RESEARCH STATEMENT

SUHAN ZHONG

Optimization has beautiful theories and broad applications. My research interests include stochastic optimization, bilevel optimization, generalized Nash equilibrium problems, loss functions for data clustering, and data science optimization. My goals are to develop new theories in optimization and apply these theories in applications.

1. DISTRIBUTIONALLY ROBUST OPTIMIZATION

The distributionally robust optimization is defined by

\[ \min_{x \in X} f(x) \quad \text{s.t.} \quad \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x, \xi)] \geq 0, \]

where \( f \) is a convex function in the decision variable \( x \in \mathbb{R}^n \) and \( h(x, \xi) = (Ax + b)^T [\xi]_d \) is a polynomial in the random variable \( \xi \in \mathbb{R}^p \) (note \([\xi]_d\) is the degree-\( d \) monomial vector of \( \xi \)). Suppose \( x \) is constrained in \( X = \{x : Gx \leq h\} \) and \( \xi \) admits a distribution of measure \( \mu \) supported in \( S \). Given a moment bound \( Y \) for \( \xi \) of order \( d \), we write that \( \mathcal{M} = \{\mu : \text{supp}(\mu) \subseteq S, \mathbb{E}_\mu([\xi]_d) \in Y\} \). Denote the conic hull

\[ K := \text{cone}(Y \cap \{\mathbb{E}_\mu([\xi]_d) : \text{supp}(\mu) \subseteq S\}). \]

Since the inf expectation constraint in (1.1) equals \( Ax + b \geq 0 \) for all \( y \in K \), the optimization (1.1) can be equivalently reformulated as

\[ \min_{x \in X} f(x) \quad \text{s.t.} \quad Ax + b \in K^*, \]

where \( K^* \) stands for the dual cone of \( K \). Assume \( S \) and \( Y \) are both closed and compact. Denote by \( \mathbb{R}[\xi]_d \) the set of polynomials in \( \xi \) of degree at most \( d \), and let \( \mathcal{P}_d(S) := \{p \in \mathbb{R}[\xi]_d : p(\xi) \geq 0, \forall \xi \in S\} \). By (1.2), we can express \( K^* \) as

\[ K^* = Y^* + \mathcal{P}_d(S). \]

For instance, consider \( \xi \) is univariate, \( d = 4 \), \( Y \) is the hypercube \([0, 1]^5\) and \( S = [a_1, a_2] \), then \( K \) can be expressed by the constraints

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

1
\[(y_0, y_1, y_2, y_3, y_4) \geq 0.\]

And \(K^*\) can be given by semidefinite programming constraints dual to the above. When \(\xi\) is multi-variate, \(K^*\) does not have a convenient characterization for computation. Instead, it can be approximated by quadratic modules, under the assumption that \(S\) is a semialgebraic set,

\[S = \{\xi \in \mathbb{R}^p : g(\xi) \geq 0\}, \quad g = (g_1, \ldots, g_m), \forall g_i \in \mathbb{R}[\xi].\]

Denote the truncated sum-of-squares (SOS) polynomial cone by \(\Sigma[\xi]_{2k} := \{p_1^2 + \cdots + p_r^2 | \deg(p_i) \leq k, i = 1, \ldots, r\}\).

For \(k \geq d/2\), the \(k\)th SOS approximation of (1.1) is

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) \\
\text{s.t.} & \quad Ax + b \in Y^* + \sum_{i=0}^{m} g_i \cdot \sigma_i, \\
& \quad Gx \leq h, g_0 = 1, \sigma_i \in \Sigma[\xi]_{2k-\deg(g_i)}; \\
& \quad i = 0, 1, \ldots, m.
\end{align*}
\]

When \(f\) is linear in \(x\), (1.5) is a conic optimization problem that can be solved directly by semidefinite programming. More generally, if \(f\) is a polynomial, (1.5) can still be solved efficiently with Moment-SOS relaxations. Under some general conditions, (1.1) and (1.5) have the same optimal value and optimizer(s) when \(k\) is large enough.

**Theorem 1.1.** Assume \(f\) is polynomial in \(x \in \mathbb{R}^n\). Suppose \(x^{(k)}\) is the minimizer of (1.5) with the order \(k\). Under some general conditions, \(x^{(k)}\) is the minimizer of (1.1) when \(k\) is big enough.

2. **Generalized Nash Equilibrium Problems**

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

\[
F_i(u_{-i}) : \begin{cases}
\min_{x_i} & \quad f_i(x_i, u_{-i}) \\
\text{s.t.} & \quad x_i \in X_i(u_{-i}),
\end{cases}
\]

for given \(u_{-i} := (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_N)\). Such \(u\) is called a generalized Nash equilibrium (GNE). Let \(x = (x_1, \ldots, x_N)\) and \(x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)\). Assume every \(f_i(x_i, x_{-i})\) and \(X_i(x_{-i})\) are given by rational functions, i.e.,

\[X_i(x_{-i}) = \{x_i \in \mathbb{R}^{n_i} : g_i(x_i, x_{-i}) \geq 0\},\]

where \(g_i\) is a rational vector. Under some constraint qualifications, every GNE satisfies the the KKT conditions. Suppose for each critical pair \((x_i, \lambda_i)\) of \(F_i(x_{-i})\), there is a tuple of rational functions \(\tau_i\) in \(x\)
such that \( \lambda_i = \tau_i(x) \). Then the KKT set of (2.1) can be given by (note \([N] = \{1, \ldots, N\}\))

\[
\mathcal{K} = \left\{ x \in \mathbb{R}^n \left| \begin{array}{l}
\nabla_{x_i} f_i(x) - \nabla_{x_i} g_i(x) \tau_i(x) = 0, \\
g_i(x) \perp \tau_i(x), \tau_i(x) \geq 0, i \in [N], \\
x = (x_1, \ldots, x_N), x_i \in X(x_{-i})
\end{array} \right. \right\}.
\]

Such \( \tau_i \) is called a rational Lagrange multipliers expression (LME), which always exists if each \( g_{i,j} \) is not constantly zero. 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, say, \( u = (u_1, \ldots, u_N) \). Then there exist \( i \in [N] \) and \( v_i \in X_i(u_{-i}) \) such that

\[
f_i(v_i, u_{-i}) - f_i(u_i, u_{-i}) < 0.
\]

However, if \( x = (x_i, x_{-i}) \) is a GNE and \( v_i \in X_i(x_{-i}) \), then \( x \) must satisfy

\[
f_i(v_i, x_{-i}) - f_i(x_i, x_{-i}) \geq 0.
\]

But the inequality (2.4) may not hold for every GNE due to the interactions among all \( x_i \) and \( F_i(x_{-i}) \).

**Example 2.1.** Consider the 2-player GNEP

\[
\begin{align*}
\min_{x_1, x_2} & (x_{1,1} - x_{1,2})x_{2,1}x_{2,2} - x_1^T x_1 & \quad \min_{x_2} & 3(x_{2,1} - x_{1,1})^2 + 2(x_{2,2} - x_{1,2})^2 \\
\text{s.t.} & 1 - e^T x \geq 0, x_1 \geq 0, & \text{s.t.} & 2 - e^T x \geq 0, x_2 \geq 0.
\end{align*}
\]

Consider the GNE \( x^* = (x_1^*, x_2^*) \) and the non-GNE \( u = (u_1, u_2) \in \mathcal{K} \), i.e.,

\[
x_1^* = x_2^* = (0.5, 0), \quad u_1 = u_2 = (0, 0).
\]

When \( i = 1 \), (2.3) is satisfied with \( v_1 = (1, 0) \). However, \( v_1 \) is not feasible for \( F_1(x_2^*) \) and that

\[
f_1(v_1, x_2^*) - f_1(x_1^*, x_2^*) = -0.75 < 0.
\]

The inequality (2.4) does not hold for \( x^* \).

Motivated by Example 2.1, 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

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

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 (2.5) 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. With rational Lagrange multiplier expressions and feasible extensions, (2.1) can be solved by a hierarchy of rational optimization problems

\[ (P_k) : \begin{cases} \min_{x \in \mathcal{K}} x^T \Theta x \\ \text{s.t. } f_i(p_i^{(j)}(x), x_{-i}) - f_i(x_i, x_{-i}) \geq 0, \\ i = 1, \ldots, N, j = 1, \ldots, k - 1. \end{cases} \]

In the above, \( \Theta \) is a generic positive definite matrix, and each \( p_i^{(j)} \) is the feasible extension produced from \((P_j)\). Note that every \((2.6)\) is a relaxation of (2.1). If \((P_k)\) is infeasible for some order \( k \), then we detect the nonexistence of GNEs. Conversely, if GNE does not exist, then \((P_k)\) must be infeasible when \( k \) is large enough, under some general assumptions. But if there exists a GNE in \( \mathcal{K} \), then \((P_k)\) must be feasible for all \( k \in \mathbb{N} \).

**Theorem 2.2.** For the rational GNEP (2.1), assume the feasible extension always exists and there are only finitely many non-GNE points in \( \mathcal{K} \). Suppose \( u^{(k)} \) is the minimizer of (2.1) with the order \( k \). Then either (2.1) has no GNE or \( u^{(k)} \) is a GNE when \( k \) is large enough.

For Example 2.1, we get the GNE with \( k = 1 \).

### 3. Bilevel polynomial optimization

The bilevel optimization problem is

\[ (P_x) : \begin{cases} \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^p} F(x, y) \\ \text{s.t. } g_i(x, y) \geq 0 (i \in \mathcal{I}_1), \\ y \in S(x), \end{cases} \]

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

\[ (P_x) : \begin{cases} \min_{z \in \mathbb{R}^p} f(x, z) \\ \text{s.t. } g_i(x, z) \geq 0 (i \in \mathcal{I}_2). \end{cases} \]

Let \( Z(x) \) denote the feasible set of \((P_x)\), and denote

\[ \mathcal{U} = \{(x, y) : g_i(x, y) \geq 0 (i \in \mathcal{I}_1 \cup \mathcal{I}_2)\}. \]

When \( F, f, h_i, g_i \) are all polynomials, (3.1) is called a bilevel polynomial optimization problem. Assume each \( y \in S(x) \) satisfies the KKT conditions for the lower level optimization, and corresponds to a rational
LME $\lambda(x, y)$. Then (3.1) can be relaxed into

$$\min_{(x, y) \in U} F(x, y)$$

s.t. $\nabla_z f(x, y) - \sum_{i \in I_2} \lambda_i(x, y) \nabla_z g_i(x, y) = 0,$

$$\lambda_i(x, y) \geq 0, \lambda_i(x, y) g_i(x, y) = 0 \quad (i \in I_2).$$

Denote by $U_0$ the feasible set of (3.3). Assume for every $(\hat{x}, \hat{y}) \in U_0$ and every $\hat{z} \in S(\hat{x})$, there exists a tuple $q = (q_1, \ldots, q_p)$, $q_i \in \mathbb{R}[x, y]$ such that

$$q(\hat{x}, \hat{y}) = \hat{z}, \quad q(x, y) \in Z(x) \quad \forall (x, y) \in U.$$  

We call such $q$ a polynomial extension of $\hat{z}$ at $(\hat{x}, \hat{y})$. Suppose $(Q_0)$ is solvable with a minimizer $(x(0), y(0))$. It is a minimizer for (3.1) if

$$f(x(0), z) - f(x(0), y(0)) \geq 0, \quad \forall z \in S(x(0)).$$

This can be verified by solving the lower level optimization with $x = x(0)$. If (3.5) does not hold, a tighter relaxation can be built with the feasible extension of polynomials. By repeating this process, we can get a hierarchy of polynomial optimization relaxations for (3.1),

$$\min_{(x, y) \in U_k} F(x, y)$$

s.t. $(x, y) \in U_k$. 

Suppose $(Q_{k-1})$ is solvable with a minimizer $(x^{(k-1)}, y^{(k-1)})$. In (3.6),

$$U_k := \{(x, y) \in U_{k-1} \mid f(x, q^{(k-1)}(x, y)) - f(x, y) \geq 0\},$$

and $q^{(k-1)}$ is the polynomial extension of some $z^{(k-1)} \in S(x^{(k-1)})$. Since each $(Q_k)$ is a relaxation of (3.1), we can detect the nonexistence of solutions for (3.1), if (3.6) is infeasible for some order $k$. Suppose every $(Q_k)$ is solvable, under some general conditions, there is an asymptotic convergence of $(x^{(k)}, y^{(k)})$ to the minimizer of (3.1). It is worthy to that that this convergence is usually finite in practice.

**Theorem 3.1.** Assume (3.1) is solvable and the polynomial extension always exists. Let $(x^{(k)}, y^{(k)})$ denote the minimizer of $(Q_k)$. Under some general conditions, the $(x^{(k)}, y^{(k)})$ or its limit point is a minimizer of (3.1).

For instance, consider the bilevel optimization problem

$$\min_{x, y \in \mathbb{R}^1} (x - 2)^2 + (y + 3)^2$$

s.t. $x \geq 1, y \geq 2,$

$$y \in \arg\min_{z \in \mathbb{R}^1} \{(z - 3)^2 : z^2 - x \geq 0\}.$$ 

It has the rational LME $\lambda(x, y) = y(y - 3)/x$. We solve the initial relaxation $(Q_0)$ for the minimizer $(x(0), y(0)) = (2.0000, 3.0000)$. It is
easy to verify that \((x^{(0)}, y^{(0)})\) is the global minimizer of the bilevel optimization.

4. Stochastic Polynomial Optimization

The stochastic polynomial optimization is

\[
\min_{x \in \mathbb{R}^n} f(x) := \mathbb{E}[F(x, \xi)] \quad \text{s.t.} \quad g(x) \geq 0,
\]

where \(F\) is a polynomial in \((x, \xi)\) and \(g = (g_1, \ldots, g_m)\) is tuple of polynomials in \(x\). Since the expectation function \(f\) is usually hard to compute directly, we approximate it by sample averages. Suppose \(\xi^{(1)}, \ldots, \xi^{(N)}\) are some identically independently distributed samples. The sample average approximation (SAA) is

\[
f_N(x) := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^{(i)}).
\]

The \(f_N\) converges to \(f\) with probability one. However, the minimizer of \(f_N\) in \(\{x : g(x) \geq 0\}\) may not be close to the optimizer of (4.1), even for cases when \(N\) is large. This is because the image of \(f\) may be sensitive with its coefficient. For example, suppose \(f(x) = 0.00001 x^2\) and \(g = \emptyset\), then a little sampling noise, i.e., \(-0.0001 x^2\), can make \(f_N\) be unbounded from below.

Assume (4.1) is solvable. To address such concern, we introduce a 2-norm perturbation to SAA. Consider

\[
\min_{x \in \mathbb{R}^n} f_N(x) + \epsilon \|[x]_{2d}\|_2 \quad \text{s.t.} \quad g(x) \geq 0,
\]

where \(\epsilon > 0\) is a parameter and \(2d\) is the smallest even integer satisfying \(2d \geq \max\{\deg(f_N), \deg(g)\}\). Under some general assumptions, when \(\epsilon\) is sufficiently small, the optimal solution of (4.2) will converge to the optimizer of (4.1) with probability one. The optimization (4.2) can be solved by a moment relaxation

\[
\begin{aligned}
\min_{x \in \mathbb{R}^n} & \quad \langle f_N, y \rangle + \epsilon \|y\|_2 \\
\text{s.t.} & \quad M_d[y] \succeq 0, \quad L_{g_i}^{(d)}[y] \succeq 0 \quad (i \in [m]), \\
& \quad y_0 = 1, \quad y \in \mathbb{R}^{2d}.
\end{aligned}
\]

In the above, \(M_d[y]\) (resp., \(L_{g_i}^{(d)}[y]\)) denotes the \(d\)th order moment matrix (resp., localizing matrix with \(g_i\)) for \(y\). Precisely, suppose
n = d = 2, then \( y = (y_{00}, y_{10}, y_{01}, y_{20}, y_{11}, y_{02}) \) is the truncated multi-sequence for \( x = (x_1, x_2) \). For \( g = 1 - x_1 \), we have

\[
M_2[y] = \begin{bmatrix}
y_{00} & y_{10} & y_{01} \\
y_{10} & y_{20} & y_{11} \\
y_{01} & y_{11} & y_{02}
\end{bmatrix}, \quad L^{(2)}_g[y] = [y_{00} - y_{10}].
\]

**Theorem 4.1.** Suppose \( \epsilon > 0 \) and \( y^* \) is a minimizer of (4.3). Then (4.3) is a tight relaxation of (4.2) if and only if \( \text{rank}M_d[y^*] = 1 \). When \( \text{rank}M_d[y^*] = 1 \), \( u = (y^*_{e_1}, \ldots, y^*_{e_n}) \) is a minimizer of (4.2).

### 5. Loss functions for finite sets

Let \( W \) be a finite set in \( \mathbb{R}^n \). A function \( f \) in \( x = (x_1, \ldots, x_n) \) is said to be a loss function for \( W \) if \( f \) is nonnegative and \( f(x) = 0 \) if and only if \( x \in W \). In particular, we say \( f \) has no spurious local minima if all local minimizers of \( f \) are also global minimizers. Suppose \( W \) is a general finite set which is given implicitly, i.e., \( W \) is given by some approximation \( V \). It is an open question that how to find a convenient loss function for \( W \). Furthermore, when does such loss function has no spurious local minima?

Assume the loss function \( f \) is in form of SOS polynomials, i.e.,

\[
(5.1) \quad f = p_1^2 + \cdots + p_m^2, \quad p_1, \ldots, p_m \in \mathbb{R}[x]
\]

Then \( f \) is a loss function for \( W \) if and only if

\[
(5.2) \quad W = \{x \in \mathbb{R}^n : p_1(x) = \cdots = p_m(x) = 0\}.
\]

It implies that \( B = \{p_1, \ldots, p_m\} \) is a basis for the vanishing ideal \( I(W) = \{q \in \mathbb{R}[x] : q(x) = 0, \forall x \in W\} \). Such \( f \) has the simplest expression when \( B \) is a minimal basis with the lowest degree. And this kind of \( B \) can be solved from a polynomial system parametrized by \( W \), say,

\[
H(p_1(x), \ldots, p_m(x)) = 0, \quad x \in W.
\]

Note that \( f \) in (5.1) has no spurious local minima when \( W \) is specified by certain points. Denote the vector function on \( p = (p_1, \ldots, p_m) \),

\[
H_W(p) = \sum_{x \in W} H(p_1(x), \ldots, p_m(x)).
\]

Suppose \( W \) has a priori cardinality \( r \), and is given implicitly by some approximation \( V \). Based on the one-to-one relation between \( W \) and the minimal basis of \( I(W) \). We can determine the loss function for \( W \).
by solving the following optimization problem

\begin{equation}
\begin{aligned}
\min_p & \quad \|H_V(p)\|^2 \\
\text{s.t.} & \quad H_Z(p) = 0, \ Z = \{z_1, \ldots, z_r\} \subseteq \mathbb{R}^n, \\
\ & \quad p = (p_1, \ldots, p_m), \ p_i \in \mathbb{R}[x], \ \forall p_i \in \mathbb{R}[x].
\end{aligned}
\end{equation}

**Theorem 5.1.** Assume \(V\) is an approximation of \(W\). Suppose \(p^*\) is the minimizer of (5.3). Under some general conditions, \(W^* := \{x \in \mathbb{R}^n : p^*(x) = 0\}\) is a close neighborhood of \(W\), i.e., \(\|W - W^*\| \leq \delta\) for a small parameter \(\delta > 0\).

6. **ONGOING & FUTURE WORK**

There is much future work to do with the above topics and their extensions or applications. Here are some problems I am particularly interested in:

- The multi-stage stochastic optimization problem.
- The stochastic optimization in portfolio selection.
- The bilevel optimization in machine learning.
- The loss functions design and application.
- The generalization of techniques like Lagrangian multiplier expression and feasible extension.

**REFERENCES**


