Research Statement

Suhan Zhong (suzhong@tamu.edu)

I work in optimization with connections to data science and machine learning. Specifically, my work uses techniques from real algebraic geometry, moment problems and optimization theory. I am particularly interested in problems related to stochastic optimization, multilevel optimization, equilibrium problems, and tensors.

In the fields of data science and machine learning, there is significant interest in optimization models that incorporate uncertainty and hierarchical structures. As these models become increasingly complex, computing or estimating global optimizers/optimal values can be extremely challenging. Yet the need for certifying global optimality is crucial due to these models' real-world implications. Recent advances in polynomial optimization offer promising solutions to bridge this gap. Polynomials are widely used in nonlinear approximations. A generic polynomial optimization can be solved globally by Moment-Sum-Of-Squares (SOS) relaxations. Hierarchies with Lagrange multiplier expressions and term sparsity facilitate the efficient adaptation to handle larger datasets. Inspired by these properties, I aim to develop *global* algorithms for solving random and multilevel optimization models.

Optimization under uncertainty is an established and evolving topic. It originated from operations research and has recently gained popularity in data science for training predictive models. Challenges for solving such problems may arise from: the objective function is implicitly defined, thus lacking a parametric expression; the underlying measure is only partially described by samples; the existence of multiple spurious (local but not global) optimizers. My work resolves above issues using polynomial optimization techniques to (i) find efficient function estimate; (ii) solve distributionally robust optimizers; (iii) construct loss functions without spurious local minimizers.

Stackelberg games and generalized Nash equilibrium problems (GNEPs) are historically difficult optimization problems. Stackelberg games optimize over a feasible set constrained by the optimizer set of another optimization problem. GNEPs aim to find strategies for N players such that each player's objective cannot be improved given the strategies of the other players. The main challenge to compute Stackelberg and Nash equilibria often lies in the difficulty of characterizing the feasible sets. My work studies parametric algebraic approximations for Stackelberg and Nash equilibria using optimality conditions. We propose semidefinite algorithms to initialize and tighten these approximations. Under proper assumptions, the algorithm will either return an equilibrium or detect nonexistence.

The following sections include projects that I am currently working on and a future research plan is given afterwards.

1 Data science optimization under uncertainty

Assume the data of interest, ξ, follows the distribution of a measure ν supported on $\Xi \subseteq \mathbb{R}^p$. A common goal is to determine parameters $x \in X \subseteq \mathbb{R}^n$ for a prescribed model such that the loss function $\mathbb{E}_\nu[f(x,\xi)]$ is minimized.

1.1 Two-stage stochastic optimization

In a training model with two layers, the objective function is implicitly defined as (i.e., F, g are given scalar and vector-valued functions respectively)

$$f(x,\xi) = \inf_{y}\{F(x,y,\xi) \,|\, g(x,y,\xi) \geq 0\}.$$

Such a function f is usually nonconvex, nonsmooth, and lacks a parametric expression. When F, g are polynomials, we can compute explicit polynomial bounds by solving the linear conic optimization (note $\mathscr{P}(K)$ is the set of polynomials that are nonnegative on K)

$$\begin{cases} \max\limits_{p(x,\xi)} & \int_{\mathcal{F}} p(x,\xi)\mathtt{d}\nu \\ \text{subject to} & F(x,y,\xi) - p(x,\xi) \in \mathscr{P}(K), \end{cases} \tag{1.1}$$

where $K = \{(x,y,\xi) \,|\, x \in X, \xi \in \Xi, g(x,y,\xi) \geq 0\}$. Suppose $\{p_k(x,\xi)\}_{k\in\mathbb{N}}$ is an optimizing sequence of (1.1) with $\deg(p_k) = 2k$. Then $f - p_k \geq 0$ on $X \times \Xi$. Under compactness, we further have

$$\int |f - p_k|\mathtt{d}\nu \to 0 \text{ as } k \to \infty.$$

Its computational complexity only depends on $n + n_1 + p$ and $\deg(p_k)$, but not the sample

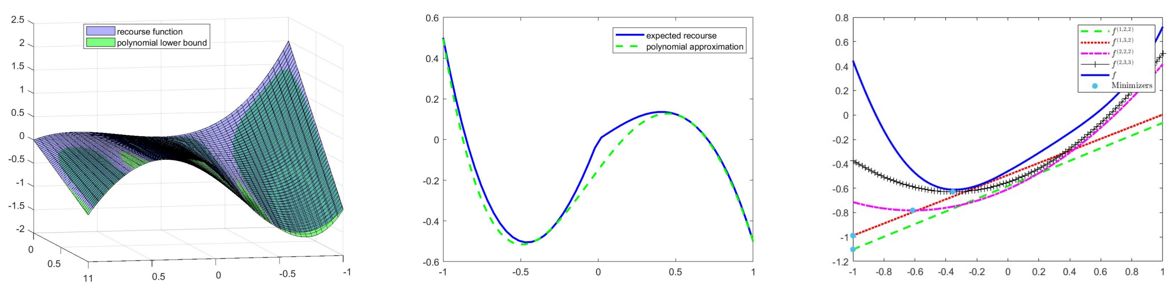

Figure 1: The left and middle figures show that $f, \mathbb{E}_\nu[f]$ are approximated by polynomials $p, \mathbb{E}_\nu[p]$. The right figure shows the improvement of approximations as k increases.

size. So this approach has computational advantages when ν is continuous. By Moment-SOS relaxations, we can solve for an optimizer $x^{\mathtt{k}}$ of $\min \mathbb{E}_\nu[p_k(x,\xi)]$. Then

$$\mathbb{E}_\nu[p_k(x^{\mathtt{k}},\xi)] \leq f_{\min} \leq \mathbb{E}_\nu[f(x^{\mathtt{k}},\xi)]$$

where $f_{\min}$ denotes the true optimal value of the original problem. When the bounding gap is small, $x^{\mathtt{k}}$ is a reliable solution estimate to the original problem (see Figure 1 for illustration).

Corresponding paper

1. S. Zhong, Y. Cui and J. Nie, Towards global solutions for nonconvex two-stage stochastic programs: a polynomial lower approximation approach, *SIAM J. Optim.*, to appear (2024). arXiv:2310.04243

1.2 Distributionally robust optimization

The underlying measure of the random variables is unknown in practice. But its support and moments can be partially estimated, i.e.,

$$\nu \in \mathcal{M} = \{\mu \,|\, \texttt{supp}(\mu) \subseteq \Xi, l_\alpha \le \mathbb{E}_\mu[\xi^\alpha] \le u_\alpha \,(\alpha \in \mathcal{A})\}.$$

Here $\Xi \subseteq \mathbb{R}^p$ is fixed, $\mathcal{A}$ is a finite index set and l_α, u_α are scalars determined by empirical data. An interesting question is to optimize an objective function under the worst-case measure in $\mathcal{M}$. It is formulated as *distributionally robust optimization* (DRO)

$$\begin{cases} \min\limits_{x\in X} & f_0(x) \\ \text{subject to} & \inf\limits_{\nu\in\mathcal{M}} \mathbb{E}_\nu[f(x,\xi)] \ge 0. \end{cases}$$

The DRO problem is generically computational intractable when the cardinality $|\mathcal{M}|$ is infinite. Assume f is polynomial in ξ. We can rewrite $f(x,\xi) = \langle f(x), [\xi]_d \rangle$ with the coefficient vector $f(x)$ and the monomial vector of ξ up to degree d. Then the inf-constraint

$$\inf_{\nu\in\mathcal{M}} \mathbb{E}_\nu[f(x,\xi)] \ge 0 \quad \Leftrightarrow \quad \langle f(x), y \rangle \ge 0 \quad \forall y \in Y,$$

where Y is a *numerically trackable* truncated moment cone. For instance, consider ξ is univariate, $d = 4$, $0 \le \xi^\alpha \le 1$ for $\alpha = 0, 1, \dots, 4$, and $\Xi = [a_1, a_2]$, then Y can be expressed by the constraints $(y_0, y_1, y_2, y_3, y_4) \ge 0$ and

$$\begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0, \quad (a_1 + a_2) \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix} \succeq a_1 a_2 \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} + \begin{bmatrix} y_2 & y_3 \\ y_3 & y_4 \end{bmatrix},$$

Using this reformulation, we develop an efficient semidefinite algorithm to solve DRO globally. Under general assumptions, it can not only return the true global optimizer, but also the worst-case measure, whose finite convergence is self-verified by flat truncation condition.

Corresponding paper

1. J. Nie, L. Yang, **S. Zhong** and G. Zhou, Distributionally robust optimization with moment ambiguity sets, *J. Sci. Comput.*, 94.12 (2023): 1-27. `doi.org/10.1007/s10915-022-02063-8`
2. J. Nie and **S. Zhong**, Distributionally robust optimization with polynomial robust constraints, under review after minor revision, *J. Glob. Optim.*, (2024). `arXiv:2308.15591`
3. B. Rao, L. Yang, G. Zhou and **S. Zhong**, Robust approximation of chance constrained optimization with polynomial perturbation, *Comput. Optim. Appl.*, (2024). `doi.org/10.1007/s10589-024-00602-7`

1.3 Financial applications

Stochastic optimization is frequently applied in financial mathematics. My following works explore its applications in financial contexts like portfolio selection problems.

Corresponding paper

1. J. Nie, L. Yang, and **S. Zhong**, Stochastic polynomial optimization, *Optim. Methods. Softw.*, 35.2 (2020): 329-347. doi.org/10.1080/10556788.2019.1649672
2. L. Yang, Y. Yang and **S. Zhong**, Global optimization for the portfolio selection model with high-order moments, *J. Oper. Res. Soc. China*, (2023). doi.org/10.1007/s40305-023-00519-8
3. B. Rao, L. Yang, G. Zhou and **S. Zhong**, Robust approximation of chance constrained optimization with polynomial perturbation, *Comput. Optim. Appl.*, (2024). doi.org/10.1007/s10589-024-00602-7

2 Loss function optimization

There has been a significant trend for using large-scale datasets to train models in machine learning and artificial intelligence. Gradient-based methods are frequently employed for computational efficiency, but their convergence to global optimality is highly dependent on convexity. It is an interesting question to find a class of convenient nonconvex *loss functions* that have no spurious local minimizers in that a descent search method will always converge to a global optimizer. An intuitive example is the binary mean-squared loss function $f(x) = x^2(x-1)^2$, which has global optimizers $x = 0, 1$ and no other local optimizers. We generalize *simplicial loss function*

$$\circledast \quad f(x) = \sum_{i=1}^{n} x_i^2(x_i - a_i)^2 + \sum_{1 \le i < j \le n} x_i^2 x_j^2$$

in $x = (x_1, \ldots, x_n)$ and prove that $f(x)$ has no spurious mimizers for nonzero scalars $a_1, \ldots, a_n$. Up to proper linear transformations, the simplicial loss functions can be used to construct a general class of loss functions for finite sets with no spurious local minimizers. They have good performance in clustering and classification (see Figure 2 for illustration).

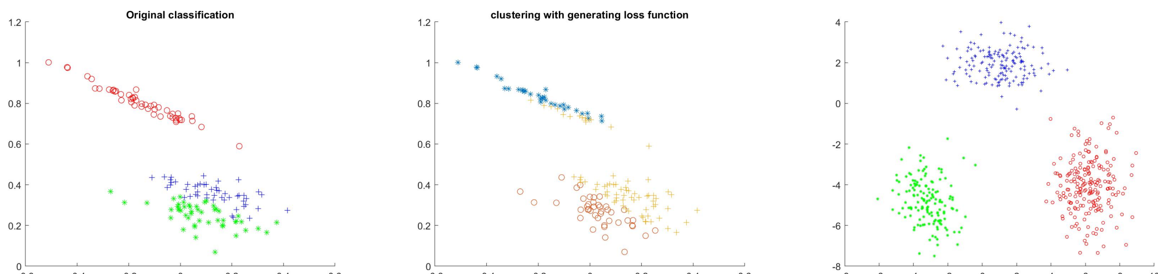

Figure 2: The left and middle figures respectively show the original classification and the trained classification with our loss functions. The right figure shows the clustering of our loss functions.

Corresponding paper

1. J. Nie and **S. Zhong**, Loss functions for finite sets, *Comput. Optim. Appl.*, 84 (2023): 421–447. doi.org/10.1007/s10589-022-00420-9
2. **S. Zhong** and J. Zhou, A new working set method for nonlinear inequality constrained minimization, *Preprint*, (2023). arXiv:2310:13199

3 Compute Nash and Stackelberg equilibria

3.1 Stakelberg games

The Stakelberg game aims to solve

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} & F(x,y) \\ \text{subject to} & h(x,y)\ge 0, \quad y\in S(x), \end{cases} \tag{3.1}$$

where $S(x)$ is the set of optimizer(s) of the lower level problem

$$(P_x): \begin{cases} \min\limits_{z\in\mathbb{R}^p} & f(x,z) \\ \text{subject to} & g(x,z)\ge 0 \end{cases} \tag{3.2}$$

Here F, f are scalar functions and h, g are function vectors. Let $\mathcal{F}$ denote feasible set of (3.1). Under constraint qualifications, every $y \in S(x)$ satisfies the Kurash-Kuhn-Tucker (KKT) conditions:

$$\nabla_z f(x,y) - G(x,y)^T\lambda = 0, \quad 0 \le g(x,y) \perp \lambda \ge 0.$$

where ∇_z denotes the partial gradient operator in z, $G(x,y)$ denotes the Jacobian matrix of g with respect to z, and $\perp$ denotes the complementarity relation. The auxiliary vector λ is the vector of *Lagrange multipliers* and the feasible (x,y) is called a KKT point. The Lagrange multiplier vector λ can be represented by rational functions, i.e., $\lambda = \lambda(x,y)$. Such $\lambda(x,y)$ are called *Lagrange multiplier expressions* (LMEs), which exist for general cases and have convenient expressions for linear constraints. Consider the special case when the rational inverse $G(x,y)^{-T}$ exists. Then we automatically obtain an LME $\lambda(x,y) = G(x,y)^{-T}\nabla_z f(x,y)$ and a semialgebraic KKT set $\mathcal{W}_1$ as follows:

$$\mathcal{W}_1 = \{(x,y)\in X \mid 0 \le g(x,y) \perp G(x,y)^{-T}\nabla_z f(x,y) \ge 0\}.$$

Clearly, $\mathcal{F} \subseteq \mathcal{W}_1$ and $\mathcal{F} = \mathcal{W}_1$ if f is convex in z and $\lambda(x,y)$ is a vector of polynomials. In this case, computing Stackelberg equilibria is equivalent to solving a polynomial optimization problem.

Suppose $(\hat{x},\hat{y})$ is a global minimizer of f_0 over $\mathcal{W}_1$. We can verify whether $(\hat{x},\hat{y}) \in \mathcal{F}$ by solving $(P_{\hat{x}})$ for an optimizer $\hat{z}$. If $f(\hat{x},\hat{z}) \ge f(\hat{x},\hat{y})$, then $\hat{y} \in S(\hat{x})$. Hence $(\hat{x},\hat{y})$ is a Stackelberg equilibrium. Otherwise, we need to find another candidate solution. In computational practice, this requires us to build a hierarchy $\{\mathcal{W}_k\}$ of (x,y) such that

$$\mathcal{F} \subseteq \cdots \subsetneqq \mathcal{W}_{k+1} \subsetneqq \mathcal{W}_k \quad \text{and} \quad \inf\{f_0(x,y) \text{ on } \mathcal{W}_k\} \xrightarrow{k\to\infty} \inf\{f_0(x,y) \text{ on } \mathcal{F}\}$$

This can be done by an exchange method with feasible extensions. Under some continuity assumptions, this hierarchy will either converge to an equilibrium or detect its nonexistence.

Corresponding paper

1. J. Nie, L. Wang, J. Ye, and **S. Zhong**, A Lagrange multiplier expression method for bilevel polynomial optimization. *SIAM J. Optim.*, 31.3 (2021): 2368-2395. doi.org/10.1137/20M1352375

2. J. Nie, J. J. Ye and **S. Zhong**, PLMEs and Disjunctive Decompositions for Bilevel Optimization, *Prepint*, (2023). arXiv:2304.00695

3.2 Generalized Nash equilibrium problems

The generalized Nash equilibrium problem (GNEP) is to determine a tuple of strategies $u = (u_1, \dots, u_N)$ such that each u_i minimizes the optimization problem

$$F_i(u_{-i}): \begin{cases} \min\limits_{x_i \in \mathbb{R}^{n_i}} & f_i(x_i, u_{-i}) \\ \text{subject to} & g_i(x_i, x_{-i}) \geq 0 \end{cases} \tag{3.3}$$

for given $u_{-i} := (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N)$, where f_i is a rational function and g_i is a rational. Such u is called a *generalized Nash equilibrium* (GNE). Under some constraint qualifications, every GNE satisfies the the KKT conditions. Let $\mathcal{K}$ denote the KKT set of the GNEP. Suppose every $F_i(x_{-i})$ is convex optimization, then $\mathcal{K}$ is the exact set of GNEs. Otherwise, $\mathcal{K}$ may contain some non-GNE point. Here is an example.

Example 3.1. *Consider the 2-player GNEP (here $e = (1,1)^T$)*

$$\begin{array}{cl|cl} \min\limits_{x_1 \in \mathbb{R}^2} & (x_{1,1} - x_{1,2})x_{2,1}x_{2,2} - x_1^T x_1 & \min\limits_{x_2 \in \mathbb{R}^2} & 3(x_{2,1} - x_{1,1})^2 + 2(x_{2,2} - x_{1,2})^2 \\ \textit{subject to} & 1 - e^T x_1 - e^T x_2 \geq 0,\ x_1 \geq 0, & \textit{subject to} & 2 - e^T x_1 - e^T x_2 \geq 0,\ x_2 \geq 0. \end{array}$$

It is easy to verify that $\hat{x} = (\hat{x}_1, \hat{x}_2)$ with $\hat{x}_1 = \hat{x}_2 = (0,0)$ is a KKT point but not a GNE. Indeed, when $i = 1$, one can find $v_1 = (1,0)$ such that $f_1(v_1, \hat{x}_2) < f_1(\hat{x}_1, \hat{x}_2)$. Interestingly, for the GNE $u = (u_1, u_2)$ with $u_1 = u_2 = (0.5, 0)$, v_1 is not feasible for $F_1(u_2)$ and that $f_1(v_1, u_2) - f_1(u_1, u_2) = -0.75 < 0$.

Assume for a given triple (u, i, v_i) with $u \in \mathcal{K}$, $i \in [N]$ and $v_i \in X_i(u_{-i})$, there exists a feasible extension p_i of v_i at u such that

$$v_i = p_i(u), \quad p_i(x) \in X_i(x_{-i}) \quad \forall x \in \mathcal{K}. \tag{3.4}$$

Then p_i can be used to preclude the non-GNE points with $f_i(p_i(x), x_{-i}) - f_i(x_i, x_{-i}) \geq 0$. This is because (3.4) guarantees $f_i(x_i, x_{-i}) \leq f_i(p_i(x), x_{-i})$ for every GNE x. The feasible extension always exists when $\mathcal{K}$ is finite. It has a universal expression when $X_i(x_{-i})$ is given by boxed, simplex or ball constraints. Then (3.3) can be solved by a hierarchy of rational optimization problems

$$(P_k): \begin{cases} \min\limits_{x \in \mathcal{K}} & x^T \Theta x \\ \text{subject to} & f_i(p_i^{(j)}(x), x_{-i}) - f_i(x_i, x_{-i}) \geq 0, \\ & i = 1, \dots, N,\ j = 1, \dots, k-1. \end{cases}$$

In the above, Θ is a generic positive definite matrix, and each $p_i^{(j)}$ is the feasible extension produced from (P_j). Assume the feasible extension always exists and there are only finitely many non-GNE points in $\mathcal{K}$. Then solving (P_k) either returns a GNE or detects its nonexistence when k is large enough.

Corresponding paper

1. J. Nie, X. Tang, and **S. Zhong**, Rational generalized Nash equilibrium problems, *SIAM J. Optim.*, 33.3 (2023): 1587-1620. doi.org/10.1137/21M1456285

2. J. Choi, J. Nie, X. Tang and **S. Zhong**, Generalized Nash equilibrium problems with quasi-linear constraints, *Preprint*, (2024). arXiv:2405.03926

4 Plan for the Future

4.1 Multi-stage stochastic optimization

Multi-stage stochastic models, like two-stage distributionally robust optimization problems, are highly valued in data science because they offer a structured way to make decisions under uncertainty over time. However, such problems are very difficult to solve since their objective functions and feasible sets are implicitly determined by value-functions and inf-constraints. This projects focuses on multi-stage stochastic optimization defined by polynomials. We propose to approximate value-functions by polynomials and feasible sets by semidefinite representable truncated moment cones. The following questions are going to be investigated in this project:

1. How do we get an efficient polynomial approximation for the implicitly-defined objective function?

2. How can we approximate feasible sets under different kinds ambiguity sets (i.e., moment-based, Wasserstein-based, mixed)?

Corresponding paper

1. S. Zhang and **S. Zhong**, Data-driven distributionally robust optimization with Wasserstein-based ambiguity, *in preparation*.

4.2 Optimization with infinite constraints

Feasible sets with infinity constraints are very difficult to characterize computationally. It is even a big challenge to approximate them accurately. Feasible extensions are promising for solving optimization with this kind of feasible sets. Such optimization problems include generalized semi-finite programming, GNEPs and bilevel optimization. The following questions are going to be investigated in this project:

1. How do we get feasible extensions for more general feasible sets? What are convenient conditions for their existence and computation?

2. How can we get higher quality initial constraining sets for feasible sets?

Corresponding paper

1. X. Hu, J. Nie and **S. Zhong**, Generalized semi-infinite programming of polynomials, *in preparation*.

4.3 Tensor optimization

Symmetric tensors are associated with homogeneous polynomials. This relationship can be used in tensor decomposition and tensor recovery problems. I aim to adapt my approach to tackle optimization problems involving high dimensional data of tensors.

Corresponding paper

1. J. Nie, X. Tang, Z. Yang and **S. Zhong**, Dehomogenization for completely positive tensors, *NACO*, 13.2 (2023): 340–363. 10.3934/naco.2022037