Systems Engineering Institute
Ministry of Education Key Lab for Intelligent Networks and Network Security

最速下降法与共轭梯度法
Method of steepest descent & Method of conjugate gradients

电信学院·自动化科学与技术系
系统工程研究所
吴江

Outline

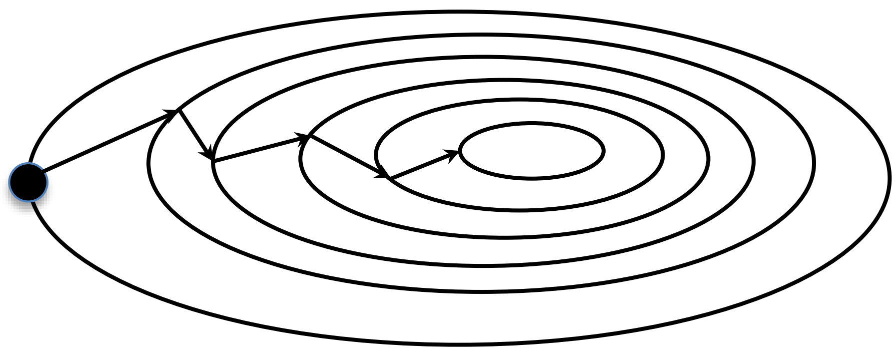
- 最速下降法
- 共轭梯度法

无约束优化问题

- n元函数的无约束非线性规划问题：

$$\min \quad f(x)$$

其中$x=(x_1, x_2, \ldots, x_n)^T \in R^n$, $f: R^n \to R^1$



最优性条件

- 设 $f: R^n \to R^1$在点 $\bar{x} \in R^n$ 处可微。若存在$p \in R^n$, 使

$$\nabla f(\bar{x})^T p < 0$$

则向量 p 是 f 在点 $\bar{x}$ 处的**下降方向**

- 设 $f: R^n \to R^1$在点$x^* \in R^n$ 处可微。若x^*是无约束优化问题的**局部最优解**，则:

$$\nabla f(x^*) = 0$$

- 设 $f: R^n \to R^1$在点$x^* \in R^n$ 处的Hesse矩阵存在。若:

  $\boldsymbol{\nabla f(x^*)=0}$, 且$\boldsymbol{\nabla^2 f(x^*)}$**正定**,

  则x^*是无约束优化问题的严格**局部最优解**.

最速下降法

- 性质：负梯度方向是比较容易获取的下降方向，同时是方向导数负向最大的方向。

$$f(x^k)-f(x^k+tp^k)=-t\nabla f(x^k)^T p^k+o(\|tp^k\|)$$

- 算法：
  - Step 0　给定$x^0 \in R^n$, $\varepsilon>0$, $k=0$
  - Step 1　while $\|\nabla f(x^k)\| > \varepsilon$

    $t_k=\text{argmin } f(x^k-t\nabla f(x^k))$

    $x^{k+1}=x^k-t_k\nabla f(x^k)$), $k=k+1$

    end
  - Step 2　输出x^k

例：

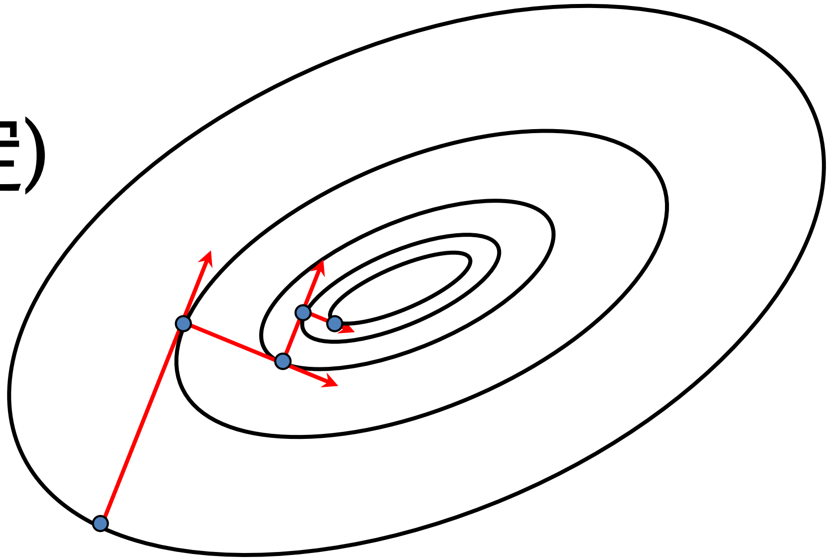
$$\min f(x_1, x_2) = x_1^2 + x_2^2, x^0 = (2,2)^T, \varepsilon = 10^{-6}$$

$$\min f(x_1, x_2) = x_1^2 + 25x_2^2, x^0 = (2,2)^T, \varepsilon = 10^{-6}$$

例

相邻两次搜索方向正交(垂直)!

$$\min\ f(x)=\frac{1}{2}x^{T}Ax\quad (A\text{正定})$$



采用最速下降法时, 有:

$$f(x^{k+1})\le\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 f(x^k)$$

λ_n, λ_1分别为A的最大, 最小特征值

搜索方向的思考

最速下降法收敛较慢，怎样改进？

最速下降法由一阶导数确定$p^{(k)}$，能否利用二阶导数信息？

利用精确Hesse阵：Newton法、信赖域法

利用近似Hesse阵：拟Newton法

间接利用Hesse阵部分信息：共轭梯度法

共轭梯度法

- 设 $f: R^n \rightarrow R^1$在点$x^* \in R^n$ 处可微。若x^*是无约束优化问题的**局部最优解**，则： $\nabla f(x^*)=0$
- 设 $f: R^n \rightarrow R^1$在点$x^* \in R^n$ 处的Hesse矩阵存在。若：

  $\nabla f(x^*)=0$, 且$\nabla^2 f(x^*)$**正定**,

  则x^*是无约束优化问题的严格**局部最优解**.

$$f(x^k)-f(x^k+tp^k)=-t\nabla f(x^k)^T p^k+o(\|tp^k\|)$$

一种无约束优化算法, 如果对凸二次函数是有效的, 那么它很有可能对于一般的无约束优化问题也是有效的

凸二次函数的最优解

目标函数	$\frac{1}{2}x^T Ax + b^T x + c$	$f(x)$
梯度	$\boldsymbol{Ax+b}$	$\nabla f(x)$
Hesse阵	$\boldsymbol{A}$	$\nabla^2 f(x)$
一阶必要条件	$\boldsymbol{Ax^*+b=0}$	$\nabla f(x^*) = 0$

不利用 $\boldsymbol{A}$ $(\nabla^2 f(x))$ 的信息，怎样得到x^*？

凸二次函数的搜索方向

$$\min_{x\in R^n} f(x)=\frac{1}{2}x^T Ax+b^T x+c\,,\ A\text{对称正定，}$$

$$x^{(0)}\in R^n,\ p^{(0)}\in R^n(p^{(0)}\text{可以是}-\nabla f(x^{(0)}))$$

第一步：采用精确LS

$$\varphi(t)=f(x^{(0)}+t\,p^{(0)})$$

$$t_0=\arg\min_{t\ge 0}\varphi(t)$$

$$x^{(1)}=x^{(0)}+t_0 p^{(0)}$$

$$\Rightarrow \varphi'(t_0)=0$$

$$\Rightarrow \nabla f(x^{(0)}+t_0\,p^{(0)})^T p^{(0)}=0$$

$$\Rightarrow \nabla f(x^{(1)})^T p^{(0)}=0$$

第二步：从$x^{(1)}$出发，沿什么样的方向搜索，才可能获得x^*？

$$\left.\begin{aligned}&\nabla f(x^{(1)})^T p^{(0)}=0\\&\nabla f(x^{(1)})=Ax^{(1)}+b\\&Ax^*+b=0\end{aligned}\right\}\Rightarrow (x^{(1)}-x^*)^T Ap^{(0)}=0$$

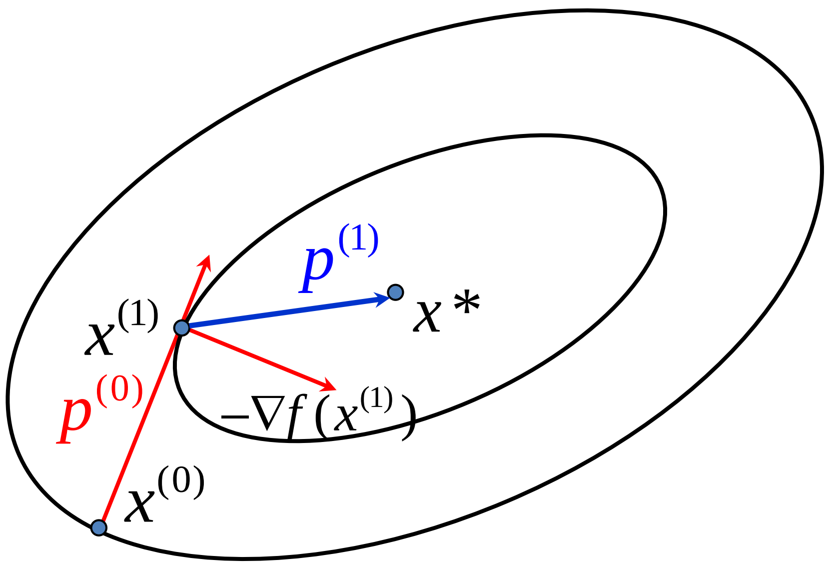
$$(p^{(1)})^T Ap^{(0)}=0$$

结论：

1. x^*位于使$p^TAp^{(0)}=0$的方向p上
2. 对n维问题，仍不能唯一确定方向，对二维问题可唯一确定

共轭方向

- 定义:
  - 设A为n阶实对称阵, 若非零向量$p,q\in R^n$满足$p^TAq=0$, 则称p和q是**关于A**共轭的(相互A共轭).
  - 若有一个非零向量组$p^{(i)}$, $i=0, 1, \ldots, n-1$满足$\forall i\neq j$, $p^{(i)\,T}Ap^{(j)}=0$, 则称$\{p^{(0)}, \ldots, p^{(n-1)}\}$是一个关于$A$共轭的**向量组**.



Conjugate direction

$$(p^{(1)})^T Ap^{(0)}=0$$

例:求所有与q关于A共轭的方向p

$$A=\begin{bmatrix}2 & 1\\ 1 & 3\end{bmatrix}, q=\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$p=\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \Rightarrow \begin{pmatrix}\alpha & \beta\end{pmatrix}\begin{bmatrix}2 & 1\\ 1 & 3\end{bmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}=0$$

$$\Rightarrow 3\alpha+4\beta=0 \Rightarrow p=t\begin{pmatrix}4\\ -3\end{pmatrix}, t\neq 0$$

共轭方向组的性质

- 定理5：设A为n阶对称正定矩阵, $p^{(i)}$, $i=0, 1, \ldots, k$是一组非零向量, 若$\{p^{(i)}\}$关于A共轭, 则$\{p^{(i)}\}$**线性无关**.

- 定理6: 设$f(x)= 1/2x^TAx+b^Tx+c$, A为n阶对称正定矩阵, $b\in R^n$, $c \in R^1$. $\{p^{(0)}, \ldots, p^{(n-1)}\}$为一组关于$A$共轭的方向. 则$x^{(0)}\in R^n$, 从$x^{(0)}$出发, 依次沿$p^{(0)}, \ldots, p^{(n-1)}$进行精确一维搜索, 至多$n$次即可获得$x^*$, x^*是满足$Ax^*+b=0$的解**(二次终止性)**

共轭方向组的性质

- 定理5：设A为n阶对称正定矩阵，$p^{(i)}$, $i=0, 1, \ldots, k$是一组非零向量，若$\{p^{(i)}\}$关于A共轭，则$\{p^{(i)}\}$**线性无关**.

为什么选共轭方向作为搜索方向？

选择共轭方向的理由？

凸二次函数无约束极小：$\min\limits_{x\in R^n} f(x)=\dfrac{1}{2}x^T Ax+b^T x+c$，$A$ 对称正定

迭代过程：$x^{(0)} \xrightarrow{t_0 p^{(0)}} x^{(1)} \xrightarrow{t_1 p^{(1)}} x^{(2)} \xrightarrow{t_2 p^{(2)}} \cdots \xrightarrow{t_{k-1} p^{(k-1)}} x^{(k)} \xrightarrow{t_k p^{(k)}} x^{(k+1)}$

不直接利用矩阵A信息前提下，能提出的最高合理要求：

$x^{(k+1)}$是$F(\alpha_0,\alpha_1,\cdots,\alpha_k)=f(x^{(0)}+\alpha_0 p^{(0)}+\alpha_1 p^{(1)}+\cdots+\alpha_k p^{(k)})$

的全局极小！

若能实现此目标，则至多经过**n**次迭代，可得凸二次函数***x****

F仍为凸函数，全局极小意味着：

$$\nabla F=0 \Leftrightarrow \left.\frac{\partial F}{\partial \alpha_i}\right|_{(t_0,t_1,\cdots,t_k)}=0\ (\forall i, 0\le i\le k)$$

$$0=\left.\frac{\partial F}{\partial \alpha_i}\right|_{(t_0,t_1,\cdots,t_k)}=\left(\nabla f(x^{(k+1)})\right)^T p^{(i)}$$

$$\forall i, \left(\nabla f(x^{(k+1)})\right)^T p^{(i)}=0?$$

选择共轭方向的理由

凸二次函数无约束极小: $\min\limits_{x\in R^n} f(x)=\frac{1}{2}x^TAx+b^Tx+c$, A 对称正定

迭代过程: $x^{(0)}\xrightarrow{t_0p^{(0)}}x^{(1)}\xrightarrow{t_1p^{(1)}}x^{(2)}\xrightarrow{t_2p^{(2)}}\cdots\xrightarrow{t_{k-1}p^{(k-1)}}x^{(k)}\xrightarrow{t_kp^{(k)}}x^{(k+1)}$

$$\forall i\ (0\le i\le k),\left(\nabla f(x^{(k+1)})\right)^T p^{(i)}=$$

全部求和

$$\left\{\begin{array}{l}\left(\nabla f(x^{(k+1)})-\nabla f(x^{(k)})\right)^T p^{(i)}\\ \left(\nabla f(x^{(k)})-\nabla f(x^{(k-1)})\right)^T p^{(i)}\\ \vdots\\ \left(\nabla f(x^{(i+2)})-\nabla f(x^{(i+1)})\right)^T p^{(i)}\\ \left(\nabla f(x^{(i+1)})\right)^T p^{(i)}\end{array}\right.$$

精确线性搜索可确保取值为0

$$\nabla f(x)=Ax+b\Rightarrow$$

$$\nabla f(x^{(j+1)})-\nabla f(x^{(j)})=A(x^{(j+1)}-x^{(j)})$$

$$\Rightarrow\left(\nabla f(x^{(j+1)})-\nabla f(x^{(j)})\right)^T p^{(i)}$$

$$=(x^{(j+1)}-x^{(j)})^T Ap^{(i)}$$

$$=t_j\left(p^{(j)}\right)^T Ap^{(i)}$$

共轭方向可确保 $i\neq j$ 时取值为0

目标实现!

核心目标

$$\forall i,\left(\nabla f(x^{(k+1)})\right)^T p^{(i)}=0?$$

共轭方向的性质

凸二次函数无约束极小：$\min\limits_{x\in R^n} f(x)=\frac{1}{2}x^T Ax+b^T x+c$，$A$ 对称正定

迭代过程：$x^{(0)} \xrightarrow{t_0 p^{(0)}} x^{(1)} \xrightarrow{t_1 p^{(1)}} x^{(2)} \xrightarrow{t_2 p^{(2)}} \cdots \xrightarrow{t_{k-1} p^{(k-1)}} x^{(k)} \xrightarrow{t_k p^{(k)}} x^{(k+1)}$

定理6： 设A为n阶对称正定矩阵, $p^{(i)}$, i =0, …, $n-1$是一组非零向量，若$\{p^{(i)}\}$关于A共轭，则从任意 $x^{(0)}\in R^n$出发，依次沿$p^{(i)}$, i =0, …, 进行精确线性搜索，至多n次迭代即可获得函数 $f(x)=\frac{1}{2}x^T Ax+b^T x+c$ 的全局极小。

出现时$\nabla f(x^{(k)})=0$即停止迭代

问题：怎样产生共轭搜索方向？

共轭方向法具有二次终止性

最速下降法无二次终止性

共轭方向怎样生成?

凸二次函数无约束极小: $\min_{x\in R^n} f(x)=\frac{1}{2}x^T Ax+b^T x+c$, A 对称正定

迭代过程: $x^{(0)} \xrightarrow{t_0 p^{(0)}} x^{(1)} \xrightarrow{t_1 p^{(1)}} x^{(2)} \xrightarrow{t_2 p^{(2)}} \cdots \xrightarrow{t_{k-1} p^{(k-1)}} x^{(k)} \xrightarrow{t_k p^{(k)}} x^{(k+1)}$

中间信息: $g^{(0)},\quad g^{(1)},\quad g^{(2)},\quad \cdots\cdots \quad g^{(k)},\quad g^{(k+1)}$

$g^{(i)}=\nabla f(x^{(i)})$

生成共轭方向? 对第二次迭代搜索方向的合理猜想: $p^{(1)}=-g^{(1)}+\lambda p^{(0)}$?

$$\left(p^{(1)}\right)^T Ap^{(0)}=0 \Rightarrow \lambda=\frac{\left(g^{(1)}\right)^T Ap^{(0)}}{\left(p^{(0)}\right)^T Ap^{(0)}}$$

表达式中仍含有A?

$$\nabla f(x)=Ax+b \Rightarrow$$

$$g^{(j+1)}-g^{(j)}=A(x^{(j+1)}-x^{(j)})$$

$$\Rightarrow Ap^{(j)}=\frac{1}{t_j}A(x^{(j+1)}-x^{(j)})$$

$$=\frac{1}{t_j}(g^{(j+1)}-g^{(j)})$$

$$\lambda=\frac{\left(g^{(1)}\right)^T (g^{(1)}-g^{(0)})}{\left(p^{(0)}\right)^T (g^{(1)}-g^{(0)})}$$

非常优美的公式

共轭方向-一般公式

凸二次函数无约束极小: $\min\limits_{x\in R^n} f(x)=\frac{1}{2}x^T Ax+b^T x+c$, A 对称正定

迭代过程: $x^{(0)} \xrightarrow{t_0 p^{(0)}} x^{(1)} \xrightarrow{t_1 p^{(1)}} x^{(2)} \xrightarrow{t_2 p^{(2)}} \cdots \xrightarrow{t_{k-1} p^{(k-1)}} x^{(k)} \xrightarrow{t_k p^{(k)}} x^{(k+1)}$

中间信息: $g^{(0)},\quad g^{(1)},\quad g^{(2)},\quad \cdots\cdots \quad g^{(k)},\quad g^{(k+1)}$

$g^{(i)}=\nabla f(x^{(i)})$　　由梯度产生共轭方向，故称共轭梯度法

$$p^{(k+1)}=-g^{(k+1)}+\lambda_k p^{(k)},\quad \text{要求 } p^{(0)}=-g^{(0)}$$

$$\lambda_k=\frac{\left(g^{(k+1)}\right)^T g^{(k+1)}}{\left(g^{(k)}\right)^T g^{(k)}}\quad \text{(Fletcher-Reeves, FR)}$$

$$\lambda_k=\frac{\left(g^{(k+1)}\right)^T \left(g^{(k+1)}-g^{(k)}\right)}{\left(g^{(k)}\right)^T g^{(k)}}\quad \text{(Polak-Ribiere-Polyak, PRP)}$$

共轭梯度法

- Step 1: 给定$x^{(0)} \in R^n$, $\varepsilon>0$, $k=0$, $p^{(0)}=-\nabla f(x^{(0)})$, $l=1$
- Step 2: while $\| \nabla f(x^{(k)}) \| > \varepsilon$

  $t_k=\arg\min f(x^{(k)}-tp^{(k)})$

  $x^{(k+1)}=x^k+t_k p^{(k)}$

  if $l=n$ then

  $l=1$, $p^{(k+1)}=-\nabla f(x^{(k+1)})$

  else

  $p^{(k+1)}=-\nabla f(x^{(k+1)})+\lambda_k p^{(k)}$

  end

  $k=k+1$, $l=l+1$

  end
- Step 3: 输出$x^{(k)}$

$$\lambda_k=\begin{cases}\dfrac{\nabla f(x^{(k+1)})^T\nabla f(x^{(k+1)})}{\nabla f(x^{(k)})^T\nabla f(x^{(k)})}\\[2ex] \dfrac{\nabla f(x^{(k+1)})^T[\nabla f(x^{(k+1)})-\nabla f(x^{(k)})]}{\nabla f(x^{(k)})^T\nabla f(x^{(k)})}\end{cases}$$

例子

例. 用FR共轭梯度法求解无约束最优化问题：

$$\min_{x\in R^2} f(x)=x_1^2+2x_2^2+4x_1+4x_2.\quad x^{(0)}=(0,0)^T$$

解. 第一次迭代： $k=1, l=1, x^{(0)}=(0,0)^T$

$$\nabla f(x)=\begin{pmatrix}2x_1+4\\4x_2+4\end{pmatrix}\Rightarrow \nabla f(x^{(0)})=\begin{pmatrix}4\\4\end{pmatrix}$$

$$p^{(0)}=\begin{pmatrix}-4\\-4\end{pmatrix},\ x^{(0)}+tp^{(0)}=\begin{pmatrix}-4t\\-4t\end{pmatrix}$$

$$\varphi(t)=f(x^{(0)}+tp^{(0)})=48t^2-32t$$

$$t_0=\arg\min_{t\ge 0}\varphi(t)=1/3$$

$$x^{(1)}=x^{(0)}+t_0p^{(0)}=\begin{pmatrix}-4/3\\-4/3\end{pmatrix}, k=2, l=2$$

$$f(x)=\frac{1}{2}x^TAx+b^Tx$$

$$A=\begin{bmatrix}2&0\\0&4\end{bmatrix}, b=\begin{pmatrix}4\\4\end{pmatrix}$$

$$x^*=-A^{-1}b=\begin{pmatrix}-2\\-1\end{pmatrix}$$

例子

例. 用FR共轭梯度法求解无约束最优化问题：

$$\min_{x\in R^2} f(x) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2. \quad x^{(0)} = (0, 0)^T$$

第二次迭代： $\nabla f(x^{(1)}) = \begin{pmatrix} 4/3 \\ -4/3 \end{pmatrix},\ p^{(1)} = \begin{pmatrix} -4/3 \\ 4/3 \end{pmatrix} + \frac{32/9}{32}\begin{pmatrix} -4 \\ -4 \end{pmatrix} = \begin{pmatrix} -16/9 \\ 8/9 \end{pmatrix}$

$$x^{(1)} + tp^{(1)} = \begin{pmatrix} -4/3 \\ -4/3 \end{pmatrix} + t\begin{pmatrix} -16/9 \\ 8/9 \end{pmatrix}$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 4 \\ 4x_2 + 4 \end{pmatrix}$$

$$\varphi(t) = f(x^{(1)} + tp^{(1)}) = \cdots\cdots$$

$$t_1 = \arg\min_{t\geq 0} \varphi(t) = 3/8, \quad x^{(2)} = x^{(1)} + t_1 p^{(1)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

$k = 2$, 重置 $l = 1$

第三次迭代： $\nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 停

从同样的初始点开始，用最速下降法求解，需无穷次迭代且会出现Zigzag现象。

作业

- P155: 19