A note on differentiating matrices

Kowal, Pawel

21 December 2006

A note on differentiating matrices

Paweł Kowal

December 21, 2006

Abstract

This paper presents a set of rules for matrix differentiation with respect to a vector of parameters, using the flattered representation of derivatives, i.e. in form of a matrix. We also introduce a new set of Kronecker tensor products of matrices. Finally we consider a problem of differentiating matrix determinant, trace and inverse.

JEL classification: C00

Keywords: matrix differentiation, generalized Kronecker products

1 Introduction

Derivatives of matrices with respect to a vector of parameters can be expressed as a concatenation of derivatives with respect to a scalar parameters. However such a representation of derivatives is very inconvenient in some applications, e.g. if higher order derivatives are considered, and or even are not applicable if matrix functions (like determinant or inverse) are present. For example finding an explicit derivative of $\det(\partial X/\partial\theta)$ would be a quite complicated task. Such a problem arise naturally in many applications, e.g. in maximum likelihood approach for estimating model parameters.

The same problems emerges in case of a tensor representation of derivatives. Additionally, in this case additional effort is required to find the flattered representation of resulting tensors, which is required, since running numerical computations efficiently is possible only in case of two dimensional data structures.

In this paper we derive formulas for differentiating matrices with respect to a vector of parameters, when one requires the flattered form of resulting derivatives, i.e. representation of derivatives in form of matrices. To do this we introduce a new set of the Kronecker matrix products as well as the generalized matrix transposition. Then, first order and higher order derivatives of functions being compositions of primitive function using elementary matrix operations like summation, multiplication, transposition and the Kronecker product, can be expressed in a closed form based on primitive matrix functions and their derivatives, using these elementary operations, the generalized Kronecker products and the generalized transpositions.

We consider also more general matrix functions containing matrix functions (inverse, trace and determinant). Defining the generalized trace function we are able to express derivatives of such functions in closed form.

2 Matrix differentiation rules

Let as consider smooth functions $\Omega \ni \theta \mapsto X(\theta) \in \mathbb{R}^{m\times n}$, $\Omega \ni \theta \mapsto Y(\theta) \in \mathbb{R}^{p\times q}$, where $\Omega \subset \mathbb{R}^k$ is an open set. Functions X, Y associate a $m \times n$ and $p \times q$ matrix for a given vector of parameters, $\theta = \mathrm{col}(\theta_1, \theta_2, \ldots, \theta_k)$. Let the differential of the function X with respect to θ is defined as

$$\frac{\partial X}{\partial \theta} = \left[\begin{array}{cccc} \frac{\partial X}{\partial \theta_1} & \frac{\partial X}{\partial \theta_2} & \cdots & \frac{\partial X}{\partial \theta_k} \end{array} \right]$$

for $\partial X/\partial \theta_i \in \mathbb{R}^{m\times n}$, $i = 1, 2, \ldots, k$.

Proposition 2.1. *The following equations hold*

1. $\frac{\partial}{\partial\theta}(\alpha X) = \alpha \frac{\partial X}{\partial \theta}$
2. $\frac{\partial}{\partial\theta}(X + Y) = \frac{\partial X}{\partial \theta} + \frac{\partial Y}{\partial \theta}$
3. $\frac{\partial}{\partial\theta}(X \times Y) = \frac{\partial X}{\partial \theta} \times (I_k \otimes Y) + X \times \frac{\partial Y}{\partial \theta}$

where $\alpha \in \mathbb{R}$ and I_k is a $k \times k$ dimensional identity matrix, assuming that differentials exist and matrix dimensions coincide.

Proof. The first two cases are obvious. We have

$$\begin{aligned} \frac{\partial}{\partial\theta}(X \times Y) &= \left[\begin{array}{ccc} \frac{\partial X}{\partial \theta_1} \times Y + X \times \frac{\partial Y}{\partial \theta_1} & \ldots & \frac{\partial X}{\partial \theta_k} \times Y + X \times \frac{\partial Y}{\partial \theta_k} \end{array} \right] \\ &= \left[\begin{array}{ccc} \frac{\partial X}{\partial \theta_1} & \ldots & \frac{\partial X}{\partial \theta_k} \end{array} \right] \times \begin{bmatrix} Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Y \end{bmatrix} + X \times \left[\begin{array}{ccc} \frac{\partial Y}{\partial \theta_1} & \ldots & \frac{\partial Y}{\partial \theta_k} \end{array} \right] \\ &= \frac{\partial X}{\partial \theta} \times (I_k \otimes Y) + X \times \frac{\partial Y}{\partial \theta} \end{aligned}$$

□

Differentiating matrix transposition is a little bit more complicated. Let us define a generalized matrix transposition

Definition 2.2. *Let $X = [X_1, X_2, \dots X_n]$, where $X_i \in \mathbb{R}^{p\times q}$, $i = 1, 2, \dots, n$ is a $p \times q$ matrix is a partition of $p \times nq$ dimensional matrix X. Then*

$$\mathrm{T}_n(X) \doteq \left[\ X'_1, X'_2, \dots, X'_n \ \right]$$

Proposition 2.3. *The following equations hold*

1. $\frac{\partial}{\partial\theta}(X') = \mathrm{T}_k(\frac{\partial X}{\partial\theta})$
2. $\frac{\partial}{\partial\theta}(\mathrm{T}_n(X)) = \mathrm{T}_{k\times n}(\frac{\partial X}{\partial\theta})$

Proof. The first condition is a special case of the second condition for $n = 1$. We have

$$\begin{aligned}\frac{\partial}{\partial\theta}(\mathrm{T}_{(n)}(X)) &= \left[\ \mathrm{T}_{(n)}(\tfrac{\partial X}{\partial\theta_1}) \quad \dots \quad \mathrm{T}_{(n)}(\tfrac{\partial X}{\partial\theta_k}) \ \right] \\ &= \left[\ \tfrac{\partial X'_1}{\partial\theta_1}, \dots, \tfrac{\partial X'_n}{\partial\theta_1} \quad \dots \quad \tfrac{\partial X'_1}{\partial\theta_k}, \dots, \tfrac{\partial X'_n}{\partial\theta_k} \ \right] = \mathrm{T}_{(k\times n)}\left(\frac{\partial X}{\partial\theta}\right)\end{aligned}$$

since

$$\frac{\partial X}{\partial\theta} = \left[\ \tfrac{\partial X_1}{\partial\theta_1}, \dots, \tfrac{\partial X_n}{\partial\theta_1} \quad \dots \quad \tfrac{\partial X_1}{\partial\theta_k}, \dots, \tfrac{\partial X_n}{\partial\theta_k} \ \right]$$

$\square$

Let us now turn to differentiating tensor products of matrices. Let for any matrices X, Y, where $X \in R^{\imath\times q}$ is a matrix with elements $x_{ij} \in \mathbb{R}$ for $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$. The Kronecker product, $X \otimes Y$ is defined as

$$X \otimes Y \doteq \begin{bmatrix} x_{11}Y & \cdots & x_{1q}Y \\ \vdots & \ddots & \vdots \\ x_{p1}Y & \cdots & x_{pq}Y \end{bmatrix}$$

Similarly as in case of differentiating matrix transposition we need to introduce the generalized Kronecker product

Definition 2.4. *Let $X = [X_1, X_2, \dots X_m]$, where $X_i \in \mathbb{R}^{p\times q}$, $i = 1, 2, \dots, m$ is a $p \times q$ matrix is a partition of $p \times mq$ dimensional matrix X. Let $Y = [Y_1, Y_2, \dots Y_n]$, where $Y_i \in \mathbb{R}^{r\times s}$, $i = 1, 2, \dots, n$ is a $r \times s$ matrix is a partition of $r \times ns$ dimensional matrix Y. Then*

$$\begin{aligned} X \otimes_n^1 Y &\doteq [X \otimes Y_1, \dots, X \otimes Y_n] \\ X \otimes_n^m Y &\doteq [X_1 \otimes_n^1 Y, \dots, X_m \otimes_n^1 Y] \\ X \otimes_{n_1,n_2,\dots,n_s}^{1,m_2,\dots,m_s} Y &\doteq [X \otimes_{n_2,\dots,n_s}^{m_2,\dots,m_s} Y_1, \dots, X \otimes_{n_2,\dots,n_s}^{m_2,\dots,m_s} Y_{n_1}] \\ X \otimes_{n_1,n_2,\dots,n_s}^{m1,m_2,\dots,m_s} Y &\doteq [X_1 \otimes_{n_1,n_2,\dots,n_s}^{1,m_2,\dots,m_s} Y, \dots, X_{m_1} \otimes_{n_1,n_2,\dots,n_s}^{1,m_2,\dots,m_s} Y] \end{aligned}$$

assuming that appropriate matrix partitions exist.

Proposition 2.5. *The following equations hold*

1. $\frac{\partial}{\partial\theta}(X \otimes Y) = \frac{\partial X}{\partial\theta} \otimes Y + X \otimes_k^1 \frac{\partial Y}{\partial\theta}$

2. $\frac{\partial}{\partial\theta}(X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y) = \frac{\partial X}{\partial\theta} \otimes_{1,n_1,\ldots,n_s}^{k,m_1,\ldots,m_s} Y + X \otimes_{k,n_1,\ldots,n_s}^{1,m_1,\ldots,m_s} \frac{\partial Y}{\partial\theta}$

Proof. We have

$$\begin{aligned}
\frac{\partial}{\partial\theta}(X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y) &= \left[\begin{array}{ccc} \frac{\partial}{\partial\theta_1}(X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y) & \cdots & \frac{\partial}{\partial\theta_k}(X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y) \end{array}\right] \\
&= \left[\begin{array}{ccc} \frac{\partial X}{\partial\theta_1} \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y & \cdots & \frac{\partial X}{\partial\theta_k} \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y \end{array}\right] \\
&+ \left[\begin{array}{ccc} X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} \frac{\partial Y}{\partial\theta_1} & \cdots & X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} \frac{\partial Y}{\partial\theta_k} \end{array}\right] \\
&= \frac{\partial X}{\partial\theta} \otimes_{1,n_1,\ldots,n_s}^{k,m_1,\ldots,m_s} Y + X \otimes_{k,n_1,\ldots,n_s}^{1,m_1,\ldots,m_s} \frac{\partial Y}{\partial\theta}
\end{aligned}$$

Since $X \otimes Y = X \otimes_1^1 Y$, in case of the standard Kronecker product we obtain

$$\frac{\partial}{\partial\theta}(X \otimes Y) = \frac{\partial X}{\partial\theta} \otimes_1^k Y + X \otimes_k^1 \frac{\partial Y}{\partial\theta} = \frac{\partial X}{\partial\theta} \otimes Y + X \otimes_k^1 \frac{\partial Y}{\partial\theta}$$

□

In proposition 2.1 we have omitted the case of multiplication of a matrix by a scalar function, using proposition 2.5 we obtain

Proposition 2.6. *Let α is a scalar function of θ and X is a matrix valued function of θ, $X(\theta) \in \mathbb{R}^{p\times q}$. Then*

$$\frac{\partial}{\partial\theta}(\alpha X) = \alpha \times \frac{\partial X}{\partial\theta} + \frac{\partial\alpha}{\partial\theta} \otimes X$$

Proof. Expression αX can be represented as $\alpha X = (\alpha \otimes I_p) \times X$, where I_p is a $p \times p$ dimensional identity matrix. Hence

$$\begin{aligned}
\frac{\partial}{\partial\theta}(\alpha X) &= \frac{\partial}{\partial\theta}((\alpha \otimes I_p) \times X) = \frac{\partial(\alpha \otimes I_p)}{\partial\theta} \times (I_k \otimes X) + (\alpha \otimes I_p) \times \frac{\partial X}{\partial\theta} \\
&= (\frac{\partial\alpha}{\partial\theta} \otimes I_p) \times (I_k \otimes X) + \alpha \times \frac{\partial X}{\partial\theta} = \frac{\partial\alpha}{\partial\theta} \otimes X + \alpha \times \frac{\partial X}{\partial\theta}
\end{aligned}$$

□

Let $\mathcal{S}$ is a set of smooth matrix valued functions $\Omega \ni \theta \mapsto X(\theta) \in \mathbb{R}^{p\times q}$, where $\Omega \subset \mathbb{R}^k$ is an open set, for any integers $p, q \geq 1$ not necessary the same for all functions in $\mathcal{S}$. Let $\operatorname{dif} \mathcal{S} \doteq \{\partial X/\partial\theta : X \in \mathcal{S}\}$. The set $\mathcal{S}$ may contain scalars and matrices, which are interpreted as constant functions.

Let $\mathrm{ext}(\mathcal{S})$ is a set of functions obtained by applying elementary matrix operations on the set $\mathcal{S}$, i.e. $\mathrm{ext}(\mathcal{S})$ is a smallest set such that if $X, Y \in \mathrm{ext}(\mathcal{S})$, then matrix valued functions $X+Y$, $X \times Y$, $\mathrm{T}_n(X)$, $X \otimes_{n1,\ldots,n_s}^{m1,\ldots,m_s} Y$ if exist belong to $\mathrm{ext}(\mathcal{S})$, where $n, n_1, \ldots, n_s, m_1, \ldots, m_s$ are any positive integers.

Theorem 2.7. $\mathrm{dif}(\mathrm{ext}(\mathcal{S})) = \mathrm{ext}(\mathcal{S} \cup \mathrm{dif}(\mathcal{S}))$.

Proof. By induction using propositions 2.1, 2.3, 2.5, 2.6. □

The theorem 2.7 states, that derivatives of matrix valued functions obtained by applying elementary operations like summation, matrix multiplication, generalized transposition and generalized Kronecker tensor product can be expressed as a combination of these functions and their derivatives using these elementary operations. Applying the theorem 2.7 to a set $\mathcal{T} = \mathrm{dif}(\mathrm{ext}(\mathcal{S}))$ we can see that also higher order derivatives can be expresses, using these elementary operations, as combinations of elementary functions $\mathcal{S}$ and their higher order derivatives.

3 Derivatives of matrix determinant, trace and inverse

Let us consider derivatives of matrix inverse, determinant and trace. We need to introduce the generalized trace defined analogously as the generalized transposition.

Definition 3.1. *Let $X = [X_1, X_2, \ldots X_n]$, where $X_i \in \mathbb{R}^{p \times p}$, $i = 1, 2, \ldots, n$ is a $p \times p$ matrix, is a partition of $p \times np$ dimensional matrix X. Then*

$$\mathrm{tr}_n(X) \doteq \left[\ \mathrm{tr}\, X_1, \mathrm{tr}\, X_2, \ldots, \mathrm{tr}\, X_n\ \right]$$

Proposition 3.2. *The following equations hold*

1. $\frac{\partial \det(X)}{\partial \theta} = \det(X) \times \mathrm{tr}_k(X^{-1} \times \frac{\partial X}{\partial \theta})$
2. $\frac{\partial \,\mathrm{tr}_n(X)}{\partial \theta} = \mathrm{tr}_{k \times n}(\frac{\partial X}{\partial \theta})$
3. $\frac{\partial X^{-1}}{\partial \theta} = -X^{-1} \times \frac{\partial X}{\partial \theta} \times (I_k \otimes X^{-1})$

Proof. We have

$$\begin{aligned}
\frac{\partial \det(X)}{\partial \theta} &= \left[\begin{array}{ccc} \frac{\partial \det(X)}{\partial \theta_1} & \ldots & \frac{\partial \det(X)}{\partial \theta_k} \end{array} \right] \\
&= \left[\begin{array}{ccc} \det(X)\operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_1}) & \ldots & \det(X) \times \operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_k}) \end{array} \right] \\
&= \det(X) \times \operatorname{tr}_k(X^{-1} \times \frac{\partial X}{\partial \theta})
\end{aligned}$$

$$\frac{\partial \operatorname{tr}_n(X)}{\partial \theta} = \left[\begin{array}{ccc} \frac{\partial \operatorname{tr}_n(X)}{\partial \theta_1} & \ldots & \frac{\partial \operatorname{tr}_n(X)}{\partial \theta_k} \end{array} \right] = \left[\begin{array}{ccc} \operatorname{tr}_n(\frac{\partial X}{\partial \theta_1}) & \ldots & \operatorname{tr}(\frac{\partial X}{\partial \theta_k}) \end{array} \right] = \operatorname{tr}_{k\times n}(\frac{\partial X}{\partial \theta})$$

Similarly

$$\begin{aligned}
\frac{\partial X^{-1}}{\partial \theta} &= \left[\begin{array}{ccc} \frac{\partial X^{-1}}{\partial \theta_1} & \ldots & \frac{\partial X^{-1}}{\partial \theta_k} \end{array} \right] = -\left[\begin{array}{ccc} X^{-1}\frac{\partial X}{\partial \theta_1}X^{-1} & \ldots & X^{-1}\frac{\partial X}{\partial \theta_k}X^{-1} \end{array} \right] \\
&= -X^{-1} \times \left[\begin{array}{ccc} \frac{\partial X}{\partial \theta_1} & \ldots & \frac{\partial X}{\partial \theta_k} \end{array} \right] \times (I_k \otimes X^{-1}) = -X^{-1}\frac{\partial X}{\partial \theta}(I_k \otimes X^{-1})
\end{aligned}$$

since in case of scalar parameter $\theta \in \mathbb{R}$, $\partial \det(X)/\partial\theta = \det(X)\operatorname{tr}(X^{-1}\partial X/\partial\theta)$, $\partial \operatorname{tr}(X)/\partial\theta = \operatorname{tr}(\partial X/\partial\theta)$, and $\partial X^{-1}/\partial\theta = -X^{-1}(\partial X/\partial\theta)X^{-1}$ (see for example Petersen, Petersen, (2006)). $\square$

Let a set S and operation dif are defined as in the previous section. Let $\operatorname{ext}_2(\mathcal{S})$ is a set of functions obtained by applying elementary matrix operations and matrix determinant, trace and inverse on the set $\mathcal{S}$, i.e. $\operatorname{ext}(\mathcal{S})$ is a smallest set such that if $X, Y \in \operatorname{ext}_2(\mathcal{S})$, then matrix valued functions $X+Y$, $X \times Y$, $\mathrm{T}_n(X)$, $X \otimes^{m1,\ldots,m_s}_{n1,\ldots,n_s} Y$, $\det(X)$, $\operatorname{tr}_n(X)$, X^{-1} if exist belong to $\operatorname{ext}_2(\mathcal{S})$, where $n, n_1, \ldots, n_s, m_1, \ldots, m_s$ are any positive integers.

Theorem 3.3. $\operatorname{dif}(\operatorname{ext}_2(\mathcal{S})) = \operatorname{ext}_2(\mathcal{S} \cup \operatorname{dif}(\mathcal{S}))$.

Proof. By induction using propositions 2.1, 2.3, 2.5, 2.6, 3.2. $\square$

4 Concluding remarks

Derived formulas requires matrix tensor products, which are absent, when representing derivatives as the concatenation of derivatives with respect to a scalar parameters. Hence, this approach may decrease numerical efficiency. This problem however can be resolved using appropriate data structures.

References

[1] T.P. Minka. Old and new matrix algebra useful for statistics. notes, December 2000.

[2] K.P. Petersen and M. S. Petersen. The matrix cookbook. notes, 2006.