ELEC-E8116 Model Based Control Systems / solution 1

Problem 1:

Let $f(x)$ be a scalar-valued function of the vector x and let A be a square matrix with an appropriate dimension. By using a simple example, study what kind of a function $f(x) = x^T Ax$ is. Prove that

$$\frac{d(Ax)}{dx} = A \qquad \text{and} \qquad \frac{df(x)}{dx} = \underline{x}^T (A + A^T)$$

when the gradient is considered to be a row vector (in the literature the gradient is sometimes regarded as a row vector and sometimes as a column vector).

Solution

General about differentiation of matrices: Let $x(t) = [x_1(t)\ x_2(t) \cdots x_n(t)]^T$ and

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}(t) & \cdots & \cdots & a_{nn}(t) \end{bmatrix}$$

Then $\quad \dfrac{dx(t)}{dt} = \dot{x}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} \qquad \dot{A}(t) = \begin{bmatrix} \dot{a}_{11}(t) & \dot{a}_{12}(t) & \cdots & \dot{a}_{1n}(t) \\ \dot{a}_{21}(t) & \dot{a}_{22}(t) & \cdots & \dot{a}_{2n}(t) \\ \vdots & \vdots & \vdots & \vdots \\ \dot{a}_{n1}(t) & \cdots & \cdots & \dot{a}_{nn}(t) \end{bmatrix}$

If f is a scalar-valued function of the vector x, the gradient can be defined to be either a row- or a column vector. As a row vector

$$\text{grad } f = \nabla f = \frac{\partial f}{\partial x} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}.$$

For example $f(x) = x_1^2 + 2x_2 + x_3 \quad \Rightarrow \quad \dfrac{\partial f}{\partial x} = [2x_1 \quad 2 \quad 1]$

Let f be a multidimensional function $f = [f_1(x) \quad f_2(x) \quad \cdots \quad f_m(x)]^T$, where $x(t) = [x_1(t)\ x_2(t) \cdots x_n(t)]^T$. Then the "derivative" has the dimension $m \times n$ and it is called the Jacobi matrix

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

To the problem: Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. We obtain

$f(x) = x^T Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + (a_{12} + a_{21})x_1 x_2 + a_{22}x_2^2$, which is a quadratic function.
In optimization the LQ-problem = linear system, quadratic cost.

Now $Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$ and the derivative is the Jacobi matrix

$\frac{d(Ax)}{dx} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A$. Note that the derivative with respect to the vector *x* has been written as a row vector. Now calculate the gradient of *f* and again consider it a row vector

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + (a_{12} + a_{21})x_2 & 2a_{22}x_2 + (a_{12} + a_{21})x_1 \end{bmatrix}$$

On the other hand

$$x^T (A + A^T) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2a_{11} & a_{12} + a_{21} \\ a_{12} + a_{21} & 2a_{22} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + (a_{12} + a_{21})x_2 & 2a_{22}x_2 + (a_{12} + a_{21})x_1 \end{bmatrix}$$

which is the same result. Hence $\frac{\partial}{\partial x}(x^T Ax) = x^T (A + A^T)$ and $\frac{\partial}{\partial x}(Ax) = A$

Problem 2

Now consider the gradient of *f*(*x*) as a column vector. Show by a simple example that

$$\frac{d(Ax)}{dx} = A^T \qquad \frac{d(x^T Ax)}{dx} = (A + A^T)x$$

Solution
Similar as in the previous problem. The result matrices are transposes to those in the previous problem.

Problem 3

Show that $x^T (A - A^T) x = 0$ holds, when *x* is a vector and *A* a square matrix with an appropriate dimension.

Solution
First some basic results from matrix calculus. With proper dimensions it holds

$$A + B = B + A$$
$$A(B + C) = AB + AC$$
$$(A + B)^T = A^T + B^T$$
$$(AB)^T = B^T A^T$$
$A^{-1}A = AA^{-1} = I$ I is the identity matrix
$$(AB)^{-1} = B^{-1}A^{-1}$$
$$(A^T)^{-1} = (A^{-1})^T$$

Note that with matrices as a rule $AB \neq BA$, $(A + B)^{-1} \neq A^{-1} + B^{-1}$
Calculus with matrices in a symbolic form is essentially more difficult than with scalar quantities..

So, the claim is $x^T Ax = x^T A^T x$. But this is a scalar, and its value does not change when the transpose is taken. But the matrix calculus rules hold anyway and therefore

$x^T Ax = (x^T Ax)^T = x^T A^T x$ as claimed.

Problem 4

Let the criterion to be minimized be given as

$$J = \int_0^T \{x(t)'Px(t) + u(t)'Qu(t)\}dt$$

where ' denotes the transpose. Show that the square matrices P and Q can always be chosen as symmetric matrices..

Solution

$x'Px = x'(P + P' - P')x = x'(P + P')x - x'P'x = x'(P + P')x - x'Px$ (compare to the previous problem)

Hence $2x'Px = x'(P + P')x \;\Rightarrow\; x'Px = x'\left(\frac{P + P'}{2}\right)x$

But $P + P'$ and also $\frac{P + P'}{2}$ are symmetric matrices. They can always be used instead of an arbitrary P without changing the value of the expression. The same holds naturally for $u'Qu$.

Problem 5

Let A, B, C and D be *n*x*n*, *n*x*m*, *m*x*n*, *m*x*m* matrices. Prove the so-called *matrix-inverion lemma*

$$(A+BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$$

where it is assumed that all inverse matrices exist.

Solution

At first sight the inversion lemma looks quite complicated. But is turns out to be valuable in many matrix calculations in control theory. A good example is, when the least-squares estimation algorithm is changed in a recursive form.

Let us start from the claim and multiply both sides with $A+BDC$

$$
\begin{aligned}
&I = (A+BDC)\left[A^{-1} - A^{-1}B\left(D^{-1}+CA^{-1}B\right)^{-1}CA^{-1}\right] = \\
&I - B\left(D^{-1}+CA^{-1}B\right)^{-1}CA^{-1} + BDCA^{-1} - BDCA^{-1}B\left(D^{-1}+CA^{-1}B\right)^{-1}CA^{-1} = \\
&I + BDCA^{-1} - B\left(I + DCA^{-1}B\right)\left(D^{-1}+CA^{-1}B\right)^{-1}CA^{-1} = \\
&I + BDCA^{-1} - BD\left(D^{-1}+CA^{-1}B\right)\left(D^{-1}+CA^{-1}B\right)^{-1}CA^{-1} = \\
&I + BDCA^{-1} - BDCA^{-1} = \\
&I
\end{aligned}
$$

and an identity followed, Ok. (Note that since we started from the claim we must go through equivalences in order the proof to be sound.)