优化模型与软件工具

主讲教师：董庆兴

华中师范大学 信息管理学院
qxdong@mail.ccnu.edu.cn

2017 年 12 月 14 日

大纲

1. 拟牛顿法
2. DFP 算法
3. BFGS 算法
4. Broyden 族算法

拟牛顿法起源

- 牛顿法需要计算函数的二阶导数，现实问题中目标函数往往相当复杂，计算二阶导数的工作量非常大，或者不可行
- 在 $\mathbf{x}$ 的维度很高的时候，计算 Hessian 矩阵的逆矩阵本身就是非常麻烦的问题
- 所以如果我们能设法构造一个矩阵 $\mathbf{B}^{(k)}$，用它来逼近矩阵 $\mathbf{H}(\mathbf{x}^{(k)})$，就可以不计算 $\mathbf{H}(\mathbf{x}^{(k)})$, 其逆矩阵 $\bar{\mathbf{H}}^{(k)}$ 可以逼近 $\mathbf{H}(\mathbf{x}^{(k)})^{-1}$
- 但是近似矩阵 $\mathbf{B}^{(k)}$ 应该满足：
  - 在某种意义上有 $\mathbf{B}^{(k)} \approx \mathbf{H}(\mathbf{x}^{(k)})$, 使得相应的算法产生的方向（拟牛顿方向）是牛顿方向的近似，以保证算法的收敛速度
  - 对所有的 k，$\mathbf{B}^{(k)}$ 对称正定，从而使得算法产生的方向在 $\mathbf{x}^{(k)}$ 处的下降方向
  - 矩阵 $\mathbf{B}^{(k)}$ 容易计算

拟牛顿条件

- 若 $f(\mathbf{x})$ 有二阶连续偏导数，$\mathbf{x}^{(k+1)}$ 为其极小点的某一近似，作 $f(\mathbf{x}^{(k+1)})$ 的二阶泰勒展开

$$f(\mathbf{x}) \approx f(\mathbf{x}^{(k+1)})+\nabla f(\mathbf{x}^{(k+1)})^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^{(k+1)})+\frac{1}{2}(\mathbf{x}-\mathbf{x}^{(k+1)})^{\mathrm{T}}\mathbf{H}(\mathbf{x}^{(k+1)})(\mathbf{x}-\mathbf{x}^{(k+1)}) \tag{1}$$

- 其梯度满足

$$\nabla f(\mathbf{x}) \approx \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{H}(\mathbf{x}^{(k+1)})(\mathbf{x} - \mathbf{x}^{(k+1)}) \tag{2}$$

- 当 $\mathbf{x} = \mathbf{x}^{(k)}$ 时，有

$$\nabla f(\mathbf{x}^{(k)}) \approx \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{H}(\mathbf{x}^{(k+1)})(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}) \tag{3}$$

拟牛顿条件

- 因此，使得 $\mathbf{B}^{(k+1)}$ 的合理逼近是，当用 $\mathbf{B}^{(k+1)}$ 替换 $\mathbf{H}(\mathbf{x}^{(k+1)})$ 时，式（3）取等号，即
$$\mathbf{B}^{(k+1)}\mathbf{s}^{(k)} = \mathbf{y}^{(k)} \tag{4}$$
其中 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \mathbf{y}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$
- 式 (4) 即为拟牛顿方程或者割线方程，也可写成
$$\bar{\mathbf{H}}^{(k+1)}\mathbf{y}^{(k)} = \mathbf{s}^{(k)} \tag{5}$$
- 注意到 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = \lambda_k \mathbf{d}^{(k)}$，从而说明 $\mathbf{B}^{(k+1)}$ 与 $\mathbf{H}(\mathbf{x}^{(k+1)})$ 沿 $\mathbf{d}^{(k)}$ 方向近似，因此拟牛顿方程规定的拟牛顿方向是牛顿方向的一个近似

DFP 法迭代：逼近逆矩阵

- 寻找近似矩阵的方式有很多，已有的方法多对 $\bar{\mathbf{H}}^{(k)}$ 修正以产生 $\bar{\mathbf{H}}^{(k+1)}$

$$\bar{\mathbf{H}}^{(k+1)} = \bar{\mathbf{H}}^{(k)} + \Delta\bar{\mathbf{H}}^{(k)} \tag{6}$$

- 在式 (6) 中，可知 $\Delta\bar{\mathbf{H}}^{(k)}$ 为对称矩阵，可以按如下方式构造

$$\Delta\bar{\mathbf{H}}^{(k)} = \alpha_k \mathbf{u}\mathbf{u}^{\mathrm{T}} + \beta_k \mathbf{v}\mathbf{v}^{\mathrm{T}} \tag{7}$$

- 将式 (6) 和 (7) 代入到割线方程中有

$$\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)} + \alpha_k \mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{y}^{(k)} + \beta_k \mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{y}^{(k)} = \mathbf{s}^{(k)} \tag{8}$$

- 观察可得 $\mathbf{u}^{\mathrm{T}}\mathbf{y}^{(k)}$ 和 $\mathbf{v}^{\mathrm{T}}\mathbf{y}^{(k)}$ 均为常数，因此式 (8) 可写成

$$\alpha_k \left(\mathbf{u}^{\mathrm{T}}\mathbf{y}^{(k)}\right) \mathbf{u} + \beta_k \left(\mathbf{v}^{\mathrm{T}}\mathbf{y}^{(k)}\right) \mathbf{v} = \mathbf{s}^{(k)} - \bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)} \tag{9}$$

- 由式 (9) 我们很难得到唯一解，对于求解 $\mathbf{u}, \mathbf{v}$，一个很自然的选择是令

$$\mathbf{u} = \mathbf{s}^{(k)}, \quad \mathbf{v} = \bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)} \tag{10}$$

DFP 法迭代

- 将式 (10) 代入到式 (9) 有

$$\alpha_k\left(\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}\mathbf{y}^{(k)}\right)\mathbf{s}^{(k)}+\beta_k\left(\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}\right)\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}=\mathbf{s}^{(k)}-\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)} \tag{11}$$

解得

$$\alpha_k=\frac{1}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}},\quad \beta_k=-\frac{1}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}} \tag{12}$$

- 从而式 (6) 可以写成如下形式

$$\bar{\mathbf{H}}^{(k+1)}=\bar{\mathbf{H}}^{(k)}+\frac{\mathbf{s}^{(k)}\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}}-\frac{\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k)}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}} \tag{13}$$

- 我们知道式 (13) 只是可能的解的一种，这种方法被称作DFP法
- 可以证明，如果 $\bar{\mathbf{H}}^{(k)}$ 对称正定，则 $\bar{\mathbf{H}}^{(k+1)}$ 也是对称正定的，从而，在计算中我们只需要保证 $\bar{\mathbf{H}}^{(0)}$ 对称正定，则按照 DFP 法，构造出来的所有矩阵均为对称正定阵

DFP 法步骤

1. 给定初始点 $\mathbf{x}^{(0)} \in \mathbb{R}^n$，精度 $\epsilon > 0$，$k \leftarrow 0$
2. 若 $\|\nabla f(\mathbf{x}^{(k)})\| \leq \epsilon$，则算法终止，得到解 $\mathbf{x}^{(k)}$。否则，令

$$\bar{\mathbf{H}}^{(k)} = \begin{cases} \mathbf{I}, & k = 0 \\ \bar{\mathbf{H}}^{(k-1)} + \dfrac{\mathbf{s}^{(k-1)}\left(\mathbf{s}^{(k-1)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k-1)}\right)^{\mathrm{T}}\mathbf{s}^{(k-1)}} - \dfrac{\bar{\mathbf{H}}^{(k-1)}\mathbf{y}^{(k-1)}\left(\mathbf{y}^{(k-1)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k-1)}}{\left(\mathbf{y}^{(k-1)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(k-1)}\mathbf{y}^{(k-1)}}, & k > 0 \end{cases}$$

   求得 $\mathbf{d}^{(k)} = -\bar{\mathbf{H}}^{(k)}\nabla f(\mathbf{x}^{(k)})$
3. 由线性搜索确定步长 $\lambda_k = \arg\min f(\mathbf{x}^{(k)} - \lambda\bar{\mathbf{H}}^{(k)}\nabla f(\mathbf{x}^{(k)}))$
4. 令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k\mathbf{d}^{(k)}, k = k + 1$。转步骤 2

DFP 法例题

牛顿法例题

取初始点 $\mathbf{x}^{(0)} = (8, 9)^{\mathrm{T}}$，初始尺度矩阵 $\bar{\mathbf{H}}^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$。采用 DFP 法求解下面的最优化问题 $\min f(\mathbf{x}) = 4(x_1 - 5)^2 + (x_2 - 6)^2$

- 直接计算可得

$$\nabla f(\mathbf{x}^{(k)}) = \begin{pmatrix} 8(x_1^{(k)} - 5) \\ 2(x_2^{(k)} - 6) \end{pmatrix} \Rightarrow \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 24 \\ 6 \end{pmatrix}$$

-

$$\mathbf{d}^{(0)} = -\bar{\mathbf{H}}^{(0)} \nabla f(\mathbf{x}^{(0)}) = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 24 \\ 6 \end{pmatrix} = -\begin{pmatrix} 24 \\ 6 \end{pmatrix}$$

- $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda_0 \mathbf{d}^{(0)} = \begin{pmatrix} 8 - 24\lambda_0 \\ 9 - 6\lambda_0 \end{pmatrix}$

DFP 法例题

- 计算最佳步长，有

$$\begin{aligned}\phi(\lambda_0) &= f(\mathbf{x}^{(0)} + \lambda_k \mathbf{d}^{(0)}) \\ &= 4[(8-24\lambda_0)-5]^2 + [(9-6\lambda_0)-6]^2\end{aligned}$$

- 令 $\phi'(\lambda)=0$，可得 $\lambda_0 = \frac{17}{130}$
- 从而有 $\mathbf{x}^{(1)} = \begin{pmatrix} 8-24\lambda_0 \\ 9-6\lambda_0 \end{pmatrix} = \begin{pmatrix} 4.862 \\ 8.215 \end{pmatrix}$
- 此时有 $\mathbf{s}^{(0)} = \mathbf{x}^{(1)} - \mathbf{x}^{(0)} = \begin{pmatrix} -3.138 \\ -0.785 \end{pmatrix}$,

  $\mathbf{y}^{(0)} = \nabla f(\mathbf{x}^{(1)}) - \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} -25.108 \\ -1.569 \end{pmatrix}$

DFP 法例题

- 构造尺度矩阵

$$
\begin{aligned}
\bar{\mathbf{H}}^{(1)} &= \bar{\mathbf{H}}^{(0)} + \frac{\mathbf{s}^{(0)}\left(\mathbf{s}^{(0)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(0)}\right)^{\mathrm{T}}\mathbf{s}^{(0)}} - \frac{\bar{\mathbf{H}}^{(0)}\mathbf{y}^{(0)}\left(\mathbf{y}^{(0)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(0)}}{\left(\mathbf{y}^{(0)}\right)^{\mathrm{T}}\bar{\mathbf{H}}^{(0)}\mathbf{y}^{(0)}} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{\begin{pmatrix} -3.138 \\ -0.785 \end{pmatrix}(-3.138, -0.785)}{(-25.108, -1.569)\begin{pmatrix} -3.138 \\ -0.785 \end{pmatrix}} \\
&\quad - \frac{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -25.108 \\ -1.569 \end{pmatrix}(-25.108, -1.569)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{(-25.108, -1.569)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -25.108 \\ -1.569 \end{pmatrix}} \\
&= \begin{pmatrix} 0.1270 & -0.0315 \\ -0.0315 & 1.0038 \end{pmatrix}
\end{aligned}
$$

DFP 法例题

- 寻找新的试算点

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - \lambda_1 \bar{\mathbf{H}}^{(1)} \nabla f(\mathbf{x}^{(1)})$$
$$= \begin{pmatrix} 4.862 \\ 8.215 \end{pmatrix} - \lambda_1 \begin{pmatrix} 0.1270 & -0.0315 \\ -0.0315 & 1.0038 \end{pmatrix} \begin{pmatrix} -1.108 \\ 4.431 \end{pmatrix}$$

- 求解最佳步长可得 $\lambda_1 = 0.4942$
- 从而有 $\mathbf{x}^{(2)} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$，可知即为极小点

拟牛顿条件

- 因此，使得 $\mathbf{B}^{(k+1)}$ 的合理逼近是，当用 $\mathbf{B}^{(k+1)}$ 替换 $\mathbf{H}(\mathbf{x}^{(k+1)})$ 时，式（3）取等号，即

$$\mathbf{B}^{(k+1)}\mathbf{s}^{(k)} = \mathbf{y}^{(k)}$$

其中 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \mathbf{y}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$

- 式 (4) 即为拟牛顿方程或者割线方程，也可写成

$$\bar{\mathbf{H}}^{(k+1)}\mathbf{y}^{(k)} = \mathbf{s}^{(k)}$$

- 注意到 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = \lambda_k \mathbf{d}^{(k)}$，从而说明 $\mathbf{B}^{(k+1)}$ 与 $\mathbf{H}(\mathbf{x}^{(k+1)})$ 沿 $\mathbf{d}^{(k)}$ 方向近似，因此拟牛顿方程规定的拟牛顿方向是牛顿方向的一个近似

BFGS 算法：逼近 Hessian 矩阵

- 之前的 DFP 方法是产生 $\bar{\mathbf{H}}^{(k+1)}$，而由式（4）可知，我们同样可以通过产生 $\mathbf{B}^{(k+1)}$，来找对应的逆矩阵

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \Delta\mathbf{B}^{(k)} \tag{14}$$

- 在式 (14) 中,$\Delta\mathbf{B}^{(k)}$ 为对称矩阵，可以按如下方式构造

$$\Delta\mathbf{B}^{(k)} = \alpha_k \mathbf{u}\mathbf{u}^{\mathrm{T}} + \beta_k \mathbf{v}\mathbf{v}^{\mathrm{T}} \tag{15}$$

- 将式 (14) 和 (15) 代入到割线方程中有

$$\mathbf{B}^{(k)}\mathbf{s}^{(k)} + \alpha_k \mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{s}^{(k)} + \beta_k \mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{s}^{(k)} = \mathbf{y}^{(k)} \tag{16}$$

- 观察可得 $\mathbf{u}^{\mathrm{T}}\mathbf{s}^{(k)}$ 和 $\mathbf{v}^{\mathrm{T}}\mathbf{s}^{(k)}$ 均为常数，因此式 (16) 可写成

$$\alpha_k \left(\mathbf{u}^{\mathrm{T}}\mathbf{s}^{(k)}\right)\mathbf{u} + \beta_k \left(\mathbf{v}^{\mathrm{T}}\mathbf{s}^{(k)}\right)\mathbf{v} = \mathbf{y}^{(k)} - \mathbf{B}^{(k)}\mathbf{s}^{(k)} \tag{17}$$

BFGS 算法：逼近 Hessian 矩阵

- 由式 (17) 我们很难得到唯一解，对于求解 $\mathbf{u}, \mathbf{v}$，一个很自然的选择是令

$$\mathbf{u} = \mathbf{y}^{(k)}, \quad \mathbf{v} = \mathbf{B}^{(k)}\mathbf{s}^{(k)} \tag{18}$$

- 将式 (18) 代入到式 (17) 有

$$\alpha_k \left(\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}} \mathbf{s}^{(k)}\right) \mathbf{y}^{(k)} + \beta_k \left(\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}} \mathbf{B}^{(k)}\mathbf{s}^{(k)}\right) \mathbf{B}^{(k)}\mathbf{s}^{(k)} = \mathbf{y}^{(k)} - \mathbf{B}^{(k)}\mathbf{s}^{(k)} \tag{19}$$

解得

$$\alpha_k = \frac{1}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}} \mathbf{s}^{(k)}}, \quad \beta_k = -\frac{1}{\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}} \mathbf{B}^{(k)}\mathbf{s}^{(k)}} \tag{20}$$

- 从而式 (14) 可以写成

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \frac{\mathbf{y}^{(k)}\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}} \mathbf{s}^{(k)}} - \frac{\mathbf{B}^{(k)}\mathbf{s}^{(k)}\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}\mathbf{B}^{(k)}}{\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}} \mathbf{B}^{(k)}\mathbf{s}^{(k)}} \tag{21}$$

- 上述方法即 BFGS 法

BFGS 法步骤

1. 给定初始点 $\mathbf{x}^{(0)} \in \mathbb{R}^n$，精度 $\epsilon > 0$，$k \leftarrow 0$
2. 若 $\|\nabla f(\mathbf{x}^{(k)})\| \le \epsilon$，则算法终止，得到解 $\mathbf{x}^{(k)}$。否则，令

$$\mathbf{B}^{(k)} = \begin{cases} \mathbf{I}, & k = 0 \\ \mathbf{B}^{(k-1)} + \dfrac{\mathbf{y}^{(k-1)}\left(\mathbf{y}^{(k-1)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k-1)}\right)^{\mathrm{T}}\mathbf{s}^{(k-1)}} - \dfrac{\mathbf{B}^{(k-1)}\mathbf{s}^{(k-1)}\left(\mathbf{s}^{(k-1)}\right)^{\mathrm{T}}\mathbf{B}^{(k-1)}}{\left(\mathbf{s}^{(k-1)}\right)^{\mathrm{T}}\mathbf{B}^{(k-1)}\mathbf{s}^{(k-1)}}, & k \ge 1 \end{cases}$$

   求得 $\mathbf{d}^{(k)} = -\left(\mathbf{B}^{(k)}\right)^{-1}\nabla f(\mathbf{x}^{(k)})$
3. 由线性搜索确定步长 $\lambda_k = \arg\min f(\mathbf{x}^{(k)} - \lambda\left(\mathbf{B}^{(k)}\right)^{-1}\nabla f(\mathbf{x}^{(k)}))$
4. 令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k\mathbf{d}^{(k)}, k = k+1$。转步骤 2

BFGS 法例题

牛顿法例题

取初始点 $\mathbf{x}^{(0)}=(1,1)^{\mathrm{T}}$, 初始尺度矩阵 $\bar{\mathbf{H}}^{(0)}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$。采用 DFP 法求解下面的最优化问题 $\min f(\mathbf{x})=\frac{1}{2}x_1^2+x_2^2-x_1x_2-2x_1$

- 直接计算可得

$$\nabla f(\mathbf{x}^{(k)})=\begin{pmatrix}x_1-x_2-2\\ -x_1+2x_2\end{pmatrix}\Rightarrow\nabla f(\mathbf{x}^{(0)})=\begin{pmatrix}-2\\ 1\end{pmatrix}$$

- 求解线性方程组 $\mathbf{B}^{(0)}\mathbf{d}^{(0)}+\nabla f(\mathbf{x}^{(0)})=0$:

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}d_1^{(0)}\\ d_2^{(0)}\end{pmatrix}+\begin{pmatrix}-2\\ 1\end{pmatrix}=0$$

可得 $\mathbf{d}^{(0)}=\begin{pmatrix}2\\ -1\end{pmatrix}$

BFGS 法例题

- 计算最佳步长，有

$$\begin{aligned}\phi(\lambda) &= f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}) \\ &= \frac{1}{2}(x_1^{(k)} + \lambda d_1^{(k)})^2 + (x_2^{(k)} + \lambda d_2^{(k)})^2 \\ &\quad - (x_1^{(k)} + \lambda d_1^{(k)})(x_2^{(k)} + \lambda d_2^{(k)}) - 2(x_1^{(k)} + \lambda d_1^{(k)})\end{aligned}$$

- 解得

$$\lambda_k = \frac{x_1^{(k)} d_1^{(k)} - x_2^{(k)} d_2^{(k)} - 2d_1 - x_1^{(k)} d_2^{(k)} + 2x_2^{(k)} d_2^{(k)}}{(d_1^{(k)} - d_2^{(k)})^2 + (d_2^{(k)})^2}$$

- 代入即可求得 $\lambda_0 = \frac{1}{2}$， $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda_0 \mathbf{d}^{(0)} = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix}$

BFGS 法例题

- 计算可得 $\nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$
- 从而有
  $\mathbf{s}^{(0)} = \mathbf{x}^{(1)} - \mathbf{x}^{(0)} = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}, \mathbf{y}^{(0)} = \nabla f(\mathbf{x}^{(1)}) - \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} \frac{3}{2} \\ -2 \end{pmatrix}$
-

$$
\begin{aligned}
\mathbf{B}^{(1)} &= \mathbf{B}^{(0)} + \frac{\mathbf{y}^{(0)}\left(\mathbf{y}^{(0)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(0)}\right)^{\mathrm{T}}\mathbf{s}^{(0)}} - \frac{\mathbf{B}^{(0)}\mathbf{s}^{(0)}\left(\mathbf{s}^{(0)}\right)^{\mathrm{T}}\mathbf{B}^{(0)}}{\left(\mathbf{s}^{(0)}\right)^{\mathrm{T}}\mathbf{B}^{(0)}\mathbf{s}^{(0)}} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{9}{10} & -\frac{6}{5} \\ -\frac{6}{5} & \frac{8}{5} \end{pmatrix} - \begin{pmatrix} \frac{4}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{pmatrix} \\
&= \begin{pmatrix} \frac{11}{10} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{12}{5} \end{pmatrix}
\end{aligned}
$$

BFGS 法例题

- 求解线性方程组 $\mathbf{B}^{(1)}\mathbf{d}^{(1)} + \nabla f(\mathbf{x}^{(1)}) = 0$:

$$\begin{pmatrix} \frac{11}{10} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{12}{5} \end{pmatrix} \begin{pmatrix} d_1^{(1)} \\ d_2^{(1)} \end{pmatrix} + \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix} = 0$$

可得 $\mathbf{d}^{(1)} = \begin{pmatrix} 1 \\ \frac{3}{4} \end{pmatrix}$

- 类似计算可得 $\lambda_1 = 2, \mathbf{x}^{(2)} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \nabla f(\mathbf{x}^{(2)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
- 从而 $\mathbf{x}^{(2)}$ 即为最优解

BFGS 法求逆矩阵

- BFGS 法是逼近 Hessian 矩阵，在计算过程中仍需求逆，因此计算时有些不便
- 不过考虑到 Sherman-Morrison 公式

Sherman-Morrison 公式

设 $\mathbf{A} \in \mathbb{R}^n$ 为非奇异方阵，$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$，若满足 $1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u} \neq 0$，则矩阵 $\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}}$ 非奇异，且其逆矩阵为

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}} \tag{22}$$

- 而

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \frac{\mathbf{y}^{(k)}\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}} - \frac{\mathbf{B}^{(k)}\mathbf{s}^{(k)}\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}\mathbf{B}^{(k)}}{\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}\mathbf{B}^{(k)}\mathbf{s}^{(k)}}$$

BFGS 法求逆矩阵

- 连续两次应用 Sherman-Morrison 公式，即可求得 $\mathbf{B}^{(k+1)}$ 的逆矩阵

$$\bar{\mathbf{H}}^{(k+1)} = \left(\mathbf{I} - \frac{\mathbf{s}^{(k)}\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}}\right)\bar{\mathbf{H}}^{(k)}\left(\mathbf{I} - \frac{\mathbf{s}^{(k)}\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}}\right) + \frac{\mathbf{s}^{(k)}\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}}$$

$$= \bar{\mathbf{H}}^{(k)} + \frac{\left(\mathbf{s}^{(k)} - \bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}\right)\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}} + \mathbf{s}^{(k)}\left(\mathbf{s}^{(k)} - \bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}\right)^{\mathrm{T}}}{\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}}$$

$$- \frac{\left(\mathbf{s}^{(k)} - \bar{\mathbf{H}}^{(k)}\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{y}^{(k)}}{\left(\left(\mathbf{y}^{(k)}\right)^{\mathrm{T}}\mathbf{s}^{(k)}\right)^{2}}\mathbf{s}^{(k)}\left(\mathbf{s}^{(k)}\right)^{\mathrm{T}}$$

其中 $\bar{\mathbf{H}}^{(k+1)} = \left(\mathbf{B}^{(k+1)}\right)^{-1}$,$\bar{\mathbf{H}}^{(k)} = \left(\mathbf{B}^{(k)}\right)^{-1}$

BFGS 法求逆矩阵

- DFP 算法和 BFGS 算法之间存在着对偶关系
- Broyden 法通过

$$\mathbf{B}^{(k+1)} = \alpha_k \mathbf{B}_{\mathrm{BFGS}}^{(k+1)} + (1-\alpha_k)\mathbf{B}_{\mathrm{DFP}}^{(k+1)} \tag{23}$$

$$\mathbf{H}^{(k+1)} = \alpha_k \mathbf{H}_{\mathrm{BFGS}}^{(k+1)} + (1-\alpha_k)\mathbf{H}_{\mathrm{DFP}}^{(k+1)} \tag{24}$$

的方式构造矩阵