Note of numerical optimaization

PB19000252 王梓睿

2022 年 11 月 12 日

1 introduction

1、连续优化问题通常比离散优化问题容易，因为连续性带来在特定点附近，限制条件不会变化太大。

2、有时候约束优化问题可以表示成无约束优化问题，方法是将约束添加到目标函数的惩罚项当中，从而阻止违反约束。

3、凸优化的局部最优解就是全局最优解。

证明. 设 x 为局部最优解（最小值），y≠x 为全局最优解，连接 y 与 x，路径与 x 的任一邻域有交，从而由凸函数性质，与 x 的局部最优性矛盾. □

本书大部分情况下我们只关心局部最优。

4、随机最优化解出的是模型的期望表现。

5、凸优化的定义：

⓪是约束优化问题的一个特例；

①目标函数为凸函数；

②约束等式中右侧函数为线性函数；

③约束不等式中右侧函数为凹函数；

2 无约束最优化

2.1 关于解

1、孤立局部最小值一定严格局部最小，反之则不一定。

2、Taylor 公式（利用两种余项），由此可以推导几个结论。

①f 在局部最小值点 x* 邻域内连续可微，则 $\nabla f(x*) = 0$；

②f 在局部最小值点 x* 邻域内，二阶导存在连续，则 $\nabla f(x*) = 0$ 且 $\nabla^2 f(x*)$ 半正定；

③（充分性）在某点 x* 邻域内二阶导存在连续，且 f(x*)=0,$\nabla^2 f(x*)$ 正定，那么 x* 为 f 的局部最小值点；

3、凸函数的局部最优值为全局最优值。若还可微，则任意稳定点都是全局最优值。

2.2 算法综述

1、线性探测法的最速下降方向是 $-\nabla f_k$，但收敛速度过慢。容易证明在与 $-\nabla f_k$ 夹角小于 $\pi/2$ 的方向上目标函数总会递减，称这些方向为下降方向。

2、牛顿方向为 $-(\nabla^2 f_k)^{-1}\nabla f_k$ 适用于 f_k 的 Hessian 矩阵正定，且函数越光滑误差越小。它收敛速度较快，但 Hessian 矩阵计算量较大。

3、准牛顿方法用近似的 B_k 矩阵代替 Hessian，给出 $-(B_k)^{-1}f_k$ 并在迭代中更新 B_k，从而减小计算量。

4、非线性共轭方法，采用公式 $p_k = -\nabla f_k + \beta_k p_{k-1}$，保留了最速下降法的计算简便性的同时提高了其收敛速度。

5、缩放，可以通过对角缩放等方式将变量大小拉到同一数量级。牛顿法能够比最速下降法更好地处理 poorly scaled 问题。

3 线性探测法

线性探测法先计算行进方向 p_k，再计算步长 α_k，迭代公式为 $x_{k+1} = x_k + \alpha_k p_k$。行进方向的选取通常要求 f_k 递降，一般令 $p_k = -(B_k)^{-1}\nabla f_k$，如果 B_k 正定，由 Taylor 公式，一次项可化为 $-\nabla f_k^T B_k^{-1}\nabla f_k < 0$，这样 f_k 能够递降。

3.1 步长

给定初值 x_k 和方向 p_k，我们记 $f_k(x_k + \alpha_k p_k)$ 为单变量函数 $\phi(\alpha_k)$，想让其尽可能的小。我们引入概念“充分递降”。

3.1.1 Wolfe 条件

1、Armijo 条件：$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1\alpha_k\nabla f_k^T p_k$。它保证了迭代过程至少线性收敛。

2、曲率条件：$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2\nabla f_k^T p_k$。事实上就是 $\phi'(\alpha_k) \geq c_2\phi'(0)$，来保证下降的梯度在某一区间内。

3、$0 < c_1 < c_2 < 1$ 时，以上两式合称 Wolfe 条件。若 2 式改写为 $|LHS| \leq |RHS|$，则称为强 Wolfe 条件（进一步缩小了梯度的可能区间）。
引理：任一连续可微有下界的函数，一定有满足（强）Wolfe 条件的区间。

3.1.2 其它

1、Goldenstein 条件：$f(x_k)+(1-c_1)\alpha_k\nabla f_k^T p_k \leq f(x_k+\alpha_k p_k) \leq f(x_k)+c_1\alpha_k\nabla f_k^T p_k$。

2、回溯线性搜索：仅用 Armijo 条件进行判断，但每次不满足都执行 $\alpha = \rho\alpha$，ρ 为收缩因子。ρ 可以在迭代过程中发生变化。

3.2 收敛性

考虑 θ_k 为方向 p_k 与最速下降方向 $-\nabla f_k$ 的夹角。

1、Zoutendijk 条件：对满足一系列条件的 f 成立

$$\sum_{0}^{\infty} cos^2\theta_k||\nabla f_k||^2 < \infty$$

Theorem 3.2.

Consider any iteration of the form (3.1), where p_k is a descent direction and α_k satisfies the Wolfe conditions (3.6). Suppose that f is bounded below in $\mathbb{R}^n$ and that f is continuously differentiable in an open set $\mathcal{N}$ containing the level set $\mathcal{L} \stackrel{\text{def}}{=} \{x : f(x) \leq f(x_0)\}$, where x_0 is the starting point of the iteration. Assume also that the gradient ∇f is Lipschitz continuous on $\mathcal{N}$, that is, there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \quad \text{for all } x, \tilde{x} \in \mathcal{N}. \tag{3.13}$$

Then

$$\sum_{k\geq 0} \cos^2\theta_k \|\nabla f_k\|^2 < \infty. \tag{3.14}$$

2、如果上述级数收敛，说明 $\cos^2\theta_k||\nabla f_k||^2 \to 0$。当 θ_k 有界地远离 90 度时，即存在 $\delta, cos\theta_k >= \delta > 0$，那么有 $||\nabla f_k \to 0$，我们称为全局收敛。

3、类牛顿方法中，如果矩阵 B_k 的条件数 $||B_k||||B_k^{-1}|| \leq M$，M 为一个常数，则可以证明全局收敛性。

3.3 收敛率

1、我们可以定义 Q 范数，并证明 $f(x) = \frac{1}{2}x^TQx - b^Tx$ 时，有 $\frac{1}{2}||x - x^*||_Q^2 = f(x) - f(x^*)$。

2、thm3.3，可以证明对严格凸的二次函数 f，有

$$||x_{k+1} - x^*||_Q^2 \leq (\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1})^2||x_k - x^*||_Q^2.$$

其中 $0 < \lambda_1 \leq \cdots \leq \lambda_n$ 为 Q 的特征值。这意味着 f_k 至少以线性收敛到 f^*。

3、thm3.5，如果 f 二次可微，Hessian 在解 x^* 附近李普希兹连续，那么 x

(i) 若起始点 x_0 离 x^* 足够近，那么迭代序列收敛到 x^*。

(ii)x_k 有二阶收敛速度，$||\nabla f_k||$ 二阶收敛到 0.

4、thm3.6，此定理说明了在 p_k 满足一个极限式的时候，k 较大时可直接将 α_k 取 1，且此时 x_k 超线性收敛到 x^*。thm3.7 说明了 α_k 恒取为 1 的情况下，3.6 的条件是充要的。而准牛顿方法满足这一条件，因此能够超线性收敛。

3.4 Hessian 修正

我们希望能够找到方阵 E_k，使得 $B_k = \nabla^2 f(x_k) + E_k$ 正定。要求这一修改尽量小，以保留 Hessian 矩阵的更多信息。1、可以考虑 $\nabla^2 f_k + \tau I$，并在迭代中不断增加 τ_k 的值。但是这样每次更新 τ 后都要重新对矩阵进行分解，计算量太大。

2、Cholesky 分解，可以在算法中引入两个新的变量列，从而边分解边进行矩阵修改。（注意并不是先分解为 $L^T DL$ 再修改 D，因为首先我们无法保证分解的存在性，其次修改后的结果不稳定）

3、更稳定的方法是进行对称不定分解 $A = L^T BL$，其中 A 须为对称矩阵，B 为准对角阵，对角块大小为 1 或 2. 这一分解必定存在。

3.5 步长选择算法

3.5.1 插值算法

1、从 $\phi(0), \phi'(0)$, 及初始步长 α_0 构造出 α_1，随后构造出三次插值函数，利用 $\phi(0), \phi'(0), \alpha_{k-1}, \alpha_k$ 来计算 α_{k+1}。这一方法并没有计算在诸 α_i 处的导数值。

2、如果导数值比较好算，我们可以用 $\phi(\alpha_{i-1}), \phi(\alpha_i), \phi'(\alpha_{i-1}), \phi'(\alpha_i)$ 来对 α_{i+1} 作三次插值。

3.5.2 Wolfe 条件下的线性搜索算法

先利用两个不等式确定一个目标值存在的区间，再调用 Zoom 函数不断缩小区间的范围，以得到满足 Wolfe 条件的值。

4 置信区域法

基本思想是先在迭代中更新置信区域及其半径，再在该区域内寻找步长和方向。

我们定义 $m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p, B_k$ 只限定对称和一致有界，来考虑 $||p|| \leq \triangle_k$ 时的 m_k 最小值。

4.1 Overview

1、Algo4.1 描述了置信区域法的一般算法。记 $\rho_k = \frac{f(x_k)-f(x_k+p_k)}{m_k(0)-m_k(p_k)}$，如果 ρ_k 为负，说明 f 增加，必须拒绝。若 ρ_k 接近 1，说明预测模型很好，可以扩大置信半径；反之若 ρ_k 接近 0 说明模型不好，需要缩小置信半径。

2、thm 4.1 刻画了求 $m_k(p)$ 取最小值的条件，即 $\exists \lambda \geq 0, s.t.$

$$(B + \lambda I)p^* = -g,$$

$$\lambda(\triangle - ||p^*||) = 0,$$

$$(B + \lambda I)\text{正定}.$$

4.2 柯西点

1、设 $p_k^s = argmin f_k + g_k^T p, ||p|| \leq \triangle_k$，再计算 τ_k 使得 $m_k(\tau p_k^s)$ 最小且 $||\tau p_k^s|| \leq \triangle_k$，我们把求得的 $p_k^c = \tau_k p_k^s$ 称为柯西点。柯西点计算比较容易，但由于不太依赖矩阵 B_k 的值，也注定了收敛速度不会太快。

（接下来考察单步迭代内部的求解过程，暂时忽略下标 k）

2、Dogleg Method（折线法），当仅 B 正定时可以使用。思想是用两条线段组成的折线代替解曲线，分别是起始点到 $p^U//g$，和 $p^B//-B^{-1}g$。thm4.2 证明了这一折线段 $\widetilde{p}(\tau)$ 上 $||\widetilde{p}(\tau)||$ 单增，且 $m(\widetilde{p}(\tau))$ 单减。

3、二维子空间法，是 dogleg 的推广，代替了 p^U, p^B 两个方向，而在 $span[p^U, p^B]$ 中找。这一方法可以进一步改为在 $span[g, (B+\alpha I)^{-1}g]$ 中找，从而不再需要 B 的正定性。

4.3 全局收敛

1、lemma4.3 给出了 m_k 下降的一个估计，式子形式类似 Wolfe 第一个条件。

2、thm4.4 说明了只要近似解 p_k 能够满足使 m_k 下降柯西点处减少值的一个固定倍数即可满足 lemma4.3。

3、thm4.5,thm4.6 给出了 Algo4.1 中 η 分别取 0 和大于 0 的某一正数时，梯度序列 g_k 的下极限和极限为 0.

4.4 子问题的迭代解

1、先介绍了如何处理 $||p(\lambda)|| = \triangle$ 问题的解。利用奇异值分解可以写出

$$||p(\lambda)||^2 = \sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_i + \lambda)^2}$$

。作出图像 (P_{104})，可以看出，当 $q_1^T g \neq 0$ 时，函数在 $-\lambda_1$ 处取 ∞, 在无穷远处取 0，从而必存在目标解。随后可以利用 Newton 迭代法进行求解（可以通过取倒数等定义新函数使迭代过程更加稳定）。而当 $q_1^T g = 0$ 时，定义新的函数 p。

2、Lemma4.7 给出了 $m(p) = g^T p + \frac{1}{2} p^T B p$ 能取到最小值的充要条件和解满足的性质。并以此证明了 thm4.1.

4.5 Newton 法的局部收敛

thm4.9 证明了，在一定的条件下，置信区域 $\triangle_k$ 会逐渐减小，序列 $\{x_k\}$ 会超线性收敛到一点。

4.6 其它改进方法

1、第二章提过，优化问题的 Scaling 可能会很糟糕，即在某些方向，目标函数变化得过于剧烈。因此我们考虑椭圆置信区间，可以描述为 $||Dp|| \leq \triangle$，其中 D 是一个对角矩阵，在对变化比较敏感的方向取较大的对角元。

然后提出了 Scaling 过的柯西点计算算法。

2、使用其它范数，比如 $||\cdot||_\infty$，这在一些约束问题和一些数据比较大的问题中可以应用。

5 共轭梯度法

共轭梯度法可以用于解决线性的和非线性的问题。

5.1 线性共轭梯度法

记残差 $r_k = Ax_k - b$。我们称一组向量 $\{p_0, p_1, \cdots, p_l\}$ 为关于 A 共轭的，当 $p_i^T A p_j = 0, \forall i \neq j$，其中 A 正定对称。

5.1.1 共轭方向法

1、thm5.1，可以证明迭代 $x_{k+1} = x_k + \alpha_k p_k$ 可以由任意起始点 $x_0 \in \mathbb{R}^n$ 迭代至多 n 步找到解，其中 $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$。

2、thm5.2，说明了 $r_k^T p_i = 0, \forall i < k$，即当前残差总正交于已搜索的全部方向。

3、Algo 5.2：共轭梯度算法.

5.2 收敛率

本节大多数定理都在数值代数中学过，略去这些定理。

5.2.1 预处理

因为收敛速度被矩阵 A 的条件数控制，而条件数取决于 A 的特征值分布，我们可以用某个非奇异矩阵 C 对向量 x 进行变换。Algo5.3 给出了增加预处理的共轭梯度算法。

随后介绍了一些选择预处理条件矩阵 M 的方法，如不完全 Cholesky 分解。

5.3 非线性共轭梯度法

5.3.1 Fletcher-Reeves 方法

1、Algo 5.4，对 Algo 5.2 的简单改造。接下来探讨了强 Wolfe 条件对 p_k 的限制，以及 β_k 的其它选取方法，如 Polak-Ribiere，FR-PR,Hestnens-Stiefel 等。

2、重启。我们可以每 n 步就重置一次迭代，并且可以证明这一重启的过程具有二次收敛性。然而由于非线性问题的规模 n 往往很大，我们实际上采用另一些策略，比如考察两个相邻的梯度是否足够偏离正交，这样重启的间隔更短。

3、Lemma 5.6，证明了满足强 Wolfe 条件（的第二式）的 α_k 能够保证下降方向满足一个不等式，该不等式蕴含着 $\nabla f_k^T p_K < 0$，即方向确为函数递降方向。同时，该不等式还能推断出 FR 方法的一个缺点，就是当 p_k 与 $-\nabla f_k$ 接近正交时（即 $\cos\theta_k$ 很小时），会出现长时间的 $p_{k+1} \approx p_k$ 的现象。因此 FR 方法通常必须和重启策略绑定使用。

5.3.2 全局收敛

1、对配备了重启的算法,可以由强 Wolfe 条件推出 $\sum_{k=0}^{\infty} \cos^2\theta_k ||\nabla f_k||^2 < \infty$（Zoutendijk 条件）。这可以推出存在 k 的子列使 $\sum_{k_i} ||\nabla f_k||^2 < \infty$，进而梯度的下确界是 0.

2、thm5.7 说明了 FR 方法中，梯度的下确界也是 0。

3、thm5.8 说明了对一般性非凸的函数,PR 方法可能永远达不到解（梯度的下确界不为 0）。

最后作者建议使用基于 PR，PR+ 或 FR-PR 的方法。

6 Quasi-Newton 方法

6.1 BFGS 方法

回顾模型 $m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$，设其梯度为 0 可以解出 $p_k = -B_k^{-1}\nabla f_k$。取该 p_k 为方向，迭代式为 $x_{k+1} = x_k + \alpha_k p_k$，其中 alpha 满足 Wolfe 条件。然而在此过程中，因为 B_k 为一近似矩阵，需要进行更新，通过计算知需要满足扇形方程 $B_{k+1}s_k = y_k, (s_k = x_{k+1} - x_k, y_k = \nabla f_{k+1} - \nabla f_k)$。该方程的解自由度很大，因此取 $\min_B ||B - B_k||, B = B^T$ 使变化最小。

如何选取范数，决定了不同的 Quasi-Newton 方法。取带权重的 Frobenius 范数，$||A||_W = ||W^{1/2}AW^{1/2}||_F$。若取 W 为平均 Hessain 矩阵之逆，则形成 DFP 方法。若在一开始就考虑 $H_k = B_k^{-1}$，并取 W 为平均 Hessain 矩阵，则得到更好的 BFGS 方法。初值 H_0 没有定论，需根据情况来设（也可以直接设为单位阵）。

BFGS 的优点是计算量小（每步 $O(n^2)$），无需计算二阶导，且超线性收敛。对 Hessian 矩阵的自纠正能力强（当线搜索方法满足 Wolfe 条件时）。

6.2 SR1(symmetric-rank-1) 方法

BFGS 和 DFP 方法每次用一个秩为 2 的矩阵更新 B_k/H_k，而 SR1 方法用秩为 1 的矩阵来更新。即 $B_{k+1} = B_k + \sigma vv^T$，则 B_{k+1} 可以由 B_k 唯一决定。然而，这一显式表示含有分母 $(y_k - B_k s_k)^T s_k$，当该式为 0 时此方法崩溃，需要引入额外判断（当上式较小时不更新 B_k）

thm6.1 说对凸二次型函数，只要保证上式不为 0，则对任意起始点和起始矩阵都能收敛。

thm6.2 说对性质较好的函数，若步进方向均匀线性独立，则 B_k 能收敛到 Hessian 矩阵。

6.3 Broyden 类

该类满足 $B_{k+1} = (1 - \Phi_k)B_{k+1}^{BFGS} + \Phi_k B_{k+1}^{DFP}$，因此也满足扇形方程和 Hessian 矩阵的正定性（后者需 $\Phi_k \in [0, 1]$）。

thm6.3 说，$A^{1/2}B^{-1}A^{1/2}$ 的特征值随着迭代不会离 1 更远。

thm6.4 说，如果使用线搜索给出的完全正确的步长，并且系数 Φ_k 始终大于一计算值 Φ_k^c，那么收敛性等将很好。

6.4 收敛性分析

thm6.5 说，若 B_0 对称正定，起始点 x_0 满足假设 6.1，那么对 Broyden 类中 $\Phi_k \in [0, 1)$，即除了 DFP 方法外，x_k 将收敛到 f 的全局最小点。

进一步的分析知 BFGS 方法有超线性收敛性。

对 SR1 方法，需要额外的有界假设 thm6.7(c3)。若满足，则也有超线性收敛性，且 B_k 在大多数时候都为半正定的。