Moment Convergence Rate in Stochastic Optimization
Student: Linwei Xin
Advisors: David Goldberg & Alexander Shapiro

1 Introduction

Recently, there has been significant interest in constructing more robust solutions to stochastic optimization problems. In particular, consider the following general stochastic optimization problem:

$$\min_{x \in \mathcal{X}} \mathbb{E}_{Q_0}\left[f(x,\xi)\right], \tag{1.1}$$

where $\mathcal{X}$ is the decision set and Q_0, the distribution of ξ, is a probability distribution supported on $[0,1]$. This kind of problem can be solved numerically or (in special cases) in closed form if we know the exact distribution of Q_0. Unfortunately, in practice one rarely knows the exact distribution Q_0. Instead, one often has only partial information about Q_0, which may include limited distributional information such as a few moments or quantiles.

In a seminal paper, Scarf proposed and solved a "robust version" of Problem (1.1). More precisely, suppose that instead of knowing Q_0 exactly, one only knew the first two moments of Q_0, and wanted to optimize against an adversary who could pick a "worst-case" distribution subject to those moment constraints. Then one is faced with the min-max optimization problem

$$\min_{x \in \mathcal{X}} \max_{Q:\int_0^1 x^k dQ(x)=\mu_k, k=1,2} \mathbb{E}_{Q}\left[f(x,\xi)\right].$$

In [10], the author solves this problem for the special case in which f is a 2-piece piece-wise linear function, and interprets his results in terms of how one would optimally manage an inventory.

Of course, one might have access to more than just two moments. Furthermore, one would suspect that as one is given access to more moments:

- the value of the robust optimization problem converges to that of the "full-information" problem (1.1);
- the computational difficulty of solving the robust problem exactly increases.

We propose a robust optimization model against a worst scenario as follows:

$$\min_{x \in \mathcal{X}} \max_{Q \in \mathfrak{M}_n} \mathbb{E}_{Q}\left[f(x,\xi)\right], \tag{1.2}$$

where $\mathfrak{M}_n$ is a set of probability measures supported on $[0,1]$ with fixed the first n moments ($\int x^k dQ(x) = \int x^k dQ_0(x) := \mu_k$, $k = 0,\dots,n$). We note that a discrete version of (1.2), called *Discrete Moment Problem*, was proposed by Préopa [8], and shown to have applications in economics and finance [7].

Although it is not difficult to show that the optimal value of (1.2) eventually converges to the optimal value of (1.1) as $n \to \infty$, it is an interesting and challenging task to study the rate of convergence.

We note that the structure of the set $\mathfrak{M}_n$ is not well understood, except for some general results, including the fact that $\mathfrak{M}_n$ is convex, and its extremal distributions are those probability measures supported on at most $n+1$ points (cf. Theorems 1.1 and 6.1 in [5]). In this work, we propose to study the relevant trade-off between accuracy and complexity, by asking

(A). What is the convergence rate for the optimal objective function value of (1.2) as $n \to \infty$?

(B). What is the convergence rate for the optimal choice of x in (1.2) as $n \to \infty$?

(C). How can we devise efficient approximation algorithms when many moments are given?

2 Research Plan

Note that if $f(x,\xi)$ was a degree-n polynomial in ξ, then the first n moments would uniquely pin down its value, reducing (1.2) to a one-dimensional (in x) optimization problem. We propose to approximate $f(x,\xi)$ by a degree-n polynomial in ξ, and formally bound the error. The following well-known polynomial approximation result is due to Jackson (cf. [6], [9]): if f is a continuous function on $[a,b]$, then

$$E_n(f;[a,b]) \le 6\omega\left(\frac{b-a}{2n}; f; [a,b]\right), \text{ where}$$

$$E_n(f;[a,b]) := \inf\left\{\|f-p\|_{L_\infty[a,b]} : \text{ p is a real polynomial function of degree at most n}\right\},$$
$$\omega(\delta; f; [a,b]) := \sup_{|x_1-x_2|\le\delta,\ x_1,x_2\in[a,b]} |f(x_1)-f(x_2)|.$$

It is not difficult to find an upper bound of the convergence rate for optimal values directly from Jackson's Approximation. For example, if $f(x,\xi)=|x-\xi|$, the rate of convergence is no worse than $O(\frac{1}{n})$. However, the difficulty is to investigate whether this upper bound is tight (Question (A)). In other words, we would like to find a distribution Q_0, an objective function $f(x,\xi)$, and a sequence $Q_n \in \mathfrak{M}_n$ such that this rate can be achieved. Two possible approaches are proposed as below, which we illustrate through the special case $f(x,\xi)=|x-\xi|$ that is typically used in inventory management (cf. [10], [12]), for the sake of concreteness.

1). Once we fix $x\in\mathcal{X}$, the dual problem of $\max_{Q\in\mathfrak{M}_n}\mathbb{E}_Q[|x-\xi|]$ is

$$\begin{array}{ll} \min\limits_{y\in\mathbb{R}^{n+1}} & \sum_{i=0}^n y_i\mu_i \\ \text{s.t.} & x-\xi \le \sum_{i=0}^n y_i\xi^i,\ \ \forall\xi\in[0,1] \\ & \xi-x \le \sum_{i=0}^n y_i\xi^i,\ \ \forall\xi\in[0,1]. \end{array}$$

It can be proved that there is no duality gap between the primal and the dual problems (cf. Corollary 3.1 in [11]). Moreover, Proposition 3.1 in [2] states that $\sum_{i=0}^n y_i\xi^i \ge 0$ for all $\xi\in[0,1]$ if and only if there exists a positive semidefinite matrix $X=[x_{ij}]_{i,j=0,\ldots,n}$ such that

$$\sum_{i,j:i+j=2\ell-1} x_{ij}=0,\ \ell=1,\ldots,n \quad \text{and} \quad \sum_{i,j:i+j=2\ell} x_{ij} = \sum_{i=0}^{\ell} y_i\binom{n-i}{\ell-i},\ \ell=0,\ldots,n$$

Hence it is possible to reformulate $\min_{x\in\mathcal{X}}\max_{Q\in\mathfrak{M}_n}\mathbb{E}_Q[|x-\xi|]$ as an equivalent semidefinite program, which can be efficiently solved by interior point methods (cf. [1]). Theoretically, we hope to take advantage of this equivalent formulation to prove that Jackson's bound can be attained, and is thus tight.

2). A different approach is based on the following observation:

$$\min_{x\in\mathcal{X}}\max_{Q\in\mathfrak{M}_n}\mathbb{E}_Q[|x-\xi|] \ge \max_{Q\in\mathfrak{M}_n}\min_{x\in\mathcal{X}}\mathbb{E}_Q[|x-\xi|] = \max_{Q\in\mathfrak{M}_n}\mathbb{E}_Q[|\text{median}(Q)-\xi|] \tag{2.1}$$

If we restrict ourselves to those distributions supported on $\left\{\frac{i}{n}\right\}_{i=0}^n$, and fix the value of the median m (e.g. $\frac{1}{2}$), then the optimal value of the following linear program gives a lower bound of the optimal value of (2.1):

$$\begin{array}{ll} \max\limits_{p\in\mathbb{R}_+^{n+1}} & \sum_{i=0}^n p_i\left|m-\frac{i}{n}\right| \\ \text{s.t.} & \sum_{i=0}^n(\frac{i}{n})^k p_i = \mu_k,\ \forall\ k=0,\ldots,n; \\ & \sum_{i=0}^{\lfloor nm\rfloor} p_i \ge .5,\ \ \sum_{i=\lceil nm\rceil}^n p_i \ge .5 \end{array}$$

We would like to show tightness of Jackson's bound by studying this linear program (or its dual).

If we could solve Question (A), hopefully the methodology would enlighten us on solving Question (B), too. We have preliminary results about finding a general upper bound of the convergence rate for optimal solutions, and believe that we can make progress on either proving tightness or improving the bound.

Finally, we would like to address Question (C): finding an efficient approximation algorithms when many moments are given. Fortunately, our approach to the problem yields a natural approximation algorithm, in the following sense. It follows from Jackson's approximation that ANY CANDIDATE DISTRIBUTION from M_n will yield nearly the same value as in the "full-information" problem (1.1). Thus one can simply solve problem (1.1), for which many efficient algorithms are known, to get a suitable approximation.

Alternatively, one could actually construct the "best" low-degree polynomial approximation (using, for example, the technique given in [9]), which reduces the optimization to a one-dimensional polynomial minimization problem, for which many techniques are known, e.g. semidefinite programming (cf. [3], [4]).

References

[1] Ben-Tal, A., A. Nemirovski. 2001. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MPS/SIAM Series on Optimization, SIAM, Philadelphia.

[2] Bertsimas, D., I. Popescu. 2005. *Optimal inequalities in probability theory: A convex optimization approach*. SIAM Journal of Optimization. **15**(3) 780-804.

[3] Thomas, R. Forthcoming. *Polynomial optimization, sums of squares and applications*. G. Blekherman, P. Parrilo, R. Thomas. eds Semidefinite Optimization and Convex Algebraic Geometry. MOS-SIAM Series on Optimization.

[4] E. de Klerk, G. Elabwabi. D. den Hertog. April 2006. *Optimization of univariate functions on bounded intervals by interpolation and semidefinite programming*. CentER Discussion paper 2006-26. Tilburg University, The Netherlands.

[5] Karlin, S., W. J. Studden. 1966. *Tchebycheff Systems: with Applications in Analysis and Statistics*. Interscience Publishers. New York.

[6] Meinardus, G. 1967. *Approximations of Functions: Theory and Numerical Methods*. New York: Springer-Verlag.

[7] Mádi-Nagy, G., A. Préopa. 2011. *Bounding Expectations of Functions of Random Vectors with Given Marginals and some Moments: Applications of the Multivariate Discrete Moment Problem*. Mathematical Inequalities and Application **15** 101-122.

[8] Préopa, A. 1990. *The discrete moment problem and linear programming*. Discrete Applied Mathematics. **27** 235-254

[9] Rivlin, T. J. 1969. *An Introduction to the Approximation of Functions*. Waltham, Massachusetts: Blaisdell Publishing Co.

[10] Scarf, H. 1958. *A min-max solution of an inventory problem*. K. Arrow, ed Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, CA, 201–209.

[11] Shapiro, A. 2001. *On duality theory of conic linear problems*. In: Miguel A. Goberna and Marco A. Lopez, eds Semi-Infinite Programming: Recent Advances. Kluwer Academic Publishers 135–165.

[12] Shapiro, A., D. Dentcheva and A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.