运筹学与控制论 DOI:10.11721/cqnuj20160116

# 非凸优化的近似束方法及对偶问题*

沈 洁,刘晓倩,陈 颖,金 希
(辽宁师范大学 数学学院,辽宁 大连 116029)

**摘要**:束方法目前是解决非光滑优化问题最有前景的方法之一。出于实际计算的需要,使用两个扰动函数共同控制真实目标函数,利用它们的信息构建增广函数,从而把凸优化迫近束方法应用到非凸问题中来。类似地建立目标函数的下近似模型,通过求解二次规划最小值点作为下一个候选点,进一步再筛选出下降点。最后利用 Lagrange 函数写出了束方法子问题的对偶问题,揭示了扰动后原问题的最优解和对偶问题最优解之间的关系。

**关键词**:非凸非光滑优化;束方法;近似函数值;Lagrange 对偶问题;lower-$C^2$ 函数
**中图分类号**:O221.2 **文献标志码**:A **文章编号**:1672-6693(2016)01-0001-05

非凸非光滑优化束方法受到很多学者重视进而展开研究,取得了许多有益成果[1-6]。Hare 和 Sagastizábal[7]研究了非凸优化的重新分配束方法,提出了对目标函数的局部凸化策略,构造了局部近似的切平面模型,并通过求解子问题,逐步迭代出原问题最优解。Kiwiel[8]考虑凸约束优化问题 $\inf\{f(x):x\in S\}$,给出一种带有近似线性化误差的迫近束方法,采用 $f$ 的近似值 $f_y\in[f(y)-\varepsilon_f,f(y)+\varepsilon_g]$和它的近似次梯度 $\partial_\varepsilon f(y):=\{g:f(\cdot)\geqslant f(y)-\varepsilon+[g,\cdot-y]\}$,构造凸函数切平面模型,最终迭代出问题的 ε-最优解。Hare 和 Sagastizábal[9]通过构造非凸函数的分片仿射模型,逐步计算出迫近点,为最终求解服务。

下面将非凸重新分配迫近束方法和目标函数近似虚拟值的选取技巧相结合,求解更一般化的非凸非光滑无约束优化问题:

$$\begin{cases}\min & f(y),\\ \text{s.t.} & y\in\mathbf{R}^n。\end{cases}\tag{1}$$

其中 $f:\mathbf{R}^n\to\mathbf{R}$ 是一个非凸函数。借鉴文献[7]中的迫近束方法思想,在目标函数值无法精确计算的情况下给出了一种扰动方法,讨论并揭示了扰动后原问题的最优解和对偶问题最优解之间的关系。而文献[7]完全是在使用函数真实信息情况下讨论了问题的求解,可以说所给出的方法是文献[7]中方法的一种扰动。这种设想初衷源于某些函数本身是由一个极小化问题来定义的,因而根本无法精确求得函数在某一点的函数值,如 Moreau-Yosida 正则化函数。既然实际计算中有时很难求得函数的真实值 $f(y)$,所以人们常用它的虚拟值$\tilde{f}(y)$代替。虚拟值有很多种取法,采用如下选取方式:假设给定误差值 $\varepsilon\geqslant0$,将 ε 平均分成 $n$ 份,令$\tilde{f}(y)\approx f(y)+\frac{\varepsilon}{n}m(y)$,其中$\tilde{f}(y)$和 $m(y)$都是非光滑函数。特别地,当 $m(y)=0$ 时,$\tilde{f}(y)=f(y)$;当 $m(y)\in[-n,n]$时,$\frac{\varepsilon}{n}m(y)\in[-\varepsilon,\varepsilon]$,即$\tilde{f}(y)$在$[f(y)-\varepsilon,f(y)+\varepsilon]$范围内,是对 $f(y)$的一种近似模拟。基于以上假设,考虑如下优化问题:

$$\begin{cases}\min & \tilde{f}(y)-\frac{\varepsilon}{n}m(y),\\ \text{s.t.} & y\in\mathbf{R}^n。\end{cases}\tag{2}$$

## 1 预备知识

**假设 1** 给定 $x^0\in\mathbf{R}^n$,$M_0\geqslant0$,存在一个开有界集 $O$ 和一个函数 $F$,使得 $F$ 在 $O$ 上是 lower-$C^2$[7]的且满足

* 收稿日期:2015-03-02 修回日期:2015-04-30 网络出版时间:2015-12-02 13:27
资助项目:国家自然科学基金(No. 11301246)
作者简介:沈洁,女,副教授,博士,研究方向为运筹学与控制论,E-mail:tt010725@163.com
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.n.20151202.1327.028.html

$F(\cdot)\equiv\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$,其中 $L_0:=\left\{x\in\mathbf{R}^n:\tilde{f}(x)-\frac{\varepsilon}{n}m(x)\leqslant\tilde{f}(x^0)-\frac{\varepsilon}{n}m(x^0)+M^0\right\}\subset O$。

**命题 1** 对于满足假设 1 的函数 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$,存在一个阈值 $\rho^{id}>0$,使得对任意 $\rho\geqslant\rho^{id}$ 和任意给定的 $y\in L_0$,函数 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)+\frac{\rho}{2}\|\cdot-y\|^2$ 在 $L_0$ 上是凸函数,并且函数 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$ 在 $L_0$ 上是 Lipschitz 连续的。

证明类似于文献[7]中命题 1 的证明。

**定义** 1(迫近映射 $p_R$) 对于 $x\in\mathbf{R}^n$,$p_R\left[\tilde{f}(x)-\frac{\varepsilon}{n}m(x)\right]:=\arg\min\limits_{y}\left\{\tilde{f}(y)-\frac{\varepsilon}{n}m(y)+\frac{R}{2}\|y-x\|^2\right\}$ 称作在 $x$ 处的迫近点映射。

由命题 1 知,当迫近参数 $R$ 足够大,也就是 $R>\rho^{id}$ 时,迫近点映射的值是唯一的,且在 $L_0$ 上是 Lipschitz 连续的。由定义可知,$\tilde{f}(x)-\frac{\varepsilon}{n}m(x)$ 的稳定点可用迫近点映射的稳定点刻画:$\bar{x}\in L_0$ 是 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$ 的稳定点当且仅当 $\bar{x}=p_R\left[\tilde{f}(\bar{x})-\frac{\varepsilon}{n}m(\bar{x})\right]$。为了更方便地迭代出 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$ 的稳定点,把定义 1 等价转化为:

$$p_R\left[\tilde{f}(x)-\frac{\varepsilon}{n}m(x)\right]=p_{R-\rho^{id}}\left[\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)+\frac{\rho^{id}}{2}\|\cdot-x\|^2\right](x)。\tag{3}$$

## 2 近似束方法

在每次迭代过程中,束方法保留的束信息为:

$$\bigcup_{i\in I_n}\{[x^i,\ \tilde{f}_i=\tilde{f}(x^i),\ m_i=m(x^i),\ \tilde{g}_f^i\in\partial\tilde{f}(x^i),\ g_m^i\in\partial m(x^i)]\},$$

其中 $I_n\subseteq\{0,1,2,\cdots,n\}$ 是当前迭代指标集。对近似函数 $\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)$,在下降点 $\hat{x}^k$ 的近似线性化误差定义为:

$$\tilde{e}_i^k:=\tilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\left(\tilde{f}_i-\frac{\varepsilon}{n}m_i+\left\langle\tilde{g}_f^i-\frac{\varepsilon}{n}g_m^i,\hat{x}^k-x^i\right\rangle\right)。\tag{4}$$

构造 $\hat{x}^k$ 点的增广函数:

$$\tilde{f}_{\eta_n}^{\hat{x}^k}(\cdot):=\tilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)+\frac{\eta_n}{2}\|\cdot-\hat{x}^k\|^2。\tag{5}$$

考虑增广束信息:

$$\bigcup_{i\in I_n}\{(\tilde{e}_i^k,\ d_i^k,\ \Delta_i^k,\ \tilde{g}_f^i,\ g_m^i)\},\tag{6}$$

其中 $\tilde{e}_i^k$ 如(4)式中定义,$d_i^k=\frac{\|x^i-\hat{x}^k\|^2}{2}$,$\Delta_i^k=x^i-\hat{x}^k$,$\tilde{g}_f^i\in\partial\tilde{f}(x^i)$,$g_m^i\in\partial m(x^i)$。

在束方法中,下降点 $\hat{x}^k$ 代表当前迭代的迫近中心。通过求解一个二次规划(QP)问题得到下一个候选点 $x^{n+1}$。为了更有效地求解(QP)问题,将当前的迫近参数 $R$ 分成两项 $\eta_n$ 和 $\mu_n$,并满足 $R=\eta_n+\mu_n$,则由(3)式、(5)式可导出等式 $p_R\left[\tilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)\right]=p_{\mu_n}\tilde{f}_{\eta_n}^{\hat{x}^k}(\hat{x}^k)$。接下来利用 $\hat{x}^k$ 和 $\mu_n$ 构造凸化函数的下近似模型 $\tilde{\varphi}_n$,分别称 $\eta_n$ 和 $\mu_n$ 为凸化参数和迫近参数,$R,\eta_n,\mu_n$ 都随着迭代的束信息改变而改变。利用束信息构造函数 $\tilde{f}_{\eta_n}^{\hat{x}^k}(\cdot)$ 的切平面模型:

$$\tilde{\varphi}_n(y):=\tilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)+\max_{i\in I_n}\left\{-(\tilde{e}_i^k+\eta_n d_i^k)+\left\langle\tilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k,\ y-\hat{x}^k\right\rangle\right\}。\tag{7}$$

下一个候选点定义为:

$$x^{n+1}:=p_{\mu_n}\tilde{\varphi}_n(\hat{x}^k)。\tag{8}$$

定义额定下降:

$$\delta_{n+1}:=\tilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)+\frac{\eta_n}{2}\|x^{n+1}-\hat{x}^k\|^2-\tilde{\varphi}_n(x^{n+1})。\tag{9}$$

那么下一个下降点将是满足重要步条件的候选点 $x^{n+1}$，即 $x^{n+1}$ 满足 $\widetilde{f}(x^{n+1})-\frac{\varepsilon}{n}m(x^{n+1})\leqslant\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\gamma\cdot\delta_{n+1}$，其中 $\gamma\in(0,1)$，记 $\hat{x}^{k+1}=x^{n+1}$；否则为零步，记 $\hat{x}^k=x^{n+1}$，并重新定义参数进行重要步的寻找。一旦取重要步，更新增广束信息如下：$\forall i\in I_n$，

$$\begin{cases}\widetilde{e}_i^{k+1}=\widetilde{e}_i^k+\widetilde{f}(\hat{x}^{k+1})-\frac{\varepsilon}{n}m(\hat{x}^{k+1})-\widetilde{f}(\hat{x}^k)+\frac{\varepsilon}{n}m(\hat{x}^k)-\left\langle \widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i,\hat{x}^{k+1}-\hat{x}^k\right\rangle,\\ d_i^{k+1}=d_i^k+\frac{\|\hat{x}^{k+1}-\hat{x}^k\|^2}{2}+\langle\Delta_i^k,\hat{x}^{k+1}-\hat{x}^k\rangle,\\ \Delta_i^{k+1}=\Delta_i^k+\hat{x}^k-\hat{x}^{k+1}。\end{cases}\tag{10}$$

结合束信息(6)式，进一步推导得到：

$$\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\in\partial_{\widetilde{e}_i^k+\eta_n d_i^k}\widetilde{\varphi}_n(\hat{x}^k),\tag{11}$$

其中 $\widetilde{e}_i^k+\eta_n d_i^k\geqslant0,\forall i\in I_n$。

为保证线性化误差非负，定义 $\eta_n^{\min}:=\max\limits_{i\in I_n,d_i^k>0}-\frac{\widetilde{e}_i^k}{d_i^k}$。显然，若 $\eta_n\geqslant\eta_n^{\min}$，有 $\widetilde{e}_i^k+\eta_n d_i^k\geqslant0,\forall i\in I_n$。

## 3 对偶问题

**定理 1**　令 $x^{n+1}$ 是(8)式定义的迫近点，则

$$x^{n+1}=\hat{x}^k-\frac{1}{\mu_n}g_{\eta_n}^{-n},\tag{12}$$

其中 $g_{\eta_n}^{-n}:=\sum\limits_{i\in I_n}\bar{\alpha}_i(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k)$，$\bar{\alpha}=(\bar{\alpha}_1,\bar{\alpha}_2,\cdots,\bar{\alpha}_{|I_n|})$ 是下述问题的最优解：

$$\begin{cases}\min\limits_{\alpha_i}\ \frac{1}{2}\left\|\sum\limits_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2+\mu_n\sum\limits_{i\in I_n}\alpha_i(\widetilde{e}_i^k+\eta_n d_i^k),\\ \alpha\in\Delta_n:=\left\{z\in[0,1]^{|I_n|}:\sum\limits_{i\in I_n}z_i=1\right\}。\end{cases}\tag{13}$$

此外，下列关系也成立：1) $g_{\eta_n}^{-n}\in\partial\widetilde{\varphi}_n(x^{n+1})$；2) $\delta_{n+1}=\omega_n+(R+\mu_n)d_{n+1}^k$，其中 $\omega_n:=\sum\limits_{i\in I_n}\bar{\alpha}_i(\widetilde{e}_i^k+\eta_n d_i^k)$；3) 当 $\eta_n\geqslant\rho^{id}$ 时，$g_{\eta_n}^{-n}\in\partial_{\omega_n}\left[\widetilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)+\frac{\eta_n}{2}\|\cdot-\hat{x}^k\|^2\right](\hat{x}^k)$。

**证明**　令 $\theta=\widetilde{\varphi}_n(y)$，则问题(8)等价于

$$\begin{cases}\min\limits_{(y,\theta)\in\mathbf{R}^n\times\mathbf{R}}\ \theta+\frac{\mu_n}{2}\|y-\hat{x}^k\|^2,\\ \theta\geqslant\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-(\widetilde{e}_i^k+\eta_n d_i^k)+\left\langle\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k,\ y-\hat{x}^k\right\rangle,\ \forall i\in I_n。\end{cases}\tag{14}$$

对 $\forall\alpha\in\mathbf{R}_+^{|I_n|}$，相应的 Lagrange 函数为：

$$L(y,\theta,\alpha)=\left(1-\sum_{i\in I_n}\alpha_i\right)\theta+\frac{\mu_n}{2}\|y-\hat{x}^k\|^2+\\ \sum_{i\in I_n}\alpha_i\left[\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-(\widetilde{e}_i^k+\eta_n d_i^k)+\left\langle\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k,y-\hat{x}^k\right\rangle\right]。$$

问题(8)的目标函数是强凸的，故有唯一解。等价问题(14)的约束是仿射的，因此存在与 $x^{n+1}$ 相应的乘子 $\bar{\alpha}$。又根据文献中[10]中的结果，原问题与对偶问题之间没有对偶间隙，所以 $(x^{n+1},\bar{\alpha})$ 可以通过解原始问题或对偶问题得到：

$$(14)式\equiv\min_{(y,\theta)\in\mathbf{R}^n\times\mathbf{R}}\ \max_{\alpha\in\mathbf{R}_+^{|I_n|}}L(y,\theta,\alpha)\equiv\max_{\alpha\in\mathbf{R}_+^{|I_n|}}\ \min_{(y,\theta)\in\mathbf{R}^n\times\mathbf{R}}L(y,\theta,\alpha)。$$

因为对偶问题中 $\theta$ 是没有限制的，所以为使对偶问题最优值有限，$L$ 中的乘子 $\theta$ 必须消失，也就是 $\alpha\in\Delta_n$，即 $\left(1-\sum\limits_{i\in I_n}\alpha_i\right)=0$。$x^{n+1}$ 和 $\bar{\alpha}$ 分别是原问题和对偶问题的解，那么上式可化为 $\min\limits_{y\in\mathbf{R}^n}\max\limits_{\alpha\in\mathbf{R}_+^{|I_n|}}L(y,\alpha)\equiv\max\limits_{\alpha\in\mathbf{R}_+^{|I_n|}}\min\limits_{y\in\mathbf{R}^n}L(y,\alpha)$，

其中 $L(y,\alpha)=\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)+\frac{\mu_n}{2}\|y-\hat{x}^k\|^2+\sum_{i\in I_n}\alpha_i\left[-(\widetilde{e}_i^k+\eta_n d_i^k)+\left\langle \widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k, y-\hat{x}^k\right\rangle\right]$。

考虑对偶问题,对每个固定的 $\alpha\in\Delta_n$,解 $y(\alpha)=\arg\min_y L(y,\alpha)$ 的最优解条件是 $0=\nabla_y L(y(\alpha),\alpha)$,即

$$0=\mu_n(y(\alpha)-\hat{x}^k)+\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)。\tag{15}$$

特别是当 $\alpha=\bar{\alpha}$ 时,有 $y(\bar{\alpha})=x^{n+1}$,所以(12)式成立。

为证明 $\bar{\alpha}$ 是(13)式的解,在(15)式两边同时乘以 $y(\alpha)-\hat{x}^k$ 及 $\frac{1}{\mu_n}\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)$,则

$$0=\mu_n\|y(\alpha)-\hat{x}^k\|^2+\sum_{i\in I_n}\alpha_i\left\langle \widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k, y(\alpha)-\hat{x}^k\right\rangle=$$

$$\sum_{i\in I_n}\alpha_i\left\langle \widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k, y(\alpha)-\hat{x}^k\right\rangle+\frac{1}{\mu_n}\left\|\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2。$$

由此得出 $\mu_n\|y(\alpha)-\hat{x}^k\|^2=\frac{1}{\mu_n}\left\|\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2$,且

$$L(y(\alpha),\alpha)=\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2-\sum_{i\in I_n}\alpha_i(\widetilde{e}_i^k+\eta_n d_i^k)。$$

结合上面两个等式,$\bar{\alpha}$ 是下面问题最优解:

$$\max_{\alpha\in\Delta_n}L(y(\alpha),\alpha)=\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\min_{\alpha\in\Delta_n}\left\{\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2+\sum_{i\in I_n}\alpha_i(\widetilde{e}_i^k+\eta_n d_i^k)\right\}。\tag{16}$$

所以(13)式成立。

为了证明1),利用(8)式的最优性条件和(12)式可得

$$0\in\partial\widetilde{\varphi}_n(x^{n+1})+\mu_n(x^{n+1}-\hat{x}^k)=\partial\widetilde{\varphi}_n(x^{n+1})-g_{\eta_n}^{-n},$$

从而有 $g_{\eta_n}^{-n}\in\partial\widetilde{\varphi}(x^{n+1})$。

下面证明2)。因为没有对偶间隙,所以原问题(8)的最优值等于对偶问题(16)的最优值,即:

$$\widetilde{\varphi}_n(x^{n+1})+\frac{\mu_n}{2}\|x^{n+1}-\hat{x}^k\|^2=\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\bar{\alpha}_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2-\sum_{i\in I_n}\bar{\alpha}_i(\widetilde{e}_i^k+\eta_n d_i^k)。$$

应用(9)式、(12)式,则:

$$\delta_{n+1}=\frac{\mu_n}{2}\|x^{n+1}-\hat{x}^k\|^2+\frac{\eta_n}{2}\|x^{n+1}-\hat{x}^k\|^2+\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\bar{\alpha}_i\left(\widetilde{g}_f^i-\frac{\varepsilon}{n}g_m^i+\eta_n\Delta_i^k\right)\right\|^2+\sum_{i\in I_n}\bar{\alpha}_i(\widetilde{e}_i^k+\eta_n d_i^k)=$$

$$\frac{R}{2}\|x^{n+1}-\hat{x}^k\|^2+\frac{1}{2\mu_n}\|g_{\eta_n}^{-n}\|^2+\omega_n=(R+\mu_n)d_{n+1}^k+\omega_n,$$

其中 $\omega_n:=\sum_{i\in I_n}\bar{\alpha}_i(\widetilde{e}_i^k+\eta_n d_i^k)$。

最后证明3),当 $\eta_n\geqslant\rho^{id}$ 时,$\forall y\in\mathbf{R}^n$,因为函数 $\widetilde{f}(\cdot)-\frac{\varepsilon}{n}m(\cdot)+\frac{\eta_n}{2}\|\cdot-\hat{x}^k\|^2$ 是凸的,所以 $\widetilde{f}(y)-\frac{\varepsilon}{n}m(y)+\frac{\eta_n}{2}\|y-\hat{x}^k\|^2\geqslant\widetilde{\varphi}_n(y)$,由1)知 $\widetilde{\varphi}_n(y)\geqslant\widetilde{\varphi}_n(x^{n+1})+\langle g_{\eta_n}^{-n},y-x^{n+1}\rangle$,应用(12)式得:

$$\widetilde{f}(y)-\frac{\varepsilon}{n}m(y)+\frac{\eta_n}{2}\|y-\hat{x}^k\|^2\geqslant\widetilde{\varphi}_n(x^{n+1})+\langle g_{\eta_n}^{-n},y-x^{n+1}\rangle=$$

$$\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)+\langle g_{\eta_n}^{-n},y-\hat{x}^k\rangle-\left[\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\widetilde{\varphi}_n(x^{n+1})-\frac{1}{\mu_n}\|g_{\eta_n}^{-n}\|^2\right]。\tag{17}$$

应用(9)式和2),上式后半部分转化成

$$\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)-\widetilde{\varphi}_n(x^{n+1})-\frac{1}{\mu_n}\|g_{\eta_n}^{-n}\|^2=\delta_{n+1}-\frac{\eta_n}{2}\|x^{n+1}-\hat{x}^k\|^2-\frac{1}{\mu_n}\|g_{\eta_n}^{-n}\|^2=\omega_n,$$

则(17)式为 $\widetilde{f}(y)-\frac{\varepsilon}{n}m(y)+\frac{\eta_n}{2}\|y-\hat{x}^k\|^2\geqslant\widetilde{f}(\hat{x}^k)-\frac{\varepsilon}{n}m(\hat{x}^k)+\langle g_{\eta_n}^{-n},y-\hat{x}^k\rangle-\omega_n$,即3)成立。 证毕

## 参考文献:

[1] Yang T B, Mahdavi M, Jin R, et al. An efficient primal dual prox method for non-smooth optimization[J]. Machine Learning, 2015, 98(3): 369-406.

[2] Noll D. Cutting plane oracles to minimize non-smooth non-convex functions[J]. Set-Valued and Variational Analysis, 2010, 18(3/4): 531-568.

[3] Do Trinh-Minh-Tri, Artières T. Regularized bundle methods for convex and non-convex risks[J]. The Journal of Machine Learning Research, 2012, 13(1): 3539-3583.

[4] Kristian B, Dirk A L, Stefan R. Minimization of non-smooth non-convex functions by iterative thresholding[J]. Journal of Optimization Theory and Applications, 2015, 165(1): 78-112.

[5] Shen J, Li D, Pang L P. A cutting plane and level stabilization bundle method with inexact data for minimizing non-smooth non-convex functions[J]. Abstract and Applied Analysis, 2014, Article ID 192893, 6 pages.

[6] Shen J, Liu X Q, Guo F F, et al. An approximate redistributed proximal bundle method with inexact data for minimizing non-smooth non-convex functions[J]. Mathematical Problems in Engineering, 2015, Article ID 215310, 9 pages.

[7] Hare W, Sagastizábal C. A redistributed proximal bundle method for non-convex optimization[J]. SIAM Journal on Optimization, 2010, 20(5): 2442-2473.

[8] Kiwiel K C. A proximal bundle method with approximate subgradient linearizations[J]. SIAM Journal on Optimization, 2006, 16(4): 1007-1023.

[9] Hare W, Sagastizábal C. Computing proximal points of non-convex function[J]. Math Program, 2009, 116(1/2, Ser. B): 221-285.

[10] 王宜举,修乃华. 非线性最优化理论与方法[M]. 北京:科学出版社,2012.
Wang Y J, Xiu N H. Nonlinear optimization theory and methods[M]. Beijing: Science Press, 2012.

**Operations Research and Cybernetics**

## An Approximate Bundle Method for Non-convex Optimization and Its Dual Problem

SHEN Jie, LIU Xiaoqian, CHEN Ying, JIN Xi

(School of Mathematics, Liaoning Normal University, Dalian Liaoning 116029, China)

**Abstract**: Proximal bundle method is considered to be one of the most promising methods for solving non-smooth optimization problems. In this paper, due to the requirement of practical calculations, two disturbance functions are used to control the real objective function together, whose information is utilized to construct an augmented function. Thereby bundle method for convex optimization is applied to non-convex optimization problem. Similarly, lower approximate model of the objective function can be constructed and by solving quadratic programming, we expect to find out the minimum point as next candidate point, and furthermore select the decreased point. Finally, the dual problem of sub-problem of bundle method is given by utilizing Lagrange function, and at the same time the relation between the solutions of the primal problem which has been disturbed and the dual problem is also revealed.

**Key words**: non-smooth non-convex optimization; bundle method; approximate function value; Lagrange dual problem; lower-$C^2$ function

(责任编辑 黄 颖)