Department of Electrical and Computer Engineering

16:332:570 ROBUST COMPUTER VISION

Some Vector Calculus

Let $\mathbf{x} \in \mathcal{R}^n$, $\mathbf{x} = [x_1 \cdots x_n]^\top$, $g(\mathbf{x})$ a scalar valued function of $\mathbf{x}$, $\mathbf{z}(\mathbf{x}) \in \mathcal{R}^p$, $\mathbf{z} = [z_1 \cdots z_p]^\top$, a vector valued function of $\mathbf{x}$, and $\mathbf{A}(\mathbf{x}) \in \mathcal{R}^{p\times q}$ a matrix valued function of $\mathbf{x}$.

The derivative of the matrix $\mathbf{A}$ with respect to a scalar x is the $p \times q$ matrix

$$\frac{\partial \mathbf{A}}{\partial x} = \begin{bmatrix} \frac{\partial A_{11}}{\partial x} & \cdots & \frac{\partial A_{1q}}{\partial x} \\ \vdots & \frac{\partial A_{ik}}{\partial x} & \vdots \\ \frac{\partial A_{p1}}{\partial x} & \cdots & \frac{\partial A_{pq}}{\partial x} \end{bmatrix} .$$

The derivative of the matrix $\mathbf{A}$ with respect to another matrix $\mathbf{B} \in \mathcal{R}^{s\times t}$ is the $ps \times qt$ matrix

$$\frac{\partial \mathbf{A}}{\partial \mathbf{B}} = \begin{bmatrix} \frac{\partial \mathbf{A}}{\partial B_{11}} & \cdots & \frac{\partial \mathbf{A}}{\partial B_{1t}} \\ \vdots & \frac{\partial \mathbf{A}}{\partial B_{ik}} & \vdots \\ \frac{\partial \mathbf{A}}{\partial B_{s1}} & \cdots & \frac{\partial \mathbf{A}}{\partial B_{st}} \end{bmatrix} .$$

Taking $\mathbf{A} = \mathbf{z}(\mathbf{x})^\top$ and $\mathbf{B} = \mathbf{x}$ we obtain the definition of the *Jacobian* matrix of $\mathbf{z}(\mathbf{x})$ with respect to $\mathbf{x}$ as the $n \times p$ matrix

$$\mathbf{J}_{\mathbf{zx}}(\mathbf{x}) = \frac{\partial \mathbf{z}(\mathbf{x})^\top}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_p}{\partial x_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial z_1}{\partial x_n} & \cdots & \frac{\partial z_p}{\partial x_n} \end{bmatrix} = \left(\frac{\partial \mathbf{z}(\mathbf{x})}{\partial \mathbf{x}^\top}\right)^\top .$$

In some books the definition of the Jacobian is the transposed $p \times n$ matrix, i.e., $\frac{\partial \mathbf{z}(\mathbf{x})}{\partial \mathbf{x}^\top}$. Note that all the formulae in the sequel will be then changed. Which definiton of the Jacobian is used is not important as long as the formulae are used consistently.

Taking $\mathbf{A} = g(\mathbf{x})$ and $\mathbf{B} = \mathbf{x}$ we obtain the definition of the *gradient* of $g(\mathbf{x})$ with respect to $\mathbf{x}$ as the n-dimensional vector

$$\mathbf{J}_{g\mathbf{x}}(\mathbf{x}) = \nabla g = \left[\frac{\partial g}{\partial x_1} \cdots \frac{\partial g}{\partial x_n}\right]^\top .$$

If we regard $[\mathbf{x}, g(\mathbf{x})]$ as a surface in $\mathcal{R}^{n+1}$, the directional derivative in the direction $\mathbf{u} \in \mathcal{R}^n$ is defined as

$$\frac{\partial g}{\partial \mathbf{u}} = (\nabla g)^\top \mathbf{u}$$

and thus the gradient vector points toward the direction of maximum change on the surface $g(\mathbf{x})$ at location $\mathbf{x}$. (Note the sin of notation abuse we just committed.) Under a rotation of $\mathbf{x}$ in the continuous domain $\mathcal{R}^n$ the gradient magnitude is invariant, and the gradient orientation is equivariant.

As a consequence, to find ∇g it is enough to compute the directional derivatives along any two orthogonal directions.

The Jacobian of a composite function $\mathbf{f}(\mathbf{z}(\mathbf{x})) \in \mathcal{R}^q$ is obtained by the chain rule of vector differentiation as

$$\mathbf{J}_{\mathbf{fx}} = \mathbf{J}_{\mathbf{zx}}\mathbf{J}_{\mathbf{fz}}$$

and thus for the scalar ($q = 1$) case $\nabla_{\mathbf{x}} f = \mathbf{J}_{\mathbf{zx}} \nabla_{\mathbf{z}} f$.

The *Taylor series* of the scalar valued function $g(\mathbf{x})$ around $\mathbf{x}_0$ is

$$g(\mathbf{x}) = g(\mathbf{x}_0) + \mathbf{J}_{g\mathbf{x}}(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \mathbf{H}_{g\mathbf{x}}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + O[\|\mathbf{x} - \mathbf{x}_0\|^3]$$

where the $n \times n$ symmetric *Hessian* matrix is

$$\mathbf{H}_{g\mathbf{x}}(\mathbf{x}) = \frac{\partial^2 g(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 g}{\partial x_n \partial x_1} & \frac{\partial^2 g}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_n^2} \end{bmatrix}.$$

and the value in which the Jacobian (gradient) and Hessian are computed is implied by the notation. For a vector valued function the computation of the second order term becomes complicated.

Some Formulae

$\mathbf{x} \in \mathcal{R}^n$; $\mathbf{X}$ is an $n \times p$ (or $p \times p$) matrix; $\mathbf{a}(\mathbf{x}),\ \mathbf{b}(\mathbf{x}) \in \mathcal{R}^p$; the vector $\mathbf{c}$ and the matrices $\mathbf{A},\ \mathbf{B}$ have dimensions suitable for the expression in which they appear.

$\mathbf{z}(\mathbf{x}) = \mathbf{x} \quad \mathbf{J}_{\mathbf{zx}} = \mathbf{I}_n \qquad \mathbf{z}(\mathbf{x}) = \mathbf{c} \quad \mathbf{J}_{\mathbf{zx}} = \mathbf{O}$

$z(\mathbf{x}) = \mathbf{a}(\mathbf{x})^\top \mathbf{b}(\mathbf{x}) \quad \mathbf{J}_{z\mathbf{x}} = \mathbf{J}_{\mathbf{ax}}\mathbf{b}(\mathbf{x}) + \mathbf{J}_{\mathbf{bx}}\mathbf{a}(\mathbf{x})$ and thus $g(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} \quad \nabla_{\mathbf{x}} g = \mathbf{c}$

$\mathbf{z}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\mathbf{b}(\mathbf{x}) \quad \mathbf{J}_{\mathbf{zx}} = [\mathbf{J}_{\mathbf{a}_1\mathbf{x}}\mathbf{b}(\mathbf{x}) \ \cdots \ \mathbf{J}_{\mathbf{a}_n\mathbf{x}}\mathbf{b}(\mathbf{x})] + \mathbf{J}_{\mathbf{bx}}\mathbf{A}(\mathbf{x})^\top$

where $\mathbf{J}_{\mathbf{a}_i\mathbf{x}}$ is the Jacobian of the i-th *row* of the $n \times p$ matrix $\mathbf{A}(\mathbf{x})$ with respect to $\mathbf{x}$.

Thus if $\mathbf{z}(\mathbf{x}) = \mathbf{A}\mathbf{x} \quad \mathbf{J}_{\mathbf{zx}} = \mathbf{A}^\top$ and $\mathbf{z}(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \quad \mathbf{J}_{\mathbf{zx}} = \mathbf{A}$.

$\nabla \left(\mathbf{a}(\mathbf{x})^\top \mathbf{B}\mathbf{a}(\mathbf{x})\right) = 2\mathbf{J}_{\mathbf{ax}}\mathbf{B}\mathbf{a}(\mathbf{x})$ where $\mathbf{B}^\top = \mathbf{B}$ and thus $\nabla \left(\mathbf{x}^\top \mathbf{B}\mathbf{x}\right) = 2\mathbf{B}\mathbf{x}$

$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^\top \mathbf{X}\mathbf{b} = \mathbf{a}\mathbf{b}^\top \qquad \mathbf{X}$ is a p × p matrix

$\frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{A}^\top \mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{X}^\top \mathbf{A}) = \mathbf{A}$ and thus $\frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{A}\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{X}\mathbf{A}) = \mathbf{A}^\top$

$\frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{X}^\top \mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{A}^\top \mathbf{X}\mathbf{B}^\top$ and thus $\frac{\partial}{\partial \mathbf{X}}\mathrm{trace}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{X}$