Derivatives involving the pseudo inverse

A wrt. pseudo inverse

$$\frac{\partial A^{\mathrm{T}}}{\partial (A^{+})_{ij}} = A^{\mathrm{T}}A(EL)_{ij} - A^{\mathrm{T}}(EL)_{ji}A^{\mathrm{T}} - A^{\mathrm{T}}A(EL)_{ij}AA^{+} \quad , \quad A \in R^{M\times N}, M \geq N \tag{1}$$

and as a special case when A is square

$$\frac{\partial A^{\mathrm{T}}}{\partial (A^{-1})_{ij}} = -A^{\mathrm{T}}(EL)_{ji}A^{\mathrm{T}} \quad , \quad A \in R^{M\times M} \tag{2}$$

proof

First, define $B = A^{+}$, i.e. also, $A = B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}$. Then

$$\begin{aligned}
\frac{\partial A^{\mathrm{T}}}{\partial (A^{+})_{ij}} =& \frac{\partial (BB^{\mathrm{T}})^{-1}B}{\partial (B)_{ij}} \\
=&(BB^{\mathrm{T}})^{-1}\frac{\partial B}{\partial (B)_{ij}} + \frac{\partial (BB^{\mathrm{T}})^{-1}}{\partial (B)_{ij}}B \\
=&(BB^{\mathrm{T}})^{-1}(EL)_{ij} - (BB^{\mathrm{T}})^{-1}\left[B\frac{\partial B^{\mathrm{T}}}{\partial (B)_{ij}} + \frac{\partial B}{\partial (B_{ij})}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}B \\
=&(BB^{\mathrm{T}})^{-1}(EL)_{ij} - (BB^{\mathrm{T}})^{-1}\left[B(EL)_{ji} + (EL)_{ij}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}B \\
=&A^{\mathrm{T}}A(EL)_{ij} - A^{\mathrm{T}}(EL)_{ji}A^{\mathrm{T}} - A^{\mathrm{T}}A(EL)_{ij}AA^{+}
\end{aligned}$$

alternative proof of (2)

$$\begin{aligned}
\frac{\partial A^{\mathrm{T}}(A^{\mathrm{T}})^{-1}}{\partial x} =& 0 \\
\frac{\partial A^{\mathrm{T}}}{\partial x}(A^{\mathrm{T}})^{-1} + A^{\mathrm{T}}\frac{\partial (A^{\mathrm{T}})^{-1}}{\partial x} =& 0 \\
\frac{\partial A^{\mathrm{T}}}{\partial x}(A^{\mathrm{T}})^{-1} =& - A^{\mathrm{T}}\frac{\partial (A^{\mathrm{T}})^{-1}}{\partial x} \\
\frac{\partial A^{\mathrm{T}}}{\partial x} =& - A^{\mathrm{T}}\frac{\partial (A^{\mathrm{T}})^{-1}}{\partial x}A^{\mathrm{T}}
\end{aligned}$$

log determinant of square wrt. A

$$\frac{\partial \log |A^{\mathrm{T}}A|}{\partial (A)_{mn}} = 2(A^{\mathrm{T}+})_{mn} \quad , \quad A \in R^{M\times N}, M \geq N \tag{3}$$

proof

$$
\begin{aligned}
&= \operatorname{Tr}\left(\frac{\partial \log |A^{\mathrm{T}}A|}{\partial A^{\mathrm{T}}A}\frac{\partial (A^{\mathrm{T}}A)^{\mathrm{T}}}{\partial (A)_{mn}}\right)\\
&= \operatorname{Tr}\left((A^{\mathrm{T}}A)'^{-1}\frac{\partial (A^{\mathrm{T}}A)^{\mathrm{T}}}{\partial (A)_{mn}}\right)\\
&= \operatorname{Tr}\left((A^{\mathrm{T}}A)^{-1}\frac{\partial A^{\mathrm{T}}A}{\partial (A)_{mn}}\right)\\
&= \operatorname{Tr}\left((A^{\mathrm{T}}A)^{-1}\left[\frac{\partial A^{\mathrm{T}}}{\partial (A)_{mn}}A + A^{\mathrm{T}}\frac{\partial A}{\partial (A)_{mn}}\right]\right)\\
&= \sum_{j=1}^{N}(A^{\mathrm{T}}A)^{-1}_{jn}A_{mj} + \sum_{j=1}^{N}(A^{\mathrm{T}}A)^{-1}_{nj}A^{\mathrm{T}}_{jm}\\
&=2\sum_{j=1}^{N}(A^{\mathrm{T}}A)^{-1}_{nj}A^{\mathrm{T}}_{jm}\\
&=2\left[(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\right]_{nm}\\
&=2(A^{+})_{nm}\\
&=2(A^{\mathrm{T}+})_{mn}
\end{aligned}
$$

log determinant of square wrt. pseudo inverse

$$
\frac{\partial \log |A^{\mathrm{T}}A|}{\partial (A^{+})_{ij}} = -2(A^{\mathrm{T}})_{ij} \quad , \quad A \in R^{M\times N}, M \geq N \tag{4}
$$

proof

Using (3) and the chain rule

$$
\begin{aligned}
\frac{\partial \log |A^{\mathrm{T}}A|}{\partial (A^{+})_{ij}} &=2\operatorname{Tr}\left((A^{\mathrm{T}})^{+}\frac{\partial A^{\mathrm{T}}}{\partial (A^{+})_{ij}}\right)\\
&\quad \text{(define } B = A^{+}\text{)}\\
&=2\operatorname{Tr}\left(B^{\mathrm{T}}\frac{\partial A^{\mathrm{T}}}{\partial (A^{+})_{ij}}\right)\\
&\quad \text{(using (1))}\\
&=2\operatorname{Tr}\left(B^{\mathrm{T}}((BB^{\mathrm{T}})^{-1}(EL)_{ij} - (BB^{\mathrm{T}})^{-1}\left[B(EL)_{ji} + (EL)_{ij}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}B)\right)\\
&=2\operatorname{Tr}\left(B^{\mathrm{T}}((BB^{\mathrm{T}})^{-1}(EL)_{ij}\right) - 2\operatorname{Tr}\left(B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}\left[B(EL)_{ji} + (EL)_{ij}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}B\right)\\
&=2\operatorname{Tr}\left(A(EL)_{ij}\right) - 2\operatorname{Tr}\left(A\left[B(EL)_{ji} + (EL)_{ij}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}B\right)\\
&=2\operatorname{Tr}\left(A(EL)_{ij}\right) - 2\operatorname{Tr}\left(\left[B(EL)_{ji} + (EL)_{ij}B^{\mathrm{T}}\right](BB^{\mathrm{T}})^{-1}\right)\\
&=2\operatorname{Tr}\left(A(EL)_{ij}\right) - 2\operatorname{Tr}\left(B(EL)_{ji}(BB^{\mathrm{T}})^{-1}\right) - 2\operatorname{Tr}\left((EL)_{ij}B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}\right)\\
&=2\operatorname{Tr}\left(A(EL)_{ij}\right) - 2\operatorname{Tr}\left(A(EL)_{ij}\right) - 2\operatorname{Tr}\left((EL)_{ij}B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}\right)\\
&=-2\operatorname{Tr}\left((EL)_{ij}A\right)\\
&=-2(A^{\mathrm{T}})_{ij}
\end{aligned}
$$

Integrals involving Dirac delta function

Scalar

$$\int p(s)\delta(vs-x)ds = \frac{1}{|v|}p(x/v) \tag{5}$$

proof

The delta function is defined by the tractable form

$$\int \delta(u-x/v)p(u)du = p(x/v).$$

To get the integral (5) to the tractable form use the transformation which satisfies $v\phi(u)-x = u-x/v$, namely,

$$\phi(u) = u/v - x/v^2 - x/v.$$

Then, transforming the integral and plugging in the Jacobian we get

$$\begin{aligned}\int p(s)\delta(vs-x)ds =& \int \left|\frac{\partial\phi(u)}{\partial u}\right| \delta(u-x/v)p(u)du\\ =&\frac{1}{|v|}p(x/v)\end{aligned}$$

Mixing matrix

$$\int p(s)\delta(As-x)ds = |\det A|^{-1}p(A^{-1}x) \tag{6}$$

proof

The delta function is defined by the tractable form

$$\int \delta(u-A^{-1}x)p(u)du = p(A^{-1}x).$$

To get the integral (6) to the tractable form use the transformation which satisfies $A\phi(u)-x = u-A^{-1}x$, namely,

$$\phi(u) = A^{-1}u - A^{-1}(A^{-1}x-x).$$

Then, transforming the integral and plugging in the Jacobian we get

$$\begin{aligned}\int p(s)\delta(As-x)ds =& \int \frac{\partial\phi(u)}{\partial u}\delta(u-A^{-1}x)p(u)du\\ =&|\det(A)|^{-1}p(A^{-1}x)\end{aligned}$$

Undercomplete mixing matrix

For $A \in R^{M\times N}, M \geq N$ we find

$$\int p(s)\delta(x-As)ds = \begin{cases} |A^{\mathrm{T}}A|^{-1/2}p(A^{+}x) & , x = AA^{+}x \\ 0 & , \text{otherwise} \end{cases} \tag{7}$$

proof

(This proof was outlined to me by Lars Kai Hansen except for the constraint part)
We shall make use of the mapping $x \mapsto (x_\perp\ ,\ x_\parallel) = (U_\perp^{\mathrm{T}} x\ ,\ U_\parallel^{\mathrm{T}} x)$, i.e. $U_\parallel^{\mathrm{T}} = (A^{\mathrm{T}}A)^{-1/2}A^{\mathrm{T}}$, such that $U_\parallel U_\parallel^{\mathrm{T}} A = A$.

$$\begin{aligned}
&= \int p(s)\delta(x_\perp)\delta(x_\parallel - (As)_\parallel)ds \\
&= \delta(x_\perp)\int p(s)\delta(x_\parallel - (As)_\parallel)ds \\
&= \delta(x_\perp)\int p(s)\delta((A^{\mathrm{T}}A)^{-1/2}A^{\mathrm{T}}x - (A^{\mathrm{T}}A)^{-1/2}A^{\mathrm{T}}As)ds \\
&\quad (\ldots\text{integral transformation}\ldots) \\
&= \delta(x_\perp)\int |\det(A^{\mathrm{T}}A)|^{-1/2}p(u)\delta((A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}x - u)du \\
&= \delta(x_\perp)|\det(A^{\mathrm{T}}A)|^{-1/2}p((A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}x) \\
&= \delta(x_\perp)|\det(A^{\mathrm{T}}A)|^{-1/2}p(A^{+}x) \\
&\quad (\ldots\text{introducing the constraint}\ldots) \\
&= \delta(x - AA^{+}x)|\det(A^{\mathrm{T}}A)|^{-1/2}p(A^{+}x) \\
&= \begin{cases} |A^{\mathrm{T}}A|^{-1/2}p(A^{+}x) & , x = AA^{+}x \\ 0 & , \text{otherwise} \end{cases}
\end{aligned}$$