The projection (dot product, vector product) of $\mathbf{x}$ onto $\mathbf{k}_i$ which may be written $y_i = \mathbf{k}_i^T\mathbf{x} = \mathbf{x}^T\mathbf{k}_i$ giving the (noise-free) measurement y_i i.e. i-th coordinate of the corresponding vector in measurement space.

Given $\mathbf{y}$ and $\mathbf{K}$ the inverse problem, or the solution of linear equations, is a matter of determining a mapping back from measurement space into state space. Each equation $y_i = \mathbf{k}_i^T\mathbf{x}$ states that the point $\mathbf{x}$ lies in a space of points whose vectors projected onto $\mathbf{k}_i$ give the value y_i. This subspace is of dimension $n-1$ and it lies at right angles to $\mathbf{k}_i$. For example in two dimensions the equation $y = k^1x_1 + k^2x_2$ defines a line in (x_1, x_2) space at right angles to the vector (k^1, k^2). Solutions to the set of linear equations lie in the intersection of all of these subspaces, which may be a point, a space of one or more dimensions, or may not exist.

We will have to deal with cases where m may be greater than, less than, or equal to n, and consider the appropriate solution method in each case. The reader may be familiar with the description of the case $m < n$ as underdetermined (underconstrained or ill-posed), and expect the case $m = n$ to be well-determined and $m > n$ to be overdetermined. Unfortunately this simple description is not always correct, as a simple example will illustrate. Consider the 2×2 case:

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned} \tag{A.2}$$

Here we have two equations for two unknowns, but they are both overdetermined and underdetermined. They give us contradictory information about the combination $x_1 + x_2$, and so are overdetermined, and give no information about the linearly independent combination $x_1 - x_2$, and so are underdetermined. This situation, which is normally not so blatantly obvious, is called *mixed-determined*. The following set is also mixed-determined, but less obviously:

$$\begin{aligned} x_1 - 2x_2 + 3x_3 &= 4 \\ 2x_1 - x_2 + 4x_3 &= 3 \\ -x_1 - 4x_2 + x_3 &= 7 \end{aligned} \tag{A.3}$$

because the difference between the first two equations contradicts the difference between the last two. Given these hints, it should be quite clear how to construct linear systems with more equations than unknowns which are underdetermined, well-determined, mixed-determined or overdetermined.

Equations A.3 describe three planes in a 3-dimensional space, each pair of which intersects in a line, but the three lines to not intersect in a point to give a solution, as in the case of a well-determined problem. They are all parallel.

A.2 Eigenvectors and Eigenvalues

The eigenvalue problem associated with an arbitrary square matrix $\mathbf{A}$ of order n is to find *eigenvectors* $\mathbf{r}$ and scalar *eigenvalues* λ which satisfy

$$\mathbf{Ar} = \lambda\mathbf{r} \tag{A.4}$$

If $\mathbf{A}$ is regarded a coordinate transformation, then $\mathbf{r}$ has the same representation in the untransformed and transformed coordinates, apart from a factor λ. This looks like a strange thing to want to know, but eigenvectors turn out to have some extremely useful features. This equation is the same as $(\mathbf{A} - \lambda\mathbf{I})\mathbf{r} = 0$, a homogeneous equation, which can only have a solution other than $\mathbf{r} = 0$ if $\mathbf{A} - \lambda\mathbf{I}$ has rank less than n, i.e. if its determinant is zero. This leads to a polynomial equation of degree n in λ, with n solutions for the eigenvalues. Each λ can be substituted back in Eq. (A.4) and a solution for the corresponding eigenvector found. Eigenvalues may be real or a member of complex conjugate pair. An eigenvector can be scaled by an arbitrary factor, and still satisfy Eq. (A.4), so it is conventional to normalise them so that $\mathbf{r}^T\mathbf{r} = 1$ or, in the case of complex eigenvectors, $\mathbf{r}^\dagger\mathbf{r} = 1$, where $\dagger$ is the Hermitian adjoint, or transpose of the complex conjugate.

The eigenvectors can be assembled into a matrix $\mathbf{R}$:

$$\mathbf{AR} = \mathbf{R}\boldsymbol{\Lambda} \tag{A.5}$$

where the columns of $\mathbf{R}$ are the eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix, with the eigenvalues on the diagonal.

We can find the vectors and values of $\mathbf{A}^T$ by transposing (A.5):

$$\mathbf{R}^T\mathbf{A}^T = \boldsymbol{\Lambda}\mathbf{R}^T \tag{A.6}$$

then premultiplying and postmultiplying by $\mathbf{L} = (\mathbf{R}^T)^{-1}$

$$\mathbf{A}^T\mathbf{L} = \mathbf{L}\boldsymbol{\Lambda} \tag{A.7}$$

Thus $\mathbf{L}$ is the matrix of normalised eigenvectors of $\mathbf{A}^T$, and the eigenvalues are the same as those of as $\mathbf{A}$. They are called the *left* eigenvectors, because they operate on $\mathbf{A}$ (rather then $\mathbf{A}^T$) on the left: $\mathbf{L}^T\mathbf{A} = \boldsymbol{\Lambda}\mathbf{L}^T$, while $\mathbf{R}$ are called the *right* eigenvectors. By postmultiplying Eq. (A.5) by $\mathbf{L}^T$ we can express $\mathbf{A}$ in terms of its eigenvectors as

$$\mathbf{A} = \mathbf{R}\boldsymbol{\Lambda}\mathbf{L}^T = \sum \lambda_i \mathbf{r}_i \mathbf{l}_i^T \tag{A.8}$$

which is described as a 'spectral decomposition' of $\mathbf{A}$.

In the case of a symmetric matrix $\mathbf{S}$, where $\mathbf{S} = \mathbf{S}^T$, we must have $\mathbf{R} = \mathbf{L}$ by symmetry, so that $\mathbf{L}^T\mathbf{L} = \mathbf{L}\mathbf{L}^T = \mathbf{I}$ or $\mathbf{L}^T = \mathbf{L}^{-1}$, and the eigenvectors are *orthonormal*. In this case the eigenvalues are all real. If the matrix is positive definite the eigenvalues are all greater than zero, and similarly for a negative definite matrix.

The following is a summary of useful relations involving eigenvectors:

Asymmetric Matrices		*Symmetric Matrices*	
$\mathbf{AR} = \mathbf{R\Lambda}$	(A.9)	$\mathbf{SL} = \mathbf{L\Lambda}$	(A.21)
$\mathbf{L}^T\mathbf{A} = \mathbf{\Lambda L}^T$	(A.10)	$\mathbf{L}^T\mathbf{S} = \mathbf{\Lambda L}^T$	(A.22)
$\mathbf{L}^T = \mathbf{R}^{-1},\ \mathbf{R}^T = \mathbf{L}^{-1}$	(A.11)	$\mathbf{L}^T = \mathbf{L}^{-1}$	(A.23)
$\mathbf{LR}^T = \mathbf{L}^T\mathbf{R} = \mathbf{I}$	(A.12)	$\mathbf{LL}^T = \mathbf{L}^T\mathbf{L} = \mathbf{I}$	(A.24)
$\mathbf{A} = \mathbf{R\Lambda L}^T = \sum_i \lambda_i \mathbf{r}_i \mathbf{l}_i^T$	(A.13)	$\mathbf{S} = \mathbf{L\Lambda L}^T = \sum_i \lambda_i \mathbf{l}_i \mathbf{l}_i^T$	(A.25)
$\mathbf{A}^T = \mathbf{L\Lambda R}^T = \sum_i \lambda_i \mathbf{l}_i \mathbf{r}_i^T$	(A.14)		
$\mathbf{A}^{-1} = \mathbf{R\Lambda}^{-1}\mathbf{L}^T$	(A.15)	$\mathbf{S}^{-1} = \mathbf{L\Lambda}^{-1}\mathbf{L}^T$	(A.26)
$\mathbf{A}^n = \mathbf{R\Lambda}^n\mathbf{L}^T$	(A.16)	$\mathbf{S}^n = \mathbf{L\Lambda}^n\mathbf{L}^T$	(A.27)
$\mathbf{L}^T\mathbf{AR} = \mathbf{\Lambda}$	(A.17)	$\mathbf{L}^T\mathbf{SL} = \mathbf{\Lambda}$	(A.28)
$\mathbf{L}^T\mathbf{A}^n\mathbf{R} = \mathbf{\Lambda}^n$	(A.18)	$\mathbf{L}^T\mathbf{S}^n\mathbf{L} = \mathbf{\Lambda}^n$	(A.29)
$\mathbf{L}^T\mathbf{A}^{-1}\mathbf{R} = \mathbf{\Lambda}^{-1}$	(A.19)	$\mathbf{L}^T\mathbf{S}^{-1}\mathbf{L} = \mathbf{\Lambda}^{-1}$	(A.30)
$\lvert\mathbf{A}\rvert = \prod_i \lambda_i$	(A.20)	$\lvert\mathbf{S}\rvert = \prod_i \lambda_i$	(A.31)

Square roots of matrices

The relation $\mathbf{A}^n = \mathbf{R\Lambda}^n\mathbf{L}^T$ can be used for arbitrary powers of a matrix, in particular the square root such that $\mathbf{A} = \mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$. This square root of a matrix is not unique, because the diagonal elements of $\mathbf{\Lambda}^{\frac{1}{2}}$ in $\mathbf{R\Lambda}^{\frac{1}{2}}\mathbf{L}^T$ can have either sign, leading to 2^n possibilities.

We only use square roots of covariance matrices in this book. In this case we can see that $\mathbf{S}^{\frac{1}{2}} = \mathbf{L\Lambda}^{\frac{1}{2}}\mathbf{L}^T$ is symmetric. As well as these roots, symmetric matrices can also have non-symmetric roots satisfying $\mathbf{S} = (\mathbf{S}^{\frac{1}{2}})^T\mathbf{S}^{\frac{1}{2}}$, of which the Cholesky decomposition, $\mathbf{S} = \mathbf{T}^T\mathbf{T}$ where $\mathbf{T}$ is upper triangular, is the most useful, see section 5.8.1.1 and Exercise 5.3. There are an infinite number of non-symmetric square roots. If $\mathbf{S}^{\frac{1}{2}}$ is a square root, then clearly so is $\mathbf{XS}^{\frac{1}{2}}$ where $\mathbf{X}$ is any orthonormal matrix. The inverse symmetric square root is $\mathbf{S}^{-\frac{1}{2}} = \mathbf{L\Lambda}^{-\frac{1}{2}}\mathbf{L}^T$, and the inverse Cholesky decomposition is $\mathbf{S}^{-1} = \mathbf{T}^{-1}\mathbf{T}^{-T}$. The inverse square root $\mathbf{T}^{-1}$ is triangular, and its numerical effect is implemented efficiently by back substitution.

A.3 Principal Axes of a Quadratic Form

Consider the scalar equation:

$$\mathbf{x}^T\mathbf{Sx} = 1 \tag{A.32}$$

where $\mathbf{S}$ is symmetric. This is the equation of a quadratic surface centered on the origin, in n-space. $\mathbf{S}$ might be for example an inverse covariance matrix, when the equation might represent a surface of constant probability density. The normal to the surface is the vector $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{Sx}) = 2\mathbf{Sx}$, and $\mathbf{x}$ is the radius vector, so

$$\mathbf{Sx} = \lambda\mathbf{x} \tag{A.33}$$

is the problem of finding points where the normal and the radius vector are parallel. These are clearly where the principal axes intersect the surface. At these points, $\mathbf{x}^T\mathbf{Sx} = 1$ too, so $\mathbf{x}^T\lambda\mathbf{x} = 1$ or:

$$\lambda = \frac{1}{\mathbf{x}^T\mathbf{x}} \tag{A.34}$$

Thus the eigenvalues of a symmetric matrix are the reciprocals of the squares of the lengths of the principal axes of the associated ellipsoid. The lengths of the axes are independent of the coordinate system, so will also be invariant under an arbitrary rotation, that is one in which $(\textit{distance})^2 = \mathbf{x}^T\mathbf{x}$ is unchanged.

Consider using the eigenvectors of $\mathbf{S}$ to transform the equation for the quadratic surface:

$$\mathbf{x}^T\mathbf{L\Lambda L}^T\mathbf{x} = 1 \ \text{ or } \ \mathbf{x}'^T\mathbf{\Lambda x}' = 1 \ \text{ or } \ \sum \lambda_i x_i'^2 = 1 \tag{A.35}$$

where $\mathbf{x}' = \mathbf{L}^T\mathbf{x}'$ or $\mathbf{x} = \mathbf{Lx}'$. The result is a quadratic surface in which the principal axes coincide with the coordinate axes.

A.4 Singular Vector Decomposition

The standard eigenvalue problem is meaningless for non-square matrices because $\mathbf{Ar}$ will be of a different dimension from $\mathbf{r}$, and Eq. (A.4) is invalid.

However an eigenvalue problem can be constructed by considering the symmetric problem:

$$\begin{pmatrix} \mathbf{O} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \tag{A.36}$$

where $\mathbf{K}$ is an arbitrary non-square matrix with m rows and n columns and $\mathbf{v}$ is of dimension n and $\mathbf{u}$ is of dimension m. We use the symbol $\mathbf{K}$, because the main application in this book is for the weighting function matrix. The vectors $\mathbf{u}$ and $\mathbf{v}$ are called 'singular vectors' of $\mathbf{K}$, and λ is a singular value. The symmetric eigenvalue problem is equivalent to the 'shifted' eigenvalue problem (Lanczos, 1961):

$$\begin{aligned} \mathbf{Kv} &= \lambda\mathbf{u} \\ \mathbf{K}^T\mathbf{u} &= \lambda\mathbf{v} \end{aligned} \tag{A.37}$$

From Eq. (A.37) we can obtain by substitution:

$$\mathbf{K}^T\mathbf{Kv} = \lambda\mathbf{K}^T\mathbf{u} = \lambda^2\mathbf{v} \tag{A.38}$$

$$\mathbf{KK}^T\mathbf{u} = \lambda\mathbf{Kv} = \lambda^2\mathbf{u} \tag{A.39}$$

showing that $\mathbf{u}$ and $\mathbf{v}$ are the eigenvectors of $\mathbf{K}\mathbf{K}^T$ $(m \times m)$ and $\mathbf{K}^T\mathbf{K}$ $(n \times n)$ respectively, which consequently both must have the same set of eigenvalues, and that the singular value is real. Furthermore, if $(\mathbf{u}, \mathbf{v})$ are a pair of singular vectors with singular value λ, then so is $(\mathbf{u}, -\mathbf{v})$, with singular value $-\lambda$. Thus we need only consider pairs with positive singular values.

A little care is needed in constructing a matrix of singular vectors, because individual $\mathbf{u}$ and $\mathbf{v}$ vectors correspond to each other, yet there are potentially different numbers of $\mathbf{v}$ and $\mathbf{u}$ vectors. However, if the rank of $\mathbf{K}$ is p, then there will be p non-zero singular values, and both $\mathbf{K}\mathbf{K}^T$ and $\mathbf{K}^T\mathbf{K}$ will have p non-zero eigenvalues. Thus the surplus eigenvectors will have zero eigenvalues, and can be discarded and we can write:

$$\begin{pmatrix} \mathbf{O} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \mathbf{\Lambda} \tag{A.40}$$

where $\mathbf{\Lambda}$ is $p \times p$, $\mathbf{U}$ is $m \times p$, and $\mathbf{V}$ is $n \times p$. There will be $n + m - p$ more eigenvectors of the composite matrix, all with zero eigenvalue. The singular vectors and values have the following properties:

$$\begin{aligned}
\mathbf{K}\mathbf{V} &= \mathbf{U}\mathbf{\Lambda} && \text{(A.41)}\\
\mathbf{K}^T\mathbf{U} &= \mathbf{V}\mathbf{\Lambda} && \text{(A.42)}\\
\mathbf{U}^T\mathbf{K}\mathbf{V} &= \mathbf{V}^T\mathbf{K}^T\mathbf{U} = \mathbf{\Lambda} && \text{(A.43)}\\
\mathbf{K} &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \textstyle\sum_i \lambda_i \mathbf{u}_i \mathbf{v}_i^T && \text{(A.44)}\\
\mathbf{K}^T &= \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T && \text{(A.45)}\\
\mathbf{V}^T\mathbf{V} &= \mathbf{U}^T\mathbf{U} = \mathbf{I}_p && \text{(A.46)}\\
\mathbf{K}\mathbf{K}^T\mathbf{U} &= \mathbf{U}\mathbf{\Lambda}^2 && \text{(A.47)}\\
\mathbf{K}^T\mathbf{K}\mathbf{V} &= \mathbf{V}\mathbf{\Lambda}^2 && \text{(A.48)}
\end{aligned}$$

For $\mathbf{V}\mathbf{V}^T$ and $\mathbf{U}\mathbf{U}^T$ to yield unit matrices, they must be extended to be square using the remaining zero eigenvectors of $\mathbf{K}^T\mathbf{K}$ or $\mathbf{K}\mathbf{K}^T$. Because $\mathbf{U}^T\mathbf{K}\mathbf{V} = \mathbf{\Lambda}$ we describe $\mathbf{U}$ as left singular vectors, and $\mathbf{V}$ as right singular vectors.

Note that a square matrix has both singular vectors and values, and eigenvectors and values. In the case of symmetric matrices they will be the same, but not in the case of unsymmetric matrices.

The right vectors $\mathbf{V}$ form an orthonormal basis (coordinate system) in the row space and the left vectors form a basis in the column space. The matrix $\mathbf{K}$ maps the row space basis vector $\mathbf{v}$ into a corresponding column space basis vector $\mathbf{u}$ (apart from a scale change given by λ), and $\mathbf{K}^T$ maps $\mathbf{u}$ back into $\mathbf{v}$. Thus $\mathbf{U}$ and $\mathbf{V}$ are a natural pair of coordinate systems for the two spaces.

A.5 Determinant and Trace

The determinant and the trace are both important quantities in quantifying information content. The following elementary properties are useful:

$$\begin{aligned}
\mathrm{tr}(\mathbf{A}) &= \textstyle\sum_i A_{ii} && \text{(A.49)}\\
\mathrm{tr}(\mathbf{A}\mathbf{A}^T) &= \textstyle\sum_{ij} A_{ij}^2 && \text{(A.50)}\\
\mathrm{tr}(k\mathbf{A}) &= k\,\mathrm{tr}(\mathbf{A}) && \text{(A.51)}\\
\mathrm{tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}) && \text{(A.52)}\\
\mathrm{tr}(\mathbf{C}\mathbf{D}) &= \mathrm{tr}(\mathbf{D}\mathbf{C}) && \text{(A.53)}\\
\mathrm{tr}(\mathbf{a}\mathbf{b}^T) &= \mathbf{b}^T\mathbf{a} && \text{(A.54)}\\
\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}\mathbf{B}) &= \mathrm{tr}(\mathbf{A}) && \text{(A.55)}\\
\mathrm{tr}(\mathbf{A}) &= \textstyle\sum_i \lambda_i && \text{(A.56)}\\
\mathrm{tr}(\mathbf{A}^{-1}) &= \textstyle\sum_i \lambda_i^{-1} && \text{(A.57)}\\
|\mathbf{A}^T| &= |\mathbf{A}| && \text{(A.58)}\\
|k\mathbf{A}| &= k^n|\mathbf{A}| && \text{(A.59)}\\
|\mathbf{A}\mathbf{B}| &= |\mathbf{B}\mathbf{A}| = |\mathbf{A}||\mathbf{B}| && \text{(A.60)}\\
|\mathbf{A}^{-1}| &= |\mathbf{A}|^{-1} && \text{(A.61)}\\
|\mathbf{B}^{-1}\mathbf{A}\mathbf{B}| &= |\mathbf{A}| && \text{(A.62)}\\
|\mathbf{A}| &= \textstyle\prod_i \lambda_i && \text{(A.63)}\\
|\mathbf{I}+\mathbf{A}| &= \textstyle\prod_i (1+\lambda_i) && \text{(A.64)}\\
|\mathbf{I}+\mathbf{a}\mathbf{b}^T| &= 1 + \mathbf{b}^T\mathbf{a} && \text{(A.65)}
\end{aligned}$$

where $\mathbf{A}$ and $\mathbf{B}$ are square matrices of the same order, λ_i is the i-th eigenvalue of $\mathbf{A}$, n is the order of $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{D}$ are rectangular matrices such that $\mathbf{C}^T$ and $\mathbf{D}$ are the same size and shape, $\mathbf{a}$ and $\mathbf{b}$ are vectors of the same order, and k is a scalar.

A.6 Calculus with Matrices and Vectors

The differential $\mathrm{d}\mathbf{x}$ is often used to indicate an infinitesimal vector. Throughout this book it is used quite differently. In integral expressions the following meaning is intended:

$$\int f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \cdots \int f(x_1, x_2, \ldots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2 \ldots \mathrm{d}x_n, \tag{A.66}$$

i.e. $\mathrm{d}\mathbf{x}$ is an element of volume of state space. In derivatives, the notation $\partial\mathbf{y}/\partial\mathbf{x}$ means the Fréchet derivative, whose value is a matrix:

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{ij} = \left(\frac{\partial y_i}{\partial x_j}\right)_{i,k\neq j} \tag{A.67}$$

where the order of the subscripts is such that a Taylor expansion looks right. The convention is to associate the row subscript with $\mathbf{y}$ and the column subscript with $\mathbf{x}$, so that we can write a Taylor expansion in a familiar form:

$$\mathbf{y} = \mathbf{y}_0 + \frac{\partial \mathbf{y}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_0) + \ldots \tag{A.68}$$

We often need to find the minimum or maximum of some matrix expression, very often with respect to a vector. For example what value of $\mathbf{x}$ minimises $\mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{b}^T\mathbf{x}$? It is convenient to have a set of rules corresponding to the familiar rules of calculus with scalars so that we do not have to carry out such manipulations in components each time.

The derivative of a scalar valued expression with respect to a vector yields a vector, because it represents the set of derivatives of the scalar with respect to each element of the vector. The vector may be expressed as a row or a column, at your convenience. In components, the i-th element of the above example gives:

$$\left[\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{b}^T\mathbf{x})\right]_i = \frac{\partial}{\partial x_i}\left(\sum_{jk} x_j A_{jk} x_k + \lambda \sum_j b_j x_j\right) \tag{A.69}$$

$$= \sum_j x_j A_{ji} + \sum_k A_{ik} x_k + \lambda b_i \tag{A.70}$$

There is a slight subtlety in returning this to a matrix notation, because the last term would produce a row vector and the other two terms would produce column vectors if done blindly. The result can be expressed either as a row or a column

$$\frac{\partial}{\partial \mathbf{x}}[\mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{b}^T\mathbf{x}] = \mathbf{A}^T\mathbf{x} + \mathbf{A}\mathbf{x} + \lambda\mathbf{b} \tag{A.71}$$

or

$$\frac{\partial}{\partial \mathbf{x}}[\mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{b}^T\mathbf{x}] = \mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T + \lambda\mathbf{b}^T \tag{A.72}$$

Confirming that:

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A} \text{ and } \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{A}) = \mathbf{A}^T \tag{A.73}$$

is left as an exercise to the reader.

Note that the symbol $\nabla_{\mathbf{x}}$ may also be used as equivalent of $\frac{\partial}{\partial \mathbf{x}}$.