

数据驱动的分布式鲁棒随机二次规划的收敛性分析

张东东*, 韩有攀*, 田孟昊

西安工程大学理学院, 陕西 西安

收稿日期: 2021年12月17日; 录用日期: 2022年1月6日; 发布日期: 2022年1月20日

摘要

本文研究数据驱动的Wasserstein模糊集下分布式鲁棒随机二次规划的收敛性问题。首先, 我们建立了目标函数的逐点Lipschitz性质。接着, 当样本量趋于无穷大时, 利用大数定律, Helly-Bray定理给出了分布式目标函数收敛于目标函数的期望值。最后, 我们建立了分布式鲁棒随机二次规划收敛于通常的随机二次规划问题。

关键词

随机二次规划, 分布式鲁棒优化, Wasserstein度量, Helly-Bray定理

Convergence Analysis of Data-Driven Distributionally Robust Stochastic Quadratic Programming

Dongdong Zhang*, Youpan Han*, Menghao Tian

School of Science, Xi'an Polytechnic University, Xi'an Shaanxi

Received: Dec. 17th, 2021; accepted: Jan. 6th, 2022; published: Jan. 20th, 2022

Abstract

In this paper, we study the convergence problem of distributionally robust stochastic quadratic programming under the data-driven Wasserstein ambiguity sets. First, we establish the point-by-point Lipschitz property of the objective function. Then, when the sample size tends to infinity, using the

*共同第一作者。

law of large numbers, the Helly-Bray theorem derives the expectation of the distributionally objective function converges to the objective function. Finally, we establish that the distributionally robust stochastic quadratic programming converges to the general stochastic quadratic programming problem.

Keywords

Stochastic Quadratic Programming, Distributionally Robust Optimization, Wasserstein Metric, Helly-Bray Theorem

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

最优解与最优值的收敛性是随机规划问题的重要性质之一, 因而关于随机规划的收敛性分析得到了广泛关注和深入研究。2000年骆建文等[1]对随机规划的逼近解的收敛性做了探讨, 证明了当随机序列依分布收敛时, 随机规划问题的任何最优解序列将收敛到原问题的最优解, 这为如何设计逼近算法来求解随机规划问题提供了理论基础。2012年霍永亮、刘三阳[2]指出关于随机规划逼近问题最优解集的上半收敛性结果依赖于概率度量(如Kolmogorov度量、Wasserstein度量、Fortet-Mourier度量)的收敛性[3] [4] [5]。

虽然随机规划的定量稳定性研究取得硕果, 却受限于复杂多变的实际问题。尤其是面对随机变量的概率分布模糊不清时非常棘手。为了更好地刻画实际问题, Ben-Tal [6]和 Ghaoui [7]在鲁棒优化方向做出了奠基工作。然而随机规划与鲁棒随机优化虽能较好地刻画随机问题, 但是前者需要知道随机变量的概率分布信息, 后者虽可求解问题, 但得到的解过于保守[8]。分布式鲁棒随机优化则介于二者之间, 它是从随机优化到鲁棒优化的桥梁, 它不需要精确知道优化问题中的随机变量的概率分布, 而且还易于求解。即使在相应的随机模型是#P-难的情况下, 分布式鲁棒优化问题也是可求解的。分布式鲁棒随机优化方法的核心是模糊集的构造。现在熟知的构造模糊集的方法有如下三种: ① 基于矩信息构造模糊集; ② 基于概率度量(如Prohorov metric、K-L divergence、Wasserstein metric等); ③ 第三种构造模糊集为拟合优度检验的置信域。

关于Wasserstein度量下的收敛性的研究工作较为热门。2013年Nicolas Fournier [9]等人给出了关于经验度量的Wasserstein距离的收敛速率, 这为后续的分布式鲁棒随机优化的收敛性分析奠定了基础。2014年Zhao Chaoyue在其博士论文[10]中给出了Wasserstein度量下样本量 N 与其构造的模糊集半径关系的系统性描述。Peyman Mohajerin Esfahani和Daniel Kuhn [11]在已有工作的基础之上, 得到了有限样本保证(有限样本情况下给出分布式鲁棒随机优化问题的置信界), 渐近一致性及分布的收敛性。

本文基于上述工作, 考虑当样本数据量趋于无穷时, Wasserstein度量下的分布式鲁棒随机二次规划的收敛性问题。

2. 模型介绍

2.1. 模型与符号

考虑如下模型:

$$\min_{x \in X} \sup_{P \in D} \int_{\Xi} f(x, \xi) P d\xi \quad (1)$$

其中, x 为决策变量, $x \in X$, $X \subseteq R^n$ 为约束集. 损失函数 $f: R^n \times R^m \rightarrow R$ 为随机二次函数, 其形式如下:

$$f(x, \xi) = \frac{1}{2} x^T Q(\xi) x + L^T(\xi) x + h(\xi),$$

$Q(\xi)$, $L(\xi)$, $h(\xi)$ 分别为关于 ξ 的仿射函数. 其中 $Q(\xi) = \sum_{i=1}^m A_i \xi_i + A_0$, $i=1, 2, \dots, m$ 为实对称矩阵且 A_0 为实半正定矩阵; $L(\xi) = B\xi + B_0$, $B \in R^{n \times m}$ 为实矩阵, $B_0 \in R^{n \times 1}$ 为常数向量; $h(\xi) = C^T \xi + c$, $C \in R^{m \times 1}$ 为常数向量, c 为常数. $\xi \in R^m$ 为随机向量, 服从分布 P , D 为包含 P 的模糊集, Ξ 是 P 的支撑集.

2.2. 数据驱动的分布式鲁棒随机二次规划模型

在诸多实际运用中, 分布 P 是未知的, 因此无法直接求解问题(1). 然而可以通过独立的历史样本数据观测到分布 P 的部分实现. 若使用 $\hat{\Xi}_N := \{\xi_i\}_{i \leq N} \subseteq \Xi$ 表示包含样本的数据集, 则可通过该数据集在某些概率距离下(如 K-L 散度、Prokhorov 度量、Wasserstein 度量等)构造的模糊集 \hat{D}_N , 使该集以较高的置信度包含真实概率分布 P . 由于 Wasserstein 度量具有良好的性质[9] [10] [11], 故本文利用 Wasserstein 距离来构造模糊集.

对于任意 $p \in [1, \infty)$, p -型 Wasserstein 度量 $W_p: M(\Xi) \times M(\Xi) \rightarrow R_+$ 定义为: 对所有的分布 $Q_1, Q_2 \in M(\Xi)$,

$$W_p(Q_1, Q_2) = \left(\inf_{\Pi} \int_{\Xi^2} \|\xi_1 - \xi_2\|^p \Pi(d\xi_1, d\xi_2) \right)^{\frac{1}{p}},$$

其中, $M(\Xi)$ 表示支撑集 Ξ 上的概率度量, Q_1 是随机变量 ξ_1 的概率分布, Q_2 是随机变量 ξ_2 的概率分布, Π 是 Q_1 和 Q_2 的联合概率分布, $\|\cdot\|$ 表示 R^m 中的任意范数.

当 $p=1$ 时, 随机变量间的单位运输成本表示为 $\|\xi_1 - \xi_2\|^1$, 相应地 $W_1(Q_1, Q_2) = \inf_{\Pi} \int_{\Xi^2} \|\xi_1 - \xi_2\|^1 \Pi(d\xi_1, d\xi_2)$. 此度量也是著名的 Kantorovich 度量.

根据 Wasserstein 度量的良好收敛性质, 本文采用 1-型 Wasserstein 度量来构造模糊集, 即

$$\hat{D}_N := \left\{ P \in M(\Xi) : W_1(P, \hat{P}_N) \leq \varepsilon \right\},$$

其中, \hat{P}_N 表示样本量为 N 的 i.i.d. 历史数据的经验分布.

故而, 数据驱动的分布式鲁棒随机二次规划模型为:

$$\min_{x \in X} \sup_{P \in \hat{D}_N} \int_{\Xi} \left[\frac{1}{2} x^T Q(\xi) x + L^T(\xi) x + h(\xi) \right] P d\xi. \quad (2)$$

3. 收敛性分析

为了得到数据驱动的分布式鲁棒随机二次规划的收敛性分析, 我们先给出如下的结论.

引理 1 [11] 当 Ξ 为紧集时, 存在一个常数 $a > 1$, $\forall N \geq 1$, $m \neq 2$, $\varepsilon > 0$ 有如下结论成立:

$$A := E^P \left[\exp(\|\xi\|^a) \right] = \int_{\Xi} \exp(\|\xi\|^a) P(d\xi) < \infty,$$

$$P^N \left\{ W_1(P, \hat{P}_N) \geq \varepsilon \right\} \leq \begin{cases} c_1 \exp(-c_2 N \varepsilon^{\max\{m, 2\}}) & \text{若 } \varepsilon \leq 1, \\ c_1 \exp(-c_2 N \varepsilon^a) & \text{若 } \varepsilon > 1, \end{cases}$$

其中, c_1, c_2 是依赖于 a, A, m 的正数。

根据上述集中不等式, 在给置信水平 β 下, 样本量 N 与 Kantorovich 度量构造的模糊集的 Wasserstein 球的半径 ε 有如下关系:

$$\varepsilon_N(\beta) := \begin{cases} \left(\frac{\log(c_1\beta^{-1})}{c_2N} \right)^{\frac{1}{\max\{m,2\}}} & \text{若 } N \geq \frac{\log(c_1\beta^{-1})}{c_2}, \\ \left(\frac{\log(c_1\beta^{-1})}{c_2N} \right)^{\frac{1}{a}} & \text{若 } N < \frac{\log(c_1\beta^{-1})}{c_2}. \end{cases}$$

引理 2 (Helly-Bray 定理) 若随机函数 $g: R^m \rightarrow R$ 为有界连续函数, 则 ξ_n 依分布收敛于 ξ 当且仅当 $E[g(\xi_n)] \rightarrow E[g(\xi)]$ 。

为了建立分布式鲁棒随机二次规划问题的收敛性结果, 首先考虑随机二次函数 $f(x, \xi)$ 的分析性质。

定理 1 对于任意给定的 $x \in X$, 任取 Ξ 中的随机向量 ξ, η 且 $\xi \neq \eta$, 则有 $\frac{|f(x, \xi) - f(x, \eta)|}{\|\xi - \eta\|} \leq K(x)$ 。

证明 首先, 对于 $\forall \xi, \eta \in \Xi$ 且 $\xi \neq \eta$,

$$f(x, \xi) - f(x, \eta) = \frac{1}{2}x^T Q(\xi)x + L^T(\xi)x + h(\xi) - \left[\frac{1}{2}x^T Q(\eta)x + L^T(\eta)x + h(\eta) \right].$$

将 $Q(\xi) = \sum_{i=1}^m A_i \xi_i + A_0$, $L(\xi) = B\xi + B_0$, $h(\xi) = C^T \xi + c$ 代入得:

$$\begin{aligned} f(x, \xi) - f(x, \eta) &= \frac{1}{2}x^T \left(\sum_{i=1}^m A_i \xi_i + A_0 \right) x - \frac{1}{2}x^T \left(\sum_{i=1}^m A_i \eta_i + A_0 \right) x + (B\xi + B_0)x - (B\eta + B_0)x + C^T \xi - C^T \eta \\ &= \frac{1}{2} \left\langle x, \left(\sum_{i=1}^m A_i \xi_i + A_0 \right) x \right\rangle - \frac{1}{2} \left\langle x, \left(\sum_{i=1}^m A_i \eta_i + A_0 \right) x \right\rangle + \langle B\xi, x \rangle - \langle B\eta, x \rangle + C^T(\xi - \eta) \\ &= \frac{1}{2} \left\langle x, \left(\sum_{i=1}^m A_i (\xi_i - \eta_i) \right) x \right\rangle + \langle B(\xi - \eta), x \rangle + C^T(\xi - \eta) \\ &= \frac{1}{2} \left\langle (x^T A_1 x, \dots, x^T A_m x), (\xi - \eta) \right\rangle + \langle \xi - \eta, B^T x \rangle + C^T(\xi - \eta) \end{aligned}$$

令 $H = [x^T A_1 x, \dots, x^T A_m x]$, H 为 m 维向量。则

$$f(x, \xi) - f(x, \eta) = \left(\frac{1}{2}H + B^T x + C^T \right) (\xi - \eta).$$

所以,

$$\frac{|f(x, \xi) - f(x, \eta)|}{\|\xi - \eta\|} = \frac{\left| \left(\frac{1}{2}H + B^T x + C^T \right) (\xi - \eta) \right|}{\|\xi - \eta\|} \leq \left\| \frac{1}{2}H + B^T x + C^T \right\| = K(x).$$

定理 2 对于每个给定的置信水平 β 和决策 $x \in X$ 。在 Ξ 为紧集的条件下, 随着样本量 $N \rightarrow \infty$ 时, 有模糊集半径 $\varepsilon \rightarrow 0$, $\lim_{N \rightarrow \infty} \sup_{Q \in \hat{D}_N} E^Q[f(x, \xi)] = E^P[f(x, \xi)]$ 成立。

证明 首先利用引理 1 的结论, 很容易得到当 $N \rightarrow \infty$ 时, $\varepsilon \rightarrow 0$ 。

根据([11], Theorem 4.2)证明过程中有如下结论:

$$\sup_{Q \in \hat{D}_N} E^Q[f(x, \xi)] = \inf_{\lambda \geq 0} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left[f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\| \right] \quad (3)$$

由于随着 $N \rightarrow \infty$, 在大数定律下经验分布 \hat{P}_N 弱收敛于分布 P . 又 $f(x, \xi)$ 关于随机变量 ξ 连续, 则 $f(x, \xi)$ 在 Ξ 上有界, 利用引理 2, 我们有如下关系:

$$\lim_{N \rightarrow \infty} \sup_{Q \in \hat{D}_N} E^Q [f(x, \xi)] \geq \lim_{N \rightarrow \infty} E^{\hat{P}_N} [f(x, \xi)] = E^P [f(x, \xi)].$$

根据(3)式, 接下来只需证明 $\liminf_{N \rightarrow \infty} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \leq E^P [f(x, \xi)]$ 成立即可。

由于

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \\ & \leq \inf_{\lambda \geq 0} \limsup_{N \rightarrow \infty} \left\{ \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} \end{aligned}$$

因此只需证明(4)成立即可:

$$\inf_{\lambda \geq 0} \limsup_{N \rightarrow \infty} \left\{ \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} \leq E^P [f(x, \xi)]. \quad (4)$$

根据定理 1 知, 存在一个常数 $L > 0$, 使得对于任意的 $x \in X$, 有

$$|f(x, \xi)| < L, \quad \forall \xi \in \Xi. \quad (5)$$

由于 Ξ 为紧集, 若令 B 为 Ξ 的直径, 则有

$$0 \leq \|\xi - z\| \leq B, \quad \forall \xi, z \in \Xi. \quad (6)$$

对于任意 $\lambda \geq 0$, 利用(5)和(6)可得:

$$-L - \lambda B \leq \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \leq L, \quad \forall z \in \Xi.$$

因此可进一步得到(7):

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \left\{ \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} \\ & = \limsup_{N \rightarrow \infty} \left\{ \lambda \varepsilon + E^{\hat{P}_N} \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} \\ & = \lim_{N \rightarrow \infty} \lambda \varepsilon + E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \\ & = E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \end{aligned} \quad (7)$$

现已将分布式鲁棒优化问题在样本量趋于无穷的情况下等价转化为鲁棒优化问题。接下来只需证明给定任意的 $x \in X$, (8)式成立即可。

$$\inf_{\lambda \geq 0} E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \leq E^P [f(x, z)]. \quad (8)$$

由于 $\inf_{\lambda \geq 0} E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \leq \limsup_{\lambda \rightarrow +\infty} E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|]$ 。因此需进一步证明下式成立即可:

$$\limsup_{\lambda \rightarrow +\infty} E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] \leq E^P [f(x, z)]. \quad (9)$$

接下来, 问题(9)中的最优解是 $\xi^*(z) = z$ 。假设 $\xi^*(z) = \hat{z} \neq z$, 根据定理 1 则有:

$$\begin{aligned} & [f(x, z) - \lambda \|z - z\|] - [f(x, \hat{z}) - \lambda \|z - \hat{z}\|] \\ & = f(x, z) - f(x, \hat{z}) + \lambda \|z - \hat{z}\| > (\lambda - K(x)) \|z - \hat{z}\| > 0 \end{aligned}$$

与 $\xi^*(z) = \hat{z} \neq z$ 矛盾, 因此可得:

$$\limsup_{\lambda \rightarrow +\infty} E^P \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - z\|] = \limsup_{\lambda \rightarrow +\infty} E^P [f(x, z)] = E^P [f(x, z)].$$

综上所述已经证得:

$$\begin{aligned} \liminf_{N \rightarrow \infty} \inf_{\lambda \geq 0} \left\{ \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} & \geq E^P [f(x, \xi)], \\ \liminf_{N \rightarrow \infty} \inf_{\lambda \geq 0} \left\{ \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [f(x, \xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} & \leq E^P [f(x, \xi)]. \end{aligned}$$

故定理 2 得证。

由于定理 2 的结论对于任意给定的 $x \in X$ 都成立。同时, $K(x)$ 是关于 x 的二次向量值函数的范数。而当 X 为紧集时, 该向量中每一个分量都是有限的。因而, 存在某个常数 $K > 0$, 使得 $K(x) \leq K$, 从而有下面的定理。

定理 3 对于每个给定的置信水平 β 及任意决策 $x \in X$ 。在 X 与 Ξ 为紧集的条件下, 我们有

$\lim_{N \rightarrow \infty} \min_{x \in X} \sup_{Q \in \hat{D}_N} E^Q [f(x, \xi)] = \min_{x \in X} E^P [f(x, \xi)]$ 成立, 即分布式鲁棒随机二次规划收敛于通常的随机二次规划。

4. 结论

当模糊集由概率分布的 Kantorovich 度量构成时, 本文对数据驱动的分布式鲁棒二次规划的收敛性问题进行了研究。在建立了目标函数关于随机变量是逐点 Lipschitz 连续性的基础上, 当样本量 N 趋于无穷时, 我们得到了最坏情况下的随机二次规划问题逐点收敛于通常的随机二次规划问题。进而, 在决策变量集合为紧集时, 我们获得分布式鲁棒随机二次规划问题收敛于通常的随机二次规划问题。这些结果不仅丰富了分布式鲁棒优化的理论, 还为进一步设计求解算法奠定了理论基础。

基金项目

本研究受国家自然科学基金 11501434 资助。

参考文献

- [1] 骆建文, 鲁世杰. 随机规划逼近解的收敛性[J]. 浙江大学学报(理学版), 2000, 27(5): 493-497.
- [2] 霍永亮, 刘三阳. 随机规划逼近问题最优解集的下半收敛性[J]. 数学进展, 2012, 41(6): 747-754.
- [3] Römisch, W. and Schultz, R. (1991) Stability Analysis for Stochastic Programs. *Annals of Operations Research*, **30**, 241-266. <https://doi.org/10.1007/BF02204819>
- [4] Klatte, D. (1994) On Quantitative Stability for Non-Isolated Minima. *Control and Cybernetics*, **23**, 183-200.
- [5] Schultz, R. (2000) Some Aspects of Stability in Stochastic Programming. *Annals of Operations Research*, **100**, 55-84. <https://doi.org/10.1023/A:1019258932012>
- [6] Ben-Tal, A. and Nemirovski, A. (1998) Robust Convex Optimization. *Mathematics of Operations Research*, **23**, 769-805. <https://doi.org/10.1287/moor.23.4.769>
- [7] ElGhaoui, L. and Le Bret, H. (1997) Robust Solutions to Least-Squares Problems with Uncertain Data. *SIAM Journal on Matrix Analysis and Applications*, **18**, 1035-1064. <https://doi.org/10.1137/S0895479896298130>
- [8] Zhao, C.Y. and Guan, Y.P. (2018) Data-Driven Risk-Averse Stochastic Optimization with Wasserstein Metric. *Opera-*

- tions Research Letters*, **46**, 262-267. <https://doi.org/10.1137/S0895479896298130>
- [9] Fournier, N. and Guillin, A. (2015) On the Rate of Convergence in Wasserstein Distance of the Empirical Measure. *Probability Theory and Related Fields*, **162**, 707-738. <https://doi.org/10.1137/S0895479896298130>
- [10] Zhao, C.Y. (2014) Data-Driven Risk-Averse Stochastic Program and Renewable Energy Integration. University of Florida, Gainesville, FL.
- [11] Esfahani, P.M. and Kuhn, D. (2018) Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations. *Mathematical Programming*, **171**, 115-166. <https://doi.org/10.1007/s10107-017-1172-1>