文章编号:1001-2486(2006)06-0058-05

参数化滤波器逼近问题的全局最优算法*

粟塔山,吴　翊

(国防科技大学 理学院,湖南 长沙　410073)

摘　要:基于 Lipschitz 下界估值和分枝定界技术,给出了一维参数化小波滤波器逼近问题的全局最优算法。由于充分利用了滤波器逼近问题的特点,本方法将原来的 Lipschitz 算法的线性收敛速率提高为二次收敛速率。

关键词:参数化正交小波滤波器;分枝定界;全局优化

中图分类号:O23　　**文献标识码**:A

The Global Optimization Algorithm for One-dimension Finite Wavelet Filters Approach

SU Ta-shan, WU Yi

(College of Science, National Univ. of Defense Technology, Changsha 410073, China)

Abstract:Based on Litschitz infimum estimate and the technology of branch-bound, the global optimization algorithm for one-dimension finite wavelet filter approach is presented, Taking advantage of the peculiarity of this filter approach problem, the convergence rate of Lipschitz algorithm was improved from linear to second-order.

Key words:formulated finite wavelet filters;technology of branch-bound;global optimization

近年来,P.P.Vaidyanathan[1]和 P.S.Long[2]研究发展的参数化构造方式,为小波滤波器的选择提供了一条广阔的途径。特别是对于一维紧支撑正交小波滤波器,参数化形式将它们一网打尽,即从理论上说,可以构造出满足任何要求的紧支撑正交小波滤波器。

利用参数化理论选择小波滤波器,都可归结为一个无约束或带约束的参数最优化问题。基于 P.S.Long 的参数化理论,文献[3]给出了从一组参数 $\theta=(\theta_1,\cdots,\theta_N)$ 到紧支撑正交小波滤波器系数 $h_k(\theta)$ $(k=0,\cdots,2N+1)$ 的解析表达式,它由一系列的正交变换完成,并导出了 $h_k(\theta)$ 关于 θ 的梯度和二阶导数矩阵的解析式及计算格式。这些工作为最优化计算提供了基础。然而,在参数化滤波器逼近问题中,理论分析和数值计算均表明,由于目标函数的非凸性,该优化问题存在很多的局部最优点(个数随滤波器的长度快速增长)。Taubman 和 A.Zakhor 等对此展开了很多研究[4-7],期望找到全局最优点。但这些方法和技巧都是基于从多个初始点出发的思想,没能得到一种理论上圆满和实践上有效的全局算法。本文根据 Lipschitz 全局最优化一般原理,结合滤波器参数优化问题的特点,给出了一个滤波器逼近问题的参数全局最优算法,改进了 Lipschitz 算法的下界估计,将原算法的线性收敛率提高到二阶收敛率。

1　维矩形域上的 Lipschitz 全局优化算法

设 $D=[a,b]$ 是 N 维矩形,Lipschitz 全局优化方法主要是通过分枝定界的思想,将矩形切割成子矩形并估计目标函数每个子矩形中的下界,再与函数在原矩形中的上界估值比较,去掉不可能含有全局极小的子矩形,从而形成一个收敛于全局极小的序列。目标函数的 Lipschitz 性质提供了在子矩形中的一种下界估计方式

$$f(x)\geqslant f(y)-L\delta(D)$$

* 收稿日期:2006-07-14
基金项目:国家自然科学基金资助项目(60673090)
作者简介:粟塔山(1961—),男,副教授,在职博士生。

其中 $x, y \in D$，L 是 Lipschitz 常数，$\delta(D)$是子矩形半径。迭代过程中函数在子矩形中的下界估值为

$$\mu = \max\{f(a) - L\delta(D), f(m) - \frac{1}{2}\delta(D), f(b) - L\delta(D)\} \tag{1}$$

其中，a, b, m 分别为矩形的左下角，右上角和中点。迭代开始时函数在原矩形中的初始上界估值为

$$\gamma_1 = \min\{f(a), f(b), f(m)\} \tag{2}$$

迭代过程中上界估值按

$$\gamma_{k+1} = \min\{\gamma_k, f(\tilde{b}^{(k)}), f(\hat{a}^{(k)}), f(\tilde{m}^{(k)}), f(\hat{m}^{(k)})\} \tag{3}$$

不断改善，其中 $\tilde{m}^{(k)}$，$\hat{m}^{(k)}$为分割出的子矩形 $\tilde{D}_k$，$\hat{D}_k$ 的中点。通过下式确定当前迭代点 x_{k+1}，

$$f(x_{k+1}) = \gamma_{k+1}, x_{k+1} \in \{x_k, \tilde{b}^{(k)}, \hat{a}^{(k)}, \tilde{m}^{(k)}, \hat{m}^{(k)}\} \tag{4}$$

为保证下界估值的不减性，还须按下式修改 $\tilde{D}_k$，$\hat{D}_k$ 的下界估值

$$\begin{cases} \tilde{\mu}_k \Leftarrow \max\{\mu_k, \tilde{\mu}_k\} \\ \hat{\mu}_k \Leftarrow \max\{\mu_k, \hat{\mu}_k\} \end{cases} \tag{5}$$

算法 1　连续函数 f 在 N 维矩形域 D 上全局最优的 Lipschitz 分枝定界算法

Step0　$\Theta = \{(D, \mu_1)\}$，其中，μ_1 是 f 在 D 中的初始下界估值，按式(1)计算。全局极小的初始上界估值 γ_1 按式(2)计算，并确定相应的 x_1，使 $f(x_1) = \gamma_1$，置 $\varepsilon > 0$，$k = 1$。

Step1　从 Θ 中取出具有最小下界估值的子矩形 D_k，下界估值为 μ_k。如果 $\gamma_k - \mu_k < \varepsilon$，算法终止，$x_k$ 为近似全局最小点，γ_k 为近似全局最小值。否则

Step2　对 D_k 作细分，产生两个子矩形 $\tilde{D}_k$，$\hat{D}_k$. 按(1)式计算 f 在子矩形 $\tilde{D}_k$，$\hat{D}_k$ 中的下界估值 $\tilde{\mu}_k$，$\hat{\mu}_k$。按式(3)、(4)改善全局极小的上界估值 γ_{k+1} 并确定相应的 x_{k+1}，再按式(5)修改 $\tilde{\mu}_k$，$\hat{\mu}_k$。

Step3　如果 $\tilde{\mu}_k < \gamma_{k+1}$，将$(\tilde{D}_k, \tilde{\mu}_k)$插入 Θ。如果 $\hat{\mu}_k < \gamma_{k+1}$，将$(\hat{D}_k, \hat{\mu}_k)$插入 Θ。置 $k \Leftarrow k+1$，返回 Step1。

R. Horst[8] 证明了算法产生的 $\{x_k\}_{k=1}^{+\infty}$ 的任意聚点都是全局最小点。算法的收敛率取决于 $\gamma_k - \mu_k$ 收敛于零的速度，由算法 1 可知，$r_{k_r} - \mu_{k_r} \leqslant L\delta(D_{k_r})$，可见，$\gamma_{k_r} - \mu_{k_r}$ 是 $\delta(D_{k_r})$ 的同级无穷小，或者说，$\gamma_{k_r} - \mu_{k_r}$ 关于 $\delta(D_{k_r})$ 线性收敛于零。

下面根据参数化滤波器逼近问题的特点，分别建立算法 1 中下界估值的一次模型和二次模型，一次模型使得 $\gamma_{k_r} - \mu_{k_r}$ 关于 $\delta(D_{k_r})$ 二阶收敛于零；二次模型的收敛率更优，但运算量比一次模型大。

2　滤波器逼近问题中的几个不等式

滤波器逼近问题归结为极小化问题[3]

$$\begin{cases} \min f(\theta) = -q^{\mathrm{T}} h(\theta) \\ \mathrm{st} \quad \theta \in D \end{cases}$$

其中，q 是已规范化的 $2N+2$ 维待逼近的滤波器系数向量，$h(\theta)$是 $2N+2$ 维参数化滤波器系数向量，$D = [a, b]$是 N 维矩形，$a = (0, \cdots, 0)^{\mathrm{T}}$，$b = (2\pi, \cdots, 2\pi)^{\mathrm{T}}$。

我们在文献[3]中已经得到$\nabla^{\mathrm{T}} f(\theta)$，$\nabla^2 f(\theta)$的计算公式和结论

$$\left\| \frac{\partial h(\theta)}{\partial \theta_j} \right\| \leqslant 2, \quad \left\| \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j} \right\| \leqslant 4, \quad \left\| \frac{\partial}{\partial \theta_s}\left(\frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j}\right) \right\| \leqslant 8$$

利用这些结论和文献[9－11]中各种形式的台劳展式，可以得到下面几个不等式，为节 3、4 做准备。

引理　记 $L_1 = 2\sqrt{N}$，$L_2 = 4N$，$L_3 = 8N^{\frac{3}{2}}$，那么，$\forall\, \theta, \theta^{(1)}, \theta^{(2)} \in R^N$，成立如下五式：

$$\max_{\theta} \| \nabla f(\theta) \| < L_1$$

$$\| f(\theta^{(1)}) - f(\theta^{(2)}) \| \leqslant L_1 \| \theta^{(1)} - \theta^{(2)} \|$$

$$\max_{\theta} \| \nabla^2 f(\theta) \|_F \leqslant L_2$$

$$\| \nabla f(\theta^{(1)}) - \nabla f(\theta^{(2)}) \| \leqslant L_2 \| \theta^{(1)} - \theta^{(2)} \|$$

$$\| \nabla^2 f(\theta^{(1)}) - \nabla^2 f(\theta^{(2)}) \|_F \leqslant L_3 \| \theta^{(1)} - \theta^{(2)} \|$$

其中 $\|\cdot\|_F$ 是指矩阵的 Frobenius 范数。引理中的第2式给出了算法1中的 Lipschitz 常数 L_1。

3 利用 $f(\theta)$ 的一次模型改进算法1中 Lipschitz 的下界估值

考察滤波器逼近问题的目标函数 $f(\theta)$ 在矩形 $D=[a,b]$ 上的下界估值问题。这里只考虑利用矩形的中点 m 产生的下界估值，左上角点 a 和右上角 b 可以完全平行地讨论。

因为 $f(\theta)$ 二阶连续可微，由中值定理，$\forall\theta\in D$，

$$f(\theta)=f(m)+\nabla^{\mathrm{T}}f(m)(\theta-m)+\int_0^1(1-t)(\theta-m)^{\mathrm{T}}\nabla^2 f(m+t(\theta-m))(\theta-m)\mathrm{d}t$$

所以

$$\begin{aligned}|f(\theta)-f(m)-\nabla^{\mathrm{T}}f(m)(\theta-m)| &= \left|\int_0^1(1-t)(\theta-m)^{\mathrm{T}}\nabla^2 f(m+t(\theta-m))(\theta-m)\mathrm{d}t\right|\\ &\leqslant\int_0^1(1-t)|(\theta-m)^{\mathrm{T}}\nabla^2 f(m+t(\theta-m))(\theta-m)|\mathrm{d}t\\ &\leqslant\int_0^1(1-t)\|\nabla^2 f(m+t(\theta-m))\|_F\cdot\|\theta-m\|^2\mathrm{d}t\end{aligned}$$

再利用节2引理中的第3式，可知

$$|f(\theta)-f(m)-\nabla^{\mathrm{T}}f(m)(\theta-m)|\leqslant\frac{L_2}{2}\|\theta-m\|^2$$

记 $\nabla^{\mathrm{T}}f(m)=(\lambda_1,\cdots,\lambda_N)$，则

$$\begin{aligned}f(\theta)&\geqslant f(m)+\nabla^{\mathrm{T}}f(m)(\theta-m)-\frac{L_2}{2}\|\theta-m\|^2\\ &\geqslant f(m)+\min_{\theta\in D}\{\nabla^{\mathrm{T}}f(m)(\theta-m)-\frac{L_2}{2}\|\theta-m\|^2\}f(m)+\sum_{i=1}^{N}\min_{a_i\leqslant\theta_i\leqslant b_i}\{\lambda_i(\theta_i-m_i)-\frac{L_2}{2}(\theta_i-m_i)^2\}\end{aligned}$$

上式最后的 N 个单变量极小化函数都是 θ_i 的凹函数，它的最小点必是区间端点 a_i 或 b_i，易知极小点是

$$\underline{\theta_i}=\begin{cases}a_i, & \lambda_i\geqslant 0\\ b_i, & \lambda_i<0\end{cases}$$

从而相应于 $\underline{\theta}$ 的最小值为

$$\nabla^{\mathrm{T}}f(m)(\underline{\theta}-m)-\frac{L_2}{2}\|\underline{\theta}-m\|^2=-\frac{1}{2}\sum_{i=1}^{N}|\lambda_i|(b_i-a_i)-\frac{L_2}{8}[\delta(D)]^2\triangleq\tau$$

于是，得到 $f(\theta)$ 在 D 中的下界估值 $\mu=f(m)+\tau$。

比较本节在 m 产生的下界 μ 与算法1在 m 产生的下界 $f(m)-\frac{L_1}{2}\delta(D)\triangleq\hat{\mu}$。

定理1 当 $\delta(D)$ 适当小，则有 $\mu>\hat{\mu}$，即本节得到的下界优于前一节的下界。

证明 由于文章篇幅所限，证明略。

定理2 若使用本节产生的下界 μ_k，则 $\gamma_k-\mu_k$ 关于 $\delta(D_k)$ 二阶收敛于零。

证明 由于文章篇幅所限，证明略。

4 利用 $f(\theta)$ 的二次模型改进算法1中 Lipschitz 的下界估值

考虑 $f(\theta)$ 的二次模型。设 $D=[a,b]$ 是 N 维矩形，这里只考虑利用 D 的中点 m 作估值。由基本公式和节2引理中的第5式，有

$$f(\theta)\geqslant f(m)+\nabla^{\mathrm{T}}f(\theta)(\theta-m)+\frac{1}{2}(\theta-m)^{\mathrm{T}}\nabla^2 f(m)(\theta-m)-\frac{L_3}{6}\|\theta-m\|^3$$

$$\geqslant f(m)+\min_{\theta\in D}\{\nabla^{\mathrm{T}}f(\theta)(\theta-m)+\frac{1}{2}(\theta-m)^{\mathrm{T}}\nabla^2 f(m)(\theta-m)\}-\frac{L_3}{6}[\delta(D)]^3$$

为得到 $f(\theta)$ 在 D 的下界估值，需求解 D 上的二次规划

$$\min_{\theta\in D}\{\nabla^{\mathrm{T}}f(\theta)(\theta-m)+\frac{1}{2}(\theta-m)^{\mathrm{T}}\nabla^2 f(m)(\theta-m)\}$$

当$\nabla^2 f(m)$正定,二次规划是严格凸的,否则是凹的或者不定的,可以对$\nabla^2 f(m)$作 Cholesky 分解判断$\nabla^2 f(m)$是否正定。当不正定时,求解二次规划是 NP 难的,但我们仍有退而求其次的办法(当迭代点接近最优点时,$\nabla^2 f(m)$必是正定的)。

下面分两种情况讨论 $f(\theta)$在 D 的下界估值问题。

4.1 $\nabla^2 f(m)$为正定矩阵

此时有成熟高效的紧约束集法求得正定二次规划的唯一全局最小解。紧约束集法针对约束集为一般的多胞形,N 维矩形作为其特殊情况,我们得到了简便得多的方法(由于文章篇幅所限,具体算法略)。

4.2 $\nabla^2 f(m)$不正定

当$\nabla^2 f(m)$不正定时,得到下界估值(由于文章篇幅所限,推导过程略)

$$\mu = f(m) + \sum_{k=1}^{N} \tau_k - \frac{L_3}{6}[\delta(D)]^3$$

其中,

$$\tau_k = \min_{\underline{a}_k \leqslant \eta_k \leqslant \bar{b}_k} \left(\rho_k \eta_k + \frac{\sigma_k}{2}\eta_k^2\right) = \begin{cases} -\dfrac{\rho_k^2}{2\sigma_k}, & \text{当 } \sigma_k > 0, |\rho_k| \leqslant \sigma_k \bar{b}_k \\ \dfrac{\sigma_k \bar{b}_k^2}{2} - |\rho_k| \bar{b}_k, & \text{否则} \end{cases}$$

定理 3 设 $\delta(D)$适当小。如果$\nabla^2 f(m)$正定,或者$\nabla^2 f(m)$不正定但 Π 上的二次规划在 D_P 上达到最小值,那么 $\mu > \hat{\mu}$,即二次模型的下界优于一次模型的下界。

证明 由于文章篇幅所限,证明略。

5 总结及三种下界估值的数值实例比较

滤波器逼近问题的 Lipschitz 全局最优算法中,下界估值决定着算法的收敛速率。本文列出了三种下界估值方法,第一种方法由原算法给出,第二和第三种方法是本文的结果。表 1 出一个数值算例比较三种下界估值。全局逼近问题

$$\begin{aligned} &\min f(\theta) = -q^{\mathrm{T}} h(\theta) \\ &\mathrm{st} \quad \theta \in D \end{aligned}$$

中的待逼近滤波器 q 是一个双正交滤波器,数据来自于文献[12],并且被规范化。

表 1 数值算例

Tab.1 The example

k	q_k	k	q_k
0	−6.509609258516232e−002	4	0.4217019604035378
1	−4.104071821573077e−002	5	−4.104071821573077e−002
2	0.4217019604035378	6	−6.509609258516232e−002
3	0.7952931720696389		0.0

这个问题中 $N=3, L_1=2\sqrt{N}, L_2=4N, L_3=8N^{\frac{3}{2}}$。

选择 D 中的一个子矩形$D_0=[a_0,b_0]$如下:

$$a^{(0)}=(1.7,1.7,3.7), \quad b^{(0)}=(2.13,2.13,4.25)$$

D_0 的中点是 $m=(1.915,1.915,3.975)$,直径 $\delta(D_0)=0.81993902212299$。利用中点 m,用三种方法估计 $f(\theta)$在 D_0 中的下界。

第一种下界估值:

$$\mu_1 = f(m) - \frac{L_1}{2}\delta(D_0) = -2.4174757327222$$

第二种下界估值：f 在 m 处的梯度向量

$$\nabla^2 f(m) = \begin{bmatrix} -0.00089451005081789 \\ -8.1216483070789e-005 \\ -0.00072522793352013 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}$$

$$\mu_2 = f(m) - \frac{1}{2}\sum_{i=1}^{N} |\lambda_i| (b_i^{(0)} - a_i^{(0)}) - \frac{L_2}{8}[\delta(D_0)]^2 = -2.0061589061833$$

第三种下界估值：f 在 m 处的二阶导数矩阵

$$\nabla^2 f(m) = \begin{bmatrix} 1.922956600846868 & 0.1249182436348519 & 1.582352198012299 \\ 0.1249182436348519 & 0.2502896111643027 & -9.031479167026521e-002 \\ 1.582352198012299 & -9.031479167026521e-002 & 1.491546069588058 \end{bmatrix}$$

它是正定矩阵。求解正定二次规划

$$\min_{\theta \in D_0} \{\nabla^T f(\theta)(\theta - m) + \frac{1}{2}(\theta - m)^T \nabla^2 f(m)(\theta - m)\}$$

得到其唯一最优值 $\tau = -2.0996439604488e-007$，于是

$$\mu_3 = f(m) + \tau - \frac{L_3}{6}(\delta(D_0))^3 = -1.4746920749309$$

可以看到 $\mu_3 > \mu_2 > \mu_1$。不能小看了这点差异，下界提升 0.1，就可能甩掉很多子矩形，从而加快了收敛速度。

参考文献：

[1] Vaidyanathan P P. Multirate Systems and Filter Banks[M]. Englewood Cliffs，NJ：Prentice-Hall，1993.

[2] Long P S. N Dimensional Finite Wavelet Filters[J]. Journal of Computational Mathematics，2003，21(5)：595－602.

[3] 粟塔山，吴翊. 一维参数化正交小波滤波器的解析性质与优化逼近[J]. 计算数学，2006，(4).

[4] Taubman D，Zakhor A. A Multi-start Algorithm for Signal Adaptive Subband Systems[J]. Proc. of the IEEE ICASSP－92，San Francisco，1992，3：213－216.

[5] Moulin P，Mihcak M. Theory and Design of Signal adapted FIR Parauitary Filter Banks [J]. IEEE Transactions on Signal Processing，1998，46(4).

[6] Kirac A，Vaidyanathan P P. Theory and Design of Optimum FIR Compaction Filters [J]. IEEE Transactions on Signal Processing，1998，46：903－919.

[7] Greiner T. Orthogonal and Briorthogonal Texture matched Wavelet Filter Banks for Hierarchical Texture Analysis [J]. Signal Processing，1996，54：1－22.

[8] Horst R，Pardalos P M，Thoai N V. 全局优化引论[M]. 黄红选，译. 北京 清华大学出版社，2003.

[9] 袁雅湘，孙文瑜. 最优化理论与方法[M]. 北京 科学出版社，1997.

[10] 李庆扬，莫孜中，祁力群. 非线性方程组的数值解法[M]. 北京 科学出版社，1987.

[11] 邓乃扬，等. 无约束最优化计算方法[M]. 北京 科学出版社，1988.

[12] Daubechies I. Ten Lectures on Wavelets[M]. CBMS，61，SIAM，Philadelphia，1992.