MACSER Activity Report 2021

October 2021, rev. February 2022

Contents

1 Introduction

The MACSER: Multifaceted Mathematics for Rare, High Impact Events in Complex Energy and Environment Systems, was started in September 2017. At inception it involved 24 principal investigators and had plans for 18 junior personnel: junior scientists, postdoctoral fellows, and graduate students. The number was reduced to 22 following the 2017 budget negotiations and it currently has 19 active principal investigators.

In 2020-2021 (one year of performance) MACSER pursued high quality mathematical research in nine different investigation areas (§A). It aimed to impact mathematical research, its community, as well as the domain area of complex energy and environmental systems. This was achieved by research contributions (§2), technical leadership, (§3), and dissemination of our findings and ideas through multiple channels and forums (§3).

To answer the organizational challenges, MACSER adjusts focus based on interactions of the director with the site leads and a technical committee. Some of the facts that guide the planning are the reviewer feedback, the evaluation of research progress, and the mathematical opportunities presented by an increased emphasis in the application areas on certain directions.

In this report, we present technical accomplishments, their impact on the mathematical and domain areas, and the impact of the MACSER researchers on the scientific community during the performance period.

2 Technical Progress

2.1 Space Time Data Analysis

In this cycle we developed new high/low tail nonstationarity distributions to model both marginal bulk and extremes, machine learning inference methods for intractable likelihoods, efficient change point detection methods for regression problems, new spectral estimation methods, and optimal experiment design strategies for data acquisition with correlated observations.

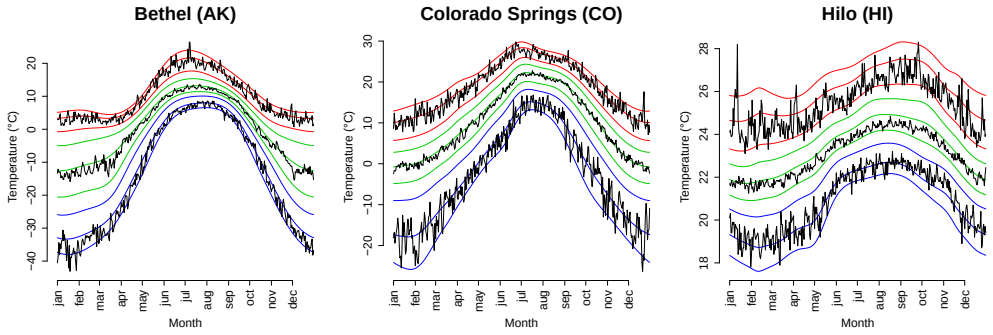


Figure 1: Flexible parametric quantile estimates for the year 2020: 0.001, 0.01, 0.1 (blue), 0.25, 0.5, 0.75 (green), and 0.9, 0.99, 0.999 (red). Black lines show the minimum/median/maximum observation for that day of year, taken over all years.

Parametric distributions and non-stationary temperature modeling. We have developed extensions of the flexible parametric model from last year [46] to take account of seasonal effects and climate change in daily temperature at locations in the United States with long temperature records [24] (see Figure 1 for distribution fits at 3 of the locations). Comparisons to empirical distributions, other parametric models, and fitted models to the tails of the distributions using generalized Pareto distributions show this approach provides a general methodology to accurately fitting the bulk and tails of a broad range of temperature distributions with widely varying seasonal characteristics and responses to climate change.

Variability of temperature and its extremes impact the energy generation and demand, as well other as infrastructures, and might arise from moderately high/cold temperature over a large spatio-temporal extent. To address the asymmetry on classical data usage when fitting tails and extremes separately, we build [24] on our recently proposed parametric model [45] by accounting for seasonality, long-term trends, and the interaction between these two characteristics. This model shows better performance when compared to several benchmark models that are used in extreme value analysis of temperature and allows simultaneous fitting *of all quantiles*, Figure 2. We have also started to extend the model in an hourly fashion accounting for daily cycles and their interaction with both seasonal and long-term trends as well as to spatial dependency.

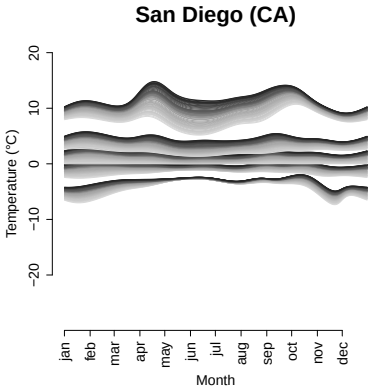


Figure 2: Estimated quantiles changing from the beginning of the observation period (light) to the end (dark). Curves are the 0.001, 0.1, 0.5, 0.9, and 0.999 quantiles.

Statistically and computationally efficient change point localization in regression. We studied multiple change-point localization in the high-dimensional regression setting, which is particularly challenging as no direct observations of the parameter of interest is available. Specifically, we assume we observe $\{x_t, y_t\}_{t=1}^n$ where $\{x_t\}_{t=1}^n$ are p-dimensional covariates, $\{y_t\}_{t=1}^n$ are the univariate responses satisfying $\mathbb{E}(y_t) = x_t^\top \beta_t^*$ for $1 \leq t \leq n$ and $\{\beta_t^*\}_{t=1}^n$ are the unobserved regression coefficients that change over time in a piecewise constant manner. We propose a novel projection-based algorithm, Variance Projected Wild Binary Segmentation (VPWBS), which transforms the original (difficult) problem of change-point detection in p-dimensional regression to a simpler problem of change-point detection in mean of a one-dimensional time series. VPWBS is shown to achieve sharp localization rate $O_p(1/n)$ up to a log factor, a significant improvement from the best rate $O_p(1/\sqrt{n})$ known in the existing literature for multiple change-point localization in high-dimensional regression. Numerical experiments demonstrate the robust and favorable performance of VPWBS over two state-of-the-art algorithms, especially when the size of change in the regression coefficients $\{\beta_t^*\}_{t=1}^n$ is small.

Vertical wind speed in the stable boundary layer. We developed a novel method for spectral estimation when data is missing and applied it to demonstrate the ability to identify unique features of the surface winds. Using the methods developed in [17] to downweight periods in which there is missing data, we were able to utilize data sequences long enough to be able to show both the diurnal and night time energy cascades of the boundary layer wind speeds (and, consequently, with much higher accuracy), Figure 3. The data in that figure are of vertical wind speeds collected via Doppler LIDAR at two heights in the atmosphere for a duration of one hour near 6am Oklahoma time. We also developed novel spectral methods for modeling high-frequency space-time data and applied these methods to Doppler LIDAR vertical wind velocity measurements [16] demonstrating again vastly improved accuracy. This paper won an honorable mention in 2021 in Student Paper Competition of the Section on Statistics and the Environment of the American Statistical Association. Moreover, we extended the `Multitaper.jl` package [18], for spectral estimation to time series with unequal temporal observing cadence and two and three dimensional data.

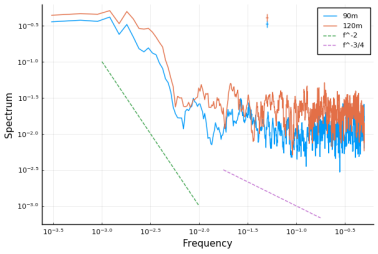


Figure 3: Spectra showing the characteristic double energy cascade of vertical wind speed during the nighttime stable boundary layer. The band in frequency between the two energy cascades can be permeated by gravity waves excited by different phenomena (7% missing observations) [17].

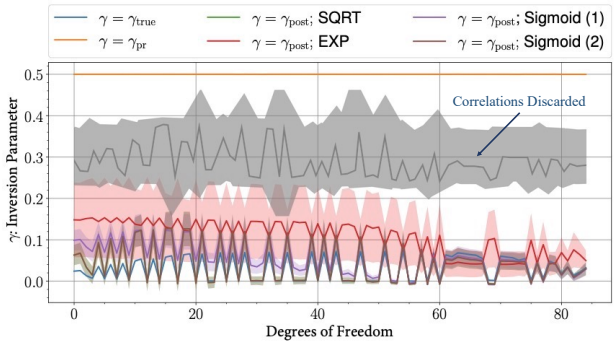


Figure 4: Solution of the inverse problem ($\pm 2\sigma_{\text{post}}$) obtained based on the A-optimal design from solving the OED problem. The number of candidate observational sensors/locations is 43 with a budget of only 8, resulting in $+10^8$ candidate observational configurations making it impossible to tune the observational configurations by brute force, and mandates efficient and accurate OED approaches. Not accounting for correlation errors leads to poor results by a factor of 4.

Optimal experimental design and data assimilation. PDE driven experimental design typically treat observational errors as being uncorrelated. In our case however, we cannot make such assumptions in the case of LIDAR instruments and discarding such evident data correlations leads to biased results, poor data collection decisions, and waste of valuable resources. We have developed new theory that allows for space-time correlations and extends current state-of-the art in the design of experiments [3]. Our framework follows a probabilistic Bayesian approach and inherently promotes a binary design, thus dramatically saving acquisition cost and and/or resources. Results (in Figure 4), shown for multiple choices of weighting kernel, demonstrate the value of our approach, compared to the case with discarded correlations, insofar solution accuracy.

2.2 Rare Events

In this cycle we have developed spatiotemporal statistical and deep learning-based models for extreme events, which perform better than the state of the art. Furthermore, we have developed perturbative methods for quantifying the failure rates of transmission networks driven by random fluctuations in generation and load. Finally, we propose a new method for estimating worst-case trajectories in transient dynamics studies.

Spatiotemporal models of extreme frequency events. In [25], we proposed a spatiotemporal statistical model for detecting early frequency disturbances and for quantifying the risk of frequency excursion. Spatial information is accounted for either as neighboring measurements in the form of covariates or with a spatiotemporal correlation model captured by a latent Gaussian field. A Bayesian decision framework is used to determine the boundary of the decision sets for under-frequency detection in real-time. Results show that the classical, purely temporal detection model has a consistently lower chance of catching a drop in frequency than our approach, suggesting the former produces higher costs for the system operator.

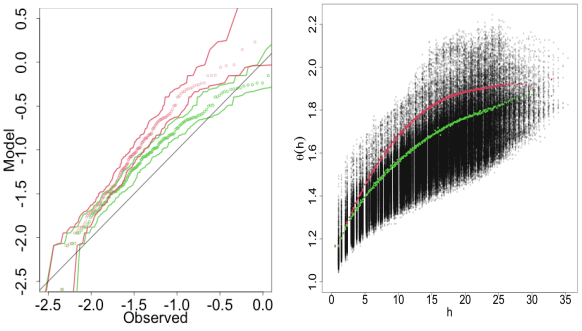


Figure 5: (Left) Observed vs predicted temperature minima with 95% confidence intervals estimated using pairwise likelihood (red) and our approach (green). (Right) F-madogram (pair-wise dependence of extremes) estimates for the validation datasets (black points) and estimated extremal coefficient functions. Pairwise likelihood underestimates the spatial dependence of extremes, while our method captures pairs of locations with high extremal dependencies.

We developed a deep neural network (DNN)-based method for estimating parameters of high dimensional max-stable processes [26], which are commonly used for modeling spatiotemporal extremes. While processes are characterized by computationally intractable full likelihoods even in moderate dimensions, our proposed DNN method allows us to accurately estimate parameters in high-dimensional settings at reduced computational cost compared to ap-

proximate likelihood approaches. Our method is also better at capturing the spatial dependence in extremes than traditional methods (Figure 5).

We developed new methods for fitting models to the tails of distributions that allow smooth downweighting of observations as they become less extreme, thus avoiding the standard approach of using sharp, arbitrary cutoffs [47]. Our methods exploits the fact that the order statistics of independent and identically distributed observations form a Markov chain, leading to a natural weighted composite likelihood for estimating models for the tails of a distribution. Furthermore, we have extended this approach to situations in which observations are not independent and identically distributed, and have demonstrate the efficacy of the proposed methods through simulations and an application to over 150 years of New York City daily rainfall data.

Failure rate analysis of transmission networks. Building on our work from previous cycles, [42], we propose a perturbative method for quantifying the failure rate of transmission network components for networks driven by random fluctuations in generation and load. We have incorporated transmission losses via a perturbative approximation of the minimum action problem in which we rewrite the deterministic flow as the sum of a lossless port-Hamiltonian flow plus a lossy contribution [5], and obtain a perturbative approximation of the F-W quasipotential in terms of the strength of the losses, the lossless F-W quasipotential, and the lossy flow. Finally, we employ this approximation to solve the minimum action problem via CasADi [1]. The computational cost of our approach is significantly smaller compared to the cost of traditional minimum action methods.

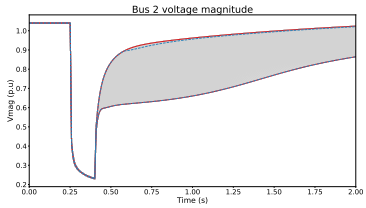


Figure 6: In red, the extreme trajectories of the voltage magnitude computed with the trust-region method. In grey, trajectories obtained from sampling the interval of a parameter using a grid with $1,000$ points.

Extreme trajectories for power dynamics. The rapid penetration of new renewable technology has resulted in lumped parameter models of distribution network with far more uncertainty and thus difficulty in assessing their effect on the system stability. We propose a new method of obtaining the extreme (worst-case) trajectories of the differential algebraic equations that model the power system in transient dynamics studies [29]. This problem is formulated as an optimization-based approach for computing the extreme trajectories dynamic systems. We also show how to appropriately derive the initial conditions for the first- and second-order sensitivity systems, which is critical for quantifying the uncertainty of the initial conditions. Results are shown in Figure 6.

2.3 Optimization Under Uncertainty

We have continued investigating new models, algorithms, and applications in the optimization under uncertainty theme. Modeling advances include the use of Distributionally Robust Optimization (DRO) for data-driven stochastic optimization models with covariate information [22, 21] and new, less conservative DRO ambiguity sets [27]. We have developed advanced algorithms for solving nonlinear/nonconvex stochastic programming problems [31, 50], DRO with Wasserstein ambiguity sets [56], stochastic integer programs [8], chance-constrained integer programs [33, 32], and stochastic mixed-integer conic programs [23]. Finally, we have explored balancing wildfire risk and electricity distribution [41] and studied the optimal restoration sequence after a major outage [40]. In the remainder of this section we highlight a few of these results in more detail.

Integrating Machine Learning (ML) and Stochastic Optimization (SO). In the last reporting year, we investigated using residuals of ML models to devise a Sample Average Approximation (SAA) in order to solve conditional SO problems in the presence of covariates [22]. In this cycle, we investigated DRO variants to regularize and robustify the residuals-based SAA *when there are*

limited observations [21]. A key finding from this study is that DRO variants—especially those that consider distributions with support that can go beyond the current observations (e.g., Wasserstein distance or sample-robust ambiguity sets)—indeed result in better performance, even when the ML model is misspecified.

Reducing conservatism in DRO. While traditional DRO models protect against the unknown distribution, they may yield conservative solutions if the ambiguity set is too large. Motivated by the observation that most uncertainty distributions for renewable generation are unimodal, we integrate unimodality into a moment-based ambiguity set to reduce the conservatism of a DRO model [27]. Figure 7 illustrates the evolution of our optimality gap and reliability measure for a stochastic optimal power flow.

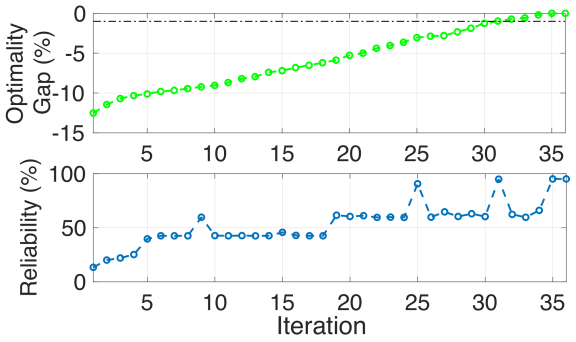


Figure 7: Optimality gap and reliability of intermediate solutions from the relaxed approximation algorithm for DR optimization.

Algorithms for security-constrained AC optimal power flow (SCACOPF). We continued to investigate scalable methods for stochastic programming with special focus on SCACOPF problems. Research and developments were done in FY21 to provide mathematically more sound (and, thus, more robust) algorithmic alternatives to classical heuristics [31], in particular stopping criteria, formulation of nonconvex line flow constraints, and nonsmooth AGC coupling. In [50] we focused on the development of a structured proximal subgradient algorithm for stochastic programming that has computationally affordable stopping criteria. The algorithm maintains the excellent decomposition and convergence properties of our SCACOPF methodology from [31]. The team has used the `HiOp-PriDec` implementation of the algorithm to solve large-scale simulations on the Summit and Lassen supercomputers, observing high parallel efficiencies [50].

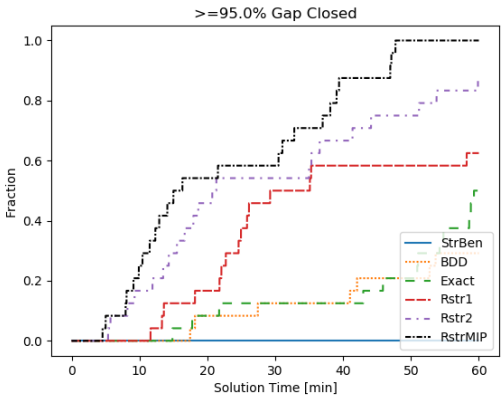


Figure 8: Number of instances for which 95% or more of the gap is closed over time using our proposed methods (RstrMIP,Rstr1,Rstr2) and existing approaches.

Dual decomposition of stochastic mixed-integer second-order cone programs (SMISOCPs). We have extended the dual decomposition of two-stage stochastic MILP to that of two-stage stochastic SMISOCPs. In particular, the parallel dual decomposition algorithm has been implemented in Argonne's open-source software DSP with Julia interface, publicly available in Github repository [23].

New cutting planes for stochastic integer programming (SIP). Traditional approaches for solving SIP either use a Lagrangian relaxation, which can be very expensive to compute, or use a branch-and-cut (BnC) approach based on LP relaxations, which can suffer due to weak bounds on the optimal value. In [8], a significant step forward was made in solving SIP by proposing an effective integration these approaches, enabling the efficient use of cuts for the BnC approach that are derived from the Lagrangian relaxation. The resulting approach has bounds stronger than the BnC approach, but is faster than Lagrangian relaxation, and leads to state-of-the-art results in solving SIP instances, as illustrated in Figure 8.

Wildfire mitigation. Building on our previous work [41], we are currently working on an optimization model which considers wildfire ignitions caused by power lines and/or other sources, and models the operation of the power system before and after the event. This is modeled as a stochastic optimization problem with integer variables in both the first and second stage. This work is a collaboration with Haoxiang Zhang and Lewis Ntaimo (Texas A&M), a collaboration that was initiated through the MACSER sponsored workshop at ICERM.

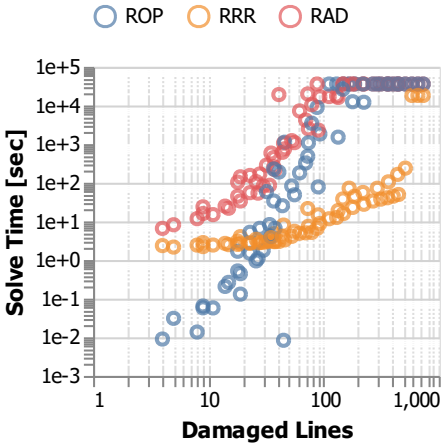


Figure 9: Solution quality vs solver time limit for four different algorithms. RRR is our proposed recursive algorithm.

Post-event restoration. We have studied the post-event restoration ordering problem, which prioritizes the repairs of damaged power system components to minimize the energy not served to customers [40]. To address the potential combinatorial explosion we have developed a new heuristic solution algorithm which obtains solutions by recursively splitting the set of damaged component into two parts, where repairs of the first set is prioritized over the second set. Numerical simulations show that our proposed algorithm obtains near-optimal solutions (within 1% gap) in the majority of cases and is 1000 times faster than existing algorithms (see Figure 9). This has allowed us to obtain better solutions for larger systems than what was previously possible.

2.4 Data-oriented Computation

In the last year, we obtained several new results in this area. In [15] we used mean-field analysis, continuous limits and gradient flow to show that, in the asymptotics of infinitely large deep architectures, the problem of training the network converges to a solution that perfectly fits the training data, resolving a phenomenon observed empirically but not explained theoretically up to our work. In [26], we proposed the first approach whereby deep neural network methodologies are utilized with the goal of performing inference on statistical models based solely on simulated data and parameters. We introduced a novel machine learning framework for learning dynamical systems in data assimilation [10] that outperforms existing methods that use expectation-maximization or particle filters to merge data assimilation and machine learning, Figure 10. We proposed a data fusion method based on multi-fidelity Gaussian process regression (GPR) framework [13]. Furthermore, we proposed a novel augmented Gaussian random field (AGRF) framework [55], which is a universal frame- work incorporating the data of observable and derivatives of any order. We defined a hybrid technique combining Bayesian inference and quantum-inspired Hamiltonian Monte Carlo (QHMC) method for imputation of missing datasets [14]. We proposed a novel an adaptive (stochastic) gradient perturbation method for differentially private empirical risk minimization [51] that considerably improves the utility guarantee compared to the standard differentially private method in which vanilla random noise is added.

In [9] we describe the first approach we are aware of that allows to do statistical inference scalably with implicit Gaussian process models, using hierarchical off diagonal low rank (HODLR) approaches *which answers the modeling and computational complexity difficulties simultaneously.* When modeling directly in data space (which is what we do with explicit Gaussian processes) there are no forms of kernels that ensure consistency and have the flexibility to deal with complex features, such as massive anisotropy and nonstationarity. When modeling in the process space we have considerably more flexibility, but the derivative of the covariance needs to be computed column by column and even with a scalable solver we need $O(n^2)$ operations. Our

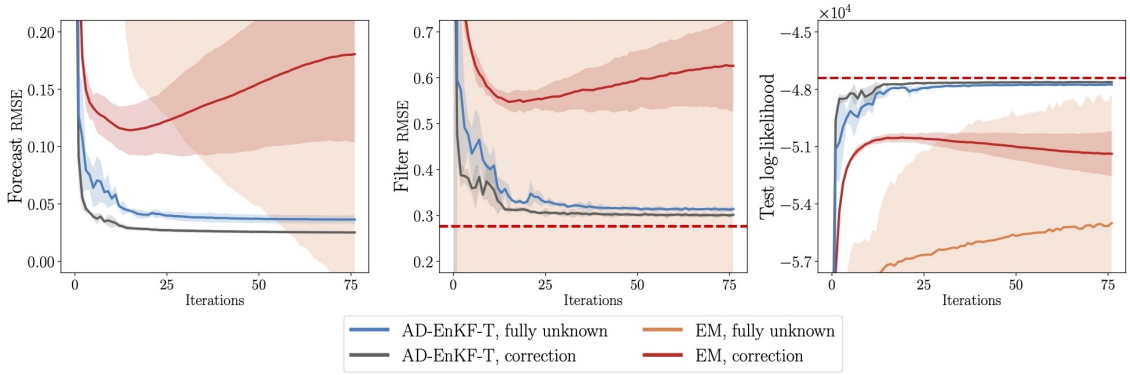


Figure 10: Learning Lorenz-96 from fully unknown dynamics v.s. model correction with partial observations ($H = [e_1, e_2, e_4, e_5, e_7, \cdots]^\top$). All performance metrics are evaluated after each training iteration. Red dashed lines correspond to metric values obtained with the reference model f^* and Q^*. The absence of lines for EM in the fully unknown setting is due to its low and unstable performance.

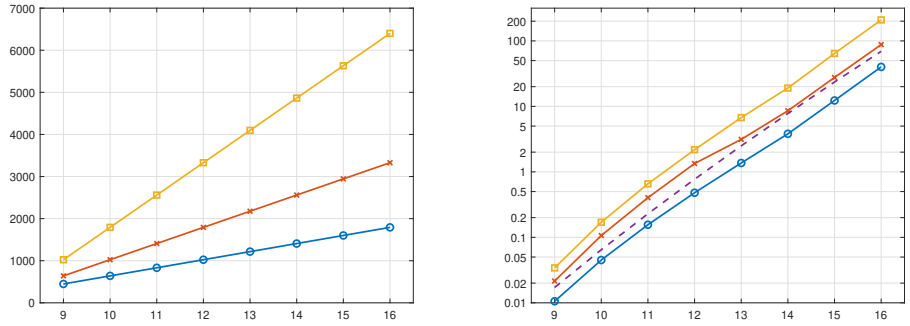


Figure 11: Computational complexity of constructing the HODLR approximation of K_U and its derivatives with respect to the parameters θ: (Left) the total number of required K_U-vector products and (right) shows the runtime (in seconds) of the rest linear algebra operations for fixed off-diagonal rank 32 (circle), 64 (x), 128 (square) over different sizes of observations. We use number of observations of size $n = 2^r$ with r ranging from 9 to 16 (x-axis). On the right, to demonstrate the scaling, the theoretical line (dashed line) corresponding to $O(n \log n)$ is added to the plot. Note that the x axis is represented in logarithmic units.

approach in [9] can use implicit models (generated for example by stochastic partial differential equations) which have *total flexibility*. However, our approach uses only $\log(n)$ forward solves with the implicit model to form an efficient approximation of the covariance matrix (Figure 11), computes the log-likelihood and its derivatives in $O(n \log^2 n)$ operations, and the Fischer matrix and consequently the relevant confidence intervals using only $O(n \log^2 n)$ computations. *Our approach allows thus both modeling flexibility and scalability.*

2.5 Bilevel Optimization

Preventing cascading failures. We have formulated bilevel optimization problem to determine the most severe cascade in a critical infrastructure network with partially or unknown connectivity. Our goal is to identify the most severe combination of failures of facilities in terms of their impact on the service level of the network. In this model, the upper-level decision maker (e.g. the hurricane) selects facilities that it closes so that the remaining service of the network is minimized. The lower-level decision maker (network operator) tries to maximize the service capacity subject to the closed facilities. Our model also allows for the inclusion of uncertainty in terms of network connectivity that is modeled as a simple affine uncertainty set. We are currently exploring stochastic optimization algorithms for solving this model that are motivated

by active-learning techniques.

New solution methods for bilevel linear programs. A classic solution approach for bilevel linear programs is to use the KKT conditions of the lower-level problem to reformulate the problem as a single level nonconvex quadratically constrained quadratic program (QCQPs). In [37], we study new formulations of nonconvex mixed-binary quadratic programs as completely positive programs. Specifically, Burer [7] showed that every nonconvex QP can be represented as a linear program over the Completely Positive (CP) cone of dimension $(n + 1)$ where n is the number of variables in the QP. We show that in the presence of m equality constraints the QP can be represented as a linear program over a CP cone of dimension $(n + 1 - m)$. We also provide conditions under which the proposed formulation has a strict interior while it is known that the formulation of Burer typically does not possess one.

An alternative formulation of a bilevel linear program introduces the lower level model into the upper level problem and constrains the objective to be at least as good as the optimal value function of the lower level problem. This leads to a (nonconvex) reverse convex constraint. Intersection cuts (ICs) are classic approach for deriving cuts to improve the relaxation of a problem containing such constraints from basic solutions of a linear relaxation that violate the reverse convex constraint. Surprisingly, in [48] we demonstrate an approach for deriving ICs from basic solutions that *satisfy* the reverse convex constraint, thus extending the power of this classic approach.

2.6 Nonlinear nonconvex local optimization

Nonlinear optimization algorithms. Quasi-Newton methods are highly effective, but for large-scale optimization problems with many sparse constraints, they rely on solving linear systems using direct methods and become impractical. In [6] we used a "shape-changing" trust-region (previously developed) and a novel reduced compact representation of the inverse quasi-Newton matrix to sharply improve the practicality of quasi-Newton methods in this context. Our method outperformed state-of-the-art methods in computational tests.

Variance-reduced primal-dual methods. Structured nonsmooth convex finite-sum optimization is ubiquitous in machine learning. For the primal-dual formulation of this problem, we proposed an algorithm called "Variance Reduction via Primal-Dual Accelerated Dual Averaging" (VRPDA2) [44]. Our approach takes coordinate steps in the dual variables and proximal steps incorporating momentum in the primal space. By exploiting a linear structure commonly encountered in machine learning applications, it achieves the optimal rate.

Complexity of smooth optimization algorithms. We continued work on worst-case complexity of algorithms for smooth optimization. Several papers of this type described in previous cycles were revised and published [11, 52]. In new work, we designed algorithms of projected gradient type for bound-constrained optimization with second-order scaling [53]. The two methods we described methods are more practical than the log-barrier approach developed in previous work, and have competitive complexity.

Additionally, we extended our Newton-CG approach for unconstrained optimization developed in earlier papers to finite-sum objectives, a structure that arises often in ML [54]. In this setting, gradients and Hessians are available only as sampled approximations, and the inexactness must be accounted for in the algorithm design and analysis.

Distributionally robust classification. In [30] we describe a model for classification in ML in which the population distribution is assumed to lie in a Wasserstein ball around the empirical distribution defined by the training set. We explore links to other classification paradigms and show that for certain interesting distribution sets, our model exhibits benign nonconvexity, that is, gradient-based algorithms will converge to the global minimizer.

Figure 12 illustrates the robustness of our formulation by comparison with the standard hinge-loss function for a binary classification problem with data is drawn from three different

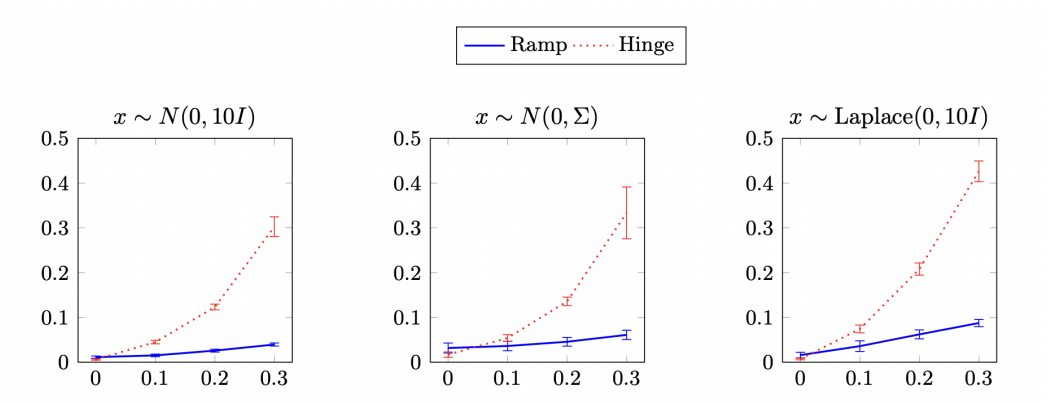


Figure 12: Test error (vertical axis) versus fraction flipped (horizontal axis) for nonseparable data, by distribution type. Averaged over 20 trials; error bars shows one standard deviation.

distributions, with a given random fraction of the labels of one class flipped to the other (incorrect) class. As we increase the fraction of points flipped, the figure shows that the test error of hinge loss (the objective traditionally used for linear classification) degrades severely, while the test error of the DRO formulation is much more stable.

2.7 Dynamic Optimization

In this cycle we made progress on multistage Distributionally Robust Optimization (DRO), dynamic complementarity and equilibrium problems, load-frequency control problems, and shaping cost trajectories for dynamic optimization. These contributions are summarized below.

Multistage DRO. In many dynamic optimization problems under uncertainty and rare events, the true distribution of the underlying stochastic process is rarely known. Multistage DRO provides an attractive approach in these cases. In [39] we investigate the question of how to define and identify critical scenarios for convex nested multistage DRO problems. We call scenarios/realizations that change the optimal value when removed as "effective". A key result shows that a scenario path (from root to end node of a scenario tree) is effective if and only if all realizations along that path are *conditionally* effective. This result provides a computational tool as well as managerial insight into this class of dynamic optimization problems. In [4], we consider general multistage DRO formed by ϕ-divergences or the Wasserstein distance and devise novel lower bounding techniques (for minimization problems) by scenario subgrouping and convolution of risk measures. The proposed methods produce high-quality bounds in a fraction of time for challenging multistage DRO lacking special structure (e.g., convexity, stagewise independence, binary state variables) that prevent decomposition.

Dynamic complementarity and equilibrium problems. In [38], we study the Discrete-Time Linear Complementarity System (DLCS)—a dynamical system in discrete time whose state evolution is governed by linear dynamics in states and algebraic variables that solve a Linear Complementarity Problem (LCP). We derive sufficient conditions for Lyapunov stability of a DLCS when using a quadratic Lyapunov function that depends only on the state variables and a quadratic Lyapunov function that depends both on the state and the algebraic variables. The sufficient conditions require checking the feasibility of a copositive program over nonconvex cones. We then devise a novel, exact cutting plane algorithm for the verification of stability and the computation of the Lyapunov functions. In ongoing work, Huber, Shen and Ferris develop a new decomposition algorithm for solving stochastic equilibrium problems with risk-averse agents. The forward step of this algorithm solves risk-adjusted equilibrium problems, and the backward step utilizes the duality structure of the risk-set

to modify each agents risk measure. Preliminary convergence results have been derived.

Dynamic control of power systems. We study Load Frequency Control (LFC) problem subject to uncertain load disturbances and stochastic wind generation. First, we devise a tractable framework for handling a chance-constrained LFC problem with Gaussian parametric uncertainties, where we use Gaussian process regression. (Figure 13). The problem allows constraint violation at a prescribed probability, and the Gaussian process exploits the probability distribution to estimate costs and to propagate mean and covariance of these functions over the prediction horizon. A comparative study between the proposed framework , denoted GP-SNMPC, and Sample Average Approximation (SAA) demonstrates that the GP-SNMPC framework is computationally more efficient and delivers a better performance in keeping load frequency balance while maintaining the system constraints. Next, we incorporate the unscented Kalman filter (UKF) [49] to this control problem to estimate the states and to propagate the mean and covariance of the states over the prediction horizon. By resorting to the Chebyshev-Cantelli inequality [28], the Schur complement, and reformulating, we obtain a conservative semidefinite program (SDP). The SDP is computationally efficient and delivers satisfactory performance with respect to chance constraints.

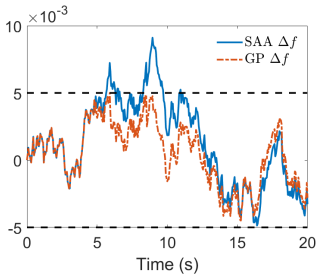
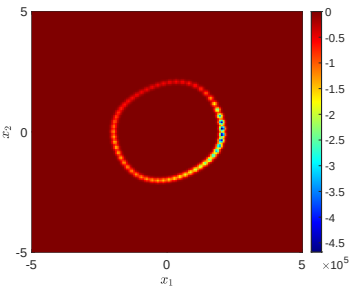
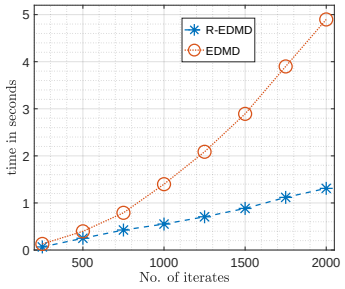


Figure 13: GP-SNMPC framework for LFC compared to SAA.

Shaping cost trajectories for dynamic optimization. As a byproduct of our work for modeling uncertainty over continuous domains (e.g., space and time) that utilizes random field theory [36], we propose new measures to shape dynamic cost trajectories by transferring risk measures from static stochastic optimization to dynamic optimization [35]. We prove that dynamic optimization problems can be classified as a special case of a two-stage stochastic program; from which it readily follows that stochastic risk measures can be used to shape dynamic cost trajectories. This key result allows us greatly enhanced modeling flexibility in posing dynamic optimization problems. For instance, we can use a time-valued conditional-value-at-risk measure to promote a smoothed dynamic cost trajectory that is shaped by penalizing high cost deviations.



(a)



(b)

Figure 14: (a) Invariant measure identified after 2000 Koopman operator iterates. (b) Computation time of R-EDMD vs EDMD.

2.8 Model Reduction

In this performance period, we have developed novel methods for discovering dynamical systems from streaming data, sparsity-promoting methods for model reduction in uncertainty quantification, and renormalized reduced-order models for nonlinear partial differential equations models.

Robust dynamics discovery from streaming noisy data. We propose an algorithm for discovering dynamical systems from noisy streaming time-series data based on Koopman operator theory and robust optimization. The proposed Recursive Extended Dynamic Mode Decomposition (R-EDMD) algorithm aims to address the problem of real-time identification of power system dynamics from data contaminated with process and measurement noise. In R-EDMD, we compute a robust Koopman operator approximation using robust optimization techniques (to mitigate

the effect of noise) recursively from streaming data. To illustrate the efficacy of the proposed approach, we identify the invariant measure of a noisy Van der Pol oscillator (Figure 14). The computational cost of R-EDMD is almost linear with respect to the number of iterates, whereas the computation time of the standard EDMD algorithm grows exponentially with the number of iterates; therefore, R-EDMD leads to a significant reduction in computation time, thus facilitating real-time identification of power system dynamics.

Sparsity-promoting uncertainty quantification methods. We propose a general framework to estimate coefficients of generalized polynomial chaos uncertainty models via rotational sparse approximation [19]. In particular, we aim to identify a rotation matrix such that the gPC expansion of a set of random variables after rotation allows a sparser representation than before rotation. To solve this problem, we employ non-convex sparsity-promoting regularization. The proposed combination of rotation and non-convex sparse-promoting regularization can yield more accurate sparse regression models using fewer data than similar existing methods.

Renormalized reduced-order models. Extreme events in complex energy and environment systems can contain many more scales than can be simulated with available computational resources. We aim to construct reduced models but popular formalisms assume that we can simulate the original system for long times. When this is not possible, we have formulated a renormalized version of model reduction. To the best of our knowledge we have constructed the first perturbatively renormalized reduced-order models for the Burgers equation and 3D Euler equations of incompressible flow without introducing any terms by hand, which has been a long-sought goal [34].

2.9 Frameworks

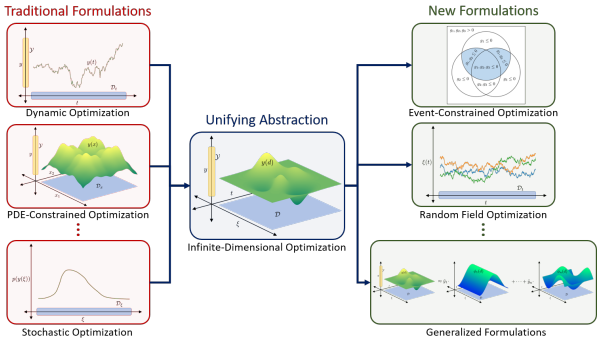


Figure 15: A visual summary of our unifying abstraction for infinite-dimensional optimization.

We have proposed a new paradigm, that we call random field optimization, for modeling uncertainty over continuous domains (e.g., space and time) that utilizes random field theory to yield infinite-dimensional optimization formulations [36]. This paradigm allows us to represent dynamic-stochastic problems in continuous time and enables us to model uncertainty that propagates over other domains (e.g., space). Our approach is general and captures a wide-breadth of existing uncertainty characterizations and techniques. Leveraging the connection between dynamic and stochastic optimization established by our unifying abstraction for infinite-dimensional optimization, we have proposed new measures to shape dynamic cost trajectories by transferring risk measures from stochastic optimization to a dynamic optimization context [35]. We have continued development of our Julia-based modeling platform `InfiniteOpt.jl` (see https://github.com/pulsipher/InfiniteOpt.jl). This package implements our unifying abstraction for infinite-dimensional optimization that is illustrated in Figure 15. We have rigorously characterized this abstraction as established in [36].

In [20] we introduce a class of optimal value function (OVF) compositions that are flexible enough to capture many applications with nested or (limited) hierarchical structure. We

relate OVF's to the more familiar framework of convex composite functions by making simplifying structural assumptions on the forms of our problems. We focus on three reformulation approaches for OVF compositions that lead to problems that have standard, reliable and implementable solution methods. We have developed a new decomposition approach for solving stochastic equilibrium problems with risk-averse agents and implemented in `ReShop`, an open-source tool for stochastic equilibrium problems. Based on the structure that is available in `ReShop`, we use a forward backward structured approach, where the forward step solves risk-adjusted equilibrium problems, and the backward step utilizes the duality structure to modify each agents risk measure. Preliminary convergence results have been developed and the approach has been successfully prototyped in `ReShop`. The group has collaborated with the rare-events team to generate data sets, as well as interacting with the Public Services Commission (PSC) of Wisconsin and the Office of Sustainability. A policy use of this framework was published in [2]. Its general scope is to solve hierarchical optimization problems through reformulations; we allow these equilibria to be hierarchical in nature, exploiting the theory developed in [20].

In [12] we describe `SUTIL`, an open-source C and C++ utility library for multi-stage stochastic programs. The library contains tools for (i) reading problems in Stochastic Mathematical Programming System (SMPS) format, a generalization of the MPS format for stochastic optimization problems, (ii) basic functions like forming deterministic equivalent and expected-value problems, (iii) handling two- and multistage stochastic optimization data structures such as scenario trees, (iv) forming subproblems either stage-wise or scenario-wise, and (v) implementations of various basic and advanced sampling methods. The library has utility for developing new decomposition- and sampling-based algorithms for solving stochastic programs. We used the `SUTIL` library in [39] and [56].

3 Visibility and Outreach

Sharing Results. The outputs of the project are shared via peer-reviewed journal and conference publications, presentations at professional workshops and conferences, and via the project website. During FY21, MACSER PI's have completed a total of 61 journal and conference papers, of which 21 have been published, 2 are in press and 38 are preprints and submissions currently under review. Furthermore, 35 submissions and preprints reported in the previous cycle have been updated during FY21, of which 27 were published and 8 are under review. The PIs have been active in presenting their work at conferences and workshops, giving a total of 100 presentations about work related to the project. Of these presentations 11 have been plenary or keynote presentations.

MACSER participants have contributed to or created 9 software items during FY21, and the code associated to 7 of the submissions mentioned above is also openly available, and the remaining 2 will be made available soon.

The MACSER web site (`https://www.mcs.anl.gov/MACSER/`) is another venue where the work of the MACSER project is shared widely. The site includes an overview of the project, including an introduction to the mathematical challenges, list of publications, team members, and news about significant project actives.

Recognition. MACSER personnel received several prestigious awards over the past year:

Rebecca Willett has been named a Fellow of the Society for Industrial and Applied Mathematics (SIAM). She is one of 28 people selected as SIAM Fellows for the class of 2021. Prof. Willett's nomination cites "contributions to mathematical foundations of machine learning, large-scale data science, and computational imaging."

Michael Stein has been appointed Editor of Applications & Case Studies for Journal of the American Statistical Association.

Sungho Shin, a MACSER postdoctoral appointee at Argonne National Laboratory, has won the Young Author Award at the 11th International Symposium on Advanced Control of System

Process (ADCHEM 2021), held virtually in June 2021. In the award paper [43], Shin presented nonlinear optimization problems as graph-structured optimization problems and shows how that structure can be exploited at both the modeling and the solver level. The approach, which he implemented as a general-purpose nonlinear programming solver called `MadNLP.jl`, was evaluated on problems arising in transient gas network optimization and multiperiod AC optimal power flow. Compared to off-the-shelf tools, `MadNLP.jl` reduced solution times by 300%.

Christopher J. Geoga, a MACSER-supported graduate student, published [16], a paper that won an honorable mention in 2021 in Student Paper Competition of the Section on Statistics and the Environment of the American Statistical Association.

References

[1] Joel A E Andersson, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl. "CasADi – A software framework for nonlinear optimization and optimal control". In: *Mathematical Programming Computation* 11.1 (2019), pp. 1–36. DOI: 10.1007/s12532-018-0139-4.

[2] Josh Arnold, Adam Christensen, and Michael Ferris. "Werewolf in London: Minding the Gap with User-Friendly Energy Optimization Tools Informing Policy Makers through the Energy Transition". In: *Accelerating the energy transition for all: evaluation's role in effective policy making*. Energy Evaluation Europe. 2021.

[3] Ahmed Attia and Emil Constantinescu. "Optimal Experimental Design for Inverse Problems in the Presence of Observation Correlations". In: *arXiv e-prints* (May 2021). arXiv: 2007.14476 [math.OC].

[4] Güzin Bayraksan, Daniel Faccini, Francesca Maggioni, and Ming Yang. *Bounds for Multistage Mixed-Integer Distributionally Robust Optimization*. To be submitted. 2021.

[5] Freddy Bouchet, Krzysztof Gawędzki, and Cesare Nardini. "Perturbative calculation of quasi-potential in non-equilibrium diffusions: a mean-field example". In: *Journal of Statistical Physics* 163.5 (2016), pp. 1157–1210.

[6] Johannes J. Brust, Roummel F. Marcia, Cosmin G. Petra, and Michael A. Saunders. "Large-scale Optimization with Linear Equality Constraints using Reduced Compact Representation". In: *arXiv e-prints*, arXiv:2101.11048 (Jan. 2021), arXiv:2101.11048. arXiv: 2101.11048 [math.OC].

[7] S. Burer. "On the Copositive Representation of Binary and Continuous Nonconvex Quadratic Programs". In: *Mathematical Programming* 120 (2009), pp. 479–495.

[8] Rui Chen and James Luedtke. "On Generating Lagrangian Cuts for Two-stage Stochastic Integer Programs". In: *arXiv e-prints*, arXiv:2106.04023 (June 2021), arXiv:2106.04023. arXiv: 2106.04023 [math.OC].

[9] Yian Chen and Mihai Anitescu. "Scalable Gaussian Process Analysis for Implicit Physics-Based Covariance Models". In: *International Journal for Uncertainty Quantification* 11.6 (2021).

[10] Yuming Chen, Daniel Sanz-Alonso, and Rebecca Willett. "Auto-differentiable Ensemble Kalman Filters". In: *arXiv e-prints*, arXiv:2107.07687 (July 2021), arXiv:2107.07687. arXiv: 2107.07687 [stat.ML].

[11] Frank E Curtis, Daniel P Robinson, Clément W Royer, and Stephen J Wright. "Trust-Region Newton-CG with Strong Second-Order Complexity Guarantees for Nonconvex Optimization". In: *SIAM Journal on Optimization* 31.1 (2021), pp. 518–544.

[12] Joe Czyzyk, Cong Han Lim, Jeff Linderoth, Zhichao Ma, Hamed Rahimian, and G. Bayraksan. *SUTIL: A Library for Stochastic Programming*. Working paper. 2021.

[13] Yixiang Deng, Guang Lin, and Xiu Yang. "Multifidelity data fusion via gradient-Enhanced Gaussian process regression". In: *Communications in Computational Physics* 28.5 (2020), pp. 1812–1837. ISSN: 1991-7120. DOI: https://doi.org/10.4208/cicp.OA-2020-0151. URL: http://global-sci.org/intro/article_detail/cicp/18397.html.

[14] Xiu Yang Didem Kochan Zheng Zhang. "A Quantum-Inspired Hamiltonian Monte Carlo Method for Missing Data Imputation". Submitted to SIAM International Conference on Data Mining. 2021.

[15] Zhiyan Ding, Shi Chen, Qin Li, and Stephen Wright. "Overparameterization of deep ResNet: zero loss and mean-field analysis". In: *arXiv e-prints*, arXiv:2105.14417 (May 2021), arXiv:2105.14417. arXiv: 2105.14417 [cs.LG].

[16] Christopher J. Geoga, Mihai Anitescu, and Michael L. Stein. "Flexible nonstationary spatiotemporal modeling of high-frequency monitoring data". In: *Environmetrics* 32.5 (2021), e2670. DOI: https://doi.org/10.1002/env.2670. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2670.

[17] C. L. Haley. "Missing-data multitaper coherence estimation". In: *IEEE Signal Processing Letters* 28 (Sept. 2021), pp. 1704–1708. DOI: 10.1109/LSP.2021.3105926. Published.

[18] Charlotte L. Haley and Christopher J. Geoga. "Multitaper.jl: A Julia package for frequency domain analysis of time series". In: *Journal of Open Source Software* 5.55 (2020), p. 2463. DOI: 10.21105/joss.02463. URL: https://doi.org/10.21105/joss.02463.

[19] Mengqi Hu, Yifei Lou, and Xiu Yang. "A general framework of rotational sparse approximation in uncertainty quantification". In: *arXiv e-prints*, arXiv:2101.04890 (Jan. 2021), arXiv:2101.04890. arXiv: 2101.04890 [stat.CO].

[20] Olivier Huber and Michael C. Ferris. "Reformulations for convex composite functions and their nested compositions". 2021.

[21] Rohit Kannan, Güzin Bayraksan, and James R. Luedtke. "Residuals-based distributionally robust optimization with covariate information". In: *arXiv e-prints*, arXiv:2012.01088 (Dec. 2020), arXiv:2012.01088. arXiv: 2012.01088 [math.OC].

[22] Rohit Kannan, Güzin Bayraksan, and James. R. Luedtke. *Data-driven sample average approximation with covariate information*. http://www.optimization-online.org/DB_HTML/2020/07/7932.html. 2020.

[23] Kibaek Kim and Victor M Zavala. *An open-source software package: DSP*. 2015. URL: https://github.com/Argonne-National-Laboratory/DSP.git.

[24] Mitchell Krock, Julie Bessac, Michael L. Stein, and Adam H. Monahan. "Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails". In: *arXiv e-prints*, arXiv:2110.10046 (Oct. 2021), arXiv:2110.10046. arXiv: 2110.10046 [stat.AP].

[25] Amanda Lenzi, Julie Bessac, and Mihai Anitescu. "Detecting Large Frequency Excursions in the Power Grid with Bayesian Decision Theory". Submitted to IEEE Open Access Journal of Power and Energy. 2021.

[26] Amanda Lenzi, Julie Bessac, Johann Rudi, and Michael L. Stein. "Neural Networks for Parameter Estimation in Intractable Models". In: *arXiv e-prints*, arXiv:2107.14346 (July 2021), arXiv:2107.14346. arXiv: 2107.14346 [stat.ME].

[27] Bowen Li, Ruiwei Jiang, and Johanna L. Mathieu. "Integrating Unimodality into Distributionally Robust Optimal Power Flow". Submitted to special issue of TOP. 2021.

[28] Lalo Magni, Daniele Pala, and Riccardo Scattolini. "Stochastic model predictive control of constrained linear systems with additive uncertainty". In: *2009 European Control Conference (ECC)*. IEEE. 2009, pp. 2235–2240.

[29] Adrian Maldonado, Emil M. Constantinescu, Hong Zhang, Vishwas Rao, and Mihai Anitescu. "Trust-region approximation of extreme trajectories in power system dynamics". Submitted to IEEE Transactions on Power Systems. 2021.

[30] Nam Ho-Nguyen and Stephen J. Wright. "Adversarial Classification via Distributional Robustness with Wasserstein Ambiguity". In: *arXiv e-prints* (May 2020). arXiv: 2005.13815 [cs.LG].

[31] Cosmin G. Petra and Ignacio Aravena Solis. "Solving realistic security-constrained optimal power flow problems". Submitted to Operations Research. 2021. URL: https://arxiv.org/abs/2110.01669.

[32] Alberto Del Pia, Jeff Linderoth, and Haoran Zhu. "Multi-cover Inequalities for Totally-Ordered Multiple Knapsack Sets". Submitted to Mathematical Programming B. 2021.

[33] Alberto Del Pia, Jeff Linderoth, and Haoran Zhu. "Multi-cover Inequalities for Totally-Ordered Multiple Knapsack Sets". In: *IPCO 2021: The 22nd Conference on Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science*. Ed. by Mohit Singh and David P. Williamson. Vol. 12707. Springer, 2021, pp. 193–207.

[34] Jacob Price, Brek Meuris, Madelyn Shapiro, and Panos Stinis. "Optimal renormalization of multiscale systems". In: *Proceedings of the National Academy of Sciences* 118.37 (2021). ISSN: 0027-8424. DOI: 10.1073/pnas.2102266118. eprint: https://www.pnas.org/content/118/37/e2102266118.full.pdf. URL: https://www.pnas.org/content/118/37/e2102266118.

[35] Joshua L Pulsipher and Victor M Zavala. "Measuring and optimizing system reliability: a stochastic programming approach". In: *TOP* (2020), pp. 1–20. URL: https://rdcu.be/b8pNY.

[36] Joshua L. Pulsipher, Weiqi Zhang, Tyler J. Hongisto, and Victor M. Zavala. "A unifying modeling abstraction for infinite-dimensional optimization". In: *Computers & Chemical Engineering* 156 (2022), p. 107567. ISSN: 0098-1354. DOI: https://doi.org/10.1016/j.compchemeng.2021.107567. URL: https://www.sciencedirect.com/science/article/pii/S0098135421003458.

[37] Arvind Ragathunan and Jeff Linderoth. *Reduced Space Completely Positive Representation Of Binary And Continuous Nonconvex Quadratic Programs*. Working paper. 2021.

[38] Arvind U. Raghunathan and Jeffrey T. Linderoth. "Stability Analysis of Discrete-Time Linear Complementarity Systems". In: *arXiv e-prints*, arXiv:2012.13287 (Dec. 2020), arXiv:2012.13287. arXiv: 2012.13287 [math.OC].

[39] Hamed Rahimian, Guzin Bayraksan, and Tito Homem-de-Mello. "Effective Scenarios in Multistage Distributionally Robust Optimization with a Focus on Total Variation Distance". In: *arXiv e-prints*, arXiv:2109.06791 (Sept. 2021), arXiv:2109.06791. arXiv: 2109.06791 [math.OC].

[40] Noah Rhodes, David M. Fobes, Carleton Coffrin, and Line Roald. "PowerModelsRestoration.jl: An open-source framework for exploring power network restoration algorithms". In: *Electric Power Systems Research* 190 (2021), p. 106736. ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2020.106736. URL: https://www.sciencedirect.com/science/article/pii/S0378779620305393.

[41] Noah Rhodes, Lewis Ntaimo, and Line Roald. "Balancing Wildfire Risk and Power Outages Through Optimized Power Shut-Offs". In: *IEEE Transactions on Power Systems* 36.4 (2021), pp. 3118–3128. DOI: 10.1109/TPWRS.2020.3046796.

[42] J. Roth, D. A. Barajas-Solano, P. Stinis, J. Weare, and M. Anitescu. "A Kinetic Monte Carlo Approach for Simulating Cascading Transmission Line Failure". In: *Multiscale Modeling & Simulation* 19.1 (2021), pp. 208–241. DOI: 10.1137/19M1306865.

[43] Sungho Shin, Carleton Coffrin, Kaarthik Sundar, and Victor M Zavala. "Graph-Based Modeling and Decomposition of Energy Infrastructures". In: *IFAC-PapersOnLine* 54.3 (2021), pp. 693–698.

[44] Chaobing Song, Stephen J. Wright, and Jelena Diakonikolas. "Variance Reduction via Primal-Dual Accelerated Dual Averaging for Nonsmooth Convex Finite-Sums". In: *arXiv e-prints*, arXiv:2102.13643 (Feb. 2021), arXiv:2102.13643. arXiv: 2102.13643 [math.OC].

[45] M. L. Stein. "Decentralized Filtering Adaptive Neural Network Control for Uncertain Switched Interconnected Nonlinear Systems". In: *Environmetrics* (2020).

[46] Michael L. Stein. "A parametric model for distributions with flexible behavior in both tails". In: *Environmetrics* (2020), e2658. DOI: 10.1002/env.2658. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2658.

[47] Michael L. Stein. "A weighted composite log-likelihood approach to parametric estimation of the extreme quantiles of a distribution". Submitted to Extremes. 2021. URL: https://www.mcs.anl.gov/MACSER/files/2021/10/WCL-extreme.pdf.

[48] Eli Towle and James Luedtke. "Intersection Disjunctions for Reverse Convex Sets". In: *Mathematics of Operations Research* (2021). DOI: 10.1287/moor.2021.1132. eprint: https://doi.org/10.1287/moor.2021.1132. URL: https://doi.org/10.1287/moor.2021.1132.

[49] Eric A Wan, Rudolph Van Der Merwe, and Simon Haykin. "The unscented Kalman filter". In: *Kalman filtering and neural networks* 5.2007 (2001), pp. 221–280.

[50] Jingyi Wang, Nai-Yuan Chiang, and Cosmin G. Petra. "An asynchronous distributed-memory optimization solver for two-stage stochastic programming problems". In: *2021 20th International Symposium on Parallel and Distributed Computing (ISPDC)*. 2021, pp. 33–40. DOI: 10.1109/ISPDC52870.2021.9521613. URL: https://ieeexplore.ieee.org/abstract/document/9521613.

[51] Xiaoxia Wu, Lingxiao Wang, Irina Cristali, Quanquan Gu, and Rebecca Willett. "Adaptive Differentially Private Empirical Risk Minimization". In: *arXiv e-prints*, arXiv:2110.07435 (Oct. 2021), arXiv:2110.07435. arXiv: 2110.07435 [cs.LG].

[52] Yue Xie and Stephen J Wright. "Complexity of proximal augmented Lagrangian for non-convex optimization with nonlinear equality constraints". In: *Journal of Scientific Computing* 86.3 (2021), pp. 1–30.

[53] Yue Xie and Stephen J. Wright. "Complexity of Projected Newton Methods for Bound-constrained Optimization". In: *arXiv e-prints*, arXiv:2103.15989 (Mar. 2021), arXiv:2103.15989. arXiv: 2103.15989 [math.OC].

[54] Zhewei Yao, Peng Xu, Fred Roosta, Stephen J. Wright, and Michael W. Mahoney. "Inexact Newton-CG Algorithms With Complexity Guarantees". In: *arXiv e-prints*, arXiv:2109.14016 (Sept. 2021), arXiv:2109.14016. arXiv: 2109.14016 [math.OC].

[55] Sheng Zhang, Xiu Yang, Samy Tindel, and Guang Lin. "Augmented Gaussian random field: Theory and computation". In: *Discrete & Continuous Dynamical Systems—S* (2021). URL: https://www.aimsciences.org/article/doi/10.3934/dcdss.2021098.

[56] Chennan Zhou and Güzin Bayraksan. *Effective Scenarios in Two-Stage Distributionally Robust Optimization with Wasserstein Distance*. Working paper. 2021.

A Up-to-date Organizational chart