

1 GSW... Matrix Calculus

Matrices are rectangular arrays of numbers. They are usually written in terms of a capital bold letter, for example \mathbf{A} . In previous chapters we've looked at matrix algebra and matrix arithmetic.

Where things start getting much more interesting is when we start differentiating expressions with matrices and vectors in them. (This is a common requirement in signal processing when we're trying to optimise something.) Here, I'm only going to write about the particular case of differentiating a scalar function of vectors and matrices with respect to a vector. This is a bit easier than the more general case (differentiating a matrix or vector with respect to a matrix or vector), it's easier to visualise, and it covers most of the interesting cases in signal processing¹.

For those who want to skip this chapter and carry on, that's fine, all you really need to know is that for the scalar expression:

$$z = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 = (\mathbf{y} - \mathbf{A}\mathbf{x})^H (\mathbf{y} - \mathbf{A}\mathbf{x}) \quad (0.1)$$

where \mathbf{A} is a given matrix, \mathbf{y} is a given vector and \mathbf{x} is a variable vector, z will have a minimum value when²:

$$\mathbf{x} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y} \quad (0.2)$$

and that you can get to this minimum value using an iterative technique, starting from any value of \mathbf{x} , and moving by a small amount in the direction of the steepest slope:

$$\mathbf{x} \rightarrow \mathbf{x} + \alpha \mathbf{A}^H (\mathbf{y} - \mathbf{A}\mathbf{x}) \quad (0.3)$$

where α is a small scalar quantity, at each iteration.

1.1 Differentiating Matrix Expressions

Differentiating a scalar function with respect to a matrix or vector implies doing a partial differentiation with respect to every element in the matrix or vector in turn, and using the results as the elements of a new matrix or vector. For example, if we had a scalar function:

$$z = \|\mathbf{x}\|^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \dots \quad (0.4)$$

then differentiating this with respect to the vector element x_1 gives:

¹ This is, after all, a "getting started" book.

² This formula works provided there is a single value of \mathbf{x} that gives a minimum value of z . It's possible that there may be a range of different values of \mathbf{x} all of which give the same minimum value of z , and in this case, this formula doesn't work: the matrix $(\mathbf{A}^H \mathbf{A})$ will not be invertible. In this case we need some more information about which value of \mathbf{x} we want. I won't discuss this case in this chapter.

$$\frac{\partial z}{\partial x_1} = 2x_1 \quad (0.5)$$

so the first element in the differentiation result vector would be $2x_1$.

Repeat this for all the elements, and the resultant row vector (sometimes written ∇z , pronounced 'grad z ') becomes:

$$\nabla z = [2x_1 \quad 2x_2 \quad 2x_3 \quad 2x_4 \quad \dots] = 2\mathbf{x}^T \quad (0.6)$$

Note that ∇z is a row vector, not another column vector. Since z is a scalar, it must result from the multiplication of a row vector by a column vector, and if we want to be able to write expressions like:

$$\delta z = \frac{\partial z}{\partial \mathbf{x}} \delta \mathbf{x} \quad (0.7)$$

to determine the small change in z resulting from a small change in the vector \mathbf{x} , then $\frac{\partial z}{\partial \mathbf{x}}$ must be a row vector.

If the scalar function $z(\mathbf{x})$ is not differentiable (in the sense that when \mathbf{x} is composed of complex elements, differentiating with respect to the real and imaginary component of \mathbf{x} gives different answers), it's harder to define a gradient.

One final note before we start on a rather confusing subject: things would be a lot easier if I could restrict all the elements in the matrices and vectors to be real. However, since complex signals are a common feature of communications engineering due to the usefulness of the equivalent baseband representation of passband signals, I'll go through all the important derivations assuming complex terms.

1.2 The First Interesting Case: $z = \mathbf{y}^H \mathbf{A} \mathbf{x}$, with \mathbf{x} real

If the scalar function we want to differentiate is:

$$z = \mathbf{y}_n^* \mathbf{A}_{nm} \mathbf{x}_m \quad (0.8)$$

then doing a partial differentiation with respect to the k^{th} element of \mathbf{x} gives³:

$$\frac{\partial z}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_n \sum_m y_n^* A_{nm} x_m = \sum_n \sum_m y_n^* A_{nm} \frac{\partial x_m}{\partial x_k}$$

The only term of all the terms x_m that changes with a small change in x_k is the term when $x = k$. All the other ones don't change, so their contribution to the differential is zero, and we don't need to do the summation over m any more – the only value of m that gives any contribution to the result is the one where $m = k$, so we can just replace A_{nm} with A_{nk} and forget all the other terms in \mathbf{A} . That gives:

[continued on next page...]

$$\frac{\partial z}{\partial x_k} = y_n^* A_{nm} \frac{\partial x_m}{\partial x_k} = y_n^* A_{nk} = \left(\mathbf{y}^* \mathbf{A} \right)_k \quad (0.9)$$

since the only element of \mathbf{x} that varies with x_k is the k^{th} element, and the k^{th} element of \mathbf{x} (x_k) when differentiated by itself just gives one. This results in the row vector:

$$\frac{\partial z}{\partial \mathbf{x}} = \mathbf{y}^H \mathbf{A} \quad (0.10)$$

and this is exactly what we'd expect, since doing the differentiation from first principles gives:

$$\begin{aligned} z + \delta z &= \mathbf{y}^H \mathbf{A} (\mathbf{x} + \delta \mathbf{x}) \\ \delta z &= \mathbf{y}^H \mathbf{A} \delta \mathbf{x} \\ \frac{\delta z}{\delta \mathbf{x}} &= \mathbf{y}^H \mathbf{A} \end{aligned} \quad (0.11)$$

Here we can define a gradient, $\nabla z = \mathbf{y}^H \mathbf{A}$ such that for a small change in the vector \mathbf{x} :

$$\delta z = \nabla z \cdot \delta \mathbf{x} \quad (0.12)$$

1.3 The First Complex Case: $z = \mathbf{y}^H \mathbf{A} \mathbf{x}$, with \mathbf{x} complex

Multiplying a complex number with a constant (in this case elements of $\mathbf{y}^H \mathbf{A}$) is a differentiable function in the mathematical sense, and therefore it doesn't matter whether we differentiate with respect to the real or the imaginary component of the vector \mathbf{x} , we'll get the same answer. Let the real part of \mathbf{x} be the vector \mathbf{u} , and the imaginary part of \mathbf{x} be the vector $j\mathbf{v}$ (so that \mathbf{v} is a real vector too).

Then, suppose I differentiate with respect to the imaginary component, where $x_k = u_k + jv_k$. This gives:

$$\frac{\partial z}{\partial jv_k} = y_n^* A_{nm} \frac{\partial x_m}{\partial jv_k} = y_n^* A_{nk} = \left(\mathbf{y}^H \mathbf{A} \right)_k \quad (0.13)$$

which is the same result as we had before. So far, so good, these matrix expressions are behaving just like normal scalar functions.

1.4 The First Problematic Case: $z = \mathbf{x}^H \mathbf{A} \mathbf{y}$, with \mathbf{x} complex

While multiplying a complex number with a constant is differentiable function, taking the complex conjugate certainly isn't. Now the result we get will depend on whether we are

$$\frac{\partial z}{\partial x_k} = \sum_n y_n^* A_{nk} \frac{\partial x_k}{\partial x_k} = \sum_n y_n^* A_{nk} = \left(\mathbf{y}^H \mathbf{A} \right)_k$$

differentiating with respect to the real or the imaginary component of \mathbf{x} , so we'll have to do both. First, with respect to the real components, u_k :

$$\frac{\partial z}{\partial u_k} = \frac{\partial x_n^*}{\partial u_k} A_{nm} y_m = A_{km} y_m = (\mathbf{A}\mathbf{y})^T_k \quad (0.14)$$

and then with respect to the imaginary component, jv_k :

$$\frac{\partial z}{\partial jv_k} = \frac{\partial x_n^*}{\partial jv_k} A_{nm} y_m = -A_{km} y_m = (-\mathbf{A}\mathbf{y})^T_k \quad (0.15)$$

and we get two different results. It's now difficult to define a gradient, since to calculate how much the value of z will change for a small change in \mathbf{x} , we'd need to know how much of the small change in \mathbf{x} is in the real part of \mathbf{x} , and how much in the imaginary part.

If \mathbf{x} is always real, then $\mathbf{x} = \mathbf{u}$, and we can define a gradient row vector $\nabla z = (\mathbf{A}\mathbf{y})^T$ such that $\delta z = \nabla z \cdot \delta \mathbf{x}$ for small changes in \mathbf{x} . If \mathbf{x} is complex, then it is no longer possible to find an expression quite as simple for ∇z , however, we can note that:

$$\begin{aligned} \delta z &= (\mathbf{A}\mathbf{y})^T \Re\{\delta \mathbf{x}\} - (\mathbf{A}\mathbf{y})^T \Im\{\delta \mathbf{x}\} \\ \delta z &= (\mathbf{A}\mathbf{y})^T (\Re\{\delta \mathbf{x}\} - \Im\{\delta \mathbf{x}\}) = (\mathbf{A}\mathbf{y})^T \delta \mathbf{x}^* \end{aligned} \quad (0.16)$$

And therefore we can extend the idea of a gradient, provided we multiply the row vector $\nabla z = (\mathbf{A}\mathbf{y})^T$ by the complex conjugate of the small change in \mathbf{x} . Note this is different from the case considered before of $z = \mathbf{y}^H \mathbf{A}\mathbf{x}$, when we could define a gradient $\nabla z = \mathbf{y}^H \mathbf{A}$ such that $\delta z = \nabla z \cdot \delta \mathbf{x}$. In that case, using the complex conjugate of the small change $\delta \mathbf{x}$ in \mathbf{x} would be wrong. Confusing, isn't it? Sometimes the idea of a gradient is best avoided.

If we're looking for a turning point, then we're trying to find a value of \mathbf{x} that gives a minimum value⁴ for z . This only makes much sense if z is real, otherwise it's a bit difficult to define what we mean by "minimum" or "maximum". For example: is $(1 + j)$ bigger or smaller than 1.2?⁵ However, if we can arrange z to be real, then we can define a turning point to be a real maximum or minimum, and then it doesn't matter whether a small change is made in the real or imaginary component of \mathbf{x} , we'd want the value of z to remain the same. In this case we'd need to use both results, and set both to zero.

The other reason we might be interested in the differential of z is if we want to know which direction is "downhill" (in other words, how to change the vector \mathbf{x} in order to make z smaller). Again, this only makes much sense if z is real. Repeatedly changing the vector \mathbf{x} by some amount in the direction of making things smaller is the basis of several interesting algorithms

⁴ Or maximum value, of course, but the problem of finding a minimum value is more common.

⁵ If you answered "bigger" then you probably want to minimise the magnitude $|z|$, or perhaps the square of this, $|z|^2$, which is usually easier. If you answered "smaller" then you probably want to minimise the real part of z , $(z + z^*)/2$. In both cases, these are real numbers.

in signal processing, where the value of \mathbf{x} that minimises some function $z(\mathbf{x})$ is found using iterative techniques, moving a little closer after each iteration. I'll write about a simple case of this in the next section, and a lot more about these cases in the chapter on Iterative Techniques.

1.5 An Interesting Quadratic Case: $z = \mathbf{x}^H \mathbf{A} \mathbf{x}$, with \mathbf{x} complex

Again, we've got a term in the complex conjugate of \mathbf{x} , so the function is not strictly differentiable, we'll have to consider the real and imaginary parts of \mathbf{x} separately. First, the real part:

$$\begin{aligned} \frac{\partial z}{\partial u_k} &= \frac{\partial x_n^*}{\partial u_k} A_{nm} x_m + x_n^* A_{nm} \frac{\partial x_m}{\partial u_k} = A_{km} x_m + x_n^* A_{nk} = \left((\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A} \right)_k \\ \frac{\partial z}{\partial \mathbf{u}} &= (\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A} \end{aligned} \quad (0.17)$$

and then the imaginary part:

$$\begin{aligned} \frac{\partial z}{\partial jv_k} &= \frac{\partial x_n^*}{\partial jv_k} A_{nm} x_m + x_n^* A_{nm} \frac{\partial x_m}{\partial jv_k} = -A_{km} x_m + x_n^* A_{nk} = \left(-(\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A} \right)_k \\ \frac{\partial z}{\partial j\mathbf{v}} &= -(\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A} \end{aligned} \quad (0.18)$$

In general, that's about as far as we can go. However, in the special case where \mathbf{A} is Hermitian, then $\mathbf{A} = \mathbf{A}^H$, and z is real (for a proof of this, see the examples). Now we can note that $\mathbf{x}^H \mathbf{A} = \mathbf{x}^H \mathbf{A}^H = (\mathbf{A} \mathbf{x})^H$, and write⁶:

$$\frac{\partial z}{\partial \mathbf{u}} = (\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A}^H = (\mathbf{A} \mathbf{x})^H + (\mathbf{A} \mathbf{x})^T = 2\Re\{(\mathbf{A} \mathbf{x})^H\} \quad (0.19)$$

$$\frac{\partial z}{\partial j\mathbf{v}} = -(\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A}^H = (\mathbf{A} \mathbf{x})^H - (\mathbf{A} \mathbf{x})^T = 2\Im\{(\mathbf{A} \mathbf{x})^H\} \quad (0.20)$$

With z real, talking about the gradient of the function $z(\mathbf{x})$ starts to make sense. A plot of the variation of z with respect to the real and imaginary components of one element x_k of \mathbf{x} might look something like this:

⁶ Using the relationships: $x + x^* = 2\Re\{x\}$ and $x - x^* = 2\Im\{x\}$. If you're unsure about these, read the chapter on Complex Numbers.

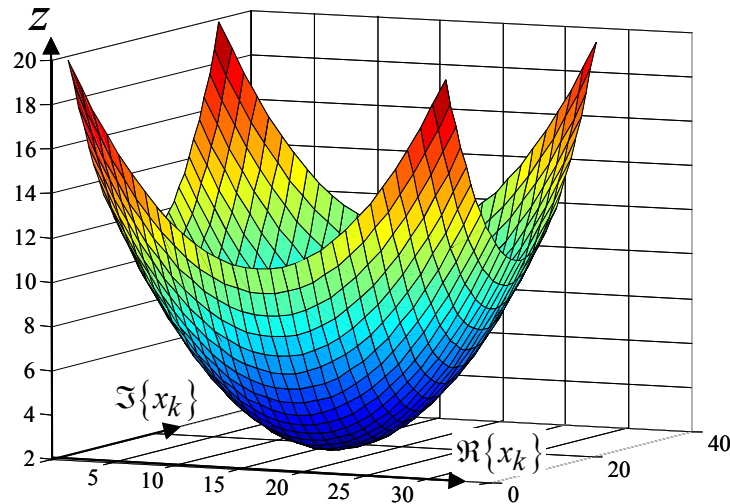


Figure 1 - Minimising The Real Function $z(\mathbf{x})$

and we could make similar looking plots for any component x_k , assuming all the other components remained constant. For any current value of vector \mathbf{x} , the change in the value of z for a small change δu_k in the real part of x_k is:

$$\delta z = 2\Re\left\{(\mathbf{A}\mathbf{x})^H\right\}_k \delta u_k \quad (0.21)$$

and for a small change in the imaginary part⁷ of x_k :

$$\delta z = 2\Im\left\{(\mathbf{A}\mathbf{x})^H\right\}_k \delta jv_k \quad (0.22)$$

Adding up the small changes in z for all the real and imaginary components of all the elements of vector \mathbf{x} then gives:

$$\begin{aligned} \delta z &= \sum_k 2\Re\left\{(\mathbf{A}\mathbf{x})^H\right\}_k \delta u_k + 2\Im\left\{(\mathbf{A}\mathbf{x})^H\right\}_k \delta jv_k \\ &= 2\Re\left\{(\mathbf{A}\mathbf{x})^H\right\} \delta \mathbf{u} + 2\Im\left\{(\mathbf{A}\mathbf{x})^H\right\} \delta j\mathbf{v} \end{aligned} \quad (0.23)$$

In addition, it's often useful to know in which 'direction' the maximum slope lies. In other words, if we can add a small vector of total length $\delta \mathbf{x}$ to \mathbf{x} , in which direction should we go in order to make the biggest possible difference to the value of z ?

⁷ Be careful when using this result. $2\Im\{\mathbf{A}\mathbf{x}\}$ is a purely imaginary number, and so is ∂jv_k . If you ask MATLAB for the imaginary part of a vector using 'imag(A*x)' it will give you a real answer: the imaginary part without the factor of j . If you want to know what the effect of adding a small amount to the imaginary part of \mathbf{x} would be using MATLAB, you have to ask for '-2*j*imag(A*x)*j*alpha' where alpha is the magnitude of the small change in the imaginary part. It's an easy mistake to forget both factors of j and end up with -1 times the answer you want.

1.5.1 The Steepest Slope

Take a look at that previous result a bit more closely:

$$\delta z = 2\Re\left\{(\mathbf{Ax})^H\right\}\delta\mathbf{u} + 2\Im\left\{(\mathbf{Ax})^H\right\}\delta j\mathbf{v} \quad (0.24)$$

We can use a result from complex number theory⁸: if f and g are complex numbers, then:

$$\Re\{f\}\Re\{g\} + \Im\{f\}\Im\{g\} = \Re\{fg\} \quad (0.25)$$

to simplify this to:

$$\delta z = 2\Re\left\{(\mathbf{Ax})^H \delta\mathbf{x}\right\} \quad (0.26)$$

Now if the inner product of two vectors of known length is to have the largest possible real component, then one vector must lie in the direction of the complex conjugate of the other one⁹. In this case, this implies that $\delta\mathbf{x}$ lies in the same direction as \mathbf{Ax} , and therefore, we could write:

$$\delta\mathbf{x} = \alpha(\mathbf{Ax}) \quad (0.27)$$

where α is a small, real, scalar value. If we want to make the largest possible increase in the value of z for any small vector of length $\delta\mathbf{x}$, then we need to make sure that $\delta\mathbf{x}$ lies in the same direction as \mathbf{Ax} .

Of course, if we want to decrease the value of z we should move in the opposite direction, which implies a small change in \mathbf{x} of:

$$\delta\mathbf{x} = -\alpha\mathbf{Ax} \quad (0.28)$$

and if we make this change in \mathbf{x} , the value of z will change according to:

$$\begin{aligned} \delta z &= -2\Re\left\{(\mathbf{Ax})^H \delta\mathbf{x}\right\} \\ &= -2\Re\left\{(\mathbf{Ax})^H (\alpha\mathbf{Ax})\right\} \\ &= -2\alpha\Re\left\{(\mathbf{Ax})^H (\mathbf{Ax})\right\} \end{aligned} \quad (0.29)$$

or more simply, since $(\mathbf{Ax})^H (\mathbf{Ax}) = \|\mathbf{Ax}\|^2$,

⁸ It's quite simple to derive: if $f = a + jb$ and $g = c + jd$, then $fg = (ac + jbjd) + j(bc + ad)$, and the real part of this is just $(ac + jbjd)$.

⁹ Consider one element of the vector, and write the complex numbers in the form $R_1\exp(j\theta_1)$ and $R_2\exp(j\theta_2)$. The contribution to the dot product of these elements is then $R_1 R_2 \exp(j(\theta_1 + \theta_2))$. The largest possible real component of this product is $R_1 R_2$ and this occurs when $\theta_1 = -\theta_2$. In other words, the two elements have opposite phases. The same goes for all the elements in the vector.

$$\delta z = -2\alpha \|\mathbf{Ax}\|^2 \quad (0.30)$$

Note this is always real, as expected (z is always real, so any small change must also be real).

It is difficult to define a gradient in the general case. For a small change in \mathbf{x} of $\delta\mathbf{x}$, we know that¹⁰:

$$\delta z = 2\Re\left\{(\mathbf{Ax})^H \delta\mathbf{x}\right\} \quad (0.31)$$

and in the optimum direction for minimising z , $(\mathbf{Ax})^H \delta\mathbf{x}$ is entirely real, and we can more simply write:

$$\delta z = (\mathbf{Ax})^H \delta\mathbf{x} \quad (0.32)$$

which suggests using $(\mathbf{Ax})^H$ as the gradient ∇z . However be careful: this does not mean that equation (0.32) holds under all circumstances. It's only true for one particular direction of the small vector $\delta\mathbf{x}$, the direction of the steepest slope of z . More generally, equation (0.31) must be used.

1.6 A Very Important Case: $z = \|\mathbf{y} - \mathbf{Ax}\|^2$

This is the last case we'll consider in this chapter, and without doubt the most important one. Note that z is a scalar, and since it is the sum of the square of the amplitudes of the components of the vector $(\mathbf{y} - \mathbf{Ax})$ it is always real, and never negative¹¹. Finding the turning points of expressions of this form is a very common problem in communications engineering.

I'll work out the answer to the problem in the general form: assuming \mathbf{y} , \mathbf{A} and \mathbf{x} are all complex. For z to be a maximum or minimum value, the differential of z with respect to all the real parts and all the imaginary parts of \mathbf{x} must be zero, so:

$$\frac{\partial(\|\mathbf{y} - \mathbf{Ax}\|^2)}{\partial u_k} = 0 \quad \frac{\partial(\|\mathbf{y} - \mathbf{Ax}\|^2)}{\partial jv_k} = 0 \quad (0.33)$$

where $x_k = u_k + jv_k$ for all values of the index k . Before we start differentiating, it's useful to note that for any vector \mathbf{w} :

$$\mathbf{w}^H \mathbf{w} = \sum_k w_k^* w_k = \sum_k |w_k|^2 = \|\mathbf{w}\|^2 \quad (0.34)$$

so we can write z as:

¹⁰ Using the result that the real part of x is equal to the real part of x^* .

¹¹ It could be zero: and in this case $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. However, if \mathbf{A} is not invertible, then the minimum value of z will be greater than zero.

$$\begin{aligned}
z &= \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 = (\mathbf{y} - \mathbf{A}\mathbf{x})^H (\mathbf{y} - \mathbf{A}\mathbf{x}) \\
&= \mathbf{y}^H \mathbf{y} - \mathbf{y}^H \mathbf{A}\mathbf{x} - (\mathbf{A}\mathbf{x})^H \mathbf{y} + (\mathbf{A}\mathbf{x})^H (\mathbf{A}\mathbf{x}) \\
&= \mathbf{y}^H \mathbf{y} - \mathbf{y}^H \mathbf{A}\mathbf{x} - \mathbf{x}^H \mathbf{A}^H \mathbf{y} + \mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x}
\end{aligned} \tag{0.35}$$

and to find a maximum or minimum point for z , we'll need to differentiate this expression with respect to the vector \mathbf{x} . The result should be a vector of zeros.

The first term in (0.35) isn't a function of \mathbf{x} at all, so when we differentiate this we'll get zero. The second term gives:

$$\frac{\partial}{\partial \mathbf{x}} (-\mathbf{y}^H \mathbf{A}\mathbf{x}) = -\mathbf{y}^H \mathbf{A} \tag{0.36}$$

from the results derived in section 1.3. The second and third terms are a bit more difficult, since they are not differentiable functions of \mathbf{x} , and we'll have to differentiate with respect to the real and imaginary components of \mathbf{x} separately.

For the third term, using the results derived in section 1.4, we get:

$$\begin{aligned}
\frac{\partial (-\mathbf{x}^H \mathbf{A}^H \mathbf{y})}{\partial \mathbf{u}} &= -(\mathbf{A}^H \mathbf{y})^T \\
\frac{\partial (-\mathbf{x}^H \mathbf{A}^H \mathbf{y})}{\partial j\mathbf{v}} &= (\mathbf{A}^H \mathbf{y})^T
\end{aligned} \tag{0.37}$$

Before we work out the result for the fourth term, note that for any matrix, $\mathbf{A}^H \mathbf{A}$ is Hermitian¹². Then, the results derived in section 1.5 give:

$$\frac{\partial}{\partial \mathbf{u}} (\mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x}) = 2\Re \left\{ (\mathbf{A}^H \mathbf{A}\mathbf{x})^H \right\} \tag{0.38}$$

$$\frac{\partial}{\partial j\mathbf{v}} (\mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x}) = 2\Im \left\{ (\mathbf{A}^H \mathbf{A}\mathbf{x})^H \right\} \tag{0.39}$$

Putting all of this together, differentiating with respect to the real component of \mathbf{x} gives:

$$\begin{aligned}
\frac{\partial z}{\partial \mathbf{u}} &= \frac{\partial (\mathbf{y}^H \mathbf{y})}{\partial \mathbf{u}} - \frac{\partial (\mathbf{y}^H \mathbf{A}\mathbf{x})}{\partial \mathbf{u}} - \frac{\partial (\mathbf{x}^H \mathbf{A}^H \mathbf{y})}{\partial \mathbf{u}} + \frac{\partial (\mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x})}{\partial \mathbf{u}} \\
&= 0 - \mathbf{y}^H \mathbf{A} - (\mathbf{A}^H \mathbf{y})^T + 2\Re \left\{ (\mathbf{A}^H \mathbf{A}\mathbf{x})^H \right\}
\end{aligned} \tag{0.40}$$

¹² See the problems and solution for the proof of this useful fact.

and this result can be simplified by noting that:

$$\left(\mathbf{A}^H \mathbf{y}\right)^T = \mathbf{y}^T \mathbf{A}^* = \left(\mathbf{y}^H \mathbf{A}\right)^* \quad (0.41)$$

so the first two terms can be expressed as:

$$-\mathbf{y}^H \mathbf{A} - \left(\mathbf{A}^H \mathbf{y}\right)^T = -\mathbf{y}^H \mathbf{A} - \left(\mathbf{y}^H \mathbf{A}\right)^* = -2\Re\left\{\mathbf{y}^H \mathbf{A}\right\} \quad (0.42)$$

so this gives:

$$\begin{aligned} \frac{\partial z}{\partial \mathbf{u}} &= 2\Re\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H\right\} - 2\Re\left\{\mathbf{y}^H \mathbf{A}\right\} \\ &= 2\Re\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \end{aligned} \quad (0.43)$$

Now, with respect to the imaginary component of \mathbf{x} :

$$\begin{aligned} \frac{\partial z}{\partial j\mathbf{v}} &= \frac{\partial\left(\mathbf{y}^H \mathbf{y}\right)}{\partial j\mathbf{v}} - \frac{\partial\left(\mathbf{y}^H \mathbf{A} \mathbf{x}\right)}{\partial j\mathbf{v}} - \frac{\partial\left(\mathbf{x}^H \mathbf{A}^H \mathbf{y}\right)}{\partial j\mathbf{v}} + \frac{\partial\left(\mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x}\right)}{\partial j\mathbf{v}} \\ &= 0 - \mathbf{y}^H \mathbf{A} + \left(\mathbf{A}^H \mathbf{y}\right)^T + 2\Im\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H\right\} \end{aligned} \quad (0.44)$$

and with similar simplification of the first two terms, we get:

$$\begin{aligned} \frac{\partial z}{\partial j\mathbf{v}} &= 2\Im\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H\right\} - 2\Im\left\{\mathbf{y}^H \mathbf{A}\right\} \\ &= 2\Im\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \end{aligned} \quad (0.45)$$

Now if we're trying to find a global minimum for z , then we need both of these expressions to be zero, so that it doesn't matter if there is a small change in any of the real or imaginary components of \mathbf{x} , the value of z won't change.

So, we need:

$$\frac{\partial z}{\partial \mathbf{u}} = 2\Re\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} = 0 \quad (0.46)$$

and also that,

$$\frac{\partial z}{\partial j\mathbf{v}} = 2\Im\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} = 0 \quad (0.47)$$

Obviously, if the real part of something is zero, and the imaginary part of the same thing is also zero, then the whole thing must be zero:

$$\begin{aligned} \left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A} &= 0 \\ \left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H &= \mathbf{y}^H \mathbf{A} \end{aligned} \quad (0.48)$$

Take the Hermitian (the complex transpose) of both sides of this equation, and we get:

$$\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{A}^H \mathbf{y} \quad (0.49)$$

or in terms of \mathbf{x} :

$$\mathbf{x} = \left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{y} \quad (0.50)$$

This formula comes up time and time again in signal processing. It's called the *normal equation*¹³.

Another useful result is to calculate the small change in z that results from a small change in the value of \mathbf{x} of $\delta \mathbf{x} = \delta \mathbf{u} + \delta j \mathbf{v}$:

$$\begin{aligned} \delta z &= \left(\frac{\partial z}{\partial \mathbf{u}}\right)^T \delta \mathbf{u} + \left(\frac{\partial z}{\partial j \mathbf{v}}\right)^T \delta j \mathbf{v} \\ &= 2\Re\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \Re\{\delta \mathbf{x}\} + 2\Im\left\{\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \Im\{\delta \mathbf{x}\} \end{aligned} \quad (0.51)$$

and using the result from complex number theory, that:

$$\Re\{a\} \Re\{b\} + \Im\{a\} \Im\{b\} = \Re\{ab\} \quad (0.52)$$

gives:

$$\delta z = 2\Re\left\{\left(\left(\mathbf{A}^H \mathbf{A} \mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right) \delta \mathbf{x}\right\} \quad (0.53)$$

1.6.1 An Iterative Solution of $z = \|\mathbf{y} - \mathbf{A} \mathbf{x}\|^2$

Solving the normal equation requires the calculation of a matrix inverse, and that can take a lot of time, particularly if the matrix \mathbf{A} is large. An alternative way to find the value of \mathbf{x} that minimises z is to use an iterative technique: start from a guess of \mathbf{x} , and find a way to modify \mathbf{x} so that the value of z is always decreasing. Eventually, we should get to the value of \mathbf{x} that minimises z .

Let \mathbf{x} be modified at each step, by the addition of a small vector $\delta \mathbf{x}$:

$$\mathbf{x} \rightarrow \mathbf{x} + \delta \mathbf{x} \quad (0.54)$$

¹³ That's "normal" in the sense of "perpendicular". For the reason why this equation is called the normal equation, see the first chapter on Linear Algebra.

in order to ensure that the new value of z is smaller than the previous value. We've just shown that for a small change in \mathbf{x} , $\delta\mathbf{x}$:

$$\delta z = 2\Re\left\{\left(\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \delta\mathbf{x} \quad (0.55)$$

so for the change $\delta\mathbf{x}$ to have the maximum effect, we need the product of $\left(\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}$ and $\delta\mathbf{x}$ to be entirely real. This is easily achieved by setting:

$$\begin{aligned} \delta\mathbf{x} &= -\alpha \left(\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A} \\ &= \alpha \mathbf{A}^H (\mathbf{y} - \mathbf{A}\mathbf{x}) \end{aligned} \quad (0.56)$$

so that:

$$\begin{aligned} \delta z &= -2\alpha \Re\left\{\left(\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\} \left(\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A} \\ &= -2\alpha \left\|\mathbf{A}^H \mathbf{A}\mathbf{x}\right)^H - \mathbf{y}^H \mathbf{A}\right\|^2 \end{aligned} \quad (0.57)$$

and since this is -2α times a positive quantity, it must always be negative: z is decreasing.

This provides an iterative way to get the minimum value of z : just keep modifying \mathbf{x} according to:

$$\mathbf{x} \rightarrow \mathbf{x} + \alpha \mathbf{A}^H (\mathbf{y} - \mathbf{A}\mathbf{x}) \quad (0.58)$$

and eventually you'll get there, without having to do any matrix inversions. This is the basic idea behind a well-known technique called the *method of steepest descent*, for more details see the chapter on Iterative Techniques.

1.7 Summary of Key Results

Differential of $\mathbf{y}^H \mathbf{A}\mathbf{x}$ with respect to real part of \mathbf{x}	$\frac{\partial}{\partial \mathbf{u}} (\mathbf{y}^H \mathbf{A}\mathbf{x}) = \mathbf{y}^H \mathbf{A}$
Differential of $\mathbf{y}^H \mathbf{A}\mathbf{x}$ with respect to imaginary part of \mathbf{x}	$\frac{\partial}{\partial \mathbf{v}} (\mathbf{y}^H \mathbf{A}\mathbf{x}) = \mathbf{y}^H \mathbf{A}$
Differential of $\mathbf{x}^H \mathbf{A}\mathbf{y}$ with respect to real part of \mathbf{x}	$\frac{\partial}{\partial \mathbf{u}} (\mathbf{x}^H \mathbf{A}\mathbf{y}) = (\mathbf{A}\mathbf{y})^T$
Differential of $\mathbf{x}^H \mathbf{A}\mathbf{y}$ with respect to imaginary part of \mathbf{x}	$\frac{\partial}{\partial \mathbf{v}} (\mathbf{x}^H \mathbf{A}\mathbf{y}) = -(\mathbf{A}\mathbf{y})^T$
Differential of $\mathbf{x}^H \mathbf{A}\mathbf{x}$ with respect to real part of \mathbf{x}	$\frac{\partial}{\partial \mathbf{u}} (\mathbf{x}^H \mathbf{A}\mathbf{x}) = (\mathbf{A}\mathbf{x})^T + \mathbf{x}^H \mathbf{A}$

Differential of $\mathbf{x}^H \mathbf{A} \mathbf{x}$ with respect to imaginary part of \mathbf{x}	$\frac{\partial}{\partial j\mathbf{v}} (\mathbf{x}^H \mathbf{A} \mathbf{x}) = -(\mathbf{A} \mathbf{x})^T + \mathbf{x}^H \mathbf{A}$
Differential of $\ \mathbf{y} - \mathbf{A} \mathbf{x}\ ^2$ with respect to real part of \mathbf{x}	$\frac{\partial (\ \mathbf{y} - \mathbf{A} \mathbf{x}\ ^2)}{\partial \mathbf{u}} = 2\Re \left\{ (\mathbf{A}^H \mathbf{A} \mathbf{x})^H - \mathbf{y}^H \mathbf{A} \right\}$
Differential of $\ \mathbf{y} - \mathbf{A} \mathbf{x}\ ^2$ with respect to imaginary part of \mathbf{x}	$\frac{\partial (\ \mathbf{y} - \mathbf{A} \mathbf{x}\ ^2)}{\partial j\mathbf{v}} = 2\Im \left\{ (\mathbf{A}^H \mathbf{A} \mathbf{x})^H - \mathbf{y}^H \mathbf{A} \right\}$
If $z = \ \mathbf{y} - \mathbf{A} \mathbf{x}\ ^2$ has a minimum, it occurs when:	$\mathbf{x} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y}$

1.8 Tutorial Questions

- 1) Show that the result of differentiating $\mathbf{y} = \mathbf{x}^T \mathbf{A}$ by \mathbf{x} is \mathbf{A} .
- 2) If \mathbf{A} is Hermitian, prove that $\mathbf{x}^H \mathbf{A} \mathbf{x}$ is real, and that for any matrix \mathbf{A} , the product $\mathbf{A}^H \mathbf{A}$ is Hermitian.
- 3) What is the differential of $(\mathbf{A} \mathbf{x})^T \mathbf{x}$ with respect to \mathbf{x} ? (Assume \mathbf{x} is real.)
- 4) Given that $\delta z = 2\Re \{ \mathbf{A} \mathbf{x} \}^T \delta \mathbf{u} - 2\Im \{ \mathbf{A} \mathbf{x} \}^T \delta j\mathbf{v}$, where $\delta \mathbf{x} = \delta \mathbf{u} + \delta j\mathbf{v}$, show that:

$$\delta z = 2\Re \left\{ (\mathbf{A} \mathbf{x})^H (\delta \mathbf{x}) \right\}$$

- 5) Given the vector \mathbf{y} and matrix \mathbf{A} as: $\mathbf{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$, what value of vector \mathbf{x} gives the minimum possible value of $\|\mathbf{y} - \mathbf{A} \mathbf{x}\|^2$? And what is this minimum value?

- 6) Given the vector \mathbf{x} and matrix \mathbf{A} as: $\mathbf{x} = \begin{bmatrix} 2+j \\ 1-3j \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 3 & 2-j \\ 2+j & 1 \end{bmatrix}$, what change in $z = \mathbf{x}^H \mathbf{A} \mathbf{x}$ results from adding a small vector $\delta \mathbf{x} = \alpha \begin{bmatrix} 1-2j \\ 2+j \end{bmatrix}$ to \mathbf{x} , where α is a small quantity?

Calculate the answers to both directly, and using the results in the text.

- 7) In the above problem, what direction should the vector $\delta \mathbf{x}$ be taken to give the maximum change in z ?
- 8) (For those who have a copy of MATLAB.) Write a MATLAB script that solves the problem of finding a minimum value of $\|\mathbf{y} - \mathbf{A} \mathbf{x}\|^2$ where \mathbf{y} and \mathbf{A} are known iteratively, without doing any matrix inversions.

Try it out with $\mathbf{A} = \begin{bmatrix} 1+j & -3j \\ -1+j & 2 \\ 2 & 2+j \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1+j \\ 2-j \\ 1 \end{bmatrix}$ and a starting value of $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

How many iterations does it take to get within 0.001 of the final answers, and which value of α is the best one to choose? Is this a more efficient technique for solving this problem than taking the matrix inverse?