

文章编号: 1001-0920(2004)03-0303-04

变直径次梯度投影函数优化方法

江永亨, 周 威, 金以慧
(清华大学 自动化系, 北京 100084)

摘 要: 拉格朗日松弛法的关键是求解对偶函数, 而在对偶函数不可微的情况下人们经常采用次梯度法, 为此提出一种变直径次梯度投影法. 该方法根据投影性质确定对偶问题定义域的有效直径, 从而使其收敛性不依赖于最优目标值和对偶问题定义域直径等任何先验知识, 并证明了其收敛性, 给出了收敛效率. 通过一个指派问题说明了所提出方法的有效性.

关键词: 对偶; 次梯度投影; 变直径

中图分类号: O 232 **文献标识码:** A

Subgradient projection method with variable diameter for optimization

J I A N G Y o n g - h e n g , Z H O U W e i , J I N Y i - h u i

(Department of Automation, Tsinghua University, Beijing 100084, China Correspondent: J I A N G Y o n g - h e n g ,
E-mail: jiangyongheng98@mails.tsinghua.edu.cn)

Abstract: A key step of Lagrangian relaxation is to optimize the dual function, and the subgradient method is frequently used when the dual function is nondifferentiable. A subgradient projection method with variable diameter is presented for the dual function of constrained programming. The efficient diameter of the dual function is determined according to the projection properties, consequently the convergence of the method is independent of any prior knowledge such as the optimal value, the diameter of the dual function definition domain, etc. The convergence is proved, and the efficiency is given. A numerical test on an assignment problem show the efficiency of the method.

Key words: dual; subgradient projection; variable diameter

1 引 言

近年来, 基于拉格朗日松弛(LR)的方法在解决复杂优化问题方面得到了研究和应用. 求解对偶问题可以获得拉格朗日乘子, 进而求解拉格朗日函数的最小化, 以获得原问题的解. 基于LR方法的关键步骤在于对偶问题的优化, 次梯度法是其常用的方案. Polyak^[1]和Allen^[2]研究了一种次梯度算法, 并证明了算法的收敛性, 但该算法依赖于对最优目标值的估计. Kim^[3]等在扩展了Polyak算法的基础

上, 提出一种变目标值的次梯度方法. Kim^[4,5]研究了凸问题优化的次梯度投影方法, 并给出了收敛效率. Kim和Kim^[4]的方法克服了对最优目标值的依赖, 但引入了对定义域直径的依赖, 这个直径既影响算法的效率又难以获得, 因而对于先验知识的依赖并没有得到解决.

本文在Kim^[4]方法的基础上提出一种变直径次梯度投影方法. 该方法克服了对除问题本身外任何先验知识的依赖, 并具有可以操作的停止准则.

收稿日期: 2003-01-24; 修回日期: 2003-06-12

基金项目: 国家自然科学基金资助项目(60174046).

作者简介: 江永亨(1974—), 男, 河北唐山人, 博士生, 从事生产系统分析和供需链管理等研究; 金以慧(1936—), 女, 浙江绍兴人, 教授, 博士生导师, 从事复杂系统的建模、控制与优化等研究.

2 基于LR的优化方案

考察约束优化问题

$$\begin{aligned} \min f(x), \\ \text{s.t. } g(x) \leq 0, x \in X. \end{aligned} \quad (1)$$

其中: X 是一个紧集, f 和 g 分别是 X 上的连续凸函数, $g(x) = L_f, g(x)$ 的维数为 r . 其对偶问题如下:

$$\max q(\lambda), \text{ s.t. } \lambda > 0$$

其中

$$q(\lambda) = \inf_{x \in X} L(x, \lambda) = \inf_{x \in X} (f(x) + \lambda^T g(x)). \quad (2)$$

称 $L(x, \lambda)$ 为拉格朗日函数, $q(\lambda)$ 为 $f(x)$ 的对偶函数, 式(2)为原问题(1)的对偶问题

下面给出基于LR的优化算法框架:

算法1(基于LR的优化方案)^[6]

Step 1: 取 $k = 0, \lambda_0 = 0$;

Step 2: 求 x_k 使得 $q(\lambda_k) = L(x_k, \lambda_k)$, 如果满足停止条件, 转 Step 4; 否则, $k = k + 1$;

Step 3: 更新 λ^k , 转 Step 2;

Step 4: x_k 和 λ_k 分别为原问题和对偶问题的目标解

算法1的关键是停止条件的选择和 λ_k 的更新. 本文以原问题和对偶问题的间距误差作为停止条件. λ_k 的更新要求解对偶问题, 对偶问题是凸紧集上的连续凹函数, 下面研究其优化算法

3 优化算法研究

算法1中, 对偶问题(2)的优化问题是关键, 下面提出一种变直径次梯度投影算法

定义1 对于凹函数 $h(x)$, 如果存在 $g(x_0)$ 使得对于任意 x 有 $h(x) \leq h(x_0) + (g(x_0) \cdot (x - x_0))$, 则称 $g(x_0)$ 为 $h(x)$ 在 x_0 的次梯度

$q(\lambda)$ 是凹函数, 任意 $\Delta \lambda$ 有 $q(\lambda + \Delta \lambda) \leq q(\lambda) + \Delta \lambda^T g(x^\lambda)$, 可知 $g(x)$ 是 $q(\lambda)$ 的次梯度函数. 为了研究 $q(\lambda)$ 的优化问题, 首先引入一个松弛算子及其重要性质^[4].

考察一个闭凸集 $C \subset R^N$ 和一个可行步长 t 使得 $0 < t_{\min} \leq t \leq t_{\max} < 2$, 定义松弛算子

$$R_{C,t}(x) = x + t(P_C(x) - x). \quad (3)$$

其中 $P_C(x)$ 为 x 在 C 上的投影, 如果 $C = \emptyset$, 则 $P_C(x) = x$.

性质1 对任意 $y \in C, x \in R^N$, 有

$$\begin{aligned} |y - R_{C,t}(x)|^2 \\ |y - x|^2 - t(2-t)|x - P_C(x)|^2 \\ |y - x|^2 - t_{\min}(2-t_{\max})d_C^2(x). \end{aligned} \quad (4)$$

其中 $d_C(x) = \inf_y |x - y|$

下面考察 $q(\lambda)$.

令 $L(q, q^{lev}) = \{\lambda | q(\lambda) = q^{lev}\}$, 则有 $M^* = L(q, q^*)$ 为最优点集. 由于 $q(\lambda)$ 的显示表达不易得到, 又考虑到次梯度的性质, 可近似表达如下:

$$\bar{q}(\lambda; \lambda) = q(\lambda) + g(x^\lambda) \cdot (\lambda - \lambda),$$

其中 $g(x^\lambda)$ 为 $q(\lambda)$ 的次梯度

令 $q^k(\lambda) = \bar{q}(\lambda; \lambda_k), H^k = \{\lambda | q^k(\lambda) = q^{lev}\}, y_k = P_{H^k}(\lambda_k)$. 考察如下投影迭代过程:

$$\lambda_{k+1} = P_+(\lambda_k + t_k(y_k - \lambda_k)). \quad (5)$$

其中: $k = 0, 1, 2, \dots; 0 < t_{\min} \leq t_k \leq t_{\max} < 2; P_+$ 是映射到正半空间的映射算子. 有

$$\lambda_{k+1} = P_+(\lambda_k + t_k \frac{q^{lev} - q(\lambda_k)g(x_k)}{|g(x_k)|^2}). \quad (6)$$

其中: $k = 0, 1, 2, \dots; 0 < t_{\min} \leq t_k \leq t_{\max} < 2$ 对式(6)应用两次式(4), 则对任意 $\lambda \in H^k \subset R^{N^+}$, 有

$$\begin{aligned} |\lambda - \lambda_{k+1}|^2 \\ |\lambda - \lambda_k|^2 - t_{\min}(2-t_{\max})d_{H^k}^2(\lambda_k) - d_{H^k}^2(\lambda_k + t_k(y - \lambda_k)). \end{aligned} \quad (7)$$

定理1 对于凹函数 $q(\lambda)$, 按式(3)进行迭代, 如果给定 λ_0 和 $q^{lev} < q^*$, 则任意 $\delta > 0$, 存在 k_m 使得 $q(\lambda_{k_m}) \leq q^{lev} - \delta$

证明 假设对于某 $\delta > 0$, 不存在 k_m 使得 $q(\lambda_{k_m}) \leq q^{lev} - \delta$, 则任意 k 有 $q(\lambda_k) > q^{lev} - \delta$. 不妨令 $q^{lev} - q(\lambda_k) = \delta = \Delta > 0$, 有

$$\begin{aligned} d_{H^k}^2(\lambda_k) = \\ (q^{lev} - q(\lambda_k))^2 / |g(x_k)|^2 > (\Delta + \delta)^2 / L_f^2. \end{aligned}$$

所以, 当 k 足够大时, $|\lambda - \lambda_{k+1}| < 0$

又由 $q^{lev} < q^*$ 和 $q(\lambda) = q^k(\lambda)$, 显然有 $M^* \subseteq H^k$, 矛盾. 结论得证.

定理2^[4] 对于凹函数 $q(\lambda)$, 按式(4)进行迭代, 给定 $d_\lambda = \sup_{\lambda, \lambda'} |\lambda - \lambda'|$ 和 q^{lev} , 如果存在 k_m 使得 $t_{\min}(2-t_{\max})d_{H^k}^2(\lambda_k) > d_\lambda^2$, 则 $q^{lev} = q^*$.

Kiviel^[4,5] 依据定理2给出了一种变目标值次梯度投影算法, 但其收敛性依赖于问题的定义域直径. 本文依据定理1和定理2给出一个新的迭代方案, 使得算法不再依赖于 d_λ .

算法2(变直径次梯度投影算法)

Step 1: 取 $k = 0; \lambda_0 = 0; \epsilon, \delta, d > 0; \rho_0 = 0; 0 < \omega < 1; k_d > 1$. 求 x_0 使得 $q(\lambda_0) = L(x_0, \lambda_0)$, 并得到相应的可行解 $x_0^f, f_0 = f(x_0^f), q_0 = q(\lambda_0), q_0^{lev} = \omega f_0 + (1 - \omega)q_0, q_0^{up} = q_0^{lev}, \bar{x} = x_0^f$

Step 2: 如果 $f_k - q_k \in \bar{x}$ 即为目标解, 算法结

束; 否则继续

$$\lambda_{k+1} = P_+ (\lambda_k + t_k (q_k^{\text{lev}} - q(\lambda_k))) / |g(x_k)|^2,$$

$$\rho_{k+1} = \rho_k + t_k (2 - t_k) d_H^k(\lambda_k) + d_{R^N}^2 (\lambda_k + t_k (P_H^k(\lambda_k) - \lambda_k)),$$

其中 $0 < t_{\min} \leq t_k \leq t_{\max} < 2$ 如果 $q(\lambda_{k+1}) = q_k$, 则 $q_{k+1} = q(\lambda_{k+1})$; 否则 $q_{k+1} = q_k$; $k = k + 1$.

Step 3: 求 x_k 使得 $q(\lambda_k) = L(x_k, \lambda_k)$, 并得到相应的可行解 $x_k^{\#}$. 如果 $f(x_k^{\#}) = f_{k-1}$, 则 $f_k = f(x_k^{\#})$, $x_k = x_k^{\#}$; 否则 $f_k = f_{k-1}$.

Step 4 1: 如果 $\rho_k > d$, 则 $q_k^{\text{up}} = \min\{q_{k-1}^{\text{lev}}, f_k\}$, $\rho_k = 0$;

Step 4 2: 如果 $\rho_k \leq d$, 则 $q_k^{\text{up}} = \min\{q_{k-1}^{\text{up}}, f_k\}$;

Step 5: $0 < \delta_k \leq \delta_{k-1}$, 如果 $q_k = q_{k-1}^{\text{lev}} - \delta_k$, 则 $q_k^{\text{up}} = f_k, d = kd$.

Step 6: $q_k^{\text{lev}} = \omega q_k^{\text{up}} + (1 - \omega)q_k$; 转 Step 2

4 算法收敛性和效率证明

算法 2 是基于 LR 的优化算法, 其收敛性和效率由 d 和 δ_k 决定, 下面进行具体讨论

引理 1 如果 $d < |\lambda_k - \lambda^*|^2$ 且对任意 k 有 $\delta_k = 0$, 则对任意 $\epsilon > 0$, 有

$$k > dL_f^2 / (t_{\min}(2 - t_{\max})\omega^2(1 - \omega)^2\epsilon^2) - ((f(x_0) - q(\lambda_0))^2 / \epsilon^2) / (2e \ln(\omega)) \Rightarrow q_k^{\text{up}} - q_k < \epsilon \quad (8)$$

证明 由于 $d < |\lambda_k - \lambda^*|^2$ 且对任意 k 有 $\delta_k = 0, q_k^{\text{up}} = q^*$. 如果算法 2 第 k 次循环第 l 次执行 Step 4 1, 则记 $k(l) = k$ 且 $k(0) = 0$ 对于 $k(l) < k(k+1)$, 有

$$\rho_k = \sum_{j=k(l)+1}^k t_j (2 - t_j) d_H^k(\lambda_j) + \sum_{j=k(l)+1}^k d_{R^N}^2 (\lambda_k + t_k (P_H^k(\lambda_k) - \lambda_k)) + \sum_{j=k(l)+1}^k t_j (2 - t_j) d_H^k(\lambda_j) + \sum_{j=k(l)+1}^k t_{\min}(2 - t_{\max}) (q_k^{\text{lev}} - q(\lambda_k))^2 / |g(x_k)|^2$$

因为 $q_k^{\text{lev}} - q(\lambda_k) = \omega q_k^{\text{up}} + (1 - \omega)q_k - q_k = \omega(q_k^{\text{up}} - q_k)$, 且对于 $k(l) < j < k(k+1)$, $q_k^{\text{up}} - q_k = q_j^{\text{up}} - q_j$, 又知道 $|g(x_k)| < L_f$, 所以 $\rho_k \leq (k - k(l)) t_{\min}(2 - t_{\max}) (\omega(q_k^{\text{up}} - q_k) / L_f)^2$.

由于 $\rho_k < d, k - k(l) > dL_f^2 / (t_{\min}(2 - t_{\max}) (\omega(q_k^{\text{up}} - q_k))^2)$, 由 Step 4 和 Step 6 知

$$q_k^{\text{up}} - q_k < \frac{(q_k^{\text{up}} - q_k) \omega}{(f(x_0) - q(\lambda_0)) \omega},$$

则给定任意 $\epsilon > 0$, 对于 $q_k^{\text{up}} - q_k > \epsilon$, 有 $m = \ln((f(x_0) - q(\lambda_0)) / \epsilon) / \ln(\omega) - ((f(x_0) - q(\lambda_0)) / \epsilon)^2 / (2e \ln(\omega))$.

再考察

$$k - k(l) > dL_f^2 / (t_{\min}(2 - t_{\max}) (\omega(q_k^{\text{up}} - q_k))^2);$$

$$k(l) - k(l-1) > dL_f^2 / (t_{\min}(2 - t_{\max}) (\omega^{m+l-1} \epsilon)^2), 0 < l \leq m;$$

$$k(m) = m + \sum_{l=1}^m (k(l) - k(l-1)) > dL_f^2 / (t_{\min}(2 - t_{\max}) \omega^2 (1 - \omega)^2 \epsilon^2) - ((f(x_0) - q(\lambda_0))^2 / \epsilon^2) / (2e \ln(\omega)).$$

所以

$$k > dL_f^2 / (t_{\min}(2 - t_{\max}) \omega^2 (1 - \omega)^2 \epsilon^2) - ((f(x_0) - q(\lambda_0))^2 / \epsilon^2) / (2e \ln(\omega)) \Rightarrow q_k^{\text{up}} - q_k < \epsilon$$

引理 2 如果 $d < |\lambda_k - \lambda^*|^2$, 则存在 k 使得 $q_k^{\text{lev}} = q_k + \delta_k$.

证明(反证法) 假设不存在 k 使得 $q_k^{\text{lev}} = q_k + \delta_k$, 则任意 k 有 $q_k < q_k^{\text{lev}} - \delta_k$. 不妨令 $q_k^{\text{lev}} - q(\lambda_k) - \delta_k = \Delta > 0$, 有 $d_H^k(\lambda_k) = (q_k^{\text{lev}} - q(\lambda_k))^2 / |g(x_k)|^2 > (\Delta + \delta_k)^2 / L_f^2$. 所以, 当 k 足够大时, $\rho_k > d$.

由 Step 4 1, 有 $q_k^{\text{up}} = \min\{q_{k-1}^{\text{lev}}, f_k\}$; 由 Step 6, 有 $q_k^{\text{lev}} = \omega q_k^{\text{up}} + (1 - \omega)q_k$. 则存在 k 使得 $q_{k+1}^{\text{lev}} - q_{k+1} = q_k^{\text{lev}} - q_k$, 矛盾. 结论得证.

定理 3 如果存在 k^* 使得 $\delta_k = 0, k \geq k^*$, 则算法 2 收敛.

证明 由引理 2 可知, 如果 $d < |\lambda_k - \lambda^*|^2$, 算法将得到 $q_k^{\text{lev}} = q_k + \delta_k$. 由 Step 5 可知, 如果 $q_k^{\text{lev}} = q_k + \delta_k$, 则 d 变大. 所以算法最终可以得到 $d < |\lambda_k - \lambda^*|^2$, 则存在 k , 使得 $d < |\lambda_k - \lambda^*|^2$ 且 $\delta_k = 0, k \geq k^*$. 由引理 1 知, 对任意 $\epsilon > 0, k - k^* > k^* > dL_f^2 / (t_{\min}(2 - t_{\max}) \omega^2 (1 - \omega)^2 \epsilon^2) - ((f(x_0) - q(\lambda_0))^2 / \epsilon^2) / (2e \ln(\omega)) \Rightarrow q_k^{\text{up}} - q_k < \epsilon$. 考虑 $q(\lambda)$ 的连续性及其定义域的紧性, 算法 2 收敛.

定理 3 保证了算法 2 的收敛性, 同时给出了算法效率的估计. 当 $d < |\lambda_k - \lambda^*|^2$ 且 $\delta_k = 0$, 算法按式(8)的效率收敛. 而整个算法的收敛效率还取决于 $d < |\lambda_k - \lambda^*|^2$ 且 $\delta_k = 0$ 的状况, 这与 d 的初始值和 δ_k 序列的选取有关. 如果 δ_k 的取值较大, 则 d 的增大会较快, d 的初值越接近 $|\lambda_k - \lambda^*|^2$ 越有利.

5 算例分析

本文通过一个具有10个节点的指派问题来说明算法的有效性,参数随机产生.图1为变直径次梯度投影算法的运算结果,图2和图3为直径固定情况下的运算结果.图1相关的参数为 $d = 0.02$, $\delta_k =$

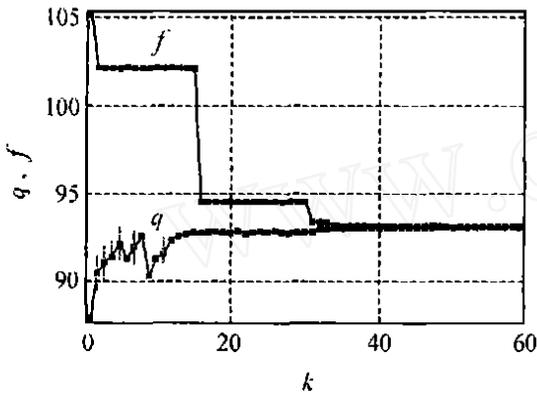


图1 变直径次梯度投影算法运算结果

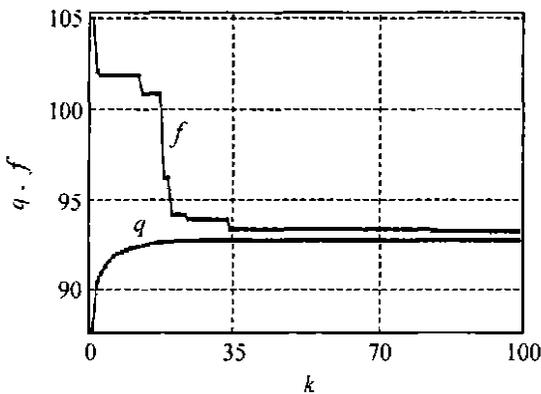


图2 小直径次梯度投影算法运算结果

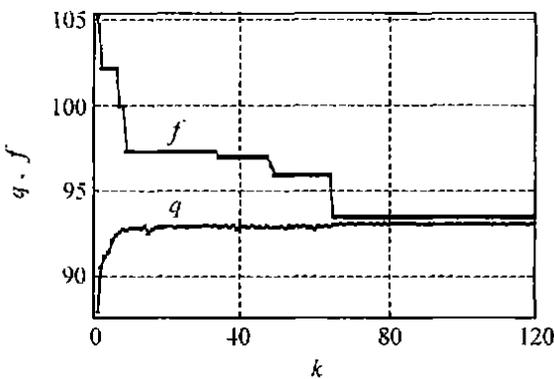


图3 大直径次梯度投影算法运算结果

$4, k_d = 2, \delta_{k+1} = 0.9\delta_k$; 图2和图3中分别取 $d = 0.02$ 和 $d = 4$

图中两条曲线分别为原问题最好可行解和对偶值的变化曲线,横坐标为执行步数.图1中,从第34步开始曲线稳定,间距误差为0.089;图2中,从第31步出现稳定现象,第81步又出现变化,间距误差为0.473;图3中,从第62步开始曲线稳定,间距误差为0.370.可见变直径次梯度投影算法可以较快地收敛于一个较好的结果.而对于直径固定的情况,如果给定直径过小则间距误差较大;如果给定直径过大则收敛速度较慢.从而说明了变直径次梯度投影算法的有效性.

6 结论

本文的变直径次梯度投影算法以原问题的值作为对偶问题的真实上界,以原问题和对偶问题的间距误差作为停止准则,在没有任何先验知识的情况下,可以求解对偶问题和原问题.算例表明,该算法的效率接近于已知准确的对偶问题定义域有效直径情况下的次梯度投影算法的算法效率.

参考文献(References):

- [1] Polyak B T. Minimization of unsmooth functionals[J]. *USSR Computational Mathematics and Mathematical Physics*, 1969, 9(3): 14-29.
- [2] Allen E, Helgason R, Kennington J, et al. A generalization of polyak's convergence result for subgradient optimization [J]. *Mathematical Programming*, 1987, 37(3): 309-317.
- [3] Kim S, Ahn H, Cho S. Variable target value subgradient method[J]. *Mathematical Programming*, 1999, 49: 359-369.
- [4] Kiwiel K C. The efficiency of subgradient projection methods for convex optimization, Part I: General level methods[J]. *SIAM J Control and Optimization*, 1996, 34(2): 660-676.
- [5] Kiwiel K C. The efficiency of subgradient projection methods for convex optimization, Part II: Implementation and optimization [J]. *SIAM J Control and Optimization*, 1996, 34(2): 677-697.
- [6] Bertsekas D P. *Nonlinear Programming* [M]. Second Edition. Belmont: Athena Scientific, 1999.