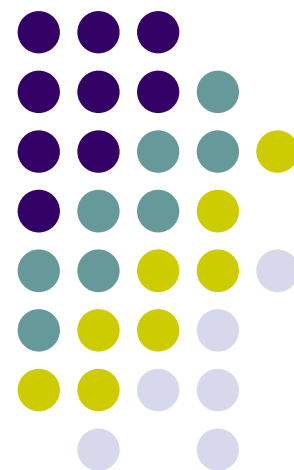


Matrix Calculus and Algebra

Yi Zhang

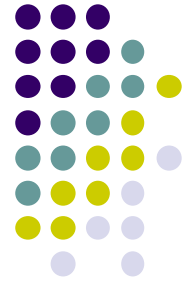




Outline

- Matrix calculus and algebra
 - Dimensions of derivatives
 - Basic calculations of matrix derivatives
 - Rules for product, chain, trace, determinant and norms

Matrix Derivatives: Dimensions



- The basic case: y scalar, x scalar
 - dy/dx : scalar
- Generalize to vectors/matrices

Matrix Derivatives: Dimensions [T. Minka]



	Scalar	Vector	Matrix
Scalar	$\frac{dy}{dx}$	$\frac{d\mathbf{y}}{dx} = \left[\frac{\partial y_i}{\partial x} \right]$	$\frac{d\mathbf{Y}}{dx} = \left[\frac{\partial y_{ij}}{\partial x} \right]$
Vector	$\frac{dy}{d\mathbf{x}} = \left[\frac{\partial y}{\partial x_j} \right]$	$\frac{d\mathbf{y}}{d\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j} \right]$	
Matrix	$\frac{dy}{d\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}} \right]$		



Basic Calculations

- Step 1: know the dimensions
- Step 2: element-wise calculations
- Step 3: put into the vector/matrix form



Basic Calculations

- Example:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}}$$

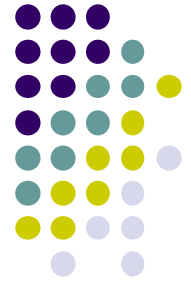
- Dimensions? Element-wise? Vector/matrix form?

Basic Calculations



- Example:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$



Basic Calculations

- Too simple? Try this ... [J. Rennie]

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{m \times k}, \text{ and } Y \in \mathbb{R}^{n \times m}$$

- We want: $\frac{\partial J}{\partial U}$



Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = \dots \dots$$

$$= 2 \sum_j (U_i V_j^T - Y_{ij}) V_{ja} + \lambda U_{ia},$$

$$= 2 (U_i V^T - Y_i) V_{\cdot a} + \lambda U_{ia},$$



Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = 2 (U_i V^T - Y_i) V_{.a} + \lambda U_{ia}$$



Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = 2 (U_i V^T - Y_i) V_{\cdot a} + \lambda U_{ia}$$

$$\frac{\partial J}{\partial U_i} =$$

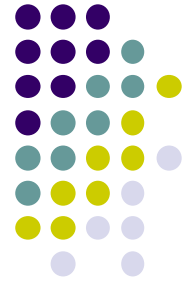


Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = 2 (U_i V^T - Y_i) V_{\cdot a} + \lambda U_{ia}$$

$$\frac{\partial J}{\partial U_i} = 2 (U_i V^T - Y_i) V + \lambda U_i$$



Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = 2 (U_i V^T - Y_i) V_{\cdot a} + \lambda U_{ia}$$

$$\frac{\partial J}{\partial U_i} = 2 (U_i V^T - Y_i) V + \lambda U_i$$

$$\frac{\partial J}{\partial U} =$$



Basic Calculations

$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$\frac{\partial J}{\partial U_{ia}} = 2 (U_i V^T - Y_i) V_{.a} + \lambda U_{ia}$$

$$\frac{\partial J}{\partial U_i} = 2 (U_i V^T - Y_i) V + \lambda U_i$$

$$\frac{\partial J}{\partial U} = 2 (UV^T - Y) V + \lambda U.$$

Some Examples



$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a} \mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b} \mathbf{a}^T$$

Gradient and Hessian



$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} =$$

Gradient and Hessian



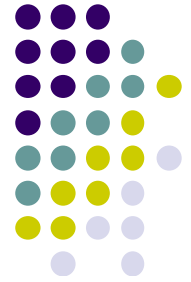
$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x} + \mathbf{b}$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} =$$

Gradient and Hessian



$$\begin{aligned} J(U, V) &= \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2} (\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2) \\ &= \sum_{i,j} \left(\sum_a U_{ia} V_{ja} - Y_{ij} \right)^2 + \frac{\lambda}{2} \left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2 \right) \end{aligned}$$

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x} + \mathbf{b}$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} = \mathbf{A} + \mathbf{A}^T$$



Some Basic Rules

- Take derivatives w.r.t. z (omitted)

$$\partial \mathbf{A} = 0 \quad (\mathbf{A} \text{ is a constant})$$

$$\partial(\mathbf{X} + \mathbf{Y}) = \partial \mathbf{X} + \partial \mathbf{Y}$$

$$\partial(\alpha \mathbf{X}) = \alpha \partial \mathbf{X}$$



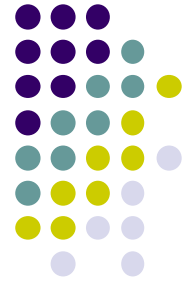
Product Rules

- Take derivatives w.r.t. z (omitted)

$$\partial(\dot{\mathbf{X}} \circ \mathbf{Y}) = (\partial \dot{\mathbf{X}}) \circ \mathbf{Y} + \dot{\mathbf{X}} \circ (\partial \mathbf{Y})$$

$$\partial(\mathbf{X}\mathbf{Y}) = (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y})$$

$$\partial(\dot{\mathbf{X}} \otimes \mathbf{Y}) = (\partial \dot{\mathbf{X}}) \otimes \mathbf{Y} + \dot{\mathbf{X}} \otimes (\partial \mathbf{Y})$$



Derivatives of Determinants

- Take derivatives w.r.t. z (omitted)

$$\partial(\det(\mathbf{X})) = \det(\mathbf{X})\text{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})$$

$$\partial(\ln(\det(\mathbf{X}))) = \text{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})$$



The Chain Rule

- Suppose $\mathbf{U} = f(\mathbf{X})$

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}}$$

- The chain rule:

$$\frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^M \sum_{l=1}^N \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}}$$



The Chain Rule

- Suppose $\mathbf{U} = f(\mathbf{X})$

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}}$$

- The chain rule:

$$\frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^M \sum_{l=1}^N \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}}$$

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \text{Tr} \left[\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \right)^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \right]$$

Derivatives of Traces [K.B. Petersen]



- Assume $F(\mathbf{X})$ is an element-wise differentiable function
 - $f()$ is the scalar derivative of $F()$

$$\frac{\partial \text{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

Derivatives of Traces [K.B. Petersen]



- Given: $\frac{\partial \text{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}^T$$

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}^T \mathbf{B}^T$$

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A}\mathbf{X}^T \mathbf{B}) = \mathbf{B}\mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^T \mathbf{A}) = \mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^2) = 2\mathbf{X}^T$$

Derivatives of Frobenius Norm



- Frobenius norm

$$||\mathbf{A}||_{\text{F}} = \sqrt{\sum_{ij} |A_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^H)}$$

- Derivatives

$$\frac{\partial}{\partial \mathbf{X}} ||\mathbf{X}||_{\text{F}}^2 = 2\mathbf{X} = \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}\mathbf{X}^H)$$



References

- T. Minka. Old and New Matrix Algebra Useful for Statistics
- K. B. Petersen. The Matrix Cookbook
- J. D. M. Rennie. A Simple Exercise on Matrix Derivatives.