

带大量凸约束的随机优化问题的随机增广拉格朗日算法*

赵文深, 韩丛英[†], 金玲子

(中国科学院大学数学科学学院, 北京 100049)

(2023 年 1 月 17 日收稿; 2023 年 5 月 15 日收修改稿)

Zhao W S, Han C Y, Jin L Z. Stochastic augmented Lagrangian method for stochastic nonconvex nonsmooth programs with many convex constraints[J]. Journal of University of Chinese Academy of Sciences, 2025, 42(1):26-42. DOI: 10.7523/j.ucas.2023.055.

摘要 随机梯度法广泛应用于机器学习并取得显著成功,但许多随机方法主要针对无约束或简单约束的优化问题。对于带有正则项和大量凸约束的非凸随机优化问题,经典增广拉格朗日法是一种解法,但精确梯度信息的要求使其难以有效应对大量约束问题。为此,提出一种随机增广拉格朗日算法,该算法用随机一阶信息代替增广拉格朗日法的精确梯度,每步迭代仅使用一组抽样梯度和部分的约束梯度。对该算法,证明其可以在 $\mathcal{O}(\epsilon^{-8})$ 次方后找到 ϵ -KKT 近似点。在多元分类 Neyman-Pearson 问题上进行数值实验,实验结果验证了算法的有效性。

关键词 随机梯度法;增广拉格朗日法;非线性优化;约束优化

中图分类号:O224 文献标志码:A DOI:10.7523/j.ucas.2023.055

Stochastic augmented Lagrangian method for stochastic nonconvex nonsmooth programs with many convex constraints

ZHAO Wenshen, HAN Congying, JIN Lingzi

(School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract The stochastic gradient methods have been widely used in machine learning, but most existing works aim for unconstrained or simple constrained problems. In this paper, we consider the nonconvex stochastic programs with many functional convex constraints. The deterministic augmented Lagrangian method is a classic algorithm for such problems, but the requirement of accurate gradient information makes the method impractical for the case with large-scale constraints. To solve such problems, we propose a novel stochastic augmented Lagrangian method, which is called CSALM (composite stochastic augmented Lagrangian method). The CSALM uses a stochastic gradient to approximate the accurate gradient, and it only samples stochastic gradients and batches constraint gradient per iteration. We establish the convergence theory of CSALM and show that the

* 国家电网公司总部科技项目(5700-202055486A-0-0-00)资助

[†] 通信作者, E-mail: hancy@ucas.ac.cn

CSALM can find an ϵ -KKT point after $\mathcal{O}(\epsilon^{-8})$ iterations. The numerical experiments on the multi-class Neyman-Pearson classification problem (mNPC) demonstrate the efficiency of CSALM.

Keywords stochastic gradient; augmented Lagrangian method; nonlinear optimization; constrained optimization

本文考虑如下带有大量复杂约束的随机优化问题。

$$\begin{aligned} \min_{x \in X} \quad & \psi_0(x) = f_0(x) \equiv \mathbb{E}_{\xi}[F(x; \xi)] + \chi_0(x), \\ \text{s. t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, M. \end{aligned} \quad (1)$$

其中: $f_0(x) := \mathbb{E}_{\xi}[F(x; \xi)]$, $\xi \in \Xi$ (概率空间); $X \subseteq \mathbb{R}^d$ 是凸紧集; $\chi_0: X \rightarrow \mathbb{R}$ 是凸函数; $f_0: X \rightarrow \mathbb{R}$ 是非凸函数; $f_i: X \rightarrow \mathbb{R}$, $i = 1, \dots, M$ 是凸函数; $F(\cdot, \xi): X \rightarrow \mathbb{R}$, $\forall \xi \in \Xi$ 是可微函数; f_i , $i = 0, \dots, M$ 在 X 上连续可微, M 充分大。

期望目标函数在当前的机器学习和数据科学中广泛出现,但常用的都是无约束有限和问题,然而采用约束优化模型的学习问题也在发展,比如基于约束优化的网络结构搜索 (NAS)^[1]。该大规模约束优化问题的一大应用是复杂优化问题的近似问题,例如概率约束问题^[2]、robust 优化^[3]、极小极大博弈^[4]。以概率约束问题为例,其问题模型是 $\min_{x \in X} f_0(x)$ s. t. $\text{Prob}(g(x, \xi_i) \leq 0) \geq 1 - \tau$, 其中 X 是凸紧集, ξ 是 Ξ 上的参数。该问题的困难在于, $E[g]$ 可能是非凸函数,即使 g 对任意 ξ 是凸函数。一类解法是通过 ξ 抽样,用抽样约束组代替概率约束,这称为抽样丢弃法^[2] (sampling and discarding approach)。具体地说,这类方法从 Ξ 中抽取 M 个独立样本,然后用抽样约束近似概率约束,得到近似的约束优化问题

$$\min_{x \in X} f_0(x) \quad \text{s. t.} \quad g(x, \xi_i) \leq 0 \quad \forall i = 1, \dots, M.$$

在一定条件下,可证明 $M \rightarrow \infty$ 时,近似问题接近于原问题,所以当精度要求较高时 M 很大^[2]。

针对随机约束优化问题目前已发展了许多算法,这些算法整体上与确定约束优化问题类似,但因目标或约束带有随机信息,所以许多算法都带随机性。因为约束性质和数量的不同,这些算法的类别差异主要体现在约束的处理上。

最简单的处理是投影类方法,这类方法处理约束的思路是每步信息更新后的迭代点投影到可行域内,保持迭代序列可行性。在约束为随机约束或约束个数极多时,每步检查所有约束是否可行以及实现投影计算是非常困难的。所以一些算

法^[5-6]引入了随机投影的思想,当约束个数极多时每步迭代只抽样一个约束,然后将迭代点投影到该约束的可行域内。在这一思路基础上,另一种处理方式^[7]是用梯度迭代代替直接投影,它同样每步迭代抽取一个约束,但转而考虑约束的一阶信息,如果约束一阶信息的范数大于给定阈值,则用该信息走一步负梯度方向。

另一类方法是子问题方法,思想是将困难问题转化为一系列的近似简单问题。经典的是二次规划方法^[8],每步求解当前点的二阶近似,但因为用到二阶信息计算量相对较高。另一种方法是水平集方法^[9-11],这类方法利用原问题的约束函数和目标函数构造水平集函数,然后将求解原问题转化为求解基于水平集的序列子问题,优势是无需二阶信息,但水平函数的构造相对更复杂,而且只能应用在凸优化问题上,如果目标非凸可以使用 ICPP 方法^[12],该方法的子问题为凸约束优化问题,转化后就可以采用已有的凸优化技巧。

第 3 类方法是罚函数类算法。这一类方法将约束以罚函数的形式加到目标函数上,从而将约束优化转为无约束优化,罚函数的构造区别了这一类算法。最简单的罚函数是 $L1$ 和 $L2$ 罚函数,前者是精确罚方法,后者则有光滑性质;另一经典的罚类方法就是增广拉格朗日算法^[13],此类算法结合了对偶理论,使得罚函数同时具备精确性和光滑性,算法交替更新迭代点和乘子,目标是寻找 KKT 点。近期针对随机目标优化的增广拉格朗日算法有 PDSG^[14],对凸目标函数, PDSG 证明了可以在 $\mathcal{O}(\epsilon^{-2})$ 次迭代后找到 ϵ 近似解。

1 引言

本文提出的算法属于增广拉格朗日算法一类。针对带大规模约束的约束优化问题,基于随机思想和已有的增广拉格朗日算法,提出一个新的随机增广拉格朗日算法,并证明从任意一个初始点出发,该算法能在 $\mathcal{O}(\epsilon^{-8})$ 次迭代后找到 ϵ -KKT 近似点。

1.1 文章贡献

本文的主要贡献有以下 3 方面:

1) 提出一个原始对偶的随机算法用以解决带大量非线性约束的随机优化问题。该方法是经典增广拉格朗日法的随机改进,其每步采取一阶近似信息更新迭代,并按照约束抽样更新对应乘子。每步的计算复杂度取决于约束抽样数与近似精确度。

2) 建立算法对非凸非光滑目标函数的全局收敛性结果,证明该算法可以在 $\mathcal{O}(\epsilon^{-8})$ 次迭代后找到 ϵ -KKT 近似点。与其他的针对非凸目标的算法不同,不假设约束的非奇异性而仅需要约束是凸的且 Lipschitz 光滑。

3) 用多分类 Neyman Pearson 问题验证算法的实际表现。数值结果表明该算法能有效收敛,同时能保证输出迭代点的不可行性的数量级较低。

1.2 标记符号定义

本文所用符号多数均为常用标记符号。使用小写字母 \mathbf{x} 表示向量,黑体的 $\mathbf{0}$ 和 \mathbf{I} 表示全 0 和全 1 向量, $[M] = \{1, 2, \dots, M\}$ 表示整数集合, \mathbb{R}_{++} 表示正实数集, $\|\mathbf{x}\|$ 表示向量 \mathbf{x} 的 ℓ_2 范数, $\|z\|_1$ 是 z 的 ℓ_1 范数, $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^T \mathbf{A} \mathbf{x}}$ 表示基于对称正定矩阵 \mathbf{A} 的椭圆范数, $[a]_+ = \max(0, a)$, $|\mathcal{I}|$ 表示集合 \mathcal{I} 的元素个数, ζ_k 表示第 k 次迭代的随机信息, $\zeta_{[k]} = \{\zeta_1, \dots, \zeta_k\}$ 表示前 k 次迭代的随机信息, 条件期望 $\mathbb{E}[\cdot | \zeta_{[k-1]}] = \mathbb{E}_{\zeta_k}[\cdot | \zeta_{[k-1]}]$ 是已知前 $k-1$ 次信息对第 k 次信息的期望, $\partial \chi(x)$ 表示 χ 在 x 处的次微分, $D_X := \sup_{x, y \in X} \|x - y\|$ 表示集合的直径, $\mathbf{d}(\cdot, \cdot)$ 表示距离。

2 CSALM

本节提出一个用于解决问题 (1) 的随机增广拉格朗日方法 (composite stochastic augmented Lagrangian method, CSALM), 算法基于问题 (1) 的非线性增广拉格朗日函数^[15-16], 表示为

$$\begin{aligned} \mathcal{L}_\beta(x, z) &= f_0(x) + \chi_0(x) + \Psi_\beta(x, z), \\ \Psi_\beta(x, z) &= \frac{1}{M} \sum_{i=1}^M \psi_\beta(f_i(x), z_i), \\ \psi_\beta(u, v) &= \begin{cases} uv + \frac{\beta}{2} u^2, & \beta u + v \geq 0, \\ -\frac{v^2}{2\beta}, & \beta u + v < 0. \end{cases} \end{aligned} \quad (2)$$

其中: $\beta > 0$ 是罚参数, z 是拉格朗日乘子或对偶变量, Ψ 是罚函数。式 (2) 是等式约束增广拉格

朗日函数的推广^[15], 其原问题是问题 (1) 的约束函数乘上系数 $1/M$ 时的等价问题。

对一般的非凸问题, 寻找全局极小点是困难的, 如果存在约束则更甚, 所以我们的目标是找到 KKT 点。KKT 点作为目标是合理的, 当优化问题满足一些约束规范性条件时如 LICQ, 优化问题就存在 KKT 点。

定义 2.1 (KKT 点) 称 $(x^*, z^*) \in X \times \mathbb{R}_+^M$ 为问题 (1) 的 KKT 点, 若其满足

$$\begin{aligned} \mathbf{0} &\in \nabla f_0(x^*) + \partial \chi_0(x^*) + \mathcal{N}_X(x^*) + \\ &\quad \frac{1}{M} \sum_{i=1}^M z_i^* \nabla f_i(x^*), \\ f_i(x^*) &\leq 0, i = 1, \dots, M, \\ z_i^* f_i(x^*) &= 0, i = 1, \dots, M. \end{aligned}$$

但因为数值计算的精度问题, 精确 KKT 点一般无法求得或者需要极高的计算量, 所以通常寻找 ϵ -KKT 近似点。

本文设计了寻找问题 (1) 的 ϵ -KKT 点算法 CSALM, 算法假设存在一个 \mathcal{SFO} (stochastic first order oracle), 即 $\forall x \in X$ 和 $i \in [M]$ 可以获得 $\nabla f_0(x)$ 基于样本 ξ_j 的随机近似 $g_{0,j} = \nabla F(x, \xi_j)$ 以及第 i 个约束的梯度和函数值 $(\nabla f_i(x), f_i(x))$ 。在第 k 次迭代, 首先从样本空间和约束集合抽取批量样本 $\mathcal{T}_k \subset \Psi$ 和 $\mathcal{I}_k \subset [M]$, 然后计算基于 \mathcal{SFO} 的 $\nabla f_0(x^k)$ 的随机近似 $g_{0,j}^k = \nabla F(x^k, \xi_j)$, $j \in \mathcal{T}_k$ 以及指标集 \mathcal{I}_k 对应约束的梯度和函数值 $(\nabla f_i(x), f_i(x))$, $i \in \mathcal{I}_k$, 获得梯度随机近似,

$$g_0^k = \frac{1}{J_k} \sum_{j \in \mathcal{T}_k} g_{0,j}^k, J_k = |\mathcal{T}_k|, \quad (3)$$

$$h^k = \frac{1}{I_k} \sum_{i \in \mathcal{I}_k} [\beta f_i(x^k) + z_i^k]_+ \nabla f_i(x^k), I_k = |\mathcal{I}_k|. \quad (4)$$

随后基于 $g_0^k + h^k$ 对 x^k 进行迭代, 最后更新 \mathcal{I}_k 对应的对偶变量。

$$\begin{aligned} x^{k+1} &= \arg \min_{x \in X} \left\{ \chi_0(x) + \langle g_0^k + h^k, x \rangle + \right. \\ &\quad \left. \frac{1}{2} \|x - x^k\|_{D_k}^2 \right\}. \end{aligned}$$

$$z_i^{k+1} = z_i^k + \rho_k \cdot \max \left(-\frac{z_i^k}{\beta}, f_i(x^k) \right), \quad i \in \mathcal{I}_k.$$

整个算法流程详见下列 CSALM 算法, 超参数有罚参 β , 步长 α , 乘子更新系数 ρ 。算法的输出是在 T 次迭代后, 随机取整数 $R \in [T]$, 输出 x^{R+1} 。

算法 CSALM (composite stochastic augmented Lagrangian method)

输入: 初始点 $x^1 \in X, z^1 = 0$, 罚参数 $\beta > 0$, 步长 $|\alpha_k| > 0, |\rho_k| \in (0, \beta)$

输出: x^{R+1} , 其中 R 是从 $[T]$ 中按均匀分布抽取的随机整数;

1. **for** $k = 1$ to T **do**

2. 按概率空间 Ξ 分布独立抽取样本集 $\mathcal{T}_k \subseteq \Xi$, 均匀分布独抽取样本集 $\mathcal{I}_k \subseteq [M]$

3. 调用一阶信息系统, 依据 (3) ~ (4) 计算 g_0^k 和 h^k ;

4. 更新原始变量 x

$$x^{k+1} = \arg \min_{x \in X} \left\{ \chi_0(x) + \langle g_0^k + h^k, x \rangle + \frac{1}{2} \|x - x^k\|_{b_k}^2 \right\}. \quad (5)$$

其中: $D_k = \alpha_k^{-1} I$;

5. 更新对偶变量 z

$$z_i^{k+1} = \begin{cases} z_i^k, & i \notin \mathcal{I}_k, \\ z_i^k + \rho_k \cdot \max \left(-\frac{z_i^k}{\beta}, f_i(x^k) \right), & i \in \mathcal{I}_k. \end{cases} \quad (6)$$

6. **End for**

当 D_k 是数量矩阵时, 子问题迭代实际是梯度下降的拓展, 但更容易处理凸项 $\chi_0(x)$ 。从式 (6) 可知, CSALM 和 PDSG^[14] 是使用 x^k 更新对偶变量 z^k , 而经典增广拉格朗日算法是使用 x^{k+1} , 这一改动体现在收敛性证明上。另可观察到, $g_0^k + h^k$ 实际是 $\mathcal{L}_\beta(x^k, z^k)$ 中可微部分的随机梯度, 所以该方法其实可类比随机梯度法, 从这一角度出发, 该方法可以通过应用已有的无约束随机优化的技巧提升理论结果与数值表现, 例如动量加速^[17] 与方差缩减^[18]。

本文需要如下 3 个假设, 分别是 KKT 存在假设、 f_i 的 Lipschitz 光滑性假设和随机梯度方差有界假设, 后两者是随机优化算法中的常见假设。

假设 2.1 问题 (1) 存在 KKT 点 $(x^*, z^*) \in X \times \mathbb{R}_+^M$, 其满足定义 2.1。

假设 2.2 $\{f_i(x), i = 0, 1, \dots, M\}$ 在 X 上是 Lipschitz 光滑的,

$$\| \nabla f_i(x) - \nabla f_i(y) \| \leq L \| x - y \|, \forall x, y \in X, \\ i = 0, 1, \dots, M, \Rightarrow |f_i(y) - f_i(x) -$$

$$\langle \nabla f_i(x), y - x \rangle| \leq \frac{L}{2} \| y - x \|^2, i = 0, \dots, M.$$

(7)

其中: $L \in \mathbb{R}_+$ 是光滑系数。

假设 2.3 记 $\sigma \geq 0$ 是方差界常数, SFO 返回的 $g_{0,j}^k$ 是无偏且方差有界的

$$\mathbb{E}[g_{0,j}^k | \zeta_{[k-1]}] = \nabla f_0(x^k), \forall j \in \mathcal{T}_k, \forall k,$$

$$\mathbb{E}[\|g_{0,j}^k - \nabla f_0(x^k)\|^2 | \zeta_{[k-1]}] \leq \sigma^2, \forall j \in \mathcal{T}_k, \forall k,$$

$$\Rightarrow \mathbb{E}[g_0^k | \zeta_{[k-1]}] = \nabla f_0(x^k), \forall k,$$

$$\mathbb{E}[\|g_0^k - \nabla f_0(x^k)\|^2 | \zeta_{[k-1]}] \leq \frac{\sigma^2}{J_k}, \forall k.$$

基于假设和本文研究问题 (1) 的设定, 可以得到关联 $f_i, \nabla f_i, \chi_0, \Psi_\beta$ 的有界性质。

引理 2.1 X 是凸紧集, 所以 $\exists F, G \in \mathbb{R}_+$ 使得

$$|f_i(x)| \leq F, \|\nabla f_i(x)\| \leq G, \\ \forall x \in X, i = 0, 1, \dots, M. \quad (8)$$

引理 2.2 X 是凸紧集且 $\chi_0(x)$ 是正常凸函数, 所以 $\exists W \in \mathbb{R}$ 使得

$$\chi_0(x) \geq W, \quad \forall x \in X. \quad (9)$$

引理 2.3 (引理 3.4^[14]) 若 $i \in [M]$ 是均匀抽样, 那么 $\forall x \in X, \forall z \in \mathbb{R}^M$ 有

$$\mathbb{E}_i[\|\beta f_i(x) + z_i\|_+ \nabla f_i(x)] = \nabla_x \Psi_\beta(x, z),$$

$$\mathbb{E}_i[\|\|\beta f_i(x) + z_i\|_+ \nabla f_i(x)\|^2] \leq 2\beta^2 F^2 G^2 + \\ \frac{2G^2}{M} \|z\|^2,$$

$$\mathbb{E}_i[\|\|\beta f_i(x) + z_i\|_+ \nabla f_i(x) - \nabla_x \Psi_\beta(x, z)\|^2] \leq \\ 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \|z\|^2.$$

则有

$$\mathbb{E}[h^k | \zeta_{[k-1]}] = \nabla_x \Psi_\beta(x^k, z^k), \quad (10)$$

$$\mathbb{E}[\|h^k - \nabla_x \Psi_\beta(x^k, z^k)\|^2 | \zeta_{[k-1]}] \leq$$

$$\frac{2\beta^2 F^2 G^2 + \frac{2G^2}{M} \|z^k\|^2}{I_k}. \quad (11)$$

其中: 由式 (8) F, G 为常数, \mathbb{E}_i 是均匀抽样的期望。

3 CSALM 收敛性分析

本节的内容是建立算法的收敛性结果。算法的目标是找到满足如下条件的 ϵ -KKT 点。

定义 3.1 (ϵ -KKT 点) $\epsilon > 0$ 为给定求解精度, $(x, z) \in \mathbb{R}^d \times \mathbb{R}_+^M$ 为 ϵ -KKT 点, 若其满足

$$\mathbb{E}_{R, \zeta_{[T]}} \left[\mathbf{d} \left(\nabla f_0(x) + \partial \chi_0(x) + \sum_{i=1}^M z_i \nabla f_i(x) + \right. \right. \\ \left. \left. \mathcal{N}_X(x), \mathbf{0} \right) \right] \leq \epsilon, \quad (12)$$

$$\mathbb{E}_{R, \zeta_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M [f_i(x)]_+ \right] \leq \epsilon, \quad (13)$$

$$\mathbb{E}_{R, \zeta_{[T]}} \left[\sum_{i=1}^M z_i [f_i(x)]_- \right] \leq \epsilon. \quad (14)$$

本节将证明在假设 2.1 ~ 假设 2.3 下,

CSALM 算法可以在 $\mathcal{O}(\epsilon^{-8})$ 次迭代后找到 ϵ -KKT 近似点。首先证明乘子 z^k 和 $\Psi_\beta(x, z)$ 的部分性质, 然后分别证明梯度 (12)、约束违反度 (13)、互补松弛违反度 (14) 的收敛性。

3.1 前置结果

本小节主要证明 CSALM 算法下拉格朗日乘子 z^k 的非负性、有界性、 $\Psi_\beta(x, z)$ 关于 x 的 Lipschitz 光滑性和 h^k 的方差有界性。非负性是 KKT 点的性质; 有界性是增广拉格朗日算法相对于原始罚函数方法 $z \rightarrow \infty$ 才收敛的改进优势; $\Psi_\beta(x, z)$ 对 x 的 Lipschitz 光滑性要求类比于 $f_0(x)$; h^k 的方差有界性类比于 g_0^k 的方差有界性。后 2 个结果是为了满足无约束优化问题下建立随机梯度算法收敛性的假设要求。

首先给出 z^k 的非负性。

引理 3.1 对任意 $k \geq 1, z^k \in \mathbb{R}_+^M$ 。

证明 $k=1$ 时, $z^1 = \mathbf{0}$ 。 $k > 1$ 时, 设 $z^k \in \mathbb{R}_+^M$, 由式 (6) 和 $\rho_k \in (0, \beta)$ 有

$$z_i^{k+1} \geq z_i^k - \frac{\rho_k}{\beta} z_i^k \geq z_i^k - z_i^k = 0, \quad \forall i \in [M].$$

由归纳法知结论成立。

然后证明 z^k 的有界性, 首先考虑 z^k 的单步变化量, 其被常数 F 和更新步长 ρ_k 控制。

引理 3.2 在假设 2.2 下, $\forall k \in [T], \forall i \in [M]$, 有

$$|z_i^{k+1} - z_i^k| \leq F\rho_k.$$

证明 式 (8) 可知 $|f(x^k)| \leq F$ 。由 z^k 的更新式 (6) 得

$$|z_i^{k+1} - z_i^k| \leq \rho_k \left| \max \left(-\frac{z_i^k}{\beta} f_i(x^k) \right) \right|, \quad \forall k \in [T], \forall i \in [M]. \quad (15)$$

$$\text{当 } f_i(x^k) \geq -\frac{z_i^k}{\beta},$$

$$\left| \max \left(-\frac{z_i^k}{\beta} f_i(x^k) \right) \right| = |f_i(x^k)| \leq F. \quad (16)$$

$$\text{当 } f_i(x^k) < -\frac{z_i^k}{\beta}, \text{ 因为 } z^k \text{ 非负 (引理 3.1), 所}$$

以

$$\left| \max \left(-\frac{z_i^k}{\beta} f_i(x^k) \right) \right| = \left| -\frac{z_i^k}{\beta} \right| \leq |f_i(x^k)| \leq F. \quad (17)$$

将式 (16)、式 (17) 代入式 (15) 结论得证。 \square

基于 z^k 的单步变化量, 以下给出 z^k 的有界性结果。

引理 3.3 记 $\sum_{t=1}^0 \rho_t = 0$, 假设 2.2 满足时,

$$\forall k \in [T+1], \|z^k\|_1 \leq MF \sum_{t=1}^{k-1} \rho_t \text{ 成立。}$$

证明 $k=1$ 时,

$$\|z^1\|_1 = 0 = \sum_{t=1}^0 \rho_t = MF \sum_{t=1}^0 \rho_t.$$

$k \geq 2$ 时, 由引理 3.2, $\forall k \in \{2, \dots, T+1\}$,

$$\begin{aligned} \|z^k\|_1 &= \sum_{i=1}^M |z_i^k| \leq \sum_{i=1}^M [|z_i^k - z_i^{k-1}| + |z_i^{k-1} - z_i^{k-2}| + \dots + |z_i^2 - z_i^1| + |z_i^1|] \leq \\ &\sum_{i=1}^M \left[\sum_{t=1}^{k-1} F\rho_t + 0 \right] = MF \sum_{t=1}^{k-1} \rho_t. \end{aligned} \quad (18)$$

引理得证。 \square

z^k 有界性结果是单步迭代量的累积, 其界取决于 ρ_k 的部分和序列。显然当 ρ_k 比 $1/k$ 低阶时, 由式 (18) 可知 z^k 有界。

利用 $\|z^k\|_1$ 的有界性, 可以推出下述 $\Psi_\beta(x, z)$ 关于 x 的 Lipschitz 光滑性。

引理 3.4 假设 2.2 满足时, 对 $\forall x, y \in X, \forall k \in [T+1]$, 以下不等式成立:

$$\|\nabla_x \Psi_\beta(x, z^k) - \nabla_x \Psi_\beta(y, z^k)\| \leq L_k \|x - y\|, \quad (19)$$

其中: $L_k = \beta G^2 + \beta FL + LF \sum_{t=1}^{k-1} \rho_t$ 。

证明 已知

$$\nabla_x \Psi_\beta(x, z) = \frac{1}{M} \sum_{i=1}^M [\beta f_i(x) + z_i]_+ \nabla f_i(x), f_i(x)$$

是 Lipschitz 光滑, 则 $\forall x, y \in X$,

$$\begin{aligned} &\|\nabla_x \Psi_\beta(x, z) - \nabla_x \Psi_\beta(y, z)\| \\ &= \left\| \frac{1}{M} \sum_{i=1}^M \left[[\beta f_i(x) + z_i]_+ \nabla f_i(x) - [\beta f_i(y) + z_i]_+ \nabla f_i(y) \right] \right\| \\ &\leq \frac{1}{M} \sum_{i=1}^M \left\| [\beta f_i(x) + z_i]_+ \nabla f_i(x) - [\beta f_i(y) + z_i]_+ \nabla f_i(y) \right\| \\ &= \frac{1}{M} \sum_{i=1}^M \left\| \left[[\beta f_i(x) + z_i]_+ - [\beta f_i(y) + z_i]_+ \right] \nabla f_i(x) + [\beta f_i(y) + z_i]_+ [\nabla f_i(x) - \nabla f_i(y)] \right\| \\ &\leq \frac{1}{M} \sum_{i=1}^M \left[\beta |f_i(x) - f_i(y)| \|\nabla f_i(x)\| + [\beta f_i(y) + z_i]_+ L \|x - y\| \right] \end{aligned}$$

$$\leq \frac{1}{M} \sum_{i=1}^M [\beta G^2 \|x - y\| +$$

$$(\beta F + |z_i|) L \|x - y\|]$$

$$= (\beta G^2 + \beta FL + L \|z\|_1/M) \|x - y\|.$$

引理 3.3 代入 $\|z\|_1$ 中结果得证。第 1 个和第 2 个不等号因为三角不等式,第 3 个不等号是对 $|f_i(x) - f_i(y)|$ 使用微分中值定理后利用 $f_i(x)$ 的梯度有界性(8)。

h^k 有界方差性质也同样利用 z^k 的有界性结果。

引理 3.5 在假设 2.2 下,

$$\mathbb{E}[\|h^k - \nabla_x \Psi_\beta(x^k, z^k)\|^2 | \zeta_{[k-1]}] \leq \frac{\sigma_k^2}{I_k}, \quad (20)$$

$$\text{其中: } \sigma_k = \sqrt{2\beta^2 F^2 G^2 + 2F^2 G^2 \left[\sum_{i=1}^{k-1} \rho_i \right]^2}, k \in [T].$$

证明 由式 (18) 可得

$$\begin{aligned} \|z^k\|^2 &= \sum_{i=1}^M |z_i^k|^2 \leq \sum_{i=1}^M \left[\sum_{t=1}^{k-1} F \rho_t \right]^2 \\ &= MF^2 \left[\sum_{t=1}^{k-1} \rho_t \right]^2, \forall k \in [T+1]. \end{aligned} \quad (21)$$

将上述不等式代入式 (11) 就能推知引理成立。□

最后考虑 ρ_k 的取值。式 (19) 的 Lipschitz 系数和式 (20) 方差界都依赖于 ρ_k 的部分和,这里给出一组 ρ_k 的设定,使得这 2 个系数被固定常数控制。

引理 3.6 假设 2.2 下,取常数 $\rho > 0$, 令

$$\rho_k \leq \frac{\rho}{\sqrt{Tk}}, k \in [T], \text{ 那么以下不等式成立:}$$

$$\|z^k\|_1 \leq 2MF\rho, \quad \|z^k\|^2 \leq 4MF^2\rho^2, \quad \forall k \in [T+1],$$

$$L_k \leq \beta G^2 + \beta FL + 2LF\rho =: L_\beta, \quad \forall k \in [T+1],$$

$$\begin{aligned} &\mathbb{E}[\mathbf{d}(\partial \mathcal{X}_0(x^{k+1}) + \nabla f_0(x^{k+1}) + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) + \mathcal{N}_X(x^{k+1}), 0) | \zeta_{[k-1]}] \\ &= \mathbb{E}[\mathbf{d}(\nabla f_0(x^{k+1}) - g_0^k + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) - h^k - \mathbf{D}_k(x^{k+1} - x^k) + \\ &\quad \mathcal{N}_X(x^{k+1}) - v^{k+1} + \partial \mathcal{X}_0(x^{k+1}) - d^{k+1}, 0) | \zeta_{[k-1]}] \\ &\leq \mathbb{E}[\|\nabla f_0(x^{k+1}) - g_0^k + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) - h^k - \mathbf{D}_k(x^{k+1} - x^k) + \\ &\quad \mathbf{d}(\mathcal{N}_X(x^{k+1}) - v^{k+1} + \partial \mathcal{X}_0(x^{k+1}) - d^{k+1}, 0) | \zeta_{[k-1]}] \\ &= \mathbb{E}[\|\nabla f_0(x^{k+1}) - g_0^k + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) - h^k - \mathbf{D}_k(x^{k+1} - x^k)\|]. \end{aligned}$$

因为 ∇f_0 的 Lipschitz 连续性和式 (19), 所以进一步有

$$\begin{aligned} &\mathbb{E}[\mathbf{d}(\nabla f_0(x^{k+1}) + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) + \mathcal{N}_X(x^{k+1}), 0) | \zeta_{[k-1]}] \\ &\leq \mathbb{E}[\|\nabla f_0(x^{k+1}) + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) - \nabla f_0(x^k) - \nabla_x \Psi_\beta(x^k, z^k)\| | \zeta_{[k-1]}] + \\ &\quad \mathbb{E}[\|\mathbf{D}_k(x^{k+1} - x^k)\| + \|\nabla f_0(x^k) - g_0^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\| | \zeta_{[k-1]}] \\ &\leq \mathbb{E}[\|\nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) - \nabla_x \Psi_\beta(x^{k+1}, z^k)\| + \|\nabla f_0(x^k) - g_0^k\| | \zeta_{[k-1]}] + \end{aligned}$$

$$\sigma_k \leq \sqrt{2\beta^2 F^2 G^2 + 8F^2 G^2 \rho^2} =: \sigma_h, \quad \forall k \in [T].$$

证明 对任意 $k \in [T+1]$,

$$\sum_{i=1}^{k-1} \rho_i \leq \frac{\rho}{\sqrt{T}} \sum_{i=1}^{k-1} \frac{1}{\sqrt{i}} \leq \frac{\rho}{\sqrt{T}} \int_{0^+}^{k-1} \frac{1}{\sqrt{u}} du =$$

$$\frac{2\rho\sqrt{k-1}}{\sqrt{T}} \leq 2\rho.$$

结合式 (18)~式 (21), 推论得证。□

3.2 最优性条件

取点 (\hat{x}, \hat{z}) 为

$$(\hat{x}, \hat{z}) = (x^{R+1}, (\beta f(x^{R+1}) + z^{R+1})_+ / M). \quad (22)$$

其中: (x^{R+1}, z^{R+1}) 是算法的输出。本节主要内容是证明在 $\mathcal{O}(\epsilon^{-4})$ 次迭代后, 点 (\hat{x}, \hat{z}) 可满足式 (12)。整体证明分为 3 个部分, 首先是找到式 (12) 的一个容易分析的上界, 然后分析 $\|x^{k+1} - x^k\|$ 的单步下降性, 该单步下降性是上界中的项, 最后证明收敛性结果。

首先, 可以从迭代式 (5)、式 (6) 得到以下结论。

引理 3.7 设假设 2.2、假设 2.3 满足, 任意步长 $\alpha_k, \rho_k > 0$, 则对任意 $k \geq 1$, 以下不等式成立:

$$\begin{aligned} &\mathbb{E}_{\zeta_{[k]}}[\mathbf{d}(\partial \mathcal{X}_0(x^{k+1}) + \nabla f_0(x^{k+1}) + \\ &\quad \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) + \mathcal{N}_X(x^{k+1}), 0)] \\ &\leq \mathbb{E}_{\zeta_{[k]}}\left[\left(\frac{1}{\alpha_k} + L + L_k\right) \|x^{k+1} - x^k\| + \frac{G}{M} \|z^{k+1} - z^k\|_1\right] + \\ &\quad \mathbb{E}_{\zeta_{[k]}}[\|\nabla f_0(x^k) - g_0^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\|]. \end{aligned}$$

证明 由式 (5) 的最优性条件, 可得 $\forall k \geq 1, \exists v^{k+1} \in \mathcal{N}_X(x^{k+1}), d^{k+1} \in \partial \mathcal{X}_0(x^{k+1})$ 使得

$$d^{k+1} + v^{k+1} + \mathbf{D}_k(x^{k+1} - x^k) + g_0^k + h^k = \mathbf{0}.$$

结合 $\mathbf{d}(u + \mathbf{A}, \mathbf{0}) \leq \|u\| + \mathbf{d}(\mathbf{A}, \mathbf{0})$, 可以推出

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{1}{\alpha_k} + L + L_k \right) \|x^{k+1} - x^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\| \mid \zeta_{[k-1]} \right] \\
& \leq \mathbb{E} \left[\frac{G}{M} \sum_{i=1}^M |(\beta f_i(x^{k+1}) + z_i^{k+1})_+ - (\beta f_i(x^{k+1}) + z_i^k)_+| + \|\nabla f_0(x^k) - g_0^k\| \mid \zeta_{[k-1]} \right] + \\
& \quad \mathbb{E} \left[\left(\frac{1}{\alpha_k} + L + L_k \right) \|x^{k+1} - x^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\| \mid \zeta_{[k-1]} \right] \\
& \leq \mathbb{E} \left[\left(\frac{1}{\alpha_k} + L + L_k \right) \|x^{k+1} - x^k\| + \frac{G}{M} \sum_{i=1}^M |z_i^{k+1} - z_i^k| \mid \zeta_{[k-1]} \right] + \\
& \quad \mathbb{E} [\|\nabla f_0(x^k) - g_0^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\| \mid \zeta_{[k-1]}].
\end{aligned}$$

第 3 个不等号因为 $\Psi_\beta(x, z)$ 在式(2)中的定义。
对上式两端取期望 $\mathbb{E}_{\zeta_{[k-1]}}$, 可得引理成立。 \square

接下来估计单次迭代带来的影响。

引理 3.8 在假设 2.2、假设 2.3 以及任意步长 $\alpha_k, \rho_k > 0$ 下, $\forall k \geq 1$ 有

$$\begin{aligned}
& \left(\frac{1}{\alpha_k} - \frac{L + L_k}{2} \right) \mathbb{E}_{\zeta_{[k]}} [\|x^{k+1} - x^k\|^2] \\
& \leq \mathbb{E}_{\zeta_{[k]}} [\mathcal{L}_\beta(x^k, z^k) - \mathcal{L}_\beta(x^{k+1}, z^{k+1}) + \\
& \quad \frac{F}{M} \|z^{k+1} - z^k\|_1 + 2\alpha_k \|g_0^k - \nabla f_0(x^k)\|^2 + \\
& \quad 2\alpha_k \|h^k - \nabla_x \Psi_\beta(x^k, z^k)\|^2].
\end{aligned}$$

证明 由 ∇f_0 和 $\nabla_x \Psi_\beta(x, z)$ 在引理 3.3 中相对于 x 的 Lipschitz 连续性可得

$$\begin{aligned}
& f_0(x^{k+1}) + \Psi_\beta(x^{k+1}, z^k) \leq f_0(x^k) + \Psi_\beta(x^k, z^k) + \\
& \quad \frac{L + L_k}{2} \|x^{k+1} - x^k\|^2 + \langle \nabla f_0(x^k) + \\
& \quad \nabla_x \Psi_\beta(x^k, z^k), x^{k+1} - x^k \rangle. \quad (23)
\end{aligned}$$

而同时,根据式(5)的最优性条件, $\exists d^{k+1} \in \partial \chi_0(x^{k+1})$, 对任意的 $x \in X$ 有

$$\langle d^{k+1} + g_0^k + h^k + D_k(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0.$$

令上式中的 $x = x^k$ 可得

$$\begin{aligned}
& \langle d^{k+1}, x^k - x^{k+1} \rangle + \langle g_0^k + h^k, x^k - x^{k+1} \rangle - \\
& \quad \frac{1}{\alpha_k} \|x^{k+1} - x^k\|^2 \geq 0. \quad (24)
\end{aligned}$$

对于凸部分,利用凸性性质得

$$\langle d^{k+1}, x^k - x^{k+1} \rangle \leq \chi_0(x^k) - \chi_0(x^{k+1}). \quad (25)$$

将式(25)代入式(24)然后和式(23)相加,得到

$$\begin{aligned}
& \mathcal{L}_\beta(x^{k+1}, z^k) \leq \mathcal{L}_\beta(x^k, z^k) + \\
& \quad \left(\frac{L + L_k}{2} - \frac{1}{\alpha_k} \right) \|x^{k+1} - x^k\|^2 + \\
& \quad \langle \nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k, x^{k+1} - x^k \rangle.
\end{aligned}$$

$$\text{记 } \hat{x}^{k+1} = \operatorname{argmin}_{x \in X} \left\{ \chi_0(x) + \langle \nabla f_0(x^k) + \right.$$

$\nabla_x \Psi_\beta(x^k, z^k), x \rangle + \frac{1}{2} \|x - x^k\|_{D_k}^2 \}$, 则由投影算

子的非扩张性(nonexpansiveness)可得

$$\begin{aligned}
& \|\hat{x}^{k+1} - x^{k+1}\| \leq \alpha_k \|\nabla f_0(x^k) + \\
& \quad \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k\|.
\end{aligned}$$

结合柯西不等式(Cauchy-Schwarz inequality),可推出

$$\begin{aligned}
& \mathcal{L}_\beta(x^{k+1}, z^k) \\
& \leq \mathcal{L}_\beta(x^k, z^k) + \left(\frac{L + L_k}{2} - \frac{1}{\alpha_k} \right) \|x^{k+1} - x^k\|^2 + \\
& \quad \langle \nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k, \hat{x}^{k+1} - x^k \rangle + \\
& \quad \langle \nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k, x^{k+1} - \hat{x}^{k+1} \rangle \\
& \leq \mathcal{L}_\beta(x^k, z^k) + \left(\frac{L + L_k}{2} - \frac{1}{\alpha_k} \right) \|x^{k+1} - x^k\|^2 + \\
& \quad \langle \nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k, \hat{x}^{k+1} - x^k \rangle + \\
& \quad \alpha_k \|\nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k\|^2.
\end{aligned}$$

对上述不等式两端取期望 $\mathbb{E}_{\zeta_k}[\cdot \mid \zeta_{[k-1]}]$, 然后由假设 2.3 和引理 2.3 可推出

$$\begin{aligned}
& \mathbb{E} [\mathcal{L}_\beta(x^{k+1}, z^k) \mid \zeta_{[k-1]}] \\
& \leq \mathbb{E} [\mathcal{L}_\beta(x^k, z^k) + \left(\frac{L + L_k}{2} - \frac{1}{\alpha_k} \right) \|x^{k+1} - x^k\|^2 + \\
& \quad \alpha_k \|\nabla f_0(x^k) + \nabla_x \Psi_\beta(x^k, z^k) - g_0^k - h^k\|^2 \mid \zeta_{[k-1]}].
\end{aligned}$$

另外,由 $\psi_\beta(u, v)$ 关于 $v \in \mathbb{R}_+$ 的 Lipschitz 连续性可得

$$\begin{aligned}
& \mathcal{L}_\beta(x^{k+1}, z^k) \\
& = \mathcal{L}_\beta(x^{k+1}, z^{k+1}) + \frac{1}{M} \sum_{i=1}^M [\psi_\beta(f_i(x^{k+1}), z_i^k) - \\
& \quad \psi_\beta(f_i(x^{k+1}), z_i^{k+1})] \\
& \geq \mathcal{L}_\beta(x^{k+1}, z^{k+1}) - \frac{1}{M} \sum_{i=1}^M |f_i(x^{k+1})| |z_i^k - z_i^{k+1}| \\
& \geq \mathcal{L}_\beta(x^{k+1}, z^{k+1}) - \frac{F}{M} \|z^k - z^{k+1}\|_1.
\end{aligned}$$

将以上 2 组不等式相加并对两端取期望 $\mathbb{E}_{\zeta_{[k-1]}}$, 再重新排列即可推知引理成立。 \square

接下来证明点 (22) 满足最优性近似 (12)。

定理 3.1 在假设 2.2、假设 2.3 下,对所有

$k \geq 1$ 设定

$$J_k = k^{\frac{1}{4}}, I_k = \min\left(\left\lceil k^{\frac{3}{4}} \right\rceil, M\right), \beta = T^{\frac{1}{4}},$$

$$\alpha_k = \alpha = \frac{T^{-\frac{1}{2}}}{L + G^2 + FL + 2LF\rho} = \frac{T^{-\frac{1}{2}}}{C}, \rho_k = \frac{\rho}{Tl_k}.$$

则对任意 $T \geq 1$, 下式成立。

$$\mathbb{E}_{R, \xi_{[T]}} \left[\mathbf{d} \left(\partial \mathcal{X}_0(x^{R+1}) + \nabla f_0(x^{R+1}) + \sum_{i=1}^M z_i \nabla f_i(x^{R+1}) + \mathcal{N}_X(x^{R+1}), 0 \right) \right] = \mathcal{O}(T^{-\frac{1}{4}}).$$

其中: $z_i = \frac{1}{M}(\beta f_i(x^{R+1}) + z_i^{R+1})_+, i \in [M]$ 。

证明 根据 $\Psi_\beta(x, z)$ 的定义和引理 3.7,

$$\begin{aligned} & \mathbb{E}_{R, \xi_{[T]}} \left[\mathbf{d} \left(\partial \mathcal{X}_0(x^{R+1}) + \nabla f_0(x^{R+1}) + \sum_{i=1}^M z_i \nabla f_i(x^{R+1}) + \mathcal{N}_X(x^{R+1}), \mathbf{0} \right) \right] \\ &= \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\mathbf{d} \left(\partial \mathcal{X}_0(x^{k+1}) + \nabla f_0(x^{k+1}) + \frac{1}{M} \sum_{i=1}^M (\beta f_i(x^{k+1}) + z_i^{k+1})_+ \nabla f_i(x^{k+1}) + \mathcal{N}_X(x^{k+1}), \mathbf{0} \right) \right] \\ &= \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\mathbf{d} \left(\partial \mathcal{X}_0(x^{k+1}) + \nabla f_0(x^{k+1}) + \nabla_x \Psi_\beta(x^{k+1}, z^{k+1}) + \mathcal{N}_X(x^{k+1}), \mathbf{0} \right) \right] \\ &\leq \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} + L + L_k \right) \|x^{k+1} - x^k\| + \frac{G}{M} \|z^{k+1} - z^k\|_1 \right] + \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\|\nabla f_0(x^k) - g_0^k\| + \|\nabla_x \Psi_\beta(x^k, z^k) - h^k\| \right] \\ &\leq \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} + L + L_\beta \right) \|x^{k+1} - x^k\| + \frac{G}{M} \|z^{k+1} - z^k\|_1 \right] + \frac{\sigma}{Tk^{\frac{1}{8}}} + \frac{\sigma_h}{Tk^{\frac{3}{8}}}. \end{aligned} \quad (26)$$

第 2 个不等式用了假设 2.3, 引理 3.6 和 $(\mathbb{E}[u])^2 \leq \mathbb{E}[u^2]$ 对任意随机变量 $u \in \mathbb{R}$ 。注意到, $\mathbb{E}[\|\nabla_x \Psi_\beta(x^k, z^k) - h^k\|] \leq \frac{\sigma_h}{k^{\frac{3}{8}}}$, 在 $\left\lceil k^{\frac{3}{4}} \right\rceil > M$ 时仍成立, 因为 $I_k = M$ 时 $h^k = \nabla_x \Psi_\beta(x^k, z^k)$ 。

接下来依次分析式 (26) 的右端项。首先考虑方差项

$$\frac{1}{T} \sum_{k=1}^T \left(\frac{\sigma}{k^{\frac{1}{8}}} + \frac{\sigma_h}{k^{\frac{3}{8}}} \right) \leq \frac{\sigma}{T} \int_{0^+}^T u^{-\frac{1}{8}} du + \frac{\sigma_h}{T} \int_{0^+}^T u^{-\frac{3}{8}} du =$$

$$\frac{8\sigma}{7T^{\frac{1}{8}}} + \frac{8\sigma_h}{5T^{\frac{3}{8}}}.$$

此外, 引理 3.6 以及 $\beta = T^{\frac{1}{4}}$ 可知 $\sigma_h = \sqrt{2\beta^2 F^2 G^2 + 8F^2 G^2 \rho^2} = \mathcal{O}(T^{\frac{1}{4}})$ 。因此

$$\frac{1}{T} \sum_{k=1}^T \left(\frac{\sigma}{k^{\frac{1}{8}}} + \frac{\sigma_h}{k^{\frac{3}{8}}} \right) \leq \frac{8\sigma}{7T^{\frac{1}{8}}} + \frac{8\sigma_h}{5T^{\frac{3}{8}}} = \mathcal{O}(T^{-\frac{1}{8}}). \quad (27)$$

然后是 $z^{k+1} - z^k$ 项, 类似于引理 3.2, 可以从式 (8) 和式 (6) 推出

$$\begin{aligned} & \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\frac{G}{M} \|z^{k+1} - z^k\|_1 \right] \\ &= \frac{G}{TM} \sum_{k=1}^T \mathbb{E}_{\xi_{[k-1]}} \left[\sum_{i=1}^M \frac{I_k}{M} \left| \rho_k \max \left(-\frac{z_i^k}{\beta}, f_i(x^k) \right) \right| \right] \\ &\leq \frac{G}{MT} \sum_{k=1}^T \rho_k I_k F = \frac{GF}{MT} \frac{\rho}{T} T = \frac{G\rho F}{MT} = \mathcal{O}(T^{-1}). \end{aligned} \quad (28)$$

最后考虑 $\|x^{k+1} - x^k\|$ 项。由步长 α_k 和罚参数 β 的设定可知对任意 $k \geq 1$ 有

$$\begin{aligned} \frac{1}{\alpha_k} &= T^{\frac{1}{2}}(L + G^2 + FL + 2LF\rho) \\ &\geq L + \beta G^2 + \beta FL + 2LF\rho = L + L_\beta \\ &\geq \frac{L + L_\beta}{2}. \end{aligned}$$

因此, 记 $\alpha_T = \sqrt{\sum_{k=1}^T \left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right)}$, 根据柯西

不等式, 即对任意 $u_k \geq 0$, $\sum u_k v_k \leq \sqrt{\sum u_k} \sqrt{\sum u_k v_k^2}$, 以及对任意随机变量 $u \in \mathbb{R}$, $(\mathbb{E}[u])^2 \leq \mathbb{E}[u^2]$, 可以推知

$$\begin{aligned} & \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\| \right] \\ &\leq \frac{\alpha_T}{T} \sqrt{\sum_{k=1}^T \left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) [\mathbb{E}_{\xi_{[k]}} [\|x^{k+1} - x^k\|]]^2} \\ &\leq \frac{\alpha_T}{T} \sqrt{\sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\|^2 \right]}. \end{aligned}$$

同时, 利用引理 3.8、引理 3.6 和式 (28) 可得

$$\begin{aligned} & \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\|^2 \right] \\ &\leq \mathbb{E}_{\xi_{[T]}} [\mathcal{L}_\beta(x^1, z^1) - \mathcal{L}_\beta(x^{T+1}, z^{T+1})] + \\ &\quad \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\frac{F}{M} \|z^{k+1} - z^k\|_1 \right] + \sum_{k=1}^T 2\alpha_k \left(\frac{\sigma^2}{k^{\frac{1}{4}}} + \frac{\sigma_h^2}{k^{\frac{3}{4}}} \right) \\ &\leq \mathbb{E}_{\xi_{[T]}} [\mathcal{L}_\beta(x^1, z^1) - \mathcal{L}_\beta(x^{T+1}, z^{T+1})] + \end{aligned}$$

$$\frac{F^2 \rho}{M} + \frac{2 \left(\frac{4}{3} \sigma^2 T^{\frac{3}{4}} + 4 \sigma_h^2 T^{\frac{1}{4}} \right)}{C \sqrt{T}}.$$

其中最后一个不等式还利用了 $\alpha_k = (CT^{\frac{1}{2}})^{-1}$,

$$\sum_{k=1}^T k^{-1/4} \leq \int_{0^+}^T u^{-1/4} du = \frac{4}{3} T^{\frac{3}{4}}, \text{ 以及 } \sum_{k=1}^T k^{-3/4} \leq \int_{0^+}^T u^{-3/4} du = 4 T^{\frac{1}{4}}.$$

考虑项 $\mathcal{L}_\beta(x^1, z^1) - \mathcal{L}_\beta(x^{T+1}, z^{T+1})$, 首先估计 $\mathcal{L}_\beta(x^1, z^1)$ 的上界, 因起始对偶点 $z^1 = \mathbf{0}$, 所以有

$$\begin{aligned} \mathcal{L}_\beta(x^1, z^1) &= \chi_0(x^1) + f_0(x^1) + \frac{\beta}{2M} \sum_{i=1}^M ([f_i(x^1)]_+)^2 \\ &\Rightarrow \mathcal{L}_\beta(x^1, z^1) \leq \chi_0(x^1) + F + \frac{\beta}{2} F^2 \\ &= \chi_0(x^1) + F + \frac{F^2}{2} T^{\frac{1}{4}}. \end{aligned}$$

然后是 $\mathcal{L}_\beta(x^{T+1}, z^{T+1})$ 的下界, 由式 (2) 可得 $\psi_\beta(u, v) \geq \frac{-v^2}{2\beta}$, 结合式 (8) 和推论引理 3.6 得

$$\begin{aligned} \mathcal{L}_\beta(x^{T+1}, z^{T+1}) &= \chi_0(x^{T+1}) + f_0(x^{T+1}) + \\ &\quad \frac{1}{M} \sum_{i=1}^M \psi_\beta(f_i(x^{T+1}), z_i^{T+1}) \\ &\geq W - F - \frac{1}{M} \sum_{i=1}^M \frac{(z_i^{T+1})^2}{2\beta} \\ &\geq W - F - \frac{4MF^2\rho^2}{2\beta M} \\ &= W - F - 2F^2\rho^2 T^{-\frac{1}{4}} \\ &\geq W - F - 2F^2\rho^2. \end{aligned}$$

所以 $\|x^{k+1} - x^k\|$ 项的上界为

$$\begin{aligned} &\sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\|^2 \right] \\ &\leq \chi_0(x^1) + W + 2F + 2F^2\rho^2 + \frac{F^2}{2} T^{\frac{1}{4}} + \\ &\quad \frac{F^2 \rho}{M} + \frac{2 \left(\frac{4}{3} \sigma^2 T^{\frac{3}{4}} + 4 \sigma_h^2 T^{\frac{1}{4}} \right)}{C \sqrt{T}}. \end{aligned}$$

考虑 $\|x^{k+1} - x^k\|$ 项。记 $C' = \chi_0(x^1) + W + 2F + 2F^2\rho^2 + \frac{F^2 \rho}{M}$, 代入 $\alpha_k \equiv \alpha = (CT^{\frac{1}{2}})^{-1}$, 所以

$$\begin{aligned} &\|x^{k+1} - x^k\| \text{ 项的上界为 } \\ &\frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\| \right] \\ &\leq \frac{\alpha_T}{T} \sqrt{C' + \frac{F^2}{2} T^{\frac{1}{4}} + \frac{2 \left(\frac{4}{3} \sigma^2 T^{\frac{3}{4}} + 4 \sigma_h^2 T^{\frac{1}{4}} \right)}{C \sqrt{T}}} \end{aligned}$$

$$\leq \frac{1}{T} \sqrt{CT^{\frac{3}{2}}} \sqrt{C' + \frac{F^2}{2} T^{\frac{1}{4}} + \frac{2 \left(\frac{4}{3} \sigma^2 T^{\frac{3}{4}} + 4 \sigma_h^2 T^{\frac{1}{4}} \right)}{C \sqrt{T}}}.$$

其中: $\alpha_T = \sqrt{\sum_{k=1}^T \left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right)}$ 。由 α_k 的定义, 有 $\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \geq \frac{L + L_\beta}{2}$ 。利用后者可以进一步得到

$$\begin{aligned} &\frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} + L + L_\beta \right) \|x^{k+1} - x^k\| \right] \\ &= \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\| \right] + \\ &\quad \frac{3L + 3L_\beta}{2} \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} [\|x^{k+1} - x^k\|] \\ &\leq \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[4 \left(\frac{1}{\alpha_k} - \frac{L + L_\beta}{2} \right) \|x^{k+1} - x^k\| \right] \\ &\leq \frac{4}{T} \sqrt{CT^{\frac{3}{2}}} \sqrt{C' + \frac{F^2}{2} T^{\frac{1}{4}} + \frac{2 \left(\frac{4}{3} \sigma^2 T^{\frac{3}{4}} + 4 \sigma_h^2 T^{\frac{1}{4}} \right)}{C \sqrt{T}}} \\ &= \mathcal{O}(T^{-\frac{1}{8}}). \end{aligned} \quad (29)$$

最后一个等式因为 $\sigma_h = \mathcal{O}(T^{\frac{1}{4}})$ 。将式 (27) ~ 式 (29) 加到式 (26) 中, 定理得证。□

3.3 约束违反度

该节证明在 $\mathcal{O}(\epsilon^{-8})$ 次迭代后, 点 (22) 中的 x^{R+1} 满足约束违反度条件 (13)

$$\mathbb{E}_{R, \xi_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M [f_i(x^{R+1})]_+ \right] \leq \epsilon.$$

分析收敛性前先给出一些需要的性质, 分别是 $\Psi_\beta(x, z)$ 对 x 的凸性, $\Psi_\beta(x, z)$ 在可行域内非正性, 以及原始对偶解的对偶间隙。

引理 3.9 当 $\{f_i, i \in [M]\}$ 均为凸函数时, $\Psi_\beta(x, z)$ 对 x 凸。

证明 Ψ_β 的定义如下,

$$\Psi_\beta(x, z) = \frac{1}{M} \sum_{i=1}^M \psi_\beta(f_i(x), z_i),$$

$$\psi_\beta(u, v) = \begin{cases} uv + \frac{\beta}{2} u^2, & \beta u + v \geq 0, \\ -\frac{v^2}{2\beta}, & \beta u + v < 0. \end{cases}$$

因为求和不改变凸性, 只需证明对任意 $i \in [M]$, $\psi_\beta(f_i(x), z_i)$ 关于 x 凸即可。

设 v 为任意常数,

当 $\beta u + v < 0$ 时, $\psi_\beta(u, v)$ 为常数, 且

$$\psi_\beta(u, v) = -\frac{v^2}{2\beta}, \nabla_u \psi_\beta(u, v) = 0;$$

当 $\beta u + v = 0$ 时

$$\psi_\beta(u, v) = \frac{u}{2}(v + v + \beta u) = -\frac{v}{2\beta} \times v = -\frac{v^2}{2\beta};$$

当 $\beta u + v \geq 0$ 时,

$$\nabla_u \psi_\beta(u, v) = \beta u + v \geq 0.$$

所以 $\psi_\beta(\cdot, v)$ 与 $\nabla_u \psi_\beta(\cdot, v)$ 都是关于 u 的单调增函数, 所以 $\psi(\cdot, v)$ 是关于 u 的凸函数。

设 z_i 为任意常数, $\lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1, x_1, x_2 \in \mathcal{X}$, 有

$$\begin{aligned} & \lambda_1 \psi_\beta(f_i(x_1), z_i) + \lambda_2 \psi_\beta(f_i(x_2), z_i) \\ & \stackrel{(1)}{\geq} \psi_\beta(\lambda_1 f_i(x_1) + \lambda_2 f_i(x_2), z_i) \\ & \stackrel{(2)}{\geq} \psi_\beta(f_i(\lambda_1 x_1 + \lambda_2 x_2), z_i). \end{aligned}$$

(1) 是 $\psi_\beta(\cdot, v)$ 凸性, (2) 是 $\psi_\beta(\cdot, v)$ 对 u 单调递增和 f_i 的凸性。引理得证。 \square

引理 3.10 对任意 $x \in X \cap \{x | f_i(x) \leq 0, \forall i \in [M]\}$, $\Psi_\beta(x, z) \leq 0$ 。

证明 若 $\beta f_i(x) + z_i \geq 0$

$$\begin{aligned} \psi_\beta(f_i(x), z_i) &= f_i(x)(z_i + \beta f_i(x)/2) = \\ & f_i(x)(z_i + \beta f_i(x) - \beta f_i(x)/2) \leq 0. \end{aligned}$$

若 $\beta f_i(x) + z_i < 0$,

$$\psi_\beta(f_i(x), z_i) = \frac{-z_i^2}{2\beta} \leq 0.$$

因此, $\Psi_\beta(x, z) = \frac{1}{M} \sum_{i=1}^M \psi_\beta(f_i(x), z_i) \leq 0$ 。 \square

引理 3.11 记 (x^*, z^*) 为满足假设 2.1 的原始-对偶解, 则 $\forall x \in X$,

$$\psi_0(x) \geq \psi_0(x^*) - \frac{1}{M} \sum_{i=1}^M z_i^* f_i(x) - \frac{L}{2} \|x - x^*\|^2.$$

证明 由 f_0 的 Lipschitz 光滑性可得, 对任意 $x \in X$,

$$\begin{aligned} f_0(x) &\geq f_0(x^*) + \langle \nabla f_0(x^*), x - x^* \rangle - \\ & \frac{L}{2} \|x - x^*\|^2. \end{aligned}$$

同时, 由 x^* 的最优性条件

$$\begin{aligned} 0 &\in \partial \mathcal{X}_0(x^*) + \nabla f_0(x^*) + \mathcal{N}_X(x^*) + \\ & \frac{1}{M} \sum_{i=1}^M z_i^* \nabla f_i(x^*), \end{aligned}$$

可知存在向量 $u \in \mathcal{N}_X(x^*), d \in \partial \mathcal{X}_0(x^*)$ 使得

$$-u - d - \frac{1}{M} \sum_{i=1}^M z_i^* \nabla f_i(x^*) = \nabla f_0(x^*).$$

将该式代入到光滑性不等式中, 并用法锥 $\mathcal{N}_X(x^*)$ 的定义, 约束函数 $f_i, i \in [M]$ 的凸性, $z_i^* \geq 0, z_i^* f_i(x^*) = 0, i \in [M]$, 和 $\mathcal{X}_0(x)$ 的凸性, 简化结果即可。 \square

下述引理类似于文献[14]的引理 3.3, 但两者的区别在于算法 CSALM 中每次迭代抽取了多个样本, 因此这里还是详细给出相应的证明。

引理 3.12 算法 CSALM 对任意确定或随机的向量 $z \in \mathbb{R}_+^M$, 有以下不等式成立:

$$\begin{aligned} & -\Psi_\beta(x^k, z^k) + \frac{1}{M} \sum_{i=1}^M z_i f_i(x^k) + \\ & \frac{1}{2\rho_k I_k} \mathbb{E}[\|z^{k+1} - z\|^2 | \zeta_{[k-1]}] \leq \frac{1}{2\rho_k I_k} \|z^k - z\|^2 - \\ & \frac{1}{2\rho_k I_k} \left(\frac{\beta}{\rho_k} - 1 \right) \mathbb{E}[\|z^{k+1} - z^k\|^2 | \zeta_{[k-1]}] + \\ & \mathbb{E} \left[\langle z^k - z, \frac{M}{I_k} e_{I_k} \odot \nabla_z \Psi_\beta(x^k, z^k) - \right. \\ & \left. \nabla_z \Psi_\beta(x^k, z^k) \rangle | \zeta_{[k-1]} \right]. \end{aligned}$$

其中: \odot 代表对应分量相乘。

证明 因为 $\nabla_{z_i} \psi_\beta(f_i(x), z_i) = \max\left(-\frac{z_i}{\beta}, f_i(x)\right)$,

所以

$$\begin{aligned} \nabla_{z_i} \Psi_\beta(x, z) &= \nabla_{z_i} \left[\frac{1}{M} \sum_{p=1}^M \psi_\beta(f_p(x), z_p) \right] \\ &= \frac{1}{M} \nabla_{z_i} \psi_\beta(f_i(x), z_i) \\ &= \frac{1}{M} \max\left(-\frac{z_i}{\beta}, f_i(x)\right). \end{aligned}$$

所以, z 的迭代公式(6)可以改写为

$$z^{k+1} = z^k + M\rho_k e_{\mathcal{I}_k} \odot \nabla_z \Psi_\beta(x^k, z^k).$$

其中: $\odot: \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^M$ 代表对应分量相乘; $e_{\mathcal{I}_k} \in \mathbb{R}^M$ 记为集合 \mathcal{I}_k 的指示向量(即对所有 $i \in \mathcal{I}_k, e_{\mathcal{I}_k}$ 对应的分量为 1, 其余分量为 0)。因此,

$$\begin{aligned} \frac{1}{I_k \rho_k} \langle z^k - z, z^{k+1} - z^k \rangle &= \langle z^k - z, \nabla_z \Psi_\beta(x^k, z^k) \rangle + \\ & \langle z^k - z, \frac{M}{I_k} e_{\mathcal{I}_k} \odot \nabla_z \Psi_\beta(x^k, z^k) - \nabla_z \Psi_\beta(x^k, z^k) \rangle. \end{aligned}$$

记 $\mathcal{I}_k^+ = \{i \in [M] | \beta f_i(x^k) + z_i^k \geq 0\}, \mathcal{I}_k^- = [M] \setminus \mathcal{I}_k^+$ 。注意到对任意 $z \geq 0$ 以及任意 $i \in \mathcal{I}_k^-$, $z_i \left(f_i(x^k) + \frac{z_i^k}{\beta} \right) \leq 0$ 成立, 结合 $\Psi_\beta(x, z)$ 在式(2)

中的定义, 所以有

$$\begin{aligned}
 & -\Psi_{\beta}(x^k, z^k) + \frac{1}{M} \sum_{i=1}^M z_i f_i(x^k) + \langle z^k - z, \nabla_z \Psi_{\beta}(x^k, z^k) \rangle \\
 = & -\frac{1}{M} \sum_{i \in \mathcal{I}_k^+} \left[z_i^k f_i(x^k) + \frac{\beta}{2} [f_i(x^k)]^2 \right] + \\
 & \frac{1}{M} \sum_{i \in \mathcal{I}_k^-} \frac{(z_i^k)^2}{2\beta} + \frac{1}{M} \sum_{i=1}^M z_i f_i(x^k) + \\
 & \frac{1}{M} \sum_{i \in \mathcal{I}_k^+} (z_i^k - z_i) f_i(x^k) - \frac{1}{M} \sum_{i \in \mathcal{I}_k^-} (z_i^k - z_i) \frac{z_i^k}{\beta} \\
 = & -\frac{1}{M} \sum_{i \in \mathcal{I}_k^+} \frac{\beta}{2} [f_i(x^k)]^2 - \\
 & \frac{1}{M} \sum_{i \in \mathcal{I}_k^-} \left[\frac{(z_i^k)^2}{2\beta} - z_i \left(f_i(x^k) + \frac{z_i^k}{\beta} \right) \right] \\
 \leq & -\frac{1}{M} \sum_{i \in \mathcal{I}_k^+} \frac{\beta}{2} [f_i(x^k)]^2 - \frac{1}{M} \sum_{i \in \mathcal{I}_k^-} \frac{(z_i^k)^2}{2\beta}.
 \end{aligned}$$

另外, 由式(6)可得

$$\begin{aligned}
 & \mathbb{E}[\|z^{k+1} - z^k\|^2 \zeta_{[k-1]}] \\
 = & \frac{I_k}{M} \left[\sum_{i \in \mathcal{I}_k^+} (\rho_k f_i(x^k))^2 + \sum_{i \in \mathcal{I}_k^-} \left(\frac{\rho_k z_i^k}{\beta} \right)^2 \right] \\
 = & \frac{I_k \rho_k^2}{M\beta} \left[\sum_{i \in \mathcal{I}_k^+} \beta (f_i(x^k))^2 + \sum_{i \in \mathcal{I}_k^-} \frac{(z_i^k)^2}{\beta} \right].
 \end{aligned}$$

结合以上 3 组(不)等式, 根据

$$\langle z^k - z, z^{k+1} - z^k \rangle =$$

$$\frac{1}{2} [\|z^{k+1} - z\|^2 - \|z^k - z\|^2 - \|z^{k+1} - z^k\|^2].$$

引理得证。□

基于引理 3.12, 可以得到以下重要结论。

定理 3.2 在假设 2.1~假设 2.3 以及任意步长 $\{\alpha_k > 0, k \in [T]\}$ 下, 下列不等式

$$\begin{aligned}
 & \mathbb{E} \left[\left(\psi_0(x^k) + \frac{1}{M} \sum_{i=1}^M z_i f_i(x^k) \right) \middle| \zeta_{[k-1]} \right] + \\
 & \mathbb{E} \left[\frac{1}{2\alpha_k} \|x^{k+1} - x\|^2 + \frac{1}{2\rho_k I_k} \|z^{k+1} - z\|^2 \middle| \zeta_{[k-1]} \right] \\
 \leq & \mathbb{E} \left[\psi_0(x) + \Psi_{\beta}(x, z^k) + \frac{1}{2} \left(\frac{1}{\alpha_k} + L \right) \right. \\
 & \left. \|x^k - x\|^2 \middle| \zeta_{[k-1]} \right] + \mathbb{E} \left[\frac{1}{2\rho_k I_k} \|z^k - z\|^2 \middle| \zeta_{[k-1]} \right] + \\
 & \alpha_k \left(\sigma^2 + G^2 + 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \mathbb{E}[\|z^k\|^2 | \zeta_{[k-1]}] \right) - \\
 & \frac{1}{2\rho_k I_k} \left(\frac{\beta}{\rho_k} - 1 \right) \mathbb{E}[\|z^{k+1} - z^k\|^2 | \zeta_{[k-1]}] - \\
 & \mathbb{E}[\langle x^k - x, (g_0^k + h^k) - (\nabla f_0(x^k) + \nabla_x \Psi_{\beta}(x^k, z^k)) \rangle | \zeta_{[k-1]}] +
 \end{aligned}$$

$$\begin{aligned}
 & \mathbb{E} \left[\langle z^k - z, \frac{M}{I_k} e_{I_k} \odot \nabla_z \Psi_{\beta}(x^k, z^k) - \nabla_z \Psi_{\beta}(x^k, z^k) \rangle \middle| \zeta_{[k-1]} \right] + \\
 & \mathbb{E}[\chi_0(x^k) - \chi_0(x^{k+1}) | \zeta_{[k-1]}].
 \end{aligned}$$

对任意 $x \in X, z \in \mathbb{R}_+^M$ 成立。

证明 由式(5)的最优性条件可得, $\exists d^{k+1} \in \chi_0(x^{k+1})$, 使得

$$\begin{aligned}
 & \langle d^{k+1} + g_0^k + h^k + D_k(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0, \\
 & \forall x \in X. \quad (30)
 \end{aligned}$$

下面将不等式左端拆为 4 项, 依次分析。

对内积项 $\langle g_0^k, x - x^{k+1} \rangle$, 根据杨氏不等式和式(7),

$$\begin{aligned}
 & \mathbb{E}[\langle g_0^k, x^{k+1} - x \rangle | \zeta_{[k-1]}] \\
 = & \mathbb{E}[\langle g_0^k, x^{k+1} - x^k \rangle + \langle g_0^k, x^k - x \rangle | \zeta_{[k-1]}] \\
 \geq & \mathbb{E} \left[-\frac{1}{4\alpha_k} \|x^{k+1} - x^k\|^2 - \alpha_k \|g_0^k\|^2 + \right. \\
 & \left. \langle \nabla f_0(x^k), x^k - x \rangle + \langle g_0^k - \nabla f_0(x^k), x^k - x \rangle \middle| \zeta_{[k-1]} \right] \\
 \geq & \mathbb{E} \left[-\frac{1}{4\alpha_k} \|x^{k+1} - x^k\|^2 - \alpha_k (\sigma^2 + G^2) + \right. \\
 & \left. f_0(x^k) - f_0(x) - \frac{L}{2} \|x^k - x\|^2 + \right. \\
 & \left. \langle g_0^k - \nabla f_0(x^k), x^k - x \rangle \middle| \zeta_{[k-1]} \right]. \quad (31)
 \end{aligned}$$

第 2 个不等号还用到了依据假设 2.3 和式(8)得到的下列不等式

$$\begin{aligned}
 & \mathbb{E}[\|g_0^k\|^2 \zeta_{[k-1]}] = \mathbb{E}[\|g_0^k - \nabla f_0(x^k)\|^2 \zeta_{[k-1]}] + \\
 & \mathbb{E}[\|\nabla f_0(x^k)\|^2 \zeta_{[k-1]}] \leq \sigma^2 + G^2.
 \end{aligned}$$

对内积项 $\langle h^k, x - x^{k+1} \rangle$, 根据 Ψ_{β} 在 x 上的凸性和杨氏不等式, 可以得到类似的估计,

$$\begin{aligned}
 & \mathbb{E}[\langle h^k, x^{k+1} - x \rangle | \zeta_{[k-1]}] \\
 = & \mathbb{E}[\langle h^k, x^{k+1} - x^k \rangle | \zeta_{[k-1]}] + \\
 & \mathbb{E}[\langle h^k, x^k - x \rangle | \zeta_{[k-1]}] \\
 \geq & \mathbb{E} \left[-\frac{1}{4\alpha_k} \|x^{k+1} - x^k\|^2 + \langle \nabla_x \Psi_{\beta}(x^k, z^k), x^k - x \rangle - \right. \\
 & \left. \alpha_k \|h^k\|^2 + \langle h^k - \nabla_x \Psi_{\beta}(x^k, z^k), x^k - x \rangle \middle| \zeta_{[k-1]} \right] \\
 \geq & \mathbb{E} \left[-\frac{1}{4\alpha_k} \|x^{k+1} - x^k\|^2 + \Psi_{\beta}(x^k, z^k) - \Psi_{\beta}(x, z^k) - \right. \\
 & \left. \alpha_k \|h^k\|^2 + \langle h^k - \nabla_x \Psi_{\beta}(x^k, z^k), x^k - x \rangle \middle| \zeta_{[k-1]} \right].
 \end{aligned}$$

由引理 2.3 和对任意向量 $\{u^i\}$, $\left\| \sum_{i=1}^M u^i \right\|^2 \leq$

$M \sum_{i=1}^M \|u^i\|^2$, 可知

$$\mathbb{E} [\|h^k\|^2 | \zeta_{[k-1]}] \leq 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \|z^k\|^2.$$

所以

$$\begin{aligned} & \mathbb{E} [\langle h^k, x^{k+1} - x \rangle | \zeta_{[k-1]}] \\ & \geq \mathbb{E} \left[-\frac{1}{4\alpha_k} \|x^{k+1} - x^k\|^2 | \zeta_{[k-1]} \right] + \\ & \mathbb{E} [\langle h^k - \nabla_x \Psi_\beta(x^k, z^k), x^k - x \rangle | \zeta_{[k-1]}] + \\ & \mathbb{E} [\Psi_\beta(x^k, z^k) - \Psi_\beta(x, z^k) | \zeta_{[k-1]}] + \\ & \mathbb{E} \left[-\alpha_k \left(2\beta^2 F^2 G^2 + \frac{2G^2}{M} \|z^k\|^2 \right) | \zeta_{[k-1]} \right]. \quad (32) \end{aligned}$$

对内积项 $\langle D_k(x^{k+1} - x^k), x^{k+1} - x \rangle$, 取期望得

$$\begin{aligned} & \mathbb{E} [\langle D_k(x^{k+1} - x^k), x^{k+1} - x \rangle | \zeta_{[k-1]}] \\ & = \frac{1}{2} \mathbb{E} [\|x^{k+1} - x\|_{D_k}^2 - \|x^k - x\|_{D_k}^2 + \\ & \|x^{k+1} - x^k\|_{D_k}^2 | \zeta_{[k-1]}]. \quad (33) \end{aligned}$$

对式 (30) 两端取期望 $\mathbb{E} [\cdot | \zeta_{[k-1]}]$, 并加上式 (31)~式 (33) 可得

$$\begin{aligned} & \mathbb{E} [f_0(x^k) - f_0(x) + \Psi_\beta(x^k, z^k) - \Psi_\beta(x, z^k) | \zeta_{[k-1]}] + \\ & \mathbb{E} \left[\frac{1}{2\alpha_k} \|x^{k+1} - x\|^2 | \zeta_{[k-1]} \right] \\ & \leq \frac{1}{2} \mathbb{E} \left[\left(L + \frac{1}{\alpha_k} \right) \|x^k - x\|^2 | \zeta_{[k-1]} \right] + \\ & \alpha_k \left(\sigma^2 + G^2 + 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \mathbb{E} [\|z^k\|^2 | \zeta_{[k-1]}] \right) - \\ & \mathbb{E} [\langle x^k - x, (g_0^k + h^k) - (\nabla f_0(x^k) + \\ & \nabla_x \Psi_\beta(x^k, z^k)) \rangle | \zeta_{[k-1]}] - \\ & \mathbb{E} [\langle d^{k+1}, x^{k+1} - x \rangle | \zeta_{[k-1]}]. \quad (34) \end{aligned}$$

考虑项 $\langle d^{k+1}, x^{k+1} - x \rangle$, 因为 $d^{k+1} \in \partial \chi_0(x^{k+1})$, 所以由 χ_0 的凸性

$$\begin{aligned} 0 & \geq \langle d^{k+1}, x^{k+1} - x \rangle \geq \chi_0(x^{k+1}) - \chi_0(x) \\ & \Leftrightarrow \chi_0(x^k) - \chi_0(x^{k+1}) \geq \chi_0(x^k) - \chi_0(x). \quad (35) \end{aligned}$$

式 (35) 加在式 (34) 上得到

$$\begin{aligned} & \mathbb{E} [\psi_0(x^k) - \psi_0(x) + \Psi_\beta(x^k, z^k) - \Psi_\beta(x, z^k) + \\ & \frac{1}{2\alpha_k} \|x^{k+1} - x\|^2 | \zeta_{[k-1]}] \\ & \leq \frac{1}{2} \mathbb{E} \left[\left(L + \frac{1}{\alpha_k} \right) \|x^k - x\|^2 | \zeta_{[k-1]} \right] + \\ & \alpha_k \left(\sigma^2 + G^2 + 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \mathbb{E} [\|z^k\|^2 | \zeta_{[k-1]}] \right) - \\ & \mathbb{E} [\langle x^k - x, (g_0^k + h^k) - (\nabla f_0(x^k) + \\ & \nabla_x \Psi_\beta(x^k, z^k)) \rangle | \zeta_{[k-1]}] + \end{aligned}$$

$$\mathbb{E} [\chi_0(x^k) - \chi_0(x^{k+1}) | \zeta_{[k-1]}]. \quad (36)$$

结合引理 3.12 和式 (36), 重排后即可得到结论. \square

在定理 3.1 的基础上, 可以得到关于 $\sum_{k=1}^T \mathbb{E}_{\zeta_{[k]}} I_k [\|z^{k+1} - z^k\|^2]$ 的估计。

引理 3.13 在假设 2.1~假设 2.3 下, 对任意 $k \geq 1$ 令

$$J_k = \left\lceil k^{\frac{1}{4}} \right\rceil, I_k = \min \left(\left\lceil k^{\frac{3}{4}} \right\rceil, M \right), \beta = T^{\frac{1}{4}},$$

$$\alpha_k = \alpha = \frac{T^{-\frac{1}{2}}}{L + G^2 + FL + 2LF\rho} = \frac{T^{-\frac{1}{2}}}{C},$$

$$\rho \in \left(0, \frac{1}{2} \right), \rho_k = \frac{\rho}{TI_k}.$$

则对任意 $T \geq 1$, 下列不等式成立:

$$\begin{aligned} & \sum_{k=1}^T I_k \mathbb{E}_{\zeta_{[k]}} [\|z^{k+1} - z^k\|^2] \\ & \leq 4\rho^2 \left[\frac{C\|x^1 - x^*\|^2}{2} + \frac{\sigma^2 + G^2 + 8G^2 F^2 \rho^2}{C} \right] T^{-\frac{7}{4}} + \\ & 4\rho^2 \left[LD_X^2 + \frac{\|z^*\|^2}{2\rho} + \frac{2F^2 G^2}{C} \right] T^{-\frac{5}{4}} + \\ & 4\rho^2 [\chi_0(x^1) - W] T^{-\frac{4}{5}}. \end{aligned}$$

证明 令定理 3.1 中 $(x, z) = (x^*, z^*)$ 、引理 3.11 中 $x = x^k$ 、引理 3.10 中 $(x, z) = (x^*, z^k)$, 加和可得

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{2\alpha_k} \|x^{k+1} - x^*\|^2 + \frac{1}{2\rho_k I_k} \|z^{k+1} - z^*\|^2 | \zeta_{[k-1]} \right] \\ & \leq \mathbb{E} \left[\frac{1}{2} \left(\frac{1}{\alpha_k} + 2L \right) \|x^k - x^*\|^2 + \right. \\ & \left. \frac{1}{2\rho_k I_k} \|z^k - z^*\|^2 | \zeta_{[k-1]} \right] + \alpha_k \left(\sigma^2 + G^2 + \right. \\ & \left. 2\beta^2 F^2 G^2 + \frac{2G^2}{M} \mathbb{E} [\|z^k\|^2 | \zeta_{[k-1]}] \right) - \\ & \frac{1}{2\rho_k I_k} \left(\frac{\beta}{\rho_k} - 1 \right) \mathbb{E} [\|z^{k+1} - z^k\|^2 | \zeta_{[k-1]}] - \\ & \mathbb{E} [\langle x^k - x^*, (g_0^k + h^k) - (\nabla f_0(x^k) + \\ & \nabla_x \Psi(x^k, z^k)) \rangle | \zeta_{[k-1]}] + \\ & \mathbb{E} \left[\langle z^k - z^*, \frac{M}{I_k} e_{I_k} \odot \nabla_z \Psi(x^k, z^k) - \right. \\ & \left. \nabla_z \Psi(x^k, z^k) \rangle | \zeta_{[k-1]} \right] + \\ & \mathbb{E} [\chi_0(x^k) - \chi_0(x^{k+1}) | \zeta_{[k-1]}]. \end{aligned}$$

代入 g_0^k, h^k 的无偏性和引理 3.6 得

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{2\alpha_k} \|x^{k+1} - x^*\|^2 + \frac{1}{2\rho_k I_k} \|z^{k+1} - z^*\|^2 \mid \zeta_{[k-1]} \right] \\ & \leq \mathbb{E} \left[\frac{1}{2} \left(\frac{1}{\alpha_k} + 2L \right) \|x^k - x^*\|^2 + \right. \\ & \quad \left. \frac{1}{2\rho_k I_k} \|z^k - z^*\|^2 \mid \zeta_{[k-1]} \right] + \\ & \quad \alpha_k (\sigma^2 + G^2 + 2\beta^2 F^2 G^2 + 8G^2 F^2 \rho^2) - \\ & \quad \frac{1}{2\rho_k I_k} \left(\frac{\beta}{\rho_k} - 1 \right) \mathbb{E} [\|z^{k+1} - z^k\|^2 \mid \zeta_{[k-1]}]. \end{aligned}$$

不等式两端取期望 $\mathbb{E}_{\zeta_{[k-1]}}[\cdot]$, 求 $k=1$ 到 T 的累加和, 由设定 $\alpha_k \equiv \alpha, \rho_k I_k \equiv \frac{\rho}{T}, \beta = T^{\frac{1}{4}}$ 得

$$\begin{aligned} & \sum_{k=1}^T \frac{T}{2\rho} \left(\frac{T^{\frac{5}{4}} I_k}{\rho} - 1 \right) \mathbb{E}_{\zeta_{[k]}} [\|z^{k+1} - z^k\|^2] \\ & \leq \mathbb{E}_{\zeta_{[T]}} \left[\frac{\|x^1 - x^*\|^2}{2\alpha} + L \sum_{k=1}^T \|x^k - x^*\|^2 + \right. \\ & \quad \left. \frac{T}{2\rho} \|z^1 - z^*\|^2 \right] + T\alpha (\sigma^2 + G^2 + 2T^{\frac{1}{2}} F^2 G^2 + \\ & \quad 8G^2 F^2 \rho^2) + \sum_{k=1}^T \mathbb{E}_{\zeta_{[k]}} [\chi_0(x^{k+1}) - \chi_0(x^k)] \\ & \leq \left[\frac{C\|x^1 - x^*\|^2}{2} + \frac{\sigma^2 + G^2 + 8G^2 F^2 \rho^2}{C} \right] T^{\frac{1}{2}} + \\ & \quad \left[LD_X^2 + \frac{z^{*2}}{2\rho} + \frac{2F^2 G^2}{C} \right] T + \\ & \quad \mathbb{E}_{\zeta_{[T]}} [\chi_0(x^1) - \chi_0(x^{T+1})]. \end{aligned}$$

其中: $D_X < +\infty$ 。已知 $\rho < \frac{1}{2}$, 所以 $1 < \frac{1}{2\rho} \leq$

$\frac{T^{\frac{5}{4}} I_k}{2\rho}$, 因而

$$\begin{aligned} & \sum_{k=1}^T \frac{T^{\frac{5}{4}} I_k}{2\rho} \mathbb{E}_{\zeta_{[k]}} [\|z^{k+1} - z^k\|^2] \\ & \leq \sum_{k=1}^T \left(\frac{T^{\frac{5}{4}} I_k}{\rho} - 1 \right) \mathbb{E}_{\zeta_{[k]}} [\|z^{k+1} - z^k\|^2] \\ & \leq 2\rho \left[\frac{C\|x^1 - x^*\|^2}{2} + \frac{\sigma^2 + G^2 + 8G^2 F^2 \rho^2}{C} \right] T^{-\frac{1}{2}} + \\ & \quad 2\rho \left[LD_X^2 + \frac{\|z^*\|^2}{2\rho} + \frac{2F^2 G^2}{C} \right] + \\ & \quad 2\rho [\chi_0(x^1) - W]. \end{aligned}$$

最后证明在 $\mathcal{O}(\epsilon^{-8})$ 次迭代后, 算法输出 x^{R+1} 在期望意义下满足式 (13)。

定理 3.3 在假设 2.1~假设 2.3 下, 按照引理 3.13 设定参数 $\alpha_k, \rho_k, \beta, I_k$ 。则对任意 $T \geq 1$, 下列等式成立:

$$\mathbb{E}_{R, \zeta_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M [f_i(x^{R+1})]_+ \right] = \mathcal{O}(T^{-\frac{1}{8}}).$$

证明 首先, 对任意实数 $u \in \mathbb{R}_+, v \in \mathbb{R}$, $[v]_+ \leq |\max\{-u, v\}|$ 。因此由引理 3.1 可知, 对任意分量 $i \in [M]$ 以及任意 $k \geq 1$, 成立

$$[f_i(x^{k+1})]_+ \leq \left| \max \left\{ -\frac{z_i^{k+1}}{\beta}, f_i(x^{k+1}) \right\} \right|$$

由 z^k 的更新公式 (6), 得

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{M} \| [f(x^{k+1})]_+ \| \mid \zeta_{[k-1]} \right] \\ & = \mathbb{E} \left[\frac{1}{M} \sum_{i=1}^M [f_i(x^{k+1})]_+^2 \mid \zeta_{[k-1]} \right]^{\frac{1}{2}} \\ & \leq \mathbb{E} \left[\frac{1}{M} \left[\sum_{i=1}^M \left| \max \left\{ -\frac{z_i^{k+1}}{\beta}, f_i(x^{k+1}) \right\} \right|^2 \right] \mid \zeta_{[k-1]} \right]^{\frac{1}{2}} \\ & = \frac{1}{M} E \left[\left(\mathbb{E} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \mid \zeta_{[k]} \right] \right)^{\frac{1}{2}} \mid \zeta_{[k-1]} \right] \\ & \leq \frac{1}{M} \left(\mathbb{E} \left[\mathbb{E} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \mid \zeta_{[k]} \right] \mid \zeta_{[k-1]} \right] \right)^{\frac{1}{2}} \\ & = \frac{1}{M} \left(\mathbb{E}_{\zeta_k, \zeta_{k+1}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \mid \zeta_{[k-1]} \right] \right)^{\frac{1}{2}}. \end{aligned}$$

第 2 个不等号因为对任意随机变量 $u \in \mathbb{R}$, $(\mathbb{E}(u))^2 \leq \mathbb{E}(u^2)$ 。再次利用 $(\mathbb{E}(u))^2 \leq \mathbb{E}(u^2)$, 对上述不等式两端取期望 $\mathbb{E}_{\zeta_{[k-1]}}$ 可得

$$\begin{aligned} & \mathbb{E}_{\zeta_{[k]}} \left[\frac{1}{M} \| [f(x^{k+1})]_+ \| \right] \\ & \leq \frac{1}{M} \mathbb{E}_{\zeta_{[k-1]}} \left[\left(\mathbb{E}_{\zeta_k, \zeta_{k+1}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \mid \zeta_{[k-1]} \right] \right)^{\frac{1}{2}} \right] \\ & \leq \frac{1}{M} \left(\mathbb{E}_{\zeta_{[k-1]}} \left[\mathbb{E}_{\zeta_k, \zeta_{k+1}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \mid \zeta_{[k-1]} \right] \right] \right)^{\frac{1}{2}} \\ & = \frac{1}{M} \left(\mathbb{E}_{\zeta_{[k+1]}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \right] \right)^{\frac{1}{2}}. \end{aligned}$$

因为 $\sqrt{\cdot}$ 是凹函数, 且对任意 $z \in \mathbb{R}^M, \|z\|_1 \leq \sqrt{M} \|z\|$, 所以

$$\begin{aligned} & \mathbb{E}_{R, \zeta_{[T]}} \left[\frac{1}{M} \| [f(x^{R+1})]_+ \|_1 \right] \\ & \leq \sqrt{M} \mathbb{E}_{R, \zeta_{[T]}} \left[\frac{1}{M} \| [f(x^{R+1})]_+ \| \right] \\ & = \frac{\sqrt{M}}{T} \sum_{k=1}^T \mathbb{E}_{\zeta_{[k]}} \left[\frac{1}{M} \| [f(x^{k+1})]_+ \| \right] \\ & \leq \frac{\sqrt{M}}{M} \frac{1}{T} \sum_{k=1}^T \left(\mathbb{E}_{\zeta_{[k+1]}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \right] \right)^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} &\leq \frac{\sqrt{M}}{M} \left(\frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k+1]}} \left[\frac{M}{\rho_{k+1}^2 I_{k+1}} \|z^{k+2} - z^{k+1}\|^2 \right] \right)^{\frac{1}{2}} \\ &= \left(\frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k+1]}} \left[\frac{T^2 I_{k+1}}{\rho^2} \|z^{k+2} - z^{k+1}\|^2 \right] \right)^{\frac{1}{2}}. \end{aligned}$$

第 2 个等号代入了步长设定 $\rho_k = \frac{\rho}{T I_k}$ 。而根

据引理 3.13,可以得到

$$\begin{aligned} &\frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k+1]}} \left[\frac{T^2 I_{k+1}}{\rho^2} \|z^{k+2} - z^{k+1}\|^2 \right] \\ &\leq \frac{1}{T} \sum_{k=1}^{T+1} \mathbb{E}_{\xi_{[k]}} \left[\frac{T^2 I_k}{\rho^2} \|z^{k+1} - z^k\|^2 \right] \\ &\leq \frac{T}{\rho^2} \times 4\rho^2 \left[\frac{C \|x^1 - x^*\|^2}{2} + \frac{\sigma^2 + G^2 + 8G^2 F^2 \rho^2}{C} \right] (T+1)^{-\frac{7}{4}} + \\ &\quad \frac{T}{\rho^2} \times 4\rho^2 \left[LD_X^2 + \frac{\|z^*\|^2}{2\rho} + \frac{2F^2 G^2}{C} \right] (T+1)^{-\frac{5}{4}} + \\ &\quad \frac{T}{\rho^2} \times 4\rho^2 [\chi_0(x^1) - W] (T+1)^{-\frac{5}{4}} \\ &\leq 4T^{-\frac{3}{4}} \left[\frac{C \|x^1 - x^*\|^2}{2} + \frac{\sigma^2 + G^2 + 8G^2 F^2 \rho^2}{C} \right] + \\ &\quad 4T^{-\frac{1}{4}} \left[LD_X^2 + \frac{\|z^*\|^2}{2\rho} + \frac{2F^2 G^2}{C} \right] + \\ &\quad 4T^{-\frac{1}{4}} [\chi_0(x^1) - W]. \end{aligned}$$

结合上述 2 组不等式,利用对任意 $a, b \geq 0$, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, 可推知定理成立。□

3.4 互补松弛条件

这一节将证明,在 $\mathcal{O}(\epsilon^{-4})$ 次迭代后,点 (22) 满足式 (14)。

定理 3.4 在定理 3.1 的条件下,对任意 $T \geq 1$, 下列不等式成立:

$$\begin{aligned} &\mathbb{E}_{R, \xi_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M (\beta f_i(x^{R+1}) + z_i^{R+1})_+ [f_i(x^{R+1})]_- \right] \\ &= \mathcal{O}(T^{-\frac{1}{4}}). \end{aligned}$$

证明 由定义可得

$$\begin{aligned} &\frac{|z_i^k|^2}{4\beta} \geq (\beta f_i(x^k) + z_i^k)_+ [f_i(x^k)]_- \\ &= \begin{cases} -(\beta f_i(x^k) + z_i^k) f_i(x^k), & 0 > f_i(x^k) \geq -\frac{z_i^k}{\beta}, \\ 0, & \text{else.} \end{cases} \end{aligned}$$

其中不等式利用了 $-u(bu + a) \leq \frac{a^2}{4b}, \forall u \in$

$\mathbb{R}, b \in \mathbb{R}_{++}$ 。类似于引理 3.3, 可得对任意分量 $i \in [M]$ 以及任意 $k \in [T+1]$, $|z_i^k| \leq F\rho$ 。进而可得

$$\begin{aligned} &\sum_{i=1}^M \mathbb{E}_{\xi_{[k-1]}} [(\beta f_i(x^k) + z_i^k)_+ [f_i(x^k)]_-] \\ &\leq \sum_{i=1}^M \mathbb{E}_{\xi_{[k-1]}} \left[\frac{|z_i^k|^2}{4\beta} \right] \leq \frac{MF^2 \rho^2}{4\beta}. \end{aligned}$$

因此,根据 $\beta = T^{\frac{1}{4}}$

$$\begin{aligned} &\mathbb{E}_{R, \xi_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M (\beta f_i(x^{R+1}) + z_i^{R+1})_+ [f_i(x^{R+1})]_- \right] \\ &= \frac{1}{T} \sum_{k=1}^T \mathbb{E}_{\xi_{[k]}} \left[\frac{1}{M} \sum_{i=1}^M (\beta f_i(x^{k+1}) + z_i^{k+1})_+ [f_i(x^{k+1})]_- \right] \\ &\leq \frac{1}{T} \sum_{k=1}^T \frac{F^2 \rho^2}{4\beta} = \frac{F^2 \rho^2}{4\beta} = \mathcal{O}(T^{-\frac{1}{4}}). \end{aligned}$$

定理得证。□

基于定理 3.1、定理 3.3 和定理 3.4, 可以得到算法 CSALM 的迭代复杂度如下。

引理 3.14 在引理 3.13 的条件下,对任意精度 $\epsilon \in (0, 1)$, 算法 CSALM 在迭代 $T = \mathcal{O}(\epsilon^{-8})$ 次后,点 (22) 是问题 (1) 的 ϵ -KKT 点

证明 利用定理 3.1、定理 3.3、定理 3.4 和 $T = \mathcal{O}(\epsilon^{-8})$, 可得到

$$\begin{aligned} &\mathbb{E}_{R, \xi_{[T]}} [\mathbf{d}(\nabla f_0(x^{R+1}) + \partial \chi_0(x^{R+1}) + \\ &\quad \sum_{i=1}^M z_i \nabla f_i(x^{R+1}) + \mathcal{N}_X(x^{R+1}), 0)] \leq \epsilon, \end{aligned}$$

$$\mathbb{E}_{R, \xi_{[T]}} \left[\frac{1}{M} \sum_{i=1}^M [f_i(x^{R+1})]_+ \right] \leq \epsilon,$$

$$\mathbb{E}_{R, \xi_{[T]}} \left[\sum_{i=1}^M z_i [f_i(x^{R+1})]_- \right] \leq \epsilon^2 \leq \epsilon.$$

其中: $z = \frac{1}{M}(\beta f(x^{R+1}) + z^{R+1})_+ \in \mathbb{R}_+^M, f = (f_1, \dots,$

$f_M)$ 。对照定义 3.1 可知结论成立。□

4 数值实验

为检验算法的实际表现,本节测试 CSALM 算法在多分类 NPC 问题 (multi-class neyman pearson classification, mNPC) 上的数值表现。mNPC 旨在最小化一类的分类损失,同时控制其他类的分类损失。设训练集有 M 个类,每类所有样本记为 \mathcal{D}_m , 每类对应参数 $x_m \in \mathbb{R}^d, m = 1, \dots, M$, 对未知类别的 ξ , 预测分类的方式为 $\arg\max_m x_m^T \xi, m = 1, 2, \dots, M$, 若要第 m 类预测精度较高,则

要求 $x_m^T \xi - x_l^T \xi, \forall m \neq l, \xi \in \mathcal{D}_m$ 的值尽可能大。因为所有类别同时极大化是矛盾的,所以 mNPC 问题取一样本类为目标,其他类别构建约束,以非降函数 ϕ 作为损失,再加入 L1 正则项 $\lambda \|x\|_1$, 最终带正则的 mNPC 问题表示为

$$\begin{aligned} \min_{\|x_m\| \leq \lambda, m=1, \dots, M} & \frac{1}{|\mathcal{D}_1|} \sum_{l>1} \sum_{\xi \in \mathcal{D}_1} \phi(x_l^T \xi - x_1^T \xi) + \\ & \lambda \|x\|_1, \\ \text{s. t. } & \frac{1}{|\mathcal{D}_m|} \sum_{l \neq m} \sum_{\xi \in \mathcal{D}_m} \phi(x_m^T \xi - x_l^T \xi) \leq r_m, \\ & m = 2, 3, \dots, M. \end{aligned}$$

其中: r_m 是其他约束的控制损失。

本文的问题设定是约束个数 M 极大,但常见的机器学习数据集类别数通常较少,而类别较多的数据集的单类规模过大如 imagenet, 所以使用 LIBSVM 的数据集 mnist、covtype 作为测试数据集,数据集的类数分别为 $M = 10, M = 7$ 。mnist 数据集使用数字 1 的类别作为目标, covtype 数据集使用标签为 1 的类别作为目标。所有数据集均使用 $\phi(x) = 1/(1 + \exp(x))$ 作为损失函数,并取 $r_m = 0.5(M - 1), \forall m = 2, \dots, M$ 作为误差界。

本文使用 ICPP 算法^[12]作为对比算法。mNPC 问题的目标函数和约束函数均可视作离散的期望函数,对目标和约束均为期望函数的优化问题,如果对目标和约束都采用随机信息,ICPP 算法能以 $\mathcal{O}(\epsilon^{-6})$ 的迭代复杂度求解,若约束是一般函数无期望函数结构,在使用随机目标信息和确定约束信息时,ICPP 算法能以 $\mathcal{O}(\epsilon^{-4})$ 的迭代复杂度求解。因为本文的研究问题设定约束为一般函数,故将与采用随机目标信息、确定约束信息的 ICPP 算法对比。

实验的具体设定因数据集和算法的理论设定有所不同。执行 CSALM 算法时,因为 M 极小,所以实际执行时,每次迭代使用的约束数量一般取 1~3,而非设定的按迭代次数增长 $I_k = \min(\lceil k^{3/4} \rceil, M)$,另外实际使用的算法步长 α_k 为下降步长而非固定步长,这是为了算法执行后期的数值稳定。具体而言,对 mnist 数据集:CSALM 算法的设定为迭代次数 $T = 500$,步长 $\alpha_k = 0.05/k^{1/4}$,罚参 $\beta = 5$,正则项 $\lambda = 0.05$,乘子步长 $\rho = 0.1$,目标函数的批量大小 $J_k = \lceil k^{1/4} \rceil$,约束个数 $I_k = 1$;对 covtype 数据集:迭代次数 $T = 1\,000$,步长 $\alpha_k = 0.01/k^{1/4}$,罚参 $\beta = 50$,正则项 $\lambda = 0.01$,乘子步长 $\rho = 0.001$,目标函数的批量大小 $J_k =$

$\lceil k^{1/4} \rceil$,约束个数 $I_k = 1$ 。

执行 ICPP 时,算法设定不变,循环输出略有不同。ICPP 算法是内外循环算法,在原算法框架中,内循环的输出点是所有内循环点的加权平均,实现时直接采用内循环的最后一次迭代点作为循环输出。具体的算法设定上,对 mnist 数据集:外循环 $K = 10$,内循环 $T = 26$,正则项 $\lambda = 0.05$,每次均使用所有约束的信息 $M = 9, t_0 = 2$,强凸项参数 $\mu_0 = 1.5$,连续性参数 $M_f = M/2$;对 covtype 数据集:外循环 $K = 10$,内循环 $T = 45$,正则项 $\lambda = 0.05$,每次均使用所有约束的信息 $M = 6, t_0 = 2$,强凸项参数 $\mu_0 = 1.5$,连续性参数 $M_f = M/2$ 。

实验前,数据集均进行了预处理。对 mnist,因为样本是灰度图像,特征全为 0~255 的正整数,故所有特征按特征归一化到 $[0, 1]$ 区间;对 covtype:前 10 种特征中除第 5 种特征外按特征归一化到 $[0, 1]$ 区间,第 5 种特征因为含负样本故归一化到 $[-1, 1]$ 区间,其他特征不变。实验时对所有算法和数据集,算法乘子初始化为 0,实验的初始点均为 $x_0 = 0$,所有实验均运行在 Windows 10, CPU intel i7-11700,内存 128 GB 的机器上, matlab 版本为 R2022a。CSALM 和 ICPP 算法均在 mnist、covtype 数据集上各运行了 10 次,取数值平均后实验结果如下。

所有实验中计算时间单位为 s,使用 tic、toc 工具测量。在函数损失实验结果中,函数指 $\phi(x)$ 不包括正则项,在不可行性实验结果中,不可行性是所有违反约束的约束违反值的算数平均。约束不可行性数据有处理,因为在对数图中 0 无法表示,为有效显示,将 mnist 的不可行性结果中小于 10^{-7} 的值替换为 10^{-7} ,将 ICPP 的不可行性结果中小于 10^{-7} 的结果替换为 10^{-7} 。ICPP 算法的数值结果仅记录外循环点。

由图 1~图 4 的结果可以看出,主要特点是 CSALM 下降快,ICPP 不可行性低。对 mnist 数据集,函数值方面 CSALM 算法初期下降显著快于 ICPP,但最终值二者接近,不可行性方面 CSALM 和 ICPP 整体接近且均有可行点出现,但整个过程 CSALM 的波动较大,原因可能是每次只抽样 1 个约束太少。对 covtype 数据集,CSALM 显著快于 ICPP,约束不可行性则显著高于 ICPP,且两算法均没有可行点,原因可能是函数值快速下降和约束不可行性低值稳定在本次实验里是相背的要求,而 ICPP 在 covtype 上的参数设置得偏向于控

制不可行性,最终导致 ICPP 下降慢于 CSALM。

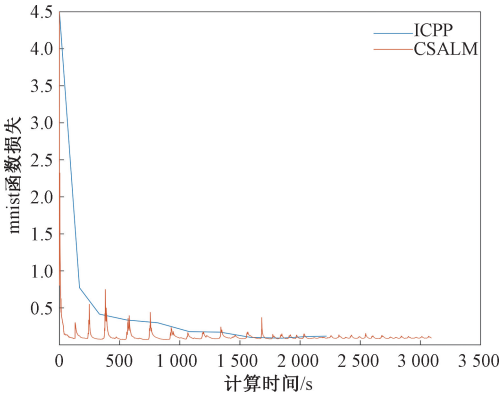


图 1 mnist 损失函数值

Fig. 1 Value of mnist loss function

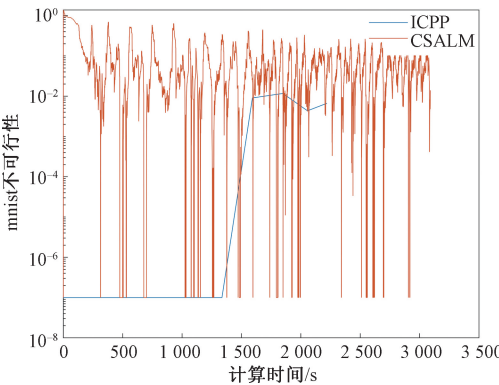


图 2 mnist 约束违反度

Fig. 2 Infeasibility of mnist

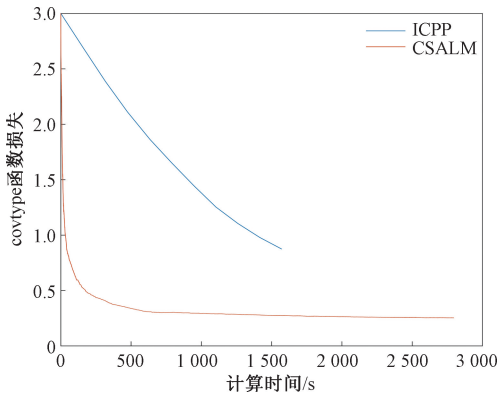


图 3 covtype 损失函数值

Fig. 3 Value of covtype loss function

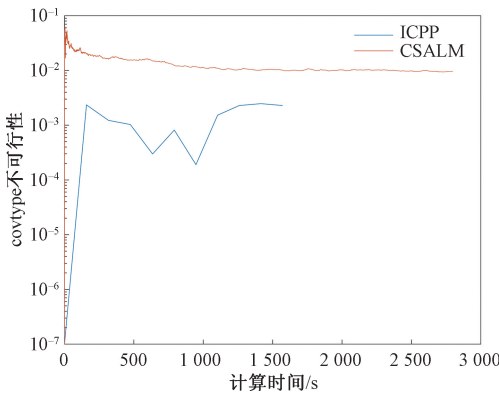


图 4 covtype 不可行性

Fig. 4 Infeasibility of covtype

设下,证明该算法可从任意初始可行点出发找到 ϵ -KKT 近似点。多分类 Neyman-Pearson 问题的数值实验表明该算法可以在实际问题中获得优秀的数值表现。

参考文献

[1] Li S Y, Sun Y N, Yen G G, et al. Automatic design of convolutional neural network architectures under resource constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021: 1-15. DOI: 10.1109/TNNLS.2021.3123105.

[2] Campi M C, Garatti S. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality [J]. Journal of Optimization Theory and Applications, 2011, 148(2): 257-280. DOI: 10.1007/s10957-010-9754-6.

[3] Calafiore G, Campi M C. Uncertain convex programs: Randomized solutions and confidence levels [J]. Mathematical Programming, 2005, 102(1): 25-46. DOI: 10.1007/s10107-003-0499-y.

[4] Seri R, Choirat C. Scenario approximation of robust and chance-constrained programs [J]. Journal of Optimization Theory and Applications, 2013, 158(2): 590-614. DOI: 10.1007/s10957-012-0230-3.

[5] Wang M D, Bertsekas D P. Stochastic first-order methods with random constraint projection [J]. SIAM Journal on Optimization, 2016, 26(1): 681-717. DOI: 10.1137/130931278.

[6] Wang M D, Chen Y C, Liu J L, et al. Random multi-constraint projection: stochastic gradient methods for convex optimization with many constraints[EB/OL]. 2015; arXiv: 1511.03760. (2015-11-12) [2023-01-16]. <https://arxiv.org/abs/1511.03760>.

[7] Lan G H, Zhou Z Q. Algorithms for stochastic optimization with function or expectation constraints [J]. Computational Optimization and Applications, 2020, 76(2): 461-498. DOI: 10.1007/s10589-020-00179-x.

[8] Nocedal J, Wright S J. Numerical optimization; springer

5 结论

针对大规模约束的非凸非光滑问题,本文提出一个基于增广拉格朗日的随机算法,该算法每步使用随机信息进行迭代更新,使得计算量与问题规模解耦。在约束凸光滑且 Lipschitz 光滑的假

- series in operations research and financial engineering[M]. New York, NY: Springer New York, 2006: 529-562.
- [9] Lin Q H, Nadarajah S, Soheili N. A level-set method for convex optimization with a feasible solution path[J]. SIAM Journal on Optimization, 2018, 28(4): 3290-3311. DOI: 10.1137/17M1152334.
- [10] Lin Q H, Ma R C, Yang T B. Level-set methods for finite-sum constrained convex optimization[C/OL]//Proceedings of the 35th International Conference on Machine Learning: Vol 80. PMLR, 2018: 3112-3121. <https://proceedings.mlr.press/v80/lin18c.html>.
- [11] Aravkin A Y, Burke J V, Drusvyatskiy D, et al. Level-set methods for convex optimization [J]. Mathematical Programming, 2019, 174(1): 359-390. DOI: 10.1007/s10107-018-1351-8.
- [12] Boob D, Deng Q, Lan G H. Stochastic first-order methods for convex and nonconvex functional constrained optimization[J]. Mathematical Programming, 2023, 197(1): 215-279. DOI: 10.1007/s10107-021-01742-y.
- [13] Xu Y Y. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming [J]. Mathematical Programming, 2021, 185(1): 199-244. DOI: 10.1007/s10107-019-01425-9.
- [14] Xu Y Y. Primal-dual stochastic gradient method for convex programs with many functional constraints[J]. SIAM Journal on Optimization, 2020, 30(2): 1664-1692. DOI: 10.1137/18M1229869.
- [15] Rockafellar R T. A dual approach to solving nonlinear programming problems by unconstrained optimization [J]. Mathematical Programming, 1973, 5(1): 354-373. DOI: 10.1007/BF01580138.
- [16] Rockafellar R T. Augmented lagrangians and applications of the proximal point algorithm in convex programming [J]. Mathematics of Operations Research, 1976, 1(2): 97-116. DOI: 10.1287/moor.1.2.97.
- [17] Kingma D, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2014-12-22) [2023-01-16]. <http://arxiv.org/abs/1412.6980>.
- [18] Defazio A, Bach F, Lacoste-Julien S. SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol 1. December 8-13, 2014, Montreal, Canada. New York. ACM, 2014: 1646-1654. DOI: 10.5555/2968826.2969010.