Machine Learning

Section 9: Matrix Differential Calculus

第9节：矩阵微分计算

Stefan Harmeling

8. November 2021

Linear regression summary

Model assumption

$$p(w) = \mathcal{N}(w|0, \tau^2 I) \qquad \text{prior}$$

$$p(y|X, w) = \mathcal{N}(y|Xw, \sigma^2 I) \qquad \text{likelihood}$$

1. Ordinary least squares (MLE):

$$w_{\text{MLE}} = (X^T X)^{-1} X^T y \qquad \text{maximizer of likelihood, no prior}$$

2. Ridge regression (MAP):

$$p(w|X, y) = \mathcal{N}(w|w_n, V_n) \qquad \text{posterior后方}$$

$$w_{\text{ridge}} = (\lambda I + X^T X)^{-1} X^T y \qquad \text{maximizer of posterior}$$

3. Bayesian linear regression:

$$p(w|X, y) = \mathcal{N}(w|w_n, V_n) \qquad \text{posterior}$$

$$w_n = (\lambda I + X^T X)^{-1} X^T y \qquad \text{posterior mean}$$

$$V_n = \sigma^2 (\lambda I + X^T X)^{-1} \qquad \text{posterior covariance}$$

Today:

How to calculate derivatives of scalar-/vector-/matrix-valued functions of scalar-/vector-/matrix-valued variables!

如何计算

标量-/向量-/矩阵值变量的标量-/向量-/矩阵值函数!

Matrix differential calculus

A tool to find complicated derivatives involving vectors and matrices.

Sources

- Download `bn142.pdf` from public sciebo folder of this lecture
- Magnus/Neudecker: Matrix differential calculus, 2007, PDF was available at `http://www.janmagnus.nl/misc/mdc2007-3rdedition.pdf`, the link doesn't work anymore, it looks like a new version is planned so the old PDF is removed. Instead much of the stuff can be found in a paper: `http://www.janmagnus.nl/papers/JRM012.pdf`
- `http://www.janmagnus.nl/misc/mdc-ch18.pdf`
- Lütkepohl: Handbook of matrices, 1996

What is a differential?

Definition 9.1 (source: https://en.wikipedia.org/wiki/Differential_of_a_function)

A differential is an infinitesimal change in some varying quantity, e.g.

- *for variable x the differential is denoted as* $\mathrm{d}x$
- *for variable $y = f(x)$ being the image of x under function f, the differential is*

$$\mathrm{d}y = f'(x)\,\mathrm{d}x$$

where $f'(x)$ is the derivative of f.

Note

- writing the derivative $f'(x) = \mathrm{d}y/\mathrm{d}x$ in Leibniz notation:

$$\mathrm{d}y = \frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{d}x$$

  (looks reasonable!)
- an integral is a summation over infinitesimal rectangles

$$\int f(x)\,\mathrm{d}x = \sum_x f(x)\,\mathrm{d}x$$

Differential calculus微分计算

Notation for derivative (not for the differential)导数的符号

$$\mathrm{D}f(x) = f'(x)$$

Identification formula: (relates derivative and differential)识别公式

$$\mathrm{d}f = \mathrm{D}f(x)\mathrm{d}x$$

How to find a derivative with the differential calculus:

如何用微积分找导数。

1. write the letter d in front of the expression.在表达式前面写上字母d
2. identity the constants and variables.确定常数和变量的身份
3. transform the expression转换表达式
4. read off the derivative using the identification table使用识别表读出导数

Notation!

For the rest of the slides:

- Scalars: small greek letters: $\phi, \psi, \alpha, \ldots$疤痕
- Vectors: small latin letters: $u, v, x, f, a, \ldots$媒介物
- Matrices: capital latin letters: $U, V, F, A, \ldots$

Rules for the calculus (1)

Only with scalars:

$$
\begin{aligned}
\mathrm{d}\alpha &= 0 \\
\mathrm{d}(\alpha\phi) &= \alpha\,\mathrm{d}\phi \\
\mathrm{d}(\phi+\psi) &= \mathrm{d}\phi + \mathrm{d}\psi \\
\mathrm{d}(\phi\,\psi) &= (\mathrm{d}\phi)\psi + \phi\,\mathrm{d}\psi \\
\mathrm{d}(\phi/\psi) &= ((\mathrm{d}\phi)\psi - \phi\,\mathrm{d}\psi)\,/\,\psi^2
\end{aligned}
$$

where α is constant.

Example:
Find the derivative of $\phi(\xi) = \xi^2$.

$$\mathrm{d}\phi = \mathrm{d}\xi^2 = 2\xi\mathrm{d}\xi$$

Thus using $\mathrm{d}\phi = \mathrm{D}\phi(\xi)\mathrm{d}\xi$ we read off the derivative:

$$\mathrm{D}\phi(\xi) = 2\xi$$

Rules for the calculus (2)

With scalars and vectors:

$$
\begin{aligned}
\mathrm{d}a &= 0_n \\
\mathrm{d}(\alpha u) &= \alpha\,\mathrm{d}u \\
\mathrm{d}(u + v) &= \mathrm{d}u + \mathrm{d}v \\
\mathrm{d}(u^\mathsf{T} v) &= (\mathrm{d}u)^\mathsf{T} v + u^\mathsf{T}\mathrm{d}v \\
\mathrm{d}(u^\mathsf{T}) &= (\mathrm{d}u)^\mathsf{T}
\end{aligned}
$$

where α and a are constant.
Example:
Find the derivative of $\phi(x) = x^\mathsf{T} x$.

$$\mathrm{d}\phi = \ldots$$

Rules for the calculus (3)

With scalars, vectors and matrices:

$$
\begin{aligned}
\mathrm{d}A &= 0_{m\times n} \\
\mathrm{d}(\alpha U) &= \alpha\, \mathrm{d}U \\
\mathrm{d}(U + V) &= \mathrm{d}U + \mathrm{d}V \\
\mathrm{d}(UV) &= (\mathrm{d}U)V + U\mathrm{d}V \\
\mathrm{d}(U^\mathsf{T}) &= (\mathrm{d}U)^\mathsf{T} \\
\mathrm{d}\,\mathrm{tr}\, U &= \mathrm{tr}\, \mathrm{d}U
\end{aligned}
$$

where α and A are constant.

Example:

Find the derivative of $\phi(X) = \mathrm{tr}(X^\mathsf{T} X)$.

$$\mathrm{d}\phi = \mathrm{d}\,\mathrm{tr}(X^\mathsf{T} X) = \mathrm{tr}((\mathrm{d}X)^\mathsf{T} X + X^\mathsf{T}\mathrm{d}X) = \mathrm{tr}(2X^\mathsf{T}\mathrm{d}X)$$

Thus

$$\mathrm{D}\phi(X) = 2(\mathrm{vec}\, X)^\mathsf{T}$$

Rules for the calculus (4)

Many more rules:

$$\begin{aligned} \mathrm{d}\alpha &= 0 & \mathrm{d}a &= 0_n & \mathrm{d}A &= 0_{m\times n} \\ \mathrm{d}(\alpha\phi) &= \alpha\,\mathrm{d}\phi & \mathrm{d}(\alpha u) &= \alpha\,\mathrm{d}u & \mathrm{d}(\alpha U) &= \alpha\,\mathrm{d}U \\ \mathrm{d}(\phi+\psi) &= \mathrm{d}\phi+\mathrm{d}\psi & \mathrm{d}(u+v) &= \mathrm{d}u+\mathrm{d}v & \mathrm{d}(U+V) &= \mathrm{d}U+\mathrm{d}V \\ \mathrm{d}(\phi\,\psi) &= (\mathrm{d}\phi)\psi+\phi\,\mathrm{d}\psi & \mathrm{d}(u^\top v) &= (\mathrm{d}u)^\top v+u^\top \mathrm{d}v & \mathrm{d}(UV) &= (\mathrm{d}U)\,V+U\mathrm{d}V \\ \mathrm{d}(\phi/\psi) &= ((\mathrm{d}\phi)\psi-\phi\,\mathrm{d}\psi)\,/\,\psi^2 & \mathrm{d}(u^\top) &= (\mathrm{d}u)^\top & \mathrm{d}(U^\top) &= (\mathrm{d}U)^\top \end{aligned}$$

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}\,U &= \mathrm{vec}\,\mathrm{d}U & \mathrm{d}\,\mathrm{tr}\,U &= \mathrm{tr}\,\mathrm{d}U \\ \mathrm{d}(U\otimes V) &= (\mathrm{d}U)\otimes V+U\otimes\mathrm{d}V & \mathrm{d}(U\odot V) &= (\mathrm{d}U)\odot V+U\odot\mathrm{d}V \\ \mathrm{d}(\phi^\alpha) &= \alpha\phi^{\alpha-1}\mathrm{d}\phi & \mathrm{d}(U^{-1}) &= -U^{-1}(\mathrm{d}U)\,U^{-1} \\ \mathrm{d}\det U &= \det(U)\,\mathrm{tr}(U^{-1}\mathrm{d}U) & \mathrm{d}\log(\det U) &= \mathrm{tr}(U^{-1}\mathrm{d}U) \\ \mathrm{d}\exp\phi &= \exp(\phi)\,\mathrm{d}\phi & \mathrm{tr}(\mathrm{d}\exp U) &= \mathrm{tr}(\exp(U)\,\mathrm{d}U) \end{aligned}$$

- α, a, A be constants
- ϕ, ψ, u, v, x, f, U, V, F be variables/functions.

Identification table (more identification formulas)

	function	differential	derivative	shape of derivative
$\phi(\xi)$	$\mathbb{R} \to \mathbb{R}$	$\mathrm{d}\phi = \alpha(\xi)\mathrm{d}\xi$	$\mathrm{D}\phi(\xi) = \alpha(\xi)$	1×1
$\phi(x)$	$\mathbb{R}^n \to \mathbb{R}$	$\mathrm{d}\phi = a(x)^\mathsf{T}\mathrm{d}x$	$\mathrm{D}\phi(x) = a(x)^\mathsf{T}$	$1 \times n$
$\phi(X)$	$\mathbb{R}^{n\times q} \to \mathbb{R}$	$\mathrm{d}\phi = \mathrm{tr}(A(X)^\mathsf{T}\mathrm{d}X)$	$\mathrm{D}\phi(X) = (\mathrm{vec}\, A(X))^\mathsf{T}$	$1 \times nq$
$f(\xi)$	$\mathbb{R} \to \mathbb{R}^m$	$\mathrm{d}f = a(\xi)\,\mathrm{d}\xi$	$\mathrm{D}f(\xi) = a(\xi)$	$m \times 1$
$f(x)$	$\mathbb{R}^n \to \mathbb{R}^m$	$\mathrm{d}f = A(x)\,\mathrm{d}x$	$\mathrm{D}f(x) = A(x)$	$m \times n$
$f(X)$	$\mathbb{R}^{n\times q} \to \mathbb{R}^m$	$\mathrm{d}f = A(X)\,\mathrm{dvec}\, X$	$\mathrm{D}f(X) = A(X)$	$m \times nq$
$F(\xi)$	$\mathbb{R} \to \mathbb{R}^{m\times p}$	$\mathrm{d}F = A(\xi)\,\mathrm{d}\xi$	$\mathrm{D}F(\xi) = \mathrm{vec}\, A(\xi)$	$mp \times 1$
$F(x)$	$\mathbb{R}^n \to \mathbb{R}^{m\times p}$	$\mathrm{dvec}\, F = A(x)\,\mathrm{d}x$	$\mathrm{D}F(x) = A(x)$	$mp \times n$
$F(X)$	$\mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$	$\mathrm{dvec}\, F = A(X)\,\mathrm{dvec}\, X$	$\mathrm{D}F(X) = A(X)$	$mp \times nq$

Note: The differential has always the same shape as the function.

Many more matrix tricks

$$\operatorname{tr}(AB) = \operatorname{tr}(BA) \qquad \operatorname{tr} A = 1_m^\top (A \odot I_{m\times n})\, 1_n = 1^T \operatorname{diag}(A) = \operatorname{tr} A^\top$$

$$\operatorname{diag}(UV^\top) = (U \odot V)1_n \qquad \operatorname{tr}(U^\top (V \odot C)) = \operatorname{tr}((U^\top \odot V^\top)C)$$

$$A \otimes 1_l = (I_m \otimes 1_l)A \qquad 1_l \otimes A = (1_l \otimes I_m)A$$

$$\operatorname{Diag} a = a1_n^T \odot I_n \qquad \operatorname{diag} A = (A \odot I_n)1_n$$

$$\operatorname{Diag}(\operatorname{diag} A) = I_n \odot A \qquad \|U\|_{\text{Fro}}^2 = \operatorname{tr}(U^\top U) = \operatorname{vec}(U)^\top \operatorname{vec}(U)$$

$$\operatorname{vec}(a) = \operatorname{vec}(a^\top) = a \qquad \operatorname{vec}(ABC) = (C^\top \otimes A)\operatorname{vec}(B)$$

$$\operatorname{tr}(u^\top v) = \operatorname{tr}(v^\top u) = v^\top u \qquad ABc = (c^\top \otimes A)\operatorname{vec} B = (A \otimes c^\top)\operatorname{vec} B^\top = \operatorname{vec}(c^\top B^\top A^\top)$$

$$\det(\exp A) = \exp(\operatorname{tr} A) \qquad \operatorname{vec} ab^\top = b \otimes a$$

Notation

$0_n, 0_{m\times n}$	n vector of zeros, $m \times n$ matrix of zeros
$1_n, 1_{m\times n}$	n vector of ones, $m \times n$ matrix of ones
$I_n, I_{m\times n}$	$n \times n$ identity matrix, $m \times n$ identity matrix
$\operatorname{vec} A$	vector containing the stacked columns of A
$\operatorname{diag} A$	vector containing the diagonal of A
$\operatorname{Diag} a$	diagonal matrix with a along the diagonal
$\exp \alpha, \exp A$	scalar exponential function, matrix exponential function
$A \odot B$	Hadamard product (component-wise product)
$A \oslash B$	component-wise division
$A \otimes B$	Kronecker product

Interlude: finite differencing (Python/Numpy version)

You should always check your derivatives with finite differencing which is an alternative way to calculate a derivative! Here is some Python/Numpy code:你应该经常用有限差分法来检查你的导数，这是计算导数的另一种方法! 这里有一些Python/Numpy代码。

```
import numpy as np
def finite_diff(f, x, delta):

    """estimate the gradient by finite-differencing method"""
    grad_f, dx = np.zeros_like(x), np.zeros_like(x)
    for i in range(x.size):
        dx.flat[i]     = delta
        grad_f.flat[i] = f(x+dx) - f(x-dx)
        dx.flat[i]     = 0.0
    return grad_f / (2*delta)
```

Interlude: finite differencing (MATLAB version)

You should always check your derivatives with finite differencing which is an alternative way to calculate a derivative! Here is some matlab code:你应该经常用有限差分法来检查你的导数，这是计算导数的另一种方法。是计算导数的另一种方法! 下面是一些matlab

代码。

```
function df = finitediff(fun, x, d, varargin)
% FINITEDIFF estimates a gradient by finite-differencing method.
% (c) Stefan Harmeling, 2012-07-10.
sx = size(x);
nx = numel(x);
df = zeros(sx);
dx = zeros(sx);
for i = 1:nx
  dx(i) = d;
  df(i) = (fun(x+dx, varargin{:})-fun(x-dx, varargin{:}))/(2*d);
  dx(i) = 0;
end
```

Matrix differential calculus矩阵微分计算

General recipe and examples

1. write the letter d in front of the expression
2. identity the constants and variables
3. transform the expression
4. read off the derivative using the identification table

Some pros and cons

+ clean notation
+ vectorized function leads to vectorized derivative (good for coding)
+ powerful: The book of Magnus/Neudecker shows how to take the derivative of eigenvalues and eigenvectors
– complicated formulas
– requires tricks and practice

More examples

Example: least squares

Find the derivative of $\phi(x) = (y - Ax)^2$.

$$\mathrm{d}\phi = \mathrm{d}((y - Ax)^\mathsf{T}(y - Ax)) = -2(y - Ax)^T A\,\mathrm{d}x$$

Thus:

$$\mathrm{D}\phi(x) = -2(y - Ax)^T A = (-2A^\mathsf{T}(y - Ax))^\mathsf{T}$$

Example: vector-valued function of matrix

Find the derivative of $f(X) = (X^\mathsf{T} X)^{-1} X^\mathsf{T} y$.
We write $A = (X^\mathsf{T} X)^{-1}$.

$$
\begin{aligned}
\mathrm{d}f &= \mathrm{d}(X^\mathsf{T} X)^{-1} X^\mathsf{T} y \\
&= -A((\mathrm{d}X)^T X + X^\mathsf{T}(\mathrm{d}X))AX^\mathsf{T} y \\
&= -A(\mathrm{d}X)^T XAX^\mathsf{T} y - AX^\mathsf{T}(\mathrm{d}X)AX^\mathsf{T} y - A(\mathrm{d}X)^\mathsf{T} y \\
&= -(A \otimes (XAX^\mathsf{T} y)^\mathsf{T})\,\mathrm{dvec}\,X - ((XAX^\mathsf{T} y)^\mathsf{T} \otimes A)\,\mathrm{dvec}\,X - (A \otimes y^\mathsf{T})\,\mathrm{dvec}\,X \\
&= \underbrace{-(A \otimes (XAX^\mathsf{T} y)^\mathsf{T} + (XAX^\mathsf{T} y)^\mathsf{T} \otimes A + A \otimes y^\mathsf{T})}_{\mathrm{D}f(X)}\,\mathrm{dvec}\,X
\end{aligned}
$$

Example: scalar-valued function of matrix例子：矩阵的标量值函数

Often it is easier to find the differential of a scalar function $\phi(X) = c^\mathsf{T}(X^\mathsf{T}X)^{-1}X^\mathsf{T}y$ where we again write $A = (X^\mathsf{T}X)^{-1}$.

$$
\begin{aligned}
\mathrm{d}\phi &= -c^\mathsf{T}A(\mathrm{d}X)^T XAX^\mathsf{T}y - c^\mathsf{T}AX^\mathsf{T}(\mathrm{d}X)AX^\mathsf{T}y + c^\mathsf{T}A(\mathrm{d}X)^\mathsf{T}y \\
&= -\operatorname{tr}(c^\mathsf{T}A(\mathrm{d}X)^T XAX^\mathsf{T}y) - \operatorname{tr}(c^\mathsf{T}AX^\mathsf{T}(\mathrm{d}X)AX^\mathsf{T}y) + \operatorname{tr}(c^\mathsf{T}A(\mathrm{d}X)^\mathsf{T}y) \\
&= -\operatorname{tr}(y^\mathsf{T}XA^\mathsf{T}X^\mathsf{T}(\mathrm{d}X)A^\mathsf{T}c) - \operatorname{tr}(c^\mathsf{T}AX^\mathsf{T}(\mathrm{d}X)AX^\mathsf{T}y) + \operatorname{tr}(y^\mathsf{T}(\mathrm{d}X)A^\mathsf{T}c) \\
&= -\operatorname{tr}(A^\mathsf{T}c\,y^\mathsf{T}XA^\mathsf{T}X^\mathsf{T}\mathrm{d}X) - \operatorname{tr}(AX^\mathsf{T}yc^\mathsf{T}AX^\mathsf{T}\mathrm{d}X) + \operatorname{tr}(A^\mathsf{T}cy^\mathsf{T}\mathrm{d}X) \\
&= \operatorname{tr}((-A^\mathsf{T}c\,y^\mathsf{T}XA^\mathsf{T}X^\mathsf{T} - AX^\mathsf{T}yc^\mathsf{T}AX^\mathsf{T} + A^\mathsf{T}cy^\mathsf{T})\mathrm{d}X) \\
&= \operatorname{tr}(\underbrace{(-Ac\,y^\mathsf{T}XAX^\mathsf{T} - AX^\mathsf{T}yc^\mathsf{T}AX^\mathsf{T} + Acy^\mathsf{T})}_{\text{call this } C(X)^\mathsf{T}}\,\mathrm{d}X)
\end{aligned}
$$

Using the identification formula: $\mathrm{D}\phi(X) = (\operatorname{vec} C(X))^\mathsf{T}$ with shape $1 \times nq$ we get:

$$
\begin{aligned}
\mathrm{D}\phi(X) &= (\operatorname{vec}(-Ac\,y^\mathsf{T}XAX^\mathsf{T} - AX^\mathsf{T}yc^\mathsf{T}AX^\mathsf{T} + Acy^\mathsf{T})^\mathsf{T})^\mathsf{T} \\
&= (\operatorname{vec}(-XAX^\mathsf{T}yc^\mathsf{T}A - XAcy^\mathsf{T}XA + yc^\mathsf{T}A))^\mathsf{T}
\end{aligned}
$$

Example: use indices or not例如：是否使用指数

Always avoiding indices can be painful:总是回避指数会让人感到痛苦。

$$\begin{aligned}\mathrm{d}\,\mathrm{tr}(A\,\mathrm{Diag}\,v) &= \mathrm{d}\,\mathrm{tr}(A(I_n \odot v1_n^{\mathsf{T}})) = \mathrm{d}\,\mathrm{tr}((A \odot I_n)v1_n^{\mathsf{T}}) \\ &= \mathrm{d}1_n^{\mathsf{T}}\,\mathrm{Diag}(A)v = \mathrm{d}\,\mathrm{diag}(A)^{\mathsf{T}}v = \mathrm{diag}(A)^{\mathsf{T}}\mathrm{d}v\end{aligned}$$

Thus better rewrite with indices in this case:因此，在这种情况下，最好用指数重写。

$$\mathrm{d}\,\mathrm{tr}(A\,\mathrm{Diag}\,v) = \mathrm{d}\sum_i A_{ii}v_i \qquad = \mathrm{d}\,\mathrm{tr}\,\mathrm{diag}(A)^{\mathsf{T}}v = \mathrm{tr}\,\mathrm{diag}(A)^{\mathsf{T}}\mathrm{d}v$$

Example: Rayleigh coefficient

Find the derivative of Rayleigh coefficient $\phi(x) = x^{\mathsf{T}}Ax/(x^{\mathsf{T}}x)$ for symmetric A.求对称A的瑞利系数 ϕ (x)=xTAx~(xTx)的导数

$$
\begin{aligned}
\mathrm{d}\phi &= \frac{2x^{\mathsf{T}}A(\mathrm{d}x)(x^{\mathsf{T}}x) - 2x^{\mathsf{T}}Axx^{\mathsf{T}}\mathrm{d}x}{(x^{\mathsf{T}}x)^2} \\
&= \frac{2(x^{\mathsf{T}}x)x^{\mathsf{T}}A(\mathrm{d}x) - 2x^{\mathsf{T}}Axx^{\mathsf{T}}\mathrm{d}x}{(x^{\mathsf{T}}x)^2} \\
&= \frac{2(x^{\mathsf{T}}x)x^{\mathsf{T}}A - 2x^{\mathsf{T}}Axx^{\mathsf{T}}}{(x^{\mathsf{T}}x)^2}\mathrm{d}x \\
&= \frac{2}{(x^{\mathsf{T}}x)^2}\; x^{\mathsf{T}}(xx^{\mathsf{T}}A - Axx^{\mathsf{T}})\,\mathrm{d}x
\end{aligned}
$$

Thus the derivative is:

$$
\begin{aligned}
\mathrm{D}\phi(x) &= \left(\frac{2}{(x^{\mathsf{T}}x)^2}\;(Axx^{\mathsf{T}} - xx^{\mathsf{T}}A)x\right)^{\mathsf{T}} \\
&= \left(2\frac{Ax}{x^Tx} - 2\frac{x^{\mathsf{T}}Ax}{(x^Tx)^2}x\right)^{\mathsf{T}}
\end{aligned}
$$

Example: derivative of steepest descent例子：最陡峭下降的导数

steepest descent:

$$x^{(k+1)} = x^{(k)} - \xi A^\mathsf{T}(y - Ax^{(k)})$$

We can derive the following differentials:我们可以推导出以下差值。

- Find the derivative of $x^{(k)}(x^{(0)})$.

$$\mathrm{d}x^{(k)} = (I_n + \xi A^\mathsf{T} A)^k \,\mathrm{d}x^{(0)}$$

- Find the derivative of $x^{(k+1)}(\xi)$.

$$\mathrm{d}x^{(k+1)} = -\left(\sum_{i=0}^{k} (\xi A^\mathsf{T} A + I_n)^i A^\mathsf{T}(y - Ax^{(k-i)})\right) \mathrm{d}\xi$$

- Find the derivative of $x^{(1)}(A)$.

$$\mathrm{d}x^{(1)} = -\xi(I_n \otimes (y - Ax^{(0)})^\mathsf{T} + (x^{(0)})^\mathsf{T} \otimes A^\mathsf{T})\,\mathrm{dvec}\,A$$

solution see `bn142.pdf`

End of Section 09