运筹学与控制论

DOI:10.11721/cqnuj20130602

解奇异无约束优化问题的改进张量法*

肖　潇，倪　勤
(南京航空航天大学 理学院，南京 210016)

摘要：给出一个解奇异无约束优化问题(极小点的 Hessian 矩阵奇异)的改进张量法。张量方法是标准牛顿模型方法的推广，它扩充目标函数的 Taylor 展式到四阶项，弥补了牛顿模型在极小点处的 Hessian 矩阵奇异时失去快速收敛性的缺陷。与标准张量法相比，本文主要的改进是，用梯度和二阶导数的差来替代函数与梯度差来构造张量模型。8 个标准函数被奇异化后进行了数值试验，数值试验结果表明这个改进张量法是有效的。

关键词：无约束优化；张量模型；奇异问题
中图分类号：O122　　**文献标志码**：A　　**文章编号**：1672-6693(2013)06-0009-04

考虑无约束最优化问题 $\min\limits_{x\in\mathbf{R}^n} f(x)$，其中 $f(x):\mathbf{R}^n\to\mathbf{R}$ 是二阶连续可微。常用的牛顿法基本思想是，利用目标函数 $f(x)$ 的二次 Taylor 展开得牛顿模型 $M_N(x_c+d)=f(x_c)+\nabla f(x_c)d+\frac{1}{2}\nabla^2 f(x_c)d^2$，进而将其极小化。如果 $\nabla^2 f(x)$ 在极小点 x_* 的邻域内非奇异，则牛顿法局部二阶收敛；然而如果 $\nabla^2 f(x)$ 在极小点 x_* 的邻域内奇异，则牛顿法往往失效[1]。

1991 年，Schnabel 与 Chow[2] 提出了解无约束优化的张量法，$f(x)$ 的 Taylor 展式扩充到四阶张量模型为

$$M_T(x_c+d)=f(x_c)+\nabla f(x_c)d+\frac{1}{2}\nabla^2 f(x_c)d^2+\frac{1}{6}T_c d^3+\frac{1}{24}V_c d^4 \tag{1}$$

其中，x_c 是当前迭代点，$d\in\mathbf{R}^n$，$\nabla f(x_c)$ 和 $\nabla^2 f(x_c)$ 是 $f(x)$ 在点 x_c 的梯度和 Hessian 阵。$T_C\in\mathbf{R}^{n\times n\times n}$ 和 $V_C\in\mathbf{R}^{n\times n\times n\times n}$，为点 x_c 处形成的三阶和四阶对称张量。这里对称张量是指元素下标的顺序不同不改变元素的值。为了简便，d^2,d^3,d^4 分别表示 $dd,ddd,dddd$，$\nabla f(x_c)d$ 表示 $\nabla f(x_c)^{\mathrm{T}}d$，$\nabla^2 f(x_c)d^2$ 表示 $d^{\mathrm{T}}\nabla^2 f(x_c)d$，三次与四次型 $T_c d^3$ 和 $V_c d^4$ 的计算格式为 $T_c d^3=\sum\limits_{i,j,k=1}^{n} t_{ijk}^c d_i d_j d_k$，$V_c d^4=\sum\limits_{i,j,k,l=1}^{n} v_{ijkl}^c d_i d_j d_k d_l$。

张量法在低秩张量计算的基础上，增强了解奇异优化问题的效率，这方面的研究参见文献[3-6]。本文便是对张量法进行改进的研究。

1 改进的张量法

为了获得模型(1)中低秩张量，Schnabel 与 Chow[2] 选择了前面 p 个不相邻的迭代点 $x_{-1},\cdots,x_{-p}$，利用这些点的函数和梯度值给出模型中张量要满足的插值条件

$$f(x_{-k})=f(x_c)+\nabla f(x_c)s_k+\frac{1}{2}\nabla^2 f(x_c)s_k^2+\frac{1}{6}T_c s_k^3+\frac{1}{24}V_c s_k^4$$

$$\nabla f(x_{-k})=\nabla f(x_c)+\nabla^2 f(x_c)s_k+\frac{1}{2}T_c s_k^2+\frac{1}{6}V_c s_k^3$$

其中 $s_k=x_{-k}-x_c$，$k=1,\cdots,p$。

这样得到的张量模型能有效解极小点是奇异的情形。然而这样得到的低秩四阶模型并没有完全应用已计算出的二阶导数矩阵信息，即插值条件中不含有二阶导数矩阵的差。本文试图改变这种情况，用前面 p 个不相邻的迭代点 $x_{-1},\cdots,x_{-p}$ 的梯度和二阶导数矩阵形成插值条件为

$$\nabla f(x_{-k})=\nabla f(x_c)+\nabla^2 f(x_c)s_k+\frac{1}{2}T_c s_k^2+\frac{1}{6}V_c s_k^3 \tag{2}$$

$$\nabla^2 f(x_{-k})=\nabla^2 f(x_c)+T_c s_k+\frac{1}{2}V_c s_k^2 \tag{3}$$

其中 $s_k=x_{-k}-x_c$，$k=1,\cdots,p$。

* 收稿日期：2013-06-06　　修回日期：2013-08-05　　网络出版时间：2013-11-20　14:46
资助项目：国家自然科学基金(No. 11071117)
作者简介：肖潇，女，硕士研究生，研究方向为线性与非线性最优化，E-mail：570637334@qq.com；通讯作者：倪勤，E-mail：niqfs@nuaa.edu.cn
网络出版地址：http://www.cnki.net/kcms/detail/50.1165.N.20131120.1446.201306.9_029.html

为了便于求解,在(3)式左右两边同时右乘 s_k 得

$$\nabla^2 f(x_{-k})s_k=\nabla^2 f(x_c)s_k+T_c s_k^2+\frac{1}{2}V_c s_k^3,k=1,\cdots,p \tag{4}$$

下面求满足(2)式和(4)式的张量 T_c 和 V_c。定义 $\alpha_k=T_c s_k^2,\beta_k=V_c s_k^3$,那么 $\alpha_k,\beta_k\in\mathbf{R}^n$, $k=1,\cdots,p$,且计算公式为 $(T_c s_k^2)_i=\sum_{i_2,i_3=1}^{n} t_{ii_2i_3}^c s_{i_2}^k s_{i_3}^k$, $(V_c s_k^3)_i=\sum_{i_2,i_3,i_4=1}^{n} v_{ii_2i_3i_4}^c s_{i_2}^k s_{i_3}^k s_{i_4}^k$, $i=1,\cdots,n$ 。

代入(2)式和(4)式得

$$\begin{cases}\frac{1}{2}\alpha_k+\frac{1}{6}\beta_k=q_{1k}\\ \alpha_k+\frac{1}{2}\beta_k=q_{2k}\end{cases} \tag{5}$$

其中 $q_{1k},q_{2k}\in\mathbf{R}^n$, $\begin{cases}q_{1k}=\nabla f(x_{-k})-\nabla f(x_c)-\nabla^2 f(x_c)\cdot s_k\\ q_{2k}=\nabla^2 f(x_{-k})\cdot s_k^2-\nabla^2 f(x_c)\cdot s_k^2\end{cases}$, $k=1,\cdots,p$。方程组(5)式是非奇异的,因此 α_k 和 β_k 被唯一确定,即

$$\begin{cases}\alpha_k=6q_{1k}-2q_{2k}\\ \beta_k=-12q_{1k}+6q_{2k}\end{cases},\ k=1,\cdots,p \tag{6}$$

为了确立张量,定义 $u\cdot v\cdot w$ 为三阶秩一张量,其下标是 ijk 的元素为 $u_iv_jw_k$,定义 $u\cdot v\cdot w\cdot x$ 为四阶秩一张量,其下标是 $ijkl$ 的元素为 $u_iv_jw_kx_l$,其中 $u,v,w,x\in\mathbf{R}^n$。由文献[5]中定理2.3,有下列低秩张量 T_c 的结论。

定理1 设 $p\leqslant n$,$\{s_k,k=1,\cdots,p\}$线性无关,$\alpha_k,k=1,\cdots,p$,由(6)式给定,那么满足 $\begin{cases}\min\limits_{T_c\in\mathbf{R}^{n\times n\times n}} \|T_c\|_F\\ \text{s.t.}\quad T_c\cdot s_k^2=\alpha_k,\ k=1,\cdots,p\end{cases}$ 的对称三阶张量 T_c 的解为

$$T_c=\sum_{k=1}^{p}(b_k\cdot s_k\cdot s_k+s_k\cdot b_k\cdot s_k+s_k\cdot s_k\cdot b_k) \tag{7}$$

其中 $b_k\in\mathbf{R}^n$,$k=1,\cdots,p$,由线性方程 $\boldsymbol{BN}+2\boldsymbol{SM}=\boldsymbol{A}$ 确立。在这个线性方程中,$B=[b_1,b_2,\cdots,b_p]$,$\boldsymbol{N},\boldsymbol{M}\in\mathbf{R}^{p\times p}$,$\boldsymbol{N}_{ij}=(s_i^{\mathrm T}s_j)^2$,$\boldsymbol{M}_{ij}=(s_i^{\mathrm T}s_j)(b_i^{\mathrm T}s_j)$,$1\leqslant i,j\leqslant p$,$\boldsymbol{A}=[\alpha_1,\alpha_2,\cdots,\alpha_p]$,$\boldsymbol{S}=[s_1,s_2,\cdots,s_p]$。

由此得到了三阶张量 T_c,而四阶张量 V_c 的确定由下面引理与定理中给出。

引理2 考虑线性方程 $\beta_i=\sum_{k=1}^{p}c_k\ (s_k^{\mathrm T}s_i)^3+3\sum_{k=1}^{p}s_k\ (s_k^{\mathrm T}s_i)^2(c_k^{\mathrm T}s_i)$,$i=1,\cdots,p$,其中 $\boldsymbol{C}=[c_1,c_2,\cdots,c_p]$是未知矩阵,$\beta_i$,$i=1,\cdots,p$,由(6)式给定。这个方程组可先转换成下列 p^2 个方程,并将未知量记成向量形式

$$\begin{bmatrix}s_1^{\mathrm T}\beta_1\\ s_1^{\mathrm T}\beta_2\\ \vdots\\ s_p^{\mathrm T}\beta_p\end{bmatrix}=\begin{bmatrix}K & \cdots & \cdots & \\ \vdots & \vdots & \vdots & \vdots\\ \vdots & \vdots & \vdots & \vdots\\ & \cdots & \cdots & K\end{bmatrix}\begin{bmatrix}y_{11}\\ y_{12}\\ \vdots\\ y_{pp}\end{bmatrix}+3\begin{bmatrix}w_{11} & & & \\ \vdots & w_{12} & & \\ w_{p1} & \vdots & & \\ & w_{p2} & \vdots & w_{1p}\\ & & & \vdots\\ & & & w_{pp}\end{bmatrix}\begin{bmatrix}y_{11}\\ y_{12}\\ \vdots\\ y_{pp}\end{bmatrix}$$

其中 $y_{ij}=c_j^{\mathrm T}s_i$ 为未知量,$w_{ij}=[(s_i^{\mathrm T}s_1)(s_1^{\mathrm T}s_j)^2,(s_i^{\mathrm T}s_2)(s_2^{\mathrm T}s_j)^2,\cdots,(s_i^{\mathrm T}s_p)(s_p^{\mathrm T}s_j)^2]$,$1\leqslant i,j\leqslant p$;求出 y_{ij},$i,j=1,\cdots,p$ 后原方程的解为$\boldsymbol{C}=(\boldsymbol{I}-3\boldsymbol{SJ})\boldsymbol{K}^{-1}$,其中 $\boldsymbol{I}=[\beta_1,\beta_2,\cdots,\beta_p]$,$\boldsymbol{J}_{ij}=(s_i^{\mathrm T}s_j)^2y_{ji}$,$\boldsymbol{K}_{ij}=(s_i^{\mathrm T}s_j)^3$,$1\leqslant i,j\leqslant p$,$\boldsymbol{S}$ 与定理1中相同。

定理3 设 $p\leqslant n$,$\{s_k,k=1,\cdots,p\}$线性无关,$\beta_k,k=1,\cdots,p$,由(6)式定义,那么满足 $\begin{cases}\min\limits_{V_c\in\mathbf{R}^{n\times n\times n\times n}} \|V_c\|_F\\ \text{s.t.}\quad V_c\cdot s_i{}^3=\beta_i,\ i=1,\cdots,p\end{cases}$ 的对称四阶张量 V_c 的解是

$$V_c=\sum_{k=1}^{p}(c_k\cdot s_k\cdot s_k\cdot s_k+s_k\cdot c_k\cdot s_k\cdot s_k+s_k\cdot s_k\cdot c_k\cdot s_k+s_k\cdot s_k\cdot s_k\cdot c_k) \tag{8}$$

其中,$c_k\in\mathbf{R}^n$,$k=1,\cdots,p$,由引理2确立。

引理2与定理3的证明可参照文献[5]中定理2.3的推导途径获得,而定理3的 V_c 结果与文献[2]中定理2.2的结果有较大的差别。

将定理1与定理3中 T_c 和 V_c 的值代入张量模型,则得到了满足插值条件(2)式与(4)式的改进张量模型为

$$\boldsymbol{M}_T(x_c+d)=f(x_c)+\nabla f(x_c)\cdot d+\frac{1}{2}\nabla^2 f(x_c)\cdot d^2+\frac{1}{2}\sum_{k=1}^{p}(b_k^{\mathrm T}d)\ (s_k^{\mathrm T}d)^2+\frac{1}{6}\sum_{k=1}^{p}(c_k^{\mathrm T}d)\ (s_k^{\mathrm T}d)^3 \tag{9}$$

接着求解上述改进张量模型。为了简便，记 $g=\nabla f(x_c)$，$H=\nabla^2 f(x_c)$。当极小点处的 Hessian 矩阵非奇异时，由局部极小点的二阶必要条件，令(9)式关于 d 的导数为 0，即$\nabla \boldsymbol{M}_T(x_c+d)=g+\boldsymbol{H}d+\boldsymbol{U}_k(d)=0$，其中 $\boldsymbol{U}_k(d)=\frac{1}{2}\sum_{k=1}^{p}(s_k^{\mathrm{T}}d)^2 b_k+\sum_{k=1}^{p}(b_k^{\mathrm{T}}d)(s_k^{\mathrm{T}}d)s_k+\frac{1}{6}\sum_{k=1}^{p}(s_k^{\mathrm{T}}d)^3 c_k+\frac{1}{2}\sum_{k=1}^{p}(c_k^{\mathrm{T}}d)(s_k^{\mathrm{T}}d)^2 s_k$，整理得

$$d=-\boldsymbol{H}^{-1}[g+\boldsymbol{U}_k(d)] \tag{10}$$

在(10)式左右两端分别左乘 $s_k^{\mathrm{T}}, b_k^{\mathrm{T}}, c_k^{\mathrm{T}}, k=1,\cdots,p$，且分别令 $\alpha_k=s_k^{\mathrm{T}}d$，$\beta_k=b_k^{\mathrm{T}}d$，$\theta_k=c_k^{\mathrm{T}}d$，$k=1,\cdots,p$ 得

$$\begin{cases} s_i^{\mathrm{T}}\boldsymbol{H}^{-1}g+\frac{1}{2}s_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}b_k\alpha_k^2+s_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k\beta_k+\frac{1}{6}s_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}c_k\alpha_k^3+\frac{1}{2}s_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k^2\theta_k+\alpha_i=0 \\ b_i^{\mathrm{T}}\boldsymbol{H}^{-1}g+\frac{1}{2}b_i^{\mathrm{T}}H^{-1}\sum_{k=1}^{p}b_k\alpha_k^2+b_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k\beta_k+\frac{1}{6}b_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}c_k\alpha_k^3+\frac{1}{2}b_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k^2\theta_k+\beta_i=0, i=1,\cdots,p \\ c_i^{\mathrm{T}}H^{-1}g+\frac{1}{2}c_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}b_k\alpha_k^2+c_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k\beta_k+\frac{1}{6}c_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}c_k\alpha_k^3+\frac{1}{2}c_i^{\mathrm{T}}\boldsymbol{H}^{-1}\sum_{k=1}^{p}s_k\alpha_k^2\theta_k+\theta_i=0 \end{cases}$$

通过求解上述 $3p$ 元 3 次方程组，可以得到 $\alpha_k,\beta_k,\theta_k,k=1,\cdots,p$。再将 $\alpha_k,\beta_k,\theta_k,k=1,\cdots,p$ 的值代入(10)式则可得到张量步 d。当极小点的 Hessian 矩阵奇异时，则求 $g+\boldsymbol{H}d+\boldsymbol{U}_k(d_c)=0$ 的极小最小二乘解，其中 d_c 是前一次迭代得到的解。

下面给出改进的张量法。

算法 1　改进张量法。步 1，计算 $g=\nabla f(x_c)$，判断是否停止，如果不停止，转步 2；步 2，计算$\nabla^2 f(x_c)$，选择 p 个过去的迭代点；步 3，根据(7)与(8)式计算张量 T_c 和 V_c；步 4，若张量模型(9)的极小近似解 d^* 是下降方向，则由线搜索求得步长 λ，令 $x_+=x_c+\lambda d^*$；若 d^* 不是下降方向，则用二次模型带线搜索算法求下一迭代点，转步 1。

注　在前 p 次迭代用二次模型；(9)式的极小解只需近似获得。

2 数值试验

本文所有的数值试验都是在 CPU 为 Intel Core2 Duo T5750，主频为 2.00 GHz，内存为 3.00 GB 的 PC 机上用 MATLAB 7.5.0(R2007b)完成的。参照文献[2]中途径，对标准函数进行奇异化。设最小化目标问题为 $f(x)=\sum_{i=1}^{m}f_i^2(x)$，其中 $f_i:\mathbf{R}^n\to\mathbf{R}$，$m$ 是函数元素的个数。令

$$\boldsymbol{F}^{\mathrm{T}}(x)=(f_i(x),\cdots,f_i(x)),\ \hat{\boldsymbol{F}}(x)=\boldsymbol{F}(x)-\boldsymbol{F}'(x_*)\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}(x-x_*)$$

其中 $\mathbf{A}\in\mathbf{R}^{n\times 1}$，$x_*$ 为 $f(x)$的极小点，$\mathbf{A}^{\mathrm{T}}=(1,0,\cdots,0)$，那么得奇异问题为$\hat{f}(x)=\frac{1}{2}\hat{\boldsymbol{F}}(x)^{\mathrm{T}}\hat{\boldsymbol{F}}(x)$，即 $\hat{f}(x)$在极小点 x_* 处二阶导数矩阵的秩至多为 $n-1$。

应用算法 1 与文献[2]中的张量法(记原算法)对来自文献[7]的 8 个标准函数奇异化后进行数值计算，为了简化和省时，取 $p=1$，且计算 10 与 30 维问题，终止条件是梯度的欧氏范数小于10^{-6}。计算结果列于表 1，其中 n 是维数，x_0 为初始迭代点，k 为迭代次数，$\|g\|$ 为最后迭代得到的梯度的欧氏范数。如果迭代次数超过 500 次即终止计算并记为 F。

表 1 表明，算法 1 能有效解较多的奇异最优化问题，在求解的问题上运算速度效率方面 2 个算法相当。
本文提出了一个解无约束优化问题的改进张量法，初步数值试验表明这一改进是有效的。改进张量法能解更多的奇异最优化问题。如果把张量模型(9)式用一些近似化成新的二次模型，那么也许会获得更好的结果。这也值得今后进一步研究。

表 1　算法 1 与原算法的数值结果比较

Tab.1　The comparison of numerical results of Algorithm 1 and original algorithm

问题	x_0^{T}	维数	k		$\|g\|$	
			原算法	算法 1	原算法	算法 1
Rosenbrock	(0.5,…,0.5)	10	81	74	8.90e−007	7.20e−007
		20	132	115	4.89e−007	5.47e−007
		30	F	162	F	9.50e−007
Powell	(3,−1,0,1,…,3,−1,0,1)	12	92	114	9.86e−007	2.76e−007
		20	97	149	8.86e−008	9.97e−007
		32	149	184	6.09e−007	6.49e−007

续表 1

问题	x_0^{T}	维数	k		$\|g\|$	
			原算法	算法 1	原算法	算法 1
Wood	(−3,−1,…,−3,−1)	12	326	246	7.56e−007	4.60e−007
		24	416	395	3.58e−007	5.98e−007
		36	F	478	F	1.82e−007
Beale	(1,1,…,1)	15	37	39	1.073e−007	9.48e−007
		30	45	38	5.53e−007	5.80e−007
		45	F	50	F	3.07e−007
Tridiagonal	(1,1,…,1)	10	29	36	6.16e−007	7.68e−007
		20	45	36	4.40e−007	3.13e−007
		30	32	40	9.50e−007	9.43e−007
Conic	(3.5,…,3.5)	10	F	239	F	9.09e−007
		20	F	148	F	9.44e−007
		30	F	255	F	5.58e−007
FH1	(0.01,…,0.01)	10	31	43	6.88e−007	7.30e−007
		20	23	62	9.93e−007	7.62e−007
		30	57	61	6.59e−007	8.95e−007
Himmelblau	(1,1,…,1)	10	23	27	2.50e−007	2.88e−007
		20	63	40	9.76e−007	7.67e−007
		30	63	138	2.08e−007	1.48e−007

参考文献:

[1] Sun W Y, Yuan Y X. Optimization theory and methods: nonlinear programming[M]. New York: Springer, 2006.

[2] Schnabel R B, Chow T. Tensor methods for unconstrained optimization using second derivatives[J]. SIAM Journal on Optimization, 1991, 1(3): 293-315.

[3] Bierlairea M, Thémansb M. Dealing with singularities in nonlinear unconstrained optimization[J]. European Journal of Operational Research, 2009, 196(1): 33-42.

[4] Bouaricha A. Tensor methods for large sparse unconstrained optimization[J]. SIAM Journal on Optimization, 1997, 7(3): 732-756.

[5] Izmailov A, Solodov M. Superlinearly convergent algorithms for solving singular equations and smooth reformulations of complementarity problems[J]. SIAM Journal on Optimization, 2002, 13(2): 386-405.

[6] Shi X, Yang L, Zhang Y. A non-monotone tensor method for unconstrained optimization problems[J]. WSEAS Transactions on Mathematics, 2012, 11(11): 1006-1017.

[7] Moré J J, Garbow B S, Hillstrom K E. Testing unconstrained optimization software[J]. ACM Transactions on Mathematical Software, 1981, 7(1): 17-41.

Operations Research and Cybernetics

A Modified Tensor Method for Singular Unconstrained Optimization

XIAO Xiao, NI Qin

(College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In this paper, we propose a modified tensor method for singular unconstrained optimization where the Hessian is singular at the minimum point. The tensor model, which is a generalization of the standard Newton model and the extension to four-order term of the Taylor expansion, fix up the weakness that the Newton model will lose the fast local convergence rate of the standard Newton method where the Hessian is singular at the minimizer. Rather than with the difference of functions and gradients, the modified tensor model is constructed with the difference of gradients and Hessian. We do the numerical experiments on eight standard test functions after singularizing. The numerical results show that the modified tensor method is effective.

Key words: unconstrained optimization; tensor model; singular problems

(责任编辑　黄　颖)