that $G(\xi) = 0$ for almost every ξ so that the recourse function is convex and piecewise affine. Nevertheless, if one is willing to trade global optimality for model fidelity, then one may be interested in obtaining a "stationary" solution to the extended modeling paradigm of two-stage SP (1). Specifically, in this paper, we are motivated to answer the following question: what type of solutions can be obtained by numerical algorithms for this class of two-stage SP problems (1)?

It is shown in [3] that the linearly bi-parameterized recourse $\psi(\bullet, \xi)$ is a difference-of-convex (dc) function, thus the combined objective function $\zeta(x)$ is also dc. In principle, the difference-of-convex algorithm [4] could be applied to solve the two-stage SP. Nevertheless, the dc decomposition of the recourse function given in [3] is only of conceptual value and practically not suitable for computation. So in the absence of the explicit dc decomposition, it is not clear what kind of limit one can expect of an iterative method for such two-stage SP problems. Furthermore, while the sampling methods have been studied extensively for convex SPs [1], when applied to nonsmooth and nonconvex SPs, the sampling technique should be combined with some convexification of the original problem with appropriate sample size control. Given the above background and challenges, we highlight two major contributions of our paper regarding to the two-stage SP with linearly bi-parameterized recourse:

1. We identify an implicit convex-concave property

of the linearly bi-parameterized recourse function $\psi(x, \xi)$ based on which the concept of a generalized critical point is defined. We present the relation of generalized critical points to the Clarke stationary points and directional stationary points. Furthermore, we derive the sufficient condition for such a point to be a directional stationary point based on the directional derivative formula in [5], which highlights the role of multipliers of the second-stage constraints.

2. By adding a Tikhonov regularization term to the second-stage program (2), the resulted regularized recourse function has an explicit dc decomposition. Hence, by combining sequential regularization, convexification, and sampling, we propose an algorithm, called the RCS algorithm, for solving the stochastic program (1). With appropriate control of regularization parameters and sample sizes in the RCS algorithm, we prove that every accumulation point of the sequence produced by the algorithm is a generalized critical point almost surely. A key technical step in the convergence analysis is the derivation of uniform bounds for various function values, subgradients, gradients, and error estimates under some matrix-theoretic assumptions. We also present numerical experiments on the joint production, pricing, and shipment planning problem to test the effectiveness of the RCS algorithm.

REFERENCES

- [1] A. Shapiro, D. Dentcheva, and A. Ruszczyński, "Lectures on stochastic programming: modeling and theory," SIAM, 2009.
- [2] D. Bertsimas, and K. Nathan, "From predictive to prescriptive analytics," *Management Science*, 66(3), 1025-1044, 2020.
- [3] M. Nouiehed, J-S. Pang, and M. Razaviyayn, "On the pervasiveness of difference-convexity in optimization and statistics," *Mathematical Programming*, 174, 195–222, 2019.
- [4] L. T. H. An, and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with dc models of real world nonconvex optimization problems," *Annals of Operations Research*, 133,23–46, 2005.

- [5] R. Janin, "Directional derivative of the marginal function in nonlinear programming," *Mathematical Programming Study*, 110-126, 1984.

The ICSP XV Dupačová-Prékopa Student Paper Prize Finalists

On Stochastic Programs without Relatively Complete Recourse

Rui Peng Liu

Georgia Institute of Technology (USA)
rayliu@gatech.edu

A finalist of the 2019 Dupačová-Prékopa Best Student Paper Prize was Rui Peng Liu at ICSP XV for the paper "On feasibility of sample average approximation solutions", published in SIAM Journal on Optimization, 30(3), 2026–2052, 2020. Below a summary of this paper.

We consider general stochastic programs of the form

$$\inf_{x \in \mathcal{X} \subseteq \mathbb{R}^n} F(x) := \mathbb{E}[f(x, \xi)]. \quad (1)$$

Here, $f(x, \cdot)$ is an extended real-valued integrable function for every decision vector $x \in \mathcal{X}$, and the expectation is taken with respect to the random vector ξ . An important class of stochastic programs is the two stage problems, in which the objective function is given by a second stage problem, i.e.,

$$f(x, \xi) = \inf_{y \in \mathcal{Y}(x, \xi)} g_{\xi}(y)$$

for real-valued functions g_{ξ} . A two stage problem is said to have relatively complete recourse (RCR) if, for every $x \in \mathcal{X}$ and almost every outcome of ξ , the set $\mathcal{Y}(x, \xi)$ is nonempty, or equivalently,

$$f(x, \cdot) < \infty \text{ a.s. } \forall x \in \mathcal{X}. \quad (2)$$

In general, we say a stochastic program has RCR if condition (2) holds.

The RCR property makes life easy. For two stage problems, this means that, in almost every scenario,

the second stage problem has a solution for whatever decision $x \in \mathcal{X}$ implemented in the first stage. Nevertheless, the RCR property does not always hold in practice. For example, when deciding the size of a reservoir, it could happen that certain size is too small to hold enough water storage to buffer potential drought impact.

In [2], we studied the sample average approximation (SAA) approach to solve stochastic programs without RCR. Given $\xi_{[N]} = (\xi_1, \dots, \xi_N)$, an i.i.d. random sample of ξ of size N , the SAA approach solves the following sample approximation of (1):

$$\inf_{x \in \mathcal{X}} \frac{1}{N} \sum_{i=1}^N f(x, \xi_i). \quad (3)$$

For every outcome of $\xi_{[N]}$, we denote $x(\xi_{[N]})$ to be an arbitrary feasible solution of (3) and $x^*(\xi_{[N]})$ to be an arbitrary optimal solution of (3).

For stochastic programs without RCR, it is important to understand how feasible a given solution x is, which is quantified by

$$d(x) = \mathbb{P}(f(x, \xi) < \infty).$$

Generally speaking, solutions x with higher $d(x)$ are preferred. For the class of problems considered below, we provided in [2] upper bounds on the probability

$$\mathbb{P}^N(d(x(\xi_N)) < 1 - \alpha), \quad \alpha \in (0, 1]. \quad (4)$$

In plain words, (4) is the portion of outcomes of $\xi_{[N]}$ for which the solutions $x(\xi_{[N]})$ have low feasibility.

We first consider the chain-constrained domain. Define

$$\text{dom } f_{\xi} = \{x : f(x, \xi) < \infty\}.$$

We say a stochastic program has chain-constrained domain of order m if there exists m functions $c_k(x)$ and m random variables $\ell_k(\xi)$ such that

$$\text{dom } f_{\xi} = \{x : c_k(x) \leq \ell_k(\xi), \forall 1 \leq k \leq m\}.$$

For example, a two stage problem has chain-constrained domain when

$$\mathcal{Y}(x, \xi) = \{y \geq 0 : Wy + Tx = h(\xi)\},$$

where W and T are deterministic matrices and $h(\xi)$ is a random variable. For stochastic programs with

chain-constrained domain of order m , we showed

$$\mathbb{P}^N(d(x(\xi_N)) < 1 - \alpha) \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

We next consider convex stochastic programs, where \mathcal{X} is a closed convex set and f is convex in x for every outcome of ξ . If the set of optimal solutions \mathcal{X}^* of (1) is compact and lives in the interior of the effective domain of F , we showed

$$\mathbb{P}^N(d(x^*(\xi_N)) < 1) \leq 1 - p_u^N,$$

where p_u^N is the probability of a uniform approximation of F by $N^{-1} \sum f(x, \xi_i)$ on a compact set around \mathcal{X}^* . Furthermore, for a convex stochastic program with chain-constrained domain, we showed

$$\mathbb{P}^N(d(x^*(\xi_N)) < 1 - \alpha) \leq \sum_{k=0}^{|J|-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k} + (1 - p_u^N),$$

where, roughly speaking, J is the set of constraints c_k that are active at \mathcal{X}^* .

REFERENCES

- [1] M. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs", *SIAM Journal on Optimization* 19 (2008), no. 3, 1211-1230.
- [2] RP. Liu, "On feasibility of sample average approximation solutions", *SIAM Journal on Optimization* 30 (2020), no. 3, 2026-2052.
- [3] R. Rockafellar and R. Wets, "Stochastic convex programming: Relatively complete recourse and induced feasibility", *SIAM Journal on Control and Optimization* 14 (1976), no. 3, 574-589.
- [4] A. Shapiro and A. Nemirovski, "On complexity of stochastic programming problems", In: *Jeyakumar V., Rubinov A. (eds) Continuous Optimization. Applied Optimization*, vol 99. Springer, Boston, MA.
- [5] A. Shapiro, D. Dentcheva, and A. Ruszczyński, "Lectures on stochastic programming: modeling and theory", second ed., *SIAM and MPS*, Philadelphia, 2014.

Advice for applying SP in the real world

Stein W. Wallace

NHH Norwegian School of Economics, Bergen (NORWAY)

Stein.Wallace@nhh.no

Impact Summary: *The purpose of this short note is to challenge those of you who consider using stochastic programming out there in the real world. By that I mean, make genuine decisions using real money, not just write a paper motivated by a real case. There are no correct ways to do this, but there are some pitfalls and some difficulties that need to be faced that are a bit different from when we write papers on applications, but nobody really cares about the results. I will focus on the stochastic side of the problem, as most questions of algebraic descriptions are shared with deterministic modeling.*

Background

How do you know that the solution you have found isn't off by, say 5%, compared to the actual best available solution for a company? In applied OR papers this question is hardly ever asked, but for a company, wasting 5% of the profit is a big issue if the reason is simply 'us'. My claim is that unless somebody is actually putting real money on your model, nobody is ever going to discover this weakness of the model, certainly not reviewers, editors or readers. I cannot answer the question either, but I can give you some thoughts on the way if you are planning to actually sell your model and its solutions. None of this is very deep, but added up it may matter.

Optimality

Though finding optimal solutions to a model is cute – and there is no reason to avoid it – there is also no major reason to require it. The reasons are obvious: The *model* is wrong by definition, relative to the problem at hand, and many of the parameters you put in are not fully known (but you wisely did not include them as random variables), either because they

are about the future, or because it is too costly to figure out the real values. And even if, for some strange reason, the model was right and all parameters you included really were known, there is the problem of the scenarios — my main issue in this short note — as they normally just approximate the underlying distribution. So don't spend too much time on worrying about optimal solutions to an incorrect model. In addition, most, if not all, real models are so large that optimality is numerically impossible — so why worry?

The people side

There is a lot of useful information for us in the work on human judgment in decision-making, represented for example by the Nobel Prize in Economic Sciences to Daniel Kahneman in 2002, see [2]. I advise all stochastic programmers to read up on that material. Here I will just mention two aspects. They might seem obvious, but they are not.

- Specialists almost always think they know more than they do. An important effect is that if you depend on expert judgments in estimating the distribution of your random variables (like the probability of machine breakdowns and their duration for a new machine with no history), they will under estimate the probability of a breakdown as well as the means and variances of their durations. Hence, uncertainty will seem less serious than it is, and investments in flexibility will be too low. Empirical research indicates that this is happening all the time, and it is not a result of people showing off.
- The world is full of misconceptions about uncertain phenomena. We all know most of them: Regression to the mean, the gambler's fallacy, the IQ of hindsight (it is hard to learn from outcomes affected by uncertainty), the law of small numbers, anchoring ... to mention but a few. Knowing this literature is very useful when discussing models or results with problem owners as it can help us understand what might seem as strange arguments and misconceptions. And it is important when trying to learn from outcomes affected by uncertainty.

Random demand

This is a very specific comment, but I see so many people struggling with it that it is worth mentioning. We often see references to 'demand data'. But that is rarely what they are, they are sales data. Demand data are very hard to come by. It is extremely unlikely that you have data on the demand for white T-shirts or the demand for bicycle rentals between a pair of rental stations. What you have is data for how many white T-shirts were sold and how many bicycles that were picked up at one station and delivered at another. You do not know how many rentals fell through as a rental station was empty or how many bicycles were handed back at the wrong station because the first-choice was full. In an academic paper this seems to be acceptable. But think carefully if you produce a method for really redistributing bicycles among stations, using sales as a substitute for demand. You might be producing a system that preserves the old sales pattern rather than facilitates the actual demand.

Scenarios

So let us turn to the main issue of applied stochastic programming, namely how to think about the uncertainty when we really care about the end results. Let me first point out that while it is fairly easy to find out how many scenarios you can *handle* in your optimization model (though it does depend a bit on the scenarios themselves), it is not so easy to work out how many scenarios you *need* to achieve a certain quality in the solution. In particular remember that the sentence 'I need s scenarios to achieve a certain quality' makes no sense, while 'I need s scenarios using scenario generation method X to achieve a certain quality' does. In other words, the number of scenarios depends both on the problem itself and how you create scenarios. For this reason I prefer to see scenario generation as part of modeling, not as part of data handling (but not everybody agrees).

The first step will always be to model the uncertainty as well as you can, using data, qualitative understanding and experts. It is beyond this note to say how, but I would like to make one important point: Try to model the uncertainty as well as you possibly can without regards for how you will even-

tually make scenarios. Do not lose quality in your description because you keep thinking about how to create scenarios from the description, and certainly do not model scenarios directly. This overall description of uncertainty will then be the basis for scenario generation (possibly using several different methods) but also for your out-of-sample testing, or possibly the use of a simulation model, to test the quality of your decisions.

Sampling is always tempting, but if you have a big model and care about quality, please be careful. Most likely you will need so many scenarios to achieve a necessary quality of the solution, that you would be numerically out of business. Cute academic models is one thing, the real world with real money a different story. That said, if you use some of the solid sampling based methods correctly, such as for example Stochastic Decomposition or Sample Average Approximation, please do so if they deliver what you need. But for reasonably large problems you are going to face some real numerical challenges. Apart from size, sampling from complicated mixed distributions, with some variables discrete (particularly binary), some continuous, some based on qualitative understanding, some on historical data, some dependent, some not, is not as easy as it might sound. In fact, it is, numerically speaking, close to impossible in most such complicated or very large cases.

Whatever you do, unless you use a scenario generation method with a built-in quality check, you need to do your best to check the quality of your results. In my view, what to do first is to check for in-sample stability. Details for this and other aspects of quality assurance, can be found in [3]. The main idea here is to make sure that if you run the scenario generation procedure several times, and then solve your model for each of the resulting scenario trees, you should get approximately the same objective function value each time. If you don't, it seems your scenario generation method does not deliver meaningful results with the given number of scenarios. This is where you would normally discover if sampling works for you; How many scenarios do you need to obtain stable results? Beyond what you can handle? This should be followed by out-of-sample testing or simulation. This is where your model of uncertainty comes back to you and defines 'truth'.

If you have a history, and thereby an empirical

distribution, you might feel safe to sample from it. Literature is full of papers doing exactly that, and it seems to be accepted. But if you are out there with real people and real money, at least stop for a moment and think: Is the empirical distribution really good enough? Do the data come from an underlying distribution? Is there any reason to assume that the empirical distribution can be seen as a sample from the distribution you have in mind? Or do the observations come from a process that is all over the place? Maybe your statistics background can help you here, I am simply pointing out that just assuming your data points constitute a valid distribution is very bold.

This is part of a problem that is mostly skipped in academic papers. Unless we are modeling betting in a Casino, there is no 'true distribution', at least not one you can get hold of. So a true out-of-sample test cannot be made relative to the real world. But by being careful when you estimate uncertainty, at least you have a basis for saying that the problem you solved is (almost) as if you had used all you know about the distribution, and not just the scenarios. And at this point defining the empirical distribution as the truth can be very risky.

Scenario generation methods that cover high dimensions, dependencies, and difficult distributions (like binary) are emerging. May I for example refer to [1] where a problem with over 25,000 dependent random variables is solved with reasonable accuracy or [4] which makes an attempt to handle complicated random variables, in particular binary ones, possibly combined with continuous. Such a combination will kill almost all existing scenario generation methods. I mention these mostly to encourage more work on scenarios in real settings, and not to tell exactly how to go about generating them.

Acknowledgements.

This work is supported by The Research Council of Norway under grant 308790.

REFERENCES

- [1] Zhaoxia Guo, Stein W. Wallace, and Michal Kaut. Vehicle routing with space- and time-correlated stochastic travel times: Evaluating the objective function. *INFORMS Journal on Computing*, 31(4):654–670, 2019.

- [2] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.
- [3] Alan J. King, Stein W. Wallace, and Michal Kaut. Scenario-tree generation. In *Modeling with Stochastic Programming*, Springer Series in Operations Research and Financial Engineering, chapter 4. Springer, 2012.
- [4] Vit Prochazka and Stein W. Wallace. Scenario tree construction driven by heuristic solutions of the optimization problem. *Computational Management Science*, 17:277–307, 2020.

Real-World Impact of Stochastic Programming: The Electricity Sector Case

Alexandre Street and Davi Valladão

LAMPS PUC-Rio (BRAZIL)

street@puc-rio.br and davimv@puc-rio.br

Impact Summary: *Stochastic programming emerged as a response to the need for more realistic and practical solutions when dealing with decision-making under uncertainty. Power system operators, planners, and agents worldwide need to coordinate and plan reliable decisions under a myriad of uncertainties every day. If consumers pay hundreds of billions of dollars every year to support such decisions, how much is left on the table when ignoring the benefits of stochastic programming decisions? And what are the big challenges this sector still offers to our community? We hope we can produce at least a differential impact on your research interest with this article. To that end, we bring a few examples of the tremendous real-world impacts our community has already produced to electrical power sectors worldwide and raise awareness of relevant challenges that still need to be addressed.*

Acknowledgement and disclaimer – We are honored and greatly appreciated the invitation to write this article. We thank Francesca Maggioni and Güzin Bayraktan, and the Committee on Stochastic Programming (COSP), for such a great opportunity. In return, we tried to provide the SP community the distillate of many discussions and ideas about how

stochastic programming (SP), observing the specificity of each application, can be impactful in real life. We focus on the specific case of the electricity sector, where billions of dollars are driven by primal and dual solutions of optimization problems every year. That said, an important disclaimer: this is an opinion article, and so we do not intend to make a literature review on covered topics. It is also worth mentioning that we intentionally use self-citations with the sole purpose of bringing further scientifically grounded material to better support our ideas. Thus, we apologize in advance for all notable related works not cited in the subsequent paragraphs.

The need for real-world impact is the underlying and impelling force behind most applied and theoretical developments in the optimization field. The necessity of improving life quality, making things better, or even survive creates the demand for new methods, models, and theories supporting decisions that will produce the actions for change. Thus, one of the first responses to this need for thrive is the art of mathematical modeling through abstract thinking. As described in [1], the modeling process was significantly systematized with the definition of traditional structures we all use today to define an optimization problem: objective function, variables, and constraints. Such systematization was key to concentrate efforts allowing the development of powerful and effective methods to solve entire classes of problems instead of relying on specialized methods for specific applications.

Frequently, when the strive for new impactful models and methods holds a sort of equilibrium between generality, encompassing a relatively large set of problems worth solving, and similarity, allowing the scientific community to rely on a common notation (or language), a new area emerges. This happened in the optimization field with the SP area. Not so long ago, when linear programming was starting to impact industry applications and change the economic thinking, decision-making under uncertainty applications started to appear (see [1]). Thus, SP emerged as a branch of the optimization field (or mathematical programming) concerned with bringing this relevant piece of realism (i.e., uncertainty) into play. In this matter, SP added a few key new structures to the modeling framework such as 1) the concept of first-stage (here-and-now) decisions and

recourse (wait-and-see) actions, 2) the explicit uncertainty characterization through probability distributions, uncertainty or ambiguity sets, and 3) the representation of the decision maker's risk preference.

Among many impactful insights, a distinctive and powerful message successfully conveyed by the SP community was the following: under certain assumptions, there is an extra value in the stochastic solution (VSS) (see [2]) that might not be obtained from any deterministic model. There are relevant cases in which the SP models find optimal solutions that can not be found by any deterministic version of the formulation, regardless the scenario used for its uncertain parameters. As we further depict in the sequel, this and other relevant findings produced profound impacts in real-world applications such as in energy, finance, logistics, just to mention a few.

The electric power sector constitutes one of the examples where SP produced relevant real-world impacts. It is one of the largest and most complex distributed machines ever built. It must be operated, maintained, and updated so that all consumers have reliable and continuous access to electricity supply almost everywhere. Such high quality of service is expected to happen despite the weather, economic situation, market frictions, equipment failures, natural hazards, fuel price fluctuations, and natural resources availability. According to the U.S. Energy Information Administration (EIA) Annual Energy Review, the U.S. total retail sales to final consumers in 2019 reached 3,811 TWh at an average price of 105.40 USD/MWh. This means that final consumers paid 401.6 billion dollars to supply the whole electricity machinery. The same number surpasses 2 trillion dollars for the countries in the Organisation of Economic Co-operation and Development (OECD). A relevant part of the costs mentioned above, roughly around 58%¹, goes to the generators to cover their investment and operating costs. Another lower, yet relevant, parcel of 13% covers transmission costs. Because both planning and operation activities are largely driven by optimization models since early 90's, and at the same time these segments are significantly challenged by uncertainties, SP has been gaining more and more attention in the last 30 years.

For instance, the VSS in the challenging task of planning the transmission system expansion has been

proven to be quite relevant [3]: *"We conclude that the cost of ignoring uncertainty (the cost of using naive deterministic planning methods relative to explicitly modeling uncertainty) is of the same order of magnitude as the cost of first-stage transmission investments."* On the operational side, SP models have been playing a crucial role in addressing the problem of mid- and long-term water resource management in hydrothermal power systems since the seminal work of Mario V. Pereira [4]. In 2000, the Stochastic Dual Dynamic Programming (SDDP) methodology was adopted as the official model in Brazil. It was used to centrally define a week-ahead generation plan and electricity spot prices based on the marginal operation cost directly derived from dual variables. Therefore, this academic work and its subsequent advances directly impacted the Brazilian power system agents and the whole chain of electric-intensive industries and final consumers. Since that time, the whole Brazilian economy started to experience electricity prices influenced by the opportunity cost of an inter-temporal constrained natural resource (water) calculated by a multistage stochastic linear model. The use of SDDP-based techniques to either minimize total system costs in centrally coordinated systems or to maximize revenues in bid-based markets was largely followed by many countries relying on high shares of hydro resources such as Chile, Peru, Colombia, and Norway, just to mention a few.

The benefits brought by SP to the electrical power sector notwithstanding, many challenges still need to be addressed to unlock a substantial part of the SP potential. Some of them are: 1) the modeling risk associated with time-inconsistent policies generated by modeling simplifications in multistage models, 2) transparency, reproducibility, and compliance issues associated with sample-dependent decisions, and 3) the connections with other areas such as machine learning that aims to harness the power of a more-and-more data-rich world.

The potentially negative impact of a time-inconsistent policy in real-world applications is real, tangible, and it really should be taken seriously. We know that multistage models better represent the reality of dynamically chained decision (under uncertainty) processes over time. Such improved representation is done by incorporating the dynamics of future planned decisions into the model. Thus, in the-

¹Shares vary according to the system characteristics.

ory (and *ceteris paribus*), this framework better characterizes future opportunity costs of key scarce resources, such as water, and time-coupling constraints through a well-approximated recourse function. As a consequence, the first-stage decisions are improved in terms of the entire time horizon. In practice, however, the decision maker must solve a temporal sequence of multistage problems to actually implement only first-stage decisions. Hence, for each problem, the first-stage solution embeds a look-ahead assessment of future flexibilities and system constraints translated by future planning decisions.

Roughly speaking, a policy is time consistent when the planning decisions, optimal for today's problem, are also optimal as first-stage decisions of future problems (in [5] a more precise and complete definition is given). The most popular and widely explored source of time inconsistency in the related literature is a bad choice of the risk measure [5]. However, another relevant source of inconsistency arises from modeling simplifications due to the intractability of more realistic and complex multistage models. The latter dramatically manifests in real-world hydrothermal operation planning problems, in which the first-stage problem (addressing implementable decisions) is full of details, while the planning part of the multistage model (recourse problem) is widely simplified. Relevant examples for the previously mentioned simplifications are: network constraints, hydro reservoir aggregation, the information level available for the decision maker in each stage, hourly constraints such as ramping limits and unit commitment constraints, and short-term uncertainties of intermittent renewables and equipment failures. Although relevant efforts have been devoted to reduce these modeling gaps, there is still much work to be done.

In this context, the inconsistency due to modeling simplifications in the planning part of multistage problems can be understood as a modeling risk caused by an opportunity-cost assessment bias, and can be measured by the time inconsistency gap [6]. Not surprisingly, the aforementioned bias is amplified by the usual approach of relaxing constraints (optimistic view) when simplifying the planning part of the multistage model. Practical consequences of relying on such simplifications are high over-costs and relevant market distortions such as price spikes. We

refer to the interested reader to [7] for further details. Although one cannot fully address this source of inconsistency, as reality will always manifest itself more complex than any model, the pursuit for new approaches aiming to reduce the detrimental effects of optimistic recourse functions is key to unlock the potential benefits of multistage SP in real-world applications.

Another important practical challenge is related to sampling-based methodologies, by far the dominant approach used to solve SP problems. For instance, despite of the well-known potential benefits of stochastic unit commitment (SUC) models (see [8]) for day-ahead markets², official short-term models still rely on deterministic approaches worldwide. This happens mainly because SUC models are generally addressed through their Sample Average Approximation (SAA) counterparts, which are scenario-dependent large-scale mixed integer linear problems (MILP). Hence, due to the computational burden associated with such problems, SUC models rely on few scenarios, leading to sample-dependent (or random) solutions. Such models play a central role in electricity markets, defining market prices and the supply and demand equilibrium at each point of the network. Therefore, market operators are overwhelmed by agents inquiring for optimality guarantees and reproducibility, hungry for any subjectivity justifying solutions in which their revenues are better off. Thus, differently from what would be acceptable for an individual agent within the extent of his own private decision-making perspective, there is no room for a SUC model delivering, e.g., sample-dependent prices. As a consequence, *ad hoc* biased forecasts and more sophisticated application-driven learning methods are being proposed by system operators. We refer to [9] for more details highlighting the Californian System Operator's case. Notwithstanding, this should sound like ringing bells calling for new ideas on exact methods observing industry's needs.

Finally, classical SAA models intrinsically rely on multivariate time-series/machine-learning models describing complex spatial-temporal dynamics

²A unit commitment model defines the on and off statuses of generating units, as well as their planned generation and (up and down) reserve commitments for the next 24 hours. It constitutes a typical step in power system operation procedures.

through Monte Carlo simulations to achieve good performances in real-world applications. Furthermore, it requires a delicate orchestration of two areas, statistics and optimization, generally not entirely dominated in the level needed to make things work properly by the same set of researchers. Thus, the pursuit of more realistic data-driven approaches [10] and contextual models [11] capable of capturing relevant information from a more and more data-rich world has become a hot topic and has a great potential of impact in real-world applications.

REFERENCES

- [1] G. B. Dantzig, "Linear programming," *Operations research*, vol. 50, no. 1, pp. 42–47, 2002.
- [2] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*. Springer Science & Business Media, 2011.
- [3] F. D. Munoz, B. F. Hobbs, J. L. Ho, and S. Kasina, "An engineering-economic approach to transmission planning under market and regulatory uncertainties: Wecc case study," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 307–317, 2013.
- [4] M. V. Pereira and L. M. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Mathematical programming*, vol. 52, no. 1, pp. 359–375, 1991.
- [5] A. Pichler, R. P. Liu, and A. Shapiro, "Risk averse stochastic programming: time consistency and optimal stopping," *To appear in Operations Research*.
- [6] B. Rudloff, A. Street, and D. M. Valladão, "Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences," *European Journal of Operational Research*, vol. 234, no. 3, pp. 743–750, 2014.
- [7] A. Brigatto, A. Street, and D. M. Valladão, "Assessing the cost of time-inconsistent operation policies in hydrothermal power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4541–4550, 2017.
- [8] A. Papavasiliou and S. S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations Research*, vol. 61, no. 3, pp. 578–592, 2013.
- [9] J. D. Garcia, A. Street, T. Homem-de Mello, and F. D. Muñoz, "Application-driven learning via joint estimation and optimization for demand and reserves requirement forecast," *arXiv preprint arXiv:2102.13273*, 2021.
- [10] A. Velloso, A. Street, D. Pozo, J. M. Arroyo, and N. G. Cobos, "Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 958–969, 2019.
- [11] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *Management Science*, vol. 66, no. 3, pp. 1025–1044, 2020.

Real World Impact of Stochastic Programming for COVID-19

Claudia Sagastizábal

IMECC-Unicamp & CEMEAI (BRAZIL)

sagastiz@impa.br

Coronavirus has been an uninvited guest in our lives for more than a year now. After a last trip to Trondheim for a mini-course (sigh), home office became my new normal from March 2020 on. The new normal also brought a different style for broadcasting news in the media. We are informed on the pandemic using plots and terms that were before considered coded language for geeks: moving averages, exponential growth, flattening the curve, the ubiquitous, yet elusive, effective reproduction rate.

Being a problem solver by nature, I became increasingly frustrated, overwhelmed by the avalanche of semi-technical information. I knew nothing about epidemiology, but wondered if the optimization could contribute to the fight against COVID-19. In May 2020 I reached out to my colleagues in CEMEAI, the Brazilian center for mathematical sciences applied to industry. About 15 years ago, São Paulo state started funding a few excellence centers to explore the frontiers of knowledge. CEMEAI's mission is to develop new transformative mathematical techniques, with an emphasis on their industrial applications. I wrote an email to the list of researchers in the center, proposing to think how to estimate the number of lives that are spared when the population adheres to social distancing measures. In pandemic times, having a positive indicator instead of reporting mortality numbers sounded like a more effective message to convey.

I sent my email without much hope, a modern cast-away shipwrecked by the pandemic, throwing a bottle with a message into the virtual sea of internet. I was surprised by the replies. I found out that CEMEAI had stepped in with force to respond to the COVID-19 pandemics. Less than two months had passed since the coronavirus outbreak in Brazil and several solutions, with powerful evocative names, had already been proposed on different fronts [1]. Info Tracker, to collect and analyze coronavirus data at municipal levels; the expert system Safe Stock to forecast PPE use and prevent a depletion of stock in hospitals; and an optimization platform to plan intermittent confinement, Robot Dance.

Robot Dance is an open-source software created by L. Gustavo Nonato, Tiago Pereira and Paulo J. S. Silva to support the decision making on COVID-19 public policies [4]. In Brazil many of such actions are taken at state levels, without federal coordination. Robot Dance provides regional governments with technical assistance on the pandemic (as in many places, unfortunately, technical recommendations are sometimes set aside for political reasons).

Partly based on Robot Dance, in the project [6] we put in place *Vidas Salvas*, a “calculator” of lives that are spared every minute in different regions of Brazil, thanks to social distancing policies. The repercussion of our relatively simple web-page was quite impressive, it was even mentioned on the national TV news. It is interesting how a shift in perspective, moving the indicator from the topic of deaths to the preservation of lives, succeeded in attracting attention and, hopefully, encourage people to maintain social isolation.

Vidas Salvas was the beginning of a fruitful collaboration with CEMEAI colleagues. The optimization model behind Robot Dance is quite sophisticated, it combines elements from various areas, and Stochastic Programming is one of them. The computational tool assesses and forecasts the consequences of interventions when there is a disease like COVID-19, whose spread depends on the circulation of people living in a region. The problem is modeled in JuMP and solved with the nonlinear optimization solver Ipopt.

As we all know, formulating a mathematical optimization problem requires defining an objective function and the feasible set. We shall comment on the

former afterwards. Regarding the latter, a set of epidemiological constraints describes different stages of the disease, considering a given population in percentual terms. This is represented mathematically by variables called “compartments” in epidemiology. A dynamical system describes the evolution in time of the percentages of individuals that are Susceptible of getting the disease, of those who have been Exposed to the coronavirus, of the Infected ones, and of those who already had COVID-19. The initial letter of each compartment, SEIR, gives the name to the model in question (R stands for “recovered”, even though the compartment includes deceased individuals, it appears that the model was originally proposed for non-lethal diseases). Those four compartments are state variables of the optimization problem, and a finite-difference discretization of the ordinary differential equations in the dynamical system constrains the corresponding trajectories, day by day, over an horizon of at least one year. For those readers familiar with hydro-power management in energy optimization, compartments are comparable to reservoir volumes, and the discretized dynamical system is akin to the well-known water balance equations (nonlinear in Robot Dance).

In the set of epidemiological constraints, the mean of new infections caused by a single infected individual determines how the disease spreads, when the pandemic ends, which portion of the population needs to be vaccinated to declare the infection controlled. This is the effective reproduction number, denoted by r_t for each time step. Since transmission occurs when people meet, keeping r_t low amounts to restricting circulation or imposing some social distancing measure. In Robot Dance, r_t is the control variable.

Brazil has large urban centers surrounded by many dormitory towns from which people commute to work. In order to account for mobility, the SEIR dynamics in Robot Dance incorporates the flow of day and night circulation through a certain contact matrix. For each sub-region i , the i th row of this matrix specifies average interactions with other sub-regions, collected from anonymized data of cellular phones [2, Section 3]. The epidemiological constraints involve variables that are now indexed by sub-region and time step.

A planning platform like Robot Dance has to bal-

ance several goals, some stated as constraints, some as metrics in the objective function. Maintaining healthcare capacity is fundamental to help minimizing the disruption to society. With this aim, we endowed the model with probabilistic constraints. In the infected compartment, there is a certain percentage that will need attention in the intensive care unit. Robot Dance models this ratio as a stochastic process $icu_t^i(\omega)$, and sets a chance constraint to keep the use of intensive care unit beds below cap_t^i , the maximum capacity in the i th sub-region at time t :

$$\mathbb{P} \left[icu_t^i(\omega) \sum_{\tau=t-7}^t I_\tau^i \leq cap_t^i \right] \geq 0.95,$$

where the sum over seven days reflects the average time spent in intensive care units. Official records, corrected by a factor that estimates under-reported cases, gives a history of ratios. In turn, this data is used to calibrate a time series that approximates the stochastic process. Having an explicit expression for the time series makes it possible to reformulate the probabilistic constraint into an equivalent deterministic inequality that is affine [2, Section 4].

Given r_0 , the basic reproductive number representing life in the “old normal”, before the outbreak, the problem is

$$\left\{ \begin{array}{l} \min_{r_t^i \in [0, r_0]} \sum_{t=1}^{360} \psi_t(r_t) \\ \text{s.t.} \quad (SEIR, r) \text{ satisfy the discretized} \\ \quad \quad \quad \text{epidemiological relations} \\ \quad \quad \quad I \text{ satisfies the explicit relations} \\ \quad \quad \quad \text{for the probabilistic constraint.} \end{array} \right.$$

In this large-scale nonconvex program, the objective function mirrors different measures the government wishes to impose. If the desired policy is to encourage maximal circulation, $\psi_t(\cdot)$ represents the mean deviation between r_t^i and r_0 . It is also possible to include a total variation term to avoid too abrupt changes in the control, as well as terms promoting an alternation of strict measures in nearby cities. This last strategy, suggested in the blog [3], that takes turns between a “hammer” of lockdown and a “dance” with open economy, explains the name Robot Dance.

For the state of São Paulo in Brazil, Robot Dance was able to pinpoint one particular weak link in the complex network of the state, with more than 20 different health districts or sub-regions. It appeared

that one minor link was crucial for the success of confinement measures in a state with more than 40 million inhabitants. São Paulo city is home to about 25% of the state’s population but harbors about 70% of the state’s intensive care unit beds. Because of the low capacity of beds in one small sub-region that has an intense flow with the state capital, São Paulo city, a very long lockdown period was necessary to contain the disease. By examining the output of Robot Dance, it became clear that if the state capital shared a small portion of beds with a few satellite dormitory towns, the trajectory of the pandemic could be controlled by imposing a shorter lockdown in the whole state, [2, Section 5.4]. Without resorting to chance-constrained optimization, performing only scenario analysis as often in the literature, the solution proposed by Robot Dance could not have been found.

The platform was extended to design a testing campaign, in a context when PCR-tests are scarce and no mechanism for contact tracing is put in place (like in Brazil). The model in [5] includes a compartment of quarantined population and additional control variables, the number of tests performed at each time step in each sub-region, together with logistic constraints on testing capacity. Once more, optimization provided a solution and insights that were superior to “natural” policies. The Figure shows the output of Robot Dance,



a strategy that tests proportionally to the number of inhabitants, and two other naive protocols: testing only on big urban centers, or prioritizing small cities (because they have reduced hospital capacity). Rather than concentrating the tests in São Paulo city, Robot Dance testing focuses on the nearby neighboring regions with little to no testing in the capital. The optimal strategy found with Robot Dance is twice more effective than all the considered alternatives.

We are currently working on optimal deployment of vaccination campaigns, trying to understand to which extent stretching the time between two doses

can help relieving pressure on intensive care units. This requires including age groups and vaccine related states in the SEIR model. Considering the current vaccine scarcity, the delivery delays, and that new virus strains seem to be more contagious, this is a pressing problem in Brazil nowadays.

REFERENCES

- [1] CEMEAI-ModCOVID, <http://www.cemeai.icmc.usp.br/ModCovid19/>.
- [2] L. G. Nonato, P. Peixoto, T. Pereira, C. Sagastizábal, and P. J. S. Silva, *Robot Dance: A mathematical optimization platform for intervention against COVID-19 in a complex network*, Optimization Online 2020.
- [3] T. Pueyo, *Coronavirus: The Hammer and the Dance, March 2020*, <https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56>.
- [4] P. J. S. Silva, T. Pereira, and L. G. Nonato, *Robot dance: a city-wise automatic control of COVID-19 mitigation levels*, medRxiv 2020.
- [5] P. J. S. Silva, T. Pereira, C. Sagastizábal, L. G. Nonato, M. Cordova, and C. J. Struchiner, *Smart testing and critical care bed sharing for COVID-19 control*, arXiv 2021.
- [6] P. J. S. Silva and C. Sagastizábal, *Vidas salvas pelo isolamento social*, <http://www.cemeai.icmc.usp.br/ModCovid19/vidas-salvas/>, May 2020.

Stochastic Programming Events in 2021

Below please find a list of events related to Stochastic Programming in 2021. **The situation due to COVID-19 is rapidly changing. Therefore, please check the provided links for the latest information.**

- Conference title: *SIAM Conference on Optimization*
Where/when: Spokane, Washington, U.S. July 20 - 23, 2021
Minisymposium devoted to Stochastic and Robust Optimization (organizers: Andy Sun and Wolfram Wiesemann)
More information: <https://www.siam.org/conferences/cm/conference/op21>.
- Conference title: *ECSSO-CMS-2021 Conference*
Where/when: Ca' Foscari University of Venice, July 7–9, 2021, **Postponed to 2022. Updates will be published in the website:** <https://lnkd.in/gG987EP>
- Conference title: *24th International Symposium on Mathematical Programming (ISMP)*
Stream devoted to Stochastic optimization (organizers: Güzin Bayraksan, Francesca Maggioni and Peter Richtárik). Where/when: Beijing, China, August 15-20, 2021, **Postponed to August 14-19, 2022. Updates will be published in the website:** <http://ismp2022.csp.escience.cn/dct/page/1>
- Conference title: *International Conference on Optimization and Decision Science (ODS2021)*
Stream: Optimization under uncertainty (organizers: Patrizia Beraldi and Francesca Maggioni)
Where/when: Rome, September 14–17, 2021
More information: <https://www.optit.net/events/ods-2021/>
- Workshop title: *Optimization under Uncertainty*
Where/when: Montréal, September 27 - October 1, 2021. **Please check the website for updates:** <https://lnkd.in/gdHHVgW>
- Conference Title: *INFORMS Annual Meeting 2021*
Stream: Optimization under uncertainty (Cluster Chair: Dr. Weijun Xie)
Where/When: Anaheim, CA (In-person or Virtual) October 24-27, 2021
More information: If you are interested in giving a talk and/or chairing a session in this cluster, please enter your talk or session at
- Conference title: *31st European Conference on Operational Research*
Where/when: Athens, 11-14 July, 2021
Stream and sessions devoted to Stochastic and Robust Optimization (organizers: Miloš Kopa, Francesca Maggioni and Steffen Rebennack)
More information: <https://euro2021athens.com/>.

<https://tinyurl.com/1tqi49pl>. Should you have any questions, please contact Dr. Weijun Xie (wxie@vt.edu).

ICSP XVI

July 24-28, 2023 Davis, California, USA
David Woodruff

UC Davis Graduate School of Management (USA)
 DLWoodruff@UCDavis.edu

Change of date! We are going to hold the sixteenth International Conference on Stochastic Programming (ICSP) during the last week of July 2023: **July 24–28, 2023**. If you are on the program committee for any other conferences, please try to avoid that week.

We are starting work on the website:

<https://gsm.ucdavis.edu/faculty-and-research/faculty-conferences/xvi-international-conference-stochastic-programming>

If you want to organize a session or a track, please let us know at DLWoodruff@UCDavis.edu



Figure 1: UC Davis campus

Committee on Stochastic Programming (COSP) Members

Chair: Güzin Bayraksan

Integrated Systems Engineering, The Ohio State University,
 Columbus, OH (USA)
 bayraksan.1@osu.edu

Secretary: Francesca Maggioni

Department of Economics University of Bergamo (Italy)
 francesca.maggioni@unibg.it

Treasurer: Bernardo Pagnoncelli

School of Business, Universidad Adolfo Ibáñez (Chile)
 bernardo.pagnoncelli@uai.cl

Alois Pichler

Faculty of Mathematics, University of Technology, Chemnitz
 (Germany)
 alois.pichler@mathematik.tu-chemnitz.de

Johannes O. Royset

Operations Research Department, Naval Postgraduate School,
 Monterey, CA (USA)
 joroyset@nps.edu

Wim van Ackooij

EDF R&D, Osiris (France)
 wim.van.ackooij@gmail.com

Phebe Vayanos

Viterbi School of Engineering, University of Southern California,
 Los Angeles, CA (USA)
 phebe.vayanos@usc.edu

Webmaster: Wolfram Wiesemann

Business School, Imperial College, London (United Kingdom)
 ww@imperial.ac.uk

David Woodruff

UC Davis Graduate School of Management, Davis, CA (USA)
 dlwoodruff@ucdavis.edu