Some useful results in linear algebra

1. If $\mathbf{A}$ is invertible, $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$. (Since it does not matter in what order we perform the inverse and the transpose, we can use the notation $\mathbf{A}^{-T}$ without risk of ambiguity.)
2. If $\mathbf{A}$ is symmetric and invertible, $\mathbf{A}^{-1}$ is also symmetric.
3. $\mathbf{B}_1 = \mathbf{A}\mathbf{A}^T$ and $\mathbf{B}_2 = \mathbf{A}^T\mathbf{A}$ are both symmetric and positive semidefinite.
4. $\mathbf{B} = \mathbf{A}\mathbf{A}^T + \lambda\mathbf{I}$ is positive definite.
5. If $\mathbf{A}$ is a square matrix with eigenvectors $\mathbf{x}_1, ..., \mathbf{x}_d$ which have the corresponding eigenvalues $\lambda_1, ..., \lambda_d$, the matrix $\mathbf{B} = \mathbf{A} + \mu\mathbf{I}$ has the same eigenvectors with corresponding eigenvalues, $\lambda_1 + \mu, ..., \lambda_d + \mu$.
6. Let $\mathbf{A}$ be a square, symmetric and positive definite matrix with eigenvectors $\mathbf{x}_1, ..., \mathbf{x}_d$ with the corresponding eigenvalues $\lambda_1, ..., \lambda_d$. Then $\mathbf{A}^{-1}$ has the same eigenvectors with the corresponding eigenvalues $1/\lambda_1, ..., 1/\lambda_d$.
7. $tr(\mathbf{A}) = tr(\mathbf{A}^T)$.
8. $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.
9. $tr(\mathbf{AB}) = tr(\mathbf{BA})$. (Note that, in general, $\mathbf{AB} \neq \mathbf{BA}$.)
10. Let $\mathbf{A}$ be a square matrix with eigenvalues $\lambda_1, ..., \lambda_d$. Then $tr(\mathbf{A}) = \sum_{i=1}^d \lambda_i$ and $|\mathbf{A}| = \prod_{i=1}^d \lambda_i$.

Derivatives of scalar functions of matrices and vectors

Let $f(\mathbf{A})$ be a scalar function of a $n \times m$ matrix $\mathbf{A}$. One of several possible definitions of the matrix derivative of $f(\mathbf{A})$ is the following:

$$\mathbf{D} = \frac{d}{d\mathbf{A}} f(\mathbf{A}) = \begin{pmatrix} \frac{d}{da_{11}} f(\mathbf{A}) & \frac{d}{da_{12}} f(\mathbf{A}) & \cdots & \frac{d}{da_{1m}} f(\mathbf{A}) \\ \frac{d}{da_{21}} f(\mathbf{A}) & \frac{d}{da_{22}} f(\mathbf{A}) & & \\ \vdots & & \ddots & \\ \frac{d}{da_{n1}} f(\mathbf{A}) & & & \frac{d}{da_{nm}} f(\mathbf{A}) \end{pmatrix} \quad (1)$$

where a_{ij} is the i, j'th element in $\mathbf{A}$.

Derivatives of expressions containing the trace

1. $\frac{d}{d\mathbf{A}} tr(\mathbf{BA}) = \frac{d}{d\mathbf{A}} tr(\mathbf{AB}) = \mathbf{B}^T$
2. $\frac{d}{d\mathbf{A}} tr(\mathbf{A}^T\mathbf{B}) = \frac{d}{d\mathbf{A}} tr(\mathbf{B}\mathbf{A}^T) = \mathbf{B}$
3. $\frac{d}{d\mathbf{A}} tr(\mathbf{A}^T\mathbf{R}\mathbf{A}) = (\mathbf{R} + \mathbf{R}^T)\mathbf{A}$
4. $\frac{d}{d\mathbf{A}} tr(\mathbf{A}\mathbf{R}\mathbf{A}^T) = \mathbf{A}(\mathbf{R} + \mathbf{R}^T)$
5. $\frac{d}{d\mathbf{A}} tr(\mathbf{B}\mathbf{A}^{-1}) = -(\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})^T$
6. $\frac{d}{d\mathbf{A}} tr((\mathbf{A}\mathbf{S}\mathbf{A}^T)^{-1}\mathbf{R}) = \frac{d}{d\mathbf{A}} tr(\mathbf{R}(\mathbf{A}\mathbf{S}\mathbf{A}^T)^{-1}) = -\mathbf{S}\mathbf{A}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}(\mathbf{R} + \mathbf{R}^T(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1})$
7. $\frac{d}{d\mathbf{A}} tr((\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A})) = \frac{d}{d\mathbf{A}} tr((\mathbf{A}^T\mathbf{S}_1\mathbf{A})(\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}) = \frac{d}{d\mathbf{A}_1} tr((\mathbf{A}_1^T\mathbf{S}_2\mathbf{A}_1)^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A}))\Big|_{\mathbf{A}_1=\mathbf{A}} + \frac{d}{d\mathbf{A}_1} tr((\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}(\mathbf{A}_2^T\mathbf{S}_1\mathbf{A}_2))\Big|_{\mathbf{A}_2=\mathbf{A}} = -2\mathbf{S}_2\mathbf{A}(\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A})(\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1} + 2\mathbf{S}_1\mathbf{A}(\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}$

Derivatives of expressions containing the determinant

1. If $\mathbf{A}$ is invertible, $\frac{d}{d\mathbf{A}}|\mathbf{A}| = |\mathbf{A}|\mathbf{A}^{-T}$.
2. If $\mathbf{S}$ is symmetric and $\mathbf{A}^T\mathbf{S}\mathbf{A}$ is invertible, $\frac{d}{d\mathbf{A}}\left|\mathbf{A}^T\mathbf{S}\mathbf{A}\right| = 2|\mathbf{A}^T\mathbf{S}\mathbf{A}|\mathbf{S}\mathbf{A}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}$
3. If $\mathbf{S}$ is symmetric and $\mathbf{A}^T\mathbf{S}\mathbf{A}$ is invertible, $\frac{d}{d\mathbf{A}}\ln\left|\mathbf{A}^T\mathbf{S}\mathbf{A}\right| = 2\mathbf{S}\mathbf{A}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}$

Some useful derivatives of the Gaussian PDF

Let

$$p(\mathbf{x}|\mathbf{m},\mathbf{C}) = \frac{1}{(2\pi)^{d/2}|\mathbf{C}|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})} \tag{2}$$

with $\mathbf{C}$ having full rank. Then

1. $\frac{d}{d\mathbf{m}}p(\mathbf{x}|\mathbf{m},\mathbf{C}) = p(\mathbf{x}|\mathbf{m},\mathbf{C})\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})$.
2. $\frac{d}{d\mathbf{m}}\ln p(\mathbf{x}|\mathbf{m},\mathbf{C}) = \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})$.
3. $\frac{d}{d\mathbf{C}}p(\mathbf{x}|\mathbf{m},\mathbf{C}) = -\frac{1}{2}p(\mathbf{x}|\mathbf{m},\mathbf{C})\mathbf{C}^{-1} + \frac{1}{2}p(\mathbf{x}|\mathbf{m},\mathbf{C})\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}$.
4. $\frac{d}{d\mathbf{C}}\ln p(\mathbf{x}|\mathbf{m},\mathbf{C}) = -\frac{1}{2}\mathbf{C}^{-1} + \frac{1}{2}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}$.