微分流形上的最优化算法

肖　刚，刘三阳，尹小艳
(西安电子科技大学 理学院，陕西 西安　710071)

摘要：指出了流形算法中利用测地线寻找最优解存在附加度量结构和计算复杂的问题，根据流形的局部与欧氏空间零点的开邻域光滑同胚这一性质，利用坐标变换把非线性等式约束优化问题转化为无约束优化问题，利用坐标变换而不是黎曼几何结构给出了函数取得极值的充分和必要条件，构造了一种映射梯度算法，并证明这种算法是线性收敛的.

关键词：微分流形；最优化算法；坐标变换
中图分类号：O221.2；O189.3　　**文献标识码**：A　　**文章编号**：1001-2400(2007)03-0472-04

Optimization algorithms on differentiable manifolds

XIAO Gang，*LIU San-yang*，*YIN Xiao-yan*
(School of Science, Xidian Univ.，Xi′an　710071，China)

Abstract:　The problem of the descent algorithm along geodesic on Riemannian manifolds is provided. A technique for the optimization algorithm for the differentiable function on differentiable manifolds is given. The constrained optimization problem is converted to the unconstrained case with the special choices of coordinate transformation and the optimality condition for the constrained optimization problem is given. Moreover, a mapping gradient method is developed and linear convergence of the mapping gradient method is established.

Key Words:　differentiable manifold；optimization algorithms；coordinate transformation

等式约束优化问题，有时可看作流形上的最优问题. 基于流形上的最优化算法始于 20 世纪七八十年代[1]. 近年来人们把这种算法应用于不同的流形解决不同的问题[2~7]. 这类算法可概括如下：首先，在流形上附加一个度量结构，使流形成为黎曼流形；然后在流形上定义一个向量场并利用测地线寻找向量场的零点. 这类算法有几点需要指出，首先，除非目标函数特别指出以外目标函数与黎曼几何之间并没有内在的联系，而流形算法附另了一个度量结构；其次，寻找向量场的零点时，测地线并非是惟一路径；最后，测地线方程的计算量很大或计算不出来. 发现在给定流形的条件下，可通过坐标变换把问题转换为欧氏空间的优化问题.

1　坐标变换

讨论一类非线性等式约束优化问题

$$\min\{f(\boldsymbol{x})\mid \boldsymbol{x}\in \mathrm{R}^n: c_i(\boldsymbol{x})=0,\quad i=1,2,\cdots,m\}\quad . \tag{1}$$

其中 $m\leqslant n$，$f(\boldsymbol{x})$ 和 $c_i(\boldsymbol{x})$ 是 R^n 上的 C^σ 可微实值函数($\sigma\geqslant 2$). 设 $c(\boldsymbol{x})$ 是 $c:\mathrm{R}^n\to\mathrm{R}^m$ 的映射，分量为 $c_i(\boldsymbol{x})$，$i=1,2,\cdots,m$. 并且假设 $c(\boldsymbol{x})$ 满足正则条件：0 是映射 c 的正则值，即对所有的点 $\boldsymbol{x}\in M=c^{-1}(0)$，$\boldsymbol{J}_c(\boldsymbol{x})\in L(\mathrm{R}^n,\mathrm{R}^m)$ 是满秩的，其中 $\boldsymbol{J}_{c(\boldsymbol{x})}$ 表示映射 $c(\boldsymbol{x})$ 的雅可比矩阵.

引理 1　参见文献[1]. 设 $c:\mathrm{R}^n\to\mathrm{R}^m$，$m<n$ 是 C^σ 可微的实值函数($\sigma\geqslant 2$)，且满足 0 是 $c(\boldsymbol{x})$ 的正则值，那么集合 $M=c^{-1}(0)$ 是 R^n 中的一个 $n-m$ 维微分流形.

收稿日期：2006-11-01
基金项目：国家自然科学基金资助(60574075)
作者简介：肖　刚(1968-)，男，西安电子科技大学博士研究生

由引理 1 可知,$M=c^{-1}(0)$ 是 R^n 中的一个 $n-m$ 维微分流形,所以式(1)实际上是微分流形上的最优问题.对于流形 M 的任意一个坐标卡 (M,U_i),总存在光滑映射 φ_i,使 U_i 与 R^{n-m} 中包含 0 点的开邻域微分同胚.由于 R^{n-m} 中包含 0 点的开邻域与 R^{n-m} 空间可以建立微分同胚.所以存在微分同胚 $\varphi_i:\mathrm{R}^{n-m}\to U_i$.这说明满足上述要求的坐标变换是存在的.特别地,若流形 $M=c^{-1}(0)$ 是紧致的,那么,存在 $r<\infty$,满足 $M=U_1\cup U_2\cup\cdots\cup U_r$,设 $F_i(\boldsymbol{u})=f(\varphi_i(\boldsymbol{u}))$,那么式(1)就转化为

$$\min_{1\leqslant i\leqslant r}\min\{F_i(\boldsymbol{u})\mid \boldsymbol{u}\in\mathrm{R}^{n-m}\}\quad,\tag{2}$$

即 r 个无约束规划问题.

例 1　考虑下述最优问题

$$\min\left\{f(x_1,x_2,x_3)\,\middle|\,x\in\mathrm{R}^3:\frac{x_1^2}{a^2}+\frac{x_2^2}{b^2}+\frac{x_3^2}{c^2}=1\right\}\quad.\tag{3}$$

可行域是 R^3 中的 2 维椭球面,如果给定椭球面一个坐标变换 $x_1=a\sin\theta\cos\varphi$, $x_2=b\sin\theta\sin\varphi$, $x_3=c\cos\theta$,就得到

$$f(x_1,x_2,x_3)=f(a\sin\theta\cos\varphi,b\sin\theta\sin\varphi,c\cos\theta):=g(\theta,\varphi)\quad,\quad 0\leqslant\theta\leqslant\pi,\ 0\leqslant\varphi\leqslant 2\pi\quad.$$

那么式(3)就变为

$$\min\{g(\theta,\varphi)\mid 0\leqslant\theta\leqslant\pi,\ 0\leqslant\varphi\leqslant 2\pi\}\quad.\tag{4}$$

这样就把椭球面(流形)上的最优问题转化为 2 维平面(R^2)上的约束优化问题,这样,就不必引入度量结构和计算测地线方程.

例 2　设有 R^3 中的 2 维球面 $S:x^2+y^2+z^2=1$,则存在 $U_1\cup U_2=S$,其中 U_1 是不包含极点(0,0,1),U_2 是不包含极点(0,0,−1).利用球极投射可得两个光滑映射[8],$\varphi_1:\mathrm{R}^2\to U_1$,其中

$$\varphi_1(u,v)=\left\{\frac{2u}{1+u^2+v^2},\ \frac{2v}{1+u^2+v^2},\ \frac{u^2+v^2-1}{1+u^2+v^2}\right\}\quad,\tag{5}$$

和 $\varphi_2:\mathrm{R}^2\to U_2$,其中

$$\varphi_2(u,v)=\left\{\frac{2u}{1+u^2+v^2},\ \frac{2v}{1+u^2+v^2},\ \frac{1-u^2-v^2}{1+u^2+v^2}\right\}\quad,\tag{6}$$

那么最优问题

$$\min\{f(x_1,x_2,x_3)\mid x\in\mathrm{R}^3:x^2+y^2+z^2=1\}\tag{7}$$

就可利用式(5),(6)的变换转化为

$$\min_{i=1,2}\ \min\{F_i(u,v)\mid (u,v)\in\mathrm{R}^2\}\quad,\tag{8}$$

其中 $F_i(u,v)=f(\varphi_i(u,v)),i=1,2$.

上述两个例子可看出,在非线性等式约束规划中,可通过坐标变换把流形上的问题转化为欧氏空间中的问题.笔者就是基于这种想法,在约束集(流形)上构建了一个坐标变换的条件下,讨论规划问题的最优性条件和最优化算法.

2　最优性条件

定理 1　设 $f(\boldsymbol{x})$ 是 $M=c^{-1}(0)$ 上的 C^σ 可微实值函数($\sigma\geqslant 2$),如果点 $\boldsymbol{x}^*$ 是式(1)的局部极小点,那么存在 $\boldsymbol{x}^*$ 的邻域 $U\subset M$ 及 U 上的任意微分同胚 $\varphi:\mathrm{R}^{n-m}\to U$,满足

$$\nabla f_{\boldsymbol{x}^*}\boldsymbol{J}_{\varphi(\boldsymbol{u}^*)}=0\quad,\tag{9}$$

其中 $\nabla f_{\boldsymbol{x}^*}$ 表示函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}^*$ 的在梯度,$\boldsymbol{J}_{\varphi(\boldsymbol{u}^*)}$ 表示函数 φ 在点 $\boldsymbol{u}^*$ 的雅可比矩阵,其中 $\boldsymbol{x}^*=\varphi(\boldsymbol{u}^*)$.

证明　因点 $\boldsymbol{x}^*\in M$,则存在 $U\subset M$,满足 $\boldsymbol{x}^*\in U$,并且存在光滑微波同胚 φ 满足 $\varphi:\mathrm{R}^{n-m}\to U$.因 $f(\boldsymbol{x})$ 是 C^σ 可微的,所以 $f(\varphi(\boldsymbol{u}))$ 也是 C^σ 可微的.又因 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的局部极小点,所以 $\boldsymbol{u}^*$ 为 $f(\varphi(\boldsymbol{u}))$ 在空间 R^{n-m} 中的局部极小点.设 $\boldsymbol{u}(t)$ 为 R^{n-m} 中任意一条经过 $\boldsymbol{u}^*=\boldsymbol{u}(0)$ 的曲线,则有

$$(\mathrm{d}/\mathrm{d}t)f(\varphi(\boldsymbol{u}(t)))\Big|_{t=0}=0\quad.$$

所以

$$(\mathrm{d}/\mathrm{d}t)f(\varphi(\boldsymbol{u}(t)))\Big|_{t=0}=\nabla f_{\boldsymbol{x}^*}J_{\varphi(\boldsymbol{u}^*)}\boldsymbol{u}'(0)=0\quad.\tag{10}$$

由 $u(t)$ 的任意性,有

$$\nabla f_{\boldsymbol{x}^*}J_{\varphi(\boldsymbol{u}^*)}=0\quad.\tag{11}$$

定理 2　设 $f(\boldsymbol{x})$ 是 $M=c^{-1}(0)$ 上 C^σ 的可微实值函数($\sigma\geqslant 2$),若点 $\boldsymbol{u}^*$ 是 $f(\varphi(\boldsymbol{u}))$ 的临界点,并且矩阵

$J_\varphi^{\mathrm{T}} H_f J_\varphi + \nabla f H_\varphi$ 在点 u^* 是正定的,其中 H_f, H_φ 分别是函数 f, φ 的 Hess 矩阵,那么点 $x^* = \varphi(u^*)$ 是 $f(x)$ 严格局部极小点.

证明 设 $u(t)$ 是 R^{n-m} 空间中的任意一条经过 $u^* = u(0)$ 的光滑曲线,满足 $x^* = \varphi(u(0))$. $f(\varphi(u(t)))$ 是关于变量 t 的($\sigma \geqslant 2$)可微函数,$f(\varphi(u(t)))$ 关于 t 的二阶导数为

$$(\mathrm{d}^2/\mathrm{d}t^2) f(\varphi(u(t))) = (\mathrm{d}/\mathrm{d}t)(\nabla f J_\varphi u'(t)) = (J_\varphi u'(t))^{\mathrm{T}} H_f (J_\varphi u'(t)) + (u'(t))^{\mathrm{T}} \nabla f H_\varphi u'(t) + \nabla f J_\varphi u''(t) \quad . \tag{12}$$

利用泰勒公式有

$$f(\varphi(u(t))) - f(\varphi(u(0))) = \nabla f_{x^*} J_{\varphi(u(0))} u'(0) t + \frac{1}{2}\{u'(0)^{\mathrm{T}} (J_{\varphi(u(0))})^{\mathrm{T}} H_{f_{x^*}} J_{\varphi(u(0))} u'(0) + (u'(0))^{\mathrm{T}} \nabla f_{x^*} H_{\varphi(u(0))} u'(0) + \nabla f_{x^*} J_{\varphi(u(0))} u''(0)\} t^2 + o(t^2) \quad . \tag{13}$$

因点 u^* 是 $f(\varphi(u))$ 临界点,所以有 $\nabla f_{x^*} J_{\varphi(u(0))} = 0$, 则有

$$f(\varphi(u(t))) - f(\varphi(u(0))) = \frac{1}{2}\{u'(0)^{\mathrm{T}} [(J_{\varphi(u(0))})^{\mathrm{T}} H_{f_{x^*}} J_{\varphi(u(0))} + \nabla f_{x^*} H_{\varphi(u(0))}] u'(0)\} t^2 + o(t^2) \quad . \tag{14}$$

又因为 $(J_{\varphi(u(0))})^{\mathrm{T}} H_{f_{x^*}} J_{\varphi(u(0))} + \nabla f_{x^*} H_{\varphi(u(0))}$ 是正定的,所以 $\exists\, \varepsilon > 0$, 当 $0 < t < \varepsilon$, $u'(t) \neq 0$ 有 $f(\varphi(u(t))) - f(\varphi(u(0))) > 0$, 由 $u(t)$ 的任意性,所以点 x^* 是 $f(x)$ 严格局部极小点.

3 映射梯度法

设 M 是紧致微分流形,由上讨论可知,式(1)可转化为 r 个无约束规划问题,现讨论单个无约束规划问题

$$\min\{ f(\varphi(u)) \mid u \in \mathrm{R}^{n-m} \} \quad . \tag{15}$$

的一种算法,称其为映射梯度法,算法如下:

step 0 选择初始点 $u_0 \in \mathrm{R}^{n-m}$, $x_0 = \varphi(u_0)$ 和向量 $d_0 = -(\nabla f_{x_0} J_{\varphi(u(0))})^{\mathrm{T}}$, $k = 0$;

step 1 如果 $\|d_k\| = \|-(\nabla f_{x_k} J_{\varphi(u_k)})^{\mathrm{T}}\| \leqslant \varepsilon$ 停止迭代,$x_k = \varphi(u_k)$ 为问题的近似最优解. 否则,转 step 2;

step 2 进行一维搜索,

$$t_k = \operatorname{argmin}\{ f(\varphi(u_k + t d_k)) \mid t > 0 \} \quad , \tag{16}$$

$$u_{k+1} = u_k + t_k d_k \quad , \tag{17}$$

$$x_{k+1} = \varphi(u_{k+1}) \quad , \tag{18}$$

$k := k + 1$, 转 step 1.

注 1 由于搜索方向 $d_k = -(\nabla f J_{\varphi(u_k)})^{\mathrm{T}}$ 是 $f(x)$ 的梯度向量 ∇f 线性映射到 R^{n-m} 空间中所得到的向量,所以称这种算法为映射梯度法.

注 2 设 $F(u) = f(\varphi(u))$, 则 $\nabla F(u) = \nabla f J_{\varphi(u)}$. 可以发现这种算法就是 R^{n-m} 空间中最速下降法,只是这里的下降方向为 $-(\nabla f J_\varphi)^{\mathrm{T}}$. 用同样的方法也可以构建与牛顿法和共轭梯度法类似的算法,这里不再一一阐述.

定理 3 设点列 $\{x_k\}$ 由上述算法得到,如果点列 $\{x_k\}$ 是无穷列,则数列 $\{f(x_k)\}$ 单调递减.

证明 设 $F(u) = f(\varphi(u))$, $u \in \mathrm{R}^{n-m}$, $d_k = -(\nabla f_{x_k} J_{\varphi(u_k)})^{\mathrm{T}}$, 由 $F(u)$ 可微性可得

$$F(u_k + t d_k) - F(u_k) = t \nabla F(u_k) d_k + o(t\|d_k\|) = t \nabla f_{x_k} J_{\varphi(u_k)} d_k + o(t\|d_k\|) = -t\|d_k\|^2 + o(t\|d_k\|) \quad , \tag{19}$$

所以当 t 充分小时,$F(u_k + t d_k) - F(u_k) < 0$, 即有 $f(x_{k+1}) - f(x_k) < 0$.

定理 4 设 $f(x)$ 是 R^n 上 C^σ 的可微实值函数,设 $F(u) = f(\varphi(u))$, $u \in \mathrm{R}^{n-m}$, 如果 $\|H_{F(u)}\| \leqslant M$, H_F 表示 $F(u)$ 的 Hess 矩阵. 则对任何给定的初始条件 $t > 0$ 和 $\varepsilon > 0$, 或算法有限终止,或 $\lim\limits_{k \to \infty} f(x_k) = -\infty$, 或 $\lim\limits_{k \to \infty} \nabla f_{x_k} J_{\varphi_{u_k}} = 0$.

证明 由函数 $F(u)$ 的泰勒公式得

$$F(\boldsymbol{u}_k+t\boldsymbol{d}_k)=F(\boldsymbol{u}_k)+t\nabla F_{\boldsymbol{u}_k}\boldsymbol{d}_k+\frac{t^2}{2}\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{H}_{F_{\boldsymbol{\theta}_k}}\boldsymbol{d}_k\leqslant F(\boldsymbol{u}_k)+t\nabla F_{\boldsymbol{u}_k}\boldsymbol{d}_k+\frac{t^2}{2}M\|\boldsymbol{d}_k\|^2\quad, \tag{20}$$

其中 $\boldsymbol{\theta}_k$ 介于 $\boldsymbol{u}_k$ 和 $\boldsymbol{u}_k+t\boldsymbol{d}_k$ 之间.

设 $\boldsymbol{d}_k=-(\nabla F_{\boldsymbol{u}_k})^{\mathrm{T}}=-(\nabla f_{\boldsymbol{x}_k}\boldsymbol{J}_{\varphi(\boldsymbol{u}_k)})^{\mathrm{T}}$,$t^*=-\nabla F_{\boldsymbol{u}_k}\boldsymbol{d}_k/(M\|\boldsymbol{d}_k\|^2)$,则

$$\begin{aligned}F(\boldsymbol{u}_k)-F(\boldsymbol{u}_k+t_k\boldsymbol{d}_k)\geqslant F(\boldsymbol{u}_k)-F(\boldsymbol{u}_k+t^*\boldsymbol{d}_k)\geqslant -t^*\nabla F(\boldsymbol{u}_k)\boldsymbol{d}_k-\frac{t^{*2}}{2}M\|\boldsymbol{d}_k\|^2=\\ \frac{1}{2M}\|\nabla F(\boldsymbol{u}_k)\|^2\quad,\end{aligned} \tag{21}$$

所以

$$F(\boldsymbol{u}_0)-F(\boldsymbol{u}_k)=\sum_{i=0}^{k-1}[F(\boldsymbol{u}_i)-F(\boldsymbol{u}_{i+1})]\geqslant\frac{1}{2M}\sum_{i=0}^{k-1}\|\nabla f_{\boldsymbol{x}_i}\boldsymbol{J}_{\varphi(\boldsymbol{u}_i)}\| \tag{22}$$

两边取极限,可得:或者 $\lim\limits_{k\to\infty}F(\boldsymbol{u}_k)=\lim\limits_{k\to\infty}f(\boldsymbol{x}_k)=-\infty$,或者 $\lim\limits_{k\to\infty}\nabla f_{\boldsymbol{x}_k}\boldsymbol{J}_{\varphi(\boldsymbol{u}_k)}=0$.

引理 2　参见文献[10].设 $F(\boldsymbol{u})=f(\varphi(\boldsymbol{u}))$,$\boldsymbol{u}\in\mathrm{R}^{n-m}$,由算法产生的序列 $\{\boldsymbol{u}_k\}$ 收敛到 $F(\boldsymbol{u})$ 的极小点 $\boldsymbol{u}^*$,且存在 $\varepsilon>0$ 和 $M>m>0$,使得当 $\|\boldsymbol{u}-\boldsymbol{u}^*\|<\varepsilon$,有

$$m\|\boldsymbol{u}\|^2\leqslant\boldsymbol{u}^{\mathrm{T}}\boldsymbol{H}_{F_{\boldsymbol{u}}}\boldsymbol{u}\leqslant M\|\boldsymbol{u}\|^2\quad, \tag{23}$$

则 $\{\boldsymbol{u}_k\}$ 是线性收敛的.

推论 1　条件同引理 2,则由算法产生的序列 $\{\boldsymbol{x}_k\}$ 是线性收敛的.

证明　因 $\{\boldsymbol{u}_k\}$ 是线性收敛的,所以 $\|\boldsymbol{u}-\boldsymbol{u}^*\|\leqslant K\theta^k$,$\theta<1$,$K=\left(\dfrac{2[F(\boldsymbol{u})-F(\boldsymbol{u}^*)]}{m}\right)^{1/2}$,见文[9].当 $\boldsymbol{u}_k\to\boldsymbol{u}^*$,存在 $\boldsymbol{u}^*$ 的闭邻域 $\bar{U}_{\boldsymbol{u}^*}$,$\boldsymbol{u}_k\in\bar{U}_{\boldsymbol{u}^*}$,因 φ 满足局部 Lipschitz 条件,所以有

$$\|\boldsymbol{x}_k-\boldsymbol{x}^*\|=\|\varphi(\boldsymbol{u}_k)-\varphi(\boldsymbol{u}^*)\|\leqslant LK\theta^k\quad, \tag{24}$$

即序列 $\{\boldsymbol{x}_k\}$ 是线性收敛的.

例 3　求函数 $f(x)=x^2+y^2$ 在球面 $S:x^2+y^2+y^2=1$ 上的极小值.

首先给出覆盖球面 S 的坐标卡,由例 2 得到 U_1 和 U_2 分别是不包含(0,0,1)和(0,0,−1)的两个覆盖以及球极投射 $\varphi_i:\mathrm{R}^2\to U_i$,$i=1,2$.则函数 $f(x)=x^2+y^2$ 变换为

$$F_1(u,v)=f(\varphi_1(u,v))=4\frac{u^2+v^2}{(1+u^2+v^2)^2}\quad,$$

和

$$F_2(u,v)=f(\varphi_2(u,v))=4\frac{u^2+v^2}{(1+u^2+v^2)^2}\quad,$$

那么求极小值问题就变换为

$$\min_{i=1,2}\min\{F_i(u,v)\mid(u,v)\in R^2\}\quad.$$

函数 F_i,$i=1,2$ 的梯度为

$$\nabla F_i(u,v)=\left\{\frac{8u(1-u^2-v^2)}{(1+u^2+v^2)^3},\frac{8v(1-u^2-v^2)}{(1+u^2+v^2)^3}\right\}\quad,\quad i=1,2\quad.$$

在映射梯度法中,设 $\boldsymbol{d}_k=-\nabla F_i(u_k,v_k)$,$(u_{k+1},v_{k+1})=(u_k,v_k)+t_k\boldsymbol{d}_k$,给定初始迭代点 $(u_0,v_0)=(1,1)$,利用现有的软件包,可以计算得到(0,0)为 F_i,$i=1,2$ 的极小点,对应到球面上为点 $(0,0,1)\in U_1$ 和 $(0,0,-1)\in U_2$.

参考文献:

[1] Gabay D. Minimizing a Differentiable Function over a Differential Manifold[J]. Journal of Optimization Theory and Applications, 1982, 37(2): 117-217.

[2] Oweren B, Welfert W. The Newton Iteration on Lie Groups[J]. BIT Numerical Mathematics, 2000, 40(1): 121-145.

[3] Mathony R, Manton J H. The Geometry of the Newton Method on Non-compact Lie Groups[J]. Journal of Global Optimization, 2002, 23(3-4): 309-327.

[4] Absol P A, Mathony R, Sepulchre R. Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation[J]. Acta Applicandae Mathematicae, 2004, 80(2): 199-220.

[5] Chang H C. Optimizarion on Nonlinear Surface[D]. West Lafayette: Purdue University, 2000: 6-28.

(下转第 480 页)