梯度投影下降算法*

费 景 高

THE GRADIENT PROJECTION DESCENT ALGORITHMS

Fei Jing-gao

Abstract

This paper considers the optimization problems with equality constraints. A new idea for constructing algorithms of the gradient projection type is presented. By means of this idea, the gradient projection restoration, gradient projection Newton and conjugate gradient projection restoration algorithms are constructed. For these algorithms, we also give the theorems of global convergence and the rates of termination convergence.

§1. 引　论

考虑具有等式约束的非线性规划问题:

$$\min\{f(z) \mid \varphi(z) = \theta\}, \tag{1.1}$$

其中 z 是 n 维欧氏空间 E^n 中的点，θ 表示各个空间的零元. $f(\cdot)$ 是由 E^n 到 E^1 中的函数,称作目标函数. $\varphi(\cdot) = (\varphi^1(\cdot), \cdots, \varphi^r(\cdot))^T$ 是由 E^n 到 $E^r(r < n)$ 中的映象,称作约束映象. 集合 $S = \{z \in E^n \mid \varphi(z) = \theta\}$ 是问题(1.1)的能行集，它是 E^n 空间中的一个曲面.

相当多的优化问题本身就具有(1.1)的形式. 当问题具有不等式约束时，可以引进附加的变量,将其归化到问题(1.1);或者在计算时只考虑约束集中的作用约束，在一定程度上也可将其看成只具有等式约束的问题. 由此可见,构造和研究求解问题(1.1)的算法是有一定的普遍意义的.

将求解无约束优化问题的梯度法移植到(1.1),可得到梯度投影方法[1,2]. 从 1960 年以来，它一直是求解问题(1.1)的重要算法. 但是对这个算法的研究是很不完善的. Luenberger[3] 给出了这个算法的终端收敛速度，但它应用了曲面的测地线这个非初等概念,并且仅对理想的情形给出了证明. 在算法的实现方面，Mukai 等[4]从较一般的观点提

* 1979 年 6 月 26 日收到.

出了一种实现格式，但仅给出了收敛性的结果．对于其它的利用梯度的下降方法，例如牛顿法、共轭梯度法、变度量法等，也仅对 $\varphi(\cdot)$ 为线性的情形作了较详细的研究[5].

本文对求解问题(1.1)的梯度投影类型的方法提出一种新的构造思想．首先，我们利用曲面的恢复算子构造了曲面的局部坐标系，将(1.1)转化成局部无约束的优化问题，可以用无约束优化方法来求解．然后，我们对所用方法中的量和计算公式引进适当的误差，即对方法进行摄动，将它们转化成与局部坐标系无关的仅与梯度投影和计算点 z 有关的量与公式． 特别是计算点 z 可以不在曲面上．从而构造出在计算机上可以实现的算法．这种构造思想的好处是可以通过摄动理论将无约束优化算法的研究成果直接应用到问题(1.1)．下面的各节将对这种构造思想及构造的算法作简要的介绍．

§2. 曲面的局部坐标系及近似曲线

作下面的假定

假定 2.1　设映象 $\varphi(\cdot)$ 满足下面的条件：

1° 对于正数 ε_φ，在含集合 $D=\{z\,|\,\|\varphi(z)\|<\varepsilon_\varphi\}$ 的空间 E^n 的一个凸域 D_c 中，映象 $\varphi(\cdot)$ 具有连续有界的一阶二阶和三阶 Fréchet 导算子 $G_{\varphi1}(\cdot)$，$G_{\varphi2}(\cdot)$ 和 $G_{\varphi3}(\cdot)$，它们的范数在 D_c 上的界分别为 $N_{\varphi1}$，$N_{\varphi2}$，$N_{\varphi3}$.

2° $G_{\varphi3}(\cdot)$ 在 D_c 上满足常数为 L_φ 的 Lipschitz 条件．

3° 对 $z\in D_c$，逆矩阵 $[G_{\varphi1}(z)G_{\varphi1}^T(z)]^{-1}$ 存在，并且存在正数 β，使对所有 $z\in D_c$ 有

$$\|[G_{\varphi1}(z)G_{\varphi1}^T(z)]^{-1}\|\leqslant\beta.$$

为了简化下文的叙述，假定 ε_φ 满足不等式：

$$\alpha_\varphi=\frac{1}{2}N_{\varphi2}N_{\varphi1}^2\beta^2\varepsilon_\varphi<1. \tag{2.1}$$

作由 $D_c\subset E^n$ 到 E^n 中的映象 $R(\cdot)$，对任何 $z\in D_c$，

$$R(z)=z-G_{\varphi1}^T(z)[G_{\varphi1}(z)G_{\varphi1}^T(z)]^{-1}\varphi(z). \tag{2.2}$$

称它为曲面 S 的恢复算子．使用类似对牛顿法的处理，可以证明，当 $z_0\in D_c$ 时，由程序

$$z_{i+1}=R(z_i),\qquad i=0,1,\cdots$$

构造的序列 $\{z_i\}_{i=0}^\infty$ 将收敛到曲面 S 上的点 z^R，并且存在与 z_0 无关的正数 A_φ，使有估计式：

$$\|z^R-z_i\|\leqslant A_\varphi\|\varphi(z_i)\|,\qquad i=0,1,\cdots, \tag{2.3}$$

$$\|\varphi(z_i)\|\leqslant\alpha_\varphi^{2^i-1}\|\varphi(z_0)\|,\qquad i=0,1,\cdots. \tag{2.4}$$

在过任意点 $\tilde{z}\in S$ 的 S 的切子空间 $T_0(\tilde{z})=\{h\in E^n\,|\,G_{\varphi1}(\tilde{z})h=\theta\}$ 中选取 $m(m=n-r)$ 个相互正交的单位向量 $h_1,\cdots,h_m$，作 $n\times m$ 矩阵 $H_{\tilde{z}}=[h_1,\cdots,h_m]$．于是 S 在 $\tilde{z}$ 处的切平面 $T_1(\tilde{z})$ 上的点均可表示为

$$z=z^0(\mu,\tilde{z})=\tilde{z}+H_{\tilde{z}}\mu,$$

其中 $\mu\in E^m$．当

$$\mu\in D_\mu=\left\{\mu\in E^m\,\middle|\,\frac{1}{2}N_{\varphi2}\|\mu\|^2\leqslant\varepsilon_\varphi\right\}$$

时，由程序

$$z^{i+1}(\mu,\tilde{z})=R(z^i(\mu,\tilde{z})),\qquad i=0,1,\cdots \tag{2.5}$$

构造的序列 $\{z^i(\mu,\tilde{z})\}_{i=0}^{\infty}$ 将收敛到 S 上的点．将其记为 $z(\mu,\tilde{z})$．$z(\mu,\tilde{z})$ 是由 D_μ 到 S 上的一个映象．对这个映象，我们证明了下面的命题．

命题 2.2 设假定 2.1 成立，则映象 $z(\mu,\tilde{z})$ 有下面的性质：

1° 存在与 $\tilde{z}$ 无关的正数 ρ, N_{z_1}, N_{z_2} 和 $L_{z\rho}$，使得当 $\mu\in D_\rho=\{\mu\in E^m\,|\,\|\mu\|\leqslant\rho\}$ 时，映象 $z(\mu,\tilde{z})$ 在 μ 处二阶连续 Fréchet 可微，并有估计

$$\left\|\frac{\partial z(\mu,\tilde{z})}{\partial\mu}\right\|\leqslant N_{z_1},\qquad \left\|\frac{\partial^2 z(\mu,\tilde{z})}{\partial\mu\partial\mu}\right\|\leqslant N_{z_2}.$$

$\dfrac{\partial^2 z(\mu,\tilde{z})}{\partial\mu\partial\mu}$ 在 D_ρ 上还满足常数为 $L_{z\rho}$ 的 Lipschitz 条件．

2° $$\left.\frac{\partial z(\mu,\tilde{z})}{\partial\mu}\right|_{\mu=\theta}=H_{\tilde{z}}, \tag{2.6}$$

$$\left.\frac{\partial^2 z(\mu,\tilde{z})}{\partial\mu\partial\mu}\right|_{\mu=\theta}=-G_{\varphi1}^T(\tilde{z})[G_{\varphi1}(\tilde{z})G_{\varphi1}^T(\tilde{z})]^{-1}\left.\frac{\partial^2\varphi(z^0(\mu,\tilde{z}))}{\partial\mu\partial\mu}\right|_{\mu=\theta}. \tag{2.7}$$

3° 存在与 $\tilde{z}$ 无关的正数 $l_{\mu z}$, $N_{\mu1}$, $N_{\mu2}$，对于曲面 S 上的每个点 $\tilde{z}$，均可建立由

$$\delta(\tilde{z},l_{\mu z})=\{z\in E^n\,|\,\|z-\tilde{z}\|\leqslant l_{\mu z}\}$$

到 E^m 中的二阶连续 Fréchet 可微映象 $\mu(z,\tilde{z})$，有 $\mu(\tilde{z},\tilde{z})=\theta$ 和当 $z\in S\cap\delta(\tilde{z},l_{\mu z})$ 时有 $\mu(z,\tilde{z})\in D_\rho$, $z(\mu(z,\tilde{z}),\tilde{z})=z$．当 $z\in\delta(\tilde{z},l_{\mu z})$ 时，有估计

$$\left\|\frac{\partial\mu(z,\tilde{z})}{\partial z}\right\|\leqslant N_{\mu_1},\quad \left\|\frac{\partial^2\mu(z,\tilde{z})}{\partial z\partial z}\right\|\leqslant N_{\mu_2}.$$

对于 S 在 $z\in S\cap\delta(\tilde{z},l_{\mu z})$ 处的 $T_0(z)$ 中的向量 h，有

$$\frac{\partial\mu(z,\tilde{z})}{\partial z}h=\left[\left[\frac{\partial z(\mu,\tilde{z})}{\partial\mu}\right]^T\frac{\partial z(\mu,\tilde{z})}{\partial\mu}\right]^{-1}\left.\left[\frac{\partial z(\mu,\tilde{z})}{\partial\mu}\right]^T\right|_{\mu=\mu(z,\tilde{z})}h. \tag{2.8}$$

4° 存在与 $\tilde{z}$ 无关的正数 l_{zH} 和 $L_{z\mu}$，使在集合 $\delta(\tilde{z},l_{zH})\cap S$ 中的任意点 z 处，可以选取适当的 $n\times m$ 矩阵 H_z，使建立的映象 $z(\mu,z)$ 满足下面的性质，对任意 $\mu\in D_\rho$ 和任意 z_1, $z_2\in\delta(\tilde{z},l_{zH})\cap S$，有

$$\|z(\mu,z_1)-z(\mu,z_2)\|\leqslant L_{z\mu}\|z_1-z_2\|,$$

$$\left\|\frac{\partial z(\mu,z_1)}{\partial\mu}-\frac{\partial z(\mu,z_2)}{\partial\mu}\right\|\leqslant L_{z\mu}\|z_1-z_2\|,$$

$$\left\|\frac{\partial^2 z(\mu,z_1)}{\partial\mu\partial\mu}-\frac{\partial^2 z(\mu,z_2)}{\partial\mu\partial\mu}\right\|\leqslant L_{z\mu}\|z_1-z_2\|.$$

我们称映象 $z(\mu,\tilde{z})$ 为曲面 S 的局部参数表示（局部坐标系），μ 称作参数．将 $z=z(\mu,\tilde{z})$ 代入 $f(z)$，可得到 μ 的函数 $F(\mu,\tilde{z})=f(z(\mu,\tilde{z}))$．这样就可将问题 (1.1) 转化成局部无约束的极小化问题

$$\min_{\mu\in E^m}F(\mu,\tilde{z}). \tag{2.9}$$

为了在计算机上实现下降类型的算法，我们还需要曲面 S 的一类近似曲线．设给定 $\hat{z}\in E^n$ 和满足 $G_{\varphi1}(\hat{z})h=\theta$ 的方向 $h\in E^n$．不要求 $\hat{z}$ 在曲面 S 上．作射线

$$z^0(\lambda)=\hat{z}+\lambda h,\qquad \lambda\geqslant 0.$$

再选取一个充分小的正数 $\tilde{\lambda}$，计算系数向量

$$a=\frac{1}{\tilde{\lambda}^2}G_{\varphi 1}^T(z^0(\tilde{\lambda}))[G_{\varphi 1}(z^0(\tilde{\lambda}))G_{\varphi 1}^T(z^0(\tilde{\lambda}))]^{-1}[\varphi(\hat{z})-\varphi(z^0(\tilde{\lambda}))]. \tag{2.10}$$

作二次曲线

$$z^{R0}(\lambda)=R(\hat{z})+\lambda h+\lambda^2 a. \tag{2.11}$$

由 $z^{R0}(\lambda)$ 出发，构造近似曲线的序列 $\{z^{Rk}(\lambda)\}_{k=0}^{\infty}$ 为

$$z^{Rk+1}(\lambda)=R(z^{Rk}(\lambda)) \qquad k=0,1,\cdots. \tag{2.12}$$

这个序列，有下面的收敛性和微分性质.

命题 2.3 设假定 2.1 成立，则对于任意给定的正数 H 和 ε_H，存在与 $\hat{z}$ 和 h 无关的正数 A_0^φ、A_1^φ、A_2^φ、$\varepsilon_{H\varphi 1}$、$\varepsilon_{\varphi 1}$、A_0^R、A_1^R、A_2^R、L_z，使序列 $\{z^{Rk}(\lambda)\}_{k=0}^{\infty}$ 有性质：

1° 对于任何 $\hat{z}\in D$，当有 $\|h\|\leqslant H$，$\|\lambda h\|\leqslant\varepsilon_H$ 和集合

$$\{z=z^{R0}(t)\,|\,t\in[0,\lambda]\}\subset D_\varepsilon$$

时，有估计式

$$\begin{aligned}\|\varphi(z^{R0}(\lambda))\|&\leqslant A_0^\varphi(\|\varphi(\hat{z})\|^2+\|\lambda h\|\|\varphi(\hat{z})\|+\|\lambda h\|^3+\|\lambda h\|^2\|(\lambda-\tilde{\lambda})h\|)\\&=E_0(\hat{z},h,\lambda);\\\left\|\frac{d\varphi(z^{R0}(\lambda))}{d\lambda}\right\|&\leqslant A_1^\varphi(\|\varphi(\hat{z})\|+\|\lambda h\|^2+\|\lambda h\|\|(\lambda-\tilde{\lambda})h\|)\|h\|\\&=E_1(\hat{z},h,\lambda),\\\left\|\frac{d^2\varphi(z^{R0}(\lambda))}{d\lambda^2}\right\|&\leqslant A_2^\varphi(\|\varphi(\hat{z})\|+\|\lambda h\|+\|(\lambda-\tilde{\lambda})h\|)\|h\|^2\\&=E_2(\hat{z},h,\lambda).\end{aligned}$$

2° 当 $\|h\|\leqslant H$，$\|\varphi(\hat{z})\|\leqslant\varepsilon_{\varphi 1}$ 和 $\|\lambda h\|\leqslant\varepsilon_{H\varphi 1}$ 时，序列 $\{z^{Rk}(\lambda)\}_{k=0}^{\infty}$ 一致收敛到 $z^R(\lambda)\in S$，$z^R(\lambda)$ 对 λ 二次连续可微，并且对任何 λ，λ_2，$\lambda\in[0,\varepsilon_{H\varphi 1}/\|h\|]$ 和

$$P=R1,R2,\cdots,R,$$

有估计

$$\begin{aligned}&\|z^P(\lambda)-z^{R0}(\lambda)\|\leqslant A_0^R E_0(\hat{z},h,\lambda),\\&\left\|\frac{dz^P(\lambda)}{d\lambda}-\frac{dz^{R0}(\lambda)}{d\lambda}\right\|\leqslant A_1^R E_1(\hat{z},h,\lambda),\\&\left\|\frac{d^2z^P(\lambda)}{d\lambda^2}\bigg|_{\lambda=\lambda_1}-\frac{d^2z^P(\lambda)}{d\lambda^2}\bigg|_{\lambda=\lambda_2}\right\|\leqslant L_z\|h\|^3|\lambda_1-\lambda_2|.\end{aligned}$$

命题 2.2 和 2.3 是我们构造求解(1.1)算法时，对曲面 S 进行处理的基础. 若将 (2.5) 和 (2.12) 式分别改成

$$z^{i+1}(\mu,\tilde{z})=z^i(\mu,\tilde{z})-G_{\varphi 1}^T(\tilde{z})[G_{\varphi 1}(\tilde{z})G_{\varphi 1}^T(\tilde{z})]^{-1}\varphi(z^i(\mu,\tilde{z})),$$

$$z^{Rk+1}(\lambda)=z^{Rk}(\lambda)-G_{\varphi 1}^T(\hat{z})[G_{\varphi 1}(\hat{z})G_{\varphi 1}^T(\hat{z})]^{-1}\varphi(z^{Rk}(\lambda)),$$

则可将假定 2.1 的 1° 中的三阶导算子项 $G_{\varphi 3}(\cdot)$ 去掉，而将 2° 中的 $G_{\varphi 3}(\cdot)$ 改成 $G_{\varphi 2}(\cdot)$. 本文为叙述简单起见，仍用 (2.5)(2.12).

§3. 梯度投影恢复算法

设下面的假定成立

假定 3.1. 1° 假定 2.1 成立.

2° 函数 f 在 D_c 上具有有界连续的一阶二阶 Fréchet 导数.

应用梯度法求解问题 (2.9). $F(\mu, \tilde{z})$ 的梯度为

$$\nabla F(\mu, \tilde{z}) = \left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T \nabla f(z(\mu, \tilde{z}))$$

$$= \left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T P_\varphi(z(\mu, \tilde{z}))\nabla f(z(\mu, \tilde{z})). \tag{3.1}$$

其中 $P_\varphi(z)$ 为 E^n 空间在其子空间 $T_0(z) = \{h \in E^n | G_{\varphi1}(z)h = \theta\}$ 上的正交投影矩阵,有

$$P_\varphi(z) = I_n - G_{\varphi1}^T(z)[G_{\varphi1}(z)G_{\varphi1}^T(z)]^{-1}G_{\varphi1}(z). \tag{3.2}$$

I_n 是 $n \times n$ 单位矩阵. 由 (2.8) 式,当 $z = z(\mu, \tilde{z})$ 时,

$$P_\varphi(z)\nabla f(z) = \left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\left[\left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^{-1}\left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T\right]P_\varphi(z)\nabla f(z). \tag{3.3}$$

在 E^m 空间沿负梯度 $-\nabla F(\mu, \tilde{z})$ 进行寻找,反映到 E^n 空间是沿切线为

$$-\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\nabla F(\mu, \tilde{z})$$

的曲面的曲线进行寻找. 由(3.1)、(3.3)式,当 $z = z(\mu, \tilde{z})$ 时,可得

$$\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\nabla F(\mu, \tilde{z}) - P_\varphi(z)\nabla f(z)$$

$$= \frac{\partial z(\mu, \tilde{z})}{\partial \mu}\left\{I_m - \left[\left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^{-1}\right\}\left[\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right]^T P_\varphi(z)\nabla f(z). \tag{3.4}$$

其中 I_m 是 $m \times m$ 单位矩阵. 由于

$$I_m = \left[\left(\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\right)^T\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\bigg|_{\mu=\theta}\right]^{-1},$$

所以由 (3.4) 和命题 2.2,可以用 $-P_\varphi(z)\nabla f(z)$ 来代替方向

$$-\frac{\partial z(\mu, \tilde{z})}{\partial \mu}\nabla F(\mu, \tilde{z})$$

作为在 E^n 中的寻找方向,引起误差的量级为 $\|\mu\|\|P_\varphi(z)\nabla f(z)\|$. 再允许 z 按适当的精度离开曲面,并按 § 2 的方式构造近似曲线. 我们构造了下面的梯度投影恢复算法.

算法 3.2 (梯度投影恢复算法). 选取充分小的正数 $\varepsilon_{R\varphi1}$ 和 ε_h, 以保证步 5 中构造的曲线段上的点在恢复算子下一致收敛. 选取 $\varepsilon_{Rg} \in (0, \frac{1}{2})$、正数 $\varepsilon_{R\varphi2}$ 和 $d_1 \in (0, \frac{1}{2})$, $d_2 \in (\frac{1}{2}, 1)$, $\varepsilon_{\Delta f} \in (0, 1)$. 并选取满足 $\varepsilon_{Rg} < \varepsilon_\rho < 1$ 的充分小的正数 ε_ρ.

步 0 选取 $z_0^0 \in D$, 令 $i = 0$.

步 1 令 $\varepsilon_{hv} = \varepsilon_h$, $j = 0$.

步 2 计算 $z_i^{j+1} = R(z_i^j)$, $\varphi(z_i^{j+1})$, $P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})$.

步 3 若存在正数 ε_g, 及对所有 $S \geqslant j + 1$, 有

$$\|P_\varphi(z_i^S)\nabla f(z_i^S)\| \geqslant \varepsilon_g,$$

$$\|P_\varphi(z_i^S)\nabla f(z_i^S) - P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\| \leqslant \varepsilon_{Rg}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|,$$

$$\|\varphi(z_i^S)\| \leqslant \min\{\varepsilon_{R\varphi 1},\ \varepsilon_{R\varphi 2}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\},$$

则转步 4；否则令 $j = j+1$，转步 2.

步 4　令

$$z_i = z_i^{j+1}, h_{z_i} = -P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1}).$$

步 5　在 § 2 中令 $\hat{z} = z_i$, $h = h_{z_i}$，构造曲线 $z^{R0}(\lambda)$. 并将其记成 $z_i^{R0}(\lambda)$, $\lambda \in [0, \varepsilon_{hv}/\|h_{z_i}\|]$.

步 6　若有

$$\langle \nabla f(z_i^{R0}(0), h_{z_i}\rangle \leqslant -(1-\varepsilon_\rho)\|h_{z_i}\|^2,$$

转步 7；否则令 $j = j+2$, $z_i^j = z_i^{R0}(0)$，转步 2.

步 7　在集合 $[0, \bar{\lambda}_i] \cap [0, \varepsilon_{hv}/\|h_{z_i}\|]$ 中寻找满足

$$d_1 \leqslant \frac{f(z_i^{R0}(\lambda_i)) - \lambda_i\langle \nabla f(z_i^{R0}(0)),\ h_{z_i}\rangle - f(z_i^{R0}(0))}{-\lambda_i\langle \nabla f(z_i^{R0}(0)),\ h_{z_i}\rangle} \leqslant d_2$$

的 λ_i，若这样的 λ_i 不存在，取 $\lambda_i = \varepsilon_{hv}/\|h_{z_i}\|$. $\bar{\lambda}_i$ 为

$$\bar{\lambda}_i = \sup\{\bar{\lambda} \geqslant 0 \mid \text{当 } \lambda \in [0, \bar{\lambda}] \text{ 时 } f(z_i^{R0}(\lambda)) \leqslant f(z_i^{R0}(0))\}.$$

步 8　令

$$l = 0,\ \triangle f_i^{R0} = f(z_i^{R0}(0)) - f(z_i^{R0}(\lambda_i)).$$

步 9　计算

$$z_i^{Rl+1}(0) = R(z_i^{Rl}(0)),$$
$$z_i^{Rl+1}(\lambda_i) = R(z_i^{Rl}(\lambda_i)),$$
$$\triangle f_i^{Rl+1} = f(z_i^{Rl+1}(0)) - f(z_i^{Rl+1}(\lambda_i)).$$

步 10　若对所有的 $S \geqslant 1$，成立不等式

$$\triangle f_i^{Rl+S} \geqslant \varepsilon_{\Delta f}\triangle f_i^{R0},$$

转步 11；若无法判定，令 $l = l+1$，转步 9；若不成立，取

$$z_i = z_i^{Rl+1}(0),$$
$$h_{z_i} = -P_\varphi(z_i^{Rl+1}(0))\nabla f(z_i^{Rl+1}(0)),$$
$$\varepsilon_{hv} = \frac{1}{2}\varepsilon_{hv},$$

转步 5.

步 11　令

$$z_{i+1}^0 = z_i^{Rl+1}(\lambda_i), \qquad i = i+1,$$

转步 1.

由假定 3.1 和命题 2.2, 2.3 可知，算法 3.2 在步 2—步 6 之间，步 5—步 10 和步 9—步 10 之间不可能无限循环.

在具体计算时，正数 ε_h 和 $\varepsilon_{R\varphi 1}$ 可通过过程

$$\varepsilon_h = \frac{1}{2}\varepsilon_h, \quad \varepsilon_{R\varphi 1} = \frac{1}{2}\varepsilon_{R\varphi 1}$$

来选取. 为了使算法 3.2 不过于繁复，我们省掉了这个过程.

下面的定理给出算法 3.2 的整体收敛性.

定理 3.3 设假定 3.1 成立，则算法 3.2 所构造的序列 $\{z_i\}$ 是无限的，其任何聚点 z^* 为 f 在曲面 S 上的稳定点，即有 $\|P_\varphi(z^*)\nabla f(z^*)\|=0$；或者 $\{z_i\}$ 是有限的，若 z_i 是其最后元，则算法 3.2 在步 2—步 3 之间无限循环，由步 2 得到的无限序列 $\{z^j_{i+1}\}_{j=0}^{\infty}$ 的极限为 f 在曲面 S 上的稳定点.

推论 3.4 设假定 3.1 成立. 若算法 3.2 构造的无限序列 $\{z_i\}_{i=0}^{\infty}$ 至少有一个聚点，则有

$$\lim_{i\to\infty}\|P_\varphi(z_i)\nabla f(z_i)\|=\lim_{i\to\infty}\|h_{z_i}\|=0.$$

下面来讨论算法 3.2 的终端收敛速度.

设 $\{z_i\}_{i=0}^{\infty}$ 是由算法 3.2 构造的无限序列，z^* 为它的一个聚点，作下面的假定.

假定 3.5 1° 假定 2.1 成立

2° 函数 f 在 D_c 上二次 Fréchet 可微，$\dfrac{\partial^2 f(z)}{\partial z\partial z}$ 在 z^* 的一个邻域中满足常数为 L_f 的 Lipschitz 条件.

3° 存在 r 维列向量 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_r^*)^T$，使有

$$3.1°\quad \nabla f(z^*)+\sum_{e=1}^{r}\lambda_e^*\nabla\varphi^e(z^*)=\theta. \tag{3.5}$$

3.2° 存在正数 a 和 $A(a\leqslant A)$，使矩阵

$$L(z^*,\lambda^*)=\frac{\partial^2 f(z)}{\partial z\partial z}\Big|_{z=z^*}+\sum_{e=1}^{r}\lambda_e^*\frac{\partial^2\varphi^e(z)}{\partial z\partial z}\Big|_{z=z^*} \tag{3.6}$$

对任意的 $h\in T_0(z^*)$ 有

$$a\|h\|^2\leqslant\langle h, L(z^*,\lambda^*)h\rangle\leqslant A\|h\|^2. \tag{3.7}$$

作 r 维列向量 $\lambda(z)=(\lambda^1(z),\cdots\lambda^r(z))^T$ 为

$$\lambda(z)=-[G_{\varphi1}(z)G_{\varphi1}^T(z)]^{-1}G_{\varphi1}(z)\nabla f(z), \tag{3.8}$$

则 $P_\varphi(z)\nabla f(z)$ 可表示成

$$P_\varphi(z)\nabla f(z)=\nabla f(z)+\sum_{l=1}^{r}\lambda^l(z)\nabla\varphi^l(z). \tag{3.9}$$

由假定 2.1，满足 (3.5) 的 λ^* 是唯一的. 由 (3.5) 解出 λ^*，可得 $\lambda^*=\lambda(z^*)$.

按 § 2 的方式在 z^* 的邻域中建立 S 的参数表示 $z(\mu,z^*)$，H_{z^*} 是所用的矩阵. 由命题 2.2 的 2°，得

$$\frac{\partial^2 F(\mu,z^*)}{\partial\mu\partial\mu}\Big|_{\mu=\theta}=H_{z^*}^T\Big[\frac{\partial^2 f(z)}{\partial z\partial z}\Big|_{z=z^*}+\sum_{l=1}^{r}\lambda^l(z^*)\frac{\partial^2\varphi^l(z)}{\partial z\partial z}\Big|_{z=z^*}\Big]H_{z^*}. \tag{3.10}$$

由假定 3.5 的 3.2°，对所有的 $h_\mu\in E^m$，有不等式

$$a\|h_\mu\|^2\leqslant\Big\langle h_\mu,\frac{\partial^2 F(\mu,z^*)}{\partial\mu\partial\mu}\Big|_{\mu=\theta}h_\mu\Big\rangle\leqslant A\|h_\mu\|^2. \tag{3.11}$$

于是，由命题 2.2，对任意小于 a 的正数 ε_a 和任意的正数 ε_A，可找到正数 ε_{μ^*} 和 ε_{z^*}，使当 $\mu\in E^m$，$\|\mu\|\leqslant\varepsilon_{\mu^*}$ 和 $\tilde z\in S$，$\|\tilde z-z^*\|\leqslant\varepsilon_{z^*}$ 时，对所有 $h_\mu\in E^m$ 有

$$(a-\varepsilon_a)\|h_\mu\|^2\leqslant\Big\langle h_\mu,\frac{\partial^2 F(\mu,\tilde z)}{\partial\mu\partial\mu}h_\mu\Big\rangle\leqslant(A+\varepsilon_A)\|h_\mu\|^2. \tag{3.12}$$

为了给出算法 3.2 的终端收敛速度，我们构造一个求解无约束优化问题的梯度法的摄动模型. 用 2^B 表示集合 B 的所有子集所成的集合. 记 $\delta_\mu=\{\mu\in E^m\mid\|\mu\|\leqslant l_\mu\}$ 为 $\mu=\theta$ 的一个邻域，l_μ 为给定的小正数. 对给定的正数 C_{FP}, r_μ, $C_{0\mu}$, $C_{1\mu}$, $C_{2\mu}$, α 构造下面的点集映象. $F_P(\cdot)$ 是由 δ_μ 到 2^{E^m} 中的点集映象，对任何 $\mu\in\delta_\mu$ 有

$$F_P(\mu)=\{\nabla F_P\in E^m\mid\|\nabla F_P-\nabla F(\mu,z^*)\|\leqslant C_{FP}\|\nabla F(\mu,z^*)\|^2\}.$$

对任意的 $\bar\mu\in\delta_\mu$, $h_\mu\in E^m$ 和正数组 $rC=\{r_\mu,C_{0\mu},C_{1\mu},C_{2\mu}\}$，用 $V(\bar\mu,h_\mu,rC)$ 表示 E^m 中一切可表成

$$\mu(\lambda)=\bar\mu+\lambda h_\mu+O_\mu(\|\lambda h_\mu\|^2)$$

的曲线的集合，其中 $O_\mu(\|\lambda h_\mu\|^2)$ 是 λ 的二次连续可微向量值函数. 当 $\|\lambda h_\mu\|\leqslant r_\mu$ 时，它和它的导数有估计式

$$\|O_\mu(\|\lambda h_\mu\|^2)\|\leqslant C_{0\mu}\|\lambda h_\mu\|^2,$$

$$\left\|\frac{dO_\mu(\|\lambda h_\mu\|^2)}{d\lambda}\right\|\leqslant C_{1\mu}\|\lambda h_\mu\|\|h_\mu\|,$$

$$\left\|\frac{d^2O_\mu(\|\lambda h_\mu\|^2)}{d\lambda^2}\right\|\leqslant C_{2\mu}\|h_\mu\|^2.$$

$e(\cdot,\cdot)$ 是由 $\delta_\mu\times E^m$ 到 2^{E^m} 中的映象，对 $\bar\mu\in\delta_\mu$, $h_\mu\in E^m$，有

$$e(\bar\mu,h_\mu)=\{\mu\in E^m\mid\mu=\mu(\lambda^*),\ \mu(\lambda)\in V(\bar\mu,h_\mu,rC),$$

$\lambda^*\in[0,\lambda_\mu]$ 且有 $F(\mu(\lambda^*),z^*)-F(\bar\mu,z^*)\leqslant-\alpha\|\nabla F(\bar\mu,z^*)\|^2\}$，其中

$$\lambda_\mu=\max\{\bar\lambda\geqslant0\mid$$

当 $\lambda\in[0,\bar\lambda]$ 时，$F(\mu(\lambda),z^*)\leqslant F(\bar\mu,z^*)\}$.

利用点集映象 $F_P(\cdot)$, $e(\cdot,\cdot)$，由 $\mu_0\in\delta_\mu$ 计算 $\nabla F_P\in F_P(\mu_0)$，令 $h_\mu=-\nabla F_P$，再计算 $\mu_1\in e(\mu_0,h_\mu)$. 这种由 μ_0 计算 μ_1 的过程是梯度法的一个摄动模型. 在 [6] 中，我们建立了比这更一般的模型.

当 z_i 充分接近 z^* 时，用恢复算子 R 得到的曲面 S 上的点 z_i^R 可表示成 $z(\mu_i^R,z^*)$，其中 $\mu_i^R\in E^m$. 通过繁复的估计，可以找到适当的自然数 i_0 和正数 l_μ, C_{FP}, r_μ, $C_{0\mu}$, $C_{1\mu}$, $C_{2\mu}$ 和 α，使得当 $i\geqslant i_0$ 时，μ_{i+1}^R 可以看成由 μ_i^R 通过上述梯度法的摄动模型得到的. 利用[6]的结果，有如下定理：

定理 3.6 设 $\{z_i\}_{i=0}^\infty$ 是由算法 3.2 构造的无限序列，z^* 是它的一个聚点，在 z^* 处假定 3.5 成立，则 $\{z_i\}_{i=0}^\infty$ 收敛到 z^*，并且终端收敛速度至少是线性的.

§ 4. 梯度投影牛顿算法

由(3.12)式，若在 z^* 处假定 3.5 成立，则当 $\tilde z\in S$ 离 z^* 充分近时，函数

$$F(\mu,\tilde z)=f(z(\mu,\tilde z))$$

对 μ 的二阶导数矩阵在 $\mu=\theta$ 处是正定的，并且这个矩阵为

$$\left.\frac{\partial^2F(\mu,\tilde z)}{\partial\mu\partial\mu}\right|_{\mu=\theta}=H_{\tilde z}^T\left[\left.\frac{\partial^2f(z)}{\partial z\partial z}\right|_{z=\tilde z}+\sum_{l=1}^r\lambda^l(\tilde z)\left.\frac{\partial^2\varphi^l(z)}{\partial z\partial z}\right|_{z=\tilde z}\right]H_{\tilde z}. \tag{4.1}$$

所以，若 f, φ^l, $l=1,\cdots,r$ 的二阶导数均能计算时，在 $\mu=\theta$ 处可施行一步牛顿法. 再

用得到的新的 z 点来代替 $\tilde{z}$，可逐次计算下去．将上述的计算过程转化到 E^n 空间进行，可得到下面的算法．

算法 4.1 （梯度投影牛顿算法）．选取正数 $\varepsilon_{\rho1}$，$\varepsilon_{\rho2}$，ε_h，α．其中 $\varepsilon_{\rho1}$，$\varepsilon_{\rho2}$ 适当小，$0<\varepsilon_{\rho1}<1$，$0<\alpha<\frac{1}{2}$．选取 ε_h 使得当 $\lambda\in[0,\varepsilon_h/\|h_{z_i}\|]$ 时，步 6 构造的序列收敛．

步 0. 选取 $z_0\in S$，令 $i=0$．

步 1. 计算 $P_\varphi(z_i)\nabla f(z_i)$．若有 $\|P_\varphi(z_i)\nabla f(z_i)\|=0$，则停止计算；否则转步 2．

步 2. 在切子空间 $T_0(z_i)$ 中选取 m 个相互正交的单位向量 h_{1i}，h_{2i}，$\cdots$，h_{mi} 组成矩阵 $H_{z_i}=[h_{1i},\cdots,h_{mi}]$，按(3.8)式计算 $\lambda(z_i)$．计算矩阵

$$L_H(z_i)=H_{z_i}^T\left[\frac{\partial^2 f(z)}{\partial z\partial z}\bigg|_{z=z_i}+\sum_{l=1}^{r}\lambda^l(z_i)\frac{\partial^2\varphi^l(z)}{\partial z\partial z}\bigg|_{z=z_i}\right]H_{z_i}.$$

步 3. 求矩阵 $L_H(z_i)$ 的逆 $L_H(z_i)^{-1}$，并计算

$$h_{z_i}=-H_{z_i}L_H(z_i)^{-1}H_{z_i}^TP_\varphi(z_i)\nabla f(z_i).$$

若逆矩阵 $L_H(z_i)^{-1}$ 不存在或者有

$$\langle P_\varphi(z_i)\nabla f(z_i),\ h_{z_i}\rangle>-\varepsilon_\rho\max\{\|P_\varphi(z_i)\nabla f(z_i)\|^2,\|P_\varphi(z_i)\nabla f(z_i)\|\|h_{z_i}\|\}.$$

时，令 $h_{z_i}=-P_\varphi(z_i)\nabla f(z_i)$，转步 6；否则转步 4．式中

$$\varepsilon_\rho=\min\{\varepsilon_{\rho1},\ \varepsilon_{\rho2}\|P_\varphi(z_i)\nabla f(z_i)\|\}.$$

步 4. 若有 $\|h_{z_i}\|\leqslant\varepsilon_h$ 转步 5；否则转步 6．

步 5. 从点 $z_i^0(1)=z_i+h_{z_i}$ 出发，由程序

$$z_i^{k+1}(1)=R(z_i^k(1))\qquad k=0,1,\cdots$$

构造 E^n 中的点列 $\{z_i^k(1)\}_{k=0}^\infty$，其极限为曲面 S 上的点 $z_i(1)$．再计算函数值 $f(z_i(1))$．若成立不等式

$$f(z_i(1))-f(z_i)\leqslant\alpha\langle P_\varphi(z_i)\nabla f(z_i),\ h_{z_i}\rangle,$$

则令 $\lambda_i=1$，转步 8；否则转步 6．

步 6. 构造直线段 $z_i^0(\lambda)=z_i+\lambda h_{z_i}$，$\lambda\in[0,\varepsilon_h/\|h_{z_i}\|]$，由

$$z_i^{k+1}(\lambda)=R(z_i^k(\lambda))\qquad k=0,1,\cdots,\lambda\in[0,\varepsilon_h/\|h_{z_i}\|]$$

构造曲面 S 上的曲线段 $z_i(\lambda)$．

步 7. 在区间 $[0,\varepsilon_h/\|h_{z_i}\|]$ 中寻找 λ_i，有

$$f(z_i(\lambda_i))=\min\{f(z_i(\lambda))\,|\,\lambda\in[0,\varepsilon_h/\|h_{z_i}\|]\}.$$

步 8. 令 $z_{i+1}=z_i(\lambda_i)$，$i=i+1$，转步 1．

这个算法实际上是梯度投影法与牛顿法的结合．当牛顿法不能施行或施行效果不好时，就应用梯度投影法．算法中 $L_H(z_i)$ 是 $(n-r)\times(n-r)$ 阶矩阵，而文献中求解(1.1)的牛顿法通常要求 $(n+r)\times(n+r)$ 阶矩阵的逆．算法 4.1 的步 5，步 6 和步 7 中均包含无限运算过程，具体计算时要进行截断．

下面二个定理分别给出算法 4.1 的整体收敛性和终端收敛速度．

定理 4.2 设假定 3.1 成立，则算法 4.1 构造的序列 $\{z_i\}$ 是有限的，其最后元 z_i 有 $\|P_\varphi(z_i)\nabla f(z_i)\|=0$；或者 $\{z_i\}$ 是无限的，则其任何聚点 z^* 是 f 在曲面 S 中的稳定点，即有 $\|P_\varphi(z^*)\nabla f(z^*)\|=0$．

定理 4.3　设 $\{z_i\}_{i=0}^{\infty}$ 是由算法 4.1 构造的无限序列，z^* 是它的一个聚点，在 z^* 处假定 3.5 成立，则序列 $\{z_i\}_{i=0}^{\infty}$ 收敛到 z^*，并且存在正数 i_0 和 q，使得当 $i \geqslant i_0$ 时，有

$$\|z_{i+1} - z^*\| \leqslant q\|z_i - z^*\|^2.$$

§5. 共轭梯度投影恢复算法

应用 Поляк-Polak-Ribiere 共轭梯度法求解问题(2.9)，并将计算过程通过 $z = z(\mu, \tilde{z})$ 转化成 E^n 空间的过程。与梯度投影算法一样，用 $P_\varphi(z(\mu,\tilde{z}))\nabla f(z(\mu,\tilde{z}))$ 代替过程中的量 $\dfrac{\partial z(\mu,\tilde{z})}{\partial\mu}\nabla F(\mu,\tilde{z})$，同时对 z 作摄动，允许它离开曲面，从而构造了下面的算法。

算法 5.1　(共轭梯度投影恢复算法)。选取自然数 $P(\geqslant m = n - r)$，选取小于 1 的正数 ε_{S1}，ε_{Rg1}，ε_{Rh1}，ε_{hg1}，A_{fd}，$\varepsilon_{\Delta f}$，要求有 $(1+\varepsilon_{Rh1})\varepsilon_{Rg1} + \varepsilon_{Rh1} < 1$。选取正数 ε_{S2}，ε_{Rg2}，ε_{Rh2}，ε_{hg2}，$A_{R\varphi}$，A_{Rfh}。选取充分大的正数 α_{rh}，选取正数 $\varepsilon_{R\varphi}$ 和 ε_h，保证由步 6 构造的曲线段 $z_i^{R0}(\lambda)$ 用恢复算子得到的序列 $\{z_i^{Rk}(\lambda)\}_{k=0}^{\infty}$，当 $\|\lambda h_{z_i}\| \leqslant \varepsilon_h$ 时，一致收敛。选取正数 $d_1 \in (0, \frac{1}{2})$，$d_2 \in (\frac{1}{2}, 1)$。

步 0　选取 $z_0^0 \in D$，令 $j = 0$，$i = 0$

步 1．计算量，

$$z_i^{j+1} = R(z_i^j),\ \varphi(z_i^{j+1}),\ P_\varphi(z_i^{j+1}),\ P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1}).$$

当 $i = 0, P, 2P, \cdots$ 时，对 $S \geqslant j+1$，令

$$\tilde{h}_i^S = -P_\varphi(z_i^S)\nabla f(z_i^S);$$

当 $i \neq 0, P, 2P, \cdots$ 时，对 $S \geqslant j+1$，令

$$\gamma_{i-1}^S = \frac{\langle P_\varphi(z_i^S)\nabla f(z_i^S) - P_\varphi(z_{i-1})\nabla f(z_{i-1}),\ P_\varphi(z_i^S)\nabla f(z_i^S)\rangle}{\langle P_\varphi(z_{i-1})\nabla f(z_{i-1}), P_\varphi(z_{i-1})\nabla f(z_{i-1})\rangle},$$

$$\tilde{h}_{i-1}^S = -P_\varphi(z_i^S)\nabla f(z_i^S) + \gamma_{i-1}^S P_\varphi(z_i^S) h_{z_{i-1}}.$$

步 2．若存在正数 ε_g，对所有 $S \geqslant j+1$ 有

$$\|P_\varphi(z_i^S)\nabla f(z_i^S)\| \geqslant \varepsilon_g,$$
$$\|P_\varphi(z_i^S)\nabla f(z_i^S) - P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\| \leqslant \varepsilon_{Rg}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|,$$
$$\|\varphi(z_i^S)\| \leqslant \varepsilon_{R\varphi},$$

则转步 3；否则令 $j = j+1$，转步 1．式中

$$\varepsilon_{Rg} = \min\{\varepsilon_{Rg1},\ \varepsilon_{Rg2}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\}.$$

步 3．若对所有的 $S \geqslant j+1$ 有

$$\|\tilde{h}_i^S\| \leqslant \min\{\varepsilon_{hg1},\ \varepsilon_{hg2}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|,$$

令

$$h_i^S = -P_\varphi(z_i^S)\nabla f(z_i^S), \qquad S \geqslant j+1.$$

并转步 4；若不成立，而对所有 $S \geqslant j+1$ 成立

$$\|\tilde{h}_i^S\| \geqslant \frac{1}{2}\min\{\varepsilon_{hg1},\ \varepsilon_{hg2}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|,$$

则令

$$h_i^S = \tilde{h}_i^S, \qquad S \geqslant j+1,$$

并转步 4; 若不能判定,令 $j=j+1$, 转步 1.

步 4. 若对 $S \geqslant j+1$, 下列估计式

$$\|h_i^{j+1} - h_i^S\| \leqslant \min\{\varepsilon_{Rh1},\ \varepsilon_{Rh2}\|h_i^{j+1}\|,\ A_{Rfh}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\}\|h_i^{j+1}\|,$$
$$\|\varphi(z_i^S)\| \leqslant A_{R\varphi}\|h_i^{j+1}\|\min\{\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|, \|h_i^{j+1}\|\},$$

均成立,转步 5;否则令 $j=j+1$, 转步 1.

步 5. 取 $\varepsilon_{Rh} = \min\{\varepsilon_{Rh1},\ \varepsilon_{Rh2}\|h_i^{j+1}\|\}$, $\varepsilon_\rho = (1+\varepsilon_{Rh})\varepsilon_{Rg} + \varepsilon_{Rh}$,
若

$$\langle P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1}), h_i^{j+1}\rangle < -\varepsilon_\rho\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|\|h_i^{j+1}\|,$$
$$\|\gamma_{i-1}^{j+1}P_\varphi(z_i^{j+1})h_{z_{i-1}}\| < \alpha_{\gamma h}\|P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1})\|,$$

均成立,转步 6; 若不全成立,令

$$h_i^{j+1} = -P_\varphi(z_i^{j+1})\nabla f(z_i^{j+1}),$$

转步 6.

步 6. 令 $z_i = z_i^{j+1}$, $h_{z_i} = h_i^{j+1}$, 并按算法 3.2 的步 5 构造曲线段 $z_i^{R0}(\lambda)$, $\lambda \in [0, \varepsilon_h/\|h_{z_i}\|]$. 令 $k=0$, 转步 7.

步7. 若有

$$\left.\frac{df(z_i^{Rk}(\lambda))}{d\lambda}\right|_{\lambda=0} \leqslant A_{fd}\langle P_\varphi(z_i)\nabla f(z_i), h_{z_i}\rangle,$$

转步 8;否则令 $j=j+k+2$, $z_i^j = z_i^{Rk}(0)$, 转步 1.

步 8. 取 $\bar{\lambda}_i = \max\{\bar{\lambda} \geqslant 0 \mid \lambda \in [0, \bar{\lambda}]$ 时 $f(z_i^{Rk}(\lambda)) \leqslant f(z_i^{Rk}(0))\}$, 在集合

$$[0, \bar{\lambda}_i] \cap [0, \varepsilon_h/\|h_{z_i}\|]$$

中寻找满足

$$1^\circ \qquad d_1 \leqslant \frac{f(z_i^{Rk}(\lambda_i)) - f(z_i^{Rk}(0)) - \lambda_i \left.\dfrac{df(z_i^{Rk}(\lambda))}{d\lambda}\right|_{\lambda=0}}{-\lambda_i \left.\dfrac{df(z_i^{Rk}(\lambda))}{d\lambda}\right|_{\lambda=0}} \leqslant d_2,$$

$$2^\circ \left|\left.\frac{df(z_i^{Rk}(\lambda))}{d\lambda}\right|_{\lambda=\lambda_i}\right| \leqslant \min\{\varepsilon_{S1}, \varepsilon_{S2}\|P_\varphi(z_i)\nabla f(z_i)\|\}\|P_\varphi(z_i^{Rk}(\lambda_i))\nabla f(z_i^{Rk}(\lambda_i))\|\|h_{z_i}\|$$

的 λ_i, 转步 9; 若找不到这样的 λ_i, 寻找只满足 1° 的 λ_i, 转步 9; 若满足 1° 的 λ_i 也找不到,令 $\lambda_i = \varepsilon_h/\|h_{z_i}\|$, 转步 9.

步 9. 令

$$\Delta f_i^{Rk} = f(z_i^{Rk}(0)) - f(z_i^{Rk}(\lambda_i)),$$
$$z_i^{Rk+0}(\lambda_i) = z_i^{Rk}(\lambda_i),$$
$$z_i^{Rk+0}(0) = z_i^{Rk}(0),$$
$$l = 0.$$

步 10. 计算

$$z_i^{Rk+l+1}(0) = R(z_i^{Rk+l}(0)),$$
$$z_i^{Rk+l+1}(\lambda_i) = R(z_i^{Rk+l}(\lambda_i)),$$
$$\Delta f_i^{Rk+l+1} = f(z_i^{Rk+l+1}(0)) - f(z_i^{Rk+l+1}(\lambda_i)).$$

步 11. 若对所有 $S \geqslant 0$ 成立 $\triangle f_i^{Rk+S} \geqslant \varepsilon_{\Delta f} \triangle f_i^{Rk}$,

$$\left|\frac{df(z_i^{Rk+S}(\lambda))}{d\lambda}\Big|_{\lambda=\lambda_l}\right|$$

$$\leqslant 2\min\{\varepsilon_{S1},\ \varepsilon_{S2}\|P_\varphi(z_i)\nabla f(z_i)\|\}\|P_\varphi(z_i^{Rk+S}(\lambda_l))\nabla f(z_i^{Rk+S}(\lambda_l))\|\|h_{z_l}\| \tag{5.1}$$

(当步 8 的 2° 不满足时，(5.1) 不需成立)转步 12；若不成立，令 $k = k + 1$ 转步 7；若尚不能判定,令 $l = l + 1$，转步 10.

步 12. 令 $z_{i+1}^0 = z_i^{Rk+l+1}(\lambda_l)$, $j = 0$, $i = i + 1$, 转步 1.

若假定 3.1 成立，则算法 5.1 不会在步 1—步 3，步 1—步 4，步 1—步 7 和步 10—步 11 之间无限循环.

对算法 5.1，有下面的定理,它给出了算法 5.1 的整体收敛性.

定理 5.2 设假定 3.1 成立. 则算法 5.1 构造的序列 $\{z_i\}$ 是无限的，这时它的每个聚点均是函数 f 在曲面 S 上的稳定点；或者 $\{z_i\}$ 是有限的,这时有二种情形,一种情形是计算在步 1—步 2 之间无限循环，由步 1 得到的无限序列 $\{z_i^j\}_{j=0}^{\infty}$ 的极限点是 f 在曲面 S 上的稳定点；另一种情形是算法在步 7—步 11 之间无限循环，由步 8 求得的点所组成序列 $\{z_i^{Rk}(\lambda_i^k)\}$ 的任何聚点均是 f 在 S 上的稳定点,其中 λ_i^k 指步 8 求得的 λ_l.

推论 3.4 对算法 5.1 也是成立的.

设 $\{z_i\}_{i=0}^{\infty}$ 是由算法 5.1 构造的无限序列，z^* 是它的一个聚点. 在 z^* 处假定 3.5 成立. 记号 $z(\mu, z^*)$, $F(\mu, z^*)$, δ_μ, $F_P(\cdot)$, $V(\bar{\mu}, h_\mu, rc)$ 均按 § 3 定义. 为了利用文[7]建立的摄动理论来给出算法 5.1 的终端收敛速度,还需构造下面的点集映象.

$\gamma(\cdot,\cdot)$ 是由 $\delta_\mu \times \delta_\mu$ 到 2^{E^1} 中的映象,对 $\mu_1 \in \delta_\mu$, $\mu_2 \in \delta_\mu$, $\|\mu_1\| \neq 0$, 有

$$\gamma(\mu_1, \mu_2) = \Big\{\gamma \in E^1 \Big| \Big|\gamma - \frac{\langle \nabla F(\mu_2, z^*) - \nabla F(\mu_1, z^*), \nabla F(\mu_2, z^*)\rangle}{\|\nabla f(\mu_1, z^*)\|^2}\Big|$$

$$\leqslant \frac{\|\nabla F(\mu_2, z^*)\|}{\|\nabla F(\mu_1, z^*)\|}[C_{\gamma_1}\|\nabla F(\mu_1, z^*)\| + C_{\gamma_2}\|\nabla F(\mu_2, z^*)\|$$

$$+ C_{\gamma_3}\|\mu_1 - \mu_2\|]\Big\}.$$

$\hat{h}(\cdot)$ 是由 E^m 到 2^{E^m} 中的映象,对 $h \in E^m$ 有

$$\hat{h}(h) = \{\hat{h} \in E^m \mid \|\hat{h} - h\| \leqslant C_{\hat{h}}\|h\|^2\}.$$

$\bar{h}(\cdot,\cdot,\cdot)$ 是由 $\delta_\mu \times \delta_\mu \times E^m$ 到 2^{E^m} 中的映象,对于 $\mu_1 \in \delta_\mu$, $\mu_2 \in \delta_\mu$, $h \in E^m$有

$$\bar{h}(\mu_1, \mu_2, h) = \{\bar{h} \in E^m \mid \|\bar{h} - h\| \leqslant C_{\bar{h}1}\|h\|\|\mu_1 - \mu_2\| + C_{\bar{h}2}\|h\|^2\}.$$

$h(\cdot,\cdot,\cdot)$ 是由 $\delta_\mu \times \delta_\mu \times E^m$ 到 2^{E^m} 中的映象,对于 $\mu_1 \in \delta_\mu$, $\mu_2 \in \delta_\mu$, $h_1 \in E^m$, 有

$$h(\mu_1, \mu_2, h_1) = \{h \in E^m \mid h = -\nabla F_P + \gamma\bar{h},\ \nabla F_P \in F_P(\mu_2),$$

$$\gamma \in \gamma(\mu_1, \mu_2),\ \bar{h} \in \bar{h}(\mu_1, \mu_2, h_1)\}.$$

$e(\cdot,\cdot)$ 是由 $\delta_\mu \times E^m$ 到 2^{E^m} 中的映象,对于 $\bar{\mu} \in \delta_\mu$, $\hat{h} \in E^m$, 有

$$e(\bar{\mu}, \hat{h}) = \{\mu \in E^m \mid \mu = \mu(\lambda^*),\ \mu(\lambda) \in V(\bar{\mu}, \hat{h}, rc),\ \lambda^* \in [0, \lambda_\mu]$$

且有

$$\left|\frac{dF(\mu(\lambda), z^*)}{d\lambda}\Big|_{\lambda=\lambda^*}\right| \leqslant \varepsilon_{S_\mu}\|\nabla F(\bar{\mu}, z^*)\|\|\nabla F(\mu(\lambda^*), z^*)\|\|\hat{h}\|\}.$$

式中 $\lambda_\mu = \max\{\bar{\lambda} \geqslant 0 \mid$ 当 $\lambda \in [0, \bar{\lambda}]$ 时, $F(\mu(\lambda), z^*) \leqslant F(\bar{\mu}, z^*)\}$. 上面的 $C_{\gamma1}$, $C_{\gamma2}$, $C_{\gamma3}$, $C_{\hat{h}}$, $C_{\bar{h}1}$, $C_{\bar{h}2}$, ε_{S_μ} 均是给定的正数.

利用上述的点集映象,可以建立 Поляк-Polak-Rihiere 算法的摄动模型.

算法模型 选取自然数 $P \geqslant m$, 正数 $l_{\mu0} \leqslant l_\mu$.

步0. 选取 $\mu_0 \in E^m$, $\|\mu_0\| \leqslant l_{\mu0}$, 令 $i = 0$.

步 1. 计算 $\nabla F_P \in F_P(\mu_i)$, 若 $\|\nabla F_P\| = 0$, 则停止计算;否则转步 2.

步 2. 当 $i = 0, P, \cdots$, 令 $h_i = -\nabla F_P$; 否则令

$$h_i \in h(\mu_{i-1}, \mu_i, h_{i-1}).$$

步 3. 计算 $\hat{h}_i \in \hat{h}(h_i)$, $\mu_{i+1} \in e(\mu_i, \hat{h}_i)$.

步 4. 令 $i = i + 1$, 转步 1.

当 z_i 充分接近于 z^* 时, 通过恢复算子对应曲面上的点 z_i^R, z_i^R 可表示成 $z_i^R = z(\mu_i^R, z^*)$. 通过繁复的推导估计,可以证明,适当选取算法模型中的常数 l_μ, $l_{\mu0}$, rc, C_{FP} 等,这些 μ_i^R 可以看成由上述算法模型构造的点. 利用文[7]的结果,可得到下面的定理.

定理 5.3 设 $\{z_i\}_{i=0}^\infty$ 是由算法 5.1 构造的无限序列, z^* 是它的一个聚点, 在 z^* 处假定 3.5 成立,则序列 $\{z_i\}_{i=0}^\infty$ 收敛到 z^*, 并且存在自然数 i_0 和正数 E_z, ϕ_z, $0 < \phi_z < 1$ 和 q_z, 当 $i \geqslant i_0$ 和 $lP \geqslant i_0$ 时,有

$$\|z_i - z^*\| \leqslant E_z \varphi_z^{i-i_0},$$
$$\|z_{(l+1)P} - z^*\| \leqslant q_z \|z_{lP} - z^*\|^2.$$

本文在写作过程中,得到童承璞同志的帮助,特此致谢.

参 考 文 献

[1] J. B. Rosen, The gradient projection method for nonlinear programming, Part I. Linear constraints, *J. Soc. Indust, Appl. Math.*, 8(1960), 181—217.

[2] J. B. Rosen, The gradient projection method for nonlinear programming, Part II, Nonlinear constraints, *J. Soc. Indust. Appl. Math.*, 9(1961), 514—532.

[3] D. G. Luenberger, Introduction to linear and nonlinear programming, Addison-Wesley, Reading, MA, 1973.

[4] H. Mukai, E. Polak, On the use of approximations in algorthms for optimization problems w. equality and inequality comstraints, *SIAM J. Numer. Anal.*, 15 (1978) 674—693.

[5] P. E. Gill, W. Murray, Numerical methods for constrained optimization, Academic Press, London, 1974.

[6] 费景高,具有线性收敛速度的最优化方法的一个算法模型,待发表.

[7] 费景高,具有实现误差的共轭梯度算法,待发表.