Mathematisches Forschungsinstitut Oberwolfach

Report No. 38/2018

DOI: 10.4171/OWR/2018/38

New Directions in Stochastic Optimisation

Organised by
Jesús De Loera, Davis
Darinka Dentcheva, Hoboken
Georg Ch. Pflug, Vienna
Rüdiger Schultz, Essen

19 August – 25 August 2018

Abstract. Recent work in stochastic programming draws on interaction with Algebraic and Combinatorial Models, Numerical Analysis of PDEs, Risk Analysis, Mathematical Equilibrium, Minimax, and Stochastic Games. For the first time ever, the workshop brought together scholars from these fields to exchange experience and identify promising topics for future research.

Mathematics Subject Classification (2010): 90C15.

Introduction by the Organisers

The workshop *New Directions in Stochastic Optimisation* organized by Jesús De Loera (Davis), Darinka Dentcheva (Hoboken), Georg Ch. Pflug (Vienna) and Rüdiger Schultz (Essen) was well attended by 54 participants with broad geographic representation. By a surprising coincidence, the workshop took place precisely 50 years after the first Oberwolfach Workshop in Operations Research (Organizers: R. Henn (Karlsruhe), H. P. Künzi (Zurich) and H. Schubert (Kiel)).

Topic: Stochastic programming offers mathematically rigorous optimisation models incorporating probabilistic information with uncertain data. Decisions are based exclusively on the information that is available at the moment of taking decisions (nonanticipativity). Depending on when missing information is unveiled and on how this interacts with decision making over time, different principal model setups arise, e.g., one-stage, two-stage, or multi-stage models. Selection and placement (in the objective or the constraints) of the statistical parameters according to which relevant random variables are to be evaluated is another important issue.

This allows to express perceptions such as reliability, risk neutrality, or risk aversion. Finally, the nature of the initial uncertain optimisation problem (linear or nonlinear, with or without integer variables, living in finite or infinite dimension) has crucial impact on the arising stochastic programming model. These aspects lead to a wide variety of stochastic optimization programs as well as to a wide variety of mathematical techniques for their analysis and algorithmic treatment.

Course of the workshop: On the one hand, the workshop reflected the diversity of the involved areas. On the other hand, there was enough overlap among individual expertise to generate new ideas and obtain input from other directions. In particular, the 35 talks together with a brainstorming session on Thursday evening provided the basic ideas for stimulating discussions covering a broad spectrum of topics. We now will discuss some contributions on specific topics.

Numerical Analysis of PDEs: The rapid development of PDE-constrained optimisation is accompanied by approaches to handle uncertainty in appropriate fashion. The approach via stochastic programming aims at finding best possible decisions under data uncertainty. Procedures for reaching optimality in terms of stochastic criteria may incorporate user attitudes such as being risk averse, risk neutral or risk seeking. A related contemporary approach to handling uncertainty, also discussed at the workshop, is uncertainty quantification that looks for "a computational framework for quantifying input and response uncertainties in a manner that facilitates predictions with quantified and reduced uncertainty"[1]. Shape optimization under stochastic loading for elastic materials have been discussed in several talks. Risk neutral and risk averse objective functionals have been investigated and the concept of stochastic dominance constraints was explored. Thereby the expected excess and the excess probability are taken into account, first as objective functional involving the compliance cost of an elastic object under stochastic loading and then as a constraint when comparing a shape with a benchmark shape.

Open problems concern the identification of subclasses allowing for duality and resulting algorithmic shortcuts; exploiting problem similarities for efficient repeated solution of PDE-constrained optimisation problems differing in the realisations of the random data; mathematical foundation and algorithm design of approximation via linearisation of full models, arriving at linear models, development of numerical PDE-solvers taking into account specifics of this problem class.

PDE-constrained optimization under uncertainty has been addressed from further points of view. We mention convergence of projected gradient methods in Hilbert spaces, reduced-order models/incorporating risk functionals /(CVaR) and risk averse optimization; optimal boundary control under uncertain initial data; general characterizations of feasibility, stationarity, optimality, and stability in PDE-constrained optimization.

[1]R.C. Smith: Uncertainty Quantification, SIAM, 2014, page ix.

Quasi-Monte-Carlo Methods: Another classical topic with relevance in stochastic programming is Quasi-Monte-Carlo Approximation. At the workshop, it was reported that randomized Quasi-Monte Carlo (RQMC) methods when used to tackle mixed-integer two-stage stochastic programs can achieve their optimal convergence order of almost $O(n^{-1})$ although integrands are discontinuous and non-differentiable in their continuity regions. Applications include randomly shifted lattice rules, randomly scrambled Sobol' point sets, and Monte Carlo methods to solve a mixed-binary electricity portfolio model with normal prices and demands.

Algebraic and Combinatorial Models: The last two decades have seen notable progress in developing structural insight and improving algorithmic capabilities in stochastic integer programming. Compared to their continuous counterparts, which enjoy convexity on a broad range and whose analysis substantially benefits from convexity, linear stochastic programs with mixed-integer variables are non-convex and even may face discontinuities in the objective and/or the constraints, so that fresh ideas are necessary. The incorporation of computational algebraic geometry to address non-convexity and integrality has become more and more vital. A research field that has emerged recently is devoted to optimality certificates provided by ideal bases, and their geometric analogues. In test-set augmentation algorithms for two-stage stochastic integer programming certificates enjoy decomposability and stabilisation with growing numbers of scenarios which is extremely useful in algorithm design. At the workshop an account on optimality certificates involving Graver sets for block-structured integer programs of which two-stage stochastic programs are a special case was given. It concluded with recent results on improved complexity estimates for augmentation algorithms with Graver-type improving vectors.

Utilization of polynomial algebra for non-convex stochastic optimisation problems whose objectives and/or constraints are given by polynomials, is another promising departure point for extending the ability to analyze and solve stochastic programs with nonlinearities. The discussion circled around relationships between nonnegative real polynomials, sums of squares, and semidefinite programming highlighting computation of sum-of-squares decompositions of polynomials using semidefinite programs. Stochastic mixed-integer programs have been considered from different perspectives: For multi-stage models, decision rules for the Lagrangian dual based on relaxation of either the non-anticipativity or the stage-wise constraints were introduced. For simple integer recourse models, a reinterpretation of approximation techniques in terms of asymptotically accurate inexact cutting planes was presented.

Single- and Multi-Period Risk: Nowadays, rigorous treatment of risk in an optimization context covers static and more and more often also dynamic optimisation models. Mostly, however, under the assumption of purely exogenous uncertainty or in Markov decision processes. At the workshop approaches to theory and application in the presence of decision dependency have been discussed.

On the one hand, investigations on chance constraints under endogenous uncertainty have been reported. Additionally, new extensions of the theory of risk measure to the case of moving probability spaces were presented. A comprehensive framework for risk models with applications in decision dependent sampling has been introduced. Further talks ranged from algorithmic and dynamic aspects of multistage optimization, via decision-dependent ambiguity sets in distributionally robust optimization, to optimality conditions for partially observable Markov Decision Processes.

Chance Constraints in Infinite Dimension: Traditional research topics like, two- and multi-stage or chance constrained stochastic programs, still today, provide the frame for stimulating research. Probabilistic (or chance) constraints, originally mostly employed in finite dimension, now occur more and more often as instruments for ensuring reliability also in infinite dimension. This raises open questions on basic structural properties including statements on stability under perturbations (of Euclidean parameters up to underlying probability measures). Some first answers regarding semi-continuity, convexity, and stability were given at the workshop for models combining PDE and chance constraints.

Robust versus stochastic paradigms: Relations between robust and stochastic modeling in optimization have been considered from various angles. Utility preference robust optimization relies on tailored integrands for risk functionals to cope with partial information about the decision maker's utility function. Concerning efficient computation, the finding of optimal distributionally robust predictors via tractable convex programs has been discussed for ambiguity sets that are optimal in a specific context. Dependence uncertainty has been considered for a portfolio selection problem where the marginal return distributions of the assets can be estimated with high accuracy, whereas the dependence structure between the assets remains ambiguous.

Equilibrium: Topics related to mathematical equilibrium ranged from algorithms for rank-1 bimatrix games, via quasi-hyperbolic discount rates to account for dynamic inconsistency in sequential decision models with changing preferences, to structure and stability of linear stochastic bilevel problems. Moreover, a computational framework for the algorithmic treatment of multiple optimization problems coupled with an equilibrium constraint and a bilevel approach to automated regularization parameter selection in image processing have been discussed.

Data and Decisions: The growing interest in data-driven analysis stimulates the study of topics seemingly apart from each other. Among the aspects discussed at the workshop there are (i) Learning enabled optimization, (ii) stochastic variational inequalities and (iii) the interplay of variational analysis and nonparametric statistics. Additionally, first results in risk-averse statistical learning in the context of clinical trails were discussed.

Markets: A class of risky capacity equilibrium problems reflecting future stochastic production in a spot market, when agents are risk-averse and can buy financial

hedges for their physical investments has been discussed in a unified Nash equilibrium framework. For risk trading in complete markets, existence of solutions has been proven. In a similar spirit, the impact of risk-aversion on participants in (incomplete) energy markets has been addressed as well.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1641185, "US Junior Oberwolfach Fellows".

Workshop: New Directions in Stochastic Optimisation

Table of Contents

Abstracts

Perspectives in probabilistic programming under continuous random distributions

René Henrion

The standard form of a probabilistic programming problem is [5]

$$\text{minimize } f(x) \quad \text{subject to}$$
$$\mathbb{P}(g_i(x,\xi) \leq 0 \ (i=1,\ldots,m) \geq p, \quad x \in\subseteq \mathbb{R}^n \quad ,$$

where ξ is a finite-dimensional (continuously distributed) random vector, $\mathbb{P}$ a probability measure and $p \in [0,1]$ some fixed probability level. Here it is assumed that x is decided on before observing ξ. This form of probabilistic programming is still prevailing in the literature. On the other hand, some formally obvious generalizations are quickly attracting interest due to new applications: This leads to the following three perspectives in probabilistic programming discussed in the talk:

- The stochastic inequality system is not finite but infinite.
- The decisions are not just finite-dimensional but elements of a Banach space
- For decisions and random vectors representing discrete time processes, the static setting above is replaced by a dynamic setting taking into account the gain of information while observing the random process.

The first two items lead us to the discussion of the model

$$\text{minimize } f(x) \quad \text{subject to} \tag{1}$$
$$\mathbb{P}(g(x,\xi,t) \leq 0 \ (t \in T \subset \mathbb{R}^m) \geq p, \quad x \in X = \text{Banach space},$$

where the cardinality of index set T may be infinite. For instance, the index set might represent some time interval (e.g., in the probabilistic level control of water reservoirs) or some domain in space (in PDE constraint optimization subject to probabilistic state constraints [6]). It could also refer to some uncertainty set of a parameter t which, in contrast to ξ is not endowed with statistical information. A typical application is optimization in gas networks subject to random loads ξ and unknown friction coefficients t in the pipes [2]. The latter aspect was motivation to call the constraint in (1) a *probust* constraint (short for: probabilistic/robust) [3].

A major challenge in the theoretical analysis and algorithmic treatment of problem (1) is to infer analytical properties of the probability function

$$\varphi(x) := \mathbb{P}(g(x,\xi,t) \leq 0 \ (t \in T \subset \mathbb{R}^m) \tag{2}$$

from the properties of the input data, i.e., the function g and the distribution of ξ. This connection is detailed in the talk for properties like lower- and upper

semicontinuity, convexity [6] and (sub-)differentiability [1, 4]. The obtained results allow one to derive existence and stability theorems of problem (1) which apply to certain types of PDE constrained optimization problems with random state constraints and to set up an algorithmic approach in the case of Gaussian-like distributions which was successfully applied to optimization problems in stationary gas networks.

The last of the three items presented above motivated the consideration of the problem

$$\begin{aligned} &\text{minimize } \mathbb{E}f(x_1, x_2(\xi_1), \dots x_m(\xi_1, \dots \xi_{m-1}), \xi) \quad \text{subject to} \\ &\mathbb{P}(g_i(x_1, x_2(\xi_1), \dots x_m(\xi_1, \dots \xi_{m-1}), \xi) \leq 0 \ (i = 1, \dots, l)) \geq p. \end{aligned} \tag{3}$$

Here, the essential difference with the static model (1) consists in understanding time-dependent decisions not in a static way but as policies reacting on previously observed random parameters. While this setting is familiar in multistage stochastic programming, the understanding of its probabilistic version is still in its infancy. The talk presents corresponding semi-continuity results for the probability function associated with (3) and provides fully explicit Lipschitz and differentiability results for the simple two-stage setting ($m = 2$) with separable (w.r.t. x and ξ) functions g_i. A major challenge discussed at the workshop addresses the concept of *time consistency* for such models.

References

[1] W. Van Ackooij anf R. Henrion, *(Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution*. SIAM/ASA J. Uncertainty Quantification **5** (2017), 63–87.

[2] C. Gotzes, H. Heitsch, R. Henrion and R. Schultz, *On the quantification of nomination feasibility in stationary gas networks with random load*, Mathematical Methods of Operations Research **84** (2016), 427–457.

[3] T. Gonzalez Grandon, H. Heitsch and R. Henrion, *A* joint model of probabilistic /robust constraints for gas transport management in stationary networks, Computational Management Science **14** (2017), 443–460.

[4] A. Hantoute, R. Henrion and P. Pérez-Aros, *Subdifferential characterization of probability functions under Gaussian distribution*, to appear in: Mathematical Programming, appeared online.

[5] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.

[6] M.H. Farshbaf-Shaker, R. Henrion and D. Hömberg, *Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization*, to appear in: Set-Valued and Variational Analysis, appeared online.

Learning Enabled Optimization (LEO)

Suvrajeet Sen

Optimization principles pervade many engineering disciplines, economic sciences, and of course computational/mathematical sciences. Whether studying classification using linear/quadratic programming for support vector machines (SVM) or logistic regression using a specialized version of Newton's methods, optimization

algorithms have provided a strong foundation for a variety of statistical/machine learning (S/ML) applications. Indeed, today's learning algorithms could be easily described as "optimization enabled learning". This approach essentially sets up a parameter estimation model which finds the "best fit" from a space of models (e.g. regression, neural networks, and others). Learning Enabled Optimization (LEO), a new paradigm to support seamless integration between two fundamental structures: a) learning models/algorithms which are based on optimizing goodness of "fit", and b) decision models which seek "near-optimal" decisions in constrained optimization applications. This new paradigm leads to composite stochastic optimization problems which support a fusion of a) and b) above. This fusion is best illustrated by the figure below. Figure 1.a) depicts the typical workflow associated with S/ML, whereas, the LEO workflow may be depicted as shown in Figure 1.b).

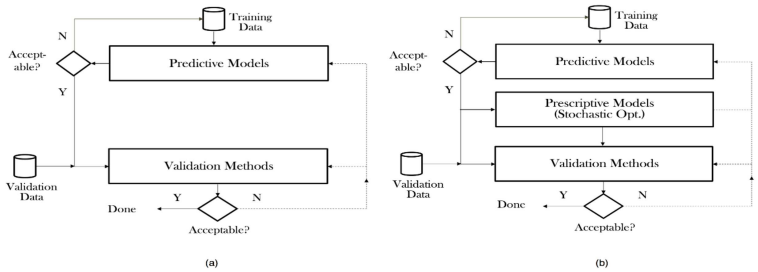


FIGURE 1. a) Statistical/Machine Learning (ML) b) Learning Enabled Optimization (LEO)

1. How is LEO different from Stochastic Programming?

In order to pose the kind of questions which motivate LEO, consider a common S/ML situation in which we have a dataset of i.i.d. observations $\{(Z_i, W_i)_{i=1}^N\}$ which represent data for random covariates (Z, W), generically referred to as predictors and responses respectively. Ordinarily, the S/ML model helps to discover a relationship (e.g., a regression) between the covariates (Z, W). Depending on the model, such a relationship can be used to predict outcomes of the response variable W, given a realization $Z = z$. This aspect of using a model to characterize the behavior of the response, given the predictor value z is usually not the purpose of SP (Birge and Louveaux 2011), although there are some examples (e.g., Support Vector Machines – SVM) which can be formulated as SP. To illustrate the connections between Empirical Risk Minimization of ML, and a *data-driven SP with Sample Average Approximation* (SAA), we state the SVM model as an instance of SP. This formulation will also help us make the distinctions between SP and LEO.

Consider a binary classification problem whose data includes a predictor Z_i and response (classification) $W_i \in \{+1, -1\}$. The goal of such an SVM is to identify

the "best-fit" parameters $(\beta_0, \beta) = \boldsymbol{\beta} \in \mathbb{R}^{p+1}$ such that $\beta_0 + \beta^\top Z_i \geq W_i - \delta_i$ for $W_i = +1$, and $\beta_0 + \beta^\top Z_i \leq W_i + \delta_i$ for $W_i = -1$; here $\delta_i \geq 0$ for all i, and the goal is to solve the following SP.

$$\min \left(\frac{\rho}{2}\right) \boldsymbol{\beta}^\top \boldsymbol{\beta} + \frac{1}{N} \sum_{i=1}^{N} \delta_i, \qquad \delta \geq 0. \tag{1}$$

In the terminology of SP, the variables$\boldsymbol{\beta}$ are known as the first-stage (or "here-and-now") variables, whereas δ_i are known as the second-stage variables which depend on the specific outcome W_i. Traditionally, it is common to solve (1) using a QP solver, although more recently, they are being solved using SP-type decomposition (Schwartz et al 2011).

From the viewpoint of mathematical structures, an SAA model of a generic two-stage SP has the following form:

$$\min_{\boldsymbol{x} \in X \subset \mathbb{R}^n} c(\boldsymbol{x}) + \mathbb{E}[h(\boldsymbol{x}, \tilde{\omega})], \tag{2}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator, and $\boldsymbol{x} \in X$ is a generic optimization decision variable. Assuming that the random variable $\tilde{\omega}$ is discrete, and takes values $\{(Z_i, W_i)\}$ each with probability $1/N$, (2) subsumes (1),and algorithms designed for (2) are directly applicable to (1); that is, it is straightforward to interpret (2) in the context of empirical observations, and replace $\mathbb{E}[h(\boldsymbol{x}, \tilde{\omega})]$ by its finite sum representation $\sum_i \frac{1}{N} h(\boldsymbol{x}, \omega_i)$. In addition, (2) will serve another purpose: it will help us clarify the distinctions between SP and LEO.

The simplest version of a LEO model, the so-called LEO with disjoint spaces, has the following form

$$\min_{\boldsymbol{x} \in X \subset \mathbb{R}^n} c(\boldsymbol{x}) + \mathbb{E}[h(\boldsymbol{x}, W | Z = z)], \tag{3}$$

where the expectation is taken with respect to random variable $W|Z = z$ and the value z is given. As in the learning literature, we assume that the data is only available in the form of covariates $\{(Z_i, W_i)\}$ and hence, one needs a statistical model to infer a distribution. A somewhat more difficult model arises in the case where decision variable $\boldsymbol{x}$ assumes values in the same space as the random variable Z. Such models arise in design optimization, marketing optimization and others where a decision may be looked upon as a "bet". In this case, the model takes the following form

$$\min_{\boldsymbol{x} \in X \subset \mathbb{R}^n} c(\boldsymbol{x}) + \mathbb{E}[h(\boldsymbol{x}, W | Z = \boldsymbol{x})], \tag{4}$$

A model of the form (4) is the so-called LEO with shared spaces.

2. Novel Research Questions

While problems (2), (3) and (4) highlight the differences between SP and LEO, they suppress the composite nature of the optimization necessary for LEO. However, it is not difficult to see that the need to infer an approximate distribution of the random variable $W|Z$ raises several novel research questions:

(a) Is deterministic optimality as suggested in (4) an appropriate goal, given that the composition of estimation and stochastic optimization creates a model which may be far too demanding for deterministic optimality? Notice that in the S/ML world, the inexactness of stochastic gradient descent (SGD) has been credited with providing solutions which are more flexible/generalizable than solving S/ML models to deterministic optimality. Is there an appropriate analog, say statistical optimality, which can quantify the probability of a solution being ϵ-optimum?
(b) How do we define generalizability for LEO models?
(c) The diamonds in Figure 1 are intended as Hypothesis Tests for Model Validation. Such tasks are well known in the statistics literature (which is captured in Figure 1.a). Are similar tests for model validation possible for LEO models.
(d) In contrast with Figure 1.a, we observe that there are several possible loops within Figure 1b. Each loop characterizes a particular combination of the manner in which the fusion of learning and optimization can be performed. For instance, the cycle between Learning and Optimization in Figure 1b suggests that there could be a tight coupling between the two aspects of LEO. Such coupled models (CLEO) are likely to be non-convex and questions regarding optimality become much more treacherous under the circumstances. Nevertheless, ideas from non-convex optimization become relevant in such cases.

These new problems are being posed in the spirit of the purpose of seeking new frontiers in Stochastic Optimization. It is my hope that the participants will find these research questions to be intriguing. Some preliminary ideas are presented in Sen and Deng 2017. This is ongoing work with my Ph.D. student Yunxiao Deng.

References

[1] Birge, J.R. and F. Louveaux 2011, "Stochastic Programming", Springer, Berlin.
[2] Sen, S. and Y. Deng 2017, "Learning Enabled Optimization", Optimization Online.
[3] Shalev-Shwartz, S., Y. Singer, N. Srebro, and A. Cotter, 2011 "Pegasos: Primal estimated sub-gradient solver for SVM", Mathematical Programming, 127(1):3-30.

Convergence of randomized Quasi-Monte Carlo methods for mixed-integer two-stage stochastic programs

Werner Römisch
(joint work with Hernan Leövey)

By extending our earlier work [2, 4] we study convergence properties of randomized Quasi-Monte Carlo (RQMC) methods for solving mixed-integer two-stage stochastic optimization models. Although the integrands are discontinuous and nondifferentiable in their continuity regions, we show that their lower order ANOVA terms can be smooth if certain geometric conditions are satisfied. If the effective dimension of the integrands is small, too, the integrands are approximately smooth and, hence, RQMC methods can achieve their optimal convergence order of almost $O(n^{-1})$ (see [1]). In case of normal distributions we show that the geometric conditions are satisfied almost everywhere with respect to the Haar measure on the topological group of real orthogonal matrices. The latter are needed to decompose the covariance matrix into diagonal form. We also show that the effective dimension of the integrand can be essentially reduced by applying principal component analysis factorization to the covariance matrix instead of classical Cholesky factorization. The theory is illustrated by presenting numerical results for randomly shifted lattice rules [6], randomly scrambled Sobol' point sets [5] and Monte Carlo methods to solve a mixed-binary electricity portfolio model with normal prices and demands (see [3]). Instead of 10^4 Monte Carlo samples one only needs only about 10^2 samples for randomly scrambled Sobol' point sets and randomly shifted lattice rules. The advantages consist in the improved accuracy for given sample size or in shorter running times for smaller sample sizes. The latter becomes crucial for high-dimensional optimization models.

References

[1] J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, Acta Numerica **22** (2013), 133–288.
[2] H. Leövey and W. Römisch: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, Mathematical Programming **151** (2015), 315–345.
[3] H. Leövey and W. Römisch: Quasi-Monte Carlo Methoden fr Optimierungsmodelle der Energiewirtschaft mit Preis- und Last-Prozessen , in: Optimierung in der Energiewirtschaft, VDI-Bericht **2266**, VDI-Verlag, Dsseldorf, 2015, 287–298.
[4] H. Heitsch, H. Leövey and W. Römisch, Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs?, Computational Optimization and Applications **65** (2016), 567–603.
[5] A. B. Owen: Scrambled net variance for integrals of smooth functions, Annals of Statistics **25** (1997), 1541–1562.
[6] I. H. Sloan, F. Y. Kuo and S. Joe: Constructing randomly shifted lattice rules in weighted Sobolev spaces, SIAM Journal on Numerical Analysis **40** (2002), 1650–1665.

Risk averse optimal stopping

Alois Pichler
(joint work with Alexander Shapiro)

Abraham Wald was the first to consider stopping rules in statistics (see [1]) and stopping times have been well studied since then.

This talk introduces stopping risk measures which allow for stopping. The Snell envelope is defined as in the classical theory, although based on the risk measures rather than the expectation. Further, risk averse, multistage stochastic optimization problems are formulated including stopping times. In this context, the Snell envelope corresponds to the solution of multistage programs with initial decisions fixed. We characterize optimal stopping policies in this general context.

We further relate the risk measures to dynamic equations. We provide dynamic equations for both situations in stochastic optimization, with and without optimal stopping.

References

[1] A. Wald. Sequential analysis. 1947.

From Data to Decisions: Distributionally Robust Optimization is Optimal

Daniel Kuhn
(joint work with Bart Van Parys and Peyman Mohajerin Esfahani)

We study stochastic programs where the decision-maker cannot observe the distribution of the exogenous uncertainties but has access to a finite set of independent samples from this distribution. In this setting, the goal is to find a procedure that transforms the data to an estimate of the expected cost function under the unknown data-generating distribution, *i.e.*, a *predictor*, and an optimizer of the estimated cost function that serves as a near-optimal candidate decision, *i.e.*, a *prescriptor*. As functions of the data, predictors and prescriptors constitute statistical estimators. We propose a meta-optimization problem to find the least conservative predictors and prescriptors subject to constraints on their *out-of-sample disappointment*. The out-of-sample disappointment quantifies the probability that the actual expected cost of the candidate decision under the unknown true distribution exceeds its predicted cost. Leveraging tools from large deviations theory, we prove that this meta-optimization problem admits a unique solution: The best predictor-prescriptor pair is obtained by solving a distributionally robust optimization problem over all distributions within a given relative entropy distance from the empirical distribution of the data.

The optimal distributionally robust predictor identified in this paper can be evaluated by solving a tractable convex optimization problem. Under standard convexity assumptions about the feasible set and the cost function of the stochastic program, the corresponding optimal prescriptor can also be evaluated in

polynomial time. Although perhaps desirable, the tractability and distributionally robust nature of the optimal predictor-prescriptor pair are not dictated *ex ante* but emerge naturally.

Relative entropy ambiguity sets have already attracted considerable interest in distributionally robust optimization [1, 2, 4, 6, 9]. Note, however, that the relative entropy constitutes an *asymmetric* distance measure between two distributions. The asymmetry implies, among others, that the first distribution must be absolutely continuous to the second one but not *vice versa*. Thus, ambiguity sets can be constructed in two different ways by designating the reference distribution either as the first or as the *second* argument of the relative entropy. All papers listed above favor the second option, and thus the emerging ambiguity sets contain only distributions that are absolutely continuous to the reference distribution. Maybe surprisingly, the optimal predictor resulting from our meta-optimization problem uses the reference distribution as the *first* argument of the relative entropy instead. Thus, the reference distribution is absolutely continuous to every distribution in the emerging ambiguity set. Relative entropy balls of this kind have previously been studied in [3, 5].

Adopting a Bayesian perspective, the paper [3] determines the smallest ambiguity sets that contain the unknown data-generating distribution with a prescribed level of confidence as the sample size tends to infinity. Both Pearson divergence and relative entropy ambiguity sets with properly scaled radii are optimal in this setting. In the terminology of the present paper, [3] thus restricts attention to the subclass of distributionally robust predictors and operates with an asymptotic notion of optimality. The meta-optimization problem proposed here entails a stronger notion of optimality, under which the distributionally robust predictor with relative entropy ambiguity set emerges as the unique optimizer. The paper [5] also seeks distributionally robust predictors that trade conservatism for out-of-sample performance. It studies the probability that the estimated expected cost function dominates the actual expected cost function uniformly across all decisions, and it calls a predictor optimal if this probability is asymptotically equal to a prescribed confidence level. Using the empirical likelihood theorem of [7], it shows that Pearson divergence and relative entropy ambiguity sets with properly scaled radii are optimal in this sense. This notion of optimality has again an asymptotic flavor because it refers to decreasing sequences of ambiguity sets that converge to a singleton, and it admits multiple optimizers.

This talk is based on the working paper [8].

References

[1] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[2] G.C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

[3] V. Gupta. Near-optimal ambiguity sets for distributionally robust optimization. *Available from Optimization Online*, 2015.

[4] Z. Hu and L.J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. *Available from Optimization Online*, 2013.
[5] H. Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *arXiv preprint arXiv:1605.09349*, 2016.
[6] H. Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016.
[7] A.B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.
[8] B. Van Parys, P. Mohajerin Esfahani and D. Kuhn. From data to decisions: Distributionally robust optimization is optimal. *Available from Optimization Online*, 2017.
[9] Z. Wang, P.W. Glynn, and Y. Ye. Likelihood robust optimization for data-driven newsvendor problems. *Computational Management Science*, 12(2):241–261, 2016.

Statistical inference of two stage stochastic programs using stochastic variational inequality techniques

Shu Lu
(joint work with Yang Yu)

Stochastic variational inequalities provide a unified framework for modeling and analyzing stochastic optimization and stochastic equilibrium problems. In this talk, we provide a method for computing confidence regions and confidence intervals for the true optimal solution of a two-stage stochastic program, based on the solution to a sample average approximation problem [1]. This method is developed by combining existing results in [2] on the asymptotic behavior of optimal solutions of two stage linear stochastic programs with recently developed tools on inference of stochastic variational inequalities. This method does not rely on the strict complementarity assumption, which has been a standard assumption in previous works.

A stochastic variational inequality (SVI) is of the form

$$0 \in f_0(x) + N_S(x), \tag{1}$$

where $f_0(x) = E[F(x,\xi)]$ is the expectation function of a random function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$, ξ is a random vector supported on a closed set $\Xi \subset \mathbb{R}^d$, S is a polyhedral convex set in $\mathbb{R}^n$, and $N_S(x) \subset \mathbb{R}^n$ denotes the normal cone to S at x:

$$N_S(x) = \{v \in \mathbb{R}^n \mid \langle v, s - x\rangle \leq 0 \text{ for each } s \in S\}.$$

The SVI (1) can arise as the first-order necessary condition of a stochastic optimization problem. More specifically, consider a stochastic optimization problem over the feasible set S whose objective is the expectation function of a random function $\Phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$:

$$\min_{x \in S} E[\Phi(x,\xi)].$$

If its objective function is differentiable at a local minimizer x_0, then x_0 satisfies the first-order necessary condition $-\nabla_x E[\Phi(x_0,\xi)] \in N_S(x_0)$, which becomes a SVI when $\nabla_x E[\Phi(x_0,\xi)] = E[\nabla_x \Phi(x_0,\xi)]$.

In many applications the expectation function f_0 does not have a closed form expression and is approximated by a sample average function. Let $\xi^1, \cdots, \xi^N$ be

i.i.d. variables with distribution same as that of ξ. Define the sample average function $f_N : \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ by

$$f_N(x,\omega) = n^{-1} \sum_{i=1}^{n} F(x, \xi^i(\omega)).$$

The sample average approximation (SAA) problem is to find a point $x \in S$ such that

$$0 \in f_N(x,\omega) + N_S(x). \tag{2}$$

A solution x_N to (2) is known to converge to a true solution x_0 to (1) almost surely as the sample size N grows, and their asymptotic distributions and exponential convergence rates have also been obtained [3, 4, 5, 6, 7]. Under conditions that ensure (1) to have a locally unique solution under small perturbations of f_0, the asymptotic distribution for x_N is known to be "piecewise normal." In other words, it is the image of a normally distributed random variable under a piecewise linear function. Under an additional condition known as "strict complementary slackness" in the standard nonlinear programming setting, that piecewise linear function reduces to a linear function and the distribution becomes normal.

In practice, because the true solution is unknown, it is difficult to check whether the strict complementary slackness condition is satisfied or not. Moreover, when this condition is not satisfied, standard methods to estimate the asymptotic distribution based on the SAA solution will fail due to the nonsmoothness associated with the piecewise normal distribution. In [8, 9, 10, 11] we provide a method to compute asymptotically exact confidence regions and confidence intervals for the true solution, based on a single SAA solution. Our method does not require making the strict complementary slackness assumption.

Next, consider a two-stage stochastic program

$$\begin{aligned} &\min\ c^T x + E[Q(x,\xi)] \\ &\text{s.t. } Ax = b,\ x \geq 0 \end{aligned} \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$, and $Q(x,\xi)$ is the optimal value of the second-stage problem in which $\xi = (q, T, W, h)$:

$$\begin{aligned} &\min\ q^T y \\ &\text{s.t. } Tx + Wy = h,\ y \geq 0. \end{aligned}$$

Let $\xi_i = (q_i, h_i, T_i, W_i), i = 1, \cdots, N$ be i.i.d. samples of ξ, and y_i be an optimal solution of the ith second stage problem. Replacing $\mathbb{E}[Q(x,\xi)]$ by $\sum_{i=1}^{N} Q(x,\xi_i)$ leads to the SAA problem which can be written as

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} c^T x + \sum_{i=1}^{N} q_i^T y_i \\ &\text{s.t. } Ax = b, x \geq 0, \\ &\quad T_i x + W_i y_i = h_i,\ y_i \geq 0,\ i = 1, \cdots, N. \end{aligned} \tag{4}$$

With $\Phi(x,\xi) = c^T x + Q(x,\xi)$ and $S = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$, (3) is in the form $\min_{x\in S} E[\Phi(x,\xi)]$, whose corresponding SVI formulation is $-E[\nabla_x \Phi(x,\xi)] \in N_S(x)$. However, for any fixed ξ, $\Phi(x,\xi)$ is a piecewise linear function of x, so $\nabla_x \Phi(x,\xi)$ is not well defined at breakpoints of the piecewise linear function. Furthermore, the piecewise constant function $\nabla_x \Phi(x,\xi)$ cannot provide a good sample average approximation for $E[\nabla_x \Phi(x,\xi)]$ in the space of continuously differentiable functions. This prevents a direct application of inference techniques developed for stochastic variational inequalities to (3).

The problem caused by the nonsmoothess of $\Phi(x,\xi)$ can be bypassed by defining a function $\delta_N(x) = \frac{1}{N}\sum_{i=1}^N \Phi(x,\xi_i) - E[\Phi(x,\xi)]$, and considering an intermediate problem

$$\min_{x\in S} E[\Phi(x,\xi)] + \langle \nabla \delta_N(x_0), x\rangle. \tag{5}$$

Under some standard assumptions, $\nabla\delta_N$ is well defined almost surely at the solution x_0 of (3), and the objective function of (5) is twice differentiable on a neighborhood of x_0. Accordingly, we can treat (5) as the SAA problem of an SVI to obtain the asymptotic distribution of its optimal solution, which in turn gives the asymptotic distribution of the optimal solution of (4) based on a result in [2]. Combining such asymptotic distribution with techniques in [8, 9, 10, 11], we can compute confidence regions and confidence intervals for the solution of (3) given a solution to (4).

References

[1] S. Lu and Y. Yu, *Statistical inference of two stage stochastic programs via stochastic variational inequalities*, Working paper, 2018.

[2] A. Shapiro, *Asymptotic behavior of optimal solutions in stochastic programming*, Mathematics of Operations Research **18** (1993), 829–845.

[3] J. Dupačová, and R. Wets, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, The annals of statistics **16** (1988), 1517–1549.

[4] G. Gürkan, A.Y. Özge and S.M. Robinson, *Sample-path solution of stochastic variational inequalities*, Mathematical Programming **84** (1999), 313–333.

[5] A.J. King and R.T. Rockafellar, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Mathematics of Operations Research **18** (1993), 148–162.

[6] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MPS–SIAM Series on Optimization, SIAM and MPS (2009).

[7] H. Xu, *Sample average approximation methods for a class of stochastic varational inequality problems*, Asia-Pacific Journal of Operational Research **27** (2010), 103–119.

[8] M. Lamm and S. Lu, *Generalized conditioning based approaches to computing confidence intervals for solutions to stochastic variational inequalities*, Mathematical Programming (2018), https://doi.org/10.1007/s10107-018-1279-z.

[9] M. Lamm, S. Lu and A. Budhiraja, *Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities*, Mathematical Programming **165** (2017), 151–196.

[10] S. Lu, *Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities*, SIAM Journal on Optimization **24** (2014), 1458–1484.

[11] S. Lu and A. Budhiraja, *Confidence regions for stochastic variational inequalities*, Mathematics of Operations Research **38** (2013), 545–568.
[12] S. M. Robinson, *Normal maps induced by linear transformations*, Mathematics of Operations Research **17** (1992), 691–714.

Variational Analysis of Infinite-Dimensional Stochastic Optimization Problems with Applications to Nonparametric Statistics

Johannes O. Royset

Variational analysis has been central in the development of optimality conditions, approximation schemes, and computational methods for stochastic optimization problems. In this lecture, we present recent advances in the formulation and solution of *infinite-dimensional problems using variational metrics*. Spaces of upper-semicontinuous (usc) functions with the hypo-distance are bounded and even compact under mild conditions [2], and exhibit low metric entropy in many cases [1]. These properties together with connections with weak convergence of distributions and convergence of minimizers and maximizers make such metric spaces attractive choices for many stochastic optimization problems [5, 7, 3]. The lecture demonstrates possibilities in the context of *statistical M-estimators*.

Given d_0-dimensional random vectors $X^1, \ldots, X^n$, we consider M-estimators

$$\hat{f}^n \in \epsilon^n\text{-}\operatorname{argmin}_{f\in F^n} \frac{1}{n}\sum_{j=1}^{n} \psi(X^j, f) + \pi^n(f), \tag{1}$$

where F^n is a class of candidate functions on $\mathbb{R}^d$, or a subset thereof, possibly varying with n, ψ is a loss function; for example $\psi(x, f) = -\log f(x)$ in maximum likelihood (ML) estimation of densities and $\psi((x, y), f) = (y - f(x))^2$ in least-squares regression, and π^n is a penalty function possibly introduced for the purpose of smoothing and regularization.

We address the challenges of existence, consistency, and computations of constrained M-estimators by viewing the class of functions under consideration as a subset of a particular metric space of usc functions equipped with the hypo-distance. Although viewing M-estimators as minimizers of empirical processes indexed by a metric space is standard, our *particular* choice is novel. The only precursors are [4, 6], which hint to developments in this direction without a systematic treatment. Three main advantages emerge from the choice of metric space: (i) A unified and disciplined approach to rich classes of functions becomes possible as the hypo-distance can be used across M-estimators. (ii) Consistency of plug-in estimators of modes of densities, maximizers of regression functions, and related quantities follows immediately from consistency of the underlying estimators. (iii) Computation of estimates becomes viable because usc functions, even when defined on unbounded sets, can be approximated by certain parametric classes to an arbitrary level of accuracy in the hypo-distance.

We bypass uniform laws of large numbers (LLN) and instead rely on a one-sided *epi-LLN* for which upper bounds on the loss function ψ becomes superfluous.

Thus, concern about density values near zero and the need for reformulations in ML estimation vanish. The epi-LLN establishes convergence in some sense across the whole class of functions as in the case of a uniform LLN, but under less burdensome conditions. Thus, consistency results are not hampered by model misspecification or other circumstances under which an estimator is constrained away from an actual function. It also becomes immaterial whether the estimator and the actual function are unique. Under misspecification in ML estimation, just to mention one case, there can be an uncountable number of densities that have the same Kullback-Leibler divergence to the one from which the data is generated. Our results still hold.

In our framework, conditions for existence and consistency of estimators essentially reduce to checking that the class of functions F^n is closed under the hypo-distance. We establish that many natural classes of functions are also closed in the hypo-distance. Specifically, we show this for classes defined by convexity, log-concavity, monotonicity, s-concavity, monotone transformations, Lipschitz continuity, pointwise upper and lower bounds, location of modes, height at modes, location of level-sets, values of moments, size of subgradients, approximate evaluation of integrals, splines, multivariate total positivity of order two, and *any combination* of the above, possibly under some minor technical conditions. No prior study has established existence and consistency of M-estimators for such a variety of constraints.

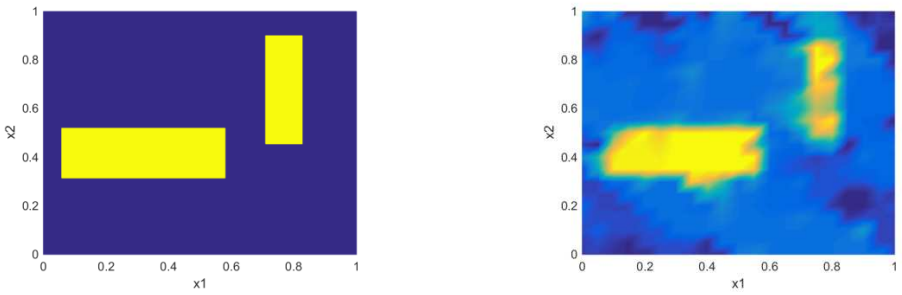


FIGURE 1. Top view of actual density (left) and estimate using sample of size $n = 1000$ (right).

Example. As a concrete example of a rich class of densities in ML estimation on $\mathbb{R}^d$, suppose that $\alpha, \kappa \geq 0$; $C, D \subset \mathbb{R}^d$; $I \subset [0, \infty]$ is closed; $g, h : \mathbb{R}^d \to [0, \infty)$, with h being usc and also satisfying $\int h(x)dx < \infty$; and

$$F = \Big\{ f : \mathbb{R}^d \to [0,\infty] \Big| f \text{ usc}, \int f(x)dx = 1,\ C \subset \operatorname{argmax}_{x \in \mathbb{R}^d} f(x), D \subset \operatorname{lev}_{\geq \alpha} f$$

$$\sup_{x \in \mathbb{R}^d} f(x) \in I,\ g(x) \leq f(x) \leq h(x),\ |f(x) - f(y)| \leq \kappa \|x - y\|_2, \forall x, y \in \mathbb{R}^d \Big\},$$

where $\operatorname{lev}_{\geq \alpha} f = \{x \in \mathbb{R}^d \mid f(x) \geq \alpha\}$ is a super-level set of f. We restrict the consideration to densities with (global) modes covering C and "high-probability

regions" covering D. Neither C nor D need to be singletons. We permit nearly arbitrary pointwise bounds on allowable densities. In settings with little data but substantial experience about what an estimate "should" look like, such constraints can be helpful modeling tools. Let $\mathbb{N} = \{1, 2, \dots\}$.

Proposition. Suppose that $X^1, X^2, \dots$ are iid random vectors, each distributed according to a density $f^0 : \mathbb{R}^d \to [0, \infty]$, F is nonempty, and $\{\epsilon^n \geq 0, n \in \mathbb{N}\} \to 0$. Then the following hold almost surely:

(i) For all $n \in \mathbb{N}$, there exists $\hat{f}^n \in \epsilon^n\text{-}\operatorname{argmin}_{f \in F} -n^{-1} \sum_{j=1}^n \log f(X^j)$.
(ii) Every hypo-cluster point of $\{\hat{f}^n, n \in \mathbb{N}\}$, of which there is at least one, minimizes the Kullback-Leibler divergence to f^0 over the class F.
(iii) If $f^0 \in F$, then f^0 is the hypo-limit of $\{\hat{f}^n, n \in \mathbb{N}\}$.

We also give rates of convergence results and approximation theory that supports computations.

Numerical Results. We consider ML estimation of the mixture of three uniform densities on $[0, 1]^2$ depicted in Figure 1(left). The resulting mixture density f^0 has height $f^0(x) = 3$ for x in the areas colored yellow and $f^0(x) = 0.6150$ elsewhere. Using a sample of size 1000, we compute a penalized ML estimate over the class

$$F = \Big\{ f : [0,1]^2 \to [\alpha, \beta] \;\Big|\; \int f(x)dx = 1,\ \{\bar{x}, \bar{y}\} \subset \operatorname{argmax}_{x \in [0,1]^2} f(x),$$
$$|f(x) - f(y)| \leq \kappa \|x - y\|_2, \forall x, y \in [0,1]^2,\ \text{pw-affine on simplicial partition} \Big\}.$$

A simplicial complex partition divides $[0, 1]^2$ into N equally sized triangles, which reduces the problem to that of finite-dimensional convex optimization. Thus, F can be viewed as an approximation, introduced for computational reasons, of the class obtained from F by relaxing the piecewise affine restriction. We adopt the penalty $\pi(f) = \lambda \sum_{k=1}^N \|g_i\|_1$, where g_i is the gradient of the ith affine function defining f. In the results reported here, $\kappa = 100$ with $\bar{x} = (0.4702, 0.4657)$ and $\bar{y} = (0.7746, 0.7773)$. We observe that F is misspecified as f^0 is not Lipschitz continuous. Figure 1 illustrates the case with $\lambda = 0.02$, $\alpha = 0.3075$, $\beta = 4.5$, and $N = 800$. Experiments show that argmax-constraints regularize the estimates.

References

[1] J. O. Royset, *Approximations of Semicontinuous Functions with Applications to Stochastic Optimization and Statistical Estimation*, ArXiv e-prints 1709.06730 (2018).
[2] J. O. Royset, *Approximations and solution estimates in optimization*, Mathematical Programming, **170**(2) (2018), 479–506.
[3] J. O. Royset and R. Wets, *Variational Analysis of Constrained M-Estimators*, ArXiv e-prints 1702.08109 (2018).
[4] J. O. Royset and R. J-B Wets, *From data to assessments and decisions: Epi-spline technology*. In A. Newman, editor, *INFORMS Tutorials*. INFORMS, Catonsville (2014).
[5] J. O. Royset and R. J-B Wets, *Fusion of hard and soft information in nonparametric density estimation*, European J. Operational Research, **247**(2) (2015), 532–547.

[6] J. O. Royset and R. J-B Wets, *Multivariate epi-splines and evolving function identification problems*, Set-Valued and Variational Analysis, **24**(4) (2016), 517–545. Erratum: 547–549.
[7] J. O. Royset and R. J-B Wets, *Variational theory for optimization under stochastic ambiguity*, SIAM J. Optimization, **27**(2) (2017), 1118–1149.

Lagrangian dual decision rules for multistage stochastic mixed-integer programming

James Luedtke
(joint work with Merve Bodur, Maryam Daryalal)

A multistage stochastic programming (MSSP) problem is an optimization problem where a vector of decisions x_1 is to made now, followed by an observation of a random variable ξ_1, which is followed by another set of decisions $x_2(\xi_1)$, and so on for T stages. In multistage stochastic mixed-integer programming (MSMIP), some of the decision variables are constrained to take on integer values. Let $\xi_1, \dots, \xi_T$ be a sequence of random variables, where $\xi_1 \equiv 1$ and let $\xi^t = [\xi_1, \dots, \xi_t]$ be the history of random variables observed at time $t \in [T] := \{1, \dots, T\}$. A MSMIP problem is formulated as follows:

$$\text{(1a)} \qquad \min \ \mathbb{E}_{\xi^T}\Big[\sum_{t\in[T]} c_t(\xi^t)^\top x_t(\xi^t)\Big]$$

$$\text{(1b)} \qquad \text{subject to} \ \ A_t(\xi^t)x_t(\xi^t) + B_t(\xi^t)x_{t-1}(\xi^{t-1}) = b_t(\xi^t), \quad t \in [T], \mathbb{P}\text{-a.s.}$$

$$\text{(1c)} \qquad x_t(\xi^t) \in X_t(\xi^t) \qquad\qquad t \in [T], \mathbb{P}\text{-a.s.}$$

where $X_t(\xi^t) := \{x \in \mathbb{Z}^{p_t} \times \mathbb{R}^{n_t} : C_t(\xi^t)x \geq d_t(\xi^t)\}$ and p_t and n_t are the number of integer and continuous variables in stage t, respectively. In this work, we assume the recourse constraint sets $X_t(\xi^t)$ are bounded for all $t \in [T]$ and all ξ^t, that the problem has *relatively complete recourse*: for each $t \in [T]$ and possible history of random variables ξ^t, and decisions $x_s(\xi^s)$ for $s = 1, \dots, t-1$, there always exists a feasible decision $x_t(\xi^t)$.

MSSP problems are generally intractable, and so an active area of research is to investigate methods to obtain approximate solutions to these problems, including feasible policies and bounds on the optimal value. In the context of multistage stochastic *linear* programming (in which all decision variables are continuous), Shapiro and Nemirovski [5] presented a primal linear decision rule (LDR) approximation, in which they restrict the decisions at each stage to be a linear function of random variables observed up to that stage. The expected value of using this policy provides an upper bound on the optimal value of the MSSP problem. On the other hand, Kuhn et al. [3] demonstrate how a lower bound can be obtained by applying the LDR restrictions to the dual policies. In our recent work [1], we introduced two-stage LDR approximations of MSSP. In the primal problem, this is accomplished by requiring *only* the state variables to be LDR, and in the dual problem, only the dual variables associated with the state equations are restricted to be LDR. These approximations provide better bounds than the static primal

and dual approximations, and may be approximately solved, for example, by using sample average approximation [4]. None of these LDR approximations extend generally to the stochastic mixed-integer programming case, because in the primal problem it is likely infeasible to constrain integer recourse decisions to be a linear function of random observations, and the dual approximations rely on (stochastic) linear programming duality.

To overcome these limitations, we propose to use LDR restrictions within *Lagrangian duals* of the MSMIP (1). Although a Lagrangian dual for a MSMIP problem does not generally have the strong duality property, it can still be useful for obtaining bounds on the optimal value of the MSMIP problem, which is our goal. We investigate the use of LDR in two specific Lagrangian duals: the stagewise dual (obtained by relaxing the state equations (1b)), and the nonanticipative dual (obtained by relaxing nonanticipativity constraints in a reformulation of the problem). In either case, once the LDR restriction is imposed, the restricted Lagrangian dual problem becomes a stochastic program in which the decision variables are the parameters of the dual LDR policy, and where subgradients for a given sample outcome can be obtained by solving deterministic mixed-integer programming (MIP) subproblems. Thus, these problems can be solved by sample average approximation or stochastic approximation methods.

For the two different Lagrangian duals we consider, we characterize the strength of the relaxation obtained by deriving an equivalent primal characterization, which has some similarity to the dual problem introduced in [3], but which also demonstrates the convexifying effect of the Lagrangian dual for the MSMIP context. In [2], it has been shown that the bound obtained from the *unrestricted* nonanticipative Lagrangian dual is at least as good as the bound obtained from the *unrestricted* stagewise Lagrangian dual. This result does not immediately extend, however, to the case when LDR restrictions are imposed in the dual problems. Using our primal characterizations, we demonstrate that, if the LDR basis functions for the nonanticipative Lagrangian dual are chosen in a particular way relative to those used in the stagewise Lagrangian dual, then the bound is at least as good.

Several interesting questions remain. One important question is how to choose the basis functions used in the LDR approximations. Some theory demonstrating that a class of basis functions is sufficient to converge to the true Lagrangian dual bound would be useful for guiding this choice. Another open problem is how to efficiently computationally optimize the LDR restricted Lagrangian dual problems. Finally, an idea that was raised at the Oberwolfach workshop was to investigate applying LDR restrictions in additional Lagrangian dual problems, such as those obtained by structural decomposition investigated in [2].

References

[1] M. Bodur and J. Luedtke. Two-stage linear decision rules for multi-stage stochastic programming. *Mathematical Programming*, 2018. in press.

[2] D. Dentcheva and W. Römisch. Duality gaps in nonconvex stochastic optimization. *Mathematical Programming*, 101(3):515–535, 2004.

[3] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.
[4] W.-K. Mak, D. P. Morton, and R. K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations research letters*, 24(1-2):47–56, 1999.
[5] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. In *Continuous optimization*, pages 111–146. Springer, 2005.

Graver basis methods for two-stage stochastic integer programs and other block structures

Matthias Köppe

The *Graver basis* $G(E)$ of an integer matrix $E \in \mathbb{Z}^{m\times n}$ is the unique inclusion-minimal finite set that generates all elements z of the integer kernel of E by nonnegative integer linear combinations of orthant-compatible vectors [7],

$$z = \sum_{i=1}^{k} \lambda_i g^i, \quad \lambda_i \in \mathbb{Z}_+, \quad g^i \in G(E), \quad g^i, z \text{ in the same orthant of } \mathbf{R}^n. \tag{1}$$

From the inclusion-minimality it follows that the elements of $G(E)$ are irreducible, i.e., they do not themselves have nontrivial decompositions of the form (1).

Consider the following basic **augmentation algorithm** for integer linear programs maximizing $c^T x$ subject to equality constraints $Ex = b$ and lower and upper bounds on the variables, $l \le x \le u$. Given the problem data E, b, c, l, u, a feasible point x^0, and Graver basis $G(E)$, start with $x \leftarrow x^0$ and run the iteration

$$\begin{aligned} &\text{while } \exists\, g \in G(E),\ c^T g > 0,\ x+g \text{ feasible} \\ &\quad x \leftarrow x + g, \end{aligned} \tag{A}$$

returning x. Assume that the problem has an optimal solution x^* and let x be an iterate, a feasible point. If x is non-optimal, i.e., $c^T x^* - c^T x > 0$, then the difference $z = x^* - x$ decomposes as (1), and thus by linearity of the objective,

$$\exists\, i \text{ such that } c^T g^i > 0. \tag{2}$$

Moreover, from the orthant-compatibility condition it follows that $x + g^i$ is feasible. Thus the algorithm finds an optimal solution x^*, terminating in finitely many steps because each improves the objective function value by at least 1.

Is this an *efficient* algorithm? Let us discuss this question in the standard framework of bit complexity, first relative to an oracle that gives suitable query access to $G(E)$. If the input size of the problem is the sum of the bitlengths of E, b, c, l, u, and x^0, then the gap $c^T x^* - c^T x^0$ is of (singly) exponential size (even if the dimension of the problem is considered a fixed constant), and thus we can only give a exponential bound on the number of iterations of the algorithm.

However, Algorithm (A) can be improved by various algorithmic **acceleration techniques** developed in the literature; see the book [3, § 3, § 4] for a comprehensive exposition. Some of these are inspired by techniques from combinatorial optimization; others are specific to Graver-based augmentation. An example of

the latter is the following. By the integer Carathéodory theorem, the number k in (1) can be bounded by $2n-2$. Thus, (2) can be sharpened to

$$\exists\ i \text{ such that } c^T\lambda_i g^i \geq \frac{1}{2n-2}(c^Tx^* - c^Tx^0). \tag{3}$$

Hence, the following *Graver-best* iteration [11],

$$\begin{aligned}&\text{while } \exists\ g \in G(E),\ c^Tg > 0,\ x+g \text{ feasible}\\ &\quad\text{find } (g^*,\lambda^*) \in G(E)\times\mathbb{Z}_+ \text{ maximizing } c^T\lambda g \text{ s.t. } x+\lambda g \text{ feasible}\\ &\quad x \leftarrow x+\lambda^* g^*,\end{aligned} \tag{$\mathrm{A_{Gb}}$}$$

achieves a geometric reduction of the gap and will therefore terminate within a polynomial number of iterations. It remains to find **efficient implementations of the Graver-best oracle** ($\mathrm{A_{Gb}}$, line 2). Two observations refine this idea: First, an approximate maximizer (g^*,λ^*) suffices; e.g., the step length λ can be chosen as a power of 2 [6]. Second, g does not actually have to be a Graver basis element [3, § 3.4]; we are allowed to optimize over a superset of $G(E)$ contained in $\ker_{\mathbb{Z}} E$. Thus, though software (`4ti2` [8]) is available for computing Graver bases explicitly, the most powerful applications of Graver basis technology only use structural results and **size estimates for Graver basis elements**. One estimation technique uses Steinitz' classic result in convexity that asserts the existence of a function $s(m)$ with the following property. Let $v_1,\dots,v_N \in \mathbf{R}^m$ with $v_1+\cdots+v_N = 0$; then there exists a reordering of the vectors so that all partial sums lie in the dilated polytope $sV = s(m)\cdot\operatorname{conv}\{v_1,\dots,v_N\}$. Using the Grinberg–Sevast'yanov estimate $s(m) \leq m$, Diaconis et al. [5] prove a bound for $\|g\|_1$ as follows: WLOG, let $g \geq 0$; take $v_1,\dots,v_N$ as the columns of E, repeated with multiplicities given by the components g_i; then $N = \|g\|_1$. The irreducibility of g implies that no two partial sums of the reordering are the same. Thus,

$$\|g\|_1 \leq \#(sV \cap \mathbb{Z}^m) \leq (2m)^m(m+1)^{m+1}\mathrm{D}(E), \tag{4}$$

where $\mathrm{D}(E)$ is a bound on the subdeterminants of E. However, using this general bound would forgo a key benefit of Graver-based analysis, namely to work with structured families in which $\mathrm{D}(E)$ is allowed to grow exponentially.

Estimates free of $\mathrm{D}(E)$, instead depending on parameters that are better to control, have become available for the **4-block structured matrices** [9] shown on the right. These include as special cases the two-stage stochastic integer programs ($C = D = O$, Hemmecke–Schultz [12]) and the N-fold programs ($B = C = O$, De Loera et al. [4]). A combination of structural results and size estimates allows us to construct and analyze **dynamic programming** techniques to implement the Graver-best oracle. The first Graver-based dynamic programming algorithm appeared in [10]; for a simplified exposition, see [3, § 4.1.3]. Moreover, **proximity**, i.e., the ℓ_∞-distance from an arbitrary optimal solution to the continuous relaxation to the closest optimal integer solution, can be bounded by $n\cdot\max\{\,\|g\|_\infty : g \in G(E)\,\}$ [3, § 4.3.3]. The resulting proximity bounds enable

$$\begin{pmatrix} C & D & D & \cdots & D\\ B & A & O & & O\\ B & O & A & & O\\ \vdots & & & \ddots & \\ B & O & O & & A \end{pmatrix}$$

generalizations to separable convex function minimization [3, § 4.3.5]; reduce the dependence of the running time on the size of the right-hand side b and the bounds l, u; and enable Frank–Tardos reduction of the linear objective c [13].

A wave of powerful new results by several groups of authors appeared in 2017–2018, providing a striking array of new applications (see [6] for a list) and contributions to the general theory. A crucial aspect is the context of multivariate algorithmics (fixed parameter tractability). For lack of space, we can only cite a few key publications and refer the interested reader to papers cited within. Chen, Xu, and Shi [2] give an improved analysis for 4-block IPs and a special case, 3-block IPs ($C = O$). Chen and Marx [1] develop Graver techniques for *tree-fold IPs*, a generalization of N-fold IPs whose block structure is the transpose of multi-stage stochastic integer programs. Eisenbrand et al. [6] find a tighter estimation analysis based on the Steinitz lemma that improves upon (4), which has strong implications for various block-structured problems. Koutecký et al. [13] give a parameterized strongly polynomial algorithm for various block-structured problems and give an interpretation of these results in terms of graph parameters such as treewidth and treedepth of the constraint matrix.

References

[1] L. Chen and D. Marx, *Covering a tree with rooted subtrees–parameterized and approximation algorithms*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, SIAM, 2018, pp. 2801–2820.

[2] L. Chen, L. Xu, and W. Shi, *On the Graver basis of block-structured integer programming*, 2018, `arXiv:1805.03741`.

[3] J. A. De Loera, R. Hemmecke, and M. Köppe, *Algebraic and geometric ideas in the theory of discrete optimization*, MOS–SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013, `https://doi.org/10.1137/1.9781611972443`, ISBN 978-1-61197-243-6.

[4] J. A. De Loera, R. Hemmecke, S. Onn, and R. Weismantel, *N-fold integer programming*, Discrete Optimization **5** (2008), no. 2, 231–241, In Memory of George B. Dantzig, `https://doi.org/10.1016/j.disopt.2006.06.006`.

[5] P. Diaconis, R. L. Graham, and B. Sturmfels, *Primitive partition identities*, Combinatorics, Paul Erdős is Eighty, Volume 2 (D. Miklós, V. T. Sós, and D. Szőnyi, eds.), Bolyai Society Mathematical Studies, vol. 2, 1996, pp. 173–192.

[6] F. Eisenbrand, C. Hunkenschröder, and K.-M. Klein, *Faster algorithms for integer programs with block structure*, 2018, `arXiv:1802.06289`.

[7] J. E. Graver, *On the foundations of linear and integer linear programming I*, Mathematical Programming **8** (1975), 207–226.

[8] R. Hemmecke, R. Hemmecke, M. Köppe, P. Malkin, and M. Walter, *4ti2 – a software package for algebraic, geometric and combinatorial problems on linear spaces*, available from `https://4ti2.github.io`.

[9] R. Hemmecke, M. Köppe, and R. Weismantel, *Graver basis and proximity techniques for block-structured separable convex integer minimization problems*, Mathematical Programming **145** (2014), no. 1–2, Ser. A, 1–18.

[10] R. Hemmecke, S. Onn, and L. Romanchuk, *N-fold integer programming in cubic time*, Mathematical Programming (2011), 1–17, `https://doi.org/10.1007/s10107-011-0490-y`.

[11] R. Hemmecke, S. Onn, and R. Weismantel, *A polynomial oracle-time algorithm for convex integer minimization*, Mathematical Programming **126** (2011), 97–117, `https://doi.org/10.1007/s10107-009-0276-7`.

[12] R. Hemmecke and R. Schultz, *Decomposition of test sets in stochastic integer programming*, Mathematical Programming **94** (2003), no. 2–3, 323–341.
[13] M. Koutecký, A. Levin, and S. Onn, *A parameterized strongly polynomial algorithm for block structured integer programs*, 2018, to appear in Proc. ICALP 2018, `arXiv:1802.05859`.

Algorithms for Rank-1 Bimatrix Games

Bernhard von Stengel
(joint work with Bharat Adsul, Jugal Garg, Ruta Mehta, and Milind Sohoni)

The rank of a bimatrix game is the matrix rank of the sum of the two payoff matrices. For a game of rank k, the set of its Nash equilibria is the intersection of a one-dimensional set of equilibria of parameterized games of rank $k-1$ with a hyperplane. For a game of rank 1, this is a monotonic path of solutions to a parameterized zero-sum game and hence LP. One intersection of this path with the hyperplane can be found in polynomial time by binary search. This gives a polynomial-time algorithm to find an equilibrium of a game of rank 1, as already shown by Adsul, Garg, Mehta, and Sohoni [1]. On the other hand, there are rank-1 games with exponentially many equilibria [5], which answers an open problem by Kannan and Theobald [3]. Hence, the exponential multiplicity of equilibria does not prevent one equilibrium in polynomial time. In this spirit, we conjecture that the *uniqueness* of an equilibrium of a rank-1 game is co-NP-hard. For games of higher rank, the parameterized equilibria of lower rank correspond to the computation of the global Newton method by Govindan and Wilson [2], which for two-player games is a special case of Lemke's algorithm [4].

References

[1] B. Adsul, J. Garg, R. Mehta, and M. Sohoni, *Rank-1 bimatrix games: a homeomorphism and a polynomial time algorithm*, Proc. 43rd Annual ACM Symposium on Theory of Computing (STOC) (2011), 195–204.
[2] S. Govindan and R. Wilson, *A global Newton method to compute Nash equilibria*, Journal of Economic Theory **110** (2003), 65–86.
[3] R. Kannan and T. Theobald, *Games of fixed rank: A hierarchy of bimatrix games*, Economic Theory **42** (2010), 157–173.
[4] C. E. Lemke, *Bimatrix equilibrium points and mathematical programming*, Management Science **11** (1965), 681–689.
[5] B. von Stengel, *Rank-1 games with exponentially many Nash equilibria*, arXiv preprint 1211.2405 (2012).

Risk-Averse Control of Partially Observable Systems

Andrzej Ruszczynski

We consider risk measurement in controlled partially observable dynamical systems in discrete time. In such systems, part of the state vector is not observed, but affects the transition kernel and the costs. Because of the dependence of the probability measure on the control and observation, we have to consider risk models involving the probability measure as one of the arguments. First, we discuss two-stage problems, derive the risk disintegration formula and hierarchical formulations of the problems. Then we pass to discrete-time Markov models and we introduce new concepts of risk filters and study their properties. We also introduce the concept of conditional stochastic time consistency. We derive the structure of risk filters enjoying this property and prove that they can be represented by a collection of law invariant risk measures on the space of function of the observable part of the state. We also derive the corresponding dynamic programming equations. Then we illustrate the results on a clinical trial problem.

Chance-Constrained Programming with Decision-Dependent Uncertainty

Miguel A. Lejeune
(joint work with François Margot, Alan D. de Oliveira)

We study a class of joint chance-constrained stochastic problems with decision-dependent uncertainty **CC-DD**. A coupling function models the dependency between decision and endogenous random variables. We propose deterministic reformulations equivalent to the general chance-constrained problem with decision-dependent uncertainty **CC-DD** and applicable to any coupling function. We expose the advantages of the proposed Boolean-based reformulations over typical scenario-based formulations. We define the properties of coupling functions and explain, illustrate their versatility, and the importance of properly modeling decision-dependent uncertainty. We then provide the explicit formulation of problem **CC-DD** in the decision-dependent service uncertainty context and derive exact mixed-integer nonlinear reformulations. We design an algorithmic framework which includes the derivation of convex MINLP relaxation problems, the use of new multiterm convexification methods for bilinear terms, the derivation of tight lower and upper bounds, and the design of a nonlinear branch-and-bound algorithm featuring a conification-relaxation method and the new smallest domain branching rule. Experiments based on real-life data show the scalability and computational efficiency of the nonlinear branch-and-bound framework. To conduct the tests, we also develop a new medical evacuation chance-constrained model with exogenous and decision-dependent uncertainty which endogenizes the calculation of individual busy probabilities.

References

[1] M.A. Lejeune, *Pattern-Based Modeling and Solution of Probabilistically Constrained Optimization Problems*, Operations Research **60**-6 (2012), 1356–1372.

[2] M.A. Lejeune, F. Margot, *Solving Chance Constrained Problems with Random Technology Matrix and Stochastic Quadratic Inequalities*, Operations Research **64**-4 (2016), 939–957.

[3] M.A. Lejeune, F. Margot, *Aeromedical Battlefield Evacuation Under Endogenous Uncertainty in Casualty Delivery Times. Management Science* (2018). Forthcoming: `https://pubsonline.informs.org/doi/10.1287/mnsc.2017.2894`.

[4] M.A. Lejeune, F. Margot, A.D. de Oliveira, *Chance-Constrained Programming with Decision-Dependent Uncertainty.* Working Paper (2018): `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3201121`.

Bounds in Multistage Stochastic Optimization Programs

Francesca Maggioni

(joint work with Georg Ch. Pflug)

Multistage stochastic optimization problems are typically formulated in terms of continuous distributions for the underlying stochastic processes describing the problem uncertainty. As such, they are usually intractable as they are originally defined and approximations in terms of finite scenario trees are needed for a numerical solution.

In this talk bounding methods for multistage stochastic optimization problems are discussed. First we consider bounds of multistage convex problems with concave risk functionals as objective and we assume that a sufficiently large discretized scenario tree describing the problem uncertainty is given but is unsolvable. We construct new refinement chains of lower bounds [2], where each bound can be computed by solving sets of group subproblems less complex than the original one, and recalculating the probabilities of each scenario in the group accordingly. Refinement chains are constructed principally in two ways: by keeping one or several scenarios fixed in all subproblems or choosing them disjoint. A monotonically nondecrasing behavior in the cardinality of scenarios of each subproblem is proved. Complexity considerations for the computation of the proposed bounds as function of tree depth and branching factor are also discussed.

Secondly, we demonstrate how one can find guaranteed bounds for the optimal values of the original *infinite* problem with continuous distribution of the problem uncertainty [4]. We construct new discrete trees directly from the simulated data of the whole scenario process and not from the discretization of the conditional distributions as done before in the literature, e.g. in [1], by means of the concept of stochastic dominance of probability measures (first order stochastic dominance and convex order dominance). The proposed method can be made arbitrarily close to optimal solution of the original infinite problem by making the approximating trees bushier. By constructing upper and lower trees, we take care of the fact that that scenario tree approximation keeps the nonanticipativity requirements of the original problem. Our bounds are the optimal values of the problem at hand on finite scenario trees and therefore are easy to get without changing the

optimization model. Moreover, we relate the solution of the approximating tree to the solution of the original continuous problem and vice-versa; that is by solving the upper and lower trees approximations, we get automatically ϵ-solutions, i.e. near optimal solutions for the original problem. Their use is shown in a multistage risk-averse production problem. Results show that the solutions based on convex order dominance construction outperform the ones obtained by first order stochastic dominance closing the gap between upper and lower bounds within a limited computational complexity and simple scenario tree structures.

Current work on optimal scenario grouping in order to obtain the best monotonic chain of lower bounds will be also discussed [3].

References

[1] K. Frauendorfer *Barycentric scenario trees in convex multistage stochastic programming*, Mathematical Programming, **75**(2) (1996), 277–293.

[2] F. Maggioni F. and G. Ch. Pflug *Bounds and approximations for multistage stochastic programs*, Siam J. Optim., **26**(1) (2016) 831–855.

[3] F. Maggioni, R. Cavagnini and S. Ahmed *Optimization Driven Monotonic Bounds in Stochastic Programming*, (in preparation).

[4] F. Maggioni and G. Ch. Pflug *Guaranteed Bounds for General Non-discrete Multistage Risk-Averse Stochastic Optimization Programs*, Optimization Online (2017), (under evaluation in Siam J. Optim.).

Open Problems Relevant to New Developments in Markov Decision Processes

Eugene A. Feinberg

Recently, there are several developments in the theory of Markov decision processes (MDPs) and in real analysis relevant to MDPs. Such progresses in MDPs include: (i) optimality conditions for problems with weakly continuous probabilities, noncompact action sets, and unbounded costs; (ii) optimality conditions for partially observable MDPs (POMDPs); (iii) applications to inventory control; and (iv) extensions to games, robust optimization, and risk. Relevant developments in real analysis focus on Berge's maximum theorem for noncompact decision sets and Fatou's lemma for varying probabilities.

These developments give rise to several open problems and questions:

1. For average-cost MDPs, it is not clear whether two sufficient conditions for average-cost optimality inequalities (sufficient conditions for the existence of relative values), Assumptions **B** and **$\underline{\mathbf{B}}$** are equivalent.
2. For POMDPs, what are the natural assumptions on the POMDP that guarantees Assumption **B** (or **$\underline{\mathbf{B}}$**) for the Completely Observable MDP (COMDP), to which POMDP is reduced.
3. The examples of data-driven applications of Fatou's lemmas for converging probabilities and Berge's theorems are not known, except their applications to POMDPs described in [6].

Let us begin with the generation of Berge's maximum theorem for noncompact decision sets. Let $\mathbb{X}$ and $\mathbb{A}$ be metric spaces, $u : \mathbb{X} \times \mathbb{A} \to \mathbf{R}$ and $A : \mathbb{X} \to 2^{\mathbb{A}} \setminus \{\emptyset\}$. Consider an optimization problem:

$$v(x) := \inf_{a \in A(x)} u(x, a) \quad \text{for each} \quad x \in \mathbb{X}. \tag{1}$$

Definition 1. *A function $f : \mathbb{X} \times \mathbb{A} \to \mathbf{R}$ is called $\mathbb{K}$-inf-compact, if for every compact set $K \subseteq \mathbb{X}$, the function $f : K \times \mathbb{A} \to \mathbf{R}$ is inf-compact.*

A $\mathbb{K}$-inf-compact function $f : \mathbb{X} \times \mathbb{A} \to \mathbf{R}$ is lower semicontinuous and inf-compact in a on $A(x)$ at each $x \in \mathbb{X}$. In fact, these assumptions are almost equivalent. Essentially, $\mathbb{K}$-inf-compactness of $f(x, a)$ is slightly stronger than: (i) lower semicontinuity in (x, a); and (ii) inf-compactness in a. However, (i, ii) do not imply lower semicontinuity of the value function. $\mathbb{K}$-inf-compactness implies semicontinuity of the value function. In addition, an inf-compact function is $\mathbb{K}$-inf-compact. Some examples are as follows. The function $f(x, a) = x^2 + a^2$ is inf-compact. The functions $f(x, a) = |x - a|$ and $f(x, a) = (x - a)^2$ are $\mathbb{K}$-inf-compact, but they are not inf-compact since $f(x, a) = 0$ for $x = a$. If the function h is inf-compact, then

$$c(x, a) = K1_{\{a>0\}} + \bar{c}a + \mathbb{E}h(x + a - D), \qquad a \geq 0,$$

is inf-compact for $\bar{c} > 0$ and $\mathbb{K}$-inf-compact for $\bar{c} = 0$. This cost function is used in inventory control with setup costs.

The following two theorems extend Berge's (maximum) theorem to possibly noncompact action sets.

Theorem 2 (Berge's theorem for possibly noncompact action sets [3])**.** *If $u : \mathbb{X} \times \mathbb{A} \to \mathbf{R}$ is a $\mathbb{K}$-inf-compact function, then the function v defined in* (1) *is lower semicontinuous and the sets $A^*(x) = \mathrm{Argmin}_{a \in A(x)} u(x, a)$ are compact.*

Theorem 3 (Berge's maximum theorem for possibly noncompact action sets [1])**.** *If*

(i) *$A : \mathbb{X} \to 2^{\mathbb{A}} \setminus \{\emptyset\}$ is a lower semicontinuous set-valued mapping,*
(ii) *$u : \mathbb{X} \times \mathbb{A} \to \mathbf{R}$ is $\mathbb{K}$-inf-compact and upper semicontinuous function,*

then the function v defined in (1) *is continuous and the set-valued mapping $A^*(x) = \mathrm{Argmin}_{a \in A(x)} u(x, a)$ is upper semicontinuous and compact-valued.*

Another development in real analysis that is relevant to MDPs is Fatou's lemma for varying probabilities. Let $\{p, p_n\}_{n=1,2,\ldots}$ be a sequence of probability measures.

Theorem 4 (Fatou's Lemma for weakly converging probabilities [4])**.** *Let $\mathbb{X}$ be a metric space, $\{p_n\}_{n=1,2,\ldots}$ converge weakly to p, and $\{f_n\}_{n=1,2,\ldots}$ be a sequence of measurable $\overline{\mathbf{R}}$-valued functions. Then the inequality*

$$\int_{\mathbb{X}} \liminf_{n \to \infty,\, y \to x} f_n(y) p(dx) \leq \liminf_{n \to \infty} \int_{\mathbb{X}} f_n(s) p_n(ds)$$

holds, if there exists a sequence of measurable real-valued functions $\{g_n\}_{n=1,2,\ldots}$ *such that* $f_n(x) \geq g_n(x)$ *for all* $n = 1, 2, \ldots$ *and* $x \in \mathbb{X}$, *and*

$$-\infty < \int_{\mathbb{X}} \limsup_{n\to\infty, y\to x} g_n(y)p(dx) \leq \liminf_{n\to\infty} \int_{\mathbb{X}} g_n(x)p_n(dx).$$

If p_n converges in total variation to p, Feinberg et al. [5] provided a more general form of Fatou's lemma and Lebesgue's convergence theorem.

Theorem 5 (Uniform Fatou's lemma [5])**.** *Let* $(\mathbb{X}, \mathcal{F})$ *be a measurable space, a sequence of finite measures* $\{p_n\}_{n=1,2,\ldots}$ *converge in total variation to a measure* p *on* $\mathbb{X}$, $f \in L^1(\mathbb{X}; p)$, *and* $f_n \in L^1(\mathbb{X}; p_n)$ *for each* $n = 1, 2, \ldots$. *Then the inequality*

$$\liminf_{n\to\infty} \inf_{X\in\mathcal{F}} \Big(\int_X f_n(x)p_n(dx) - \int_X f(x)p(dx) \Big) \geq 0$$

holds if and only if the following two statements hold:
(i) *for each* $\epsilon > 0$

$$p(\{x \in \mathbb{X} : f_n(x) \leq f(x) - \epsilon\}) \to 0 \text{ as } n \to \infty;$$

(ii) $\lim_{K\to+\infty} \inf_{n=1,2,\ldots} \int_{\mathbb{X}} f_n(x)\mathbf{I}\{x \in \mathbb{X} : f_n(x) \leq -K\}p_n(dx) \geq 0.$

Theorem 6 (Uniform Lebesgue's convergence theorem [5])**.** *Let* $(\mathbb{X}, \mathcal{F})$ *be a measurable space, a sequence of finite measures* $\{p_n\}_{n=1,2,\ldots}$ *converge in total variation to a measure* p *on* $\mathbb{X}$, $f \in L^1(\mathbb{X}; p)$, *and* $f_n \in L^1(\mathbb{X}; p_n)$ *for each* $n = 1, 2, \ldots$. *Then the equality*

$$\lim_{n\to\infty} \sup_{X\in\mathcal{F}} \Big| \int_X f_n(x)p_n(dx) - \int_X f(x)p(dx) \Big| = 0$$

holds if and only if the following two statements hold:
(i) *the sequence of functions* $\{f_n\}_{n=1,2,\ldots}$ *converges in measure* p *to* f; *and*
(ii) $\lim_{K\to+\infty} \limsup_{n\to\infty} \int |f_n(x)|\mathbf{I}\{x \in \mathbb{X} : |f_n(x)| \geq K\}p_n(dx) = 0.$

Let us consider an MDP defined by $\{\mathbb{X}, \mathbb{A}, q, c\}$, where (i) $\mathbb{X}$ is a state space; (ii) $\mathbb{A}$ is an action space; (iii) $q(dy|x, a)$ is the transition probability; and (iv) $c(x, a)$ is the one-step cost function. The objective is to minimize

(i) finite-horizon costs: $v^\pi_{N,\alpha}(x) := \mathbb{E}^\pi_x \left[\sum_{t=0}^{N-1} \alpha^t c(x_t, a_t)\right]$, where $\alpha \in [0, 1)$ is the discount factor and $N = 1, 2, \ldots$ is the horizon length;
(ii) infinite-horizon costs: $v^\pi_\alpha(x) := \mathbb{E}^\pi_x \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right]$; or
(iii) average costs per unit time: $w^\pi(x) := \limsup_{N\to\infty} \frac{1}{N}\mathbb{E}^\pi_x \left[\sum_{t=0}^{N-1} c(x_t, a_t)\right]$.

Each of the following two conditions, being added to the assumptions that the transition probabilities q is weakly continuous and the cost function c is $\mathbb{K}$-inf-compact, implies the existence of stationary optimal policies; see [2] for details.

Assumption B ([7]). (i) $w^* := \inf_x \inf_\pi w^\pi(x) < +\infty$; (ii) $\sup_{\alpha\in[0,1)} u_\alpha(x) < \infty$, $x \in \mathbb{X}$.

Assumption <u>B</u> ([2]). (i) Assumption B(i) holds; (ii) $\liminf_{\alpha\uparrow 1} u_\alpha(x) < \infty$, $x \in \mathbb{X}$.

Assumption B is more general and leads to the different optimality inequality than the optimality inequality that holds under Assumption **B**.

Let us consider a POMDP defined by $(\mathbb{X}, \mathbb{Y}, \mathbb{A}, P, Q, c)$, where $\mathbb{Y}$ is the observation set and $Q(dy|a,x)$ is the observation kernel. Let $P_0(dx'|x,a)$ be the initial state distribution and $Q_0(dy|a,x)$ be the distribution of the initial observation.

The following theorem states the existence of optimal policies for POMDPs.

Theorem 7 ([6])**.** *Suppose the following conditions hold for the POMDP:*

(i) *the cost function c is $\mathbb{K}$-inf-compact;*
(ii) *the transition probability $P(\cdot|x,a)$ is weakly continuous in (x,a);*
(iii) *the observation kernel $Q(\cdot|a,x)$ is continuous in total variation in (a,x).*

Then optimal policies exist and convergence of value iteration takes place for discounted cost problems.

References

[1] E. A. Feinberg, P. O. Kasyanov, M. Voorneveld. Berge's maximum theorem for noncompact image sets. *J. Math. Anal. Appl.* 413(2), 1040–1046, 2014.

[2] E. A. Feinberg, P. O. Kasyanov, N. V. Zadoianchuk. Average cost Markov decision processes with weakly continuous transition probabilities. *Math. Oper. Res.* 37(4), 591–607, 2012.

[3] E. A. Feinberg, P. O. Kasyanov, N. V. Zadoianchuk. Berge's theorem for noncompact image sets. *J. Math. Anal. Appl.* 37(1), 255–259, 2013.

[4] E. A. Feinberg, P. O. Kasyanov, N. V. Zadoianchuk. Fatou's lemma for weakly converging probabilities. *Theory Probab. Appl.* 58(4), 683–689, 2014.

[5] E. A. Feinberg, P. O. Kasyanov, M. Z. Zgurovsky. Uniform Fatou's lemma. *J. Math. Anal. Appl.* 444(1), 550–567, 2016.

[6] E. A. Feinberg, P. O. Kasyanov, M. Z. Zgurovsky. Partially observable total-cost Markov decision processes with weakly continuous transition probabilities. *Math. Oper. Res.* 41(2), 656–681, 2016.

[7] M. Schäl. Average optimality in dynamic programming with general state space. *Math. Oper. Res.* 18(1):163–172, 1993.

Risk averse stochastic shape optimization

Martin Rumpf

(joint work with Sergio Conti, Benedikt Geihe, Harald Held, Rüdiger Schultz, Sascha Tölkes)

In this extended abstract shape optimization for elastic materials [2, 1] is considered under stochastic loading. Risk averse objective functionals are investigated and the concept of stochastic dominance constraints is explored. Thereby the expected excess and the excess probability are taking into account, first as objective functional involving the complience cost of a elastic object under stochastic loading and then as a constraint when comparing a shape with a benchmark shape.

Shape optimization with stochastic loading. Shape optimization based on an explicit parametric description of the mechanical object is algorithmically demanding. Hence, we consider an approximating phase-field representation of the elastic object $\mathcal{O} \subset \Omega$, where $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ denotes the computational domain. To this end, we take into account a phase-field function $v : \Omega \to \mathbb{R}$ of Allen–Cahn or Modica–Mortola type with a phase field energy functional $\mathcal{L}^\varepsilon[v] := \frac{1}{2}\int_\Omega \varepsilon|\nabla v|^2 + \frac{1}{\varepsilon}\Psi(v)\,\mathrm{d}\,x$, where the scale parameter ε describes the width of the interfacial region. In our context of shape modeling, we set $\Psi(v) := \frac{9}{16}(v^2-1)^2$, which has two minima at $v = -1$ and $v = 1$ representing the two phases outside and inside of $\mathcal{O}$, respectively. In the limit $\varepsilon \to 0$, the phase field v is forced towards the pure phases -1 and 1 and $\mathcal{L}^\varepsilon$ Γ-converges to the total interface area. Furthermore, we approximate the total volume by the smooth functional $\mathcal{V}[v] := \int_\Omega \chi(v)\,\mathrm{d}\,x$, where $\chi(v) := \frac{1}{4}(v+1)^2$ is an approximation to the characteristic function $\chi_\mathcal{O}$. Next, let us recall the model of linearized elasticity taylored to the phase field description of the elastic object $\mathcal{O}$. Instead of considering a void phase on $\Omega \setminus \mathcal{O}$, we follow common practice in shape optimization assuming the presence of a very soft material on the part of the computational domain outside of $\mathcal{O}$. The elastic energy stored inside the material under a given displacement $u : \Omega \to \mathbb{R}^d$ is then defined as

$$\mathbb{W}^\delta[v,u] := \frac{1}{2}\int_\Omega ((1-\delta)\chi(v)+\delta)C\epsilon[u] : \epsilon[u]\mathrm{d}\,x\,,$$

where $\epsilon[u] = \frac{1}{2}(Du^T + Du)$ is the strain tensor. Here, C is the elasticity tensor of the material, in our case that of the Lamé-Navier elasticity model. Furthermore, we take into account a Dirichlet boundary $\Gamma_D \subset \partial\Omega$ and an inhomogeneous Neumann boundary $\Gamma_N \subset \partial\Omega$ with $\Gamma_D \cap \Gamma_N = \emptyset$. The equilibrium displacement $u \in H^1(\Omega;\mathbb{R}^d)$ for fixed phase field v is then given as the minimizer of the free energy $\mathbb{E}[v,u] := \mathbb{W}^\delta[v,u] - \mathbb{C}[u]$ within a set of admissible displacements u with trace $u|_{\Gamma_D} = 0$, where the functional $\mathbb{C}[u] := \int_{\Gamma_N} g\cdot u\,\mathrm{d}\,a$ is the (negative) potential of the surface load for $g \in L^2(\Gamma_N;\mathbb{R}^d)$.

Finally, we define the cost functional

$$\mathcal{J}[v,u] := 2\mathbb{W}^\delta[v,u] + \nu\mathcal{V}[v] + \gamma\mathcal{L}^\varepsilon[v]\,,$$

where $\mathcal{V}$ approximately measures the volume of the elastic object $\mathcal{O}$ and $\mathcal{L}^\varepsilon$ the object perimeter within the computational domain Ω and ν, $\gamma > 0$. The shape optimization problem is now to minimize $\mathcal{J}[v,u[v]]$ subject to the constraint that $u[v]$ is a minimizer of $u \mapsto \mathbb{E}[v,u]$.

Stochastic loading and risk averse functionals. If the surface load g is uncertain, i .e. described by a random variable $g[\omega]$ on some probability space $(\Omega,\mathcal{A},\mathbb{P})$, then so are the displacement $u[\omega]$ and the cost functional $\mathcal{J}[v,\omega]$. Optimizing or selecting shapes therefore amounts to optimizing the random variable representing the cost, $\left\{\mathcal{J}[v,u[v][\cdot]] \;:\; v \in H^{1,2}_{\Gamma_N}(\Omega)\right\}$, where the Sobolev space $H^{1,2}_{\Gamma_N}(\Omega)$ is the set of admissible phase fields which are measurable functions obeying $v = 1$ on Γ_N in the sense of traces.

In mean-risk models [3] objective functions are formulated as weighted sums of the expected value and some risk measure, which is chosen as a quantity that formalizes the user's perception of risk. For the expected value one obtains $\mathcal{Q}_{ev}(v) := \int_\Omega \mathcal{J}[v, u[v][\omega]]\, \mathbb{P}(d\omega)$. General risk measures are nonlinear quantities [4, 6]. Typical examples of risk measures are the excess probability $\mathcal{Q}_{ep}$, which is the probability of exceeding a preselected target value $\eta \in \mathbb{R}$: $\mathcal{Q}_{ep}(v) := \mathbb{P}\left[\mathcal{J}[v, u[v][\omega]] > \eta\right]$, and the expected excess $\mathcal{Q}_{ee}$ given by the mean-value of the outcomes above a preselected $\eta \in \mathbb{R}$: $\mathcal{Q}_{ee}(v) := \int_\Omega \max\{\mathcal{J}[v, u[v][\omega]] - \eta, 0\}\, \mathbb{P}(d\omega)$ In [4, 6] risk neutral and risk averse stochastic shape optimization were discussed for mean-risk models with the above specifications. To this end the stochastic energies were used as stochastic cost functions for shape optimization directly.

Stochastic dominance constraints. Now, we use a notion of "being stochastically smaller " to state dominance of first order and "being smaller with respect to the increasing convex order" in the presence of second order dominance [5]. The definitions of dominance of first order ($\preceq_{\mathrm{st}}$) and second order ($\preceq_{\mathrm{icx}}$), respectively, are given as follows:

$$\mathsf{X} \preceq_{\mathrm{st}} \mathsf{Y} \quad \text{iff } \mathbb{P}\left[\{\omega \in \Omega : \mathsf{X}[\omega] \le \eta\}\right] \ge \mathbb{P}[\{\omega \in \Omega : \mathsf{Y}[\omega] \le \eta\}]\ \forall \eta \in \mathbb{R}\,,$$

$$\mathsf{X} \preceq_{\mathrm{icx}} \mathsf{Y} \quad \text{iff } \int_\Omega \max\{\mathsf{X}[\omega] - \eta, 0\}\, \mathbb{P}(d\omega) \le \int_\Omega \max\{\mathsf{Y}[\omega] - \eta, 0\}\, \mathbb{P}(d\omega)\ \forall \eta \in \mathbb{R}\,.$$

In our shape optimization context, first order dominance can be understood as turning the excess probability approach based on the cost function $\mathcal{Q}_{ep}$ into a constraint. Analogously, second order dominance is conceptually related to the expected excess approach with the cost function $\mathcal{Q}_{ee}$.

Having fixed a benchmark phase field v_b (which describes an elastic object $\mathcal{O}_b$) and a random benchmark surface load $g[\omega_b]$ one can formulate the abstract optimization problem

$$\min\{\mathcal{G}(v) \,:\, \mathcal{J}[v, u[v][\omega]] \preceq \mathcal{J}[v_b, u[v_b][\omega_b]],\ v \in H^{1,2}(\Omega),\ v = 1 \text{ on } \Gamma_N\}.$$

Here $\preceq$ is specified as either $\preceq_{\mathrm{st}}$ or $\preceq_{\mathrm{icx}}$, and for the sake of notational simplicity we identify v_b with $\mathcal{O}_b$ and neglect the difference between soft material ($\delta > 0$) and void ($\delta = 0$).

The entity $\mathcal{J}[v_b, u[v_b][\omega_b]]$ corresponds to the cost-representing random variable that arises when exposing the benchmark)shape $\mathcal{O}_b$ to a random surface load $g[\omega_b]$. In this model $\mathcal{G}$ stands for an objective function which, in the present context mainly consists of the volume or the elastic energy of $\mathcal{O}$ complemented with a small perimeter functional. I.e. we choose

$$\mathcal{G}(v) := \mathcal{V}(v) + \beta \mathcal{L}^\varepsilon(v)$$

for a $\beta > 0$. Furthermore, in the application we assume a finite set of realizations of the stochastic loading, each with different boundary data, in which every scenario is represented by a random variable ω_k and comes with a certain probability π_k and a force g_k for $k = 1, \ldots, K$, with $\pi_k \in [0, 1]$ and $\sum_{k=1}^K \pi_k = 1$. Then, it suffices

to consider a finite set of constraints. For a discussion of the numerical findings we refer to [5].

Acknowledgements. M. Rumpf acknowledges support of the Hausdorff Center for Mathematics and the Collaborative Research Center 1060 funded by the German Research Foundation.

References

[1] G. Allaire. *S*hape optimization by the homogenization method, volume 146 of *A*pplied Mathematical Sciences. Springer-Verlag, New York, 2002.
[2] M. P. Bendsøe. *O*ptimization of structural topology, shape, and material. Springer-Verlag, Berlin, 1995.
[3] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz. Shape optimization under uncertainty - a stochastic programming perspective. *S*IAM Journal on Optimization, 19(4):1610–1632, 2009.
[4] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz. Risk averse shape optimization. *S*IAM Journal on Control and Optimization, 49(3):927–947, 2011.
[5] S. Conti, M. Rumpf, R. Schultz, and S. Tölkes. Stochastic dominance constraints in elastic shape optimization. *S*IAM Journal on Control and Optimization, 2018. accepted.
[6] B. Geihe, M. Lenz, M. Rumpf, and R. Schultz. Risk averse elastic shape optimization with parametrized fine scale geometry. *M*athematical Programming, 141(1-2):383–403, 2013.

Uncertainty Quantification for Bayesian Inverse Problems

Claudia Schillings

(joint work with Dirk Blömker, Andrew Stuart, Philipp Wacker, Simon Weissmann)

Uncertainty Quantification is an interesting, fast growing research area aimed at developing methods to address the impact of parameter, data and model uncertainty in complex systems. Here, we will focus on the identification of parameters through observations of the response of the system - the inverse problem. The uncertainty in the solution of the inverse problem will be described via the Bayesian approach.

The forward response operator $\mathcal{G} \in \mathcal{C}(\mathcal{X}, \mathbb{R}^K)$ maps the unknown parameters $u \in \mathcal{X}$ in a separable Hilbert space $\mathcal{X}$ to the data space $\mathbb{R}^K$. The number of observation is assumed to be finite and denoted by $K \in \mathbb{N}$. The inverse problem then consists of recovering the unknown parameters $u \in \mathcal{X}$ from noisy data y

$$y = \mathcal{G}(u) + \eta\,,$$

where the observation noise η is assumed to be Gaussian, i.e. $\eta \sim \mathcal{N}(0, \Gamma)$ for a given covariance matrix $\Gamma \in \mathbb{R}^{K\times K}$. The unknown parameter u is modeled as a $\mathcal{X}$-valued random variable with prior distribution μ_0. Assuming that $\eta \sim \mathcal{N}(0, \Gamma)$ is independent of $u \sim \mu_0$, the solution of the inverse problem is the $\mathcal{X}$-valued random variable $u^y \sim \mu^y$ with

$$\mu^y(du) = \frac{1}{Z} \exp(-\Phi(u; y))\mu_0(du)$$

by Bayes' theorem [8]. The constant Z is the normalization constant

$$Z := \int_{\mathcal{X}} \exp(-\Phi(u;y))\mu_0(du)$$

and the functional $\Phi : \mathcal{X} \to \mathbb{R}$ is the least-squares functional

$$\Phi(u;y) = \frac{1}{2}\|(y - \mathcal{G}(u))\|_\Gamma^2 .$$

with $\|\cdot\|_\Gamma := \|\Gamma^{-\frac{1}{2}}\cdot\|_2$. In cases, where the model evaluations are prohibitively expensive, ad hoc methods such as the ensemble Kalman filter for inverse problems [3, 4] are widely and successfully used by practitioners in order to approximate the solution of the Bayesian problem. Since the method is usually used with a rather small ensemble size (in the range of 20 - 100 ensemble members), the goal of our work is to build analysis for a fixed ensemble size. To do so, we discuss the ensemble Kalman inversion based on the continuous time scaling limits of the method as a derivative free optimization method for the least-squares misfit functional. The continuous time limit of the ensemble Kalman inversion is formally given by

$$\mathrm{d}u_t^{(j)} = C^{up}(u_t)\Gamma^{-1}(y - \mathcal{G}(u_t^{(j)}))\,\mathrm{d}t + C^{up}(u_t)\Gamma^{-1/2}\,\mathrm{d}W_t^{(j)}\,, \tag{1}$$

where $W^{(j)}$ are independent Brownian motions on $\mathbb{R}^K$ and

$$C^{up}(u) = \frac{1}{J}\sum_{j=1}^{J}(u^{(j)} - \overline{u}) \otimes (\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}})$$

with $\overline{u} = \frac{1}{J}\sum_{j=1}^{J} u^{(j)}$, $\overline{\mathcal{G}} = \frac{1}{J}\sum_{j=1}^{J}\mathcal{G}(u^{(j)})$. The strong convergence of the discrete ensemble Kalman inversion iteration to continuous paths of (1) in the general nonlinear setting is an interesting and still open question. Preliminary results for a simplified model can be found in [1]. Under the assumption that the forward operator is linear, i.e. $\mathcal{G}(\cdot) = A\cdot$ with $A \in \mathcal{L}(\mathcal{X}, \mathbb{R}^K)$, the continuous time limit (1) is given by

$$\mathrm{d}u_t^{(j)} = \frac{1}{J}\sum_{k=1}^{J}\langle A(u_t^{(k)} - \overline{u}_t), (y - Au_t^{(j)})\,\mathrm{d}t + \sqrt{\Gamma}\,\mathrm{d}W_t^{(j)}\rangle_\Gamma(u_t^{(k)} - \overline{u}_t)\,. \tag{2}$$

The structure of the continuous time limit (2) can be exploited to establish existence and uniqueness results, i.e. well-posedness of the ensemble Kalman inversion. In particular, in [2], it is shown that the subspace property holds true in the continuous time limit, which allows to use the theory of stochastic differential equations in finite dimensional spaces. Furthermore, the ensemble collapse in the observation as well as parameter space is characterized in terms of moments and almost sure convergence with given rates. In case of exact data, variance inflation can be shown to lead to convergence to the true parameters in terms of moments and almost sure convergence. We refer to [6, 7, 2] for more details. The interpretation of the ensemble Kalman inversion as a derivative free optimization method opens up the perspective to use the method in various fields of application, e.g. imaging,

groundwater flow problems, biological problems as well as in the context of the training of neural networks [5]. Even though the analysis in [6, 7, 2] is confined to the linear setting, it can guide the analysis of nonlinear problems. Numerical experiments show a good convergence behavior also for highly nonlinear problems. The generalization of the well-posedness and accuracy results will be subject to future work.

References

[1] D. Blömker, C. Schillings, and P. Wacker, *A strongly convergent numerical scheme from EnKF continuum analysis*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 2537–2562.

[2] D. Blömker, C. Schillings, P. Wacker, and S. Weissmann, *Well-posedness and convergence analysis of the ensemble Kalman inversion*, (in preparation).

[3] G. Evensen, *The Ensemble Kalman filter: theoretical formulation and practical implementation*, Ocean Dynamics, 53 (2003), pp. 343–367.

[4] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29 (2013), p. 045001.

[5] N. Kovachki and A. M. Stuart, *Ensemble Kalman inversion: A derivative-free technique for machine learning tasks*, ArXiv e-prints, (2018).

[6] C. Schillings and A. M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 1264–1290.

[7] C. Schillings and A. M. Stuart, *Convergence analysis of ensemble Kalman inversion: the linear, noisy case*, Applicable Analysis, 97 (2018), pp. 107–123.

[8] A.M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica 19 (2010).

Reduced-Order Models for CVaR Estimation and Risk Averse Optimization

Matthias Heinkenschloss

(joint work with Boris Kramer, Timur Takhtaganov, Karen Willcox)

This talk introduces two approaches to improve the sampling efficiency in risk averse optimization governed by partial differential equations (PDEs) with uncertain parameters. For these problems, the PDE solution and the quantity of interest (QoI), which depends on the PDE solution and would be the objective function in the deterministic case, now depend on the random parameters. Therefore, a so-called risk measure is applied to the QoI to define the objective function risk averse optimization. We consider the Conditional-Value-at-Risk (CVaR) risk measure. Evaluating the CVaR of the QoI requires sampling in the tail of the distribution of the QoI and typically requires many solutions of the expensive PDE. We use reduced-order models (ROM) of the PDE to substantially reduce the computational expense of naive Monte Carlo (MC) sampling.

The parameter region that corresponds to the tail of the distribution of the QoI that enters the evaluation of the risk measure is a small region in the original parameter space and is called the risk region. For CVaR the probability of the risk region is $1 - \beta \ll 1$, where $\beta \in (0, 1)$ is the quantile level at which CVaR is computed. Unfortunately, this risk region depends on the QoI, the PDE and the

random variables that enter the PDE and is not known a-priori known. When using naive MC to estimate the CVaR, about $100\beta\%$ of the samples, each of which requires an expensive PDE solve, fall outside the risk region and therefore do not contribute to the CVaR estimate. We use ROMs in two ways to substantially reduce the number of expensive high fidelity PDE solves (also referred to as full-order model (FOM) evaluations) needed to estimate CVaR. We summarize our approaches next. Details can be found in [1].

The first approach replaces the computationally expensive FOM by an inexpensive ROM. This idea has been used before, but our refined error estimate shows that the resulting CVaR estimation error is proportional to the ROM error in the small risk region. This improves previous error estimates and also opens opportunities for adaptive ROM generation.

The second approach uses a combination of FOM and ROM evaluations via importance sampling, and can be effective even if the ROM has large errors, but the risk region computed with the ROM is a good approximation of the true risk region. In the importance sampling approach, ROM samples are used to estimate the risk region and to construct a biasing distribution. Few FOM samples are then drawn from this biasing distribution. Asymptotically as the ROM error goes to zero, the importance sampling estimator reduces the variance by a factor $1-\beta \ll 1$, where β is again the quantile level at which CVaR is computed.

Numerical experiments of CVaR estimation of a QoI related to heat release modeled by a system of semilinear diffusion-advection-reaction PDEs derived from a simple combustion model demonstrate the substantial computational savings (measured in FOM evaluations) of our approaches. In one comparison the number of FOM evaluations is reduced by a factor of 100. The importance sampling framework led to substantially better CVaR estimates when coarser ROMs are used compared to simply replacing FOM samples by ROM samples, but at the expense of using 10^2 additional FOM samples. Since the ROM construction also requires FOM evaluations, these are accounted for in our computational cost estimates. Overall, our numerical results showed that it appears beneficial to invest FOM samples to train a ROM—which is then used to compute a biasing distribution—than to sample from the FOM directly.

References

[1] M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox, *Conditional-Value-at-Risk Estimation via Reduced-Order Models*, SIAM/ASA Journal on Uncertainty Quantification, in press.

How turnpike theory helps to find good controls

Martin Gugat
(joint work with René Henrion)

The computation of optimal controls that solve a dynamic optimal control problem is often more time-consuming than the solution of the corresponding static problems. Turnpike theorems provide a relation between the solution of the dynamic problems and the solution of the static problem. Recently, some progress in turnpike theory with pdes has been achieved, see [6, 5, 4, 3, 2]. Here we present abstract turnpike theorems that can be applied in the context of boundary control of gas pipeline operations. First we present an integral turnpike result that states that the norm of the difference between the dynamic and the static optimal control remains uniformly bounded for all time horizons. Then we state an exact turnpike result that gives conditions that imply that after a fixed time t_0 the dynamic optimal control is equal to the static optimal control. This exact turnpike result only holds under rather strong assumptions.

In the application to the gas flow through a network of pipes the state in each pipe is governed by a hyperbolic partial differential equation. At the nodes of the network graph the system is governed by algebraic node conditions. The consumer behavior at the boundary nodes of the graph is uncertain. This gives rise to uncertainty in the data in the corresponding Dirichlet boundary conditions. The customers require that the state at the boundary nodes satisfies certain state constraint. In order to take the uncertainty into account, this requirement can be modeled in the form of a probabilistic constraint. Under the assumption that the distribution is log-concave, this probabilistic constraint is equivalent to a deterministic convex inequality constraint. This is the form that we consider in the following analysis. The corresponding static optimal control problems are studied in [1]. For the numerical solution, the spheric-radial decomposition is used.

1. Definitions and notation

Let a natural number n be given. For $T > 0$ define $H(T) = (L^2(0, T))^n$. For all $T > 0$, let a convex continuously differentiable function $J_T : H(T) \to [0, \infty)$ be given. Assume that J_T is strongly convex in the sense that there a constant $\kappa > 0$ that is independent of T such that for all $T > 0$

$$\langle J_T'(u) - J_T'(v),\, u - v\rangle_{H(T)} \geq \kappa \|u - v\|^2_{H(T)}. \tag{1}$$

Let a convex continuously differentiable function $G_T : H(T) \to \mathbb{R}$ be given. Define the convex set $K_T = \{u \in H(T) : G_T(u) \leq 0\}$. Assume that there exists a Slater point $r_s \in \mathbb{R}^n$ such that $G_T(r_s) < 0$. We consider the time-parametric *dynamic optimization problem*

$$P(T) : \min_{u \in K_T} J_T(u).$$

Assume that for all $T_1 > 0$, $T_2 > 0$ we have $K_{T_1} \cap \mathbb{R}^n = K_{T_2} \cap \mathbb{R}^n$. Let a strongly convex continuously differentiable function $f : \mathbb{R}^n \to [0, \infty)$ be given. Assume

that for all $r \in \mathbb{R}^n$ there exist constants $C_0 > 0$, $C_1 > 0$ such that for all $T > 0$, for all $h \in H(T)$ we have

$$\langle f'(r) - J_T'(r),\, h\rangle_{H(T)} \leq C_0 + C_1 \|h\|_{H(T)}. \tag{2}$$

Define the *static optimization problem*

$$P_{(\sigma)} : \min_{r \in \mathbb{R}^n \cap K_1} f(r).$$

2. An integral turnpike result

The solutions of $P(T)$ and $P_{(\sigma)}$ are related by the following turnpike result:

Proposition 1: *For $T > 0$ let $u^{(\delta,T)}$ denote the solution of $P(T)$. Let $u^{(\sigma)}$ denote the solution of $P_{(\sigma)}$. Then there exists a constant $\eta_0 > 0$ that does not depend on T such that for all $T > 0$ the following inequality holds:*

$$\left\| u^{(\delta,T)} - u^{(\sigma)} \right\|_{H(T)} \leq \eta_0 \tag{3}$$

Proof: Due to (1), we have

$$\langle J_T'(u^{(\delta,T)}) - J_T'(u^{(\sigma)}),\, u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \geq \kappa \|u^{(\delta,T)}) - u^{(\sigma)}\|^2_{H(T)}. \tag{4}$$

Due to the Slater condition the necessary optimality conditions imply that for all $T > 0$ there exist multipliers $\mu^{(\delta,T)} \geq 0$, $\mu^{(\sigma)} \geq 0$ such that

$$\begin{aligned} -J_T'(u^{(\delta,T)}) &= \mu^{(\delta,T)}\, G_T'(u^{(\delta,T)}) \\ -f'(u^{(\sigma)}) &= \mu^{(\sigma)} G_T'(u^{(\sigma)}) \end{aligned}$$

and

$$\mu^{(\delta,T)} \langle G_T'(u^{(\delta,T)}), u^{(\sigma)} - u^{(\delta,T)}\rangle_{H(T)} \leq 0, \tag{5}$$

$$\mu^{(\sigma)} \langle G_T'(u^{(\sigma)}), u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \leq 0. \tag{6}$$

Hence due to (4) we have

$$\begin{aligned} \kappa \|u^{(\delta,T)}) - u^{(\sigma)}\|^2_{H(T)} &\leq \langle J_T'(u^{(\delta,T)}) - J_T'(u^{(\sigma)}),\, u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \\ = \langle \Big[J_T'(u^{(\delta,T)}) - f'(u^{(\sigma)}) \Big] &+ \Big[f'(u^{(\sigma)}) - J_T'(u^{(\sigma)}) \Big],\, u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \\ &= \langle f'(u^{(\sigma)}) - J_T'(u^{(\sigma)}),\, u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \\ + \mu^{(\delta,T)} \langle G_T'(u^{(\delta,T)}), u^{(\sigma)} - u^{(\delta,T)}\rangle_{H(T)} &+ \mu^{(\sigma)} \langle G_T'(u^{(\sigma)}), u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \\ &\leq \langle f'(u^{(\sigma)}) - J_T'(u^{(\sigma)}),\, u^{(\delta,T)} - u^{(\sigma)}\rangle_{H(T)} \end{aligned}$$

where the last inequality follows using (5) and (6). Due to (2) this implies

$$\kappa \|u^{(\delta,T)}) - u^{(\sigma)}\|^2_{H(T)} \leq C_0 + C_1 \|u^{(\delta,T)} - u^{(\sigma)}\|_{H(T)}.$$

This implies (3).

Example: Let a linear operator $F_{(\sigma)} : \mathbb{R}^n \to \mathbb{R}^n$ and for all $T > 0$ a linear operator $F_T : H(T) \to H(T)$ be given. Let $d \in \mathbb{R}^n$ denote a given desired

state. For $r \in \mathbb{R}^n$, define $f(r) = \frac{1}{2} r^\top r + \frac{1}{2}\|F_{(\sigma)} r - d\|_{\mathbb{R}^n}$. For $u \in H(T)$, define $J_T(u) = \frac{1}{2}\|u\|^2_{H(T)} + \frac{1}{2}\|F_T u - d\|^2_{H(T)}$.

Assume that there exists $C_2 > 0$ such that for all $T > 0$ we have $\|F_T\| \le C_2$.

Assume that for all $r \in \mathbb{R}^n$ there exist $C_3 > 0$ and $C_4 > 0$ such that for all $T > 0$ we have $\|(F_T - F_{(\sigma)})r\|_{H(T)} \le C_3$ and $\|(F^*_T - F^*_{(\sigma)})(F_T r - d)\|_{H(T)} \le C_4$. Then for all $r \in \mathbb{R}^n$ we have

$$\begin{aligned} \|f'(r) - J'_T(r)\|_{H(T)} &= \|F^*_{(\sigma)}(F_{(\sigma)} r - d) - F^*_T(F_T r - d)\|_{H(T)} \\ &= \|F^*_{(\sigma)}(F_{(\sigma)} - F_T)r\|_{H(T)} + (F^*_{(\sigma)} - F^*_T)(F_T r - d)\|_{H(T)} \\ \le \|F^*_{(\sigma)}\|\,\|(F_{(\sigma)} - F_T)r\|_{H(T)} &+ \|(F^*_{(\sigma)} - F^*_T)(F_T r - d)\|_{H(T)} \le C_2\, C_3 + C_4. \end{aligned}$$

Hence (2) holds. Thus Proposition 1 implies that the turnpike inequality (3) holds.

3. Exact controllability and an exact turnpike result

We consider again the time-parametric *dynamic optimization problem* $P(T)$, but without assumption (1). Assume that the system is exactly controllable in the time t_0 in the sense that for all $r \in K_1$ there exists an exact control $u^{(r)}_{exact}$ such that for the control

$$\tilde{u}^{(r)}(t) = \begin{cases} u^{(r)}_{exact}(t), & t \le t_0, \\ r, & t > t_0 \end{cases}$$

for all $T \ge t_0$ we have $\tilde{u}^{(r)} \in K(T)$ and $J_T(\tilde{u}^{(r)}) \le f(r)$.

Moreover, assume that for all $u \in K_T$ we have

$$J_T(u) \ge f(u^{(\sigma)}). \tag{7}$$

Then we have the following exact turnpike result:

Proposition 2: *Let $u^{(\sigma)}$ denote the solution of $P_{(\sigma)}$. Then for all $T \ge t_0$ the control $u^{(\delta,T)} = \tilde{u}^{(u^{(\sigma)})}$ is a solution of $P(T)$.*

Proof: Let $\omega(T)$ ($\omega_{(\sigma)}$)) denote the optimal value of $P(T)$ ($P_{(\sigma)}$ respectively). Then $\omega(T) \le J_T(\tilde{u}^{(u^{(\sigma)})}) \le f(u^{(\sigma)}) = \omega_{(\sigma)}$. On the other hand, (7) implies that $\omega(T) \ge \omega_{(\sigma)}$. Therefore $\omega_{(\sigma)} = \omega(T) = J_T(\tilde{u}^{(u^{(\sigma)})})$ which implies the assertion.

This work was supported by DFG grant CRC/TRR 154, project C03 and B04.

References

[1] C. Gotzes, H. Heitsch, R. Henrion, R. Schultz, *Feasibility of nominations in stationary gas networks with random load*, Math. Meth. of Operations Research **84** (2016), 427–457.

[2] M. Gugat, E. Trélat, E. Zuazua, *Optimal Neumann control for the 1D wave equation: Finite horizon, infinite horizon, boundary tracking terms and the turnpike property*, Syst. & Control Lett. **90** (2016), 61–70.

[3] E. Trélat, C. Zhang, E. Zuazua, *Measure-turnpike property for dissipative infinite-dimensional optimal control problems*, Preprint (2017)

[4] E. Trélat, C. Zhang, E. Zuazua, *Steady-State and Periodic Exponential Turnpike Property for Optimal Control Problems in Hilbert Spaces*, SIAM J. Control Optim. **56** (2018), 1222-1252.

[5] A. Porretta, E. Zuazua, *Long time versus steady state optimal control*, SIAM J. Control Optim. **51** (2013), 4242–4273.
[6] A. Zaslavski, *Turnpike properties in the calculus of variations and optimal control*, Springer, Heidelberg, 2006.

Risky Capacity Equilibrium Models in Electricity Markets

Daniel Ralph
(joint work with Gauthier de Maere d'Aertrycke, Andeas Ehrenmann, Yves Smeers)

We present a set of investment models, the class of *risky capacity equilibrium problems*, reflecting future (stochastic) production in a spot market, when agents are risk averse and can buy financial hedges for their physical investments. These models are structured in a unified stochastic Nash game framework. Each model is the concatenation of a short-term electricity market (perfect competition or Cournot) with long-term investment behaviour (risk neutral or risk averse behaviour in different risk trading settings). For risk trading with complete markets, we prove existence of solutions and report numerical results to illustrate the relevance of market imperfections on investments and welfare.

Our focus is two stage models but, time permitting, we show that the extension to multistage is essentially notational via a standard approach to multistage risk functions. This approach is viable for industrial scale multistage models.

The working paper for this topic and some extensions to existence of equilibria in the incomplete case is available at [1].

References

[1] G. de Maere d'Aertrycke, A. Ehrenmann, D. Ralph and Y. Smeers, *Risk trading in capacity equilibrium models*, EPRG Working Paper 1720 (2017), https://www.eprg.group.cam.ac.uk/eprg-working-paper-1720/

Inexact cutting planes for two-stage mixed-integer stochastic programs

Ward Romeijnders
(joint work with Niels van der Laan, Suvrajeet Sen)

We consider two-stage mixed-integer stochastic programs, with randomness in the right-hand side only. For so-called simple integer recourse models, we reinterpret α-approximations of [1] as inexact cutting plane approximations, connecting two existing solution approaches for solving stochastic mixed-integer programs. Moreover, we derive conditions under which inexact cutting planes are asymptotically accurate. Intuitively, this means that the cutting planes yield good approximations if the variability of the random parameters in the model is large enough. We illustrate our results for an inexact mixed-integer rounding cut on a two-dimensional example with a single random constraint.

Open problems include how to algorithmically implement several classes of inexact cutting planes and how to derive tighter error bounds for the approximations obtained using inexact cutting planes.

References

[1] W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk, *Simple integer recourse models: convexity and convex approximations*, Mathematical Programming **108** (2006), 435–473.

Risk averse linear bilevel programming under stochastic uncertainty

Matthias Claus
(joint work with Johanna Burtscheidt, Stephan Dempe)

Two-stage linear stochastic programs and linear bilevel problems under stochastic uncertainty bear significant conceptual similarities. However, the step from the first to the latter mirrors the step from optimal values to optimal solutions and entails a loss of convexity and desirable analytical properties.

Taking into account that the lower level problem may have more than one optimal solution, the talk focusses on properties of the optimistic formulation under stochastic uncertainty. Assuming that only the follower can observe the realization of the randomness, Lipschitz continuity and continuous differentiability of the objective function are established for the risk neutral and various risk averse models.

The second part of the talk examines stability of local optimal solution sets under perturbations of the underlying probability measure w.r.t. to the topology of weak convergence for models involving law-invariant convex risk measures.

Finally, extended formulations for the stochastic linear bilevel problem with finite discrete distributions and their numerical handling shall be discussed.

References

[1] M. Claus, V. Krätschmer, and R. Schultz. Weak continuity of risk functionals with applications to stochastic programming. *S*IAM Journal on Optimization, 27(1):91–108, 2017.

Shape optimization under uncertainty

Helmut Harbrecht

Introduction. Shape optimization is indispensable for designing and constructing industrial components. Many problems that arise in application, particularly in structural mechanics and in the optimal control of distributed parameter systems, can be formulated as the minimization of functionals which are defined over a class of admissible domains.

Shape optimization problems can be solved by means of gradient based minimization algorithms, which involve the shape functionals' derivative with respect to the domain under consideration. The computation of the shape gradient and the implementation of appropriate numerical optimization algorithms is meanwhile well understood, provided that the state equation's input data are given exactly. In practice, however, input data for numerical simulations in engineering are often not exactly known. One must thus address how to account for uncertain input data in the state equation.

Uncertainty in the state equation might arise from three different sources:

- Uncertainty might arise from geometric entities like a certain part of boundary which has not to be optimized but is prescribed.
- The right-hand side of the state equation might be random.
- The material parameters, entering the partial differential operator, might be not exactly known.

We separately consider these sources of uncertainty and discuss their impact on the shape optimization problem. Especially, we show the well-posedness of the problem formulations and present numerical solution methods.

Shape optimization under geometric uncertainty. Bernoulli's free boundary problem is concerned with finding the exterior (free) boundary Γ of an annular domain $D \subset \mathbb{R}^n$ for a given interior (fixed) boundary Σ such that, besides Dirichlet boundary conditions at both boundaries, also a Neumann boundary condition is satisfied at the exterior boundary. The problem under consideration models for example the growth of anodes in electrochemical processes and can be seen as the prototype of a free boundary problem arising in many applications.

We shall consider the situation that the interior boundary is random, i.e., it is $\Sigma = \Sigma(\omega)$ with an additional parameter $\omega \in \Omega$. Such an assumption arises when treating tolerances in fabrication processes or when the interior boundary is only known by measurements which typically contain errors. We are thus looking for a tuple $\big(D(\omega), u(\omega)\big)$ such that there holds

$$\begin{aligned} \Delta u(\omega) &= 0 && \text{in } D(\omega), \\ u(\omega) &= 1 && \text{on } \Sigma(\omega), \\ -\frac{\partial u}{\partial \mathbf{n}}(\omega) = g,\ u(\omega) &= 0 && \text{on } \Gamma(\omega). \end{aligned} \tag{1}$$

The questions which have to be addressed are the following:

(1) What is a suitable model for the domain $D(\omega)$? Is the problem well-posed in the sense of $D(\omega)$ being almost surely well-defined?
(2) How to define the expectation of a random domain? The main difficulty to deal with here is that the space of domains is not linear.
(3) How to compute the solution to the random free boundary problem numerically?

We provide the theoretical background that ensures the well-posedness of the problem under consideration and describe two different frameworks to define the expectation and the deviation of the resulting annular domain. The first approach is based on the Vorob'ev expectation, which can be defined for arbitrary sets. The second approach is based on a particular parametrization. We compare these approaches by analytical computations for circular interior domains and by numerical experiments for more general geometric configurations. For the numerical approximation of the domain's expectation and deviation, we propose a sampling method like the (quasi-) Monte Carlo quadrature. Then, each particular realization Σ_i of the interior boundary leads to an exterior boundary Γ_i via the solution of Bernoulli's free boundary problem. It is computed by solving the shape optimization problem

$$\begin{aligned} J(D_i) &= \int_{D_i} \left\{ \|\nabla u_i\|^2 + g^2 \right\} \mathrm{d}\mathbf{x} \to \min \\ \text{subject to} \quad & \Delta u_i = 0 \text{ in } D_i, \quad u_i = 1 \text{ on } \Sigma_i, \quad u_i = 0 \text{ on } \Gamma_i. \end{aligned}$$

Shape optimization in case of random diffusion. We consider again Bernoulli's free boundary problem (1), but now the situation that the material contained in the domain D is not perfectly homogeneous. Hence, we arrive at the following random free boundary problem: seek the free boundary Γ, such that

$$\begin{aligned} \operatorname{div}\big(\alpha(\omega)\nabla u(\omega)\big) &= 0 && \text{in } D, \\ u(\omega) &= 1 && \text{on } \Sigma, \\ -\alpha(\omega)\frac{\partial u}{\partial \mathbf{n}}(\omega) = g,\ u(\omega) &= 0 && \text{on } \Gamma, \end{aligned} \tag{2}$$

holds for almost all $\omega \in \Omega$. Since we intend to model a uniformly elliptic random perturbation of the Laplace operator, the random diffusion is assumed to satisfy

$$0 < \alpha_{\min} \le \alpha(\omega) \le \alpha_{\max} < \infty$$

almost everywhere in D.

For solving the random free boundary problem (2), we first show that the minimizer of the shape optimization problem

$$\begin{aligned} J(D,\omega) &= \int_D \left\{ \alpha(\omega)\|\nabla u(\omega)\|^2 + \frac{g^2}{\alpha(\omega)} \right\} \mathrm{d}\mathbf{x} \to \min \\ \text{subject to} \quad & \operatorname{div}\big(\alpha(\omega)\nabla u(\omega)\big) = 0 \text{ in } D, \quad u(\omega) = 1 \text{ on } \Sigma, \quad u(\omega) = 0 \text{ on } \Gamma \end{aligned} \tag{3}$$

solves the free boundary problem (2) for each particular $\omega \in \Omega$. This is done by deriving the Hadamard representation of the associated shape gradient and proving

that the necessary optimality condition imposes the desired Neumann boundary condition.

Therefore, since the random diffusion induces a random state and thus a random shape functional, we are going to minimize the ensemble average $\mathbb{E}\big[J(D,\omega)\big]$ of the random shape functional. Note that this shape optimization problem is well-posed since we are minimizing a continuous energy functional.

Shape optimization in case of random right-hand sides. We shall next consider shape optimization problems where the right-hand side in the state equation is random. Then, the state is random, but depends *linearly* on the randomness. In this situation, the expectation and variance of a quadratic objective can be reformulated as deterministic expressions by exploiting the state's moments. This leads to cheap, deterministic algorithms to minimize such objectives.

Consider, for example, Bernoulli's free boundary problem in case of random Dirichlet data at the interior boundary:

$$\begin{aligned} \Delta u(\omega) &= 0 && \text{in } D,\\ u(\omega) &= f(\omega) && \text{on } \Sigma,\\ -\frac{\partial u}{\partial \mathbf{n}}(\omega) &= g,\ u(\omega) = 0 && \text{on } \Gamma. \end{aligned} \tag{4}$$

We show that the random shape optimization problem

$$\begin{aligned} &\mathbb{E}[J(D,\omega)] = \mathbb{E}\bigg[\int_D \big\{\|\nabla u(\omega)\|^2 + g^2\big\}\,\mathrm{d}\mathbf{x}\bigg] \to \min\\ &\text{subject to}\quad \Delta u(\omega) = 0 \text{ in } D,\quad u(\omega) = f(\omega) \text{ on } \Sigma,\quad u(\omega) = 0 \text{ on } \Gamma \end{aligned} \tag{5}$$

can be reformulated into the deterministic shape optimization problem

$$\mathbb{E}[J(D,\omega)] = \int_D \left\{ \left(\|(\nabla \otimes \nabla)\,\mathrm{Cor}[u](\mathbf{x},\mathbf{y})\|\Big|_{\mathbf{y}=\mathbf{x}} \right)^2 + g^2 \right\} \mathrm{d}\mathbf{x} \to \min,$$

where $\mathrm{Cor}[u](\mathbf{x},\mathbf{y})$ denotes the two-point correlation of random state $u(\omega)$ given by (5).

References

[1] R. Brügger, R. Croce, and H. Harbrecht, *Solving a Bernoulli type free boundary problem with random diffusion*, *Preprint 2018-09*, Fachbereich Mathematik, Universität Basel, Switzerland, 2018.

[2] M. Dambrine, H. Harbrecht, and B. Puig, *Incorporating knowledge on the measurement noise in electrical impedance tomography*, *Preprint 2017-14*, Fachbereich Mathematik, Universität Basel, Switzerland, 2017 (to appear in ESAIM Control Optim. Calc. Var.).

[3] M. Dambrine, H. Harbrecht, M. Peters, and B. Puig, *On Bernoulli's free boundary problem with a random boundary*, Int. J. Uncertain. Quantif. **7**(4) (2017) 335–353.

[4] M. Dambrine, C. Dapogny, and H. Harbrecht, *Shape optimization for quadratic functionals and states with random right-hand sides*, SIAM J. Control Optim. **53**(5) (2015) 3081–3103.

[5] H. Harbrecht and M. Peters, *Solution of free boundary problems in the presence of geometric uncertainties*, in M. Bergounioux et al., editors, *Topological Optimization and Optimal Transport in the Applied Sciences*, pages 20–39, de Gruyter, Berlin-Bosten, 2017.

Variance-based stochastic extragradient methods with line search for stochastic variational inequalities

Alejandro Jofré

We propose a dynamic sampled stochastic approximated (DS-SA) extragradient method for stochastic variational inequalities (SVI) that is robust with respect to an unknown Lipschitz constant. We propose, to the best of our knowledge, the first provably convergent robust SA method with variance reduction, either for SVIs or stochastic optimization jointly with optimal rates of convergence using dynamic sampling.

References

[1] Iusem, A., Jofré, A., Oliveira, R. I., and Thompson, P. Variance-based stochastic extragradient methods with line search for stochastic variational inequalities. arXiv preprint arXiv:1703.00262. (2017).

On a dynamic decision model with quasi-hyperbolic discounting

Anna Jaśkiewicz
(joint work with Ł. Balbus, A.S. Nowak)

The issue of dynamic inconsistency in sequential decision models with changing preferences in time was pointed out in the seminal paper of [4]. The time preference studies found that discount rates are much grater in the short run than in the long run. To model this phenomena researchers have adopted discount functions from the class of generalized hyperbolas, [2] and references therein. The discrete-time analog of quasi-hyperbolic discounting involves the functions: $1, \delta\beta, \delta\beta^2, ...$, where $\beta \in (0,1)$ is a long run discount factor and $\delta > 0$ is a short run discount factor. Such discounting was first used by [3]. Specifically, they noticed that finding a time-consistent solution may be obtained by looking for a Nash equilibrium in certain games played by countably many short-lived players assuming that each player can act only once.

Define $S := [0, +\infty)$, $S_+ := (0, +\infty)$ and, for each $s \in S$, $A(s) := [0, s]$. The set S is referred to as the *state space*. It represents the *set of "levels" for the renewable resource* and $A(s)$ is the set of *available actions* (possible consumption levels) in state $s \in S$. In a *dynamic choice model* with *quasi-hyperbolic preferences* and the state space S we envision an individual decision maker as a sequence of autonomous temporal *selves*. These selves are indexed by period numbers $t \in T := \mathbb{N}$ in which they make their choices. More precisely, for a given state $s_t \in S$ at the beginning of t-th period, self t chooses a consumption level $a_t \in A(s_t)$ and the remaining part $y_t := s_t - a_t$ is invested for future *selves*. Self t's satisfaction is measured by a bounded continuous *period utility function* $u_\tau : S \to S$ for all $\tau \geq t$.

Let q_t be a transition probability from S to S. Then, the state s_{t+1} is generated by $q_t(\cdot|y_t)$ depending on the investment $y_t \in A(s_t)$.

Let Φ be the set of all Borel measurable functions $\phi : S \to S$ such that $\phi(s) \in A(s)$ for each $s \in S$. A *Markov strategy* for self t is a function $c_t \in \Phi$. We put $i_t(s) = s - c_t(s)$, $s \in S$. This is an investment strategy (or saving) of self t for following selves. For any sequence $(c_n) \in \Phi^\infty := \Phi \times \Phi \times \cdots$ of strategies of all selves and any $t \in T$, we define $c^t := (c_t, c_{t+1}, ...)$. For any state s_t and any $c^t \in \Phi^\infty$, the transition probabilities $q_\tau(\cdot|i_\tau(s))$ induced by q_τ with $\tau \geq t$ generate due to the Ionescu-Tulcea theorem a unique probability measure $P_{s_t}^{c^t}$ on S^∞ endowed with the product σ-algebra. Let $E_{s_t}^{c^t}$ denote the expectation operator corresponding to the measure $P_{s_t}^{c^t}$.

The *expected utility* of self t is

$$U_t(c^t)(s_t) := E_{s_t}^{c^t}\left(u_t(c_t(s_t)) + \delta\beta \sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} u_\tau(c_\tau(s_\tau))\right). \tag{1}$$

Clearly, the expression in (1) is well-defined.

For any $c^n = (c_n, c_{n+1}, ...) \in \Phi^\infty$, $n \geq 2$ and $s_n \in S$, let

$$J_n(c^n)(s_n) = E_{s_n}^{c^n}\left(\sum_{\tau=n}^{\infty} \beta^{\tau-n} u_\tau(c_\tau(s_\tau))\right).$$

Then we have

$$U_t(c^t)(s_t) = u_t(c_t(s_t)) + \delta\beta \int_S J_{t+1}(c^{t+1})(s_{t+1}) q_t(ds_{t+1}|s_t - c_t(s_t)).$$

Define

$$P_t(a, c^{t+1})(s) := u_t(a) + \delta\beta \int_S J_{t+1}(c^{t+1})(s_{t+1}) q_t(ds_{t+1}|s - a),\ s \in S,\ a \in A(s).$$

Definition 1. A *Markov Perfect Equilibrium* (MPE) is a sequence $\hat{c} = (\hat{c}_t)_{t\in T} \in \Phi^\infty$ such that for every $s \in S$ and $t \in T$, we have

$$\sup_{a\in A(s)} P_t(a, \hat{c}^{t+1})(s) = P_t(\hat{c}_t(s), \hat{c}^{t+1})(s) = U_t(\hat{c}^t)(s).$$

Definition 2. A *Stationary Markov Perfect Equilibrium* ($SMPE$) is an MPE $\hat{c} = (\hat{c}_t)_{t\in T} \in \Phi^\infty$ such that $\hat{c}_t = c_0$ for some $c_0 \in \Phi$ and for all $t \in T$.

In a stationary MPE every self uses the same consumption strategy.

One can think that every self is a short-lived player in a non-cooperative game and acts only once. The payoff function of self $t \in T$ is given by (1). Then an MPE $\hat{c} = (\hat{c}_t)_{t\in T} \in \Phi^\infty$ is a Nash equilibrium in this game.

Let $c = (c_t)_{t\in T} \in \Phi^\infty$ and

$$\tilde{u}_{c_{t+n}}(s_{t+n}) := u_{t+n}(c_{t+n}(s_{t+n})), \quad n \in T,\ s_{t+n} \in S.$$

For any bounded Borel measurable function $v : S \to S$, define

$$Q_{c_n} v(s_n) := \int_S v(s) q_n(ds|s_n - c_n(s_n)).$$

Consider the composition $Q_{c_{t+1}} \cdots Q_{c_{t+n}} \tilde{u}_{c_{t+1+n}}(s_{t+1})$. By the Ionescu-Tulcea theorem, it follows that

$$J_{t+1}(c^{t+1})(s_{t+1}) = u_{t+1}(c_{t+1}(s_{t+1})) + \sum_{n=1}^{\infty} \beta^n Q_{c_{t+1}} \cdots Q_{c_{t+n}} \tilde{u}_{c_{t+1+n}}(s_{t+1}). \tag{2}$$

Clearly, (2) is bounded and non-negative for any $s_{t+1} \in S$, $c^{t+1} \in \Phi^\infty$ and $t \in T$.

We now formulate our main assumptions.

(U) For each $t \in T$, the function $u_t : S \to S$ is increasing and strictly concave.

(Q) The transition probability q_t is weakly continuous on S, that is, for each $y_0 \in S$ and $y_m \to y_0$, we have $q_t(\cdot|y_m) \Rightarrow q_t(\cdot|y_0)$ as $m \to \infty$. Moreover, for each $y \in S_+$, the probability measure $q_t(\cdot|y)$ is non-atomic and $q_t(\cdot|0)$ has no atoms in S_+.

We now define some special classes of strategies of the players. By F, we denote the set of all continuous from the left mappings $c : S \to S$ such that the function $i(s) := s - c(s)$ is non-decreasing and $0 \le c(s) \le s$ for all $s \in S$. Note that i is lower semicontinuous. Thus, $c \in F$ is upper semicontinuous.

We can now state our first two main results. The proofs can be found in [1].

Theorem 1. Assume that (U) and (Q) are satisfied.
(a) Then, there exists an MPE $\hat{c} = (\hat{c}_t)$ that belongs to the space F^∞
(b) If, in addition, the model is stationary, i.e., $q_t = q$ and $u_t = u$ are independent of $t \in T$, then there exists an $SMPE$ $\hat{c} = (c_0, c_0, ...)$ with $c_0 \in F$.

In this talk, we presented the solution for non-atomic transition probabilities. The open issue is the existence of a $SMPE$ with deterministic transition probabilities in a stationary decision model.

References

[1] Ł. Balbus, A. Jaśkiewicz, A.S. Nowak, Markov perfect equilibria in a dynamic decision model with quasi-hyperbolic discounting, *Annals of Operations Research* (2018), https://doi.org/10.1007/s10479-018-2778-2.

[2] C. Harris, D. Laibson, Dynamic choices of hyperbolic consumers, *Econometrica* **69** (2001), 935–957.

[3] E. Phelps, R. Pollak, On second best national savings, and game equilibrium growth, *Review of Economic Studies* **35** (1968), 195–199.

[4] R.H. Strotz, Myopia and inconsistency in dynamic utility maximization, *Review of Economic Studies* **23** (1956), 165–180.

A Dynamic Approach to Scaling in Bundle Methods for Convex Optimization

Christoph Helmberg
(joint work with Alois Pichler)

For a detailed account and proofs, see [4]. Bundle methods [1, 6, 2] are designed to minimize a nonsmooth convex $f\colon \mathbf{R}^n \to \mathbf{R}$ given by a first order oracle, which for a point $x \in \mathbf{R}^n$ returns function value $f(x)$ and any subgradient $g \in \partial f(x)$ in the subdifferential of f in x. By the subgradient inequality the data $\omega = (\gamma, g) = (f(x) - g^\top x, g)$ describes a global affine minorant $f_\omega(y) := \gamma + g^\top y \le f(y)$, called minorant ω for brevity. The set of all affine minorants $W_f := \{(\gamma, g)\colon f(y) \ge \gamma + g^\top y \text{ for all } y \in \mathbf{R}^n\}$ is convex and $f = \sup_{\omega \in W_f} f_\omega$. Any subset $\widehat{W}_f \subseteq W_f$ yields a *cutting model* $f_{\widehat{W}_f} := \sup_{\omega \in \widehat{W}_f} f_\omega \le f$.

In iteration $k \in \mathbb{N}$ a bundle method determines for a given compact (often finite) model $\widehat{W}_f^k \subseteq W_f$ and a *center of stability* $\hat{y}^k$ a next *candidate* $\bar{y}^k$ by solving a *bundle subproblem*

$$\bar{y}^k = \operatorname*{argmin}_{y \in \mathbf{R}^n} f_{\widehat{W}_f^k}(y) + \frac{1}{2}\|y - \hat{y}^k\|_H^2,$$

where the quadratic (or proximal) term $\|y-\hat{y}^k\|_H^2 = (y-\hat{y}^k)^\top H(y-\hat{y}^k)$ with positive definite *scaling* matrix H serves as proximity or trust region control. Strong convexity of the proximal term and compactness of $\widehat{W}_f^k$ ensure the existence of a unique saddle point $(\bar{y}^k, \bar{\omega}^k) \in \mathbf{R}^n \times \operatorname{conv} \widehat{W}_f^k$ to the Lagrangian $L(y, \omega{=}(\gamma, g)) = \gamma + g^\top y + \frac{1}{2}\|y - \hat{y}^k\|_H^2$. The saddle point minorant $\bar{\omega}^k = (\bar{\gamma}^k, \bar{g}^k) \in \operatorname{conv} \widehat{W}_f^k$ is called the *aggregate* and by optimality $\bar{y}^k = \hat{y}^k - H^{-1}\bar{g}^k$. The function is then evaluated at $\bar{y}^k$ by a first order oracle which returns $f(\bar{y}^k)$ and some subgradient $g^k \in \partial f(\bar{y}^k)$ giving the candidate minorant $\omega^k = (\gamma^k = f(\bar{y}^k) - (g^k)^\top \bar{y}^k, g^k) \in W_f$. Then, depending on a *descent test* $f(\hat{y}^k) - f(\bar{y}^k) \ge \kappa[f(\hat{y}^k) - f_{\bar{\omega}^k}(\bar{y}^k)]$ for given $\kappa \in (0, 1)$, a *descent step* sets $\hat{y}^{k+1} = \bar{y}^k$ or a *null step* keeps $\hat{y}^{k+1} = \hat{y}^k$ while improving the model by requiring $\{\bar{\omega}^k, \omega^k\} \subseteq \widehat{W}_f^{k+1}$. The framework ensures $f(\hat{y}^k) \to f^* := \inf_{y \in \mathbf{R}^n} f(y)$ for $k \to \infty$ [1]; if minimizers of f exist, the sequence $\hat{y}^k$ converges to some minimizer; if f is bounded from below, $\liminf \|\bar{g}^k\| = 0$.

The same still holds if H of the quadratic term or the proximal term is dynamically updated with some care, see [1], and various approaches inspired by quasi-Newton methods have been proposed for this, *e. g.*, [7, 9, 8, 3].

We propose a different approach that derives H directly from a current collection $\widehat{W}_f$ of minorants, the current aggregate $\bar{\omega}$ and a *violation parameter* $\varepsilon > 0$: for any $\omega \in \widehat{W}_f$ satisfying $f_{\bar{\omega}}(\hat{y}) > f_\omega(\hat{y}) - \varepsilon$ the quadratic model of $\bar{\omega}$ should not violate f_ω by more than ε, *i. e.*, $f_{\bar{\omega}} + \frac{1}{2}\|\cdot - \hat{y}\|_H^2 \ge f_\omega - \varepsilon$. Denoting the set of (symmetric) positive semidefinite matrices of order n by S_+^n and, for symmetric A, B, the Loewner partial order by $A \succeq B :\Leftrightarrow A - B \in S_+^n$ ($A \succ B :\Leftrightarrow A - B$ positive definite), and the trace inner product by $\langle A, B\rangle = \operatorname{tr} AB$ this gives rise to the following requirement.

Lemma 1. *Given $\hat{y} \in \mathbf{R}^n$, violation parameter $\varepsilon > 0$ and $\bar{\omega} = (\bar{\gamma}, \bar{g}), \omega = (\gamma, g) \in \mathbf{R} \times \mathbf{R}^n$ with $\gamma + g^\top \hat{y} < \bar{\gamma} + \bar{g}^\top \hat{y} + \varepsilon$, then for $\bar{H} \in S_+^n$*

$$f_{\bar{\omega}} + \frac{1}{2}(\cdot - \hat{y})^\top \bar{H}(\cdot - \hat{y}) \geq f_\omega - \varepsilon \quad \Leftrightarrow \quad \bar{H} \succeq \frac{1}{2}\frac{(\bar{g} - g)(\bar{g} - g)^\top}{\bar{\gamma} + \varepsilon - \gamma + (\bar{g} - g)^\top \hat{y}}.$$

The search for a "best" H may be formulated as a semidefinite program.

Theorem 2. *Given $\hat{y} \in \mathbf{R}^n$, violation parameter $\varepsilon > 0$, $\bar{\omega} = (\bar{\gamma}, \bar{g}), \omega^i = (\gamma^i, g^i) \in \mathbf{R} \times \mathbf{R}^n$ with $\gamma^i + (g^i)^\top \hat{y} < \bar{\gamma} + \bar{g}^\top \hat{y} + \varepsilon$ for $i = 1, \ldots, k$ and a positive definite $C \in S^n$, let $\bar{H}$ be an optimal solution of the semidefinite program*

$$\begin{array}{lll} \textit{minimize} & \langle C, H \rangle & \\ \textit{subject to} & H \succeq \frac{1}{2}\frac{(\bar{g} - g^i)(\bar{g} - g^i)^\top}{\bar{\gamma} + \varepsilon - \gamma^i + (\bar{g} - g^i)^\top \hat{y}}, & i = 1, \ldots, k, \\ & H \succeq 0. & \end{array} \tag{1}$$

Then $f_{\bar{\omega}} + \frac{1}{2}(\cdot - \hat{y})^\top \bar{H}(\cdot - \hat{y}) \geq \max_{i=1,\ldots,k} f_{\omega^i} - \varepsilon$. Furthermore, $\mathcal{D} := \operatorname{span}\{\bar{g} - g^i \colon i = 1, \ldots, k\} \subseteq \mathcal{R}(\bar{H})$ for range space $\mathcal{R}$. If $C = I$, equality holds, $\mathcal{D} = \mathcal{R}(\bar{H})$.

$C = I$ minimizes the sum of the eigenvalues of H or the overall curvature and restricts H to the space explored. In order to relate the optimal solution to the Hessian of a smooth convex function, consider the case of quadratic functions.

Lemma 3. *Let $f(x) = \frac{1}{2}x^\top Ax + b^\top x + \rho$ with $A \in S_+^n$ and suppose $\hat{y} \in \mathbf{R}^n$, $\bar{\omega} = (\bar{\gamma}, \bar{g} = \nabla f(\hat{y}))$ and violation parameter $\varepsilon \geq 0$ satisfy $\bar{\gamma} + \bar{g}^\top \hat{y} \leq f(\hat{y}) \leq \bar{\gamma} + \bar{g}^\top \hat{y} + \varepsilon$. Given $y^i \in \mathbf{R}^n$ and $\omega^i = (\gamma^i, g^i) = (f(y^i) - \nabla f(y^i)^\top y^i, \nabla f(y^i))$, $i = 1, \ldots, k$, let $P \in \mathbf{R}^{n \times h}, P^\top P = I_h$ have range space $\mathcal{R}(P) = \operatorname{span}\{A(\hat{y} - y^i) \colon i = 1, \ldots, k\}$, then $PP^\top APP^\top$ is feasible for (1).*

Thus, the projected Hessian is always feasible but it is not necessarily optimal. If, however, the $\hat{y} - y^i$ include a full set of conjugate directions and $\bar{g}$ is the gradient at $\hat{y}$ and $\varepsilon = 0$, the optimal H can be shown to be the Hessian for, *e. g.*, $C = I$.

Theorem 4. *Let $f(x) = \frac{1}{2}x^\top Ax + b^\top x + \rho$ with $A \in S_+^n$ and $v_i := \hat{y} - y^i$ contain a family of conjugate directions of A satisfying $v_i^\top A v_j = \delta_{ij}$ for $i, j = 1, \ldots, n$ with Kronecker delta δ_{ij}. Then for $\bar{g} = A\hat{y} + b$, $g^i = Ay^i + b$, $\bar{\gamma}^i = f(\hat{y}) - \bar{g}^\top \hat{y}$, $\gamma^i = f(y^i) - {g^i}^\top y^i$, $\varepsilon = 0$ and $C = \sum_{i=1}^n \zeta_i v_i v_i^\top$ for any choice of $\zeta_i > 0$, $i = 1, \ldots, n$, the Hessian $\bar{H} = A$ is the unique optimal solution of (1).*

For practical purposes solving (1) is computationally too expensive. Instead, we try to choose a $Q \in \mathbf{R}^{n \times h}$, whose columns form an orthonormal basis of the hopefully most relevant subspace of $\mathcal{D}$ of Th. 2. In the nonsmooth case a natural choice for this are the normalized eigenvectors to the h largest eigenvalues of the matrix $VV^\top$ for $V = [d_1, \ldots, d_k]$ with $d_i = (\bar{g} - g^i)/\sqrt{2(\bar{\gamma} + \varepsilon - \gamma^i + (\bar{g} - g^i)^\top \hat{y})}$ for $i = 1, \ldots, k$. Alternatively, compute the singular value decomposition of V and choose the left eigenvectors to the h largest singular values. Next, we restrict $\bar{H}$ to be of the form $\bar{H} = Q\Lambda_{\bar{H}}Q^\top$ with diagonal $\Lambda_{\bar{H}} = \operatorname{diag}(\lambda_1^{\bar{H}}, \ldots, \lambda_h^{\bar{H}})$. Finally, instead of simplifying the semidefinite constraints of (1) to $\Lambda_{\bar{H}} \succeq Q^\top d_i d_i^\top Q$ we put $\lambda_j^{\bar{H}} = \max\{\operatorname{diag}(Q^\top d_i d_i^\top Q)_j = (Q^\top d_i)_j^2 \colon i = 1, \ldots, k\}$ $j = 1, \ldots, h$.

Again, for smooth f the choice of $\lambda_j^{\bar{H}}$ and Q via the singular value decomposition of V may be justified to some extent. Indeed, for the quadratic f of Lemma 3 the columns of V read $d_i = A(\hat{y} - y^i)/\|A^{\frac{1}{2}}(\hat{y} - y^i)\|$. Whenever the generating directions $s_i = \hat{y} - y^i$ are spread rather randomly the next result gives some hope that the columns of Q will be close to an eigenvector basis of A.

Theorem 5. *Given $\delta > 0$ and a positive definite $A \in S^n$, let directions $s_i \in \mathbf{R}^n \setminus \{0\}$, $i = 1, \ldots, k$, be chosen by a rotationally invariant distribution with second moments, let $V = [d_1, \ldots, d_k]$ with $d_i = As_i/\|A^{\frac{1}{2}}s_i\|$ and let $Q\Sigma_V P^\top = V$ denote a singular value decomposition with $Q^\top Q = I_n$. With probability going to one for $k \to \infty$ there is an orthogonal matrix $\bar{Q} = Q + O(\delta)$ diagonalizing $A = \bar{Q}\Lambda_A\bar{Q}^\top$.*

In practice the differences $\hat{y} - y^i$ are unlikely to be distributed randomly. Still, together with Th. 4 the result indicates why in the smooth case the use of the singular value decomposition in this more practical heuristic might indeed produce a reasonable estimate of the projected Hessian.

Several further implementational decisions are required to develop this into a fully automatic dynamic scaling approach via computing such a low rank approximation $\bar{H}$. Preliminary computational results indicate significant benefits even if only the diagonal of $\bar{H}$ is used in the bundle method when solving large scale real world problems [5] arising in scheduling trucks for transporting pallets between warehouses so as to minimize the expected number of future transports.

References

[1] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization*. Springer, 2nd edition, 2006.

[2] W. de Oliveira and C. Sagastzábal. Bundle methods in the XXIst century: a bird's-eye view. *Pesquisa Operacional*, 34(3):647–670, 2014.

[3] N. Haarala, K. Miettinen, and M. M. Mäkelä. Globally convergent limited memory bundle method for large-scale nonsmooth optimization. *Math. Programming*, 109(1):181–205, 2007.

[4] C. Helmberg and A. Pichler. Dynamic scaling and submodel selection in bundle methods for convex optimization. Preprint 2017, Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany, Aug. 2017.

[5] C. Helmberg and S. Röhl. A case study of joint online truck scheduling and inventory management for multiple warehouses. *Operations Research*, 55(4):733–752, July 2007.

[6] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I+II*, volume 305 and 306 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1993.

[7] C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: From conceptual to implementable forms. *Math. Programming*, 76:393–410, 1997.

[8] L. Lukšan and J. Vlček. Globally convergent variable metric method for convex nonsmooth unconstrained minimization. *J. Optim. Theory Appl.*, 102:593–613, 1999.

[9] R. Mifflin and C. Sagastizábal. A $\mathcal{VU}$-algorithm for convex minimization. *Math. Programming*, 104(2-3):583–608, 2005.

A $\mathcal{VU}$-point of view of nonsmooth optimization

Claudia Sagastizábal

For functions that fail to be differentiable, nonsmoothness appears according to some structure that can be exploited algorithmically. Such is the case, for instance, for functions defined as the pointwise maximum of, say, C^2 convex functions:

$$f(x) := \max\{f_i(x) : i \in \{1, \ldots, m\}\} .$$

Suppose one wants to minimize such function over the whole of $\Re^n$ The *active index set*

$$I(x) := \{i \in \{1, \ldots, m\} : f(x) = f_i(x)\} ,$$

characterizes a manifold along which the function appears to be smooth:

$$\mathcal{M} = \{y \approx x : f(y) = f_i(y) \text{ for } i \in I(x)\}$$

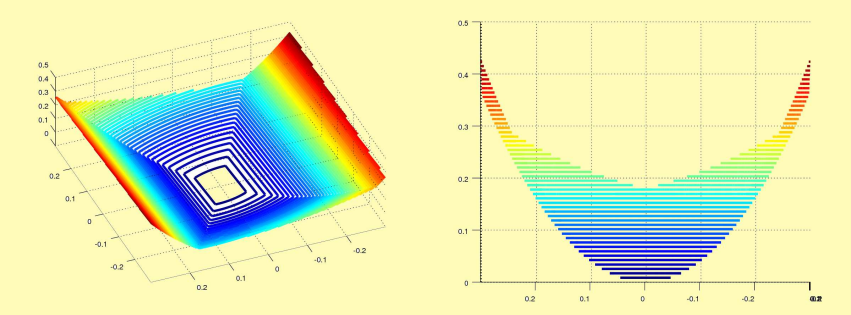
Near a minimizer $\bar{x}$, the identification of such a manifold is therefore of interest, since along trajectories in $\mathcal{M}$ it is possible to make Newton-like steps, and converge to $\bar{x}$ with superlinear speed, even if the function is not smooth.

This is the basics of the $\mathcal{V}U$-space decomposition method introduced in [LOS00]; see also [MS03] and [Lew02]. To illustrate the methodology, consider the following simple function, with a unique minimizer at $x^* = 0 \in \Re^n$, depending on gives matrices A and B, for A with nontrivial kernel, and $B \succ 0$:

$$f(x) = \max\Big(|x^\top Ax| , x^\top Bx\Big) .$$

On the null space of A, $\mathcal{N}(A)$, the function is not differentiable; on the same subspace the first term vanishes: $f|_{\mathcal{N}(A)}$ looks smooth. All the nonsmoothness of f near x is concentrated in the orthog0nal subspace defined by the range subspace, $\mathcal{R}(A)$. The respective "smooth" and "sharp" subspaces are the so-called $\mathcal{U}$ and $\mathcal{V}$ subspaces. Figure 1 shows such views for a function like in the example, with $n = 2$.

Figure 1. Views of a nonsmooth function



The $\mathcal{V}U$-algorithm [MS05] is the a globally convergent algorithm with Q-superlinear rate designed according to these premises. We refer to [Sag18] for references and more details. Roughly speaking, to track the smooth manifold, special trajectories, denoted by $\chi(u)$ in [MS02] are shown to have a $\mathcal{V}$-component that goes to zero faster t han the $\mathcal{U}$-component. As a result, along the smooth trajectory a Newton-like method (acting solely on the $\mathcal{U}$-component) drives the convergence speed of the iterates, therefore yielding the desired superlinear speed.

A fundamental result in the identification process is [MS02, Thm.5.1], reproduced below, and where riS stands for the relative interior of a convex set S. The result links the proximal point operator of a convex function f at a given $x \in \Re^n$

$$p^f(x) := \arg\min_{p\in\Re^n} \left\{ f(p) + \frac{1}{2}\mu\|p-x\|_2^2 \right\},$$

with the special trajectories in the smooth manifold, denoted by $\chi(u)$.
Theorem: [Proximal points are on the fast track] Let $f : \Re^n \to \Re$ be a convex function with minimizer $\bar{x} \in \Re^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\chi(u)$ be a fast track. Given a positive parameter μ, for all x close enough to $\bar{x}$ there exists $u(x)$ such that $p^f_\mu(x) = \chi\Big(u(x)\Big)$.

The interest of parameterizing the function via $\chi(u)$, is that it makes f appear as if it were smooth. The corresponding function, defined only on the $\mathcal{U}$-subspace, is the $\mathcal{U}$-Lagrangian introduced in [LOS00]:

$$L_{\mathcal{U}}(u) = f(\chi(u)).$$

Associated with $\chi(u)$ there is a dual element, called the $\mathcal{U}$-gradient along which the Newton step is done. In particular, Corollary 3.5 in [LOS00] shows that the $\mathcal{V}$-component of the trajectory goes to zero faster than the $\mathcal{U}$-component, so the speed of convergence of the overall process is directed by the speed of the $\mathcal{U}$-component.

Back to the proximal point operator, the optimality condition for $p^f(x)$ gives the equivalence

$$0 \in \partial f(p^f(x)) + \mu\Big(p^f(x) - x\Big) \quad \Longleftrightarrow \quad p^f(x) = x - \frac{1}{\mu} g^p \text{ for some } g^p \in \partial f(p^f(x)).$$

In the inclusion above the implicit nature of the subgradient makes it impossible to compute the proximal point of f, unless the function f is not available analytically.

In many optimization problems, the only information available is given by a *first-order oracle*, delivering the functional value and one subgradient, say $f^i = f(x^i)$ and $g^i \in \partial f(x^i)$ for a given $x^i \in \Re^n$. At the kth iteration of a minimization algorithm, the output of past oracle iterations (the "bundle of information") can be used to define a cutting-planes model for the function:

$$f(x) \approx \mathfrak{m}^{\mathrm{k}}(x) \quad \text{for} \quad \mathfrak{m}^{\mathrm{k}}(x) = \max\{ f^i + (x - x^i)^\top g^i \,, i = 1, \ldots, k\}.$$

It was shown in [CL93] that

$$\lim_{k\to\infty} p^{\mathfrak{m}^{\mathrm{k}}}(\hat{x}) = \lim_{k\to\infty} \arg\min_{p\in\Re^n} \{\mathfrak{m}^{\mathrm{k}}(p) + \frac{1}{2}\mu\|p-\hat{x}\|_2^2\} = p^f(\hat{x}),$$

see also [Aus87], [Fuk84]. This nice result shows that enriching the cutting-plane model without moving the prox-center x eventually gives the proximal point for f. Together with the theorem linking proximal points to the fast track, this opens the door towards implementability of the $\mathcal{VU}$ decomposition. More precisely, the track is approximated using the proximal point of the cutting-plane model:

$$\chi(u) \approx p_\mu^{\mathfrak{m}^k}(\hat{x}^k) \quad \text{and} \quad \nabla L_{\mathcal{U}}(u) \approx \hat{g}^k \in \partial_{\hat{\varepsilon}^k} f(\hat{x}^k)\,,$$

where $\hat{x}^k$ is the prox-center at the kth iteration.

References

[Aus87] A. Auslender, *Numerical methods for nondifferentiable convex optimization*, Mathematical Programming Studies **30** (1987), 102–126.

[CL93] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming **62** (1993), no. 2, 261–275.

[Fuk84] M. Fukushima, *A descent algorithm for nonsmooth convex optimization*, Mathematical Programming **30** (1984), 163–175.

[Lew02] A. S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM Journal on Optimization **13** (2002), no. 3, 702–725.

[LOS00] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc. **352** (2000), no. 2, 711–729.

[MS02] R. Mifflin and C. Sagastizábal, *Proximal points are on the fast track*, Journal of Convex Analysis **9** (2002), no. 2, 563–579.

[MS03] ———, *Primal-Dual Gradient Structured Functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM Journal on Optimization **13** (2003), no. 4, 1174–1194.

[MS05] ———, *A $\mathcal{VU}$-algorithm for convex minimization*, Mathematical Programming **104** (2005), no. 2–3, 583–608.

[Sag18] Claudia Sagastizábal, *A $\mathcal{VU}$-Spoint of view of nonsmooth optimization*, Proceedings of the International Congress of Mathematicians 2018- Invited Lectures, vol. 3, 2018, pp. 3785–3806.

Stochastic Gradients for PDE Constrained Optimization under Uncertainty

Caroline Geiersbach

(joint work with G. Pflug)

Convergence of a projected stochastic gradient method is demonstrated for convex objective functionals with convex and bounded constraints in Hilbert spaces. In the convex case, the sequence of iterates u_n converges weakly to a point in the set of minimizers with probability one. In the strongly convex case, the sequence converges strongly to the unique optimum with probability one. An application to a class of PDE constrained problems with a convex objective, convex and bounded constraint and random elliptic PDE constraints is shown. Theoretical results are demonstrated numerically. Future work includes the extension to a Banach space setting with a nonconvex objective. This is motivated by a model problem in structural topology optimization, where phase fields represent shapes and the objective is nonconvex.

We consider problems of the form

$$\min_{u\in C}\{j(u) = \mathbb{E}[J(u,\xi)]\}, \tag{1}$$

where C is a nonempty, closed and convex subset of a Hilbert space H. The random variable $\xi : \Omega \to \mathbf{R}^m$ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; it is assumed that for every ξ, $J(\cdot,\xi)$ is convex on C and L^2-differentiable on an open neighborhood of C, making $j(\cdot)$ convex and differentiable as well. Therefore

$$j(u) = \mathbb{E}[J(u,\xi)] = \int_\Omega J(u,\xi)\ \mathrm{d}\mathbb{P}(\xi) \tag{2}$$

is well-defined and finite for each $u \in C$. Typically, direct calculation of the integral (2) is not tractable. In stochastic approximation, one relies on the notion of a stochastic gradient, or a function $G(u,\xi)$ such that $\mathbb{E}[G(u,\xi)] \approx \nabla j(u)$. Iterates take the form

$$u_{n+1} = \pi_C(u_n - \tau_n G(u_n, \xi_n)),$$

where the proper choice of the step size $\tau_n \geq 0$ is essential to efficient performance. Such methods originated in a paper by Robbins and Monro in 1951 [5], where authors developed an iterative method for finding the root of a function where only noisy estimates of the function are available. A related work for finding the maximum of a regression function followed in a paper by Kiefer and Wolfowitz in 1952 [4].

Models incorporating uncertain inputs, such as random forces or material properties, have been of increasing interest in PDE constrained optimization. Gradient-based methods are already widely used in PDE constrained optimization [3], and it is the goal of our work [2] to establish the stochastic counterpart using stochastic gradients as a viable method for the solution to problems involving uncertainty. Under certain assumptions, (almost sure) strong convergence of the sequence $\{u_n\}$ can be demonstrated. Using efficiency estimates, proper step sizes can be chosen that lead to robust performance even in the case where the objective function is not strongly convex.

The approach was reviewed for a model problem involving a convex objective function with a random elliptic PDE constraints and box constraints. Convergence behavior aligning with the theory can be demonstrated. The next step will be to investigate under what conditions convergence can be shown in Banach spaces and in the nonconvex case. This is motivated by a model problem in structural topology optimization, where phase fields represent shapes and the objective function is nonconvex. Convergence of a generalized projected gradient method in Banach spaces was proven by Blank and Rupprecht [1] in the deterministic setting. Initial simulations show promising results using a single random search direction combined with a Armijo line search at each iteration.

References

[1] L. Blank, C. Rupprecht, An extension of the projected gradient method to a Banach space setting with application in structural topology optimization, *SICON* **55(3)** (2017), 1481–1499.
[2] C. Geiersbach, G. Pflug, Projected Stochastic Gradients for Convex Constrained Problems in Hilbert Spaces, *arXiv:1807.09132*, submitted (2018).
[3] M. Hinze, R. Pinnau, M. Ulbrich, Optimization with PDE Constraints, Springer (2009).
[4] J. Kiefer, J. Wolfowitz, Stochastic Estimation of the Maximum of a Regression Function, *The Annals of Mathematical Statistics* **23(3)** (1952), 462-466.
[5] H. Robbins, S. Monro, A Stochastic Approximation Method, *Ann. Math. Statist.* **22(3)** (1951), 400–407.

Solving Stochastic Equilibria: EMP, SELKIE, and Optimal Value Functions

Michael C. Ferris
(joint work with Olivier Huber and Youngdae Kim)

Equilibrium problems are used extensively to model phenomena in economics and engineering applications. Such problems usually involve collections of optimization problems and variational inequalities, solved concurrently in a Nash manner. A concrete example is a collection of multiple optimization problems coupled with an equilibrium constraint (or variational inequality), so called MOPECs:

$$\begin{array}{lll} \text{find} & (x_1^*, \ldots, x_N^*, \pi^*) & \text{satisfying,} \\ x_i^* \in & \arg\min_{x_i} & \theta_i(x_i, x_{-i}^*, \pi^*), \\ & \text{s.t.} & h_i(x_i, x_{-i}^*, \pi^*) = 0, \\ & & g_i(x_i, x_{-i}^*, \pi^*) \le 0, \\ \hline \pi^* \in & \mathrm{SOL}(K(x^*), F(\cdot; x^*)). & \end{array}$$

Here $\mathrm{SOL}(K(x^*), F(\cdot; x^*))$ represents the solution set of a quasi-variational inequality defined by the set $K(x)$ and the function $F(y, x)$. Essentially we have added to a Generalized Nash Equilibrium [3] an additional VI agent that solves a variational inequality, i.e. market clearing conditions

$$0 \le \text{supply} - \text{demand} \quad \perp \quad \pi \ge 0.$$

We have developed a framework, called Extended Mathematical Programming (EMP), that enables such problems to be effectively described in modeling systems such as GAMS, AMPL and Julia and transformed and communicated to solvers [4]. The framework allows modelers to formulate their problem naturally, with appropriate mathematical constructs, to convey known problems structures, and automate problem transformations to tractable forms for solution engines. More recent extensions [5] allow shared constraints and variational solutions [8], implicit functions and state equations to be added to problem descriptions, along with

duality tools to formulate appropriate primal-dual models. Hierarchical constructs, such as optimal value functions [1] are also part of the EMP framework.

Real problems are messy, and involve lots of different constructs that a modeler often has difficulty in conveying to a solver, and consequently these constructs are not available to the solver. This frequently leads to poorly specified implementations of a nice problem (that are much harder to solve), and an inability of solvers to detect the substructures in the formulation (that are well known to the modeler) and therefore adds significantly to the complexity of solving the underlying problem. The new framework communicates this structure, and improves solution robustness as documented for example in [7].

We also introduce Selkie [6], a general-purpose solver for equilibrium problems that exploits problem structures in a flexible and adaptable way to achieve a more robust and faster solution path. To accomplish this, it transforms a given model into a set of structure-exploiting sub-models via structure analysis and accounting for information given via EMP. Many algorithms from the literature can be implemented via options to the solver, including standard Gauss-Seidel and Jacobi methods, and decomposition methods (for variational inequalities for example [9]). New applications also become easier to implement, such as dynamic programming schemes for large scale economic analyses, e.g. [2]. Also, it is possible to exploit underlying sparsity in the models since a modeler can define intermediate variables that appear in multiple subproblems but can be treated computationally as an implicit function.

In applications of economics and engineering, it is important to allow modelers to formulate and solve stochastic equilibrium problems within modeling systems. While there are some tools to do this in existing modeling systems, we believe they have several shortcomings that preclude their use by modelers with such problems. Such an example can be found for instance in [11]. This involves the notion of a multi-stage stochastic process, over which a number of agents solve a risked stochastic program, that is linked at different time scales and in different realizations by market clearing constraints (of an equilibrium type):

$$
\begin{aligned}
\text{CP:} \quad & \min_{d^1, d^2_\omega \geq 0, t^C} && \sigma t^C + p^1 d^1 - W(d^1) + \rho_C \left[p^2_\omega d^2_\omega - W(d^2_\omega) - t^C_\omega \right] \\
\text{TP:} \quad & \min_{v^1, v^2_\omega \geq 0, t^T} && \sigma t^T + C(v^1) - p^1 v^1 + \rho_T \left[C(v^2_\omega) - p^2_\omega v^2_\omega - t^T_\omega \right] \\
\text{HP:} \quad & \min_{\substack{u^1, x^1 \geq 0 \\ u^2_\omega, x^2_\omega \geq 0, t^H}} && \sigma t^H - p^1 U(u^1) + \rho_H \left[-p^2_\omega U(u^2_\omega) - V(x^2_\omega) - t^H_\omega \right] \\
& \text{s.t.} && x^1 = x^0 - u^1 + h^1, \\
& && x^2_\omega = x^1 - u^2_\omega + h^2_\omega
\end{aligned}
$$

$$
\begin{aligned}
& 0 \leq p^1 \perp U(u^1) + v^1 \geq d^1 \\
& 0 \leq p^2_\omega \perp U(u^2_\omega) + v^2_\omega \geq d^2_\omega, \forall \omega \\
& 0 \leq \sigma_\omega \perp t^C_\omega + t^T_\omega + t^H_\omega \geq 0, \forall \omega \;\; \sigma = (\sigma_\omega)
\end{aligned}
$$

Note that this example has two stages only (so is simpler than the general case!), involving three agents (a consumer C, thermal generator T, hydro generator H) with cost, welfare and operational functions C, W, U and V, state variables d, v, u and x, risk functions ρ and trades t, coupled by prices of goods p and trades σ over scenarios indexed by ω. Note this is a concrete example of a MOPEC.

It is imperative that we build mechanisms to allow these models to be formulated appropriately in modeling systems, that incorporate all the constructs necessary for stochasticity and equilibrium definition, and enable such specified models to be transformed into tractable computational pieces that can be manipulated in new algorithms for their solution. Much of the new algorithmic work outlined in this Oberwolfach workshop will be effective for problem solving in a myriad of new applications if we can accomplish the formulation and implementation of stochastic equilibria in modeling systems.

References

[1] A. Aravkin, J. V. Burke, G. Pillonetto, *Sparse / Robust Estimation and Kalman Smoothing with Nonsmooth Log-Concave Densities: Modeling, Computation, and Theory*, Journal of Machine Learning Research, 14(1), (2013), 2689–2728.

[2] W. Chang, M. C. Ferris, Y. Kim, T. F. Rutherford, *Solving Stochastic Dynamic Programming Problems: a Mixed Complementarity Approach*, Computational Economics, submitted, (2018).

[3] F. Facchinei, C. Kanzow, *Generalized Nash Equilibrium Problems*, 4or, 5(3) (2007), 173–210.

[4] M. C. Ferris, S. P. Dirkse, J–H. Jagla, A. Meeraus, *An Extended Mathematical Programming Framework*, Computers and Chemical Engineering, 33, (2009), 1973–1982.

[5] Y. Kim, M. C. Ferris, *Solving equilibrium problems using extended mathematical programming*, Mathematical Programming C, submitted, https://arxiv.org/abs/1806.02255, (2018).

[6] Y. Kim, M. C. Ferris, M. C., *SELKIE: a model transformation and distributed solver for structured equilibrium problems*, in preparation, (2018).

[7] Y. Kim, O. Huber, M. C. Ferris *A Structure-Preserving Pivotal Method for Affine Variational Inequalities*, Mathematical Programming, 168(1) (2017), 93–121.

[8] A. A. Kulkarni, U. V. Shanbhag, *On the variational equilibrium as a refinement of the generalized Nash equilibrium*, Automatica, 48(1), (2012), 45–55.

[9] J. P. Luna, C. Sagastizábal, M. Solodov,

[10] , Mathematical Programming, 143(1-2), (2012), 177–209.

[11] A. B. Philpott, M. C. Ferris, R. J.-B. Wets, *Equilibrium, Uncertainty and Risk in Hydrothermal Electricity Systems*, Mathematical Programming B, 157(2), (2016), 483–512.

Distributionally robust sample average approximation with small sample sizes

Andy Philpott
(joint work with Eddie Anderson)

1. Introduction

Consider a stochastic optimization problem

$$\min_x \mathbb{E}_{\mathbb{P}}[c(x,y)],$$

with known cost function $c(x,y)$. Here expectations are taken over the random variable Y, with instance $y \in \mathbb{R}^m$, and probability distribution $\mathbb{P}$. Assume the decision variable x has optimal value x^*. In Sample Average Approximation (SAA) the decision maker optimizes using a sample $S = \{y_1, y_2, ..., y_N\}$ of Y. We assume that N is small and it is expensive or impossible to obtain more sample points. For example S could be a record of past outcomes for Y, e.g. a historical record of annual demand for a set of products where x is a capacity decision. Given a sample S the decision maker can approximate the value of $\mathbb{E}_{\mathbb{P}}[c(x,y)]$ by $(1/N)\sum_{i=1}^N c(x,y_i)$ and solve the problem

$$\text{SAA:} \quad \min_{x\in X}(1/N)\sum_{i=1}^N c(x,y_i).$$

We write $x^*(S)$ for the solution of SAA which depends on sample S. The finite distribution with probability $(1/N)$ for each y_i is denoted ν_S.

Distributionally robust optimization solves the problem

$$\text{DRO:} \quad \min_{x\in X} \sup_{\mathbb{Q}\in\mathcal{P}_\delta} \mathbb{E}_{\mathbb{Q}}\left[c(x,y)\right],$$

for some choice of $\mathcal{P}_\delta$ being a ball of size δ centered at ν_S. We write $x^*_{DRO}(S)$ for the optimal solution of DRO. It has been observed by many authors (see e.g. [8] [4],[6],[9],[10]) that $x^*_{DRO}(S)$ is often a better solution on average than $x^*(S)$ when evaluated out of sample. This is partly related to $x^*(S)$ overfitting to S. Here distributionally robust optimization can be viewed as a regularization of SAA.

The set $\mathcal{P}_\delta$ can be defined in several ways. A popular approach is to use a ϕ-*divergence* ([2],[3]) which confines $\mathbb{Q}$ to probability distributions over the sampled set of points $\{y_1, y_2, ..., y_N\}$. The ϕ-divergence of $\mathbb{Q} = (q_1, q_2, ..., q_N)$ relative to $\nu(S) = (1/N, 1/N, ..., 1/N)$ is defined as

$$d_\phi(\mathbb{Q}) = \sum_{i=1}^N (1/N)\phi\left(Nq_i\right) \tag{1}$$

for ϕ a convex function defined on $[0,\infty)$ with $\phi(1) = 0$ (and achieving its minimum there). This gives

$$\mathcal{P}_\delta = \{\mathbb{Q} : d_\phi(\mathbb{Q}) < \delta\}.$$

An alternative definition of $\mathcal{P}_\delta$ uses the *Wasserstein* distance defined by the minimum cost transportation plan from one probability distribution to the other. Details are discussed in [4].

2. Unbiased solutions

As observed by [6] the estimate $x^*_{DRO}(S)$ has a lower variance than $x^*(S)$ for small δ, and this reduction compensates for any extra bias in $x^*_{DRO}(S)$, so $x^*_{DRO}(S)$ performs better than $x^*(S)$ out of sample. This will become most obvious when $x^*_{DRO}(S)$ is unbiased. A simple example is provided by the estimation problem

$$\min_x \mathbb{E}_{\mathbb{P}}[(x-y)^2],$$

which has solution $x^* = \mathbb{E}[Y]$. The SAA problem

$$\text{SAA:} \quad \min_{x\in X}(1/N)\sum_{i=1}^N (x-y_i)^2).$$

has solution $x^*(S) = (1/N)\sum_{i=1}^N y_i$, the sample average. This estimate is unbiased, and is the minimum variance unbiased estimator when Y is normally distributed. For other distributions, consider solving

$$\text{DRO:} \quad \min_{x\in X} \sup_{\mathbb{Q}\in\mathcal{P}_\delta} \mathbb{E}_{\mathbb{Q}}\left[(x-y_i)^2\right],$$

where $\mathcal{P}_\delta$ is defined using (1) with $\phi(t) = |t-1|$. For example if Y has a uniform distribution and $\delta \geq 2-(4/N)$, the solution $x^*_{DRO}(S)$ can be shown to be $(1/2)(\min S + \max S)$ i.e. the solution identified by Fisher [5] and shown by Lloyd [7] to be the minimum variance linear estimator for $\mathbb{E}[Y]$ for uniformly distributed Y. This is a much better estimator than $x^*(S)$.

3. Biased solutions

If the cost function is not symmetric in y then DRO might introduce a bias in $x^*_{DRO}(S)$. This could improve out-of-sample performance, or make it worse. Consider a simple model in which the cost to produce x is $kx+x^2$, and selling price is y, so we minimize $c(x,y) = kx+x^2-xy$. Suppose y is random, where $\mu = \mathbb{E}_{\mathbb{P}}(y)$ is the average price, so $\mathbb{E}_{\mathbb{P}}[c(x,y)] = kx + x^2 - x\mu$ with minimum value of $-(\mu-k)^2/4$ attained at $x^* = (\mu-k)/2$.

Given a sample $S = \{y_1, y_2, ..., y_N\}$ with sample average $\overline{y}(S) = (1/N)\sum_{i=1}^N y_i$, the SAA solution minimizes $(1/N)\sum_{i=1}^N c(x,y_i)$ yielding $x^*(S) = (\overline{y}(S)-k)/2$. The SAA solution is an unbiased estimator for the true solution x^*. The (out-of-sample) cost of the SAA solution is

$$\begin{aligned} C_{SAA}(S) &= \mathbb{E}_{\mathbb{P}}[c(x^*(S),y)] = kx^*(S) + x^*(S)^2 - x^*(S)\mu \\ &= -\frac{(\mu-k)^2}{4} + \frac{(\mu-\overline{y}(S))^2}{4} \end{aligned}$$

giving an increase in cost of $(\mu-\overline{y}(S))^2/4$ in comparison with the correct value.

Assume $x > 0$. Suppose we solve DRO with $\mathcal{P}_\delta$ defined using (1) with $\phi(t) = |t-1|$ and small δ. The solution shifts probability $\delta/2$ from the y_i with the lowest $c(x,y)$ value (i.e. the highest price in the sample, y_M) and moves it to the y_i with the highest $c(x,y)$ value (which is the lowest price in the sample, y_L). The solution

we obtain satisfies

$$(1/N)\sum_{i=1}^{N}(k+2x-y_i)+(\delta/2)(y_M-y_L)=0$$

so

$$x^*_{DRO}(S)=\frac{(\overline{y}-k)}{2}-\frac{\delta(y_M-y_L)}{4}=x^*(S)-\frac{\delta R(S)}{4}$$

where $R(S)=y_M-y_L$. This is always less than $x^*(S)$ so $x^*_{DRO}(S)$ is biased downwards. The cost of $x^*_{DRO}(S)$ is

$$\begin{aligned}C_{DRO}(S)&=\mathbb{E}_{\mathbb{P}}[c(x^*_{DRO}(S),y)]\\&=C_{SAA}(S)-(\overline{y}(S)-\mu)R(S)\frac{\delta}{4}+R(S)^2\frac{\delta^2}{16}.\end{aligned}$$

For small δ, $\mathbb{E}[C_{DRO}(S)]<\mathbb{E}[C_{SAA}(S)]$ when

$$\mathbb{E}[(\overline{y}(S)-\mu)R(S)]>0, \tag{2}$$

where the expectation is over the sampling distribution. Random variables Y satisfying (2) can be characterized by conditions on their distributions which need to be asymmetric with a right-handed skew (see [1]).

Finally, we mention the case where $x^*(S)$ is biased. Here, using (1) with $\phi(t)=|t-1|$, we can contruct examples of situations where $x^*_{DRO}(S)$ has less bias than $x^*(S)$ and performs better on average out of sample, as well as examples of situations where $x^*_{DRO}(S)$ has more bias than $x^*(S)$ and performs worse on average out of sample. Details, derivations and extensions can be found in [1], where we also compare the above results with those using the Wasserstein metric.

References

[1] Anderson, E.J. and Philpott, A.B., Using robust optimization to improve on Sample Average Approximation. Technical report, EPOC, 2018.

[2] Bayraksan G. and Love, D.K., Data-driven stochastic programming using phi-divergences. In *The Operations Research Revolution*, chapter 1, pages 1-19. INFORMS, 2015.

[3] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G., Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341-357, 2013.

[4] Esfahani, P.M. and Kuhn, D., 2017. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, pp.1-52.

[5] Fisher, R.L., On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London A*, 22:309–368, 1922.

[6] Gotoh, J.Y., Kim, M.J. and Lim, A.E., Robust empirical optimization is almost the same as mean–variance optimization. *Operations Research Letters*, 2018.

[7] Lloyd, E.H., Least-squares estimation of location and scale parameters using order statistics, *Biometrika*, 39, 88-95, 1952.

[8] Pflug, G. and Wozabal, D., Ambiguity in portfolio selection. *Quantitative Finance* 7(4) 435-442, 2007.

[9] Van Parys, B.P., Esfahani, P.M. and Kuhn, D., From data to decisions: Distributionally robust optimization is optimal. arXiv preprint arXiv:1704.04118, 2017.

[10] Wozabal, D., 2014. Robustifying convex risk measures for linear portfolios: A nonparametric approach. Operations Research, 62(6), pp.1302-1315.

Distributionally Robust Optimization with Decision-Dependent Ambiguity Set

Nilay Noyan (Sabanci University, Turkey)
(joint work with Gábor Rudolf and Miguel Lejeune)

The classical stochastic programming literature relies on the assumption that the probability distribution of uncertain model parameters is given as a model input, often as set of scenarios along with their probabilities. However, in many decision-making applications the true parameter distribution is unknown. *Distributionally robust optimization* (DRO) is a recent and appreciated approach to hedge against such distributional uncertainty. Instead of assuming that there is a known underlying probability distribution, in DRO one considers an *ambiguity set* that consists of probability distributions, and solves a minimax-type problem to determine decisions that provide hedging against the worst-case parameter distribution in the ambiguity set.

Another common fundamental assumption in the stochastic programming literature is that the underlying probability space is independent of the decisions. In other words, it is usually assumed that the probability distributions of random model parameters are exogenously given. In the DRO setting this attitude translates to the assumption that the specified ambiguity set of distributions is decision-independent. However, in certain situations decisions can directly affect the distribution of the random parameters, either by changing the parameter realizations or by changing the probabilities of underlying random events that occur after the decisions are taken. This phenomenon is known as *endogeneous uncertainty* (see, e.g., [7, 13]). For example, in the context of pre-disaster planning, if the links of a transportation network are subject to random failure in case of a disaster, then the investment decisions on strengthening such links (seismic retrofitting of bridges/viaducts on links) can reduce the failure probabilities and improve network survivability ([9]).

In our study we aim to address both distributional and endogeneous uncertainty. In this regard, we incorporate endogenous uncertainty into distributionally robust stochastic programming problems via decision-dependent ambiguity sets. Until recently, DRO with decision-dependent ambiguity sets has been an almost untouched research area ([12, 14]).

We limit our attention to statistical distance-based ambiguity sets. These sets consist of probability distributions that are in the vicinity of a nominal distribution—often the empirical one—thought to approximate the true distribution. The vicinity is defined here as a ball centered on the nominal distribution. A wide variety of statistical distances, which provide a measure of dissimilarity between two probability distributions, have been employed to construct such balls. We refer to [4] for an elaborate discussion of the pros and cons associated with various ambiguity sets, and in particular the advantages of Wasserstein metrics (see, e.g., [3, 10]) over ϕ-divergences (see, e.g., [1]). In line with the recent sharp

increase in the use of Wasserstein distances in DRO, our focus will be on a general class of *earth mover's distances*, introduced in a discrete context by Rubner et al. [11]. Our chosen class includes both the total variation distance and the Wasserstein-1 metric (also known as the Kantorovich-Rubinstein metric), allows the construction of ambiguity sets based on higher-order Wasserstein distances, and also has desirable statistical and tractability properties.

While decision-dependent uncertainty—endogenous uncertainty—is straightforward to express in the framework of Markov decision processes, its use in stochastic programming remains a tough endeavor, and is far from being a well-resolved issue. The relevant literature primarily focuses on two types of optimization problems: problems with decision-dependent information revelation, and problems with decision-dependent probabilities. In problems of the first type, decisions can partially resolve the uncertainty, affect the timing of uncertainty resolution, and alter the set of possible future random outcomes. In problems of the second-type, decisions alter the probability measures. The first problem type has been addressed more widely (see, e.g., [5, 13]) in the literature. Accordingly, in our study, we aim to contribute to the literature by focusing on problems of the second type, where decisions can affect the likelihood of underlying random future events and/or can affect the possible realizations of the random parameters. Stochastic problems with decision-dependent probability measures are notoriously difficult to model and solve, and, not surprisingly, the relevant literature is quite sparse. Dupacova [2] briefly discusses optimization under endogenous uncertainty, without providing specific formulations or solution methods. Studies that feature algorithmic developments are relatively recent, and typically rely on additional structural properties that are specific to their problems of interest (see, e.g., [6, 8]).

Our contributions. We present a unified modeling framework for a new class of distributionally robust optimization problems under decision-dependent ambiguity sets. In particular, as our ambiguity sets we consider balls centered on a decision-dependent probability distribution. The balls are based on a class of earth mover's distances that includes both the total variation distance and the Wasserstein metrics. Our models typically give rise to non-convex non-linear programs, which are in general very hard to solve. However, we provide an overview of various settings where it is possible to obtain tractable formulations. Some of the arising side results that make these formulations possible are also of independent interest, including novel mathematical programming expressions for robustified risk measures in a discrete space. We also discuss how our results can be utilized to provide tractable formulations for a practical pre-disaster investment planning problem in humanitarian logistics.

Further avenues of research. One of the main distinguishing features of our approach is that the nominal distribution at the center of the ambiguity set is decision-dependent. In this regard, it would be essential to investigate possible ways to describe this dependence, and in particular to develop meaningful and tractable characterizations of decision-dependent nominal parameter realizations and/or scenario probabilities for practical applications.

The DRO formulations we presented mostly rely on a full (exponential sized) scenario set, which usually makes it impossible to solve problem instances of practical interest. It is of crucial importance to reduce the number of scenarios. For certain problem classes one could benefit from scenario bundling methods. However, bundling raises an important concern in DRO: Considering an EMD ball around the reduced version of the original distribution (based on the original individual scenarios) is in general not equivalent to considering the reduced versions of the distributions in the EMD ball around the original distribution. Thus, replacing the scenario distances in the original formulation with suitably defined bundle distances is highly non-trivial, and typically leads to undesired consequences, such as formulations whose optimal solution depends on the choice of bundling. Alternatively, one might consider sampling methods to reduce the number of scenarios. The use of sampling, however, comes with two important caveats. First, sampling typically sacrifices exact solutions in exchange for tractability. Second, even in cases where a particular sampling method (such as importance sampling) is known to work well for the underlying (non-DRO) problem, this does not automatically translate to a performance guarantee for the DRO variant. For example, when using an ambiguity set based on the total variation distance, the worst-case distribution will be highly sensitive to the worst scenario included in a sample. Accordingly, developing sampling approaches suited for DRO with decision-dependent ambiguity sets appears to be an open and interesting research topic.

References

[1] G. Bayraksan and D. Love, *Data-Driven Stochastic Programming using Phi-Divergences*, INFORMS TutORials in Operations Research (2015), 1–15.

[2] J. Dupacova, *Optimization under exogenous and endogenous uncertainty*, University of West Bohemia in Pilsen, Working paper (2006): http://www.karlin.mff.cuni.cz/ dupacova/papers/MME06sty.pdf

[3] P.M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*, Mathematical Programming (2018), Online first: https://link.springer.com/article/10.1007/s10107-017-1172-1.

[4] R. Gao and A.J. Kleywegt, *Distributionally Robust Stochastic Optimization with Wasserstein Distance*, Technical Report 1604.02199, ArXiv (2016), https://arxiv.org/abs/1604.02199.

[5] V. Goel and I.E. Grossmann, *A Class of stochastic programs with decision dependent uncertainty*, Mathematical Programming **108(2)** (2006), 355–394.

[6] U.-U. Haus, C. Michini, and M. Laumanns, *Scenario Aggregation using Binary Decision Diagrams for Stochastic Programs with Endogenous Uncertainty*, Working paper (2017): https://arxiv.org/pdf/1701.04055.pdf.

[7] L. Hellemo, *Managing Uncertainty in Design and Operation of Natural Gas Infrastructure*, Ph.D. thesis, Norwegian University of Science and Technology (2016).

[8] M. Laumanns, S. Prestwich, and B. Kawas, *Distribution shaping and scenario bundling for stochastic programs with endogenous uncertainty*, Working paper (2014): https://edoc.huberlin.de/bitstream/handle/18452/9095/5.pdf.

[9] S. Peeta, F.S. Salman, D. Gunnec, and K. Viswanath, *Pre-disaster investment decisions for strengthening a highway network*, Computers and Operations Research **37** (2010), 1708–1719.

[10] G.C. Pflug and D. Wozabal, *Ambiguity in portfolio selection*, Quantitative Finance **7(4)** (2007), 435–442.
[11] Y. Rubner, C. Tomasi, and L.J. Guibas, *A metric for distributions with applications to image databases*, Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271) (1998), 59–66.
[12] J. O. Royset and R.J.-B. Wets, *Variational Theory for Optimization under Stochastic Ambiguity*, SIAM Journal on Optimization **27(2)** (2017), 1118–1149.
[13] T.W. Jonsbråten, R.J-B. Wets, and D.L. Woodruff, *A class of stochastic programs with decision dependent random elements*, Annals of Operations Research **82** (1998), 83–106.
[14] J. Zhang, H. Xu, and L. Zhang, *Quantitative Stability Analysis for Distributionally Robust Optimization with Moment Constraints*, SIAM Journal on Optimization **26(3)** (2016), 1855–1882.

Ambiguity in stochastic optimization

Mathias Pohl
(joint work with Stefan Eckstein, Michael Kupper, Georg Pflug)

As a motivation we consider the portfolio selection problem in some detail. In mean-risk portfolio optimization, it is typically assumed that the returns of the assets follow a known distribution P_0, which is estimated from observed data. We aim at an investment strategy which is robust against possible misspecification of P_0. Hence, the portfolio selection problem is considered with respect to the worst-case distribution within a Wasserstein-neighborhood of P_0. The investor's trust in the *reference distribution* P_0 is captured by the size of this Wasserstein-neighborhood. We review tractable formulations of the so-called *portfolio selection problem under model ambiguity* by [1], [2] and [3]. For instance, the latter two references established that high model ambiguity leads to equally-weighted portfolio diversification.

Then, we introduce the portfolio selection problem under dependence uncertainty. This novel viewpoint is motivated by a situation in which the marginal return distributions of the assets can be estimated with high accuracy, whereas the dependence structure between the assets remains ambiguous. We therefore fix the marginal distributions and assume the model uncertainty only concerns the dependence structure. It is thus natural to define the Wasserstein-neighborhood, in which the worst case distribution has to lie, directly on the level of the dependence structure. We show that if the investment decision is based on the worst case among all kinds of dependencies, which corresponds to the case where the investor does not trust the reference dependence structure at all, then concentration of the portfolio into a single asset is optimal. Moreover, we develop an algorithm which allows us to approximately solve the portfolio selection problem for a fixed level of dependence uncertainty.

An empirical study compares the two approaches and their opposing implications.

Finally, we study stochastic optimization problems under dependence uncertainty in a more general framework. Inspired by Gao and Klewegt [4], we establish a duality result in this context and show that the derived dual problem can be

reformulated as an LP under restrictive assumptions. Proposing a new direction in stochastic programming, we show that artificial neural networks allow us to approximately solve the dual problem in a general framework. Moreover, we can also compute the primal optimizers and derive bounds.

This talk is based on [5] as well as on the working paper [6].

References

[1] Pflug, G. and Wozabal, D. *Ambiguity in portfolio selection* Quantitative Finance, 7.4: 435–442, 2017

[2] Pflug G. Ch., Pichler A. and Wozabal D. *The 1/N investment strategy is optimal under high model ambiguity* Journal of Banking & Finance, 36, 410–417, 2012

[3] Esfahani P. M. and Kuhn D. *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations* Mathematical Programming, 1–52, 2017

[4] Gao R. and Kleywegt A.J. *Data-driven robust optimization with known marginal distribution* `http://www.prism.gatech.edu/ rgao32/copula.pdf`, 2017

[5] Pflug G. Ch. and Pohl M. *A review on ambiguity in stochastic portfolio optimization* Set-Valued and Variational Analysis, 1–25, 2017

[6] Eckstein S., Kupper M. and Pohl M. *Robust risk aggregation with neural networks* In preparation, 2018

Utility Preference Robust Optimization: Piecewise Linear Approximation, Error Bounds and Qualitative Stability

Huifu Xu

(joint work with Shaoyan Guo)

Consider the following one-stage expected utility maximization problem

$$\max_{x\in X} \ \mathbb{E}_{\mathbb{P}}[u(f(x,\xi(\omega)))],$$

where $u : \mathbb{R} \to \mathbb{R}$ is a real-valued utility function, x is a decision vector restricted to take values over a specified compact set $X \subset \mathbb{R}^n$, $\xi : \Omega \to \Xi$ is a vector of random variables defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a bounded support set $\Xi \subset \mathbb{R}^k$ and f is a continuous function of x and ξ. In practice, f may represent a financial position or the performance of an engineering design. Our focus here is on the situation where a decision maker does not have complete information on the utility function u, i.e., utility preference, but it is possible to elicit partial information to construct an ambiguity set of utility functions denoted by $\mathcal{U}$ such that the true utility function which reflects precisely the decision maker's preference lies in $\mathcal{U}$ with high likelihood. Under such a circumstance, it might be sensible to consider a maximin preference robust optimization model:

$$\text{(PRO)} \qquad \vartheta := \max_{x\in X} \min_{u\in\mathcal{U}} \ \mathbb{E}_{\mathbb{P}}[u(f(x,\xi(\omega)))].$$

Let $\mathscr{U}$ denote a class of nonconstant increasing functions defined over $[a,b]$ with $u(a)=0, u(b)=1$, let

$$\mathcal{U} := \left\{ u \in \mathscr{U} : \int_a^b \psi_j(t) du(t) \le c_j, \text{for } j = 1, \cdots, m \right\}, \tag{1}$$

where $\psi_j : [a,b] \to \mathbb{R}$, $j = 1, \cdots, m$ are a class of u-integrable functions, and c_j $j = 1, \cdots, m$ are fixed constants. It is easy to verify that $\mathcal{U}$ is a convex set.

An important component of the research is to develop efficient computational schemes for solving (PRO) model. A key step that we take in this paper is to develop a piecewise linear approximation of the utility functions defined in the ambiguity set $\mathcal{U}$. We consider an approximation of the ambiguity set $\mathcal{U}$ defined in (1) by

$$\mathcal{U}_N := \left\{ u_N \in \mathscr{U}_N : \int_a^b \psi_j(t) du_N(t) \le c_j, \text{for } j = 1, \cdots, m \right\}$$

over interval $[a,b]$, where $\mathscr{U}_N$ is a class of continuous, piecewise linear functions defined over interval $[a,b]$ with kinks $\{t_1, \cdots, t_N\} \subset [a,b]$ and $t_1 = a, t_N = b$. With $\mathcal{U}_N$, we propose an approximation of (PRO) model by

$$\text{(PRO-N)} \qquad \vartheta_N := \max_{x \in X} \min_{u \in \mathcal{U}_N} \ \mathbb{E}_{\mathbb{P}}[u(f(x, \xi(\omega)))].$$

Let $\mathscr{G} := \left\{ g : [a,b] \to \mathbb{R} |\ g \text{ is measurable, } \sup_{\xi \in [a,b]} |g(t)| \le 1 \right\}$ be a set of measurable functions defined over $[a,b]$. For $u, v \in \mathcal{U}$, define the pseudo-metric between u and v by $\mathsf{dl}_{\mathscr{G}}(u,v) := \sup_{g \in \mathscr{G}} |\langle g, u \rangle - \langle g, v \rangle|$. The following theorem quantifies the difference between (PRO-N) model and (PRO) model.

Theorem 1 (Error bound on the optimal value)**.** *Assume: (a) each function $u \in \mathcal{U}$ is differentiable and Lipschitz continuous over $[a,b]$ with modulus being bounded by L, and its derivative function ∇u is also Lipschitz continuous with modulus being bounded by $\tilde{L}$; (b) there exist a positive constant α and a function $u_0 \in \mathcal{U}$ such that for $j = 1, \cdots, m$. $\int_a^b \psi_j(t) du_0(t) - c_j < -\alpha$; Then there exists a positive number N_0 such that*

$$|\vartheta_N - \vartheta| \le \left[\tilde{L}(b-a) + \frac{2}{\alpha} \left(\sqrt{m}(b-a) \sup_{j=1,\cdots,m, t \in [a,b]} |\psi_j(t)| \tilde{L} \right) \right] \beta_N$$

for all $N \ge N_0$, where $\beta_N := \max_{i=2,\cdots,N}(t_i - t_{i-1})$.

To explain the idea of qualitative robustness, let P_N and Q_N be the empirical distributions with supports generated respectively by the true probability distribution P and its perturbation Q. Note that we are supposed to observe samples generated by Q rather than P. We consider a statistical estimator $T(\cdot)$ which is based on Q_N, that is, $T(Q_N)$ is a statistical quantity we are able to obtain in practice. Our interest is whether $T(Q_N)$ is close to $T(P_N)$ under some appropriate metric for all N sufficiently large. Here $T(P_N)$ should be understood as the corresponding statistical estimator when the noise in the samples is detached. If

$T(Q_N)$ is close to $T(P_N)$, then it is safe to use $T(Q_N)$ as an estimator of $T(P)$ (because we are unable to obtain $T(P_N)$ in practice). In what follows, the particular estimator that we are looking at is the optimal value of the (PRO) problem. Specifically we consider

$$\text{(PRO-}Q_N\text{)} \qquad \vartheta(Q_N) := \max_{x\in X} \min_{u\in\mathcal{U}} \mathbb{E}_{Q_N}[u(f(x,\xi))].$$

Obviously $\vartheta(Q_N)$ is a statistical estimator of $\vartheta(Q)$ rather than the true robust optimal value $\vartheta(P)$, so we are interested in how close $\vartheta(Q_N)$ is to $\vartheta(P_N)$ (which is the statistical estimator of $\vartheta(P)$) under some metric. Let

$$S(P_N) := \arg\sup_{x\in X} \min_{u\in\mathcal{U}} \mathbb{E}_{P_N}[u(f(x,\xi))],\ S(Q_N) := \arg\sup_{x\in X} \min_{u\in\mathcal{U}} \mathbb{E}_{Q_N}[u(f(x,\xi))]$$

be the sets of optimal solutions.

Theorem 2 (Qualitative robustness of the optimal value and optimal solutions)**.** *Assume: (a) there are an exponent $\gamma > 0$ and a positive constant $C \geq 1$ such that*

$$|f(x,t)| \leq C(\|t\|^\gamma + 1), \forall (x,t) \in \mathbb{R}^n \times \mathbb{R}^k;$$

(b) the set $\mathcal{M} \subset \mathcal{M}_{k,\kappa}^{\phi^p}$, where $\mathcal{M}_{k,\kappa}^{\phi^p}$ is the class of all $P \in \mathscr{P}(\mathbb{R}^k)$ such that

$$\int_{\mathbb{R}^k} \phi(t)^p P(dt) \leq \kappa, \qquad \text{for } \kappa \geq 0 \text{ and } p > 1,$$

with $\phi(t) := \max(\sup_{u\in\mathcal{U}} u(C(\|t\|^\gamma + 1)), \sup_{u\in\mathcal{U}} -u(-C(\|t\|^\gamma + 1)))$; (c) the set of functions $\{u(f(x,\cdot)) : \mathbb{R}^k \to \mathbb{R} | x \in X, u \in \mathcal{U}\}$ are equi-continuous, that is, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for every g in the set, $|g(\xi') - g(\xi'')| \leq \epsilon$ when $\|\xi' - \xi''\| \leq \delta$; (d) $f(\cdot,\xi)$ is uniformly Lipschitz continuous in x, that is, there exists a positive constant L such that $|f(x',\xi) - f(x'',\xi)| \leq L\|x' - x''\|$ for all $x', x'' \in X$ and $\xi \in \Xi$. Then the following assertions hold.

(i) *For any small number $\epsilon > 0$, there exist positive numbers $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$*

$$Q \in \mathcal{M}, \mathsf{dl}_\phi(P,Q) \leq \delta \Longrightarrow \mathsf{dl}_{Proh}\left(P^{\otimes N} \circ \vartheta(P_N)^{-1}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}\right) \leq \epsilon.$$

(ii) *If, in addition, for each $Q \in \mathcal{M}$, the sets of optimal solutions $S(Q_N)$ and $S(Q)$ are singleton, then for any small number $\epsilon > 0$, there exist positive numbers $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$*

$$Q \in \mathcal{M}, \mathsf{dl}_\phi(P,Q) \leq \delta \Longrightarrow \mathsf{dl}_{Proh}\left(P^{\otimes N} \circ S(P_N)^{-1}, Q^{\otimes N} \circ S(Q_N)^{-1}\right) \leq \epsilon.$$

$P^{\otimes N}$ denotes a probability measure on the measurable space $\left(\Xi^{\otimes N}, \mathcal{B}(\mathbb{R}^k)^{\otimes N}\right)$ with marginal P on each $(\Xi, \mathcal{B}(\mathbb{R}^k))$ and $Q^{\otimes N}$ with marginal Q, where $\Xi^{\otimes N}$ denotes the Cartesian product $\Xi \otimes \cdots \otimes \Xi$ and $\mathcal{B}(\mathbb{R}^k)^{\otimes N}$ its Borel sigma algebra. Let $\mathcal{M}_k^\phi := \{P \in \mathscr{P}(\mathbb{R}^k) : \int_{\mathbb{R}^k} \phi(t)P(dt) < \infty\}$. The distance $\mathsf{dl}_\phi : \mathcal{M}_k^\phi \times \mathcal{M}_k^\phi \to \mathbb{R}$ is defined by

$$\mathsf{dl}_\phi(P', P'') := \mathsf{dl}_{Proh}(P', P'') + \left| \int_{\mathbb{R}^k} \phi(t)P'(dt) - \int_{\mathbb{R}^k} \phi(t)P''(dt) \right|,$$

for $P', P'' \in \mathcal{M}_k^\phi$ *where* $\mathsf{dl}_{Proh} : \mathscr{P}(\mathbb{R}^k) \times \mathscr{P}(\mathbb{R}^k) \to \mathbb{R}_+$ *is the Prohorov metric defined as follows:*

$$\mathsf{dl}_{Proh}(P', P'') := \inf\{\epsilon > 0 : P'(A) \leq P''(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(\mathbb{R}^k)\},$$

where $A^\epsilon := A + B_\epsilon(0)$ *denotes the Minkowski sum of* A *and the open ball centred at* 0 *(w.r.t. Euclidean norm).*

The research opens a new direction for robust optimization. An interesting further development will be to extend it to risk management particularly in the multi-stage setting, it will also be helpful to link the new robust framework to distributionally robust optimization and stochastic dominance.

References

[1] B. Armbruster and E. Delage, *Decision making under uncertainty when preference information is incomplete*, Management Science *61* (2015), 111-128.

[2] R. Cont, R. Deguest and G. Scandolo, *Robustness and sensitivity analysis of risk measurement procedures*, Quantitative Finance *10* (2010): 593–606.

[3] S. Guo and H. Xu, *Utility preference robust optimization: piecewise linear approximation and statistical robustness*, manuscript (2018).

[4] S. Guo and H. Xu, *Statistical robustness on utility preference robust optimization*, manuscript (2018).

[5] J. Hu and S. Mehrotra, *Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization*, IIE Transaction *47* (2015), 358–372.

[6] V. Krätschmer, A. Schied and H. Zähle, *Qualitative and infinitesimal robustness of tail-dependent statistical functionals*, Journal of Multivariate Analysis *103* (2012) 35-47.

Perspectives on PDE-Constrained Optimization under Uncertainty

Thomas M. Surowiec

(joint work with Drew P. Kouri, Sandia National Laboratories)

A diverse array of science and engineering applications necessitate the solution of optimization problems constrained by partial differential equations (PDE) with uncertain or random inputs. As in all traditional stochastic programming models, it is crucial to find solutions that are in some sense robust to catastrophic events.

The purpose of this talk is two-fold. First, we motivate the standard practices in modern PDE-constrained optimization as in, e.g., [1]. This includes a brief discussion of the necessary analytical tools and algorithmic considerations along with their implications for numerics. Second, we present an abstract class of stochastic PDE-constrained optimization problems that we now largely understand from a theoretical perspective. The subsequent results can be found in [8, 7, 5, 2].

In what follows, $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space; $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ ($p \in [1, \infty]$) is the space of p-integrable random variables on Ω; $D \subset \mathbb{R}^n$, $n \in \mathbb{N}$ is the physical domain for the PDE, which is open, bounded and has a Lipschitz boundary ∂D; the space Z of decision variables is a Hilbert space and is deterministic; $Z_{\text{ad}} \subseteq Z$ is the set of feasible decisions; and the solution space for the deterministic PDE model $U = U(D)$, is a reflexive Banach space.

Given some $z \in Z_{\text{ad}}$, we represent the PDE constraint by the abstract nonlinear equation: For $\mathbb{P}$-a.e. $\omega \in \Omega$, find $u(\omega) \in U$ such that

$$e(u(\omega), z, \omega) = 0. \tag{1}$$

Here, we set $u(\omega) = [S(z)](\omega)$ and refer to the latter as the solution mapping and sometimes the control-to-state mapping. In the absence of uncertainty, the usual workflow requires the study of the well-posedness of (1) as well as higher regularity properties of the PDE solution $u : D \to \mathbb{R}$. In addition, we typically require continuity (e.g., compactness) and differentiability properties of S as a function of z. These requirements are augmented in the stochastic setting by the need to establish measurability and uniform integrability properties of $S(z) : \Omega \to U$. This can be challenging even for semilinear elliptic PDEs with uncertain inputs, see [2].

As in the deterministic setting, we aim to optimize an objective function F that depends on $u \in U$ and $z \in Z_{\text{ad}}$. In addition to the constraints defined in Z_{ad}, one can consider constraints on the PDE solution u such as $u \geq 0$. The treatment of these state constraints in the stochastic context is currently an open topic. One major difficulty is the lack of regularity of $S(z) : \Omega \to U$, e.g., continuity, which is needed to verify constraint qualifications. In the deterministic setting, one obtains continuity by appealing to classical regularity theory for PDEs. This can only be done on a case-by-case basis for the dependencies on ω in the stochastic setting. Regardless of the presence of control or state constraints, the fact that u is a random field means that we cannot directly minimize F as a function of u and z. For this we appeal to the wide variety of models found in stochastic programming.

Of the many possibilities offered by stochastic programming, e.g., stochastic order constraints, distributionally robust optimization, and optimization of risk measures, we choose the latter. This allows us to partially leverage existing results from the deterministic case for existence of solutions, optimality conditions, and the development of function-space-based numerical methods. Moreover, there is a rich theory of risk measures that provides a diversity of modeling possibilities. We therefore consider problems of the type:

$$\min_{z \in Z_{\text{ad}}} \mathcal{R}[J(S(z))] + \wp(z). \tag{2}$$

Here, $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is a risk measure that assigns a numerical value to the random objective $J(S(z))$ where $J : L^q(\Omega, \mathcal{F}, \mathbb{P}; U) \to \mathcal{X}$, $q \in [1, \infty]$, and $\wp : Z \to \mathbb{R}$ is the cost of the decision z. Extensions to other stochastic programming models are interesting directions of future research.

Though many possibilities for $\mathcal{R}$ exist, we primarily focus on coherent risk measures, as defined in [3]. In this setting, $\mathcal{R}$ is a positively homogeneous and subadditive functional that respects the usual partial ordering on $\mathcal{X}$ along with so-called "translation equivariance": $\mathcal{R}[X + C] = \mathcal{R}[X] + C$ for any $X \in \mathcal{X}$ and $C \in \mathbb{R}$. A related class of risk measures that works well in our framework are the convex risk measures, cf. [4], where positive homogeneity and subadditivity are replaced by convexity.

The main difficulties in analyzing (2) are due to the potential nonsmoothness of $\mathcal{R}$ and the differentiability of the nonlinear composition $(J \circ S)$. We provide sufficient conditions, based on the theory of superposition operators, for treating these difficulties in [5]. In particular, the main difficulty in proving existence of solutions to (2) is to prove weak lower semicontinuity of the composition $(\mathcal{R} \circ J \circ S)$; whereas the derivation of optimality conditions hinge on the subdifferentiability of $\mathcal{R}$ and the Fréchet differentiability of $(J \circ S)$. These issues are further exacerbated when the governing PDEs have multiple solutions, i.e., S is set valued.

Letting $F := (J \circ S)$, one can show that (2) does in fact possess an optimal solution $z^\star$ under appropriate regularity conditions [5]. Moreover, for any optimal solution $z^\star$ to (2), there exists $\vartheta^\star \in \mathcal{X}^*$ such that

$$\langle \mathbb{E}[\vartheta^\star \, \nabla F(z^\star)], z - z^\star \rangle + \wp'(z^\star; z - z^\star) \geq 0, \ \forall \, z \in Z_{\mathrm{ad}} \tag{3a}$$

$$\mathcal{R}[X] - \mathcal{R}[F(z^\star)] \geq \mathbb{E}[\vartheta^\star (X - F(z^\star))], \ \forall \, X \in \mathcal{X}. \tag{3b}$$

We can partially reinterpret this system using the notion of an adjoint state $\lambda^\star$ as is done in the deterministic case. Here, we have

$$\nabla F(z^\star) = e_z(u^\star, z^\star, \cdot)^* \lambda^\star,$$

where $u^\star$ and $\lambda^\star$ solve the coupled system

$$e(u^\star, z^\star, \cdot) = 0 \ \text{ a.s.} \tag{4a}$$

$$e_u(u^\star, z^\star, \cdot)^* \lambda^\star + J_u(u^\star) = 0 \ \text{ a.s.} \tag{4b}$$

The terms e_u, e_z, and J_u denote partial derivatives of e and J.

Finally, we address the numerical solution of (2). The nonlinear couplings in (3) make the application of direct solvers, e.g., semismooth Newton, rather complicated in general. Hence, fast solvers exploiting second-order information are still an open topic. In some cases, e.g., for $\mathcal{R} = \mathbb{E}$ and $Z_{\mathrm{ad}} = Z$, one possible solution method is presented in [6]. This is a trust-region algorithm that makes use of inexact gradient and objective information along with an adaptive sparse grid strategy for sampling the uncertainty. Although this method was developed for the risk-neutral case, it can be extended for some smooth $\mathcal{R}$.

In our recent work, we have had significant success with globalized derivative-based solvers for smooth $\mathcal{R}$ using sample average approximation for the uncertainties. In contrast, we have observed generally poor performance of subgradient and bundle methods for nonsmooth risk measures such as mean-plus-AVaR, cf. the discussions in [5, 2]. As a consequence, we developed a comprehensive variational smoothing technique in [7] that can be used in conjunction with globalized derivative-based solvers by employing a continuation approach. We briefly discuss our smoothing technique here.

Given a proper, closed and convex functional $\Phi : \mathcal{X} \to \overline{\mathbb{R}}$ and a proper, closed, and convex risk measure $\mathcal{R}$, we define the epi-regularized risk measure $\mathcal{R}_\varepsilon^\Phi$ by

$$\mathcal{R}_\varepsilon^\Phi[X] := \inf_{Y \in \mathcal{X}} \left\{ \mathcal{R}[X - Y] + \varepsilon \Phi\left[\varepsilon^{-1} Y\right] \right\}.$$

Under suitable assumptions, cf. [7], one can show that $\mathcal{R}_\varepsilon^\Phi$ enjoys a number of favorable properties including pointwise error bounds, Mosco convergence to $\mathcal{R}$ and consistency of minimizers and stationary points as $\varepsilon \downarrow 0$. Furthermore, under additional properties, given a solution $z_\varepsilon^\star$ for the epi-regularized problem and $z^\star$ for the original problem, we have

$$\|z_\varepsilon^\star - z^\star\| = O(\sqrt{\varepsilon}). \tag{5}$$

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

[1] Hinze, M. and Pinnau, R. and Ulbrich, M. and Ulbrich, S., *Optimization with PDE constraints*, Mathematical Modelling: Theory and Applications **23** (2009), xii+270.
[2] D. P. Kouri and T. M. Surowiec, *Risk-Averse Optimal Control of Semilinear Elliptic PDEs*, submitted (2018).
[3] Ph. Artzner and F. Delbaen and J.-M. Eber and D. Heath, *Coherent measures of risk*, Math. Finance **3** (1999), 203–228.
[4] H. Föllmer and A. Schied, *Convex measures of risk and trading constraints*, Finance Stoch. **4** (2002), 429–447.
[5] D. P. Kouri and T. M. Surowiec, *Existence and Optimality Conditions for Risk-Averse PDE-Constrained Optimization*, SIAM/ASA Journal on Uncertainty Quantification **7** (2018), 787-815.
[6] Kouri, D. P. and Heinkenschloss, M. and Ridzal, D. and van Bloemen Waanders, B. G., *Inexact Objective Function Evaluations in a Trust-Region Algorithm for PDE-Constrained Optimization Under Uncertainty*, SIAM Journal on Scientific Computing **6** (2014), A3011-A3029.
[7] D. P. Kouri and T. M. Surowiec, *Epi-Regularization of Risk Measures*, submitted (2018).
[8] D. P. Kouri and T. M. Surowiec, *Risk-Averse PDE-Constrained Optimization Using the Conditional Value-At-Risk*, SIAM Journal on Optimization **26** (2016), 365-396.

Reporter: Matthias Claus and Mathias Pohl

Participants

Prof. Dr. Shabbir Ahmed
School of Industrial and Systems Engineering
Georgia Institute of Technology
765 Ferst Drive
Atlanta GA 30332-0205
UNITED STATES

Johanna Burtscheidt
Fakultät für Mathematik
Universität Duisburg-Essen
45117 Essen
GERMANY

Dr. Matthias Claus
Fakultät für Mathematik
Universität Duisburg-Essen
45117 Essen
GERMANY

Prof. Dr. Giorgio Consigli
Institute of Mathematics, Statistics and Computer Science
University of Bergamo
Via dei Caniana 2
24127 Bergamo (BG)
ITALY

Prof. Dr. Jesús A. De Loera
Department of Mathematics
University of California, Davis
1, Shields Avenue
Davis, CA 95616-8633
UNITED STATES

Prof. Dr. Stephan Dempe
Fakultät für Mathematik und Informatik
Institut für Numerische Mathematik und Optimierung
Technische Universität Bergakademie Freiberg
09596 Freiberg
GERMANY

Prof. Dr. Darinka Dentcheva
Department of Mathematics
Stevens Institute of Technology
Castle Point Station
Hoboken, NJ 07030
UNITED STATES

Dr. Julio Alejandro Deride Silva
Discrete Mathematics and Optimization
Sandia National Laboratories
Albuquerque, NM 85171
UNITED STATES

Prof. Dr. Laureano F. Escudero
Departamento de Matematica Aplicada
Universidad Rey Juan Carlos
Calle Tulipan
28933 Mostoles, Madrid
SPAIN

Prof. Dr. Eugene A. Feinberg
Department of Applied Mathematics
State University of New York at Stony Brook
Stony Brook, NY 11794-3600
UNITED STATES

Prof. Dr. Michael Ferris
Computer Sciences Department
University of Wisconsin-Madison
1210 West Dayton Street
Madison, WI 53706-1685
UNITED STATES

Caroline Geiersbach
Department of Statistics and Operations Research
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
AUSTRIA

Prof. Dr. Martin Gugat
Institut für Angewandte Mathematik
Universität Erlangen
Martensstrasse 3
91058 Erlangen
GERMANY

Prof. Dr. Akshay Gupte
Department of Mathematical Sciences
Clemson University
Clemson, SC 29634-0975
UNITED STATES

Dr. Jamie Haddock
Department of Mathematics
University of California at Los Angeles
MSB 7354
520, Portola Plaza
P.O. Box 951555
Los Angeles, CA 90095-1555
UNITED STATES

Helmut Harbrecht
Departement Mathematik und
Informatik
Universität Basel
Spiegelgasse 1
4051 Basel
SWITZERLAND

Prof. Dr. Matthias Heinkenschloss
Department of Computational Sciences
and Applied Mathematics
Rice University
MS-134
6100 S. Main Street
Houston, TX 77005-1892
UNITED STATES

Prof. Dr. Christoph Helmberg
Fakultät für Mathematik
Technische Universität Chemnitz
09107 Chemnitz
GERMANY

Prof. Dr. Rene Henrion
Weierstrass-Institute for Applied
Analysis and Stochastics
Mohrenstrasse 39
10117 Berlin
GERMANY

Prof. Dr. Michael Hintermüller
Institut für Mathematik
Humboldt-Universität zu Berlin
10099 Berlin
GERMANY

Ihor Indyk
Department of Mathematics
Stevens Institute of Technology
Castle Point Station
Hoboken, NJ 07030
UNITED STATES

Prof. Dr. Anna Jaśkiewicz
Wydziat Matematyki
Politechnika Wroclawska
Wybrzeze Wyspianskiego 27
50-370 Wrocław
POLAND

Prof. Dr. Alejandro Jofré
Centro de Modelamiento Matemático
Universidad de Chile
Beauchef 851
CP 8370456
Blanco Encalada 2120
Santiago de Chile
CHILE

Prof. Dr. Matthias Koeppe
Department of Mathematics
University of California, Davis
1, Shields Avenue
Davis, CA 95616-8633
UNITED STATES

Prof. Dr. Ekaterina A. Kostina
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
GERMANY

Prof. Dr. Daniel Kuhn
EPFL CDM MTE-GE
ODY 1 01.2 (Odyssea)
Station 5
1015 Lausanne
SWITZERLAND

Prof. Dr. Miguel Lejeune
Decision Sciences
The George Washington University
Office 406, Funger Hall
2201 G Street, NW
Washington, DC 20052
UNITED STATES

Yang Lin
Department of Mathematics
Stevens Institute of Technology
Castle Point Station
Hoboken, NJ 07030
UNITED STATES

Prof. Dr. Shu Lu
Department of Statistics and Operations Research
University of North Carolina, Chapel Hill
355 Hanes Hall
Chapel Hill, NC 27599-3175
UNITED STATES

Prof. Dr. Jim Luedtke
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
3236 Mechanical Engineering Building
1513 University Avenue
Madison, WI 53706
UNITED STATES

Dr. Anna Ma
Department of Mathematics
University of California, San Diego
9500 Gilman Drive
San Diego, CA 92093
UNITED STATES

Prof. Dr. Francesca Maggioni
Institute of Mathematics, Statistics and Computer Science
University of Bergamo
Via dei Caniana 2
24127 Bergamo (BG)
ITALY

Prof. Dr. Nilay Noyan
Industrial Engineering Program
Faculty of Engineering and Natural Sciences
Sabanci University
Orta Mahalle, Orhanli
34956 Tuzla, Istanbul
TURKEY

Prof. Dr. Pablo A. Parrilo
Department of Electrical Engineering and Computer Science
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307
UNITED STATES

Prof. Dr. Georg Pflug
Institut für Statistik und OR
Universität Wien
Oskar-Morgenstern Platz 1
1090 Wien
AUSTRIA

Prof. Dr. Andy Philpott
The University of Auckland
Uniservices House
Building 439, Level 4, Room 415
70 Symonds Street
Auckland 1010
NEW ZEALAND

Prof. Dr. Alois Pichler
Fakultät für Mathematik
Technische Universität Chemnitz
09009 Chemnitz
GERMANY

Mathias Pohl
Department of Statistics and Operations Research
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
AUSTRIA

Prof. Dr. Daniel Ralph
The Judge Institute of Management
University of Cambridge
Trumpington Street
Cambridge CB2 1AG
UNITED KINGDOM

Dr. Ward Romeijnders
Mathematisch Instituut
Rijksuniversiteit Groningen
Postbus 800
9700 AV Groningen
NETHERLANDS

Prof. Dr. Werner Römisch
Fachbereich Mathematik
Humboldt Universität Berlin
10099 Berlin
GERMANY

Prof. Dr. Johannes Royset
Operations Research Department
Naval Postgraduate School
Glasgow Hall 212
Monterey CA 93943
UNITED STATES

Prof. Dr. Martin Rumpf
Institut für Numerische Simulation
Universität Bonn
Endenicher Allee 60
53115 Bonn
GERMANY

Prof. Dr. Andrzej Ruszczynski
Department of Management Science and Information Systems
Rutgers University
94 Rockafeller Road
Piscataway, NJ 08854
UNITED STATES

Prof. Dr. Ekkehard Sachs
Abteilung Mathematik
Fachbereich IV
Universität Trier
54286 Trier
GERMANY

Prof. Dr. Claudia Sagastizábal
Instituto de Matematica Pura e Aplicada (IMPA)
Jardim Botânico
Estrada Dona Castorina, 110
22460 Rio de Janeiro, RJ 320
BRAZIL

Prof. Dr. Claudia Schillings
Fakultät für Mathematik und Informatik
Universität Mannheim
68131 Mannheim
GERMANY

Prof. Dr. Rüdiger Schultz
Fachbereich Mathematik
Universität Duisburg-Essen
Thea-Leymann-Strasse 9
45127 Essen
GERMANY

Prof. Dr. Suvrajeet Sen
Epstein Department of Industrial and Systems Engineering
Data Driven Decisions Laboratory
University of Southern California
3650 McClintock Avenue
Los Angeles CA 90089-2565
UNITED STATES

Prof. Dr. Mikhail V. Solodov
Instituto de Matematica Pura e Aplicada (IMPA)
Jardim Botânico
Estrada Dona Castorina, 110
22460 Rio de Janeiro, RJ 320
BRAZIL

Kai Arne Spürkel
Fakultät für Mathematik
Universität Duisburg-Essen
Thea-Leymann-Straße 9
45127 Essen
GERMANY

Prof. Dr. Thomas M. Surowiec
Fachbereich Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Straße 6
35043 Marburg
GERMANY

Prof. Dr. Bernhard von Stengel
Department of Mathematics
The London School of Economics and Political Science
10, Houghton Street
London WC2A 2AE
UNITED KINGDOM

Prof. Dr. Huifu Xu
School of Mathematics
University of Southampton
Highfield Campus
Southampton SO17 1BJ
UNITED KINGDOM