最优化方法

授课教师：徐海涛

2024年秋

第三章 无约束优化问题

3.6 共轭梯度法

- 问题：对于最速下降法，迭代的次数可能会较多
- 相对于最速下降法，共轭梯度法的性能有了巨大的提升
- 针对n元二次型函数，共轭梯度法能够通过n次迭代得到极小点，这是最速下降法所不具备的
- 已经证明，共轭梯度法对于一般形式函数的极小点求解也非常有效

3.6 共轭梯度法

考虑二次型函数的极小化问题：

$$\text{minimize} \quad q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\,\mathbf{x} + \mathbf{c}^{\mathrm{T}}\mathbf{x}$$

其中，$\mathbf{A}$ 为对称正定矩阵。满足

$$\mathbf{d}_i^{\mathrm{T}}\mathbf{A}\mathbf{d}_j = 0, \quad i \neq j, \quad 0 \leqslant i, \ j \leqslant n$$

的向量或方向 $\mathbf{d}^i$ 和 $\mathbf{d}^j$ 是关于 $\mathbf{A}$ 共轭的，称为共轭方向。

$$\boldsymbol{d}_i{}^T\boldsymbol{d}_j = 0$$

$$\boldsymbol{d}_i{}^T\boldsymbol{A}\boldsymbol{d}_j = 0$$

3.6 共轭梯度法（算法）

共轭方向法首先要求确定初始点 $\mathbf{x}_0$ 和一组共轭方向 $\mathbf{d}_0,\mathbf{d}_1,\cdots,\mathbf{d}_{n-1}$，沿着方向 $\mathbf{d}_0$ 使得 $q(\mathbf{x})$ 达到最小，从而获得 $\mathbf{x}_1$。以此类推，沿着方向 $\mathbf{d}_2$ 使得 $q(\mathbf{x})$ 最小，获得 $\mathbf{x}_2$。最后，沿着方向 $\mathbf{d}_{n-1}$ 可获得 $\mathbf{x}_n$，这就是函数的极小点 $q(\mathbf{x})$。这意味着能够通过 n 次搜索获得二次型函数的极小点。该方法的关键就是如何获得共轭方向，在接下来介绍的算法中，利用 $q(\mathbf{x})$ 的梯度来产生共轭方向。

基本的共轭方向算法。给定初始点 $\boldsymbol{x}^{(0)}$ 和一组关于 $\boldsymbol{Q}$ 共轭的方向 $\boldsymbol{d}^{(0)}$，$\boldsymbol{d}^{(1)}$，…，$\boldsymbol{d}^{(n-1)}$，迭代公式为($k\geqslant 0$ 表示迭代次数)

$$\boldsymbol{g}^{(k)}=\nabla f(\boldsymbol{x}^{(k)})=\boldsymbol{Q}\boldsymbol{x}^{(k)}-\boldsymbol{b}$$

$$\alpha_k=-\frac{\boldsymbol{g}^{(k)\top}\boldsymbol{d}^{(k)}}{\boldsymbol{d}^{(k)\top}\boldsymbol{Q}\boldsymbol{d}^{(k)}}$$

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}$$

明理，精工，笃学，致远 5

3.6 共轭梯度法（证明）

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}-\boldsymbol{x}^\top\boldsymbol{b}$$

其中，$\boldsymbol{Q}=\boldsymbol{Q}^\top>0$，$\boldsymbol{x}\in\mathbb{R}^n$。注意，由于 $\boldsymbol{Q}>0$，因此函数 f 有一个全局极小点，可通过求解 $\boldsymbol{Q}\boldsymbol{x}=\boldsymbol{b}$ 得到。

定理 10.1 对于任意初始点 $\boldsymbol{x}^{(0)}$，基本的共轭方向算法都能在 n 次迭代之内收敛到唯一的全局极小点 $\boldsymbol{x}^*$，即 $\boldsymbol{x}^{(n)}=\boldsymbol{x}^*$。 □

证明：由于方向 $\boldsymbol{d}^{(i)}$，$i=0,1,\cdots,n-1$ 线性无关，因此，$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}\in\mathbb{R}^n$ 可由它们线性表出，即

$$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}=\beta_0\boldsymbol{d}^{(0)}+\cdots+\beta_{n-1}\boldsymbol{d}^{(n-1)}$$

其中，β_i，$i=0,1,\cdots,n-1$ 为常数。

明理，精工，笃学，致远 6

3.6 共轭梯度法（证明）

证明：由于方向 $\boldsymbol{d}^{(i)}$，$i=0,1,\cdots,n-1$ 线性无关，因此，$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}\in\mathbb{R}^n$ 可由它们线性表出，即

$$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}=\beta_0\boldsymbol{d}^{(0)}+\cdots+\beta_{n-1}\boldsymbol{d}^{(n-1)}$$

其中，β_i，$i=0,1,\cdots,n-1$ 为常数。

$$x^*-x^{(0)}=\beta_0 d^{(0)}+\cdots+\beta_{n-1}d^{(n-1)}$$

$$x^{(n)}-x^{(0)}=\alpha_0 d^{(0)}+\cdots+\alpha_{n-1}d^{(n-1)}$$

明理，精工，笃学，致远 7

3.6 共轭梯度法（证明）

证明：由于方向 $\boldsymbol{d}^{(i)}$，$i=0,1,\cdots,n-1$ 线性无关，因此，$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}\in\mathbb{R}^n$ 可由它们线性表出，即

$$\boldsymbol{x}^*-\boldsymbol{x}^{(0)}=\beta_0\boldsymbol{d}^{(0)}+\cdots+\beta_{n-1}\boldsymbol{d}^{(n-1)}$$

其中，β_i，$i=0,1,\cdots,n-1$ 为常数。

在等号两端同时左乘 $\boldsymbol{d}^{(k)\top}\boldsymbol{Q}$，$0\leqslant k<n$，由 $\boldsymbol{Q}$ 共轭的性质可知，$\boldsymbol{d}^{(k)\top}\boldsymbol{Q}\boldsymbol{d}^{(i)}=0$，$k\neq i$，因此有

$$\boldsymbol{d}^{(k)\top}\boldsymbol{Q}(\boldsymbol{x}^*-\boldsymbol{x}^{(0)})=\beta_k\boldsymbol{d}^{(k)\top}\boldsymbol{Q}\boldsymbol{d}^{(k)}$$

整理后，可得

$$\beta_k=\frac{\boldsymbol{d}^{(k)\top}\boldsymbol{Q}(\boldsymbol{x}^*-\boldsymbol{x}^{(0)})}{\boldsymbol{d}^{(k)\top}\boldsymbol{Q}\boldsymbol{d}^{(k)}}$$

明理，精工，笃学，致远 8

3.6 共轭梯度法（证明）

$$\beta_k = \frac{d^{(k)\top}Q(x^* - x^{(0)})}{d^{(k)\top}Qd^{(k)}}$$

迭代点 $x^{(k)}$ 可写为

$$x^{(k)} = x^{(0)} + \alpha_0 d^{(0)} + \cdots + \alpha_{k-1}d^{(k-1)}$$

即

$$x^{(k)} - x^{(0)} = \alpha_0 d^{(0)} + \cdots + \alpha_{k-1}d^{(k-1)}$$

而

$$x^* - x^{(0)} = (x^* - x^{(k)}) + (x^{(k)} - x^{(0)})$$

等式两端同时左乘 $d^{(k)\top}Q$，由于 $g^{(k)} = Qx^{(k)} - b$，$Qx^* = b$，可得

$$d^{(k)\top}Q(x^* - x^{(0)}) = d^{(k)\top}Q(x^* - x^{(k)}) = -d^{(k)\top}g^{(k)}$$

明理，精工，笃学，致远 9

3.6 共轭梯度法（证明）

$$\beta_k = \frac{d^{(k)\top}Q(x^* - x^{(0)})}{d^{(k)\top}Qd^{(k)}}$$

$$d^{(k)\top}Q(x^* - x^{(0)}) = d^{(k)\top}Q(x^* - x^{(k)}) = -d^{(k)\top}g^{(k)}$$

因此有

$$\beta_k = -\frac{d^{(k)\top}g^{(k)}}{d^{(k)\top}Qd^{(k)}} = \alpha_k$$

这说明 $x^* = x^{(n)}$，证明完成。

$$\boxed{x^* - x^{(0)} = \beta_0 d^{(0)} + \cdots + \beta_{n-1}d^{(n-1)}}$$

明理，精工，笃学，致远 10

3.6 共轭梯度法（根据梯度来获得迭代条件）

$$f(x) = \frac{1}{2}x^\top Qx - x^\top b$$

基本的共轭方向算法。给定初始点 $x^{(0)}$ 和一组关于 Q 共轭的方向 $d^{(0)}$，$d^{(1)}$，…，$d^{(n-1)}$，迭代公式为（$k \geqslant 0$ 表示迭代次数）

$$g^{(k)} = \nabla f(x^{(k)}) = Qx^{(k)} - b$$

$$\alpha_k = -\frac{g^{(k)\top}d^{(k)}}{d^{(k)\top}Qd^{(k)}}$$

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$$

明理，精工，笃学，致远 11

3.6 共轭梯度法（如何获得共轭方向）

考虑二次型函数的极小化问题：

$$\text{minimize} \quad q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\,\mathbf{x} + \mathbf{c}^\top \mathbf{x}$$

其中，$\mathbf{A}$ 为对称正定矩阵。满足

$$\mathbf{d}_i^\top \mathbf{A}\mathbf{d}_j = 0, \quad i \neq j, \quad 0 \leqslant i, j \leqslant n$$

记 $q(\mathbf{x})$ 的梯度为 $\mathbf{g}$，则 $\mathbf{g}_k = \nabla q(\mathbf{x}_k) = \mathbf{A}\mathbf{x}_k + \mathbf{c}$。点 $\mathbf{x}_k$ 为第 k 次迭代对应的迭代点，初始点对应的下降方向 $\mathbf{d}_0$ 为最速下降方向，即 $-\mathbf{g}_0$。沿着方向 $\mathbf{d}_k$ 得到函数 $q(\mathbf{x})$ 达到最小，以此产生新点 $\mathbf{x}_{k+1}$：

$$\boxed{\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k} \Leftarrow \alpha_k = -\frac{\mathbf{d}_k^\top \mathbf{g}_k}{\mathbf{d}_k^\top \mathbf{A}\,\mathbf{d}_k}$$

明理，精工，笃学，致远 12

3.6 共轭梯度法（如何获得共轭方向）

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \quad \Leftarrow \quad \alpha_k = -\frac{\mathbf{d}_k^{T}\mathbf{g}_k}{\mathbf{d}_k^{T}\mathbf{A}\,\mathbf{d}_k}$$

按照微分法则，由 $dq(\alpha)/d\alpha = 0$ 可推导出

$$\mathbf{d}_k^{T}\mathbf{g}_{k+1} = 0$$

$$\frac{dq(\alpha)}{d\alpha} = \nabla f(x_k + \alpha d_k)\cdot d_k = g_{k+1}\cdot d_k = 0$$

接下来就是最关键的步骤了，按照下式确定下降方向 $\mathbf{d}_{k+1}$：

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$$

上式的含义为对最速下降方向 $-\mathbf{g}_{k+1}$ 进行适当的“纠偏”，如图 3.6 所示。由于要求 $\mathbf{d}_{k+1}$ 和 $\mathbf{d}_k$ 关于 $\mathbf{A}$ 共轭，即 $\mathbf{d}_{k+1}^{T}\mathbf{A}\mathbf{d}_k = 0$ 由此可得

$$-\mathbf{g}_{k+1}^{T}\mathbf{A}\,\mathbf{d}_k + \beta_k \mathbf{d}_k^{T}\mathbf{A}\,\mathbf{d}_k = 0$$

3.6 共轭梯度法（如何获得共轭方向）

上式的含义为对最速下降方向 $-\mathbf{g}_{k+1}$ 进行适当的“纠偏”，如图 3.6 所示。由于要求 $\mathbf{d}_{k+1}$ 和 $\mathbf{d}_k$ 关于 $\mathbf{A}$ 共轭，即 $\mathbf{d}_{k+1}^{T}\mathbf{A}\mathbf{d}_k = 0$ 由此可得

$$-\mathbf{g}_{k+1}^{T}\mathbf{A}\,\mathbf{d}_k + \beta_k \mathbf{d}_k^{T}\mathbf{A}\,\mathbf{d}_k = 0$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{T}\boldsymbol{A}\boldsymbol{d}_k}{\boldsymbol{d}_k{}^{T}\boldsymbol{A}\boldsymbol{d}_k}$$

$$g^{(k)} = \nabla f(\boldsymbol{x}^{(k)}) = \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}$$

$$\alpha_k = -\frac{\boldsymbol{g}^{(k)\top}\boldsymbol{d}^{(k)}}{\boldsymbol{d}^{(k)\top}\boldsymbol{Q}\boldsymbol{d}^{(k)}}$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$$

3.6 共轭梯度法（一般性）

对于一般形式函数的极小点求解也非常有效

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{T}\boldsymbol{A}\boldsymbol{d}_k}{\boldsymbol{d}_k{}^{T}\boldsymbol{A}\boldsymbol{d}_k}$$

由式（3.20）可得，$\mathbf{d}_k = (\mathbf{x}_{k+1} - \mathbf{x}_k)/\alpha_k$，因此，$\mathbf{A}\mathbf{d}_k = (\mathbf{g}_{k+1} - \mathbf{g}_k)/\alpha_k$，代入上式并整理后可得

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{T}(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\alpha_k \boldsymbol{d}_k{}^{T}\boldsymbol{A}\boldsymbol{d}_k}$$

3.6 共轭梯度法（一般性）

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{T}(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\alpha_k \boldsymbol{d}_k{}^{T}\boldsymbol{A}\boldsymbol{d}_k} \qquad \alpha_k = \frac{-\boldsymbol{d}_k^{T}\boldsymbol{g}_k}{\boldsymbol{d}_k{}^{T}\boldsymbol{A}\boldsymbol{d}_k}$$

$$\boldsymbol{d}_k^{T}\boldsymbol{g}_{k+1} = 0$$

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k \quad \Rightarrow \quad \boldsymbol{d}_{k+1}^{T}\boldsymbol{g}_{k+1} = -\boldsymbol{g}_{k+1}^{T}\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k^{T}\boldsymbol{g}_{k+1}$$

3.6 共轭梯度法（一般性）

$$\boldsymbol{d}_k^T \boldsymbol{g}_{k+1} = 0$$

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k \Rightarrow \boldsymbol{d}_{k+1}^T \boldsymbol{g}_{k+1} = -\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k^T \boldsymbol{g}_{k+1}$$

$$\boldsymbol{d}_{k+1}^T \boldsymbol{g}_{k+1} = -\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1} \qquad \boldsymbol{d}_k^T \boldsymbol{g}_k = -\boldsymbol{g}_k^T \boldsymbol{g}_k$$

$$\alpha_k = \frac{-\boldsymbol{d}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k{}^T \boldsymbol{A} \boldsymbol{d}_k} \Rightarrow \alpha_k = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k{}^T \boldsymbol{A} \boldsymbol{d}_k}$$

3.6 共轭梯度法（一般性）

$$\alpha_k = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k{}^T \boldsymbol{A} \boldsymbol{d}_k} \Rightarrow \beta_k = \frac{\boldsymbol{g}_{k+1}^T (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\alpha_k \boldsymbol{d}_k{}^T \boldsymbol{A} \boldsymbol{d}_k}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^T (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$$

由式（3.32）可以看出，β_k 与矩阵 **A** 和向量 **c** 无关，这意味着该算法也适用于非二次型函数的极小点求解问题。但是，对于非二次型函数，式（3.31）所示的步长公式就不再适用了，必须采用数值线性搜索方法来求解步长 α_k。当然，通过 n 步迭代就可以得到最优解这一特性只适用于二次型函数。对于非二次型函数，应该每 n 次迭代完成之后，都需要将下降方向重置为当前梯度的负方向，即最速下降方向。采用式（3.32）来计算 β_k 的共轭梯度法称为 Polak – Rebiere 算法。

3.6 共轭梯度法（一般性）

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^T (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\boldsymbol{g}_k^T \boldsymbol{g}_k} \Rightarrow \beta_k = \frac{\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$$

$$\begin{aligned} \boldsymbol{g}_{k+1}^T \boldsymbol{g}_k &= \boldsymbol{g}_{k+1}^T (-\boldsymbol{d}_k + \beta_{k-1} \boldsymbol{d}_{k-1}) = \beta_{k-1} \boldsymbol{g}_{k+1}^T \boldsymbol{d}_{k-1} \\ &= \beta_{k-1} (\boldsymbol{g}_k^T + \alpha_k \boldsymbol{d}_k^T \boldsymbol{A}) \boldsymbol{d}_{k-1} \\ &= 0 \end{aligned}$$

$$\boxed{\mathbf{A}\mathbf{d}_k = (\mathbf{g}_{k+1} - \mathbf{g}_k)/\alpha_k}$$

这称为 Fletcher-Reeves 算法［Fletcher and Reeves，1964］

3.6 共轭梯度法（例题）

例 3.9 函数为 $f = x_1^2 + 4x_2^2$，初始点为 $\mathbf{x}_0 = [1,1]^{\mathrm{T}}$。利用共轭梯度法开展两次迭代，即可得到极小点。第一次迭代对应的下降方向为最速下降方向，即

$$\mathbf{d}_0 = -\nabla f(\mathbf{x}_0) = -[2, 8]^{\mathrm{T}}$$

此处未将方向向量进行归一化处理，不影响后续计算。求函数

$$f(\alpha) = f(\mathbf{x}_0 + \alpha \mathbf{d}_0) = (1 - 2\alpha)^2 + 4(1 - 8\alpha)^2$$

的极小点，可得 $\alpha_0 = 0.130\,769\,2$，构造新点为 $\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 = [0.738\,461\,5, -0.046\,153\,8]^{\mathrm{T}}$。进入到下一次迭代，计算

$$\beta_0 = \frac{\|\nabla f(\mathbf{x}^1)\|^2}{\|\nabla f(\mathbf{x}^0)\|^2} = 2.3176/68 = 0.034\,082\,8$$

$$\mathbf{d}_1 = -\nabla f^{\mathrm{T}}(\mathbf{x}_1) + \beta_0 \mathbf{d}_0 = \begin{pmatrix} -1.476\,923 \\ 0.369\,231 \end{pmatrix} + 0.034\,082\,8 \begin{pmatrix} -2 \\ -8 \end{pmatrix} = \begin{pmatrix} -1.545\,08 \\ 0.096\,56 \end{pmatrix}$$

3.6 共轭梯度法（例题）

求函数

$$f(\alpha)=f(\mathbf{x}_1+\alpha\mathbf{d}_1)=(0.738\,461\,5-1.545\,08\,\alpha)^2+4(-0.046\,153\,8+0.096\,56\,\alpha)^2$$

的极小点，得到步长和下一个迭代点为

$$\alpha_1=0.477\,941$$

$$\mathbf{x}_2=\mathbf{x}_1+\alpha_1\mathbf{d}_1=\begin{pmatrix}0.738\,461\,5\\-0.046\,153\,8\end{pmatrix}+0.477\,941\begin{pmatrix}-1.545\,08\\0.096\,56\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

这验证了前面的结论，即经过 $n=2$ 次迭代之后，算法收敛于极小点。

本节结束，谢谢

Thank you for listening

课程名称：最优化方法

习题

作业题： P3.1 P3.3 P3.4 P3.10

思考题：P3.5 P3.6