非凸优化与数值算法专题

DOI:10.11721/cqnuj20260208

非凸复合优化的镜像惯性修正前向-反射-后向分裂算法*

传军霖，罗洪林
(重庆师范大学 数学科学学院，重庆 401331)

摘要:本文针对非凸复合优化问题,通过引入外推步,对镜像惯性前向-反射-后向分裂算法进行改进,提出了一种带有外推步的镜像惯性修正前向-反射-后向分裂算法(mirror inertial modified forward-reflected-backward method,i * MFRB)。推广了带外推步的前向-后向分裂算法在非凸情形下的应用,为前向-反射-后向分裂算法提供了一种更简洁的证明思路。通过构造适当且简单的价值函数,建立了算法函数值的收敛性;进一步结合 Kurdyka-Łojasiewicz 性质,证明了算法迭代序列的收敛性。将所提算法应用于一类非凸可行性问题,设计了对应的子问题求解方式。最后通过数值实验展示了 i * MFRB 分裂算法与现有其他类似算法在高维情况的比较结果。

关键词:非凸复合优化;Tseng 类邻近算法;前向-反射-后向分裂算法;Bregman 邻近映射;Kurdyka-Łojasiewicz 性质
中图分类号:O221.2　　**文献标志码**:A　　**文章编号**:1672-6693(2026)02-0095-12

非凸复合优化问题在机器学习、信号处理与图像重建等领域有着广泛的应用。该问题基本形式为:

$$\min_{x\in\mathbf{R}^n} f(x)+g(x), \tag{1}$$

其中:函数 $f:\mathbf{R}^n\to\mathbf{R}$ 为可微函数(不必凸),梯度具有 Lipschitz 连续性(Lipschitz 常数为 L_f);函数 g 为真下半连续函数。现有研究已发展出多种算法,如 Douglas-Rachford 分裂算法[1]、前向-反射-后向分裂算法[2]及该算法的 Bregman 惯性(Bregman inertial forward-reflected-backward,BiFRB)[3]或镜像惯性[4]推广形式、惯性 Tseng 类分裂算法(Inertial Tseng's type splitting algorithm,iTseng)[5]等。

其中,Tseng 类分裂算法也称为修正前向-后向分裂算法,在高维优化问题中展现出良好的优越性,但该类算法在非凸情况下的理论研究尚不完善。例如,Radu Ioan Boţ 与 Ernö Robert Csetnek 虽于 2016 年对 Tseng 类分裂算法进行了非凸推广,但未提出相应的子问题求解方式与具体数值应用。相比之下,BiFRB[3]及它的镜像惯性版本[4]在理论与数值运用方面的研究更加充分:前者通过数值实验验证了小规模问题的稳定收敛性,这恰是 Tseng 类分裂算法所欠缺的;后者放宽了问题(1)中目标函数对梯度 Lipschitz 连续性的严格要求,但对参数要求苛刻(如要求惯性系数为负数[4])。此外,前向-反射-后向分裂算法加速技巧(如外推步)在非凸框架下的潜力仍有待系统挖掘。

因此,为弥补上述研究空白,本文提出了一种镜像惯性修正-前向-反射-后向分裂算法(mirror inertial forward-reflected-backward,i * FRB)用于求解问题(1),下称算法 1,计算框架可表述为:

$$\begin{aligned} x_{k+1}\in\ & (I-\lambda_k\nabla f)(\lambda_k\partial g+\nabla h)^{-1}(\nabla h(x_k)+\lambda_k\nabla f(x_{k-1})- \\ & 2\lambda_k\nabla f(x_k)+\alpha_k(\nabla h(x_k)-\nabla h(x_{k-1})))+\lambda_k\nabla f(x_k)。\end{aligned}$$

该算法设计思路是将 Tseng 类分裂算法的"外推步"嵌入镜像惯性前向-反射-后向分裂算法(mirror inertial modified forward-reflected-backward method,i * MFRB)。本文通过理论证明与数值实验验证了 i * MFRB 处理高维非凸问题(1)的有效性,从而首次系统地将外推步技术拓展至前向-反射-后向类分裂算法的应用框架。而理论方面,现有前向-反射-后向分裂算法[3-4]的价值函数形式隐式且严重依赖参数。针对这一问题,本文基于文

* **收稿日期**:2025-10-14　**修回日期**:2025-12-15　**网络出版时间**:2026-04-10T18:55
资助项目:国家自然科学基金面上项目(No. 11771064);重庆市教育委员会科学技术研究计划重点项目(No. KJZD-K202500507)
第一作者简介:传军霖,男,研究方向为最优化理论与算法,E-mail:junlinchuan@stu.cqnu.edu.cn;**通信作者简介**:罗洪林,男,教授,博士,E-mail:luohonglin@cqnu.edu.cn
网络出版地址:https://link.cnki.net/urlid/50.1165.n.20260409.1406.002

献[5]设计了一种不随参数迭代变化、显式且简洁的价值函数,简化了收敛证明,为此类分裂算法的收敛性分析提供了新思路。

本文后续内容安排如下:在第1节给出相关记号与预备知识、性质与引理前提;在第2节提出了算法框架,并进行算法收敛性分析;最后第3节针对一个非凸可行性问题设计了子问题求解过程,进行了数值实验。

1 记号与预备知识

1.1 基本记号

本文在标准欧几里得空间 $\mathbf{R}^n$ 中开展研究,该空间配备标准内积 $\langle x,y\rangle=x^{\mathrm{T}}y$ 和欧几里得范数 $\|x\|=\sqrt{\langle x,y\rangle}$,其中 $x,y\in\mathbf{R}^n$。如无特别说明,$\|\cdot\|$ 表示2范数。$\|\cdot\|_0=\sum_{i=1}^{n}|\mathrm{sgn}(\cdot^{(i)})|$,其中:$\cdot^{(i)}$ 记为向量 $\cdot$ 的第 i 个分量,$\mathrm{sgn}(\cdot)$ 为符号函数,满足 $|\mathrm{sgn}(\theta)|=\begin{cases}0, & \theta=0\\ 1, & \theta\neq 0\end{cases}$。$\overline{\mathbf{R}}=\mathbf{R}\cup\{+\infty\}$ 表示扩展实数集。对于扩展实值函数 $f:\mathbf{R}^n\to\overline{\mathbf{R}}$ 的定义域为:当 $\mathrm{dom}\ f\neq\varnothing$ 时,称 f 为真函数。

1.2 次微分相关理论

本节提供次微分相关理论[6-8],设 $f:\mathbf{R}^n\to(-\infty,+\infty]$ 为真下半连续函数,有以下结论:

1) Fréchet 次微分($x\in\mathrm{dom}\ f$):

$$\hat{\partial}f(x):=\left\{v\in\mathbf{R}^n:\liminf_{y\to x}\frac{f(y)-f(x)-\langle v,y-x\rangle}{\|y-x\|}\geqslant 0\right\},$$

对 $x\notin\mathrm{dom}\ f$,规定 $\hat{\partial}f(x):=\varnothing$。

2) 极限次微分($x\in\mathrm{dom}\ f$):

$$\partial f(x)=\{v\in\mathbf{R}^n:\exists\, x_k\to x, f(x_k)\to f(x)\ \&\ \exists\, v_k\in\hat{\partial}f(x_k), v_k\to v(k\to+\infty)\},$$

对 $x\notin\mathrm{dom}\ f$,规定 $\partial f(x):=\varnothing$。

3) 当 f 为凸函数时,上述次微分概念与凸次微分一致:

$$\hat{\partial}f(x)=\partial f(x)=\{v\in\mathbf{R}^n:f(y)\geqslant f(x)+\langle v,y-x\rangle,\forall\, y\in\mathbf{R}^n\}。$$

4) 包含关系:$\hat{\partial}f(x)\subseteq\partial f(x)$ 对所有 $x\in\mathbf{R}^n$ 成立。

5) 图像闭性:若序列 (x_k,v_k) 满足:

$$v_k\in\partial f(x_k),\text{且}(x_k,v_k)\to(x,v),f(x_k)\to f(x),$$

则 $v\in\partial f(x)$。

6) 费马法则:若 x 是 f 的局部极小点,则 $0\in\partial f(x)$。

7) 可微情形:若 f 在 x 处连续可微,则 $\partial f(x)=\{\nabla f(x)\}$。

8) 临界点集:$\mathrm{crit}(f):=\{x\in\mathbf{R}^n:0\in\partial f(x)\}$。

9) 次微分求和法则:若 g 为真下半连续函数,f 连续可微,则:

$$\partial(f+g)(x)=\partial g(x)+\nabla f(x)。$$

1.3 Kurdyka-Łojasiewicz 性质与 Bregman 距离

令 $\eta\in(0,+\infty]$,定义函数类:$\Theta_\eta:=\varphi:[0,\eta)\to[0,+\infty)$。其中:$\varphi$ 为凹连续函数,且有:

$$\varphi\in C^1(0,\eta),\varphi(0)=0,\varphi'(s)>0(\forall\, s\in(0,\eta))。$$

集合 $A\subset\mathbf{R}^n$ 对应的距离函数定义为:

$$\mathrm{dist}(x,A):=\inf_{y\in A}\|x-y\|,\forall\, x\in\mathbf{R}^n。$$

定义 1(Kurdyka-Łojasiewicz 性质)[9-10] 设 $f:\mathbf{R}^n\to(-\infty,+\infty]$ 为真下半连续函数。称 f 在点 $\overline{x}\in\mathrm{dom}\ \partial f$ 处满足 KL 性质,若存在 $\eta>0$,邻域 U 和函数 $\varphi\in\Theta_\eta$,使得对所有

$$x\in U\cap\{x\in\mathbf{R}^n:f(\overline{x})<f(x)<f(\overline{x})+\eta\}$$

都有:$\varphi'(f(x)-f(\overline{x}))\mathrm{dist}\ (0,\partial f(x))\geqslant 1$。

引理 1[10] 设 $\Omega\subseteq\mathbf{R}^n$ 为紧集,$f:\mathbf{R}^n\to(-\infty,+\infty]$ 为真下半连续函数。若 f 在 Ω 上为常数且满足 KL 性

质，则存在 $\varepsilon,\eta>0$ 和 $\varphi\in\Theta_\eta$，使得对所有

$$\overline{x}\in\Omega$$

和

$$x\in\{x\in\mathbf{R}^n:\operatorname{dist}(x,\Omega)<\varepsilon\}\cap\{x\in\mathbf{R}^n:f(\overline{x})<f(x)<f(\overline{x})+\eta\},$$

都有：

$$\varphi'(f(x)-f(\overline{x}))\operatorname{dist}(0,\partial f(x))\geqslant 1。$$

满足 KL 性质的函数类极具研究价值，这类函数包含半代数函数、实次解析函数、半凸函数、一致凸函数以及满足增长条件的凸函数[9,11-14]。

由可微强凸函数 h 诱导的 Bregman 距离为：

$$D_h(x,y)=h(x)-h(y)-\langle\nabla h(y),x-y\rangle,$$

由此得下述结论。

引理 2[15]　若 h 为 σ_h-强凸函数且梯度 Lipschitz 连续（常数 $L_{\nabla h}$），则：

$$\frac{\sigma_h}{2}\|x-y\|^2\leqslant D_h(x,y)\leqslant\frac{L_{\nabla h}}{2}\|x-y\|^2\ \forall\, x,y\in\mathbf{R}^n。$$

1.4　收敛性引理

引理 3[15]　设 (ζ_k) 和 (ϑ_k) 为实序列，满足 $\vartheta_k\geqslant 0(\forall\, k\in\mathbf{N})$，$(\zeta_k)$ 有下界，且 $\zeta_{k+1}+\vartheta_k\leqslant\zeta_k$ 对 $\forall\, k\in\mathbf{N}$ 成立。则 (ζ_k) 单调递减收敛，且 $\sum\limits_{k=0}^{\infty}\vartheta_k<+\infty$。

引理 4[5]　设 (ξ_k) 和 (ε_k) 为非负序列，满足 $\sum\varepsilon_k<+\infty$ 且 $\xi_{k+1}\leqslant a\xi_k+b\xi_{k-1}+\varepsilon_k(k\geqslant 1)$，其中 $a\in\mathbf{R}$，$b\geqslant 0$，$a+b<1$，则 $\sum\xi_k<+\infty$。

引理 5（下降引理）[15]　若 f 梯度 Lipschitz 连续（常数 $L_{\nabla f}$），则：

$$f(y)\leqslant f(x)+\langle\nabla f(x),y-x\rangle+\frac{L_{\nabla f}}{2}\|y-x\|^2\ \forall\, x,y\in\mathbf{R}^n。$$

2　i * MFRB 分裂算法及其收敛性分析

本节介绍了 i * MFRB 分裂算法的框架，并给出了该算法在函数值收敛性与序列收敛性方面的证明。

2.1　i * MFRB 分裂算法

本文后续分析基于以下基本假设。

假设 1　1）函数 $f:\mathbf{R}^n\to\overline{\mathbf{R}}$ 具有 Lipschitz 连续梯度，常数为 $L_{\nabla f}$。

2）函数 $g:\mathbf{R}^n\to\mathbf{R}$ 为真下半连续函数。

3）函数 $h:\mathbf{R}^n\to\mathbf{R}$ 具有 σ_h-强凸性（参数 $\sigma_h>0$）且梯度 Lipschitz 连续（模 $L_{\nabla h}$）。

4）目标函数 $F=f+g$ 下有界。

在基础假设 1 下，本文提出以下 i * MFRB 分裂算法。

算法 1（i * MFRB）　选择 $x_0,x_1\in\mathbf{R}^n$，$p_0=x_0$，$\underline{\lambda},\overline{\lambda}>0$，$\alpha\geqslant 0$。以及序列 $(\lambda_k)_{k\geqslant 1}$，$(\alpha_k)_{k\geqslant 1}$，这些序列需满足：

$$0\leqslant\alpha_k\leqslant\alpha,0<\underline{\lambda}\leqslant\lambda_k\leqslant\overline{\lambda}\ \forall\, k\geqslant 1,$$

考虑如下迭代方案（$\forall\, k\geqslant 1$）：

步骤 1，设置 y_k，使得 $\nabla h(y_k)=\nabla h(x_k)-\lambda_k(\nabla f(x_k)-\nabla f(x_{k-1}))$，

步骤 2，选择

$$p_k\in\arg\min_{x\in\mathbf{R}^n}\left\{g(x)+\left\langle x,\nabla f(x_k)-\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))\right\rangle+\frac{1}{\lambda_k}D_h(x,y_k)\right\},\tag{2}$$

步骤 3，设置 $x_{k+1}=p_k+\lambda_k(\nabla f(x_k)-\nabla f(p_k))$。

2.2　i * MFRB 分裂算法的价值函数及其参数规则

在本文后续内容中，除非另有说明，记：

$$M_1=\frac{\sigma_h}{2\bar{\lambda}}-\frac{\bar{\lambda}L_{\nabla f}^2+3L_{\nabla f}}{2}-\nu-\frac{\alpha}{\underline{\lambda}}\mu,$$

$$M_2=\frac{L_{\nabla h}}{2\underline{\lambda}}\left[\left(\frac{\bar{\lambda}L_{\nabla f}}{\sigma_h}\right)(1+\bar{\lambda}L_{\nabla f})+(\bar{\lambda}L_{\nabla f})\right]^2+\frac{L_{\nabla f}}{2}(1+\bar{\lambda}L_{\nabla f})+$$

$$\bar{\lambda}^2L_{\nabla f}^2\left(\nu+\frac{L_{\nabla f}^2}{2\nu}+L_{\nabla f}\right)+\frac{\alpha}{\underline{\lambda}}\left[\mu\bar{\lambda}^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+\bar{\lambda}L_{\nabla f})^2\right],$$

其中:$\mu,\nu>0$。下面给出了一个不等式,这将在后续进一步分析中起作用。

命题 1 若假设 1 成立,考虑由算法 1 生成的序列(x_k,p_k)。则对任意 $\mu,\nu>0$,下列不等式成立:

$$(f+g)(p_k)+M_1\|x_k-p_k\|^2\leqslant(f+g)(p_{k-1})+M_2\|x_{k-1}-p_{k-1}\|^2。$$

证明 由算法 1 步骤 2,有:

$$g(p_k)+\frac{1}{\lambda_k}D_h(p_k,y_k)+\langle p_k,\nabla f(x_k)\rangle+\frac{\alpha_k}{\lambda_k}\langle p_k,\nabla h(x_{k-1})-\nabla h(x_k)\rangle\leqslant$$

$$g(p_{k-1})+\frac{1}{\lambda_k}D_h(p_{k-1},y_k)+\langle p_{k-1},\nabla f(x_k)\rangle+\frac{\alpha_k}{\lambda_k}\langle p_{k-1},\nabla h(x_{k-1})-\nabla h(x_k)\rangle。\tag{3}$$

由引理 5,有:

$$f(p_k)\leqslant f(p_{k-1})+\langle\nabla f(p_{k-1}),p_k-p_{k-1}\rangle+\frac{L_{\nabla f}}{2}\|p_k-p_{k-1}\|^2,\tag{4}$$

将式(3)与(4)相加,得:

$$(f+g)(p_k)\leqslant(f+g)(p_{k-1})+\frac{1}{\lambda_k}(D_h(p_{k-1},y_k)-D_h(p_k,y_k))+\frac{L_{\nabla f}}{2}\|p_k-p_{k-1}\|^2+$$

$$\langle\nabla f(p_{k-1})-\nabla f(x_k),p_k-p_{k-1}\rangle+\frac{\alpha_k}{\lambda_k}\langle p_{k-1}-p_k,\nabla h(x_{k-1})-\nabla h(x_k)\rangle。\tag{5}$$

通过引理 2 结合函数 h 在假设 1 的 3)中的描述,得:

$$D_h(p_{k-1},y_k)-D_h(p_k,y_k)\leqslant\frac{L_{\nabla h}}{2}\|p_{k-1}-y_k\|^2-\frac{\sigma_h}{2}\|p_k-y_k\|^2,\tag{6}$$

又由算法 1 步骤 1,有:

$$\|x_k-y_k\|=\|\nabla h^*(\nabla h(x_k)-\lambda_k(\nabla f(x_k)-\nabla f(x_{k-1})))-\nabla h^*(\nabla h(x_k))\|\leqslant$$

$$\frac{\lambda_kL_{\nabla f}}{\sigma_h}\|x_k-x_{k-1}\|,\tag{7}$$

以及步骤 3,可得:

$$\|x_k-p_{k-1}\|=\lambda_{k-1}\|\nabla f(x_{k-1})-\nabla f(p_{k-1})\|\leqslant\lambda_{k-1}L_{\nabla f}\|x_{k-1}-p_{k-1}\|,\tag{8}$$

由此,有:

$$\|x_k-x_{k-1}\|\leqslant(1+\lambda_{k-1}L_{\nabla f})\|x_{k-1}-p_{k-1}\|,\tag{9}$$

且有:

$$\|p_k-p_{k-1}\|^2\leqslant2(\|x_k-p_k\|^2+\lambda_{k-1}^2L_{\nabla f}^2\|x_{k-1}-p_{k-1}\|^2),\tag{10}$$

依次运用式(7)、(9),可得:

$$\|x_k-y_k\|\leqslant\frac{\lambda_kL_{\nabla f}}{\sigma_h}\|x_k-x_{k-1}\|\leqslant\frac{\lambda_kL_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})\|x_{k-1}-p_{k-1}\|,\tag{11}$$

此时利用式(8)、(11)且皆转化为以$\|x_{k-1}-p_{k-1}\|^2$ 项后,将系数写成完全平方形式可以得到:

$$\|p_{k-1}-y_k\|^2\leqslant\|x_k-y_k\|^2+\|x_k-p_{k-1}\|^2+2\|x_k-y_k\|\|x_k-p_{k-1}\|=$$

$$\left[\frac{\lambda_kL_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})+(\lambda_{k-1}L_{\nabla f})\right]^2\|x_{k-1}-p_{k-1}\|^2。\tag{12}$$

对于$\|p_k-y_k\|^2$,有:

$$\|p_k-y_k\|^2\geqslant(\|p_k-x_k\|-\|x_k-y_k\|)^2\geqslant$$

$$\|p_k-x_k\|^2+\|x_k-y_k\|^2-2\|p_k-x_k\|\|x_k-y_k\|。\tag{13}$$

由式(11)以及基本不等式有：

$$2\|p_k-x_k\|\|x_k-y_k\|\leqslant\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})(\|x_{k-1}-p_{k-1}\|^2+\|p_k-x_k\|^2),\tag{14}$$

此时结合式(13)、(14)并利用$\|x_k-y_k\|^2\geqslant 0$，有：

$$\|p_k-y_k\|^2\geqslant\|p_k-x_k\|^2+\|x_k-y_k\|^2-\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})(\|x_{k-1}-p_{k-1}\|^2+\|p_k-x_k\|^2)\geqslant$$

$$\left[1-\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})\right]\|p_k-x_k\|^2-\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})\|x_{k-1}-p_{k-1}\|^2。\tag{15}$$

最后结合式(6)、(12)和(15)，有：

$$D_h(p_{k-1},y_k)-D_h(p_k,y_k)\leqslant\frac{L_{\nabla h}}{2}\left[\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})+(\lambda_{k-1}L_{\nabla f})\right]^2\|x_{k-1}-p_{k-1}\|^2-$$

$$\frac{\sigma_h}{2}\left[1-\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})\right]\|x_k-p_k\|^2+\frac{\sigma_h}{2}\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})\|x_{k-1}-p_{k-1}\|^2。\tag{16}$$

此外，由 Cauthy-Schwarz 不等式与杨氏不等式可以得到存在 $\mu>0,\nu>0$，以下不等式成立：

$$\langle\nabla f(p_{k-1})-\nabla f(x_k),p_k-p_{k-1}\rangle\leqslant\frac{\nu}{2}\|p_k-p_{k-1}\|^2+\frac{L_{\nabla f}^2}{2\nu}\|x_k-p_{k-1}\|^2,$$

$$\langle p_k-p_{k-1},\nabla h(x_{k-1})-\nabla h(x_k)\rangle\leqslant\frac{\mu}{2}\|p_k-p_{k-1}\|^2+\frac{L_{\nabla h}^2}{2\mu}\|x_k-x_{k-1}\|^2。$$

分别结合式(8)、(10)与式(9)、(10)，有：

$$\langle\nabla f(p_{k-1})-\nabla f(x_k),p_k-p_{k-1}\rangle\leqslant\nu\|x_k-p_k\|^2+\left(\nu\lambda_{k-1}^2L_{\nabla f}^2+\frac{\lambda_{k-1}^2L_{\nabla f}^4}{2\nu}\right)\|x_{k-1}-p_{k-1}\|^2,\tag{17}$$

$$\langle p_k-p_{k-1},\nabla h(x_{k-1})-\nabla h(x_k)\rangle\leqslant\left[\mu\lambda_{k-1}^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+\lambda_{k-1}L_{\nabla f})^2\right]\|x_{k-1}-p_{k-1}\|^2+\mu\|x_k-p_k\|^2。\tag{18}$$

最后，由式(5)、(10)、(16)、(17)和(18)，重新整理移项后可以得到：

$$(f+g)(p_k)+M_{1,k}\|x_k-p_k\|^2\leqslant(f+g)(p_{k-1})+M_{2,k}\|x_{k-1}-p_{k-1}\|^2,$$

其中：

$$M_{1,k}=\frac{\sigma_h}{2\lambda_k}-\frac{\lambda_{k-1}L_{\nabla f}^2+3L_{\nabla f}}{2}-\nu-\frac{\alpha_k}{\lambda_k}\mu,$$

$$M_{2,k}=\frac{L_{\nabla h}}{2\lambda_k}\left[\frac{\lambda_k L_{\nabla f}}{\sigma_h}(1+\lambda_{k-1}L_{\nabla f})+(\lambda_{k-1}L_{\nabla f})\right]^2+\frac{L_{\nabla f}}{2}(1+\lambda_{k-1}L_{\nabla f})+$$

$$\lambda_{k-1}^2L_{\nabla f}^2\left(\nu+\frac{L_{\nabla f}^2}{2\nu}+L_{\nabla f}\right)+\frac{\alpha_k}{\lambda_k}\left[\mu\lambda_{k-1}^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+\lambda_{k-1}L_{\nabla f})^2\right]。$$

由 $\underline{\lambda}\leqslant\lambda\leqslant\overline{\lambda}$，有 $M_{1,k}\geqslant M_1,M_{2,k}\leqslant M_2$。　　证毕

定义 2(i * MFRB 价值函数)　本文引入 i * MFRB 价值函数 $H:\mathbf{R}^n\times\mathbf{R}^n\to\overline{\mathbf{R}}$，定义为：

$$H(x,y)=(f+g)(x)+M_2\|x-y\|^2\ \forall(x,y)\in\mathbf{R}^n\times\mathbf{R}^n。$$

可以看出，当 x 与 y 趋于相等的有界点时，$H(x,y)$与$(f+g)(x)$会趋于相同的函数值。通过一直满足的 $H(x,y)\geqslant(f+g)(x)$，H 的下降性与收敛性一定程度上描述了 $f+g$ 的收敛情况。因此后续可以通过研究 H 相关性质得到 $f+g$ 的收敛情况。

2.3　函数值收敛性

为了保证价值函数下降性，接下来给出所需的步长规则。

命题 2(步长规则)　若假设 1 成立，考虑任意 $\nu,\mu>0$ 并选择 $\alpha\geqslant 0$，使得：

$$2\alpha\left(\mu+\frac{L_{\nabla h}^2}{2\mu}\right)<\sigma_h,$$

则存在 $\overline{\lambda}>\underline{\lambda}$，使得命题 1 中引入的常数满足 $M_1>M_2$。

证明　由条件 $2\alpha(\mu+L_{\nabla h}^2/2\mu)<\sigma_h$，必存在足够小的 $\underline{\lambda}>0$ 使得：

$$\left(L_{\nabla h}\left[\frac{\underline{\lambda}L_{\nabla f}}{\sigma_h}(1+\underline{\lambda}L_{\nabla f})+(\underline{\lambda}L_{\nabla f})\right]^2+L_{\nabla f}\underline{\lambda}(1+\underline{\lambda}L_{\nabla f})+2\underline{\lambda}^3L_{\nabla f}^2\left(\nu+\frac{L_{\nabla f}^2}{2\nu}+L_{\nabla f}\right)+\right.$$
$$\left.2\alpha\left[\mu\underline{\lambda}^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+\underline{\lambda}L_{\nabla f})^2\right]+\underline{\lambda}(\underline{\lambda}L_{\nabla f}^2+3L_{\nabla f})+2\underline{\lambda}\nu+2\alpha\mu\right)<\sigma_h。\tag{19}$$

式(19)可等价地写为:

$$2\underline{\lambda}\left(\frac{L_{\nabla h}}{2\underline{\lambda}}\left[\frac{\underline{\lambda}L_{\nabla f}}{\sigma_h}(1+\underline{\lambda}L_{\nabla f})+(\underline{\lambda}L_{\nabla f})\right]^2+\frac{L_{\nabla f}}{2}(1+\underline{\lambda}L_{\nabla f})+\lambda^2L_{\nabla f}^2\left(\nu+\frac{L_{\nabla f}^2}{2\nu}+L_{\nabla f}\right)+\right.$$
$$\left.\frac{\alpha}{\underline{\lambda}}\left[\mu\underline{\lambda}^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+\underline{\lambda}L_{\nabla f})^2\right]+\frac{(\underline{\lambda}L_{\nabla f}^2+3L_{\nabla f})}{2}+\nu+\frac{\alpha}{\underline{\lambda}}\mu\right)<\sigma_h。$$

因此,存在 $\beta>0$ 使得:

$$2(\underline{\lambda}+\beta)\left(\frac{L_{\nabla h}}{2\underline{\lambda}}\left[\left(\frac{(\underline{\lambda}+\beta)L_{\nabla f}}{\sigma_h}\right)(1+(\underline{\lambda}+\beta)L_{\nabla f})+((\underline{\lambda}+\beta)L_{\nabla f})\right]^2+\frac{L_{\nabla f}}{2}(1+(\underline{\lambda}+\beta)L_{\nabla f})+\right.$$
$$(\underline{\lambda}+\beta)^2L_{\nabla f}^2\left(\nu+\frac{L_{\nabla f}^2}{2\nu}+L_{\nabla f}\right)+\frac{\alpha}{\underline{\lambda}}\left[\mu(\underline{\lambda}+\beta)^2L_{\nabla f}^2+\frac{L_{\nabla h}^2}{2\mu}(1+(\underline{\lambda}+\beta)L_{\nabla f})^2\right]+$$
$$\left.\frac{(\underline{\lambda}+\beta)L_{\nabla f}^2+3L_{\nabla f}}{2}+\nu+\frac{\alpha}{\underline{\lambda}}\mu\right)<\sigma_h。$$

现在定义 $\bar{\lambda}:=\underline{\lambda}+\beta$,由上述不等式可直接得出关系 $M_1>M_2$。 证毕

现在基于步长规则命题2,给出价值函数下降性质与一些其他相关性质,为后续分析做铺垫。

命题3(价值函数下降性质) 若假设1成立,考虑算法1生成的序列(x_k,p_k),其中 $\nu,\mu,\underline{\lambda},\bar{\lambda}$,和 α 按命题2选择。则以下陈述成立:

1) $\sum\limits_{k\geqslant1}\|x_k-p_k\|^2<+\infty$,$\sum\limits_{k\geqslant1}\|x_{k+1}-x_k\|^2<+\infty$ 且 $\sum\limits_{k\geqslant1}\|p_k-y_k\|^2<+\infty$;

2) 序列$(H(x_k,p_k))_{k\geqslant1}$ 单调递减且收敛;

3) 序列$((f+g)(p_k))_{k\geqslant1}$ 收敛。

证明 由命题1可得:

$$(f+g)(p_k)+M_2\|x_k-p_k\|^2+(M_1-M_2)\|x_k-p_k\|^2\leqslant(f+g)(p_{k-1})+M_2\|x_{k-1}-p_{k-1}\|^2,$$

通过命题2可得$(M_1-M_2)\|x_k-p_k\|^2>0$,此时2)成立。运用$(M_1-M_2)\|x_k-p_k\|^2>0$ 结合假设1的4)可得3)成立。结合引理3可得 $\sum\limits_{k\geqslant1}\|x_k-p_k\|^2<+\infty$ 成立,此时由式(9)得 $\sum\limits_{k\geqslant1}\|x_{k+1}-x_k\|^2<+\infty$。下证 $\sum\limits_{k\geqslant1}\|p_k-y_k\|^2<+\infty$,结合式(10)和(12)以及 $\lambda\leqslant\bar{\lambda}$ 有:

$$\|p_k-y_k\|^2=\|p_k-p_{k-1}+p_{k-1}-y_k\|^2\leqslant$$
$$2\|p_k-p_{k-1}\|^2+2\|p_{k-1}-y_k\|^2\leqslant$$
$$4\|x_k-p_k\|^2+\left\{4\bar{\lambda}^2L_{\nabla f}+2\left[\frac{\bar{\lambda}L_{\nabla f}}{\sigma_h}(1+\bar{\lambda}L_{\nabla f})+(\bar{\lambda}L_{\nabla f})\right]^2\right\}\|x_{k-1}-p_{k-1}\|^2。$$

由算法1的设置,有 $p_0=x_0$,结合 $\sum\limits_{k\geqslant1}\|x_k-p_k\|^2=\sum\limits_{k\geqslant1}\|x_{k-1}-p_{k-1}\|^2+x_0-p_0^2$ 则有:

$$\sum_{k\geqslant1}\|p_k-y_k\|^2\leqslant\left(4+\left\{4\bar{\lambda}^2L_{\nabla f}+2\left[\frac{\bar{\lambda}L_{\nabla f}}{\sigma_h}(1+\bar{\lambda}L_{\nabla f})+(\bar{\lambda}L_{\nabla f})\right]^2\right\}\right)\sum_{k\geqslant1}\|x_k-p_k\|^2<+\infty,$$

由此得到所有结论成立。 证毕

以下命题给出了极限次微分中某些元素的估计,在后续收敛性证明中会被使用。

命题4(次微分元素估计) 若假设1成立,考虑算法1生成的序列(x_k,p_k)。则对每个 $k\geqslant2$ 有:

$$s_k\in\partial(f+g)(p_k),$$

其中:

$$s_k=\frac{1}{\lambda_k}(\nabla h(x_k)-\nabla h(p_k))+\nabla f(p_k)-2\nabla f(x_k)+\nabla f(x_{k-1})+\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))。$$

此外还满足：

$$\|s_k\|\leqslant\left(\frac{L_{\nabla h}}{\lambda_k}+L_{\nabla f}\right)\|p_k-x_k\|+\left(\frac{\alpha_k L_{\nabla h}}{\lambda_k}+L_{\nabla f}\right)(1+\lambda_{k-1}L_{\nabla f})\|p_{k-1}-x_{k-1}\|。\tag{20}$$

证明　取 $k\geqslant 2$，由算法 1 可得：

$$0\in\partial g(p_k)+\frac{1}{\lambda_k}(\nabla h(p_k)-\nabla h(y_k))+\nabla f(x_k)+\frac{\alpha_k}{\lambda_k}(\nabla h(x_{k-1})-\nabla h(x_k))=$$

$$\partial g(p_k)+\frac{1}{\lambda_k}(\nabla h(p_k)-\nabla h(x_k))+2\nabla f(x_k)-\nabla f(x_{k-1})+\frac{\alpha_k}{\lambda_k}(\nabla h(x_{k-1})-\nabla h(x_k))，$$

第 2 个关系式源于 $\nabla h(y_k)=\nabla h(x_k)-\lambda_k(\nabla f(x_k)-\nabla f(x_{k-1}))$。

因此，利用 $f+g$ 的次微分公式，有：

$$0\in\partial(g+f)(p_k)+\frac{1}{\lambda_k}(\nabla h(p_k)-\nabla h(y_k))+\nabla f(x_k)-\nabla f(p_k)+\frac{\alpha_k}{\lambda_k}(\nabla h(x_{k-1})-\nabla h(x_k))=$$

$$\partial(g+f)(p_k)+\frac{1}{\lambda_k}(\nabla h(p_k)-\nabla h(x_k))+2\nabla f(x_k)-\nabla f(p_k)-\nabla f(x_{k-1})+\frac{\alpha_k}{\lambda_k}(\nabla h(x_{k-1})-\nabla h(x_k))，$$

然后定义序列 $(s_k)_{k\geqslant 2}$ 如下：

$$s_k=\frac{1}{\lambda_k}(\nabla h(y_k)-\nabla h(p_k))+\nabla f(p_k)-\nabla f(x_k)+\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))=$$

$$\frac{1}{\lambda_k}(\nabla h(x_k)-\nabla h(p_k))+\nabla f(p_k)-2\nabla f(x_k)+\nabla f(x_{k-1})+\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))，$$

此外，由 f，h 的梯度 Lipschitz 连续性和关系式(9)，有：

$$\|s_k\|\leqslant\left(\frac{L_{\nabla h}}{\lambda_k}+L_{\nabla f}\right)\|p_k-x_k\|+\left(\frac{\alpha_k L_{\nabla h}}{\lambda_k}+L_{\nabla f}\right)(1+\lambda_{k-1}L_{\nabla f})\|p_{k-1}-x_{k-1}\|。$$　　证毕

在后续描述中，本文使用符号 $\omega((p_k)_{k\geqslant 1})$ 表示序列 $((p_k)_{k\geqslant 1})$ 的聚点集。接下来将类似文献[9]，在额外假设下给出该集合的一些性质。

假设 2　函数 $f+g$ 是强制的，即 $\lim_{\|x\|\to+\infty}(f+g)(x)=+\infty$。

命题 5(聚点集相关性质)　若假设 1、2 成立，考虑算法 1 中生成的序列 (x_k,p_k)，其中 ν，μ，$\bar{\lambda}$，$\underline{\lambda}$ 和 α 按命题 2 选择。则以下陈述成立：

1) $\varnothing\neq\omega((p_k)_{k\geqslant 1})\subseteq\mathrm{crit}(f+g)$。

2) $\lim\limits_{k\to+\infty}\mathrm{dist}(p_k,\omega((p_k)_{k\geqslant 1}))=0$。

3) $\omega((p_k)_{k\geqslant 1})$ 是非空、紧且连通的集。

4) $f+g$ 在 $\omega((p_k)_{k\geqslant 1})$ 上有限且为常数，$\lim\limits_{k\to+\infty}(f+g)(p_k)=(f+g)(p^*)=\rho$。

证明　由于 $f+g$ 是真、下半连续且强制的函数，故 $\inf\limits_{x\in\mathbf{R}^n}[f(x)+g(x)]$ 有限且下确界可达(见文献[6])。因此 $f+g$ 有下界。

1) 根据命题 3 的 2)，有：

$$(f+g)(p_k)\leqslant(f+g)(p_k)+M_2\|x_k-p_k\|^2\leqslant(f+g)(p_1)+M_2\|x_1-p_1\|^2\ \forall k\geqslant 1,$$

由于函数 $f+g$ 是强制的，水平集有界，可以得出结论 $(p_k)_{k\geqslant 1}$ 有界，因此 $\omega((p_k)_{k\geqslant 1})\neq\varnothing$。取任意 $p^*\in\omega((p_k)_{k\geqslant 1})$，存在子序列 $((p_{k_l})_{l\in\mathbf{N}})$ 使得 $p_{k_l}\to p^*$ 当 $l\to+\infty$。

接下来证明 $\lim\limits_{l\to+\infty}g(p_{k_l})=g(p^*)$。注意到函数 g 的下半连续性确保：

$$\liminf_{l\to+\infty}g(p_{k_l})\geqslant g(p^*)。$$

此外，由算法 1 步骤 2、引理 2，以及三角不等式，可以得到对于每个 $k\geqslant 1$ 有：

$$g(p_k)\leqslant g(p^*)+\left\langle p^*,\nabla f(x_k)-\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))\right\rangle+\frac{1}{\lambda_k}D_h(p^*,y_k)-$$

$$\left[\left\langle p_k,\nabla f(x_k)-\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))\right\rangle+\frac{1}{\lambda_k}D_h(p_k,y_k)\right]\leqslant$$

$$g(p^*)+\langle p^*-p_k,\nabla f(x_k)\rangle+\left\langle p^*-p_k,\frac{\alpha_k}{\lambda_k}(\nabla h(x_{k-1})-\nabla h(x_k))\right\rangle+$$

$$\frac{L_{\nabla h}}{\lambda_k}\|p^*-p_k\|^2+\left(\frac{L_{\nabla h}}{\lambda_k}-\frac{\sigma_h}{2\lambda_k}\right)\|p_k-y_k\|^2。$$

由命题3的1),并考虑到所涉及序列的有界性,则可得$\lim\limits_{l\to+\infty}\sup g(p_{k_l})\leqslant g(p^*)$,因此$\lim\limits_{l\to+\infty}g(p_{k_l})=g(p^*)$。

进一步,利用命题4,对所有$l\geqslant 2$有:

$$s_{k_l}\in\partial(f+g)(p_{k_l})。$$

由式(20)和命题3的1),从$p_{k_l}\to p^*$可得,当$l\to+\infty$时有$s_{k_l}\to 0$。再由$\lim\limits_{l\to\infty}(f+g)(p_{k_l})=(f+g)(p^*)$以及极限次微分算子的图的闭性,可保证$0\in\partial(f+g)(p^*)$,因此$p^*\in\mathrm{crit}(f+g)$。

2)和3)的证明可参照文献[9]的引理5进行,同时考虑到文献[9]的注5,其中注意到性质2)和3)对于满足当$k\to+\infty$有$p_{k+1}-p_k\to 0$的序列是通用的。

4)由命题3的3),$((f+g)(p_k))_{k\geqslant 1}$是收敛序列,记此序列的极限为$\rho\in\mathbf{R}$。取任意$p^*\in\omega((p_k)_{k\geqslant 1})$,存在子序列$(p_{k_l})_{l\in\mathbf{N}}$使得当$l\to+\infty$有$p_{k_l}\to p^*$。如项1)所示,有:

$$\lim_{l\to+\infty}(f+g)(p_{k_l})=(f+g)(p^*)。$$

另一方面,$\lim\limits_{l\to+\infty}(f+g)(p_{k_l})=\rho$,因此$(f+g)(p^*)=\rho$。于是,$f+g$在$\omega((p_k)_{k\geqslant 1})$上等于$\rho$,且有:

$$\lim_{k\to+\infty}(f+g)(p_k)=\lim_{l\to+\infty}(f+g)(p_{k_l})=(f+g)(p^*)=\rho。$$ 证毕

接下来,运用前面的命题,得到i*MFRB分裂算法框架下函数值的收敛性结果。

定理1(函数值收敛性) 若假设1、2成立,考虑算法1生成的序列(x_k,p_k),其中$\nu,\mu,\underline{\lambda},\overline{\lambda}$,和$\alpha$按命题2选择,且常数$M_1$和$M_2$如命题1中所定义。则以下陈述成立:

1)$\varnothing\neq\omega((p_k,x_k)_{k\geqslant 1})\subseteq\mathrm{crit}(H)=\{(x,x)\in\mathbf{R}^n\times\mathbf{R}^n:x\in\mathrm{crit}(f+g)\}$。

2)$\lim\limits_{k\to+\infty}\mathrm{dist}((p_k,x_k),\omega((p_k,x_k)_{k\geqslant 1}))=0$。

3)$\omega((p_k,x_k)_{k\geqslant 1})$是非空、紧且连通的集。

4)H在$\omega((p_k,x_k)_{k\geqslant 1})$上有限且为常数,且$\lim\limits_{k\to+\infty}H(p_k,x_k)=H(p^*,p^*)=\rho$。

证明 由命题3的1)有:$\lim\limits_{l\to+\infty}x_{k_l}=\lim\limits_{l\to+\infty}p_{k_l}=p^*$,以及对每个$k\geqslant 2$(见命题3的2))有:

$$H(p_k,x_k)+(M_1-M_2)\|x_k-p_k\|^2\leqslant H(p_{k-1},x_{k-1}),\tag{21}$$

$$(s_k+2M_2(p_k-x_k),2M_2(x_k-p_k))\in\partial H(p_k,x_k)。\tag{22}$$

其中$(s_k)_{k\geqslant 2}$是命题4中引入的序列。式(22)源于:

$$\partial H(x,y)=\{\partial(f+g)(x)+2M_2(x-y)\}\times\{2M_2(y-x)\},\forall(x,y)\in\mathbf{R}^n\times\mathbf{R}^n。$$

再运用命题5对$(f+g)(p_k)$的证明过程,类似证明$H(p_k,x_k)$便可得到结论。 证毕

2.4 序列收敛性

现在可以证明当H是KL函数时i*MFRB分裂算法的序列收敛性结果。

定理2(序列收敛性) 若假设1、2成立,考虑算法1中生成的序列(x_k,p_k),其中其中$\nu,\mu,\underline{\lambda},\overline{\lambda}$,和$\alpha$的选择如命题2所示,以及命题1中的常数$M_1$和$M_2$。并且假设定义2中定义的i*MFRB价值函数$H(x,y)$是KL函数。那么,下列陈述成立:

1)$\sum\limits_{k\geqslant 1}\|x_k-p_k\|<+\infty$且$\sum\limits_{k\in\mathbf{N}}\|x_{k+1}-x_k\|<+\infty$。

2)存在$x\in\mathrm{crit}(f+g)$使得$\lim\limits_{k\to+\infty}x_k=\lim\limits_{k\to+\infty}p_k=x$。

证明 1)根据定理1的1),可以考虑一个元素$p^*\in\mathrm{crit}(f+g)$使得:

$$(p^*,p^*)\in\omega((p_k,x_k)_{k\geqslant 1})。$$

类比命题5的证明,可以容易地证明

$$\lim_{k\to+\infty}H(p_k,x_k)=H(p^*,p^*)。$$

接下来考虑2种情况,

情况 1，存在 $\bar{k}\in\mathbf{N}$，使得 $H(p_{\bar{k}},x_{\bar{k}})=H(p^*,p^*)$。式(21)中的递减性质意味着对于每个 $k\geqslant\bar{k}$ 都有 $H(p_k,x_k)=H(p^*,p^*)$。可以归纳地证明序列 $(p_k,x_k)_{k\geqslant\bar{k}}$ 是常数，结论随之成立。

情况 2，对于所有 $k\geqslant1$，有 $H(p_k,x_k)>H(p^*,p^*)$。取 $\Omega:=\omega((p_k,x_k)_{k\geqslant1})$。由于 H 是 KL 函数，根据定理 1 的 3)、4)以及引理 1，存在 $\varepsilon,\eta>0$ 和 $\varphi\in\Theta_\eta$ 使得对于所有在交集

$$\{(x,y)\in\mathbf{R}^n\times\mathbf{R}^n:\operatorname{dist}((x,y),\Omega)<\varepsilon\}\bigcap$$
$$\{(x,y)\in\mathbf{R}^n\times\mathbf{R}^n:H(p^*,p^*)<H(x,y)<H(p^*,p^*)+\eta\}\tag{23}$$

中的 (x,y)，具有以下不等式成立：

$$\varphi'(H(x,y)-H(p^*,p^*))\operatorname{dist}((0,0),\partial H(x,y))\geqslant1。\tag{24}$$

设 $k_1\geqslant1$ 使得对于每个 $k\geqslant k_1$ 有 $H(p_k,x_k)<H(p^*,p^*)+\eta$。此外，根据定理 1 的 2)，存在 $k_2\in\mathbf{N}$ 使得对于每个 $k\geqslant k_2$ 有 $\operatorname{dist}((p_k,x_k),\Omega)<\varepsilon$。因此，序列 $(p_k,x_k)_{k\geqslant K}$ 属于式(23)中的交集，其中 $K=\max\{k_1,k_2\}$。

由式(24)，有：

$$\varphi'(H(p_k,x_k)-H(p^*,p^*))\operatorname{dist}((0,0),\partial H(p_k,x_k))\geqslant1\quad\forall k\geqslant K,\tag{25}$$

进一步，由于 φ 是凹函数，对于每个 $k\geqslant1$，可以得到下列不等式：

$$\varphi(H(p_k,x_k)-H(p^*,p^*))-\varphi(H(p_{k+1},x_{k+1})-H(p^*,p^*))\geqslant$$
$$\varphi'(H(p_k,x_k)-H(p^*,p^*))\cdot(H(p_k,x_k)-H(p_{k+1},x_{k+1}))。\tag{26}$$

此外，结合式(21)、(25)有：

$$\varphi'(H(p_k,x_k)-H(p^*,p^*))\geqslant\frac{1}{\|(s_k+2M_2(p_k-x_k),2M_2(x_k-p_k))\|}\quad\forall k\geqslant K,\tag{27}$$

此时对每个 $k\geqslant1$ 记 $\Delta_{k,k+1}:=\varphi(H(p_k,x_k)-H(p^*,p^*))-\varphi(H(p_{k+1},x_{k+1})-H(p^*,p^*))$，从式(21)、(26)和(27)可以得到，$\forall k\geqslant K$，有：

$$\Delta_{k,k+1}\geqslant\frac{H(p_k,x_k)-H(p_{k+1},x_{k+1})}{\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|x_k-p_k\|^2}}\geqslant$$
$$\frac{(M_1-M_2)\|x_{k+1}-p_{k+1}\|^2}{\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|x_k-p_k\|^2}}。$$

由此进一步得到，$\forall k\geqslant K$，有：

$$\|x_{k+1}-p_{k+1}\|\leqslant\sqrt{\frac{\Delta_{k,k+1}}{(M_1-M_2)}\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|x_k-p_k\|^2}}=$$
$$\left[\frac{1}{\delta}\frac{\Delta_{k,k+1}}{(M_1-M_2)}\right]^{\frac{1}{2}}(\delta\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|x_k-p_k\|^2})^{\frac{1}{2}}\leqslant$$
$$\frac{\delta}{2}\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|x_k-p_k\|^2}+\frac{\Delta_{k,k+1}}{2\delta(M_1-M_2)},\tag{28}$$

其中选择 $\delta>0$ 且满足下列不等式：

$$\frac{\delta\sqrt{2}}{2}\left\{\sqrt{\left[\left(\frac{L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}+2M_2\right)^2+2M_2^2\right]}+\left[\left(\frac{\alpha L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}\right)(1+\bar{\lambda}L_{\nabla f})\right]\right\}<1。\tag{29}$$

此外，由式(20)对于每个 $k\geqslant1$ 有：

$$\sqrt{\|s_k+2M_2(p_k-x_k)\|^2+4M_2^2\|p_k-x_k\|^2}\leqslant$$
$$\left[\left\|\left(\frac{L_{\nabla h}}{\lambda_k}+L_{\nabla f}+2M_2\right)\|p_k-x_k\|+\left(\frac{\alpha_k L_{\nabla h}}{\lambda_k}+L_{\nabla f}\right)(1+\lambda_{k-1}L_{\nabla f})\|p_{k-1}-x_{k-1}\|\right\|^2+4M_2^2\|p_k-x_k\|^2\right]^{\frac{1}{2}}\leqslant$$
$$\sqrt{\left[2\left(\frac{L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}+2M_2\right)^2+4M_2^2\right]}\|p_k-x_k\|+\sqrt{2}\left[\left(\frac{\alpha L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}\right)(1+\bar{\lambda}L_{\nabla f})\right]\|p_{k-1}-x_{k-1}\|,$$

此时结合式(28)可以推导出：

$$\|x_{k+1}-p_{k+1}\|\leqslant a\|x_k-p_k\|+b\|x_{k-1}-p_{k-1}\|+\frac{\Delta_{k,k+1}}{2\delta(M_1-M_2)}\quad\forall k\geqslant K,\tag{30}$$

其中：

$$a:=\frac{\delta\sqrt{2}}{2}\sqrt{\left[\left(\frac{L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}+2M_2\right)^2+2M_2^2\right]},\ b:=\frac{\delta\sqrt{2}}{2}\left[\left(\frac{\alpha L_{\nabla h}}{\bar{\lambda}}+L_{\nabla f}\right)(1+\bar{\lambda}L_{\nabla f})\right],$$

由式(29)可以得到 $a+b<1$。现在,对于固定的 $j\geqslant 1$ 并结合 φ 仅取非负值可得:

$$\sum_{k=1}^{j}\Delta_{k,k+1}=\varphi(H(p_1,x_1)-H(p^*,p^*))-\varphi(H(p_{j+1},x_{j+1})-H(p^*,p^*))\leqslant$$
$$\varphi(H(p_1,x_1)-H(p^*,p^*)),$$

因此 $\sum\limits_{k\geqslant 1}\frac{\Delta_{k,k+1}}{2\delta(M_1-M_2)}<+\infty$。结合式(30)和引理4可以得出结论 $\sum\limits_{k\geqslant 1}\|x_k-p_k\|<+\infty$。进一步结合式(9),可以得到 $\sum\limits_{k\in\mathbf{N}}\|x_{k+1}-x_k\|<+\infty$。

2) 由结论1)可知$(x_k)_{k\in\mathbf{N}}$是柯西序列,因此是收敛的。由于当 $k\to+\infty$时,有 $x_k-p_k\to 0$,由定理1的结论1)可证。 证毕

3 子问题设计与数值实验

本文主要针对以下极小化问题进行子问题设计与数值实验:

$$\min_{x\in D}\frac{1}{2}\mathrm{dist}^2(x,C),$$

这里 $C=\{x\in\mathbf{R}^n:\mathbf{A}x=\mathbf{b}\}$,$D=\{x\in\mathbf{R}^n:\|x\|_0\leqslant r,\|x\|\leqslant R\}$,$\mathbf{A}\in\mathbf{R}^{m\times n}$,$\mathbf{b}\in\mathbf{R}^m$,$r\in\mathbf{N}\backslash\{0\}$,且 $R>0$。

显然 C 是半代数集。又因为集合 D 可等价表示为 $D=\{x\in\mathbf{R}^n:\|x\|_0\leqslant r\}\bigcap\{x\in\mathbf{R}^n:\|x\|^2\leqslant R^2\}$,可得这是半代数集的交集,由文献[16]的公式(27d),可得 D 也是半代数集。令 $f=\frac{1}{2}\mathrm{dist}^2(\cdot,C)$以及 $g=\delta_D$,则该问题便与式(1)对应。

3.1 子问题设计

根据文献[3],可以获得求解子问题的研究框架。第1步是简化子问题算子 $T_\lambda(u)$的表达式。设 $\kappa_1\geqslant 0$,$\kappa_2\geqslant 0$。固定 $\omega\in\mathbf{R}^n$ 和 $\lambda>0$。在本节后续内容中,令

$$h(x)=\kappa_1\sqrt{1+\|x\|^2}+\frac{\kappa_2}{2}\|x\|^2,$$

容易看出 $h(x)$是强凸函数且其梯度具有 Lipschitz 连续性,强凸系数为 κ_2,梯度 Lipschitz 系数为 $\kappa_1+\kappa_2$。

对任意 $u\in\mathbf{R}^n$,定义 $p_\lambda(u)=\lambda\omega-\nabla h(u)$,并建立如下关系:

$$T_\lambda(u)=\arg\min_{x\in\mathbf{R}^n}\left\{g(x)+\langle x,\omega\rangle+\frac{1}{\lambda}D_h(x,y_n)\right\}=\arg\min_{x\in\mathbf{R}^n}\{\lambda g(x)+\langle x,p_\lambda(u)\rangle+h(x)\}。$$

i * MFRB 子问题式(2)中,令 $\omega=\nabla f(x_k)-\frac{\alpha_k}{\lambda_k}(\nabla h(x_k)-\nabla h(x_{k-1}))$,$u=y_k$ 以及 $\lambda=\lambda_k$,便与 $T_\lambda(u)$对应。对于子问题,可以结合以下定义和引理得到。

定义3(硬阈值算子) $H_r(x)$为 $x\in\mathbf{R}^n$ 的硬阈值算子(参数为 r),$\forall i\in\{1,\cdots,n\}$,$H_r^{(i)}(x)$的定义为:

$$H_r^{(i)}(x)=\begin{cases}x^{(i)},|x^{(i)}|\text{在}|x^{(1)}|,\cdots,|x^{(n)}|\text{中排前}\ r\ \text{位};\\ 0,|x^{(i)}|\text{在}|x^{(1)}|,\cdots,|x^{(n)}|\text{中不排前}\ r\ \text{位}。\end{cases}$$

引理6[3] 设 $g=\delta_D$,其中 $D=\{x\in\mathbf{R}^n:\|x\|_0\leqslant r,\|x\|\leqslant R\}$,整数 $r\leqslant n$,实数 $R>0$。定义

$$x^*=\begin{cases}0,\ \nabla h(u)=\lambda\omega;\\ -t^*\|H_r(p_\lambda(u))\|^{-1}H_r(p_\lambda(\mu)),\ \nabla h(u)\neq\lambda\omega,\end{cases}$$

上式中的 $t*$ 是如下优化问题的唯一解:

$$\min_{0\leqslant t\leqslant R}\left\{\varphi(t)=\kappa_1\sqrt{1+t^2}+\frac{\kappa_2}{2}t^2-\|H_r(p_\lambda(u))\|t\right\}。$$

其中:当 $\|H_r(p_\lambda(u))\|<\kappa_1(1+R^2)^{-\frac{1}{2}}R+\kappa_2R$ 时,t^* 是方程

$$\kappa_1(1+t^2)^{-\frac{1}{2}}t+\kappa_2t-\|H_r(p_\lambda(u))\|=0$$

的唯一解，且 $t^* \in (0,R)$；否则，$t^* = R$。则此时 $x^* \in T_\lambda(u)$。

3.2　数值实验

本文将对 i * MFRB、BiFRB[3] 和 iTseng[5] 进行基准测试，参数设置为 $R = 1\ 000$ 和 $r = \lfloor m/5 \rfloor$。i * MFRB 和 iTseng 分别采用命题 2 和文献[5]引理 3.3 给出的步长，BiFRB 则按照文献[3]命题 4.2 的步长规则实现。当达到最大迭代次数(10 000)或者满足

$$\frac{\max\{\|x_k - x_{k-1}\|, \|p_k - p_{k-1}\|\}}{\max\{1, \|x_{k-1}\|, \|p_{k-1}\|\}} < o$$

时终止迭代，其中 o 为预设的收敛容差。此外，对于子问题中的函数 h，i * MFRB 和 BiFRB 分别使用：

$$h_1(x) = 0.05 \cdot \sqrt{1 + \|x\|^2} + 0.05 \cdot \frac{\|x\|^2}{2},$$

$$h_2(x) = 0.1 \cdot \sqrt{1 + \|x\|^2} + 2.51 \cdot \frac{\|x\|^2}{2}。$$

关于惯性参数，根据命题 2 和文献[5]引理 3.3 中惯性要求并且令系数 $\mu = 1$，i * MFRB 使用

$$\alpha_1 = 0.9 \times 0.05 \div (2 \times (1 + (0.05 + 0.05)^2 \div 2)),$$

iTseng 使用

$$\alpha_2 = 0.9 \times 1 \div (2 \times (1 + (1 \div 2))) = 0.3,$$

参照文献[3]数值实验部分内容，BiFRB 使用 $\alpha_3 = 0.9$，3 个算法惯性的选取都靠近惯性上界。所有算法均从原点初始化，步长选取方式为各自采取步长上界的基础上，采用文献[1]备注 4 中描述的步长启发式更新方法。涉及代码在 MATLAB(R2022a)中实现，实验运行环境为 64 位 Windows 系统，硬件为 Intel(R)Core(TM)i5-8300H 2.30 GHz 处理器及 16 GB 内存。

矩阵 $\boldsymbol{A} \in \mathbf{R}^{m \times n}$ 随机生成且满足标准高斯分布。针对不同规模需求，随机生成了 50 组算例。最终，获得了各规模下 50 次算例的平均迭代次数和 50 次算例所求得最终目标函数值的最小值($f_{\min}^{\text{value}}$)。仿真结果如表 1 所示；函数值随迭代次数(最大迭代设置为 2 000)的平均变化情况如图 1 所示。

观察到以下现象：

1) 与同类外推算法 iTseng 相比，i * MFRB 分裂算法明显提高了收敛速度。在所有参数设置情况下，iTseng 几乎无法收敛，但 i * MFRB 仍能在 2 000 次迭代内实现收敛。

2) 与 BiFRB 分裂算法相比，在 $m \geqslant 500$ 时，i * MFRB 的迭代次数仅为 BiFRB 的一半左右。

3) 在目标函数值方面，i * MFRB 分裂算法在较大 m($m \geqslant 500$)的情况下能够获得最优结果。

4) 从图 1 可以看出，算法 i * MFRB 在所有测试场景下均能快速收敛且收敛效果最优；算法 iTseng 短时间几乎无法收敛；而算法 BiFRB 的收敛效果介于两者之间。

表 1　i * MFRB 与 BiFRB 和 iTseng 的对比

Tab.1　Comparing i * MFRB to BiFRB and iTseng

规模		i * MFRB		BiFRB		iTseng	
m	n	平均迭代数/次	$f_{\min}^{\text{value}}$	平均迭代数/次	$f_{\min}^{\text{value}}$	平均迭代数/次	$f_{\min}^{\text{value}}$
100	4 000	1 000	2.171e−04	590	3.170e−03	10 000	5.579e−03
100	5 000	1 230	2.23e−18	428	2.40e−19	10 000	6.520e−03
100	6 000	1 468	1.143e−03	515	2.52e−19	10 000	6.746e−03
500	4 000	1 188	9.87e−19	2 053	5.05e−17	8 995	8.63e−06
500	5 000	304	1.30e−18	2 586	1.205e−04	9 003	1.417e−03
500	6 000	392	1.55e−18	3 022	6.157e−02	9 856	5.758e−02
900	4 000	1 697	6.38e−17	2 444	1.46e−16	6 489	4.90e−16
900	5 000	1 957	1.97e−17	2 996	2.42e−16	6 161	9.276e−03
900	6 000	1 948	2.92e−07	3 546	5.17e−05	6 857	6.441e−01

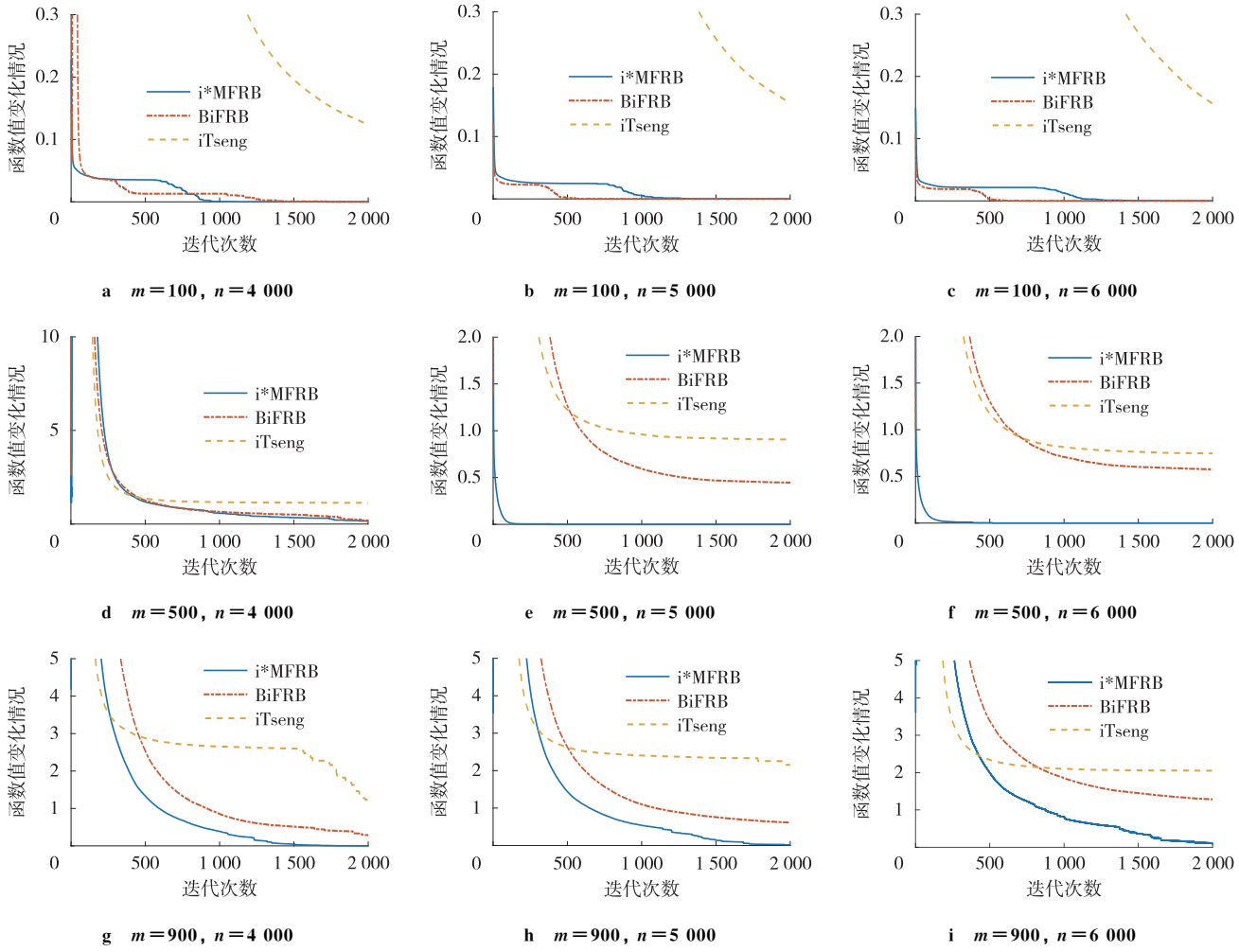

图 1 经过 50 次重复后,函数值在迭代过程中的平均变化

Fig. 1 Average change of function values over itertion after 50 repetition

4 结论

本文以 i * FRB 为研究对象,针对非凸复合优化问题,提出了一种带有外推步的 i * MFRB。首先,设计了与迭代过程中的惯性、步长参数相关性较低且具有显式形式的价值函数,并在适当步长规则下证明了该算法的函数值收敛性;其次,结合 KL 性质建立了迭代序列的收敛性;进一步地,针对一类非凸可行性问题设计了有效的子问题求解策略;最后的数值实验表明,该算法在高维情形下具有明显的数值优势。

参考文献:

[1] Li G Y, Pong T K. Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems[J]. Mathematical Programming, 2016, 159 (1/2): 371-401.

[2] Wang X F, Wang Z Y. Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems[J]. Computational Optimization and Applications, 2022, 82 (2): 441-463.

[3] Wang X F, Wand Z Y. A Bregman inertial forward-reflected-backward method for nonconvex minimization[J]. Journal of Global Optimization, 2024, 89 (2): 327-354.

[4] Wang Z Y, Themelis A, Ou H J, et al. A mirror inertial forward-reflected-backward splitting: convergence analysis beyond convexity and Lipschitz smoothness[J]. Journal of Optimization Theory and Applications, 2024, 203 (2): 1127-1159.

[5] Boţ R I, Csetnek E R. An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems[J]. Journal of Optimization Theory and Applications, 2016, 171 (2): 600-616.

[6] Rockafellar R T, Wets R J B. Variational analysis[M]. Berlin, Heidelberg: Springer, 1998.
[7] Mordukhovich B S. Variational analysis and generalized differentiation, I: basic theory[M]. Berlin: Springer-Verlag, 2006.
[8] Mordukhovich B S. Variational analysis and generalized differentiation, II: applications[M]. Berlin: Springer-Verlag, 2006.
[9] Attouch H, Bolte J, Redont P, et al. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the kurdyka-łojasiewicz inequality[J]. Mathematics of Operations Research, 2010, 35 (2): 438-457.
[10] Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems[J]. Mathematical Programming, 2014, 146 (1/2): 459-494.
[11] Bolte J, Daniilidis A, Lewis A, et al. Clarke subgradients of stratifiable functions[J]. SIAM Journal on Optimization, 2007, 18 (2): 556-572.
[12] Bolte J, Daniilidis A, Ley O, et al. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity[J]. Transactions of the American Mathematical Society, 2010, 362 (6): 3319-3363.
[13] Łojasiewicz S. Une propriété topologique des sous-ensembles analytiques réels[C]//Les équations aux dérivées partielles. Paris: éditions du Centre National de la Recherche Scientifique, 1963: 87-89.
[14] Kurdyka K. On gradients of functions definable in o-minimal structures[J]. Annales de L' Institut Fourier, 1998, 48 (3): 769-783.
[15] Bauschke H H, Combettes P L. Convex analysis and monotone operator theory in Hilbert spaces[M]. Cham: Springer International Publishing, 2017.
[16] Bauschke H H, Luke D R, Phan H M, et al. Restricted normal cones and sparsity optimization with affine constraints[J]. Foundations of Computational Mathematics, 2014, 14 (1): 63-83.

Special Section on Non-Convex Optimization and Numerical Algorithms

A Mirror Inertial Modified Forward-Reflected-Backward Splitting for Nonconvex Composite Optimization

CHUAN Junlin, LUO Honglin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: This paper focuses on non-convex composite optimization problems and improves the mirror inertial forward-reflected-backward (i * FRB) algorithm by introducing an extrapolation step, proposing a mirror inertial modified forward-reflected-backward method (i * MFRB). This work extends the application of the forward-backward splitting algorithm with an extrapolation step to non-convex scenarios and offers a more concise proof approach for the forward-reflected-backward splitting algorithm. By constructing an appropriate and simple value function, the convergence of the algorithm's function values is established. Furthermore, by combining the Kurdyka-Łojasiewicz property, the convergence of the algorithm's iterative sequence is proved. The proposed algorithm is applied to a class of non-convex feasibility problems, and a corresponding sub-problem solving approach is designed. Finally, numerical experiments are conducted to compare the i * MFRB splitting algorithm with existing similar algorithms in high-dimensional cases.

Keywords: non-convex composite optimization; Tseng-type proximal algorithms; forward-reflected-backward splitting algorithm; Bregman proximal mapping; Kurdyka-Łojasiewicz property

(责任编辑 陈新颖)