arXiv:2603.00851v1 [math.OC] 1 Mar 2026

部分不确定性下混合设定中的稳健最优投资组合

N. D. Shyamalkumar 和 Tianrun Wang*

2026年3月3日

摘要

在数据有限的情况下管理保险和金融风险，是保险业的一项重要任务。本文聚焦于这样的情形：风险分布被建模为混合分布，其中一些组成部分可被高精度估计或已知，而另一些组成部分及其权重则未知。本文讨论两个存在部分不确定性的稳健投资组合优化问题，其损失函数涉及方差或条件风险价值（CVaR）。我们采用投影次梯度下降算法求解这些优化问题。该问题可归结为一个凸-非凹极小极大问题。我们证明，尽管一般问题的收敛速率为 $O(1/\sqrt{k})$，其中 k 表示迭代次数，但在某些情况下可以实现指数收敛。最后，我们提供数值示例，展示该方法的有效性以及几何收敛速率的实现。本研究旨在为模型不确定性下的精算决策提供更有效的解决方案。

1 引言

自2000年代中期以来，美国保险公司一直受到C3第二阶段框架下基于风险的资本监管（见美国精算师学会 [2005]），该框架要求通过随机模拟计算可变年金等产品的监管资本。目前，美国保险公司每季度都会开展大量随机模拟，以便根据VM-20、VM-21和VM-22计算多种产品的法定准备金和资本，见美国保险监督官协会 [2024]。这种运用随机模拟来评估风险度量的做法本质上是被动的，因为它并非旨在推动战略性业务决策。作为替代方案，让我们考虑一项内部资产负债管理演练，以便在若干可能在下一年显著改变负债组合的业务策略中进行选择。目标是确定一种能够最大化风险调整后利润指标的策略。请注意，每一种此类策略还将涉及其最优资产组合的适应性选择。这一资产负债管理任务与上述基于风险的资本计算在两个重要方面有所不同。首先，风险调整后利润的计算涉及一项随机规划演练，用以确定最优资产组合。其次，尽管随机

*通讯作者：tianrun-wang@uiowa.edu

规划演练将使用与RBC任务类似的有限样本集，而最优资产组合的质量将关键取决于这些样本对未来经济环境的代表性。

现在，我们给出上述问题的数学表述。假设我们知道随机向量 $\mathbf{X}$ 的真实分布 P_*，对于函数 h，我们希望求解

$$\inf_{\theta\in\Theta} \mathbb{E}_{P_*}\left[h(\theta,\mathbf{X})\right],$$

其中 θ 是决策变量，Θ 是 θ 可以取值的集合。然而，真实分布往往未知，我们只有过去的数据，或者一组情景（用精算监管术语来说），例如 $\mathbf{x}_1$，.. .，$\mathbf{x}_n$。一种自然的解决方案是使用过去数据的经验分布，或者以情景为支撑集的离散均匀分布，作为真实分布的近似分布。换言之，我们得到样本均值近似（SAA）问题

$$\inf_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} h(\theta,\mathbf{x}_i).$$

遗憾的是，SAA 问题的最优解往往会受到过拟合的影响，这也称为优化器诅咒（见 Kuhn 等人 [2019]）。具体而言，在我们所关注的问题背景下，这样确定的最优资产配置可能依赖于近似分布中的特异性特征来取得良好表现。而真实分布中若不存在这些特征，其表现可能会差得多。

缓解这一问题的一种方法是修改目标函数，使其考虑近似分布邻域内各分布上的表现，希望借此减弱对个别特性的依赖，从而获得更优的稳健配置。这一问题的一种解决方案是极小极大分布鲁棒优化（DRO）问题

$$\inf_{\theta\in\Theta} \sup_{Q\in\mathcal{N}(\hat{P}_{\mathbf{X}})} \mathbb{E}_{Q}\left[h(\theta,\mathbf{X})\right],$$

其中，$\hat{}\, P_{\mathbf{X}}$ 是定义在$\mathbf{x}_1$、.....、$\mathbf{x}_n$上的离散均匀分布，而 $\mathcal{N}(\hat{}\, P_{\mathbf{X}})$表示其周围的分布邻域，称为模糊集。选择模糊集时需要权衡。例如，模糊集应足够大，以减轻过拟合；同时也应足够小，以免在真实分布上的表现显著下降，尤其要考虑到在不可信的分布上实现稳健表现所需付出的代价。

在许多领域和应用中，指定围绕最佳猜测 $P_{\mathbf{X}}$ 分布的邻域 $\mathcal{N}(\hat{}\, P_{\mathbf{X}})$ 有多种方法。经典统计稳健性领域最早的方法之一是 Huber [1964]，的 ϵ-污染模型，后来 Berger 和 Berliner [1986] 将其引入贝叶斯先验稳健性；另请参阅 Berger [1994]，了解贝叶斯先验稳健性中使用的多种不同邻域规范的概述。在精算学领域，有关如何求取保守风险界（即稳健风险度量）的研究表明，当分布的某些部分受到约束或被指定时，例如边缘分布受到约束而依赖结构任意，或矩受到约束而分布为任意单峰分布，便可进行此类分析。请参阅

Embrechts 等人 [2005], 、Bernard 等人 [2020], 、Li 等人 [2018], 、Cornilly 等人 [2018], ，以及其中的参考文献。在鲁棒随机规划领域，请参阅 Delage 和 Ye [2010]：其中一阶矩被限制在给定椭球内，二阶矩则满足半正定约束。正如 Tang 和 Yang [2023], 所指出的，这类界的一个缺点是，用于定义这些邻域的函数值（例如矩和依赖度量）即使受到合理约束，也往往仍会允许不合理的分布，从而导致风险界过于保守。

另一种指定邻域 $\mathcal{N}(\hat{P}_{\mathbf{X}})$ 的方法，是将其设为分布空间上所选度量空间拓扑中的一个球，或者用 f-散度来定义（见 Csisz´ar [1972]）。f-散度中最常见的是 Kullback-Leibler 散度，它与统计推断有着根本联系；例如，参见 Kullback 和 Leibler [1951],Akaike [1973] 以及 Berk [1966]。此外，在鲁棒随机规划中，一个重要性质是可以将渐近置信集构造为经验分布的 f-散度邻域，见 Ben-Tal 等人[2013]。尽管 f-散度有这些优点，但它通常会在分布不绝对连续于参考分布时趋于无穷；若以经验分布为中心，这会排除真实分布，因此并不理想。出于这一原因，我们在保持数学上可处理性的同时，在本文中采用 Wasserstein距离；关于使用 Wasserstein距离进行分布鲁棒优化的教程，参见 Kuhn 等人 [2019]和 Blanchet 等人 [2021]。其他选择包括基于核的最大均值差异（MMD）距离（见 Staib 和 Jegelka [2019]），以及积分概率度量（IPM）这一类度量（见 M¨uller [1997] 和 Sriperumbudur 等人 [2012], ）；该类度量包括 f- 散度中的全变差距离、MMD 和 Wasserstein-1距离。关于在分布鲁棒优化中使用 IPM，参见 Husain [2020] 。对于 μ 上的概率测度，阶数为 ρ、半径为 r且以 μ 为中心的 Wasserstein 球定义为 $\mathbb{B}_{W_\rho}(\mu, r) = \{\nu\colon W_p(\mu,\nu) < r\}$，其中 $W_p(\mu,\nu)^p = \inf_{\mathbf{X}\sim\mu,\, \mathbf{Y}\sim\nu} \mathbb{E}\left(||\mathbf{X} - \mathbf{Y}||_p^p\right)$，下确界取遍 μ 和 ν 的所有耦合。

我们将 Huber [1964] 的 ϵ-污染方法与将污染分布限制在 Wasserstein 球内相结合，以构造 DRO 框架下的模糊集。当然，在 Huber [1964], 的设定中，污染分布代表离群值的分布，因此自然不包含有用信息。相反，遵循 Tang 和 Yang [2023], 的方法，在我们的设定中，假设未知分布 $\mathbf{X}$ 具有混合结构，其中一些组成部分已知，另一些未知，且分量权重也未知。Tang 和 Yang[2023] 针对 $\mathbb{E}_{P_*}[h(X)]$ 给出了一个上界，形式为

$$\sup_{Q\in\mathcal{N}(\hat{F}_X)} \mathbb{E}_Q[h(X)], \tag{1}$$

其中 h 是某个多项式函数， $\mathcal{N}$ 是如上所述的模糊集。

我们为相关文献作出了若干贡献。我们将这种混合分布模糊框架应用于鲁棒均值-方差和均值-CVaR投资组合优化问题，提供求解这些问题的算法，并分析其收敛速率。具体而言，令决策变量为投资组合$\mathbf{x} \in \mathbb{R}^d$，收益为随机向量$\mathbf{R} \in \mathbb{R}^d$。我们希望选择一个投资组合$\mathbf{x}$，以使最坏情形风险最小化：

$$\inf_{\mathbf{x}:\mathbf{1}'\mathbf{x}=1} \sup_{F_{\mathbf{R}}\in\mathcal{N}(\hat{F}_{\mathbf{R}})} D(-\mathbf{x}'\mathbf{R}), \tag{2}$$

ˆ 其中， D 是一个效用损失函数， $F_{\mathbf{R}}$ 具有双混合结构，源自可能存在的两种经济状态。例如，收益$\mathbf{R}$ 可以建模为混合分布，其权重未知，且高利率状态下的收益因数据有限而具有模糊性。我们采用一个尊重这种混合结构的模糊集 $\mathcal{N}$ ，并将在第2节和第3节中详细讨论。

在第2节中，我们研究带有 $D(\cdot) =$ 方差 $(\cdot) + \gamma\mathbb{E}(\cdot)$ 的问题 (2)，即均值-方差分布鲁棒优化问题，推导其等价的极小极大形式，并使用投影次梯度下降算法求解。该算法保证以 $O(\frac{1}{\sqrt{T}})$ 速率收敛；在附加条件下，我们还证明它可以达到指数收敛速率。在第3节中，我们将框架扩展到带有 $D(\cdot) = \mathbb{E}(\cdot) + \rho\text{CVaR}_p(\cdot)$ 的均值-CVaR问题。在第4节中，我们给出数值示例，表明即使某些假设不成立，也能达到指数速率。此外，我们通过将样本外效用损失与样本平均近似方法的结果进行比较，展示了分布鲁棒优化方法的有效性。在这两种情况下，我们从理论和实证两方面证明，随着球的半径趋于无穷大，分布鲁棒优化投资组合会收敛到1/N投资组合。在第5节中，我们总结全文，并讨论允许卖空的情形。

记号： 令 $\rightrightarrows$ 表示一致收敛。若 f 是一个函数，则 $f_x := \frac{\partial}{\partial x} f$ 表示 f 关于 x的偏导数。 $\|\cdot\|$、 $\|\cdot\|_1$和 $\|\cdot\|_\infty$ 分别表示向量的 2-范数、1-范数和 ∞-范数。对于矩阵 A， $\|A\|_2 := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$表示 A的算子2-范数。

2 均值-方差效用损失函数的混合分布 $\mathbf{n}$

考虑从 d项资产中构建投资组合的问题。我们用权重向量 $\mathbf{x} \in \Delta$表示投资于这些资产的财富比例，其中

$$\Delta = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \geq 0, \mathbf{1}'\mathbf{x} = 1\},$$

是 d-单纯形。投资组合收益为 $\mathbf{x}'\mathbf{R}$，其中 $\mathbf{R}$是资产收益向量。我们的目标是选择 $\mathbf{x}$以最小化效用损失函数，其中 $\mathbf{R}$的真实分布服从双成分混合结构，但其中一个成分及其权重未知。请注意，尽管下文的推导中不允许卖空头寸，但我们会在第5节讨论如何放宽这一限制。

在本节中，我们首先考虑以下均值-方差效用损失函数，它反映了投资组合的预期收益与风险之间的权衡，其中风险以方差衡量，权衡程度由参数 $\gamma > 0$ 控制：

$$\text{Var}\left(\mathbf{x}'\mathbf{R}\right) - \gamma\mathbb{E}(\mathbf{x}'\mathbf{R}). \tag{3}$$

注意， $\frac{2}{\gamma}$ 是均值-方差效用中传统的风险厌恶系数，例如，参见 Ang [2014] 第 2 章。假设资产收益的未知真实分布 P 属于一个概率测度模糊集，记为 $\mathcal{S}$。因此，投资组合选择问题就是以下最坏情况效用损失最小化问题：

$$\inf_{\mathbf{x}\in\Delta} \sup_{P\in\mathcal{S}} \left(\text{Var}_P\left(\mathbf{x}'\mathbf{R}\right) - \gamma\mathbb{E}_P(\mathbf{x}'\mathbf{R})\right). \tag{Opt 1}$$

模糊集 $\mathcal{S}$ 是由一组概率测度构成的，用于表示建模中的模糊性。对于一个存在两种潜在机制的市场，其中机制以及两种机制的相对发生频率均存在模糊性，我们考虑以下形式的模糊集：

$$\mathcal{S} = \{P = (1-q)P_N + qP_S \mid q \in [q_0 - \varepsilon, q_0 + \varepsilon],\, P_S \in \mathbb{B}_{W_2}(P_{S_0}, r(q))\}, \tag{4}$$

其中，P_N，$P_{S_0} \in \mathcal{P}(\mathbb{R}^d)$ 表示定义在 $\mathbb{R}^d$ 上的博雷尔概率测度集合，且 $\varepsilon > 0$。在下文的推导中，我们进一步假设 $r \in C^2([0,1])$，并假设两种机制下的方差-协方差矩阵 Σ_{P_N} 和 $\Sigma_{P_{S_0}}$ 均为正定矩阵。换言之，我们假设资产在两种机制下均受到约束，只能构成一个不含多重共线性的最小集合。需要指出的是，要纳入无风险资产，只需将概率单纯形 Δ 替换为次概率单纯形即可。由于这一替换仍保留了紧性，我们的推导同样适用于这种情况。

$\mathcal{S}$ 的定义取决于判断：根据领域知识和样本外环境，对哪些分布是合理的进行判断，同时还要考虑由此产生的计算问题在数学上的可处理性。通过这种方式，可以在可接受的保守性成本下，实现对未知抽样环境的稳健性，参见 Blanchet 等人 [2025]。上述定义的 $\mathcal{S}$ 可以嵌入一个 Wasserstein 球中，因为 r是有界的。然而，这样的 Wasserstein 球会包含许多偏离上述结构形式的分布，尤其是当 P_N 和P_{S_0} 在 W_2 距离下相距甚远时。因此，为实现所期望的稳健性，Wasserstein 球会引入不可接受程度的保守性。

在上述定义中，$\mathbb{B}_{W_2}(P_{S_0},\ r(q))$ 是以参考分布 P_{S_0} 为中心、半径为 $r(q)$ 的 Wasserstein-2 球。P_S（其中 S 表示“压力情景”）用于描述非典型或受压的投资环境，例如利率较高、市场崩盘前后的环境，以及市场基本面显著偏离正常水平的其他情景。显然，从对这种状态的描述本身就可以看出，建模者缺乏足够的数据来对 P_S 建模，尤其是与可用于对 P_N 建模的数据相比（其中 N 表示“正常”）。受压市场环境出现的相对频率记为 q_0。例如，可以进行如下粗略估算：过去40年间，发生过三次重大市场崩盘：2002年的互联网泡沫崩盘、2008年的次贷危机以及2020年的新冠疫情，因此可以采用 $q_0 = (3/40) = 0.075$。

下面的定理1将 (Opt 1) 化为一个极小极大问题，其证明见下一小节。为此，我们定义两个将在本节中始终使用的辅助函数：

$$\psi_{a,\gamma}(y) := (y-a)^2 - \gamma y, \tag{5}$$

and

$$V(q, \mathbf{x}, a) = \left(r(q)\|\mathbf{x}\| + \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2} \right)^2 - a\gamma - \frac{\gamma^2}{4}. \tag{6}$$

定理1。(Opt1) 中的优化问题等价于以下极小极大问题:

$$\min_{\mathbf{x}\in\Delta, a\in I} \max_{q\in[q_0-\varepsilon,\, q_0+\varepsilon]} \Big[(1-q)\cdot \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + q\cdot V(q,\mathbf{x},a)\Big], \tag{7}$$

其中 I 是一个不依赖于$\mathbf{x}$的紧区间。 □

我们提出一种投影次梯度下降算法（算法 1）来求解 (7)。定义极小极大目标

$$h(q,\mathbf{x},a) := (1-q)\cdot \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + q\cdot V(q,\mathbf{x},a), \tag{8}$$

及其部分最大值

$$J(\mathbf{x},a) := \sup_{q\in[q_0-\varepsilon, q_0+\varepsilon]} h(q,\mathbf{x},a), \tag{9}$$

其中 $\Omega := [q_0-\varepsilon,\, q_0+\varepsilon]\times I\times\Delta$ 是 h 的紧域。我们假设 $\varepsilon < q_0\wedge(1-q_0)$，使 q 与 0 和 1 均保持一定距离。该算法的投影步骤是将一个向量投影到 Δ 上，投影算子可方便地表示为以下二次规划的解:

$$\mathrm{Proj}_{\Delta}(\mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}\in\Delta} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \mathbf{y}\in\mathbb{R}^d. \tag{10}$$

算法 1 鲁棒均值-方差的投影次梯度下降

1: **初始化:** 选择 $(\mathbf{x}_0, a_0) \in \mathbb{R}\times\Delta$，迭代次数 $T > 0$ 和步长 es
$\eta_k > 0$ 对于 $k = 0, 1, \ldots, T$. 2: 对于 $k = 0, 1, \ldots, T$执行 3: 求最坏情况
概率 $q_k^* = \mathsf{argmax}\ h(q, \mathbf{x}_k, a_k)_{.q\ni|q-q_0|\le\epsilon}$4: 计算 h 在 $(q_k^*, \mathbf{x}_k, a_k)$ 处关于 $(\mathbf{x}\ a)$
的梯度 $\mathsf{g}_{k,\mathbf{x}}, g_{k,a}$。 5: 更新变量:

$$\mathbf{x}_{k+} \leftarrow \mathbf{x}_k - \eta_k \mathbf{g}_{k,\mathbf{x}}$$
$$a_{k+1} \leftarrow \tau_k - \eta_k g_{k,a}$$

6: 投影 $\mathbf{x}_{k+1}$ ←投影$_{\Delta}(\mathbf{x}^{k+})$.7: **结束**
循环8: **返回** $(a_1, \mathbf{x}_1)\ldots, (a_T, \mathbf{x}_T)$.

在投影下降算法中，投影步骤通常是主要的计算瓶颈；然而，在我们的情形下，投影算子有一个闭式表达式。对于 $\mathbf{y}\in\mathbb{R}^d$, $\mathbf{x}^*$:=投影$_{\Delta}(\mathbf{y})$ 的定义如下:

$$x_i^* = \max(y_i - \lambda, 0), \quad \text{where } \lambda = \frac{1}{m}\left(\sum_{j=1}^{m} y_{(j)} - 1\right), \tag{11}$$

其中 m 是满足以下条件的最大整数

$$y_{(m)} - \frac{1}{m}\left(\sum_{j=1}^{m} y_{(j)} - 1\right) > 0.$$

在上文中，$y_{(1)} \geq y_{(2)} \geq \cdots \geq y_{(d)}$ 表示向量$\mathbf{y}$的有序组成部分，并可在 $O(d\log\ d)$时间内计算。这也是投影步骤的复杂度。

请注意，在算法 1 的第 4 步中，我们真正寻求的是 J 在$(\mathbf{x}_k, a_k)$处的一个次梯度$\mathbf{g}_k$ 。鉴于引理 1，根据 Danskin 定理（见附录中的定理 9），这样的次梯度$\mathbf{g}_k$ 由内部函数 h 关于$(\mathbf{x}\ a)$的梯度给出，并在$(q_k^*, \mathbf{x}_k, a_k)$处求值。下面的命题给出了该次梯度的表达式，其证明见附录。为陈述该命题，我们将 S_γ 定义为，

$$S_\gamma(\mathbf{x}, a) = \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2},$$

并注意 h 可用它表示为

$$h(q, \mathbf{x}, a) = (1-q)\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + q\left[(r(q)\|\mathbf{x}\| + S_\gamma(\mathbf{x}, a))^2 - a\gamma - \frac{\gamma^2}{4}\right].$$

命题1. J 在$(\mathbf{x}_k, a_k)$ 处的一个次梯度为$(\mathbf{g}_{k,\mathbf{x}}, g_{k,a})$，其中

$$\begin{aligned}\mathbf{g}_{k,\mathbf{x}} &= (1-q_k^*)\left[2(\Sigma_{P_N} + \mu_{P_N}\mu_{P_N}')\mathbf{x}_k - 2a_k\mu_{P_N} - \gamma\mu_{P_N}\right] \\ &\quad + q_k^*\left[2\left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right)\left(r(q_k^*)\frac{\mathbf{x}_k}{\|\mathbf{x}_k\|} + \frac{\Sigma_{P_{S_0}}\mathbf{x}_k + \left(\mathbf{x}_k'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right)\cdot\mu_{P_{S_0}}}{S_\gamma(\mathbf{x}_k, a_k)}\right)\right],\end{aligned}$$

and

$$g_{k,a} = (1-q_k^*)[-2(\mathbf{x}_k'\mu_{P_N} - a_k)] + q_k^*\left[-2\frac{r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)}{S_\gamma(\mathbf{x}_k, a_k)}\left(\mathbf{x}_k'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right) - \gamma\right].$$

□

注意，由于 $r \in C^2([0,1])$，我们有 $h \in C^2(\Omega)$。因此，h_q，h_{qq}，$\frac{\partial^2}{\partial q \partial(\mathbf{x}\ a)}h$和 $\frac{\partial^2}{\partial^2\ \partial(\mathbf{x}\ a)}h$ 是在紧域 Ω 上的连续函数。令 $(\mathbf{g}_{k,\ \mathbf{x}},\ g_{k,\ a})$ 在 $(q_k^*,\ \mathbf{x}_k,\ a_k) \in \Omega$ 上的最大 2-范数为 K_1。令 K_2，K_3 为正数

$$K_2 := \max_{(q,\mathbf{x},a)\in\Omega}\left\|\frac{\partial^2}{\partial(\mathbf{x}, a)^2}h\right\|_2, \quad K_3 := \max_{(q,\mathbf{x},a)\in\Omega}\left\|\frac{\partial^2}{\partial q\partial(\mathbf{x}, a)}h\right\|.$$

注意，对于 K_1，K_2，K_3，使用上界而非最大值即可。

在下面的引理 1 中，我们将证明 h 和 J的凸性与光滑性。注意，如果 f 二阶可微，那么 α 的强凸性（见附录中的定义 2）等价于其海森矩阵的特征值均至少为 α。另一方面，β-光滑性（见附录中的定义 3）等价于这些特征值均至多为 β。比值 β/α 称为 f的条件数。如果 $-f$ 是强凸的，那么我们称 f 是强凹的。引理 1 的证明见下一小节。

引理 1. 关于函数 h 和 J，我们有以下结论：

i. h 在$(\mathbf{x}\ a)$上是强凸的。

ii. J 是 K_1-利普希茨连续的，且是$((1-q_0-\varepsilon)\lambda_M)$-强凸的，其中 λ_M 是.....的最小特征值

$$M := \begin{pmatrix} \Sigma_{P_N} + \mu_{P_N}\mu'_{P_N} & -\mu_{P_N} \\ -\mu'_{P_N} & 1 \end{pmatrix}.$$

iii. 当对任意$(\mathbf{x}\ a)$，h关于 q 在 $[q_0-\varepsilon, q_0+\varepsilon]$上强凹时，存在$\alpha>0$使得 $h_{qq} \le -\alpha < 0$，且 J是 $\binom{K_3^2}{\alpha} + K_2$-光滑的。—

□

投影次梯度下降算法的三个基本复杂度结果是附录中的定理 10、11 和定理 12，它们分别给出了算法以 $O(1/\sqrt{k})$、 $O(1/k)$ 和指数速率收敛的条件。由引理 1 ii. 可知， J 是强凸的，因此存在唯一一点，以下记为（$\mathbf{x}^*$， a^*），使其取得最小值。此外，由于如命题1所示， J的次梯度有界于 K_1，因此 J 满足定理 11 的条件。因此，我们的算法至少具有 $O(1/k)$ 的收敛速率。此外，如果引理 1 iii. 的条件成立，∇J 将是利普希茨连续的；由定理 12 可知，此时收敛速率将是指数级的。我们将在以下两个定理中正式阐述收敛速率。

定理2. 对于按(4)定义的模糊集，采用时变步长的算法 *1*

$$\eta_k = \frac{2}{\lambda_M(1-q_0-\varepsilon)(k+1)}$$

满足

$$J\left(\sum_{k=1}^{T} \frac{2k}{T(T+1)}(\mathbf{x}_k, a_k)\right) - J((\mathbf{x}^*, a^*)) \ \le\ \frac{2K_1^2}{\lambda_M(1-q_0-\varepsilon)(T+1)}.$$

证明. 这是引理 1 ii. 与定理 11 的结合。 □

在下面的定理 3 中，我们证明算法 1 能够达到指数收敛速率。为方便起见，我们在表述该定理时选择 r，使其峰值位于 q_0。此类函数可从多用途的贝塔分布族中选取

$$r(q) \propto q^{\alpha-1}(1-q)^{\beta-1},$$

其中 r 在 $\alpha>1, \beta>1$ 时于 $\frac{\alpha-1}{\alpha+\beta-2}$ 处达到最大值，而单峰性意味着其在众数附近局部严格凹。定理 3 的证明见下一小节。

定理 3. 若 $r'(q_0) = 0$, $r''(q_0) < 0$，则存在一个 $\varepsilon_0 > 0$，使得对于 (4) 中定义且满足 $\varepsilon \le \varepsilon_0$ 的那些模糊集，采用恒定步长的算法 *1*

$$\eta_k = \frac{K_4}{K_3^2 + K_2 K_4}$$

满足

$$J(\mathbf{x}_T, a_T) - J(\mathbf{x}^*, a^*) \le \frac{K_3^2 + K_2 K_4}{2K_4} \exp\left(-\frac{K_4\, \lambda_M (1 - q_0 - \varepsilon_0)}{K_3^2 + K_2 K_4}\, T\right) \|(\mathbf{x}_0, a_0) - (\mathbf{x}^*, a^*)\|^2,$$

其中 $K_4 := -q_0 r(q_0) r''(q_0)/d > 0$。 □

最后，定理4的证明将在下一小节给出。定理4表明，随着模糊性增大，$\mathbf{x}^*$趋近于 $(\frac{1}{d}, \ldots, \frac{1}{d})$，即等权重投资组合，也称为 $1/N$ 投资组合，参见 Guo 等人 [2019]。

定理4。 设$(\mathbf{x}_C^*, a_C^*)$ 是 J 的一个极小化解，其中 *(4)* 中的模糊半径定义为$r(q) = C \cdot r_{base}(q)$，其中 $C > 0$ 是一个缩放参数，r_{base} 是在区间 $[0, 1]$, 上固定的非负连续函数，且在 $[q_0 - \varepsilon, q_0 + \varepsilon]$ 上不恒为零。于是，当 $C \to \infty$时，$\mathbf{x}_C^*$ 收敛于等权重投资组合 $\frac{1}{d} \cdot \mathbf{1}$。

2.1 结果的证明

2.1.1 定理1的证明

为证明定理1，我们需要以下引理。我们先推导（Opt 1）中内部最大化问题的一种等价形式。

引理 2. 对于投资组合权重$\mathbf{x}$，（*Opt 1*）的内部问题 满足:

$$\sup_{P \in \mathcal{S}} \left(\mathrm{Var}_P\left(\mathbf{x}'\mathbf{R}\right) - \gamma \mathbb{E}_P(\mathbf{x}'\mathbf{R})\right) = \min_{a \in I} \sup_{|q - q_0| \le \varepsilon} \left\{(1-q)\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + qV(q, \mathbf{x}, a)\right\}, \quad (12)$$

其中

$$\psi_{a,\gamma}(y) := (y-a)^2 - \gamma y, \quad V(q, \mathbf{x}, a) := \sup_{G \in \mathbb{B}_{W_2}(P_{S_0}, r(q))} \mathbb{E}_{P_S}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})),$$

且 I 是一个不依赖于$\mathbf{x}$的紧区间。

证明。我们使用方差的变分形式，即 $\mathrm{Var}\,(Y) = \min_{a \in \mathbb{R}} \mathbb{E}(Y-a)^2$,

$$\sup_{P \in \mathcal{S}} \left(\mathrm{Var}_P\left(\mathbf{x}'\mathbf{R}\right) - \gamma \mathbb{E}_P(\mathbf{x}'\mathbf{R})\right) = \sup_{P \in \mathcal{S}} \min_{a \in \mathbb{R}} \left(\mathbb{E}_P\left((\mathbf{x}'\mathbf{R} - a)^2 - \gamma \cdot \mathbf{x}'\mathbf{R}\right)\right). \quad (13)$$

我们首先证明，内部优化可以限制在一个不含 $P \in \mathcal{S}$ 且不含 $\mathbf{x} \in \Delta$ 的紧区间 I 内。只需证明存在一个包含最小点 $\mathbb{E}_P(\mathbf{x}'\mathbf{R})$ 的 I，使其对所有 $P \in \mathcal{S}$ 和 $\mathbf{x} \in \Delta$ 均成立。由于 f 连续，令 M 为 f 在 $[q_0 - \varepsilon, q_0 + \varepsilon]$ 上的最大值。那么，对于任意属于 $\mathbb{B}_{W_2}(P_{S_0}, M)$ 的 P_S 以及 $\mathbf{x} \in \Delta$，若记 π 为 P_S 与 P_{S_0} 之间的最优 2-Wasserstein 耦合，并且 $(\mathbf{R}, \mathbf{R}_0) \sim \pi$，则有

$$\begin{aligned}\mathbb{E}_{P_S}\left(\|\mathbf{R}\|^2\right) = \mathbb{E}_\pi\left(\|\mathbf{R}_0 + \mathbf{R} - \mathbf{R}_0\|^2\right) &\le 2\left(\mathbb{E}_\pi\left(\|\mathbf{R}_0\|^2\right) + \mathbb{E}_\pi\left(\|\mathbf{R} - \mathbf{R}_0\|^2\right)\right) \\ &\le 2\left(\mathbb{E}_{P_{S_0}}\left(\|\mathbf{R}\|^2\right) + M^2\right),\end{aligned}$$

and

$$[\mathbb{E}_{P_S}(\mathbf{x}'\mathbf{R})]^2 \le \mathbb{E}_{P_S}(\mathbf{x}'\mathbf{R})^2 \le \|\mathbf{x}\|^2\mathbb{E}_{P_S}\big(\|\mathbf{R}\|^2\big) \le 2(\mathbb{E}_{P_{S_0}}\big(\|\mathbf{R}\|^2\big) + M^2).$$

同理,

$$(\mathbb{E}_{P_N}(\mathbf{x}'\mathbf{R}))^2 \le \big(\mathbb{E}_{P_N}\big((\mathbf{x}'\mathbf{R})^2\big)\big) \le \|\mathbf{x}\|^2\mathbb{E}_{P_N}\big(\|\mathbf{R}\|^2\big) \le \mathbb{E}_{P_N}\big(\|\mathbf{R}\|^2\big).$$

由于 $\mathcal{S}$中的任意 P都是 P_N与 $\mathbb{B}_{W_2}$（P_{S_0}, M）中某个 P_S的混合分布，其混合权重 q属于 $[q_0-\varepsilon, q_0+\varepsilon]$，综合上述陈述，我们有,

$$|\mathbb{E}_P(\mathbf{x}'\mathbf{R})| = |(1-q)\mathbb{E}_{P_N}(\mathbf{x}'\mathbf{R}) + q\mathbb{E}_{P_S}(\mathbf{x}'\mathbf{R})| \le \sqrt{2(\mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|^2) + M^2) \vee \mathbb{E}_{P_N}(\|\mathbf{R}\|^2)}.$$

上述结果表明我们所求的 I存在。

由于 I是紧的， $\mathbb{E}_P((\mathbf{x}'\mathbf{R} - a)^2 - \gamma \cdot \mathbf{x}'\mathbf{R})$关于 a是凸的且连续的，并且关于 P在2-瓦瑟斯坦度量拓扑下是凹的（线性的）且连续的（关于 W_2 收敛的含义，参见Villani [2003] 中的定理 7.12），根据Sion 极小极大定理（参见附录中的定理 14），我们可以交换最小值与上确界算子:

$$\sup_{P\in\mathcal{S}} \min_{a\in I} \big(\mathbb{E}_P\big((\mathbf{x}'\mathbf{R}-a)^2 - \gamma\cdot\mathbf{x}'\mathbf{R}\big)\big) = \min_{a\in I} \sup_{P\in\mathcal{S}} \big(\mathbb{E}_P\big((\mathbf{x}'\mathbf{R}-a)^2 - \gamma\cdot\mathbf{x}'\mathbf{R}\big)\big). \tag{14}$$

由此可得，对于固定的$\mathbf{x}$ a,

$$\sup_{P\in\mathcal{S}} \big(\mathbb{E}_P\big((\mathbf{x}'\mathbf{R}-a)^2 - \gamma(\mathbf{x}'\mathbf{R})\big)\big) = \sup_{|q-q_0|\le\varepsilon} \{(1-q)\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + qV(q,\mathbf{x},a)\}.$$

□

上述引理表明，对于固定的$\mathbf{x}$ a，内部问题可以依次求解；首先找出Wasserstein 球中的最坏情形分布 G_q ，然后找出最坏情形混合权重 q。

我们利用 Gao 和 Kleywegt [2023] 的定理 1 的对偶性结果（参见附录中的定理 8），推导出 $V(\cdot,\cdot,\cdot)$ 的如下表示:

$$V(q,\mathbf{x},a) = \inf_{\lambda\ge 0}\left\{\lambda r(q)^2 + \mathbb{E}_{P_{S_0}}\left(\sup_{\mathbf{r}}\left(\psi_{a,\gamma}(\mathbf{x}'\mathbf{r}) - \lambda\|\mathbf{r}-\mathbf{R}\|^2\right)\right)\right\}, \tag{15}$$

其中，$\mathbf{R}$ 是从参考分布 P_{S_0} 中抽取的随机向量。式 (15) 中下确界的可处理性，关键在于内层上确界的闭式表达式。为此，我们证明以下两个引理。

引理 3。

$$\mathbb{E}_{P_{S_0}}\left(\sup_{\mathbf{r}}\left(\psi_{a,\gamma}(\mathbf{x}'\mathbf{r}) - \lambda\|\mathbf{r}-\mathbf{R}\|^2\right)\right) = \begin{cases} \frac{\lambda}{\lambda-\|\mathbf{x}\|^2}\left(\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2\right) - a\gamma - \frac{\gamma^2}{4}, & \lambda \ge \|\mathbf{x}\|^2 \\ +\infty, & \lambda < \|\mathbf{x}\|^2, \end{cases} \tag{16}$$

其中 $\mu_{P_{S_0}}, \Sigma_{P_{S_0}}$ 分别表示 P_{S_0} 的均值和方差-协方差矩阵，并约定$0\ 0 := 0, \frac{c}{0} := \infty$ 适用于 $c > 0$。

证明。 我们先令投资组合收益 $\mathbf{x}'$ 等于 y，对$\mathbf{r}$ 进行优化，然后再对 y 进行优化，从而推导出上确界的表达式。

$$\sup_{\mathbf{r}}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{r}) - \lambda\|\mathbf{r}-\mathbf{R}\|^2) = \sup_{y\in\mathbb{R}}\left\{\psi_{a,\gamma}(y) - \inf_{\mathbf{r}:\mathbf{x}'\mathbf{r}=y}\lambda\|\mathbf{r}-\mathbf{R}\|^2\right\}. \tag{17}$$

下确界在以下位置取到

$$\mathbf{r}^* = \mathbf{R} - \left(\frac{\mathbf{x}'\mathbf{R}-y}{\|\mathbf{x}\|^2}\right)\mathbf{x},$$

从而

$$\inf_{\mathbf{r}:\mathbf{x}'\mathbf{r}=y}\|\mathbf{r}-\mathbf{R}\| = \frac{|y-\mathbf{x}'\mathbf{R}|}{\|\mathbf{x}\|}. \tag{18}$$

代入后，我们得到一个关于 y 的一维最大化问题。令$\mathbf{x}'\mathbf{R}$ 为 Y，现在我们推导出其闭式表达式，

$$\begin{aligned} H(\lambda, Y) &:= \sup_{y\in\mathbb{R}}\left\{(y-a)^2 - \gamma y - \frac{\lambda}{\|\mathbf{x}\|^2}(y-Y)^2\right\} \\ &= \sup_{y}\left\{\left(1-\frac{\lambda}{\|\mathbf{x}\|^2}\right)y^2 + \left(2\frac{\lambda Y}{\|\mathbf{x}\|^2} - 2a - \gamma\right)y + \left(a^2 - \frac{\lambda Y^2}{\|\mathbf{x}\|^2}\right)\right\}. \end{aligned} \tag{19}$$

上述　　二次最大化问题在 $\lambda > \|\mathbf{x}\|^2$ 时是凹的。简单计算可得

$$H(\lambda, Y) = \begin{cases} \frac{\lambda}{\lambda-\|\mathbf{x}\|^2}\left(Y - \frac{2a+\gamma}{2}\right)^2 - a\gamma - \frac{\gamma^2}{4}, & \lambda \geq \|\mathbf{x}\|^2 \\ +\infty, & \lambda < \|\mathbf{x}\|^2 \end{cases} \tag{20}$$

采用上述定义的扩展实数约定。结合起来，利用以下事实可得 (16):

$$\mathbb{E}_{P_{S_0}}[Y] = \mathbf{x}'\mu_{P_{S_0}}, \quad \text{and} \quad \mathrm{Var}_{P_{S_0}}(Y) = \mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x},$$

在 $\mathbb{E}_{P_{S_0}}(H(\lambda,\ Y))$. □

引理 4。

$$V(q,\mathbf{x},a) = \left(r(q)\|\mathbf{x}\| + \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2}\right)^2 - a\gamma - \frac{\gamma^2}{4}$$

证明。 根据引理 3，对于固定的 $\mathbf{x}$ a，q，

$$V(q,\mathbf{x},a) = \inf_{\lambda\geq\|\mathbf{x}\|^2}\left\{\lambda r(q)^2 + \frac{\lambda}{\lambda-\|\mathbf{x}\|^2}\left(\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2\right) - a\gamma - \frac{\gamma^2}{4}\right\}. \tag{21}$$

令 $\alpha := \mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2$。使函数最小化的最优 λ

$$L(\lambda) = \lambda r(q)^2 + \left(\frac{\lambda}{\lambda-\|\mathbf{x}\|^2}\right)\alpha, \quad \text{subject to } \lambda \geq \|\mathbf{x}\|^2. \tag{22}$$

由下式给出

$$\lambda^* = \|\mathbf{x}\|^2 + \frac{\|\mathbf{x}\|\sqrt{\alpha}}{r(q)}, \tag{23}$$

最优值为

$$\min_{\lambda \geq \|\mathbf{x}\|^2} L(\lambda) = (r(q)\|\mathbf{x}\| + \sqrt{\alpha})^2.$$

代入 α back 的表达式，可得以下闭式解:

$$V(q, \mathbf{x}, a) = \left(r(q)\|\mathbf{x}\| + \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2} \right)^2 - a\gamma - \frac{\gamma^2}{4}. \tag{24}$$

□

结合引理 2、3、4，定理1得证

2.1.2 命题1的证明

J在点 $(\mathbf{x}_k, a_k)$ 处的一个次梯度为 $\mathbf{g}_k = \nabla_{(\mathbf{x},a)} h(q_k^*,\ \mathbf{x}_k,\ a_k)$。回顾我们的记号

$$S_\gamma(\mathbf{x}_k, a_k) = \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right)^2}.$$

计算关于 $\mathbf{x}$ 的次梯度:

$$\nabla_{\mathbf{x}} \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) = 2(\Sigma_{P_N} + \mu_{P_N}\mu'_{P_N})\mathbf{x}_k - 2a_k\mu_{P_N} - \gamma\mu_{P_N}.$$

$$\begin{aligned}
\nabla_{\mathbf{x}} V(q_k^*, \mathbf{x}_k, a_k) &= 2\left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right) \cdot \nabla_{\mathbf{x}} \left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right) \\
&= 2\left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right) \left(r(q_k^*)\frac{\mathbf{x}_k}{\|\mathbf{x}_k\|} + \nabla_x S_\gamma(\mathbf{x}_k, a_k) \right),
\end{aligned}$$

其中

$$\nabla_x S_\gamma(\mathbf{x}_k, a_k) = \frac{\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right) \cdot \mu_{P_{S_0}}}{S_\gamma(\mathbf{x}_k, a_k)}$$

相对于$\mathbf{x}$ 而言， J 的次梯度是以下加权和:

$$\begin{aligned}
\mathbf{g}_{k,\mathbf{x}} = (1 - q_k^*)\left[2(\Sigma_{P_N} + \mu_{P_N}\mu'_{P_N})\mathbf{x}_k - 2a_k\mu_{P_N} - \gamma\mu_{P_N}\right] + \\
q_k^* \left[2\left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right) \left(r(q_k^*)\frac{\mathbf{x}_k}{\|\mathbf{x}_k\|} + \frac{\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right) \cdot \mu_{P_{S_0}}}{S_\gamma(\mathbf{x}_k, a_k)} \right) \right].
\end{aligned}$$

计算关于 a的次梯度:

$$\frac{\partial}{\partial a} \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) = -2(\mathbf{x}'_k\mu_{P_N} - a_k).$$

$$\frac{\partial S_\gamma(\mathbf{x}_k, a_k)}{\partial a_k} = \frac{1}{2S_\gamma(\mathbf{x}_k, a_k)} \cdot 2\left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right)\cdot\left(-\frac{2}{2}\right) = -\frac{1}{S_\gamma(\mathbf{x}_k, a_k)}\left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right).$$

对 V应用链式法则:

$$\begin{aligned}\frac{\partial}{\partial a_k} V(q_k^*, \mathbf{x}_k, a_k) &= 2\left(r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)\right)\frac{\partial S_\gamma(\mathbf{x}_k, a_k)}{\partial a_k} - \gamma \\ &= -2\,\frac{r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)}{S_\gamma(\mathbf{x}_k, a_k)}\left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+k}{2}\right) - \gamma.\end{aligned}$$

相对于 a， J 的次梯度是以下加权和:

$$g_{k,a} = (1-q_k^*)[-2(\mathbf{x}_k'\mu_{P_N} - a_k)] + q_k^*\left[-2\,\frac{r(q_k^*)\|\mathbf{x}_k\| + S_\gamma(\mathbf{x}_k, a_k)}{S_\gamma(\mathbf{x}_k, a_k)}\left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a_k+\gamma}{2}\right) - \gamma\right].$$

2.1.3 引理 1 的证明

证明。 i. 为了证明 h 在 (**x** a) 上是强凸的，我们注意到

$$\begin{aligned}\nabla^2_{(\mathbf{x},a)}\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) &= \nabla^2_{(\mathbf{x},a)}\mathbb{E}_{P_N}\left((\mathbf{x}'\mathbf{R}-a)^2 - \gamma\mathbf{x}'\mathbf{R}\right) \\ &= 2\mathbb{E}_{P_N}\left(\begin{pmatrix}\mathbf{R}\\ -1\end{pmatrix}(\mathbf{R}', -1)\right) \\ &= \begin{pmatrix}\Sigma_{P_N} + \mu_{P_N}\mu_{P_N}' & -\mu_{P_N}\\ -\mu_{P_N}' & 1\end{pmatrix}.\end{aligned}$$

对该海森矩阵进行分块分解 $M = \begin{pmatrix}A & b\\ b' & c\end{pmatrix}$， 其中

$$A = \Sigma_{P_N} + \mu_{P_N}\mu_{P_N}',\ b = -\mu_{P_N},\ \text{and}\ c = 1,$$

矩阵 M 的分块 c 的舒尔补为 $A - bc^{-1}b' = \Sigma_{P_N}$，根据假设它是正定的；由于 $c > 0$， 由舒尔补判据可得$M \succ 0$。令 $t := \frac{2a+\gamma}{2}$ 且 $z := (\mathbf{x}\ t)'$，则有

$$S_\gamma(\mathbf{x}, a)^2 = \mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + (\mathbf{x}'\mu_{P_{S_0}} - t)^2 = z'\begin{pmatrix}\Sigma_{P_{S_0}} + \mu_{P_{S_0}}\mu_{P_{S_0}}' & -\mu_{P_{S_0}}\\ -\mu_{P_{S_0}}' & 1\end{pmatrix}z =: z'Bz,$$

根据舒尔补判据得到 $B \succ 0$。由于 $\mathbf{x}$ 受限于概率单纯形 Δ，定义域中任意两个不同点都不会位于经过原点的同一条正射线上。因此， $S_\gamma(\mathbf{x}\ a) = \sqrt{z'Bz}$作为一个范数，在定义域上是严格凸的，不过凸性已足以满足我们的推导需求。由此可得

$$V(q, \mathbf{x}, a) = \left(r(q)\|\mathbf{x}\| + S_\gamma(\mathbf{x}, a)\right)^2$$

是关于 (x a) 的凸函数，因为递增凸函数与凸函数复合后仍为凸函数。对于 $q < 1$, 现在可以得出结论： h 在 (x a) 上是强凸的，因为

$$\nabla^2_{(\mathbf{x},a)} h(q, \mathbf{x}, a) \succeq (1-q)\nabla^2_{(\mathbf{x},a)} \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) = (1-q)M.$$

因此，对于固定的 $q < 1$, h 在 (x a) 上是 ($(1-q)\lambda_M$)-强凸的，其中 λ_M 是 M 的最小特征值。

ii. 由于我们假设 $q_0 + \varepsilon < 1$, J 是强凸函数的上确界，因此它是 ($(1 - q_0 - \varepsilon)\lambda_M$)-强凸的。对于每个 (x a), J 的次梯度均由 K_1 界定，因此 J 是 K_1-利普希茨的。

iii. h 的强凹性意味着 $\operatorname{argmax}_q h(q,$ x $a)$ 是单元素集合，不妨设其包含 q^*(x a)。根据 Danskin 定理，我们知道 ∇J(x a) $= \nabla_{(\mathbf{x},a)} h(q^*$(x a), x $a)$。由于 ∇h 是利普希茨的，只需证明 q^* 是利普希茨的。定义

$$F(q, \mathbf{x}, a) = h_q(q, \mathbf{x}, a).$$

由于 $\frac{\partial^2}{\partial q^2} h$ 在紧域上是连续函数，且 h 是强凹的，因此对于某个适当的正数常量 α, 有 $F_q = h_{qq} \le -\alpha < 0$ 。

回顾一下

$$\left\| \frac{\partial}{\partial(\mathbf{x}, a)} F \right\| = \left\| \frac{\partial^2}{\partial q \partial(\mathbf{x}, a)} h \right\| \le K_3.$$

如果对所有 x a 都有 $F(q^*$(x a), x $a) = 0$，则不难看出 q^* 是一个$\frac{K_3}{\alpha}$–利普希茨函数。然而，由于 q位于有界区间内，上述一阶条件未必成立。尽管如此，我们将利用考虑了区间约束的KKT条件来证明这一点成立。考虑 (x1, a_1) 与 (x2, a_2), 其中 q^* (x1, a_1) $< q^*$(x2, a_2)。由于 q^*(x1, a_1) 可能等于 $q_0 - \varepsilon$, 且 q^*(x2, a_2) 可能等于$q_0 + \varepsilon$, KKT条件意味着

$$F(q^*(\mathbf{x}_1, a_1), \mathbf{x}_1, a_1) - F(q^*(\mathbf{x}_2, a_2), \mathbf{x}_2, a_2) \le 0.$$

因此，利用中值定理，我们有，

$$\begin{aligned}
0 &\ge F(q^*(\mathbf{x}_1, a_1), \mathbf{x}_1, a_1) - F(q^*(\mathbf{x}_2, a_2), \mathbf{x}_2, a_2) \\
&= (F(q^*(\mathbf{x}_1, a_1), \mathbf{x}_1, a_1) - F(q^*(\mathbf{x}_1, a_1), \mathbf{x}_2, a_2)) \\
&\quad + (F(q^*(\mathbf{x}_1, a_1), \mathbf{x}_2, a_2) - F(q^*(\mathbf{x}_2, a_2), \mathbf{x}_2, a_2)) \\
&= \frac{\partial F}{\partial(\mathbf{x}, a)}(q^*(\mathbf{x}_1, a_1), \mathbf{x}_0, a_0) \cdot ((\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2)) \\
&\quad + F_q(q_0, \mathbf{x}_2, a_2) \cdot (q^*(\mathbf{x}_1, a_1) - q^*(\mathbf{x}_2, a_2)) \\
&\ge -K_3 \left\| (\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2) \right\| + \alpha((q^*(\mathbf{x}_2, a_2) - q^*(\mathbf{x}_1, a_1))),
\end{aligned}$$

其中 ($\mathbf{x}_0, a_0$) 是位于 ($\mathbf{x}_1, a_1$) 和 ($\mathbf{x}_2, a_2$) 之间线段上的一个点，并且

$$q_0 \in (q^*(\mathbf{x}_1, a_1), q^*(\mathbf{x}_2, a_2)).$$

因此，

$$|q^*(\mathbf{x}_1, a_1) - q^*(\mathbf{x}_2, a_2)| \leq \frac{K_3}{\alpha} \|(\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2)\|$$

因此， q^* 是一个 $\frac{K_3}{\alpha}$—利普希茨函数。回顾一下 $\nabla J(\mathbf{x}\ a) = \nabla_{(\mathbf{x},a)} h(q^*(\mathbf{x}\ a), \mathbf{x}\ a)$。根据向量值映射的中值定理（参见 Apostol[1974] 中的定理 12.9），

$$\begin{aligned}
&\|\nabla J(\mathbf{x}_1, a_1) - \nabla J(\mathbf{x}_2, a_2)\| \\
&\leq \left\|\nabla_{(\mathbf{x},a)} h(q^*(\mathbf{x}_1, a_1), \mathbf{x}_1, a_1) - \nabla_{(\mathbf{x},a)} h(q^*(\mathbf{x}_2, a_2), \mathbf{x}_1, a_1)\right\| \\
&\quad + \left\|\nabla_{(\mathbf{x},a)} h(q^*(\mathbf{x}_2, a_2), \mathbf{x}_1, a_1) - \nabla_{(\mathbf{x},a)} h(q^*(\mathbf{x}_2, a_2), \mathbf{x}_2, a_2)\right\| \\
&\leq K_3 \cdot \frac{K_3}{\alpha} \|(\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2)\| + K_2 \|(\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2)\| \\
&\leq \left(\frac{K_3^2}{\alpha} + K_2\right) \|(\mathbf{x}_1, a_1) - (\mathbf{x}_2, a_2)\|,
\end{aligned}$$

这里我们回顾一下.....的定义

$$K_2 := \max_{(q,\mathbf{x},a)\in\Omega} \left\| \frac{\partial^2}{\partial(\mathbf{x}, a)^2} h \right\|_2.$$

因此， ∇J 是 $\left(\frac{K_3^2}{\alpha} + K_2\right)$-利普希茨的。 J 是 $\left(\frac{K_3^2}{\alpha} + K_2\right)$-光滑的。

□

2.1.4 定理 3 的证明

证明。我们首先 s 存在 ε_0，使得 h 关于 $[q_0 - \varepsilon_0, q_0 + \varepsilon_0]$ 在 q 上是强凹的 .

$$\begin{aligned}
h_{qq}(q, \mathbf{x}, a) &= 2V_q(q, \mathbf{x}, a) + qV_{qq}(q, \mathbf{x}, a) \\
&= 4r'(q)\Big(r(q)\|\mathbf{x}\|^2 + \|\mathbf{x}\| S_\gamma(\mathbf{x}, a)\Big) + 2qr''(q)\Big(r(q)\|\mathbf{x}\|^2 + \|\mathbf{x}\| S_\gamma(\mathbf{x}, a)\Big) \\
&\quad + 2q\big(r'(q)\big)^2\|\mathbf{x}\|^2.
\end{aligned}$$

Let

$$K_5 := \max_{q\in[0,1]} r(q), \text{ and } K_6 := \max_{\mathbf{x}\in\Delta, a\in I} S_\gamma(\mathbf{x}, a).$$

定义一个辅助函数

$$\bar{h}_{qq}(q) := 4\,|r'(q)|\,(K_5 + K_6) + 2qr''(q)r(q)/d + 2q\big(r'(q)\big)^2.$$

由于 由于 r'' 连续，存在一个 $\varepsilon_1 > 0$，使得 $h_{qq}(q) \geq h_{qq}(\{\mathbf{v}7, \mathbf{x}\ a)$ 在

$$q \in [q_0 - \varepsilon_1, q_0 + \varepsilon_1]$$

对于所有 $\mathbf{x} \in \Delta$， $a \in I$。

由于 $r'(q_0) = 0$，我们有

$$\bar{h}_{qq}(q_0) = 2q_0 r(q_0) r''(q_0)/d < 0.$$

由于 h_{qq} 连续，因此存在 $\varepsilon_2 > 0$，使得 $h_{qq}(q) \leq -K_4 < 0$ 在 $q \in [q_0 - \varepsilon_2, q_0 + \varepsilon_2]$上成立。取 $\varepsilon_0 = \varepsilon_1 \wedge \varepsilon_2$ 可知， h 关于 $[q_0 - \varepsilon_0, q_0 + \varepsilon_0]$在 q 上是强凹的。结合引理 1 ii.、iii. 和定理 12，即得证明。 □

2.1.5 定理4的证明

证明。 我们首先注意到，单纯形中范数最小的向量显然是$\mathbf{x} = (1/d) \cdot \mathbf{1}$，对应于等权投资组合。

$$\begin{aligned}\frac{h(q,\mathbf{x},a)}{C^2} &= \frac{(1-q)\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R}))}{C^2} + q\left(r_{\text{base}}(q)^2\|\mathbf{x}\|^2 + \frac{2r_{\text{base}}(q)\|\mathbf{x}\|S_\gamma(\mathbf{x},a)}{C}\right.\\ &\qquad\left. + \frac{S_\gamma(\mathbf{x},a)^2 - a\gamma - \gamma^2/4}{C^2}\right)\\ &\Rightarrow q\cdot r_{\text{base}}(q)^2\|\mathbf{x}\|^2,\end{aligned}$$

当 $C \to \infty$ 时，由于分子中其他项有界，因此在 q，$\mathbf{x}$ a 上一致收敛。

我们令 $C > 1$， $K_7 < \infty$ ，使得

$$\left|\frac{(1-q)\mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R}))}{C^2} + q\left(\frac{2r_{\text{base}}(q)\|\mathbf{x}\|S_\gamma(\mathbf{x},a)}{C} + \frac{S_\gamma(\mathbf{x},a)^2 - a\gamma - \gamma^2/4}{C^2}\right)\right| \le \frac{K_7}{C}.$$

由于 r_{base} 的连续性，我们记

$$\max_{q\in[q_0-\varepsilon,q_0+\varepsilon]} q\cdot r_{base}(q)^2 = K_8 > 0,$$

其中 q^* 为最大点。记非空集合 $I_q = \{q\colon\ q\cdot r_{\text{base}}(q)^2 \ge \frac{K_8}{2}\}$。当 $C > \frac{4K_7 n}{K_8} \vee 1$ 时，很容易看出

$$(\mathbf{x}_C^*, a_C^*) = \operatorname*{argmin}_{\mathbf{x}\in\Delta, a\in I} \max_{q\in I_q} \frac{h(q,\mathbf{x},a)}{C^2},$$

对于固定的 $\mathbf{x}$ a 和 $q_0 \notin I_q$,

$$\frac{h(q^*,\mathbf{x},a)}{C^2} - \frac{h(q_0,\mathbf{x},a)}{C^2} \ge \frac{K_8}{2}\cdot\frac{1}{d} - 2\frac{K_7}{C} \ge 0.$$

令 $\mathbf{x}^* = (\frac{1}{d},\ \ldots\ldots,\ \frac{1}{d})$ 且 $\mathbf{1}'\boldsymbol{\eta} = 0$。对于任意 $q \in I_q$, a, $a' \in I$如果 $\|\boldsymbol{\eta}\| > \sqrt{\frac{4K_7}{CK_8}}$，

$$\frac{h(q,\mathbf{x}^*+\boldsymbol{\eta},a')}{C^2} - \frac{h(q,\mathbf{x},a^*)}{C^2} \ge \frac{K_8}{2}\|\boldsymbol{\eta}\|^2 - 2\frac{K_7}{C} > 0.$$

因此，$\mathbf{x}^* + \boldsymbol{\eta}$ 不可能是最小点。因此，随着 C 增加，最小点

$$\mathbf{x}_C^* \in B_{\sqrt{\frac{4K_7}{CK_8}}}(\mathbf{x}^*),$$

保证其位于等权重投资组合周围的一个不断缩小的球内，半径$\sqrt{\frac{4K_7}{CK_8}}$ 逐渐趋于 0。

□

3 均值-CVaR效用损失函数

方差作为一种风险度量存在局限性；例如，它对均值两侧的偏离一视同仁，不具正齐次性，也不具平移等变性。出于这一原因及其他原因，保险监管更倾向于采用一致性风险度量——条件风险价值（CVaR），它也称为尾部风险价值（T-VaR）和条件尾部期望（CTE）。本节将探讨上一节问题的一种替代表述，其中以 CVaR 取代方差作为风险度量。

CVaR 有几种不同的定义；在本文的推导中，Rockafellar 等人 [2000] 的以下变分定义最为相关：

定义 1. 设 X 为具有有限均值的随机变量。那么对于 $p \in (0,\ 1)$，p-水平下 X 的条件风险价值记为 $\mathrm{CVaR}_p(X)$，其定义如下：

$$\mathrm{CVaR}_p\left(X\right) = \inf_{\tau\in\mathbb{R}} \left\{ \tau + \frac{1}{1-p}\mathbb{E}(\max(X-\tau, 0)) \right\}. \tag{25}$$

同样，收益$\mathbf{R}$ 的真实分布受限于双成分混合结构，其中成分分布和权重均未知。投资者的均值-CVaR 效用损失函数是期望损失与置信水平为 $p \in (0,\ 1)$ 时的条件风险价值的加权和：

$$\mathbb{E}(-\mathbf{x}'\mathbf{R}) + \rho\mathrm{CVaR}_p\left(-\mathbf{x}'\mathbf{R}\right), \tag{26}$$

其中，$\rho > 0$ 为风险厌恶参数。虽然收益的真实分布（记为 P）未知，但投资者的信念将其限制在模糊集 $\mathcal{S}$中。投资者的稳健投资组合设计问题可归结为求解以下优化问题：

$$\inf_{\mathbf{x}:\mathbf{x}\in\Delta} \sup_{P\in\mathcal{S}} \left(\mathbb{E}_P(-\mathbf{x}'\mathbf{R}) + \rho\mathrm{CVaR}_p\left(-\mathbf{x}'\mathbf{R}\right)\right). \tag{Opt 2}$$

与第4节类似，我们将模糊集 $\mathcal{S}$ 定义如下。

$$\mathcal{S} = \{(1-q)P_N + qP_S \mid q \in [q_0-\epsilon, q_0+\epsilon],\ P_S \in \mathbb{B}_{W_1}(P_{S_0}, r(q))\}. \tag{27}$$

与式 (4) 的唯一区别在于，以 P_{S_0} 为中心、半径为 $r(q)$ 的 Wasserstein-1球。我们假设两种分布 P_N 和 P_{S_0} 下的资产收益具有有限的一阶矩，即 $\mathbb{E}_{P_N}(\|\mathbf{R}\|) < \infty$ 和 $\mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|) < \infty$。由于对于 $\mathcal{S}$ 中的任意分布，$-\mathbf{x}'\mathbf{R}$ 都具有有限均值，因此其均值-CVaR负效用对这类分布也有明确定义。本节的主要结果如下：

定理5. 定义损失函数 $\ell_{\mathbf{x},\tau}(\mathbf{R}) := -\mathbf{x}'\mathbf{R} + \frac{\rho}{1-p}$最大值$(-\mathbf{x}'\mathbf{R} - \tau,\ 0)$。(优化问题2) 等价于以下问题：

$$\min_{\mathbf{x}\in\Delta,\tau\in I} J(\mathbf{x},\tau), \tag{28}$$

其中

$$J(\mathbf{x},\tau) = \left\{ \rho\tau + \sup_{q\in[q_0\pm\epsilon]} \left((1-q)\mathbb{E}_{P_N}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + q\left[\mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)(1+\frac{\rho}{1-p})\|\mathbf{x}\|_\infty \right] \right) \right\}. \tag{29}$$

此外， J关于 $(\mathbf{x}\ \tau)$是凸的。

□

上述定理的证明将在下一小节给出。其陈述中的目标函数J 是一个不可微凸函数；因此，为了最小化该函数，我们提出如下算法 2 所示的投影次梯度下降算法。

算法 2 鲁棒均值-CVaR 的投影次梯度下降

1: **初始化：**选择 $(\mathbf{x}_0, \tau_0) \in \Delta \times \mathbb{R}$，迭代次数 $T > 0$ 和步长$\eta_k > 0$，其中 $k = 0, 1, \ldots, T$。2: **对于** $k = 0, 1, \ldots, T$ **执行** 3: 求出最坏情况概率 q_k^*。4: 计算 J 在 $(\mathbf{x}_k.\tau_k)$ 处的次梯度分量：$\mathbf{g}_{k,\ \mathbf{x}}$， $g_{k,\ \tau}$。5: 更新变量：

$$\mathbf{x}_{k+} \leftarrow \mathbf{x}_k - \eta_k \mathbf{g}_{k,\mathbf{x}}$$
$$\tau_{k+1} \leftarrow \tau_k - \eta_k g_{k,\tau}$$

6: 将 $\mathbf{x}_{k+1} \leftarrow$ 投影到 $\text{Proj}_\Delta(\mathbf{x}_{k+})$。7: **结束循环**8: **返回**$(\mathbf{x}_1, \tau_1) \ldots, (\mathbf{x}_T, \tau_T)$。

在每次迭代 k 中，给定 $(\mathbf{x}_k, \tau_k)$，我们首先求出在 (29) 中达到上确界的 q_k^*。根据 Danskin 定理，次梯度 $\mathbf{g}_k \in \partial J(\mathbf{x}_k, \tau_k)$ 由在 q_k^* 处求得的内部函数的一个次梯度给出。下面给出关于 J 的次梯度的命题。

命题 2。 J 在$(\mathbf{x}_k.\tau_k)$ 处的一个次梯度 $\mathbf{g}_k$ 具有以下组成部分：

$$\begin{aligned}\mathbf{g}_{k,\mathbf{x}} = &-(1-q_k^*)\left[\mathbb{E}_{P_N}(\mathbf{R}) + \left(\frac{\rho}{1-p}\right)\mathbb{E}_{P_N}(\mathbf{R}\cdot\mathbf{1}_{-\mathbf{x}'\mathbf{R}>\tau})\right] \\ &+ q_k^*\left[-\left(\mathbb{E}_{P_{S_0}}(\mathbf{R}) + \left(\frac{\rho}{1-p}\right)\mathbb{E}_{P_{S_0}}(\mathbf{R}\cdot\mathbf{1}_{-\mathbf{x}'\mathbf{R}>\tau})\right) + r(q)(1+\frac{\rho}{1-p})g_\infty(\mathbf{x})\right],\end{aligned}$$

其中 $g_\infty(\mathbf{x}) \in \partial\|$ 为 $\mathbf{x}\|_\infty$，并且

$$g_{k,\tau} = \rho - (1-q_k^*)\left(\frac{\rho}{1-p}\right)P_N(-\mathbf{x}'\mathbf{R} > \tau) - q_k^*\left(\frac{\rho}{1-p}\right)P_{S_0}(-\mathbf{x}'\mathbf{R} > \tau).$$

□

注意，在命题 2 的陈述中， $\|\mathbf{x}\|_\infty$ 的次梯度集合记为 $\partial\|\mathbf{x}\|_\infty$，其定义为

$$\partial\|\mathbf{x}\|_\infty := \text{conv}(\{\text{sign}(\mathbf{x}_i)e_i : |\mathbf{x}_i| = \|\mathbf{x}\|_\infty\}),$$

其中， e_i 是 $\mathbb{R}^n$ 的第 i个标准基向量；对于集合 A，conv(A)表示其凸包。下面的定理为算法 2提供收敛保证。

定理6。令 $K_6 = \mathbb{E}_{P_N}(\|\mathbf{R}\|)$ $K_7 = \mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|)$ $K_8 =$最大值$^{q\in[0,1]}\, r(q)$ 且

$$K_9 = \sqrt{\left[\left(1+\frac{\rho}{1-p}\right)\left((K_6 \vee K_7)+K_8\right)\right]^2 + \rho^2\left(\frac{p}{1-p}\vee 1\right)^2}.$$

设 R 满足 $\|(\mathbf{x}_0,\tau_0) - (\mathbf{x}^*,\tau^*)\| \le R$。若在算法 *2* 中采用恒定步长 $\eta_k = \frac{R}{K_9\sqrt{T}}$，则有

$$J\left(\frac{1}{T}\sum_{i=1}^{T}(\mathbf{x}_i,\tau_i)\right) - J(\mathbf{x}^*,\tau^*) \le \frac{RK_9}{\sqrt{T}}.$$

换言之，算法 *2* 的收敛速率至少为 $O(1/\sqrt{T})$。

证明。

$$\|\mathbf{g}_{k,\mathbf{x}}\| \le \left(1+\frac{\rho}{1-p}\right)\left((K_6\vee K_7)+K_8\right),\ |g_{k,\tau}| \le \rho \vee \left(\frac{\rho}{1-p}-\rho\right) = \rho\cdot\left(\frac{p}{1-p}\vee 1\right),$$

其中我们对 2-范数应用了詹森不等式。因此， J 是 K_9-利普希茨的。由定理10可完成证明。

□

最后，定理7表明，随着模糊性增大，$\mathbf{x}^*$趋近于等权重投资组合 $(\frac{1}{d},\ldots,\frac{1}{d})$。由于证明与第2节中的定理4类似，此处省略证明。

定理7。 设$(\mathbf{x}_C^*, a_C^*)$ 是 J 的一个极小值点，其中式(27) 所定义的模糊性半径为$r(q) = C\cdot r_{base}(q)$，其中 $C>0$ 是缩放参数， r_{base} 是区间 $[0,1]$, 上一个固定的非负连续函数，且在 $[q_0-\varepsilon, q_0+\varepsilon]$ 上不恒为零。那么，当 $C\to\infty$ 时，$\mathbf{x}_C^*$ 收敛到等权重投资组合 $\frac{1}{d}\cdot\mathbf{1}$。

3.1 结果证明

为证明定理5，我们需要以下两个引理。

引理5。 均值-*CVaR*问题（优化问题2）具有以下等价刻画：

$$\inf_{\mathbf{x}\in\Delta}\sup_{P\in\mathcal{S}}\left(\mathbb{E}_P(-\mathbf{x}'\mathbf{R}) + \rho\mathrm{CVaR}_p\left(-\mathbf{x}'\mathbf{R}\right)\right)$$
$$= \inf_{\mathbf{x}\in\Delta,\tau\in I}\left\{\rho\tau + \sup_{P\in\mathcal{S}}\mathbb{E}_P\left(-\mathbf{x}'\mathbf{R}+\frac{\rho}{1-p}\max(-\mathbf{x}'\mathbf{R}-\tau,0)\right)\right\}, \quad (30)$$

其中 I 是一个与$\mathbf{x}$ 和 P无关的紧区间。

证明。 根据定义（25）：

$$\sup_{P\in\mathcal{S}}\left(\mathbb{E}_P(-\mathbf{x}'\mathbf{R}) + \rho\mathrm{CVaR}_p\left(-\mathbf{x}'\mathbf{R}\right)\right) = \sup_{P\in\mathcal{S}}\inf_{\tau\in\mathbb{R}}\left\{\rho\tau + \mathbb{E}_P\left(-\mathbf{x}'\mathbf{R}+\frac{\rho}{1-p}\max(-\mathbf{x}'\mathbf{R}-\tau,0)\right)\right\},$$

将损失随机变量记为 $L := -\mathbf{x}'\mathbf{R}$。众所周知，对于 $p \in (0,1)$，使τ达到下确界的集合等于 $[\vec{F}_L(1-p),\ F_L(1-p)],$，其中

$$F_L(1-p) := \sup\{x : F_L(x) \le 1-p\}, \text{ and } \vec{F}_L(1-p) := \inf\{x : F_L(x) \ge 1-p\},$$

分别是 F_L（ L的累积分布函数）的右广义逆和左广义逆，它取决于$\mathbf{x}$ 和P。右广义逆和左广义逆分别是非递减且右连续、非递减且左连续的。我们首先证明，内部优化可以限制在一个紧区间 I 内，该区间不包含 $P \in \mathcal{S}$ 和$\mathbf{x} \in \Delta$。只需证明存在一个 I ，使其对所有 $P \in \mathcal{S}$和$\mathbf{x} \in \Delta$都包含$[\vec{F}_L(1-p),\ F_L(1-p)]$。由于f连续，令 M为 $[q_0-\varepsilon, q_0+\varepsilon]$上 f的最大值。那么，对于 $\mathbb{B}_{W_1}$中的任意 P_S（P_{S_0}，M）以及$\mathbf{x} \in \Delta$，若记 π为 P_S 与 P_{S_0} 之间最优的1-Wasserstein耦合，且$(\mathbf{R},\ \mathbf{R}_0) \sim \pi$，则有

$$\mathbb{E}_{P_S}(\|\mathbf{R}\|_1) = \mathbb{E}_\pi(\|\mathbf{R}_0 + \mathbf{R} - \mathbf{R}_0\|_1) \le \mathbb{E}_\pi(\|\mathbf{R}_0\|_1) + \mathbb{E}_\pi(\|\mathbf{R} - \mathbf{R}_0\|_1) \le \mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|_1) + M,$$

and

$$\mathbb{E}_{P_S}(|\mathbf{x}'\mathbf{R}|) \le \|\mathbf{x}\|_\infty \mathbb{E}_{P_S}(\|\mathbf{R}\|_1) \le \mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|_1) + M.$$

同理,

$$\mathbb{E}_{P_N}(|\mathbf{x}'\mathbf{R}|) \le \|\mathbf{x}\|_\infty \mathbb{E}_{P_N}(\|\mathbf{R}\|_1) \le \mathbb{E}_{P_N}(\|\mathbf{R}\|_1).$$

由于 $\mathcal{S}$中的任意 P都是 P_N与 $\mathbb{B}_{W_1}$（P_{S_0}，M）中某个 P_S的混合分布，且混合权重 q属于 $[q_0-\varepsilon, q_0+\varepsilon]$，结合上述结论可得，

$$\mathbb{E}_P(|\mathbf{x}'\mathbf{R}|) = (1-q)\mathbb{E}_{P_N}(|\mathbf{x}'\mathbf{R}|) + q\mathbb{E}_{P_S}(|\mathbf{x}'\mathbf{R}|) \le (\mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|_1) + M) \vee \mathbb{E}_{P_N}(\|\mathbf{R}\|_1).$$

对于任意 $t \in [\vec{F}_L(1-p),\ F_L(1-p)]$,

$$|t| \le \left(\frac{1}{1-p} \vee \frac{1}{p}\right) \cdot \left[(\mathbb{E}_{P_{S_0}}(\|\mathbf{R}\|_1) + M) \vee \mathbb{E}_{P_N}(\|\mathbf{R}\|_1)\right].$$

上述内容表明，我们所需的 I存在。

由于 I是紧集，$\rho\tau + \mathbb{E}_P\left(-\mathbf{x}'\mathbf{R} + \frac{\rho}{1-p}\max(-\mathbf{x}'\mathbf{R} - \tau, 0)\right)$关于$\tau$是凸且连续的，并且关于 P相对于 1-Wasserstein度量拓扑是凹（线性）且连续的（关于 W_1 收敛的含义，参见 Villani [2003] 中的定理 7.12），由 Sion 极小极大定理（参见附录中的定理 14），我们可以交换 min 和 sup 运算符，从而完成证明。

□

T问题的核心是求解内层上确界 $\sup_{P\in\mathcal{S}} \mathbb{E}_P(\ell_{\mathbf{x},\tau}(\mathbf{R})$).

引理 6. $\ell_{\mathbf{x},\tau}$ 关于 L_1 范数是利普希茨连续的，其利普希茨常数为

$$\left(1 + \frac{\rho}{1-p}\right)\|\mathbf{x}\|_\infty$$

证明。 L_1-范数的对偶范数是 L_∞-范数。利普希茨常数是损失函数梯度的对偶范数最大值。损失函数关于向量$\mathbf{R}$ 的梯度为:

$$\nabla_{\mathbf{R}}\ell_{\mathbf{x},\tau}(\mathbf{R}) = \begin{cases} -\mathbf{x} & \text{if } -\mathbf{x}'\mathbf{R} - \tau < 0 \\ -\left(1 + \frac{\rho}{1-p}\right)\mathbf{x} & \text{if } -\mathbf{x}'\mathbf{R} - \tau > 0 \end{cases} \tag{31}$$

利普希茨常数是该梯度的最大 L_∞-范数：

$$\mathrm{Lip}(\ell_{\mathbf{x},\tau}) = \max\left(\|-\mathbf{x}\|_\infty, \left\|-\left(1+\frac{\rho}{1-p}\right)\mathbf{x}\right\|_\infty\right) = \left(1+\frac{\rho}{1-p}\right)\|\mathbf{x}\|_\infty.$$

□

引理7。 对于固定的$\mathbf{x}$ τ, q， *Wasserstein-1*球上损失的最坏情形期望：

$$V(q,\tau,\mathbf{x}) := \sup_{P_S\in\mathbb{B}_{W_1}(P_{S_0},r(q))} \mathbb{E}_{P_S}(\ell_{\mathbf{x},\tau}(\mathbf{R})) = \mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)\cdot Lip(\ell_{\mathbf{x},\tau}). \tag{32}$$

证明。 根据（8），

$$V(q,\tau,\mathbf{x}) = \inf_{\lambda\geq 0}\left\{\lambda r(q) + \mathbb{E}_{P_{S_0}}\left(\sup_{\mathbf{r}}\left(\ell_{\mathbf{x},\tau}(\mathbf{r}) - \lambda\|\mathbf{r}-\mathbf{R}\|_1\right)\right)\right\},$$

其中$\mathbf{R}\sim P_{S_0}$。注意，当 $\lambda \geq$**Lip(**$\ell_{\mathbf{x},\tau}$**)**时，$\sup_{\mathbf{r}}\{\ell_{\mathbf{x},\tau}(\mathbf{r})-\lambda\|\mathbf{r}-\mathbf{R}\|_1\} = \ell_{\mathbf{x},\tau}(\mathbf{R})$；否则，上确界为无穷大。因此，对偶问题变为：

$$V(q,\tau,\mathbf{x}) = \inf_{\lambda\geq \mathrm{Lip}(\ell_{\mathbf{x},\tau})}\left\{\lambda r(q) + \mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R}))\right\} = \mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)\cdot\mathrm{Lip}(\ell_{\mathbf{x},\tau}).$$

□

备注。 上述引理并不令人意外，因为根据 *Wasserstein-1* 距离的康托罗维奇-鲁宾斯坦对偶，

$$d_{W_1}(P_{S_0},P_S)\cdot Lip(\ell_{\mathbf{x},\tau}) \geq \mathbb{E}_{P_S}(\ell_{\mathbf{x},\tau}(\mathbf{R})) - \mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R}))$$

.

因此，

$$V(q,\tau,\mathbf{x}) \leq \mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)\cdot Lip(\ell_{\mathbf{x},\tau}).$$

现在我们回到定理5的证明：

证明。 由（30）和（32）

$$\begin{aligned}
&\inf_{\mathbf{x}\in\Delta}\sup_{P\in\mathcal{S}}\left(\mathbb{E}_P(-\mathbf{x}'\mathbf{R}) + \rho\mathrm{CVaR}_p\left(-\mathbf{x}'\mathbf{R}\right)\right)\\
&= \inf_{\mathbf{x}\in\Delta,\tau\in I}\left\{\rho\tau + \sup_{P\in\mathcal{S}}\mathbb{E}_P\left(-\mathbf{x}'\mathbf{R} + \frac{\rho}{1-p}\max(-\mathbf{x}'\mathbf{R}-\tau,0)\right)\right\}\\
&= \inf_{\mathbf{x}\in\Delta,\tau\in I}\left\{\rho\tau + \sup_{q\in[q_0\pm\epsilon]}\left((1-q)\mathbb{E}_{P_N}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + q\left[\mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)(1+\frac{\rho}{1-p})\|\mathbf{x}\|_\infty\right]\right)\right\}\\
&= \inf_{\mathbf{x}\in\Delta,\tau\in I} J(\mathbf{x},\tau).
\end{aligned}$$

由于 $\ell_{\mathbf{x},\,\tau}$**(R)** $= -\mathbf{x}'\mathbf{R} + \frac{\rho}{1-p}$**max(**$-\mathbf{x}'\mathbf{R}-\tau$，**0)** 关于 **(**$\mathbf{x}$ τ**)** 是凸函数，因此 J 的凸性源于以下事实：任意范数都是凸的，且凸函数的上确界也是凸的。

根据 Berge 最大值定理， J是连续的，因为 $[q_0-\varepsilon, q_0+\varepsilon]$ 是一个紧集，且目标函数关于 **(**q，$\mathbf{x}$ τ**)** 连续。由于 J在紧集上连续，因此可以将 inf 改为 min。

□

注。以下论证可以更简单地说明 J 的连续性。注意， J 在包含 $\Delta \times I$的开球上是凸的。根据 ***Rockafellar*** 的定理 ***10.1*** *[1970]*（定义在 $\mathbb{R}^n$ 上的凸函数相对于其定义域的相对内部连续）， J 在 $\Delta \times I$上连续。

4 模拟实验

我们首先介绍生成模拟数据的设置。我们考虑 10 种资产 ($d = 10$) 的 1,000 个收益独立同分布样本，并将压力状态的真实概率设为 3% ($q = 3\%$)。在正常状态下，第 i 种资产的收益可分解为特质成分 $N(\mu = 0.03i,\ \sigma = 0.025i)$ 和系统性成分$N(\mu = 0,\ \sigma = 0.02)$，后者为所有资产共有，这与 Mohajerin Esfahani 和 Kuhn [2018] 第 7.2 节中的模拟设置相同。换言之，资产在 P_N 下的收益服从多元 $N_d(\boldsymbol{\mu},\ \Sigma)$分布，其中以 $\delta_{\cdot}$ 表示克罗内克δ函数，

$$\mu_i = 0.03i,\ \text{and}\ \sigma_{ij} = 0.02^2 + 0.025^2 i^2 \delta_{i,j}.$$

在压力状态下，资产的收益服从具有五个自由度的重尾多元 t 分布， $t_d(\boldsymbol{\mu},\ \Sigma,\ \nu = 5)$，其中

$$\mu_i = -0.05(i+1),\ \text{and}\ \sigma_{ij} = (0.1 + 0.03i) \cdot (0.1 + 0.03j) \cdot (0.7 + 0.3\delta_{ij}),$$

由此，资产之间的基准相关性高达 0.7，且标准差随指数线性递增。它还体现了一个典型事实：在正常状态下收益和方差较高的资产（例如规模较小的公司），在压力状态下往往会受到更严重的冲击。我们设定中的另一个重要特点是，资产在正常状态下尾部独立，而在压力状态下尾部相关，尾部相关系数为 0.43。图1通过对资产 1 和资产 2 的收益进行散点图展示，直观呈现了这些分布，其中包含 1,000 个样本。

现在，我们将介绍实验中考虑的不同 $\mathcal{S}$ 选择。我们允许 $\varepsilon \in \{0, 0.01, 0.02, 0.03\}$，并将半径 $r(\cdot)$ 限定为以下函数形式

$$r(q) = c \cdot q^{\alpha-1}(1-q)^{\beta-1},$$

为使 q_0 达到众数，我们选择

$$\alpha = M \cdot q_0 + 1,\ \beta = M \cdot (1 - q_0) + 1.$$

在实验中，我们选择 $M = 10$。我们特意选择了一些区间 $[q_0 - \varepsilon, q_0 + \varepsilon]$ ，使其包含 0，以检验当某些假设被违反时，我们理论的实证表现。

4.1 均值-方差效用损失

我们选择 $\gamma = 0.1$ 以表示高度风险厌恶。为分析我们的投影次梯度下降算法的收敛速率，我们选取固定步长 (η) 为 0.001，并选择

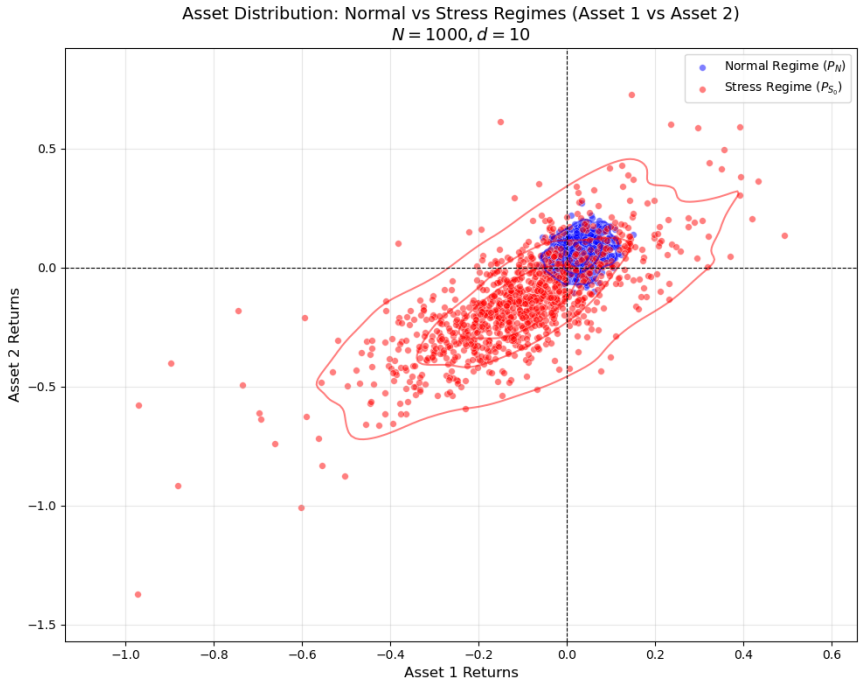

图1：资产 1 和资产 2 的 1,000 个收益样本的散点图，并展示了两种机制下联合密度核密度估计的等高线。

c、 ε 和 q_0 分别取 0.1、0.03 和 0.024。我们以迭代次数 k为横轴，在对数尺度上绘制次优差距 $J(x_k, a_k) - J^*$。由于 $0 \in [q_0 - \varepsilon, q_0 + \varepsilon]$，因此 h 在 $[q_0 - \varepsilon, q_0 + \varepsilon]$上并非强凹。然而，在图 2 中，收敛速率仍快于指数速率，因为图线位于一条直线下方。

我们将分布鲁棒优化估计量的样本外表现与标准的样本均值近似（SAA）基准方法进行比较。为构建准确的测试环境，我们从真实分布中抽取大量 3×10^6 收益样本，作为样本外分布。我们进行了 $N = 100$ 次模拟。在每次模拟中，抽取一个包含 1,000 个样本的新训练集，用于拟合分布鲁棒优化和 SAA 估计量，然后在大型样本外测试集上对它们进行评估。重复试验 100 次后，我们可以得到样本外效用损失的均值及第 20 至第 80 百分位数。图3左侧绘出了这些结果，展示了不同半区间宽度 ε下，样本外表现（左列）及相应的平均投资组合配置（右列）如何随半径参数 c 变化。图中清楚显示，分布鲁棒优化估计量的表现对这些参数较为敏感，并且在达到某个最优 c之前，往往优于 SAA（蓝线位于红色虚线下方）。绿线表示真实分布下最优投资组合的负效用。

4.2 均值-CVaR负效用

我们采用与均值-方差问题相同的设定，并像在均值-方差问题中一样，将 $\rho = 10$ 和 $p = 0.95$ 设为表示较高风险厌恶程度。图 3 右侧再次表明，

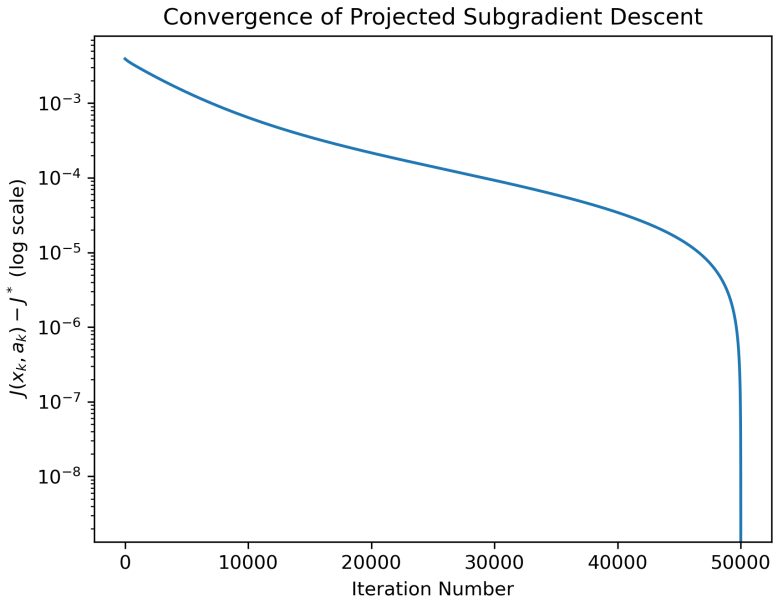

图 2：步长固定为 $\eta = 0.001$ 时投影次梯度下降算法的收敛情况，以及 $c = 0.1$，$\varepsilon = 0.03$，$q_0 = 0.024$。纵轴以对数尺度绘制次优差距 $J(x_k,\ a_k) - J^*$，横轴为迭代次数 k。尽管 h 并非强凹，该算法仍以超指数速率收敛。

分布鲁棒优化估计量的表现往往优于样本均值近似，直至 c 达到最优水平。

5 结论

本文针对具有混合结构的均值-方差和均值-CVaR分布鲁棒优化问题（优化问题1）（优化问题2），得到了极小极大形式，随后可采用投影次梯度下降法求解。我们的算法保证了 O（$\frac{1}{\sqrt{T}}$）收敛速率；对于均值-方差问题，算法还能达到指数收敛速率。在第4节中，我们给出了数值示例，表明即使某些假设不成立，仍能达到指数速率。此外，我们通过比较样本外负效用与样本平均近似方法的结果，展示了分布鲁棒优化方法的有效性。在这两种情形下，我们从理论和实证两方面表明，随着球的半径趋于无穷大，分布鲁棒优化投资组合会收敛至1/N投资组合。

我们简要讨论允许卖空的情形。通常，杠杆头寸会受到限制，即我们将投资组合限制为 $\mathbf{1}'\mathbf{x} = 1;\mathbf{x} \geq v$，其中 v 为负数。在这种情况下，所有结果，包括极小极大重构和收敛速率，仍然成立，因为 $\mathbf{x}$ 属于紧集，而且我们仍可将 a, τ 限制在紧集内。在允许无限制卖空的理想化情形下，也就是说，我们对 $\mathbf{x}$ 施加的唯一约束是 $\mathbf{1}'\mathbf{x} = 1$，我们无法将 a 和 τ 限制在紧集内。要获得与式（7）和（28）类似的极小极大重构（仅需改变极小化问题的定义域），一个关键步骤是证明模糊集 $\mathcal{S}$ 是紧的，且.....

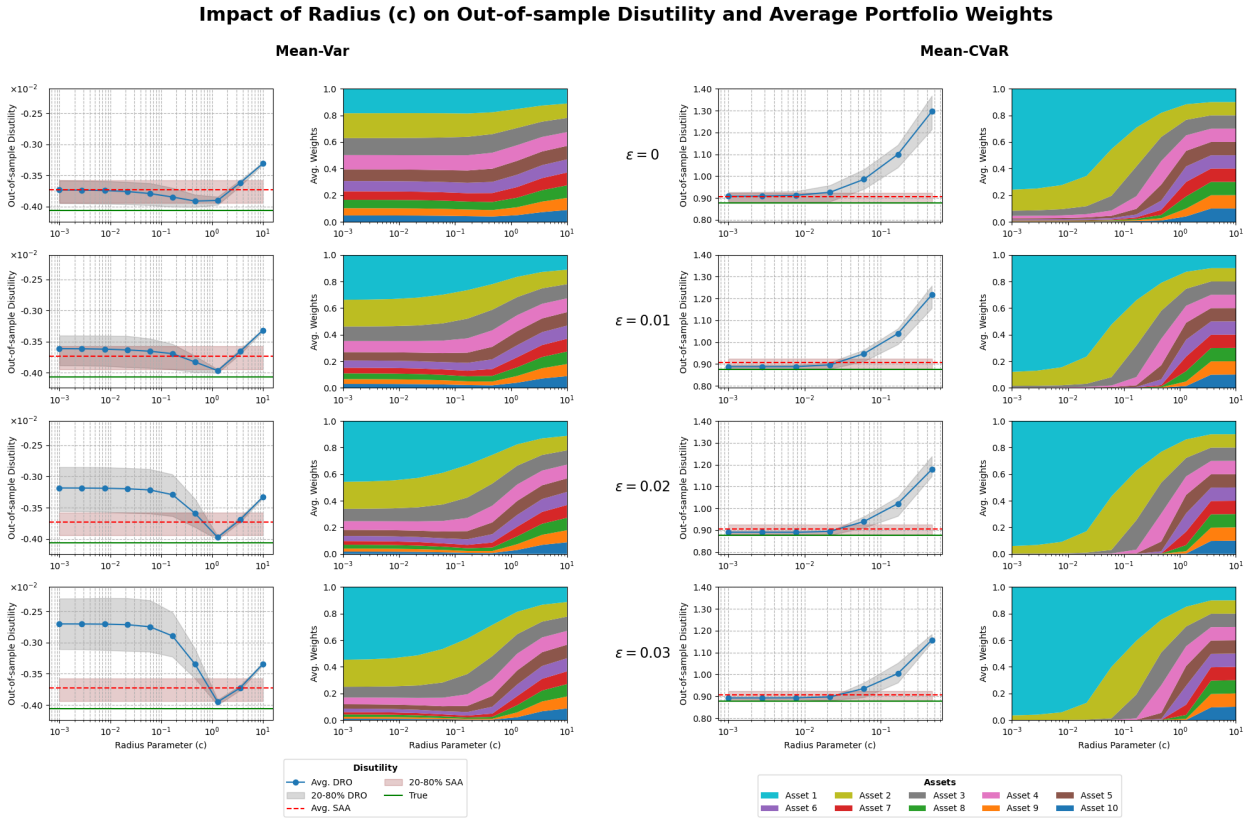

图 3：均值-方差模型（a）和均值-CVaR模型（b）的样本外表现与投资组合配置对比。每一行对应一个不同的半区间宽度 ε，取值范围为集合 $\{0, 0.01, 0.02, 0.03\}$。第一列和第三列分别显示均值分布鲁棒优化的样本外效用损失（蓝线）及其在 100 次模拟中的第 20 至第 80 百分位区间（阴影区域；每次模拟均使用 1000 个样本内数据点）。样本均值近似的平均样本外效用损失以红色虚线表示，其第 20 至第 80 百分位区间以红色阴影表示。第二列和第四列分别显示相应的平均投资组合配置。绿线表示真实分布下最优投资组合的负效用。

对应的 Wasserstein度量拓扑，并应用 Sion 极小极大定理（定理 14）。在均值-方差问题中，最终的等价改写为

$$\inf_{\mathbf{1}'\mathbf{x}=1, a\in\mathbb{R}} \max_{q\in[q_0-\varepsilon,\, q_0+\varepsilon]} \Big[(1-q)\cdot \mathbb{E}_{P_N}(\psi_{a,\gamma}(\mathbf{x}'\mathbf{R})) + q\cdot V(q,\mathbf{x},a)\Big], \tag{33}$$

其中

$$\psi_{a,\gamma}(y) := (y-a)^2 - \gamma y,$$

and

$$V(q,\mathbf{x},a) = \left(r(q)\|\mathbf{x}\| + \sqrt{\mathbf{x}'\Sigma_{P_{S_0}}\mathbf{x} + \left(\mathbf{x}'\mu_{P_{S_0}} - \frac{2a+\gamma}{2}\right)^2}\right)^2 - a\gamma - \frac{\gamma^2}{4},$$

在均值-CVaR问题中，最终的极小极大问题为

$$\inf_{\mathbf{1}'\mathbf{x}=1, \tau\in\mathbb{R}} \left\{\rho\tau + \sup_{q\in[q_0\pm\epsilon]} \left((1-q)\mathbb{E}_{P_N}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + q\left[\mathbb{E}_{P_{S_0}}(\ell_{\mathbf{x},\tau}(\mathbf{R})) + r(q)(1+\frac{\rho}{1-p})\|\mathbf{x}\|_\infty\right]\right)\right\}. \tag{34}$$

在 (33) 中，所有结果仍然成立，因为我们可以将 $\mathbf{x}$ 和 a 限制在紧集内，而不影响下确界。然而，在 (34) 中，最小值可能为 $-\infty$，且在杠杆头寸任意大的情况下取到。

6 附录

6.1 第2节和第3节中使用的一些定理

定理 8 (Gao 和 Kleywegt 中的定理1 [2023])。 对于任意波兰空间(Ξ, d) 及可测函数 Ψ ，其中 $p \geq 1$,

$$\sup_{\mu\in\mathcal{P}(\Xi)} \{\mathbb{E}_\mu[\Psi(x,\xi)] : W_p(\mu,\nu) \leq \theta\} = \min_{\lambda\geq 0}\left\{\lambda\theta^p - \int_\Xi \inf_{\xi\in\Xi}\left[\lambda d^p(\xi,\zeta) - \Psi(x,\xi)\right]\nu(d\zeta)\right\}.$$

□

定理 9 (Bertsekas 中的命题 A.3.2 [2009]: Danskin 定理)。 设 $Z \subset \mathbb{R}^m$ 为紧集，且 ϕ : $\mathbb{R}^n \times Z \mapsto \mathbb{R}$ 连续，并满足对每个 $z \in Z$， $\phi(\cdot, z) : \mathbb{R}^n \mapsto \mathbb{R}$ 为凸函数。

(a) 函数 f： $\mathbb{R}^n \mapsto \mathbb{R}$ 定义为

$$f(x) = \max_{z\in Z} \phi(x,z) \tag{35}$$

是凸函数，其方向导数为

$$f'(x;y) = \max_{z\in Z(x)} \phi'(x,z;y),$$

其中， $\phi'(x, z;\ y)$ 是函数 $\phi(\cdot, z)$ 在 x 沿y方向的方向导数，而 $Z(x)$ 是（35）中的最大化点集：

$$Z(x) = \left\{ z \,\middle|\, \phi(x,z) = \max_{z \in Z} \phi(x,z) \right\}.$$

特别地，如果 $Z(x)$ 由唯一一点 $\bar{z}$ 组成，并且 $\phi(\cdot, \bar{z})$ 在 x处可微，那么f 在 x处可微，且

$$\nabla f(x) = \nabla_x \phi(x, \bar{z}),$$

其中 $\nabla_x \phi(x, \bar{z})$ 是坐标为以下各项的向量

$$\frac{\partial \phi(x, \bar{z})}{\partial x_i}, \quad i = 1, \ldots, n.$$

(b) 如果对所有 $z \in Z$， $\phi(\cdot, z)$ 均可微，并且 $\nabla_x \phi(x, \cdot)$ 对每个 x都在 Z 上连续，那么

$$\partial f(x) = \operatorname{conv}\left\{ \nabla_x \phi(x,z) \,\middle|\, z \in Z(x) \right\}, \qquad \forall x \in \mathbb{R}^n.$$

引理 8 (Berge 最大值定理，Berge [1997]). 设 X 和 Θ 是拓扑空间，$f:\ X \times \Theta \to \mathbb{R}$ 是在笛卡尔积 $X \times \Theta$上的连续函数，且 $C:\ \Theta \rightrightarrows X$ 是紧值对应，并满足对所有 $\theta \in \Theta$， $C(\theta) \neq \emptyset$ 。

定义边际函数（或值函数） $f^*: \Theta \to \mathbb{R}$ 为

$$f^*(\theta) = \sup\{ f(x,\theta) : x \in C(\theta) \}$$

以及最大化解集 $C^*: \Theta \rightrightarrows X$ 为

$$C^*(\theta) = \arg\max_{x \in C(\theta)} f(x,\theta) = \{ x \in C(\theta) : f(x,\theta) = f^*(\theta) \}.$$

如果 C 在 θ 处连续（即同时上半连续和下半连续），那么值函数 f^* 连续，且最大化解集 C^* 上半连续，其取值非空且紧。因此，上确界 可以替换为 最大值。

□

6.2 投影次梯度下降的收敛速率

投影次梯度下降算法的基本复杂度结果如下：

定理10 (Bubeck 等人的定理 3.2 [2015])。 设J是定义在 $\mathcal{X}$（一个紧致凸欧几里得集合）上的凸函数。设 x_k 为迭代k时的决策变量向量， x^* 为一个最优解。假设J关于 $\mathcal{X}$ 和 $\|x_0 - x^*\| \leq R$ 是 G-利普希茨的。采用恒定步长 $\eta_k = \frac{R}{G\sqrt{T}}$ 的投影梯度下降（即 $x_{k+1} = \mathit{Proj}_{\mathcal{X}}(x_k - \eta_k \mathbf{g}_k)$）满足

$$J\left(\frac{1}{T} \sum_{i=1}^{T} x_i \right) - J(x^*) \leq \frac{RG}{\sqrt{T}}.$$

□

然而，上述定理的收敛速率 $O(1/\sqrt{T})$ 相对较慢。为加强结论，我们需要对 $\bar{J}$ 作出更强的假设，并为此引入以下定义：

定义 2（强凸性）。对于 $\alpha > 0$，函数 $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ 是 $\alpha-$强凸的，当且仅当 $f - \frac{\alpha}{2}\|\mathbf{x}\|^2$ 是凸的。

定义 3（β-光滑）。如果连续可微凸函数 f 的梯度 ∇f 是 β-利普希茨的，我们就称其为β-光滑，即

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta\|x - y\|.$$

注意，如果 f是二阶可微的，那么强凸性等价于海森矩阵的特征值至少为 α。另一方面，β-光滑性等价于海森矩阵的特征值至多为 β。比值 β/α 称为 f的条件数。

定理 11（Bubeck 等人的定理 3.9 [2015]）。设 J 在 $\mathcal{X}$上是 α-强凸且 L-利普希茨的，其中 $\mathcal{X}$是一个紧致凸欧几里得集合。则采用时变步长的投影次梯度下降

$$\eta_s = \frac{2}{\alpha(s+1)}$$

满足

$$J\left(\sum_{s=1}^{T} \frac{2s}{T(T+1)} x_s\right) - J(x^*) \;\leq\; \frac{2L^2}{\alpha(T+1)}.$$

□

定理 12 (Bubeck 等人定理 3.10 [2015])。设 J 为 $\mathcal{X}$上的 α-强凸且 β-光滑函数，其中 $\mathcal{X}$是一个紧致凸欧几里得集合，κ 表示其条件数。那么，步长为$(1/\beta)$ 的投影梯度下降满足：对于 $T \geq 0$,

$$\|\theta_T - \theta^*\|^2 \leq \exp\left(-\frac{T}{\kappa}\right)\|\theta_0 - \theta^*\|^2.$$

注意，

$$J(\theta_T) - J(\theta^*) \leq \frac{\beta}{2}\|\theta_T - \theta^*\|^2 \leq \frac{\beta}{2}\exp\left(-\frac{T}{\kappa}\right)\|\theta_0 - \theta^*\|^2.$$

□

6.3 极小极大定理

定理 13 (TERKELSEN 中的定理 3 [1972])。设 X 为一个紧致连通拓扑空间，Y 为一个集合，且 f： $X \times Y \to \mathbb{R}$ 为一个满足以下条件的函数：

(i) 对任意 $y_1, y_2 \in Y$，存在 $y_0 \in Y$ 使得

$$f(x, y_0) \geq \left(\frac{1}{2}\right)(f(x, y_1) + f(x, y_2))$$

对所有 $x \in X$。

(ii) 所有形如 $\{x \in X\colon\ f\left(x, y\right) \leq \alpha\}$ 的集合的有限交，若满足 $\left(y, \alpha\right) \in Y \times \mathbb{R}$，则为闭且连通。

Then

$$\sup_{y \in Y} \min_{x \in X} f(x, y) = \min_{x \in X} \sup_{y \in Y} f(x, y).$$

定理 14 (Sion 的定理 3.4 的推论 3.3 [1958])。 设 X 和 Y 是拓扑向量空间中的凸子集，其中 X 是紧集，且令 $f\colon\ X \times Y \to \mathbb{R}$ 满足：

(i) 对每个 $x \in X$，函数 $y \mapsto f(x, y)$ 在 Y上是上半连续且拟凹的；

(ii) 对每个 $y \in Y$，函数 $x \mapsto f(x, y)$ 在 X上是下半连续且拟凸的。

Then

$$\sup_{y} \min_{x} f = \min_{x} \sup_{y} f.$$

参考文献

赤池弘次。信息论与最大似然原理的扩展。载于 Boris N. Petrov 和 Frigyes Cs´aki 编，《第二届国际信息论研讨会》，第 267–281 页，布达佩斯，1973 年。Akad´emiai Kiad´o。

美国精算师学会。可变年金及类似产品基于风险的资本要求的建议方法。技术报告，美国精算师学会，2005 年 6 月。网址： `https://actuary.org/c3-phase-ii-rbc-and-reserves-project/`。

Andrew Ang。 资产管理：因子投资的系统方法。牛津大学出版社，2014年。

Tom M Apostol。《数学分析》。Addison-Wesley数学系列。Addison-Wesley，马萨诸塞州雷丁，1974年。网址 `https://cds.cern.ch/record/105425`。

Aharon Ben-Tal、Dick den Hertog、Anja De Waegenaere、Bertrand Melenberg 和 Gijs Rennen。受不确定概率影响的优化问题的稳健解。《管理科学》，59(2)：341–357，2013年。doi：10.1287/mnsc.1120.1641。

Claude Berge。《拓扑空间》。Dover出版社公司，纽约州米尼奥拉，1997年。ISBN 0-486-69653-7。书中还论述了多值函数、向量空间和凸性；由 E. M. Patterson 从法文原著翻译，重印自1963年译本。

James O. Berger。稳健贝叶斯分析综述。*Test*，3(1):5–124，1994。doi: 10.1007/BF02562676。

James O. Berger与L. Mark Berliner。带ϵ污染先验的稳健贝叶斯与经验贝叶斯分析。《统计年鉴》，14(2):461–486，1986。doi: 10.1214/aos/ 1176349933。Robert H. Berk。模型设定错误时后验分布的极限行为。《数学统计年鉴》，37(1):51–58，1966。doi: 10.1214/aoms/1177699597。Carole Bernard、Rodrigue Kazzi与Steven Vanduffel。部分信息下单峰分布的区间风险价值界。《保险：数学与经济学》，94:9–24，2020年。Dimitri Bertsekas。凸优化理论，第1卷。Athena Scientific，2009。Jose Blanchet、Karthyek Murthy与Viet Anh Nguyen。Wasserstein分布鲁棒估计量的统计分析。载于《运筹学教程：新兴优化方法及应用建模技术》，第227–254页。INFORMS，2021。Jose Blanchet、Jiajin Li、Sirui Lin与Xuhui Zhang。分布鲁棒优化与稳健统计。《统计科学》，40(3):351–377，2025年。S´ebastien Bubeck等人。凸优化：算法与复杂性。基础与趋势® 机器学习，8(3-4):231–357，2015。D. Cornilly、L. R¨uschendorf与S. Vanduffel。矩空间上严格凹失真风险度量的上界。《保险：数学与经济学》，82:141–151，2018。Imre Csisz´ar。观测信道的一类信息量度。《匈牙利数学期刊》，2(1-4):191–213，1972。Erick Delage与Yinyu Ye。矩不确定性下的分布鲁棒优化及其在数据驱动问题中的应用。《运筹学》，58(3):595–612，2010。Paul Embrechts、Andrea H¨oing与Giovanni Puccetti。最坏情形VaR。《保险：数学与经济学》，37(1):115–134，2005。Rui Gao与Anton Kleywegt。基于Wasserstein距离的分布鲁棒随机优化。《运筹学数学》，48(2):603–655，2023。Danqiao Guo、Phelim P Boyle、Chengguo Weng与Tony S Wirjanto。1/n规则何时有效？可于*SSRN 3111531*获取，2019。Peter J. Huber。位置参数的稳健估计。《数学统计年鉴》，35(1):73–101，1964。doi: 10.1214/aoms/1177703732。Hisham Husain。基于积分概率度量的分布鲁棒性及其与正则化和生成对抗网络的联系。《神经信息处理系统进展》，33:11816–11827，2020年。

Daniel Kuhn、Peyman Mohajerin Esfahani、Viet Anh Nguyen 和 Soroosh Shafieezadeh- Abadeh。《数据分析时代的运筹学与管理科学》，第130–166页。INFORMS，2019年。Solomon Kullback 和 Richard A. Leibler。论信息与充分性。《数学统计年鉴》，22(1):79–86，1951年。doi: 10.1214/aoms/1177729694。Lujun Li、Hui Shao、Ruodu Wang 和 Jingping Yang。部分信息下的最坏情形范围风险价值。SIAM金融数学杂志，9(1):190–218，2018年。doi: 10.1137/17M1126138。URL https://doi.org/10.1137/17M1126138。Peyman Mohajerin Esfahani 和 Daniel Kuhn。利用 Wasserstein度量进行数据驱动的分布鲁棒优化：性能保证与易处理的重构。《数学规划》，171(1):115–166，2018年。Alfred M¨uller。积分概率度量及其生成函数类。《应用概率进展》，29(2):429–443，1997年。doi: 10.2307/1428011。美国保险监督官协会。《估值手册：2024年1月1日版》。美国保险监督官协会，2024年。URL https://content.naic.org/sites/default/files/pbr-data-valuation-manual-2024-edition.pdf。2024年1月1日起生效。R Rockafellar。《凸分析》。普林斯顿数学丛书，第28卷，1970年。

R Tyrrell Rockafellar、Stanislav Uryasev 等。条件风险价值的优化。《风险期刊》，2:21–42，2000。

Maurice Sion。关于一般极小极大定理。1958。Bharath K. Sriperumbudur、Kenji Fukumizu、Arthur Gretton、Bernhard Sch¨olkopf 与 Gert R. G. Lanckriet。关于积分概率度量的经验估计。电子统计期刊，6:1550–1599，2012。doi: 10.1214/12-EJS722。Matthew Staib 与 Stefanie Jegelka。核方法中的分布鲁棒优化与泛化。《神经信息处理系统进展》，32，2019。URL https://arxiv.org/abs/1905.10943。Qihe Tang 与 Yunshen Yang。部分不确定性下的最坏情形矩。*ASTIN*公报：国际精算协会期刊，53(2):443–465，2023。FRODE TERKELSEN。若干极小极大定理。斯堪的纳维亚数学，31(2):405–413，1972。ISSN 00255521, 19031807。URL http://www.jstor.org/stable/24490558。C´edric Villani。最优传输专题，第58卷。美国数学学会，2003。