数值优化（Numerical Optimization）学习系列-信赖域方法

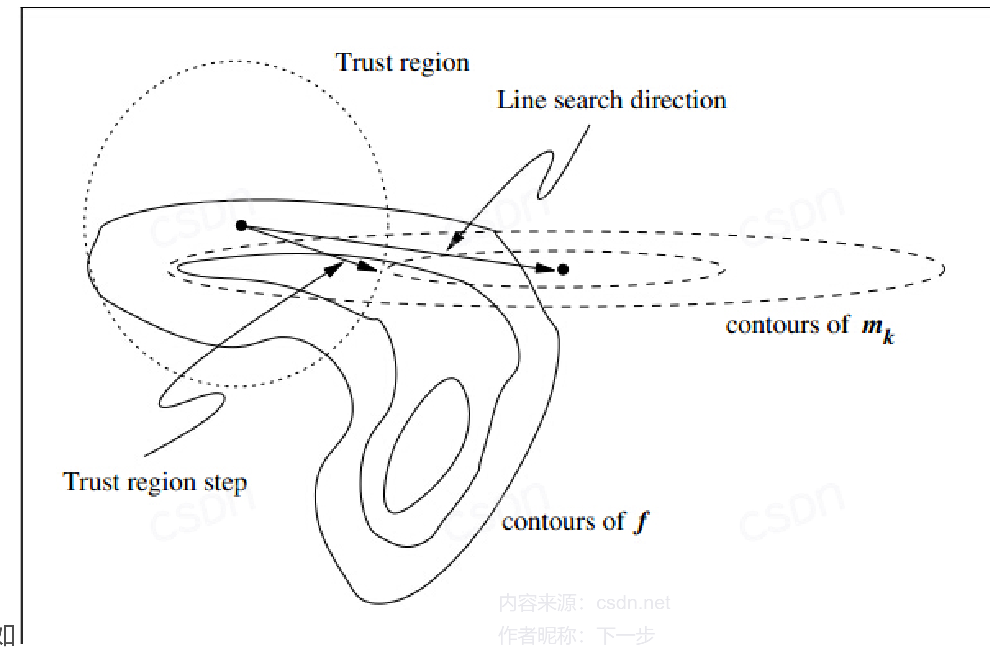
下一步 于 2015-12-27 18:45:43 发布

信赖域方法和线搜索类似都是迭代方法，与其不同的是，每次迭代时，在一个选定的**可信赖区域**内，选择当前迭代点的近似模型 m_k ，然后计算最优步长；如果步长不合适，可以对区域进行缩放。该小结主要介绍：

1. 信赖域方法的基本形式
2. 求解信赖域的基础方法
3. 信赖域方法的收敛性和收敛速度
4. 信赖域方法的扩展

信赖域方法的基本形式

在信赖域方法中，可信赖的区域（Region）的选择很重要，一般都会根据上一步结果进行动态变化；如果上一步太大则缩小，否则进行扩大。



在模型最优化问题中，选择TR方法比LS方法能够较快的收敛，例如
在该例子中，在非凸函数F中，当前步骤TR方法要优于LS。
信赖域方法有几个参数需要选择：

1. 近似模型 m_k
2. 可信赖区域 Δ_k
3. 求解参数 p_k

基本形式

在本节中模型选择为二次近似模型，采用函数二阶泰勒展开，即

$$f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f_k p$$

一般情况下会用 B_k 去近似Hessian矩阵，即

$$m_k = f_k + g_k^T + \frac{1}{2} p^T B_k p$$

其中 B_k为对称矩阵。

信赖域的基本形式为：

$$\min m_k(p) = f_k + g_k^T + \frac{1}{2} p^T B_k p \quad s.t\ ||p|| \le \Delta_k$$

该问题为关于p的带约束的最优化问题，参数p被限制在一个球形区域内。如果B_k选择为Hessian，则为TR的牛顿方法。
如果$||B_k^{-1} g_k|| \le \Delta_k$则 $p_k = -B_k^{-1} g_k$为**完全步(Full Step)**，即球形约束没有作用。

Δ_k的选择

参数Δ_k的选择一般会根据上一步的结果进行调整，定义

$$\rho_k = \frac{f_k - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

其中分子表示函数实际减小的值；分母表示近似模型减少的值。分析 ρ_k：

1. 如果 ρ_k 小于0，一般情况下分母不可能小于0，因为目标函数求解的是最小值；此时说明分子小于0，即下一个目标点比上一步大，此时需要舍弃。
2. 如果 ρ_k 大于0并且接近1，说明模型和实际的预期比较相符合，此时可以考虑扩大Δ_k
3. 如果ρ_k大于0但是明显小于1，此时可以不用调整

4. 如果ρ_k大于0，但是接近0，说明模型变化范围比较大，但是实际改变比较小，此时应该收缩或者减少Δ_k

Algorithm 4.1 (Trust Region).

Given $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:

for $k = 0, 1, 2, \ldots$

  Obtain p_k by (approximately) solving (4.3);

  Evaluate ρ_k from (4.4);

  if $\rho_k < \frac{1}{4}$

    $\Delta_{k+1} = \frac{1}{4}\Delta_k$

  else

    if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$

      $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$

    else

      $\Delta_{k+1} = \Delta_k$;

  if $\rho_k > \eta$

    $x_{k+1} = x_k + p_k$

  else

    $x_{k+1} = x_k$;

end (for).

具体算法如下：

子问题的最优解

为简化形式，将信赖域问题的子问题表示为：

$$\min \mathrm{m(p)} = \mathrm{f} + \mathrm{g^T p} + \frac{1}{2}\mathrm{p^T B p} \quad \mathrm{s.t}\ ||\mathrm{p}|| \leq \Delta$$

该问题为标准的带不等式约束的二次优化问题，可以根据KKT条件（后面会深入介绍）得到该问题的最优解

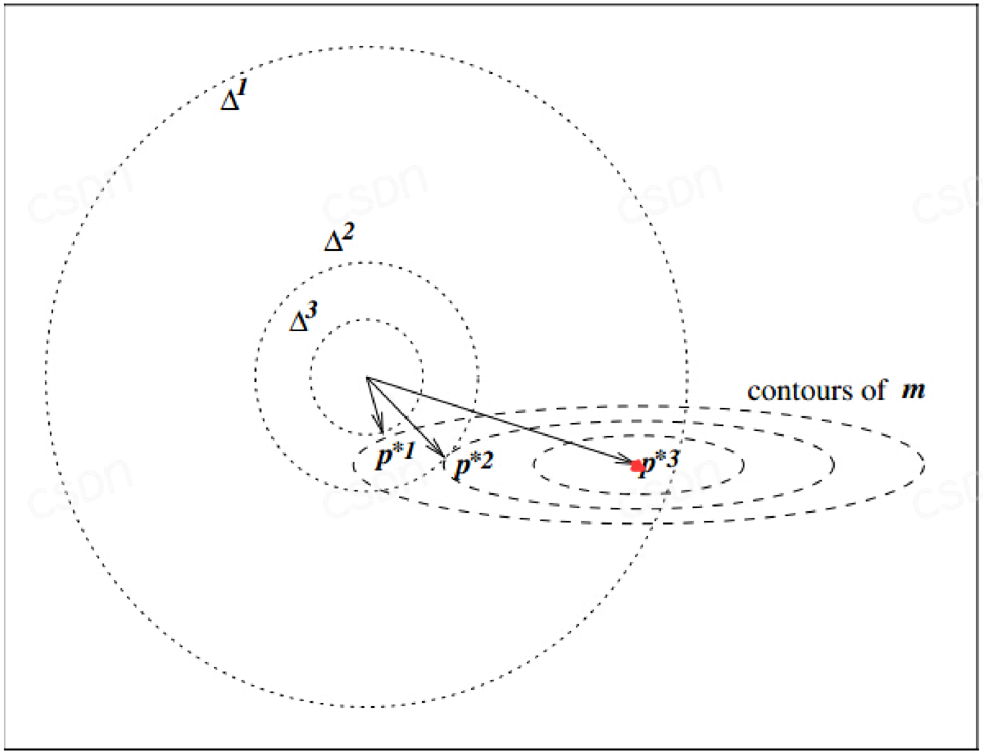
定理

如果向量 p^* 为子问题的最优解，当且仅当满足 p^* 为可行解，并且存在标量 λ 满足

KaTeX parse error: No such environment: align at position 7: \begin{align} & (B+\lambda …

其中条件（b）为补充条件，即要么λ为0要么$\Delta = ||\mathrm{p}^*||$。

从下图中可以看出最优解和参数λ的关系



当参数$\Delta = \Delta^1$时，最优解为p^3此时相当于没有约束，此时$\lambda = 0$
当参数$\Delta = \Delta^1$或者Δ^2时，最优解被球形约束限制，此时满足$\Delta = ||p^*||$，根据上面条件（a）有

$$\lambda p^* = -Bp^* - g$$

如果能够找到这样的p满足这些条件就能找到最优解。

基于柯西点（Cauchy-Point）的算法

在实际求解过程中不一定找到最优解，而是找到一个充分下降的点满足全局收敛即可。柯西点就是满足该条件的点（p_k^c）

柯西点（Caychy-Point）算法

求解步骤如下

1. 计算原子问题的线性蜕化问题，寻找向量p_k^s满足

$$p_k^s = \arg\min f_k + g_k^T p \quad s.t. ||p|| \le \Delta_k$$

2. 寻找标量$\tau_k > 0$满足 $\min m_k(\tau p_k^s)$ 同时满足信赖域的约束，即

$$\tau_k = \arg\min m_k(\tau p_k^s) \quad \text{s.t.}\ ||\tau p_k^s|| \le \Delta_k$$

3. $p_k^c = \tau_k p_k^s$

分别解释如下：

1. 计算p_k^s 从上述步骤1中可以看出求解步骤1有必使解，$p_k^s = -\frac{\Delta_k}{||g_k||}g_k$。两种思路，一是退化后的函数为线性函数，而且是递减的，只要取下界即可。二是利用KKT条件也可以推出。
2. 将求解得到的p_k^s代入子问题可以得到，

$$\min m(p) = f + g^T p + \frac{1}{2}p^T B p \quad \text{s.t}\ ||p|| \le \Delta$$
$$\min m(\tau p_k^s) = f - g^T \tau \frac{\Delta_k}{||g_k||}g_k + \frac{1}{2}\tau^2 \frac{\Delta_k}{||g_k||}g_k^T B_k \frac{\Delta_k}{||g_k||}g_k$$
$$\text{s.t.}\ ||\tau p_k^s|| \le \Delta_k$$

   是一个关于τ 的二次函数，如果
   1）$g_k^T B_k g_k \le 0$，是一个关于τ的递减函数，直接得到$\tau = 1$
   2)如果$g_k^T B_k g_k > 0$ 求导可以得到KaTeX parse error: Undefined control sequence: \e at position 68: …时\tau需要满足 \tau \e 1

柯西点很容易计算，但是如果只利用柯西点，相当于只利用了梯度方向，相当于线搜索的扩展而已，即**收敛速度为线性收敛。**

Dogleg算法

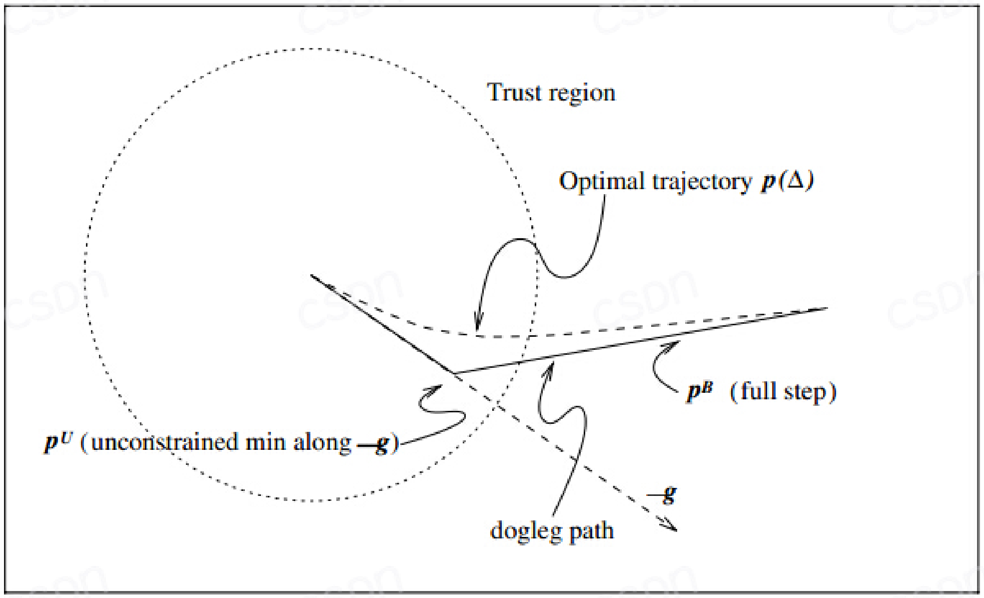
该方法适用于**当**B_k正定时，寻找半径Δ和最优解$p^*(\Delta)$之间的关系。
在之前定义了完全步(Full Step)，即$||B_k^{-1}g_k|| \le \Delta_k$则 $p_k^B = -B_k^{-1}g_k$
该算法的思路为，$p^*(\Delta)$表示不同Δ条件下的最优解，
1）$p^*(\Delta) = p^B \quad ||p^B|| \le \Delta$，否则
2）

$$p(\tau) = \begin{cases} \tau p^u, & 0 < \tau < 1 \\ p^u + (\tau - 1)(p^B - p^u), & 1 < \tau < 2 \end{cases}$$

其中p^u定义为

$$p^u = -\frac{g^T g}{g^T B g} g$$

，沿着梯度方向的最优解。
此为Dogleg方法，即寻找一个折线即两个线段的交点，即一条线段是沿着负梯度方向的最优值；二是沿着p^U到p^B方向。在此折线上寻找最优解。
由定理可以证明$||p(\tau)||$是一个关于τ的递增函数，而$m(p(\tau))$是一个关于τ的递减函数，又是一个线性函数，因此只要计算$p(\tau)$和$||p|| = \Delta$的交点即可。从下图可以清晰看到



二维子空间最优化

在DogLeg算法中，可以理解为最优解p^*可以表示为p^U和p^B的扩展子空间，即$p^* = \lambda_1 g + \lambda_2 B^{-1} g$，则原子问题可以表示为：

$$\min m(p) = f + g^T p + \frac{1}{2} p^T B p$$
$$s.t\ ||p|| \le \Delta$$
$$p \in span[g, -B^{-1} g]$$

迭代算法

根据子问题的最优解形式可以得到
1）当$\lambda = 0$时需要满足

$$Bp^* = -g$$

$$B正定$$
$$||p^*|| \neq \Delta$$

2) 当$\lambda \neq 0$时需要满足

$$(B + \lambda \mathbf{I})p^* = -g$$
$$B + \lambda \mathbf{I}正定$$
$$||p^*|| = \Delta$$

即

$$||p^*|| = || - (B + \lambda \mathbf{I})^{-1} g|| = \Delta$$

通过定义$p(\lambda) = -(B + \lambda \mathbf{I})^{-1} g$，寻找特定的$\lambda$使得$||p(\lambda)|| = \Delta$

相关定义

由于B正定因此可以进行正交分解,$B = Q\Lambda Q^T$； $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$并且$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

$$B + \lambda \mathbf{I} = Q(\Lambda + \lambda \mathbf{I})Q^T$$
$$p(\lambda) = Q(\Lambda + \lambda \mathbf{I})^{-1} Q^T g$$
$$p(\lambda)^2 = -\sum_1^n \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}$$

由于是正交分解因此$q_i q_j = 1$

迭代算法

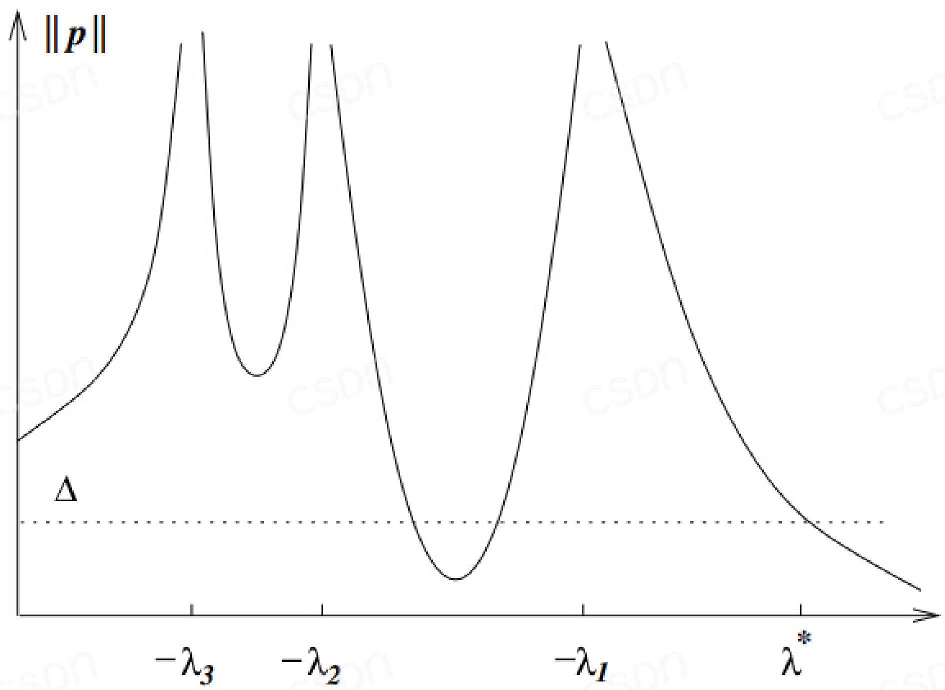
1）由上述定义

$$p(\lambda)^2 = -\sum_1^n \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}$$

可以看出

2） 当$\lambda > -\lambda_1$时函数值是一个单调递减函数，特别当$\lim_{\lambda \to \infty}(p(\lambda)^2) \to 0$

3)因此我们需要寻找$\lambda \in (-\infty, \lambda_1)$使得$||\mathrm{p}|| = \Delta$



λ求解

在$\mathrm{q_i^T g} \neq 0$

1）通过牛顿迭代法求解

$$\phi(\lambda) = ||\mathrm{p}(\lambda)|| - \Delta = 0$$

2）近似算法求解

$$\phi_1(\lambda) = \frac{\mathrm{C1}}{\lambda + \lambda_1} + \mathrm{C2}$$
$$\mathrm{phi}_1(\lambda) = \frac{1}{\Delta} - \frac{\lambda + \lambda_3}{\mathrm{C3}}$$

当$\mathrm{q_i^T g} = 0$是一个Hard Case，此时可以添加一个误差因子进行近似

$$p(\lambda) = Q(\Lambda + \lambda \mathbf{I})^{-1} Q^T g + \tau \mathbf{E}$$

信赖域的其他扩展

1. poor scaling问题，可以扩展为球形或者椭圆形
2. 可以构造Scaled的算法

$$\min m(p) = f + g^T p + \frac{1}{2} p^T B p \quad s.t\ ||Dp|| \leq \Delta$$

3. 矩阵D的构造和$\partial^2 f / \partial x_i$相关
4. 可以构造通用的柯西点废
5. 其他Norm方法，例如

$$||p||_1 \leq \Delta$$
$$||p||_\infty \leq \Delta$$

总结

通过该节需要了解

1. 信赖域方法和线搜索方法的不同
2. 信赖域方法的基本形式
3. 信赖域方法的柯西点算法、DogLeg算法和最优解迭代算法
4. 信赖域方法收敛