Next: Sherman-Morrison formula **Up:** Appendices **Previous:** Review of Linar Algebra

Taylor series expansion

- **Single-variable functions:**

  A single-variable function $f(x)$ can be expanded around a given point x by the Taylor series:

  $$f(x+\delta x)=f(x)+f'(x)\delta x+\frac{1}{2!}f''(x)\delta x^2+\frac{1}{3!}f'''(x)\delta x^3+\cdots$$

  When δx is small, the higher order terms can be neglected so that the function can be approximated as a quadratic function

  $$f(x+\delta x)=f(x)+f'(x)\delta x+\frac{1}{2!}f''(x)\delta x^2+\varepsilon(\delta x^3)\approx f(x)+f'(x)\delta x+\frac{1}{2!}f''(x)\delta x^2$$

  or even a linear function

  $$f(x+\delta x)=f(x)+f'(x)\delta x+\varepsilon(\delta x^2)\approx f(x)+f'(x)\delta x$$

- **Multi-variable scalar-valued functions:**

  A multi-variable function $f(x_1,\cdots,x_N)=f(\mathbf{x})$ can also be expanded by the Taylor series:

  $$f(\mathbf{x}+\delta\mathbf{x})=f(\mathbf{x})+\sum_{j=1}^N\frac{\partial f(\mathbf{x})}{\partial x_j}\delta x_j+\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\frac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}\delta x_i\delta x_j+\cdots$$

  which can be expressed in vector form as:

  $$f(\mathbf{x}+\delta\mathbf{x})=f(\mathbf{x})+\mathbf{g}^T\,\delta\mathbf{x}+\frac{1}{2}\delta\mathbf{x}^T\mathbf{H}\,\delta\mathbf{x}+\cdots$$

where $\delta\mathbf{x}=[\delta x_1,\cdots,\delta x_N]^T$ is a vector and $\mathbf{g}$ and $\mathbf{H}$ are respectively the gradient vector and the *Hessian* matrix (first and second order derivatives in single variable case) of the function defined as:

$$\mathbf{g}=\mathbf{g}_f(\mathbf{x})=\bigtriangledown f(\mathbf{x})=\frac{d}{d\mathbf{x}}f(\mathbf{x})=\left[\begin{array}{c}\frac{\partial f\mathbf{x})}{\partial x_1}\\ \vdots\\ \frac{\partial f\mathbf{x})}{\partial x_N}\end{array}\right]_{N\times 1},$$

$$\mathbf{H}=\mathbf{H}_f(\mathbf{x})=\frac{d}{d\mathbf{x}}\mathbf{g}_f(\mathbf{x})=\left[\begin{array}{ccc}\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_N}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f(\mathbf{x})}{\partial x_N\partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_N^2}\end{array}\right]_{N\times N}$$

If the second derivatives of $f(\mathbf{x})$ are continuous, then $\partial^2 f(\mathbf{x})/\partial x_i\partial x_j=\partial^2 f(\mathbf{x})/\partial x_j\partial x_i$, i.e., $\mathbf{H}^T=\mathbf{H}$ is symmetric.

When $\delta\mathbf{x}$ is small (i.e., $||\delta\mathbf{x}||$ is small), then $f(\mathbf{x}+\delta\mathbf{x})$ can be approximated by a quadratic function with a third order error term

$$f(\mathbf{x}+\delta\mathbf{x})=f(\mathbf{x})+\mathbf{g}^T\,\delta\mathbf{x}+\frac{1}{2}\delta\mathbf{x}^T\mathbf{H}\,\delta\mathbf{x}+\varepsilon(||\delta\mathbf{x}||^3)$$

or even a linear function with a second order error term

$$f(\mathbf{x}+\delta\mathbf{x})=f(\mathbf{x})+\mathbf{g}^T\,\delta\mathbf{x}+\varepsilon(||\delta\mathbf{x}||^2)$$

- **Multi-variable vector-valued functions:**

  A set of M multi-variable functions $f_i(\mathbf{x})$ ($i=1,\cdots,M$) can be expressed as a vector function

$$\mathbf{f}(\mathbf{x})=[f_1(\mathbf{x}),\cdots,f_M(\mathbf{x})]^T$$

The Taylor expansion of the ith ($i=1,\cdots,M$) component is:

$$f_i(\mathbf{x}+\delta\mathbf{x})=f_i(\mathbf{x})+\mathbf{g}_i^T\,\delta\mathbf{x}+\frac{1}{2}\delta\mathbf{x}^T\mathbf{H}_i\,\delta\mathbf{x}+\varepsilon(\vert\vert\delta\mathbf{x}\vert\vert^3),$$

The first two terms of these M components can be written in vector form:

$$\mathbf{f}(\mathbf{x}+\delta\mathbf{x})\approx\left[\begin{array}{c}f_1(\mathbf{x})\\ \vdots\\ f_M(\mathbf{x})\end{array}\right]+\left[\begin{array}{ccc}\frac{\partial f_1}{\partial x_1}&\cdots&\frac{\partial f_1}{\partial x_N}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_M}{\partial x_1}&\cdots&\frac{\partial f_M}{\partial x_N}\end{array}\right]\left[\begin{array}{c}\delta x_1\\ \vdots\\ \delta x_N\end{array}\right]=\mathbf{f}(\mathbf{x})+\mathbf{J_f}(\mathbf{x})\,\delta\mathbf{x}$$

where $\mathbf{J_f}(\mathbf{x})$ is the *Jacobian matrix* defined over the vector function $\mathbf{f}(\mathbf{x})$:

$$\mathbf{J_f}(\mathbf{x})=\frac{d}{d\mathbf{x}}\mathbf{f}(\mathbf{x})=\left[\begin{array}{c}\frac{\partial}{\partial x_1}\\ \vdots\\ \frac{\partial}{\partial x_N}\end{array}\right][f_1\cdots f_N]=\left[\begin{array}{ccc}\frac{\partial f_1}{\partial x_1}&\cdots&\frac{\partial f_1}{\partial x_N}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_M}{\partial x_1}&\cdots&\frac{\partial f_M}{\partial x_N}\end{array}\right]_{M\times N}$$

However, the 2nd order term can no longer expressed in matrix form, as it requires tensor notation.

Note that the Hessian matrix of a function $f(\mathbf{x})$ can be obtained as the Jacobian matrix of the gradient vector of $f(\mathbf{x})$:

$$\mathbf{g}_f(\mathbf{x})=\left[\begin{array}{c}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_N}\end{array}\right],\qquad \mathbf{J_g}(\mathbf{x})=\left[\begin{array}{ccc}\frac{\partial^2 f}{\partial x_1^2}&\cdots&\frac{\partial^2 f}{\partial x_1\partial x_N}\\ \vdots&\ddots&\vdots\\ \frac{\partial^2 f}{\partial x_N\partial x_1}&\cdots&\frac{\partial^2 f}{\partial x_N^2}\end{array}\right]=\mathbf{H_f}(\mathbf{x})$$

Note that M and N may not be always the same. When $M>N$, there are more equations than variables, the problem is over-constrained.

Next: Sherman-Morrison formula **Up:** Appendices **Previous:** Review of Linar Algebra
Ruye Wang 2015-02-12