维基百科
自由的百科全书

黑塞矩阵

维基百科，自由的百科全书
（重定向自黑塞矩陣）

黑塞矩阵（德语：Hesse-Matrix；英语：Hessian matrix 或 Hessian），又译作**海森矩阵**、**海塞（赛）矩阵**或**海瑟矩阵**等，是一个由多变量实值函数的所有二阶偏导数组成的方阵，由德国数学家奥托·黑塞引入并以其命名。

定义

假设有一实值函数$f(x_1, x_2, \ldots, x_n)$，如果 f 的所有二阶偏导数都存在并在定义域内连续，那么函数f 的黑塞矩阵为

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

或使用下标记号表示为

$$\mathbf{H}_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

显然黑塞矩阵 $\mathbf{H}$ 是一个$n \times n$ 方阵。黑塞矩阵的行列式被称为黑塞式（英语：Hessian），而需注意的是英语环境下使用Hessian一词时可能指上述矩阵也可能指上述矩阵的行列式[1]。

性质

由高等数学知识可知，若一元函数$f(x)$ 在$x = x_0$ 点的某个邻域内具有任意阶导数，则函数$f(x)$ 在$x = x_0$ 点处的泰勒展开式为

$$f(x) = f(x_0) + f'(x_0)\Delta x + \frac{f''(x_0)}{2!}\Delta x^2 + \cdots$$

其中，$\Delta x = x - x_0$ 。

同理，二元函数$f(x_1, x_2)$ 在$x_0(x_{10}, x_{20})$ 点处的泰勒展开式为

$$f(x_1, x_2) = f(x_{10}, x_{20}) + f_{x_1}(x_{10}, x_{20})\Delta x_1 + f_{x_2}(x_{10}, x_{20})\Delta x_2 + \frac{1}{2}[f_{x_1 x_1}(x_{10}, x_{20})\Delta x_1^2 + 2 f_{x_1 x_2}(x$$

其中，$\Delta x_1 = x_1 - x_{10}$，$\Delta x_2 = x_2 - x_{20}$，$f_{x_1} = \frac{\partial f}{\partial x_1}$，$f_{x_2} = \frac{\partial f}{\partial x_2}$，$f_{x_1 x_1} = \frac{\partial^2 f}{\partial x_1^2}$，$f_{x_2 x_2} = \frac{\partial^2 f}{\partial x_2^2}$，$f_{x_1 x_2} = \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$。

将上述展开式写成矩阵形式，则有

$$f(x) = f(x_0) + \nabla f(x_0)^{\mathrm{T}} \Delta x + \frac{1}{2} \Delta x^{\mathrm{T}} G(x_0) \Delta x + \cdots$$

其中，$\Delta x = \begin{bmatrix} \Delta x_1 \\ \\ \Delta x_2 \end{bmatrix}$，$\Delta x^{\mathrm{T}} = [\,\Delta x_1 \quad \Delta x_2\,]$是$\Delta x$的转置，$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$是函数$f(x_1, x_2)$在$x_0(x_{10}, x_{20})$的梯度，矩阵

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{x_0}$$

即函数$f(x_1, x_2)$在$x_0(x_{10}, x_{20})$点处的2×2黑塞矩阵。它是由函数$f(x_1, x_2)$在$x_0(x_{10}, x_{20})$点处的所有二阶偏导数所组成的方阵。

由函数的二次连续性，有

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$$

所以，黑塞矩阵$G(x_0)$为对称矩阵。

将二元函数的泰勒展开式推广到多元函数，函数$f(x_1, x_2, \cdots, x_n)$在$x_0(x_1, x_2, \cdots, x_n)$点处的泰勒展开式为

$$f(x) = f(x_0) + \nabla f(x_0)^{\mathrm{T}} \Delta x + \frac{1}{2} \Delta x^{\mathrm{T}} G(x_0) \Delta x + \cdots$$

其中，

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}_{x_0}^{T}$$

为函数$f(x)$在$x_0(x_1, x_2, \cdots, x_n)$点的梯度，

$$G(x_0)=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n}\\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n}\\ \\ \vdots & \vdots & \ddots & \vdots\\ \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{bmatrix}_{x_0}$$

为函数$f(x)$在$x_0(x_1,x_2,\cdots,x_n)$点的$n\times n$黑塞矩阵。若函数有n次连续性，则函数的$n\times n$黑塞矩阵是对称矩阵。

说明：在优化设计领域中，黑塞矩阵常用G表示，且梯度有时用g表示。[2]

函数f的黑塞矩阵和雅可比矩阵有如下关系：

$$\mathbf{H}(f)=\mathbf{J}(\nabla f)^T$$

即函数f的黑塞矩阵等于其梯度的雅可比矩阵。

应用

函数的极值条件

对于一元函数$f(x)$，在给定区间内某$x=x_0$点处可导，并在$x=x_0$点处取得极值，其必要条件是

$$f'(x_0)=0$$

即函数$f(x)$的极值必定在驻点处取得，或者说可导函数$f(x)$的极值点必定是驻点；但反过来，函数的驻点不一定是极值点。检验驻点是否为极值点，可以采用二阶导数的正负号来判断。根据函数$f(x)$在$x=x_0$点处的泰勒展开式，考虑到上述极值必要条件，有

$$f(x)=f(x_0)+\frac{f''(x_0)}{2!}\Delta x^2+\cdots$$

若$f(x)$在$x=x_0$点处取得极小值，则要求在$x=x_0$某一邻域内一切点x都必须满足

$$f(x)-f(x_0)>0$$

即要求

$$\frac{f''(x_0)}{2!}\Delta x^2>0$$

亦即要求

$$f''(x_0)>0$$

$f(x)$在$x=x_0$点处取得极大值的讨论与之类似。于是有极值充分条件：

设一元函数$f(x)$在$x=x_0$点处具有二阶导数，且$f'(x_0)=0$，$f''(x_0)\neq 0$，则

1. 当$f''(x_0)>0$时，函数$f(x)$在$x=x_0$处取得极小值；
2. 当$f''(x_0)<0$时，函数$f(x)$在$x=x_0$处取得极大值。

而当$f''(x_0)=0$时，无法直接判断，还需要逐次检验其更高阶导数的正负号。由此有一个规律：若其开始不为零的导数阶数为偶数，则驻点是极值点；若为奇数，则为拐点，而不是极值点。

对于二元函数$f(x_1,x_2)$，在给定区域内某$x_0(x_{10},x_{20})$点处可导，并在$x_0(x_{10},x_{20})$点处取得极值，其必要条件是

$$f_{x_1}(x_0)=f_{x_2}(x_0)=0$$

即

$$\nabla f(x_0)=0$$

同样，这只是必要条件，要进一步判断$x_0(x_{10},x_{20})$是否为极值点需要找到取得极值的充分条件。根据函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$点处的泰勒展开式，考虑到上述极值必要条件，有

$$f(x_1,x_2)=f(x_{10},x_{20})+\frac{1}{2}[f_{x_1x_1}(x_0)\Delta x_1^2+2f_{x_1x_2}(x_0)\Delta x_1\Delta x_2+f_{x_2x_2}(x_0)\Delta x_2^2]+\cdots$$

设$A=f_{x_1x_1}(x_0)$，$B=f_{x_1x_2}(x_0)$，$C=f_{x_2x_2}(x_0)$，则

$$f(x_1,x_2)=f(x_{10},x_{20})+\frac{1}{2}[A\Delta x_1^2+2B\Delta x_1\Delta x_2+C\Delta x_2^2]+\cdots$$

或

$$f(x_1,x_2)=f(x_{10},x_{20})+\frac{1}{2A}[(A\Delta x_1+B\Delta x_2)^2+(AC-B^2)\Delta x_2^2]+\cdots$$

若$f(x_1,x_2)$在$x_0(x_{10},x_{20})$点处取得极小值，则要求在$x_0(x_{10},x_{20})$某一邻域内一切点x都必须满足

$$f(x_1,x_2)-f(x_{10},x_{20})>0$$

即要求

$$\frac{1}{2A}[(A\Delta x_1+B\Delta x_2)^2+(AC-B^2)\Delta x_2^2]>0$$

亦即要求$A>0$，$AC-B^2>0$

即

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x_0}>0$$

$$\left[\frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2}-(\frac{\partial^2 f}{\partial x_1\partial x_2})^2\right]_{x_0}>0$$

此条件反映了$f(x_1,x_2)$在$x_0(x_{10},x_{20})$点处的黑塞矩阵$G(x_0)$的各阶主子式都大于零，即对于

$$G(x_0)=\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{x_0}$$

要求

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x_0}>0$$

$$|G(x_0)|=\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix}_{x_0}>0$$

$f((x_1,x_2)$在$x_0(x_{10},x_{20})$点处取得极大值的讨论与之类似。于是有极值充分条件：

设二元函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$点的邻域内连续且具有一阶和二阶连续偏导数，又有$f_{x_1}(x_0)=f_{x_2}(x_0)=0$，同时令$A=f_{x_1x_1}(x_0)$，$B=f_{x_1x_2}(x_0)$，$C=f_{x_2x_2}(x_0)$，则

1. 当$A>0$，$AC-B^2>0$时，函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$处取得极小值；
2. 当$A<0$，$AC-B^2>0$时，函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$处取得极大值。

此外可以判断，当$AC-B^2<0$时，函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$点处没有极值，此点称为鞍点。而当$AC-B^2=0$时，无法直接判断，对此，补充一个规律：当$AC-B^2=0$时，如果有$A\equiv 0$，那么函数$f(x_1,x_2)$在$x_0(x_{10},x_{20})$有极值，且当$C>0$有极小值，当$C<0$有极大值。

由线性代数的知识可知，若矩阵$G(x_0)$满足

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x_0}>0$$

$$\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix}_{x_0}>0$$

则矩阵$G(x_0)$是正定矩阵，或者说矩阵$G(x_0)$正定。

若矩阵$G(x_0)$满足

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x_0} < 0$$

$$\left.\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix}\right|_{x_0} > 0$$

则矩阵$G(x_0)$是负定矩阵，或者说矩阵$G(x_0)$负定。[3]

于是，二元函数$f(x_1, x_2)$在$x_0(x_{10}, x_{20})$点处取得极值的条件表述为：二元函数$f(x_1, x_2)$在$x_0(x_{10}, x_{20})$点处的黑塞矩阵正定，则取得极小值；在$x_0(x_{10}, x_{20})$点处的黑塞矩阵负定，则取得极大值。

对于多元函数$f(x_1, x_2, \cdots, x_n)$，若在$x_0(x_1, x_2, \cdots, x_n)$点处取得极值，则极值存在的必要条件为

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}_{x_0}^T = 0$$

取得极小值的充分条件为

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{x_0}$$

正定，即要求$G(x_0)$的各阶主子式都大于零，即

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x_0} > 0$$

$$\left.\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix}\right|_{x_0} > 0$$

$$\vdots$$

$$|G(x_0)| > 0$$

取得极大值的充分条件为

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{x_0}$$

负定。[4][5][6]

拓展阅读

- 雅可比矩阵
- 梯度

参考文献

1. Binmore, Ken; Davies, Joan. Calculus Concepts and Methods. Cambridge University Press. 2007: 190. ISBN 9780521775410. OCLC 717598615.
2. 白清顺; 孙靖明; 梁迎春 (编). 机械优化设计（第6版）. 北京: 机械工业出版社. 2017.6（2018.11重印）: 35~36页. ISBN 978-7-111-56643-4.
3. 刘二根; 谢霖铨 (编). 线性代数. 江西高校出版社. 2015.7: 164~166页. ISBN 978-7-5493-3588-6.
4. 白清顺; 孙靖明; 梁迎春 (编). 机械优化设计（第6版）. 北京: 机械工业出版社. 2017.6（2018.11重印）: 37~39页. ISBN 978-7-111-56643-4.
5. 同济大学数学系 (编). 高等数学（第七版）上册. 高等教育出版社. 2014.7: 155页. ISBN 978-7-04-039663-8.
6. 同济大学数学系 (编). 高等数学（第七版）下册. 高等教育出版社. 2014.7: 113页. ISBN 978-7-04-039662-1.