2.5 The Singular Vector Expansion

Least-squares is a very powerful, very useful method for finding solutions of linear simultaneous equations of any dimensionality and one might wonder why it is necessary to discuss any other form of solution. But even in the simplest form of least-squares, the solution is dependent upon the inverses of $\mathbf{E}^T\mathbf{E}$, or $\mathbf{E}\mathbf{E}^T$. In practice, their existence cannot be guaranteed, and we need to understand what that means, the extent to which solutions can be found when the inverses do not exist and the effect of introducing weight matrices $\mathbf{W}$, $\mathbf{S}$. This problem is intimately related to the issue of controlling solution and residual norms. Second, the relationship between the equations and the solutions is somewhat impenetrable, in the sense that structures in the solutions are not easily relatable to particular elements of the data y_i. For many purposes, particularly physical insight, understanding the structure of the solution is essential. We will return to examine the least-squares solutions using some extra machinery.

2.5.1 Simple Vector Expansions

Consider again the elementary problem (2.1) of representing an L–dimensional vector $\mathbf{f}$ as a sum of a complete set of L–orthonormal vectors $\mathbf{g}_i$, $1 \le i \le L$, $\mathbf{g}_i^T\mathbf{g}_j = \delta_{ij}$. Without error,

$$\mathbf{f} = \sum_{j=1}^{L} a_j \mathbf{g}_j, \quad a_j = \mathbf{g}_j^T \mathbf{f}\,. \tag{2.185}$$ {34001}

But if for some reason, only the first K coefficients a_j are known, we can only approximate $\mathbf{f}$ by its first K terms:

$$\begin{aligned} \tilde{\mathbf{f}} &= \sum_{j=1}^{K} b_j \mathbf{g}_j \\ &= \mathbf{f} + \delta\mathbf{f}_1, \end{aligned} \tag{2.186}$$ {34002}

and there is an error, $\delta\mathbf{f}_1$. From the orthogonality of the $\mathbf{g}_i$, it follows that $\delta\mathbf{f}_1$ will have minimum l_2 norm only if it is orthogonal to the K vectors retained in the approximation, and then only if $b_j = a_j$ as given by (2.185). The only way the error could be reduced further is by increasing K.

Define an $L \times K$ matrix, $\mathbf{G}_K$ whose columns are the first K of the $\mathbf{g}_j$. Then $\mathbf{b} = \mathbf{a} = \mathbf{G}_K^T\mathbf{f}$ is the vector of coefficients $a_j = \mathbf{g}_j^T\mathbf{f}$, $1 \le j \le K$, and the finite representation (2.186) is (one should write it out),

$$\tilde{\mathbf{f}} = \mathbf{G}_K\mathbf{a} = \mathbf{G}_K(\mathbf{G}_K^T\mathbf{f}) = (\mathbf{G}_K\mathbf{G}_K^T)\mathbf{f}\,, \quad \mathbf{a} = \{a_i\}, \tag{2.187}$$ {34003}

where the third equality follows from the associative properties of matrix multiplication. This expression shows that a *representation of a vector in an incomplete orthonormal set produces a resulting approximation which is a simple linear combination of the elements of the correct values* (i.e., a weighted average, or "filtered" version of them). Column i of $\mathbf{G}_K\mathbf{G}_K^T$ produces the weighted linear combination of the true elements of $\mathbf{f}$ which will appear as $\tilde{f}_i$.

Because the columns of $\mathbf{G}_K$ are orthonormal, $\mathbf{G}_K^T\mathbf{G}_K = \mathbf{I}_K$, that is, the $K \times K$ identity matrix; but $\mathbf{G}_K\mathbf{G}_K^T \neq \mathbf{I}_L$ unless $K = L$. (That $\mathbf{G}_L\mathbf{G}_L^T = \mathbf{I}_L$ for $K = L$ follows from the theorem for *square* matrices that shows a left inverse is also a right inverse.) If $K < L$, $\mathbf{G}_K$ is "semi-orthogonal." If $K = L$, it is "orthogonal"; in this case, $\mathbf{G}_L^{-1} = \mathbf{G}_L^T$. If it is only semi-orthogonal, $\mathbf{G}_K^T$ is a left inverse, but not a right inverse. Any orthogonal matrix has the property that its transpose is identical to its inverse.

$\mathbf{G}_K\mathbf{G}_K^T$ is known as a "resolution matrix," with a simple interpretation. Suppose the true value of $\mathbf{f}$ were $\mathbf{f}_{j_0} = [\,0\;0\;0\;\ldots\;0\;1\;0\;.\;0\;..\;0\,]^T$, that is, a Kronecker delta δ_{jj_0}, with unity in element j_0 and zero otherwise. Then the incomplete expansion (2.186) or (2.187) would not reproduce the delta function but,

{34004}
$$\tilde{\mathbf{f}}_{j_0} = \mathbf{G}_K\mathbf{G}_K^T\mathbf{f}_{j_0}\,, \tag{2.188}$$

which is column j_0 of $\mathbf{G}_K\mathbf{G}_K^T$. Each column (or row) of the resolution matrix tells one what the corresponding form of the approximating vector would be, if its true form were a Kronecker delta.

To form a Kronecker delta requires a complete set of vectors. An analogous elementary result of Fourier analysis shows that a Dirac delta function demands contributions from all frequencies to represent a narrow, very high pulse. Removal of some of the requisite vectors (sinusoids) produces peak broadening and sidelobes. Here, depending upon the precise structure of the $\mathbf{g}_i$, the broadening and sidelobes can be complicated. If one is lucky, the effect could be a simple broadening (schematically shown in figure 2.9) without distant sidelobes), leading to the tidy interpretation of the result as a local average of the true values, called "compact resolution."[40]

A resolution matrix has the property,

{34005}
$$\text{trace}(\mathbf{G}_K\mathbf{G}_K^T) = K\,, \tag{2.189}$$

which follows from noting that,

$$\text{trace}\left(\mathbf{G}_K\mathbf{G}_K^T\right) = \text{trace}(\mathbf{G}_K^T\mathbf{G}_K) = \text{trace}(\mathbf{I}_K) = K.$$

2.5.2 Square-Symmetric Problem. Eigenvalues/Eigenvectors

Orthogonal vector expansions are particularly simple to use and interpret, but might seem irrelevant to dealing with simultaneous equations where neither the row nor column vectors of

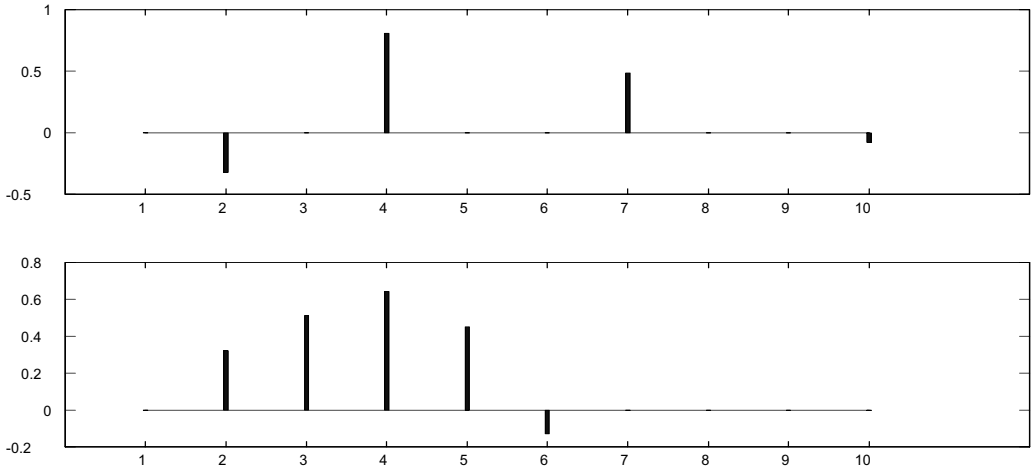


Figure 2.9: Example of a row, j_0, of a 10×10 resolution matrix, perhaps the fourth one, showing widely distributed averaging in forming $\mathbf{f}_{j0}$ (upper panel). Lower panel shows so-called compact resolution, in which the solution e.g., is a readily interpreted local average of the true solution. Such situations are not common.

{fig3_9.eps}

the coefficient matrix are so simply related. What we will show, however, is that we can always find sets of orthonormal vectors to greatly simplify the job of solving simultaneous equations. To do so, we digress to recall the basic elements of the "eigenvector/eigenvalue problem" mentioned in passing on P. 24.

Consider a *square*, $M \times M$ matrix $\mathbf{E}$ and the simultaneous equations,

$$\mathbf{E}\mathbf{g}_i = \lambda_i \mathbf{g}_i, \quad 1 \le i \le M\,, \tag{2.190}$$

{34006}

that is, the problem of finding a set of vectors $\mathbf{g}_i$ whose dot products with the rows of $\mathbf{E}$ are proportional to themselves. Such vectors are "eigenvectors," and the constants of proportionality are the "eigenvalues." Under special circumstances, the eigenvectors form an orthonormal spanning set: *Textbooks show that if* $\mathbf{E}$ *is square and symmetric, such a result is guaranteed.* It is easy to see that if two λ_j, λ_k are distinct, then the corresponding eigenvectors are orthogonal:

$$\mathbf{E}\mathbf{g}_j = \lambda_j \mathbf{g}_j, \tag{2.191}$$

$$\mathbf{E}\mathbf{g}_k = \lambda_k \mathbf{g}_k \tag{2.192}$$

Left multiply the first of these by $\mathbf{g}_k^T$, and the second by $\mathbf{g}_j^T$ and subtract:

$$\mathbf{g}_k^T \mathbf{E}\mathbf{g}_j - \mathbf{g}_j^T \mathbf{E}\mathbf{g}_k = (\lambda_j - \lambda_k)\, \mathbf{g}_k^T \mathbf{g}_j. \tag{2.193}$$

But because $\mathbf{E} = \mathbf{E}^T$, the left-hand side vanishes, and hence $\mathbf{g}_k^T \mathbf{g}_j$ by the assumption $\lambda_j \neq \lambda_k$. A similar construction proves that the λ_i are all real, and an elaboration shows that for coincident λ_i, the corresponding eigenvectors can always be made orthogonal.

Example

To contrast with the above result, consider the non-symmetric, square matrix,

$$\left\{ \begin{array}{ccc} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{array} \right\}.$$

Solution to the eigenvector/eigenvalue problem produces $\lambda_i = 1$, *and* $\mathbf{u}_i = [1, 0, 0]^T$, $1 \leq i \leq 3$. T*he eigenvectors are not orthogonal, and are certainly not a spanning set. On the other hand, the eigenvector/eigenvalues of,*

$$\left\{ \begin{array}{ccc} 1 & -1 & -2 \\ -1 & 2 & -1 \\ 1.5 & 2 & -2.5 \end{array} \right\}$$

are,

$$\begin{aligned} \mathbf{u}_1 &= \begin{bmatrix} -0.29 + 0.47i \\ -0.17 + 0.25i \\ 0.19 + 0.61i \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -0.29 - 0.47i \\ -0.17 - 0.25i \\ 0.19 - 0.61i \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} -0.72 \\ 0.90 \\ 0.14 \end{bmatrix}, \\ \lambda_j &= [-1.07 + 1.74i, -1.07 - 1.74\,i, 2.64] \end{aligned}$$

(rounded) are not orthogonal, but are a spanning set. The complex eigenvalues/eigenvectors appear in complex conjugate pairs and in some contexts are called "principal oscillation patterns" (POPs).

Suppose for the moment that we have the square, symmetric, special case, and recall how eigenvectors can be used to solve (2.16). By convention, the pairs $(\lambda_i, \mathbf{g}_i)$ are ordered in the sense of decreasing λ_i. If some λ_i are repeated, an arbitrary order choice is made.

With an orthonormal, spanning set, both the known $\mathbf{y}$ and the unknown $\mathbf{x}$ can be written

as,

{34007a}
$$\mathbf{x} = \sum_{i=1}^{M} \alpha_i \mathbf{g}_i, \quad \alpha_i = \mathbf{g}_i^T \mathbf{x}\,, \tag{2.194}$$

{34007b}
$$\mathbf{y} = \sum_{i=1}^{M} \beta_i \mathbf{g}_i, \quad \beta_i = \mathbf{g}_i^T \mathbf{y}\,. \tag{2.195}$$

By convention, $\mathbf{y}$ is known, and therefore β_i can be regarded as given. If the α_i could be found, $\mathbf{x}$ would be known.

Substitute (2.194) into (2.16),

$$\mathbf{E} \sum_{i=1}^{M} \alpha_i \mathbf{g}_i = \sum_{i=1}^{M} \left(\mathbf{g}_i^T \mathbf{y}\right) \mathbf{g}_i, \tag{2.196}$$

or, using the eigenvector property,

{34008}
$$\sum_{i=1}^{M} \alpha_i \lambda_i \mathbf{g}_i = \sum_{i}^{M} \left(\mathbf{g}_i^T \mathbf{y}\right) \mathbf{g}_i. \tag{2.197}$$

But the expansion vectors are orthonormal and so

{34009a}
$$\lambda_i \alpha_i = \mathbf{g}_i^T \mathbf{y}, \tag{2.198}$$

{34009b}
$$\alpha_i = \frac{\mathbf{g}_i^T \mathbf{y}}{\lambda_i}, \tag{2.199}$$

{34009c}
$$\mathbf{x} = \sum_{i=1}^{M} \frac{\mathbf{g}_i^T \mathbf{y}}{\lambda_i} \mathbf{g}_i\,. \tag{2.200}$$

Apart from an obvious difficulty if an eigenvalue vanishes, the problem is now completely solved. Define a diagonal matrix, $\mathbf{\Lambda}$, with elements, λ_i, in descending numerical value, and the matrix $\mathbf{G}$, whose columns are the corresponding $\mathbf{g}_i$ in the same order, the solution to (2.16) can be written, from (2.194), (2.198)–(2.200) as

{34010a}
$$\boldsymbol{\alpha} = \mathbf{\Lambda}^{-1} \mathbf{G}^T \mathbf{y}, \tag{2.201}$$

{34010b}
$$\mathbf{x} = \mathbf{G} \mathbf{\Lambda}^{-1} \mathbf{G}^T \mathbf{y} \tag{2.202}$$

where $\mathbf{\Lambda}^{-1} = \mathrm{diag}(1/\lambda_i)$.

Vanishing eigenvalues, $i = i_0$, cause trouble and must be considered. Let the corresponding eigenvectors be $\mathbf{g}_{i_0}$. Then any part of the solution which is proportional to such an eigenvector is "annihilated" by $\mathbf{E}$, that is, $\mathbf{g}_{i_0}$ is orthogonal to all the rows of $\mathbf{E}$. Such a result means that there is no possibility that anything in $\mathbf{y}$ could provide any information about the coefficient α_{i_0}. If $\mathbf{y}$ corresponds to a set of observations (data), then $\mathbf{E}$ represents the connection ("mapping")

between system unknowns and observations. The existence of zero eigenvalues shows that the act of observation of $\mathbf{x}$ removes certain structures in the solution which are then indeterminate. Vectors $\mathbf{g}_{i_0}$ (and there may be many of them) are said to lie in the "nullspace" of $\mathbf{E}$. Eigenvectors corresponding to non-zero eigenvalues lie in its "range." The simplest example is given by the "observations,"

$$x_1 + x_2 = 3\,,$$
$$x_1 + x_2 = 3\,.$$

Any structure in $\mathbf{x}$ such that $x_1 = -x_2$ is destroyed by this observation, and by inspection, the nullspace vector must be $\mathbf{g}_2 = [1, -1]^T/\sqrt{2}$. (The purpose of showing the observation twice is to produce an $\mathbf{E}$ which is square.)

Suppose there are $K < M$ non-zero λ_i. Then for $i > K$, Eq. (2.198) is

{34011}
$$0\alpha_i = \mathbf{g}_i^T\mathbf{y}, \quad K+1 \le i \le M\,, \tag{2.203}$$

and two cases must be distinguished.

Case (1):

{34012}
$$\mathbf{g}_i^T\mathbf{y} = 0\,, \quad K+1 \le i \le M\,. \tag{2.204}$$

We could then put $\alpha_i = 0$, $K+1 \le i \le M$, and the solution can be written

{34013}
$$\tilde{\mathbf{x}} = \sum_{i=1}^{K} \frac{\mathbf{g}_i^T\mathbf{y}}{\lambda_i}\mathbf{g}_i, \tag{2.205}$$

and $\mathbf{E}\tilde{\mathbf{x}} = \mathbf{y}$, *exactly*. We have put a tilde over $\mathbf{x}$ because a solution of the form,

{34014}
$$\tilde{\mathbf{x}} = \sum_{i=1}^{K} \frac{\mathbf{g}_i^T\mathbf{y}}{\lambda_i}\mathbf{g}_i + \sum_{i=K+1}^{M} \alpha_i\mathbf{g}_i\,, \tag{2.206}$$

with the remaining α_i taking on arbitrary values also satisfies the equations exactly. That is, the true value of $\mathbf{x}$ *could* contain structures proportional to the nullspace vectors of $\mathbf{E}$, but the equations (2.16) neither require their presence, nor provide information necessary to determine their amplitudes. We thus have a situation with a "solution nullspace." Define the matrix $\mathbf{G}_K$ to be $M \times K$, carrying only the first K of the $\mathbf{g}_i$, that is the range vectors, $\mathbf{\Lambda}_K$ to be $K \times K$ with only the first K, non-zero eigenvalues, and the columns of $\mathbf{Q}_G$ are the $M - K$ nullspace vectors (it is $M \times (M-K)$), then the solutions (2.205) and (2.206) are,

$$\tilde{\mathbf{x}} \;=\; \mathbf{G}_K\mathbf{\Lambda}_K^{-1}\mathbf{G}_K^T\mathbf{y}\,, \tag{2.207}$$
$$\tilde{\mathbf{x}} \;=\; \mathbf{G}_K\mathbf{\Lambda}_K^{-1}\mathbf{G}_K^T\mathbf{y} + \mathbf{Q}_G\boldsymbol{\alpha}_G, \tag{2.208}$$

where $\boldsymbol{\alpha}_G$ is the vector of unknown nullspace coefficients, respectively. Eq. (2.204) is often known as a "solvability condition." The solution in (2.207), with no nullspace contribution will be called the "particular" solution.

If $\mathbf{G}$ is written as a partitioned matrix,

$$\mathbf{G} = \{\mathbf{G}_K \quad \mathbf{Q}_G\}\,,$$

it follows from the column orthonormality that

$$\mathbf{G}\mathbf{G}^T = \mathbf{I} = \mathbf{G}_K\mathbf{G}_K^T + \mathbf{Q}_G\mathbf{Q}_G^T \tag{2.209}$$ {34016a}

or

$$\mathbf{Q}_G\mathbf{Q}_G^T = \mathbf{I} - \mathbf{G}_K\mathbf{G}_K^T\,. \tag{2.210}$$ {34016b}

Vectors $\mathbf{Q}_G$ span the nullspace of $\mathbf{G}$.

Case (2):

$$\mathbf{g}_i^T\mathbf{y} \neq 0\,, \quad i > K+1\,, \tag{2.211}$$ {34017}

for one or more of the nullspace vectors. In this case, Eq. (2.198) is the contradiction,

$$0\alpha_i \neq 0,$$

and Eq. (2.197) is actually,

$$\sum_{i=1}^{K}\lambda_i\alpha_i\mathbf{g}_i = \sum_{i=1}^{M}(\mathbf{g}_i^T\mathbf{y})\mathbf{g}_i, \quad K < M\,, \tag{2.212}$$ {34018}

that is, with differing upper limits on the sums. Owing to the orthonormality of the $\mathbf{g}_i$, there is no choice of α_i, $1 \leq i \leq K$ on the left which can match the last $M-K$ terms on the right. Evidently there is no solution in the conventional sense unless (2.204) is satisfied, hence the name "solvability condition.." What is the best we might do? Define "best" to mean that the solution $\tilde{\mathbf{x}}$ should be chosen such that,

$$\mathbf{E}\tilde{\mathbf{x}} = \tilde{\mathbf{y}},$$

where the difference, $\tilde{\mathbf{n}} = \mathbf{y} - \tilde{\mathbf{y}}$, which we call the "residual," should be as small as possible (in the l_2 norm). If this choice is made, then the orthogonality of the $\mathbf{g}_i$ shows immediately that the best choice is still (2.199), $1 \leq i \leq K$. No choice of nullspace vector coefficients, nor any other value of the coefficients of the range vectors, can reduce the norm of $\tilde{\mathbf{n}}$. The best solution is then also (2.205) or (2.207).

In this situation, we are no longer solving the equations (2.16), but rather are dealing with a set that could be written,

$$\mathbf{E}\mathbf{x} \sim \mathbf{y}, \tag{2.213}$$ {34019}

where the demand is for a solution that is the "best possible," in the sense just defined. Such statements of approximation are awkward, and it is more useful to always rewrite (2.213) as,

$$\mathbf{E}\mathbf{x} + \mathbf{n} = \mathbf{y}, \tag{2.214}$$ {34020}

where $\mathbf{n}$ is the residual. If $\tilde{\mathbf{x}}$ is given by (2.206) then,

$$\tilde{\mathbf{n}} = \sum_{i=K+1}^{M} (\mathbf{g}_i^T \mathbf{y}) \mathbf{g}_i, \tag{2.215}$$ {34021}

by (2.212). Notice that $\tilde{\mathbf{n}}^T \tilde{\mathbf{y}} = \mathbf{0} : \tilde{\mathbf{y}}$ is orthogonal to the residuals.

Example *Let*

$$\begin{aligned} x_1 + x_2 &= 1, \\ x_1 + x_2 &= 3. \end{aligned}$$

Then using $\lambda_1 = 2,\ \mathbf{g}_1 = [1,1]^T/\sqrt{2},\ \lambda_2 = 0,\ \mathbf{g}_2 = [1,-1]^T/\sqrt{2},$ *one has* $\tilde{\mathbf{x}} = [1/2, 1/2]^T \propto \mathbf{g}_1$, $\tilde{\mathbf{y}} = [2,2]^T \propto \mathbf{g}_1,\ \tilde{\mathbf{n}} = [-1,1]^T \propto \mathbf{g}_2$.

This outcome, where M-equations in M-unknowns were found in practice not to be able to determine some solution structures, is labeled "formally just-determined." The expression "formally" alludes to the fact that the appearance of a just-determined system did not mean that the characterization was true in practice. One or more vanishing eigenvalues mean that neither the rows nor columns of $\mathbf{E}$ are spanning sets.

Some decision has to be made about the coefficients of the nullspace vectors in (2.208). The form could be used as it stands, regarding it at as the "general solution." The analogy with the solution of differential equations should be apparent—typically, there is a particular solution and a homogeneous solution—here the nullspace vectors. When solving a differential equation, determination of the magnitude of the homogeneous solution requires additional information, often provided by boundary or initial conditions; here additional information is also necessary, but missing.

Despite the presence of indeterminate elements in the solution, a great deal is known about them: They are proportional to the nullspace vectors. Depending upon the specific situation, we might conceivably be in a position to obtain more observations, and would seriously consider observational strategies directed at observing these missing structures. The reader is also reminded of the discussion of the Neumann problem in Chapter 1.

Another approach is to define a "simplest" solution, appealing to what is usually known as "Ockham's Razor," or the "principle of parsimony," that in choosing between multiple explanations of a given phenomenon, the simplest one is usually the best. What is "simplest" can

be debated, but here there is a compelling choice: The solution (2.207), that is without any nullspace contributions, is less structured than any other solution. (It is often, but not always true that the nullspace vectors are more "wiggly" than those in the range. The nullspace of the Neumann problem is a counter example. In any case, including any vector not required by the data is arguably producing more structure than is required.) Setting all the unknown α_i to zero is thus one plausible choice. It follows from the orthogonality of the $\mathbf{g}_i$ that this particular solution is also the one of minimum solution norm. Later, other choices for the nullspace vectors will be made.

If the nullspace vector contributions are set to zero, the true solution has been expanded in an incomplete set of orthonormal vectors. Thus, $\mathbf{G}_K\mathbf{G}_K^T$ is the resolution matrix, and the relationship between the true solution and the minimal one is just

$$\tilde{\mathbf{x}} = \mathbf{G}_K\mathbf{G}_K^T\mathbf{x} = \mathbf{x} - \mathbf{Q}_G\boldsymbol{\alpha}_G, \quad \tilde{\mathbf{y}} = \mathbf{G}_K\mathbf{G}_K^T\mathbf{y}, \quad \tilde{\mathbf{n}} = \mathbf{Q}_G\mathbf{Q}_G^T\mathbf{y}\,. \tag{2.216}$$

{34022}

{pagesqsymm}

These results are so important, we recapitulate them: (2.206) or (2.208) is the general solution. There are three vectors involved, one of them, $\mathbf{y}$, known, and two of them, $\mathbf{x}$, $\mathbf{n}$, unknown. Because of the assumption that $\mathbf{E}$ has a complete orthonormal set of eigenvectors, all three of these vectors can be expanded, exactly, as,

$$\mathbf{x} = \sum_{i=1}^{M}\alpha_i\mathbf{g}_i, \quad \mathbf{n} = \sum_{i=1}^{M}\gamma_i\mathbf{g}_i, \quad \mathbf{y} = \sum_{i=1}^{M}(\mathbf{y}^T\mathbf{g}_i)\mathbf{g}_i\,. \tag{2.217}$$

{34023}

Substituting into ((2.214)), and using the eigenvector property produces,

$$\sum_{i=1}^{M}\alpha_i\mathbf{E}\mathbf{g}_i + \sum_{i=1}^{M}\gamma_i\mathbf{g}_i = \sum_{i=1}^{M}(\mathbf{y}^T\mathbf{g}_i)\mathbf{g}_i$$

or,

$$\sum_{i=1}^{K}\lambda_i\alpha_i\mathbf{g}_i + \sum_{i=1}^{M}\gamma_i\mathbf{g}_i = \sum_{i=1}^{M}(\mathbf{y}^T\mathbf{g}_i)\mathbf{g}_i\,.$$

From the orthogonality property,

$$\lambda_i\alpha_i + \gamma_i = \mathbf{y}^T\mathbf{g}_i, \quad 1 \le i \le K\,, \tag{2.218}$$

{34025a}

$$\gamma_i = \mathbf{y}^T\mathbf{g}_i, \quad K+1 \le i \le M\,. \tag{2.219}$$

{34025b}

In dealing with the first relationship, a choice is required. If we set,

$$\gamma_i = \mathbf{g}_i^T\mathbf{n} = 0\,, \quad 1 \le i \le K\,, \tag{2.220}$$

{34025c}

the residual norm is made as small as possible, by completely eliminating the range vectors from the residual. This choice is motivated by the attempt to satisfy the equations as well as possible, but is seen to have elements of arbitrariness. A decision about other possibilities depends upon knowing more about the system and will be the focus of attention later.

The relative contributions of any structure in $\mathbf{y}$, determined by the projection, $\mathbf{g}_i^T\mathbf{y}$ will depend upon the ratio $\mathbf{g}_i^T\mathbf{y}/\lambda_i$. Comparatively weak values of $\mathbf{g}_i^T\mathbf{y}$ may well be amplified by small, but non-zero, elements of λ_i. One must keep track of both $\mathbf{g}_i^T\mathbf{y}$, and $\mathbf{g}_i^T\mathbf{y}/\lambda_i$.

Before leaving this special case, note one more useful property of the eigenvector/eigenvalues. For the moment, let $\mathbf{G}$ have all its columns, containing both the range and nullspace vectors, with the nullspace vectors being last in arbitrary order. It is thus an $M \times M$ matrix. Correspondingly, let $\mathbf{\Lambda}$ contain all the eigenvalues on its diagonal, including the zero ones; it too, is $M \times M$. Then the eigenvector definition (2.190) produces,

{34026}
$$\mathbf{EG} = \mathbf{G\Lambda}\,. \tag{2.221}$$

Multiply both sides of (2.221) by $\mathbf{G}^T$:

{34027}
$$\mathbf{G}^T\mathbf{EG} = \mathbf{G}^T\mathbf{G\Lambda} = \mathbf{\Lambda}. \tag{2.222}$$

$\mathbf{G}$ is said to "diagonalize" $\mathbf{E}$. Now multiply both sides of (2.222) on the left by $\mathbf{G}$ and on the right by $\mathbf{G}^T$:

{34028}
$$\mathbf{GG}^T\mathbf{EGG}^T = \mathbf{G\Lambda G}^T. \tag{2.223}$$

Using the orthogonality of $\mathbf{G}$,

{34029}
$$\mathbf{E} = \mathbf{G\Lambda G}^T\,, \tag{2.224}$$

a useful representation of $\mathbf{E}$, consistent with its symmetry, known as the "singular value decomposition" or SVD.

Recall that $\mathbf{\Lambda}$ has zeros on the diagonal corresponding to the zero eigenvalues, and the corresponding rows and columns are entirely zero. Writing out (2.224), these zero rows and columns multiply all the nullspace vector columns of $\mathbf{G}$ by zero, and it is found that the nullspace columns of $\mathbf{G}$ can be eliminated, $\mathbf{\Lambda}$ reduced to its $K \times K$ form, and the decomposition (2.224) is still exact—in the form,

{34030}
$$\mathbf{E} = \mathbf{G}_K\mathbf{\Lambda}_K\mathbf{G}_K^T\,, \tag{2.225}$$

also known as the SVD. It is readily confirmed that the representation (decomposition) in either Eq. (2.224, or 2.225) is identical to

{svd5}
$$\mathbf{E} = \lambda_1\mathbf{g}_1\mathbf{g}_1^T + \lambda_2\mathbf{g}_2\mathbf{g}_2^T + ... + \lambda_K\mathbf{g}_K\mathbf{g}_K^T. \tag{2.226}$$

That is, a square symmetric matrix can be exactly represented by a sum of products orthonormal vectors $\mathbf{g}_i\mathbf{g}_i^T$ multiplied by a scalar, λ_i.

Example.

Consider the matrix from the last example,

$$\mathbf{E} = \left\{ \begin{array}{cc} 1 & 1 \\ & \\ 1 & 1 \end{array} \right\}.$$

We have

$$\mathbf{E} = \frac{2}{\sqrt{2}} \left[\begin{array}{c} 1 \\ \\ 1 \end{array} \right] \left[\begin{array}{cc} 1 & 1 \end{array} \right] \frac{1}{\sqrt{2}} + \frac{0}{\sqrt{2}} \left[\begin{array}{c} 1 \\ \\ -1 \end{array} \right] \left[\begin{array}{cc} 1 & -1 \end{array} \right] \frac{1}{\sqrt{2}}.$$

The simultaneous equations (2.214) are,

$$\mathbf{G}_K \mathbf{\Lambda}_K \mathbf{G}_K^T \mathbf{x} + \mathbf{n} = \mathbf{y}\,. \tag{2.227}$$ {34031}

Left multiply both sides by $\mathbf{\Lambda}_K^{-1}\mathbf{G}_K^T$ (existence of the inverse is guaranteed by the removal of the zero eigenvalues) and,

$$\mathbf{G}_K^T \mathbf{x} + \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \mathbf{n} = \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \mathbf{y}\,. \tag{2.228}$$ {34032}

But $\mathbf{G}_K^T\mathbf{x}$ are the projection of $\mathbf{x}$ onto the range vectors of $\mathbf{E}$, and $\mathbf{G}_K^T\mathbf{n}$ is the projection of the noise. We have agreed to set the latter to zero, and obtain,

$$\mathbf{G}_K^T \mathbf{x} = \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \mathbf{y}\,,$$

the dot products of the range of $\mathbf{E}$ with the solution. Hence, it must be true, since the range vectors are orthonormal, that

$$\tilde{\mathbf{x}} \equiv \mathbf{G}_K \mathbf{G}_K^T \mathbf{x} \equiv \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \mathbf{y}\,, \tag{2.229}$$ {34033a}

$$\tilde{\mathbf{y}} = \mathbf{E}\tilde{\mathbf{x}} = \mathbf{G}_K \mathbf{G}_K^T \mathbf{y}\,, \tag{2.230}$$ {34033b}

which is identical to the particular solution (2.205). The residuals are

$$\tilde{\mathbf{n}} = \mathbf{y} - \tilde{\mathbf{y}} = \mathbf{y} - \mathbf{E}\tilde{\mathbf{x}} = (\mathbf{I}_M - \mathbf{G}_K \mathbf{G}_K^T)\mathbf{y} = \mathbf{Q}_G \mathbf{Q}_G^T \mathbf{y}, \tag{2.231}$$ {34034}

with $\tilde{\mathbf{n}}^T\tilde{\mathbf{y}} = 0$. Notice that matrix $\mathbf{H}$ of Eq. (2.98) is just $\mathbf{G}_K\mathbf{G}_K^T$, and hence $(\mathbf{I} - \mathbf{H})$ is the projector of $\mathbf{y}$ onto the nullspace vectors.

The expected value of the solution (2.205) or (2.229) is,

$$\langle \tilde{\mathbf{x}} - \mathbf{x} \rangle = \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \langle \mathbf{y} \rangle - \sum_{i=1}^{N} \alpha_i \mathbf{g}_i = -\mathbf{Q}_G \boldsymbol{\alpha}_G, \tag{2.232}$$

{34036}

and so the solution is biassed unless $\boldsymbol{\alpha}_G = 0$.

The uncertainty is,

$$\begin{aligned} \mathbf{P} = D^2(\tilde{\mathbf{x}} - \mathbf{x}) &= \langle \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T (\mathbf{y}_0 + \mathbf{n} - \mathbf{y}_0)(\mathbf{y}_0 + \mathbf{n} - \mathbf{y}_0)^T \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \rangle \\ &\quad + \langle \mathbf{Q}_G \boldsymbol{\alpha}_G \boldsymbol{\alpha}_G^T \mathbf{Q}_G^T \rangle \\ &= \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \langle \mathbf{n}\mathbf{n}^T \rangle \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T + \mathbf{Q}_G \langle \boldsymbol{\alpha}_G \boldsymbol{\alpha}_G^T \rangle \mathbf{Q}_G^T \\ &= \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T \mathbf{R}_{nn} \mathbf{G}_K \mathbf{\Lambda}_K^{-1} \mathbf{G}_K^T + \mathbf{Q}_G \mathbf{R}_{\alpha\alpha} \mathbf{Q}_G^T \\ &= \mathbf{C}_{xx} + \mathbf{Q}_G \mathbf{R}_{\alpha\alpha} \mathbf{Q}_G^T , \end{aligned} \tag{2.233}$$

{34037a}

defining the second moments, $\mathbf{R}_{\alpha\alpha}$, of the coefficients of the nullspace vectors. Under the special circumstances that the residuals, $\mathbf{n}$, are white noise, with $\mathbf{R} = \sigma_n^2 \mathbf{I}$, (2.233) reduces to,

{34037b}

$$\mathbf{P} = \sigma_n^2 \mathbf{G}_K \mathbf{\Lambda}_K^{-2} \mathbf{G}_K^T + \mathbf{Q}_G \mathbf{R}_{\alpha\alpha} \mathbf{Q}_G^T . \tag{2.234}$$

Either case shows that the uncertainty of the minimal solution is made up of two distinct parts. The first part, the solution covariance, $\mathbf{C}_{xx}$, arises owing to the noise present in the observations, and generates uncertainty in the coefficients of the range vectors; the second contribution arises from the "missing" nullspace vector contribution. Either term can dominate. The magnitude of the noise term depends largely upon the ratio of the noise variance, σ_n^2, to the smallest non-zero eigenvalue, λ_K^2.

Example

Suppose

$$\begin{aligned} \mathbf{E}\mathbf{x} &= \mathbf{y}, \\ \begin{Bmatrix} 1 & 1 \\ 1 & 1 \end{Bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \end{aligned} \tag{2.235}$$

which is inconsistent and has no solution in the conventional sense. $\mathbf{E}$ *is a square symmetric matrix. Solving,*

$$\mathbf{E}\mathbf{g}_i = \lambda_i \mathbf{g}_i, \tag{2.236}$$

or

$$\begin{Bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{Bmatrix} \begin{bmatrix} g_{i1} \\ g_{i2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{2.237}$$

{eig1}

This equation requires

$$g_{i1}\begin{bmatrix}1-\lambda\\ \\ 1\end{bmatrix}+g_{i2}\begin{bmatrix}1\\ \\ 1-\lambda\end{bmatrix}=\begin{bmatrix}0\\ \\ 0\end{bmatrix}$$

or,

$$\begin{bmatrix}1-\lambda\\ \\ 1\end{bmatrix}+\frac{g_{i2}}{g_{i1}}\begin{bmatrix}1\\ \\ 1-\lambda\end{bmatrix}=\begin{bmatrix}0\\ \\ 0\end{bmatrix},$$

which is

$$\begin{aligned}\frac{g_{i2}}{g_{i1}} &= -(1-\lambda)\\ \frac{g_{i2}}{g_{i1}} &= -\frac{1}{1-\lambda}.\end{aligned}$$

Both equations are satisfied only if $\lambda = 2, 0$. *This method, which can be generalized, in effect derives the usual statement that for Eq. (2.237) to have a solution, the determinant,*

$$\begin{vmatrix}1-\lambda & 1\\ \\ 1 & 1-\lambda\end{vmatrix},.$$

must vanish. The first solution is labelled $\lambda_1 = 2$, *and substituting back in produces* $\mathbf{g}_1 = \frac{1}{\sqrt{2}}[1,1]^T$, *when given unit length. Also* $\mathbf{g}_2 = \frac{1}{\sqrt{2}}[-1,1]^T$, $\lambda_2 = 0$. *Hence,*

$$\mathbf{E} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \\ 1\end{bmatrix}2\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \\ 1\end{bmatrix}^T = \begin{bmatrix}1\\ \\ 1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}. \tag{2.238}$$

The equations have no solution in the conventional sense. There is, however, a sensible "best" solution:

$$\tilde{\mathbf{x}} = \frac{\mathbf{g}_1^T\mathbf{y}}{\lambda_1}\mathbf{g}_1 + \alpha_2\mathbf{g}_2, \tag{2.239}$$

$$= \left(\frac{4}{2\sqrt{2}}\right)\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \\ 1\end{bmatrix} + \alpha_2\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\ \\ 1\end{bmatrix} \tag{2.240}$$

$$= \begin{bmatrix}1\\ \\ 1\end{bmatrix} + \alpha_2\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\ \\ 1\end{bmatrix}. \tag{2.241}$$

Notice that

$$\mathbf{E}\tilde{\mathbf{x}} = \begin{bmatrix} 2 \\ \\ 2 \end{bmatrix} + 0 \neq \begin{bmatrix} 1 \\ \\ 3 \end{bmatrix}. \tag{2.242}$$

The solution has compromised the inconsistency. No choice of α_2 *can reduce the residual norm. The equations would more sensibly have been written*

$$\mathbf{E}\mathbf{x} + \mathbf{n} = \mathbf{y},$$

and the difference, $\mathbf{n} = \mathbf{y} - \mathbf{E}\tilde{\mathbf{x}}$ *is proportional to* g_2*. A system like (2.235) would most likely arise from measurements (if both equations are divided by 2, they represent two measurements of the average of* (x_1, x_2)*, and* $\mathbf{n}$ *would be best regarded as the noise of observation.*

Example

Suppose the same problem as in the last example is solved using Lagrange multipliers, that is, minimizing,

$$J = \mathbf{n}^T\mathbf{n} + \gamma^2\mathbf{x}^T\mathbf{x} - \mathbf{2}\boldsymbol{\mu}^T(\mathbf{y} - \mathbf{E}\mathbf{x} - \mathbf{n}).$$

Then, the normal equations are

$$\begin{aligned} \frac{1}{2}\frac{\partial J}{\partial \mathbf{x}} &= \gamma^2\mathbf{x} + \mathbf{E}^T\boldsymbol{\mu} = \mathbf{0} \\ \frac{1}{2}\frac{\partial J}{\partial \mathbf{n}} &= \mathbf{n} + \boldsymbol{\mu} = \mathbf{0} \\ \frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}} &= \mathbf{y} - \mathbf{E}\mathbf{x} - \mathbf{n} = \mathbf{0}, \end{aligned}$$

which produces,

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{E}^T\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right)^{-1}\mathbf{y} \\ &= \begin{Bmatrix} 1 & 1 \\ \\ 1 & 1 \end{Bmatrix}\left\{\left\{\begin{Bmatrix} 2 & 2 \\ \\ 2 & 2 \end{Bmatrix} + \gamma^2\begin{Bmatrix} 1 & 0 \\ \\ 0 & 1 \end{Bmatrix}\right\}^{-1}\begin{bmatrix} 1 \\ \\ 3 \end{bmatrix}. \end{aligned}$$

The limit $\gamma^2 \to \infty$ *is readily evaluated. Letting* $\gamma^2 \to 0$ *involves inverting a singular matrix. To understand what is going on, let us use,*

$$\mathbf{E} = \mathbf{G}\boldsymbol{\Lambda}\mathbf{G}^T = \mathbf{g}_1\lambda_1\mathbf{g}_1^T + 0 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} 2\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \\ 1 \end{bmatrix}^T + 0 \tag{2.243}$$

Hence,

$$\mathbf{E}\mathbf{E}^T = \mathbf{G}\boldsymbol{\Lambda}^2\mathbf{G}^T$$

Note that the full $\mathbf{G}, \mathbf{\Lambda}$ *are being used. Note also that* $\mathbf{I} = \mathbf{G}\mathbf{G}^T$. *Thus,*

$$\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right) = \left(\mathbf{G}\mathbf{\Lambda}^2\mathbf{G}^T + \mathbf{G}\left(\gamma^2\right)\mathbf{G}^T\right) = \mathbf{G}\left(\mathbf{\Lambda}^2 + \gamma^2\mathbf{I}\right)\mathbf{G}^T.$$

By inspection, the inverse of this last matrix is necessarily,

$$\left(\mathbf{E}\mathbf{E}^T + \mathbf{I}/\gamma^2\right)^{-1} = \mathbf{G}\left(\mathbf{\Lambda}^2 + \gamma^2\mathbf{I}\right)^{-1}\mathbf{G}^T.$$

But $\left(\mathbf{\Lambda}^2 + \gamma^2\mathbf{I}\right)^{-1}$ *is the inverse of a diagonal matrix,*

$$\left(\mathbf{\Lambda}^2 + \gamma^2\mathbf{I}\right)^{-1} = \text{diag}\left\{1/\left(\lambda_i^2 + \gamma^2\right)\right\}$$

Then

$$\begin{aligned}
\tilde{\mathbf{x}} &= \mathbf{E}^T\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right)^{-1}\mathbf{y} = \mathbf{G}\mathbf{\Lambda}\mathbf{G}^T\left(\mathbf{G}\,\text{diag}\left\{1/\left(\lambda_i^2 + \gamma^2\right)\right\}\mathbf{G}^T\right)\mathbf{y} \\
&= \mathbf{G}\,\text{diag}\left\{\lambda_i/\left(\lambda_i^2 + \gamma^2\right)\right\}\mathbf{G}^T\mathbf{y} \\
&= \sum_{i=1}^{K}\mathbf{g}_i\frac{\lambda_i}{\lambda_i^2 + \gamma^2}\mathbf{g}_i^T\mathbf{y} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \\1\end{bmatrix}\frac{2}{2+\gamma^2}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \\1\end{bmatrix}^T\begin{bmatrix}1\\ \\3\end{bmatrix} + 0 \\
&= \frac{4}{2+\gamma^2}\begin{bmatrix}1\\ \\1\end{bmatrix}
\end{aligned}$$

And the solution always exists as long as $\gamma^2 > 0$. *It is a tapered-down form of the solution with* $\gamma^2 = 0$ *if all* $\lambda_i \neq 0$.

$$\mathbf{n} = \begin{bmatrix}1\\ \\3\end{bmatrix} - \frac{4}{2+\gamma^2}\mathbf{E}\begin{bmatrix}1\\ \\1\end{bmatrix} = \begin{bmatrix}1\\ \\3\end{bmatrix} - \frac{4}{2+\gamma^2}\begin{bmatrix}2\\ \\2\end{bmatrix}$$

so that $\gamma^2 \to \infty$, *the solution* $\tilde{\mathbf{x}}$ *is minimized, becoming 0 and the residual is equal to* $\mathbf{y}$.

2.5.3 Arbitrary Systems

The Singular Vector Expansion and Singular Value Decomposition

It may be objected that this entire development is of little use, because most problems, including those outlined in Chapter 1, produced $\mathbf{E}$ matrices which could not be guaranteed to have complete orthonormal sets of eigenvectors. Indeed, the problems considered produce matrices which are usually non-square, and for which the eigenvector problem is not even defined.

For arbitrary *square* matrices, the question of when a complete orthonormal set of eigenvectors exists is not difficult to answer, but becomes somewhat elaborate.[41] When a square matrix of dimension N is not symmetric, one must consider cases in which there are N distinct eigenvalues and where some are repeated, and the general approach requires the so-called Jordan form. But we will next find a way to avoid these intricacies, and yet deal with sets of simultaneous equations of arbitrary dimensions, not just square ones. Although the mathematics are necessarily somewhat more complicated than is employed in solving the just-determined simultaneous linear equations using a complete orthonormal eigenvector set, this latter problem provides full analogues to all of the issues in the more general case, and the reader will probably find it helpful to refer back to this situation for insight.

Consider the possibility, suggested by the eigenvector method, of expanding the solution $\mathbf{x}$ in a set of orthonormal vectors. Eq. (2.88) involves one vector, $\mathbf{x}$, of dimension N, and two vectors, $\mathbf{y}$, $\mathbf{n}$, of dimension M. We would like to use spanning orthonormal vectors, but cannot expect, with two different vector dimensions involved, to use just one set: $\mathbf{x}$ can be expanded exactly in N, N-dimensional orthonormal vectors; and similarly, $\mathbf{y}$ and $\mathbf{n}$ can be exactly represented in M, M-dimensional orthonormal vectors. There are an infinite number of ways to select two such sets. But using the structure of $\mathbf{E}$, a particularly useful pair can be identified.

The simple development of the solutions in the square, symmetric case resulted from the theorem concerning the complete nature of the eigenvectors of such a matrix. So construct a new matrix,

{34038}
$$\mathbf{B} = \begin{Bmatrix} \mathbf{0} & \mathbf{E}^T \\ \mathbf{E} & \mathbf{0} \end{Bmatrix}, \tag{2.244}$$

which by definition is square (dimension $M+N$ by $M+N$) and symmetric. Thus, $\mathbf{B}$ satisfies the theorem just alluded to, and the eigenvalue problem,

{34039}
$$\mathbf{B}\mathbf{q}_i = \lambda_i \mathbf{q}_i \tag{2.245}$$

will give rise to $M+N$ orthonormal eigenvectors $\mathbf{q}_i$ (an orthonormal spanning set) whether or

not the λ_i are distinct or non-zero. Writing out (2.245),

{34040}
$$\begin{Bmatrix} \mathbf{0} & \mathbf{E}^T \\ \mathbf{E} & \mathbf{0} \end{Bmatrix} \begin{bmatrix} q_{1i} \\ . \\ q_{Ni} \\ q_{N+1,i} \\ . \\ q_{N+M,i} \end{bmatrix} = \lambda_i \begin{bmatrix} q_{1i} \\ . \\ q_{Ni} \\ q_{N+1,i} \\ . \\ q_{N+M,i} \end{bmatrix}, \quad 1 \le i \le M+N \tag{2.246}$$

where q_{pi} is the pth element of $\mathbf{q}_i$. Taking note of the zero matrices, (2.246) may be rewritten,

$$\mathbf{E}^T \begin{bmatrix} q_{N+1,i} \\ . \\ q_{N+M,i} \end{bmatrix} = \lambda_i \begin{bmatrix} q_{1i} \\ . \\ q_{Ni} \end{bmatrix}, \tag{2.247}$$ {34041a}

$$\mathbf{E} \begin{bmatrix} q_{1i} \\ . \\ q_{Ni} \end{bmatrix} = \lambda_i \begin{bmatrix} q_{N+1,i} \\ . \\ q_{N+M,i} \end{bmatrix}, \quad 1 \le i \le M+N \tag{2.248}$$ {34041b}

Define,

$$\mathbf{u}_i = \begin{bmatrix} q_{N+1,i} \\ . \\ q_{N+M,i} \end{bmatrix}, \qquad \mathbf{v}_i = \begin{bmatrix} q_{1i} \\ . \\ q_{Ni} \end{bmatrix}, \text{ or, } \mathbf{q}_i = \begin{bmatrix} \mathbf{v}_i \\ \mathbf{u}_i \end{bmatrix}, \tag{2.249}$$ {34042}

that is, defining the first N elements of $\mathbf{q}_i$ to be called $\mathbf{v}_i$ and the last M to be called $\mathbf{u}_i$. Then (2.247)–(2.248) are

$$\mathbf{E}\mathbf{v}_i = \lambda_i \mathbf{u}_i \tag{2.250}$$

$$\mathbf{E}^T \mathbf{u}_i = \lambda_i \mathbf{v}_i \,. \tag{2.251}$$

If (2.250) is left multiplied by $\mathbf{E}^T$, and using (2.251), one has,

$$\mathbf{E}^T\mathbf{E}\mathbf{v}_i = \lambda_i^2 \mathbf{v}_i, \quad 1 \le i \le N \tag{2.252}$$ {34044a}

Similarly, left multiplying (2.251) by $\mathbf{E}$ and using (2.250) produces,

$$\mathbf{E}\mathbf{E}^T\mathbf{u}_i = \lambda_i^2\mathbf{u}_i \quad 1 \le i \le M. \tag{2.253}$$ {34044b}

These last two equations show that the $\mathbf{u}_i$, $\mathbf{v}_i$ each separately satisfy two independent eigenvector/eigenvalue problems of the square symmetric matrices $\mathbf{E}\mathbf{E}^T$, $\mathbf{E}^T\mathbf{E}$ and they can be separately given unit norm. The λ_i come in pairs as $\pm\lambda_i$ and the convention is made that only the positive ones are retained, as the corresponding $\mathbf{u}_i, \mathbf{v}_i$ also differ at most by a minus sign, and hence are not independent of the ones retained.[42] If one of M, N is much smaller than the other, only the smaller eigenvalue/eigenvector problem needs to be solved for either of $\mathbf{u}_i$, $\mathbf{v}_i$; the other set is immediately calculated from (2.250) or (2.251). Evidently, in the limiting cases, of either a single equation or a single unknown, the eigenvalue/eigenvector problem is completely trivial, involving a pure scalar, no matter how large is the other dimension.

In going from (2.247, 2.248) to (2.252, 2.253), the range of the index i has dropped from $M + N$ to M or N. The missing "extra" equations correspond to negative λ_i and carry no independent information. By definition, $\lambda_i \ge 0$.

Example.

Consider the non-square, non-symmetric matrix,

$$\mathbf{E} = \begin{Bmatrix} 0 & 0 & 1 & -1 & 2 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \end{Bmatrix}$$

Form the larger matrix $\mathbf{B}$, *and solve the eigenvector/eigenvalue problem which produces*

$$\mathbf{Q} = \begin{Bmatrix} -0.31623 & 0.63246 & -1.1796 \times 10^{-16} & -0.63246 & 0.31623 \\ -0.63246 & -0.31623 & -2.0817 \times 10^{-16} & 0.31623 & 0.63246 \\ 0.35857 & -0.22361 & 0.80178 & -0.22361 & 0.35857 \\ -0.11952 & -0.67082 & -0.26726 & -0.67082 & -0.11952 \\ 0.59761 & 0.00000 & -0.53452 & 0.00000 & 0.59761 \end{Bmatrix}$$

$$\mathbf{S} = \left\{ \begin{array}{ccccc} -2.6458 & 0 & 0 & 0 & 0 \\ 0 & -1.4142 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.4142 & 0 \\ 0 & 0 & 0 & 0 & 2.6458 \end{array} \right\}$$

where $\mathbf{Q}$ *is the matrix whose columns are* $\mathbf{q}_i$ *and* $\mathbf{S}$ *is the diagonal matrix whose values are the corresponding eigenvalues. Note that one of the eigenvalues vanishes identically, and that the others occur in positive and negative pairs. The corresponding* $\mathbf{q}_i$ *differ only by sign changes in parts of the vectors, but they are all linearly independent. Defining a* $\mathbf{V}$ *matrix from the first two rows of* $\mathbf{Q}$,

$$\mathbf{V} = \left\{ \begin{array}{ccccc} -0.31623 & 0.63246 & 0 & -0.63246 & 0.31623 \\ -0.63246 & -0.31623 & 0 & 0.31623 & 0.63246 \end{array} \right\}$$

and again, only two of the vectors are linearly independent (the zero-vector is not physically realizeable). Similarly, the last three rows of $\mathbf{Q}$ *define a* $\mathbf{U}$ *matrix,*

$$\mathbf{U} = \left\{ \begin{array}{ccccc} 0.35857 & -0.22361 & 0.80178 & -0.22361 & 0.35857 \\ -0.11952 & -0.67082 & -0.26726 & -0.67082 & -0.11952 \\ 0.59761 & 0.00000 & -0.53452 & 0.00000 & 0.59761 \end{array} \right\}$$

in which only three columns are linearly independent. Retaining only the last two columns of $\mathbf{V}$ *and the last three of* $\mathbf{U}$, *and column normalizing each to unity, produces the singular vectors.*

The $\mathbf{u}_i$, $\mathbf{v}_i$ are called "singular vectors," and the λ_i are the "singular values." By convention, the λ_i are ordered in decreasing numerical value. . Equations (2.250)–(2.251) provide a relationship between each $\mathbf{u}_i$ and each $\mathbf{v}_i$. But because in general, $M \neq N$, there will be more of one set than another. The only way equations (2.250)–(2.251) can be consistent is if $\lambda_i = 0$, $i > \min(M, N)$ (where $\min(M, N)$ is read as "the minimum of M and N"). Suppose $M < N$. Then (2.253) is solved for $\mathbf{u}_i$, $1 \leq i \leq M$, and (2.250) is used to find the corresponding $\mathbf{v}_i$. There are $N - M$ $\mathbf{v}_i$ not generated this way, but which can be found using the Gram-Schmidt method described on page 20.

Let there be K non-zero λ_i; then

$$\mathbf{E}\mathbf{v}_i \neq 0\,, \quad 1 \leq i \leq K\,. \tag{2.254}$$ {34045a}

These $\mathbf{v}_i$ are known as the "range vectors of $\mathbf{E}$" or the "solution range vectors." For the remaining $N-K$ vectors $\mathbf{v}_i$,

{34045b}
$$\mathbf{E}\mathbf{v}_i = 0\,, \quad K+1 \le i \le N\,, \tag{2.255}$$

known as the "nullspace vectors of $\mathbf{E}$" or the "nullspace of the solution." If $K < M$, there will be K of the $\mathbf{u}_i$ such that,

{34046a}
$$\mathbf{E}^T\mathbf{u}_i \neq 0\,, \quad 1 \le i \le K\,, \tag{2.256}$$

which are the "range vectors of $\mathbf{E}^T$" and $M-K$ of the $\mathbf{u}_i$ such that

{34046b}
$$\mathbf{E}^T\mathbf{u}_i = 0\,, \quad K+1 \le i \le M\,, \tag{2.257}$$

the "nullspace vectors of $\mathbf{E}^T$" or the "data, or observation, nullspace vectors." The "nullspace" of $\mathbf{E}$ is spanned by its nullspace vectors, the "range" of $\mathbf{E}$ is spanned by the range vectors, etc., in the sense, for example, that an arbitrary vector lying in the range is perfectly described by a sum of the range vectors. We now have two complete orthonormal sets in the two different spaces. Note that (2.255, 2.257) imply that,

$$\mathbf{E}\mathbf{v}_i = 0,\ \mathbf{u}_i^T\mathbf{E} = 0, \quad K+1 \le i \le N, \tag{2.258}$$

expressing hard relationships among the columns and rows of $\mathbf{E}$.

Because the $\mathbf{u}_i$, $\mathbf{v}_i$ are complete in their corresponding spaces, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{n}$ can be expanded without error:

{34047}
$$\mathbf{x} = \sum_{i=1}^{N} \alpha_i \mathbf{v}_i\,, \quad \mathbf{y} = \sum_{j=1}^{M} \beta_i \mathbf{u}_i\,, \quad \mathbf{n} = \sum_{i=1}^{M} \gamma_i \mathbf{u}_i\,, \tag{2.259}$$

where $\mathbf{y}$ has been measured, so that we know $\beta_j = \mathbf{u}_j^T\mathbf{y}$. To find $\mathbf{x}$, we need α_i, and to find $\mathbf{n}$, we need the γ_i. Substitute (2.259) into the equations (2.88), and using (2.250)–(2.251),

$$\begin{aligned} \sum_{i=1}^{N} \alpha_i \mathbf{E}\mathbf{v}_i + \sum_{i=1}^{M} \gamma_i \mathbf{u}_i \;&=\; \sum_{i=1}^{K} \alpha_i \lambda_i \mathbf{u}_i + \sum_{i=1}^{M} \gamma_i \mathbf{u}_i \\ &=\; \sum_{i=1}^{M} (\mathbf{u}_i^T\mathbf{y})\mathbf{u}_i\,. \end{aligned} \tag{2.260}$$

Notice the differing upper limits on the summations. Because of the orthonormality of the singular vectors, (2.260) can be solved as,

{34049a}
$$\alpha_i\lambda_i + \gamma_i = \mathbf{u}_i^T\mathbf{y}\,, \quad i = 1 \text{ to } M\,, \tag{2.261}$$

{34049b}
$$\alpha_i = (\mathbf{u}_i^T\mathbf{y} - \gamma_i)/\lambda_i\,, \quad \lambda_i \neq 0\,,\ 1 \le i \le K\,. \tag{2.262}$$

In these equations, if $\lambda_i \neq 0$, nothing prevents setting $\gamma_i = 0$, that is,

{34049c}
$$\mathbf{u}_i^T \mathbf{n} = 0\,, \quad 1 \leq i \leq K, \tag{2.263}$$

should we wish, and which will have the effect of making the noise norm as small as possible (there is arbitrariness in this choice, and later we will choose γ_i differently). Then (2.262) produces,

$$\alpha_i = \frac{\mathbf{u}_i^T \mathbf{y}}{\lambda_i}\,, \quad 1 \leq i \leq K\,. \tag{2.264}$$ {34050}

But, because $\lambda_i = 0$, $i > K$, the only solution to (2.261) for these values of i is $\gamma_i = \mathbf{u}_i^T \mathbf{y}$, and α_i is indeterminate. These γ_i are non-zero, except in the event (unlikely with real data) that,

$$\mathbf{u}_i^T \mathbf{y} = 0\,, \quad K+1 \leq i \leq N\,. \tag{2.265}$$ {34051}

This last equation is a solvability condition— in direct analogy to (2.204).

The solution obtained in this manner now has the following form:

$$\tilde{\mathbf{x}} = \sum_{i=1}^{K} \frac{\mathbf{u}_i^T \mathbf{y}}{\lambda_i} \mathbf{v}_i + \sum_{i=K+1}^{N} \alpha_i \mathbf{v}_i \tag{2.266}$$ {34052a}

$$\tilde{\mathbf{y}} = \mathbf{E}\tilde{\mathbf{x}} = \sum_{i=1}^{K} (\mathbf{u}_i^T \mathbf{y}) \mathbf{u}_i \tag{2.267}$$ {34052b}

$$\tilde{\mathbf{n}} = \sum_{i=K+1}^{M} (\mathbf{u}_i^T \mathbf{y}) \mathbf{u}_i\,. \tag{2.268}$$ {34052c}

The coefficients of the last $N - K$ of the $\mathbf{v}_i$ in Eq. (2.266), the solution nullspace vectors, are arbitrary, representing structures in the solution about which the equations provide no information. A nullspace is always present unless $K = N$. The solution residuals are directly proportional to the nullspace vectors of $\mathbf{E}^T$ and will vanish only if $K = M$, or the solvability conditions are met.

Just as in the simpler square symmetric case, no choice of the coefficients of the solution nullspace vectors can have any effect on the size of the residuals. If we choose once again to exercise Ockham's razor, and regard the simplest solution as best, then setting the nullspace coefficients to zero,

$$\tilde{\mathbf{x}} = \sum_{i=1}^{K} \frac{\mathbf{u}_i^T \mathbf{y}}{\lambda_i} \mathbf{v}_i, \tag{2.269}$$ {34053}

along with (2.268), this is the "particular-SVD solution." It minimizes the residuals, and simultaneously produces the corresponding $\tilde{\mathbf{x}}$ with the smallest norm. If $\langle \mathbf{n} \rangle = 0$, the bias of (2.269) is evidently,

$$\langle \tilde{\mathbf{x}} - \mathbf{x} \rangle = - \sum_{i=K+1}^{N} \alpha_i \mathbf{v}_i \, . \tag{2.270}$$ {34054}

The solution uncertainty is

{34055a}
$$\mathbf{P} = \sum_{i=1}^{K} \sum_{j=1}^{K} \mathbf{v}_i \frac{\mathbf{u}_i^T \mathbf{R}_{nn} \mathbf{u}_j}{\lambda_i \lambda_j} \mathbf{v}_i^T + \sum_{i=K+1}^{N} \sum_{j=K+1}^{N} \mathbf{v}_i \langle \alpha_i \alpha_j \rangle \mathbf{v}_j^T \, . \tag{2.271}$$

If the noise is white with variance σ_n^2 or, if a row-scaling matrix $\mathbf{W}^{-T/2}$ has been applied to make it so, then (2.271) becomes,

{34055b}
$$\mathbf{P} = \sum_{i=1}^{K} \frac{\sigma_n^2}{\lambda_i^2} \mathbf{v}_i \mathbf{v}_i^T + \sum_{i=K+1}^{N} \langle \alpha_i^2 \rangle \mathbf{v}_i \mathbf{v}_i^T , \tag{2.272}$$

where it was also assumed that $\langle \alpha_i \alpha_j \rangle = \langle \alpha_i^2 \rangle \delta_{ij}$ in the nullspace. The influence of very small singular values on the uncertainty is very clear: In the solution (2.266) or (2.269) there are error terms $\mathbf{u}_i^T \mathbf{y} / \lambda_i$ which are greatly magnified by small or nearly vanishing singular values, introducing large terms proportional to σ_n^2 / λ_i^2 into (2.272).

The structures dominating $\tilde{\mathbf{x}}$ are clearly a competition between the magnitudes of $\mathbf{u}_i^T \mathbf{y}$ and λ_i, given by the ratio, $\mathbf{u}_i^T \mathbf{y} / \lambda_i$. Large λ_i can suppress comparatively large projections onto $\mathbf{u}_i$, and similarly, small, but non-zero λ_i may greatly amplify comparatively modest projections. In practice,[43] one is well-advised to study the behavior of both $\mathbf{u}_i^T \mathbf{y}$, $\mathbf{u}_i^T \mathbf{y} / \lambda_i$ as a function of i to understand the nature of the solution.

The decision to omit contributions to the residuals by the range vectors of $\mathbf{E}^T$, as we did in Eqs. (2.263), (2.268) needs to be examined. Should some other choice be made, the $\tilde{\mathbf{x}}$ norm would decrease, but the residual norm would increase. Determining the desirability of such a trade-off requires understanding of the noise structure—in particular, (2.263) imposes rigid structures, and hence covariances, on the residuals.

2.5.4 The Singular Value Decomposition

The singular vectors and values have been used to provide a convenient pair of orthonormal spanning sets to solve an arbitrary set of simultaneous equations. The vectors and values have another use, however, in providing a decomposition of $\mathbf{E}$.

Define $\mathbf{\Lambda}$ as the $M \times N$ matrix whose diagonal elements are the λ_i, in order of descending values in the same order, $\mathbf{U}$ as the $M \times M$ matrix whose columns are the $\mathbf{u}_i$, $\mathbf{V}$ as the $N \times N$ matrix whose columns are the $\mathbf{v}_i$. As an example, suppose $M = 3$, $N = 4$; then

$$\mathbf{\Lambda} = \begin{Bmatrix} \lambda_i & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \end{Bmatrix} .$$

Alternatively, if $M = 4$, $N = 3$

$$\begin{Bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \\ 0 & 0 & 0 \end{Bmatrix} ,$$

therefore extending the definition of a diagonal matrix to non-square ones.

Precisely as with matrix $\mathbf{G}$ considered above, column orthonormality of $\mathbf{U}$, $\mathbf{V}$ implies that these matrices are orthogonal,

$$\mathbf{U}\mathbf{U}^T = \mathbf{I}_M , \qquad \mathbf{U}^T\mathbf{U} = \mathbf{I}_M , \tag{2.273}$$ {svd3}

$$\mathbf{V}\mathbf{V}^T = \mathbf{I}_N , \qquad \mathbf{V}^T\mathbf{V} = \mathbf{I}_N . \tag{2.274}$$ {svd4}

(It follows that $\mathbf{U}^{-1} = \mathbf{U}^T$, etc.) As with $\mathbf{G}$ above, should one or more columns of $\mathbf{U}$, $\mathbf{V}$ be deleted, the matrices will become semi-orthogonal.

The relations (2.250) to (2.253) can be written compactly as:

$$\mathbf{E}\mathbf{V} = \mathbf{U}\mathbf{\Lambda} , \qquad \mathbf{E}^T\mathbf{U} = \mathbf{V}\mathbf{\Lambda}^T , \tag{2.275}$$

$$\mathbf{E}^T\mathbf{E}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^T\mathbf{\Lambda} , \qquad \mathbf{E}\mathbf{E}^T\mathbf{U} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^T . \tag{2.276}$$

Left multiply the first of (2.275) by $\mathbf{U}^T$ and right multiply it by $\mathbf{V}^T$, and invoking Eq. (2.274),

$$\mathbf{U}^T\mathbf{E}\mathbf{V} = \mathbf{\Lambda} . \tag{2.277}$$ {34058}

So $\mathbf{U}$, $\mathbf{V}$ diagonalize $\mathbf{E}$ (with "diagonal" having the extended meaning for a rectangular matrix as defined above.)

Right multiplying the first of (2.275) by $\mathbf{V}^T$,

$$\mathbf{E} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \,. \tag{2.278}$$ {34059}

This last equation represents a product, called the "singular value decomposition" (SVD), of an arbitrary matrix, of two orthogonal matrices, $\mathbf{U}$, $\mathbf{V}$, and a usually non-square diagonal matrix, $\mathbf{\Lambda}$.

There is one further step to take. Notice that for a rectangular $\mathbf{\Lambda}$, as in the examples above, one or more rows or columns must be all zero, depending upon the shape of the matrix. In addition, if any of the $\lambda_i = 0$, $i < \min(M, N)$, the corresponding rows or columns of $\mathbf{\Lambda}$ will be all zeros. Let K be the number of non-vanishing singular values (the "rank" of $\mathbf{E}$). By inspection (multiplying it out), one finds that the last $N - K$ columns of $\mathbf{V}$ and the last $M - K$ columns of $\mathbf{U}$ are multiplied by zeros only. If these columns are dropped entirely from $\mathbf{U}$, $\mathbf{V}$ so that $\mathbf{U}$ becomes $M \times K$ and $\mathbf{V}$ becomes $N \times K$, and reducing $\mathbf{\Lambda}$ to a $K \times K$ square matrix, then the representation (2.278) remains exact, in the form,

$$\mathbf{E} = \mathbf{U}_K\mathbf{\Lambda}_K\mathbf{V}_K^T = \lambda_1\mathbf{u}_1\mathbf{v}_1^T + \lambda_2\mathbf{u}_2\mathbf{v}_2^T + ... + \lambda_K\mathbf{u}_K\mathbf{v}_K^T \,, \tag{2.279}$$ {34060}

with the subscript indicating the number of columns, where $\mathbf{U}_K$, $\mathbf{V}_K$ are then only semi-orthogonal, and $\mathbf{\Lambda}_K$ is now square. Eq. (2.279) should be compared to (2.225).[44]

The SVD solution can be obtained by direct matrix manipulation, rather than vector by vector. Consider once again finding the solution to the simultaneous equations ((2.88)), but first write $\mathbf{E}$ in its reduced SVD,

$$\mathbf{U}_K\mathbf{\Lambda}_K\mathbf{V}_K^T\mathbf{x} + \mathbf{n} = \mathbf{y} \,. \tag{2.280}$$ {34061}

Left multiplying by $\mathbf{U}_K^T$ and invoking the semi-orthogonality of $\mathbf{U}_K$ produces

$$\mathbf{\Lambda}_K\mathbf{V}_K^T\mathbf{x} + \mathbf{U}_K^T\mathbf{n} = \mathbf{U}_K^T\mathbf{y} \,. \tag{2.281}$$ {34062}

The inverse of $\mathbf{\Lambda}_K$ (square with all non-zero diagonal elements) is easily computed and,

$$\mathbf{V}_K^T\mathbf{x} + \mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\mathbf{n} = \mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\mathbf{y} \,. \tag{2.282}$$ {34063}

But $\mathbf{V}_K^T\mathbf{x}$ is the dot product of the first K of the $\mathbf{v}_i$ with the unknown $\mathbf{x}$. Eq. (2.282) thus represents statements about the relationship between dot products of the unknown vector, $\mathbf{x}$, with a set of orthonormal vectors, and therefore must represent the expansion coefficients of the solution in those vectors. If we set,

$$\mathbf{U}_K^T\mathbf{n} = 0 \,, \tag{2.283}$$ {34064}

then,

{34065}
$$\mathbf{V}_K^T\mathbf{x} = \mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\mathbf{y}\,, \tag{2.284}$$

and hence,

{34066}
$$\tilde{\mathbf{x}} = \mathbf{V}_K\mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\mathbf{y}\,, \tag{2.285}$$

identical to the solution (2.269), which the reader is urged to confirm by writing it out explicitly. As with the square symmetric case, the contribution of any structure in $\mathbf{y}$ proportional to $\mathbf{u}_i$ depends upon the ratio of the projection, $\mathbf{u}_i^T\mathbf{y}$ to λ_i. Substituting solution (2.285) into (2.280),

$$\mathbf{U}_K\mathbf{\Lambda}_K\mathbf{V}_K^T\mathbf{V}_K\mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\mathbf{y} + \mathbf{n} = \mathbf{U}_K\mathbf{U}_K^T\mathbf{y} + \mathbf{n} = \mathbf{y}$$

or

{34067}
$$\tilde{\mathbf{n}} = (\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^T)\mathbf{y}\,. \tag{2.286}$$

Let the full $\mathbf{U}$ and $\mathbf{V}$ matrices be rewritten as

{9697}
$$\mathbf{U} = \{\mathbf{U}_K \quad \mathbf{Q}_u\}\,,\ \mathbf{V} = \{\mathbf{V}_K \quad \mathbf{Q}_v\} \tag{2.287}$$

where $\mathbf{Q}_u$, $\mathbf{Q}_v$ are the matrices whose columns are the corresponding nullspace vectors. Then,

{34069a}
$$\mathbf{E}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} = \mathbf{y},\ \mathbf{E}\tilde{\mathbf{x}} = \tilde{\mathbf{y}} \tag{2.288}$$

{34069b}
$$\tilde{\mathbf{y}} = \mathbf{U}_K\mathbf{U}_K^T\mathbf{y},\ \tilde{\mathbf{n}} = \mathbf{Q}_u\mathbf{Q}_u^T\mathbf{y} = \sum_{j=K+1}^{N}\left(\mathbf{u}_j^T\mathbf{y}\right)\mathbf{u}_j \tag{2.289}$$

which is identical to (2.267). Note,

{Q1}
$$\mathbf{Q}_u\mathbf{Q}_u^T = (\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^T),\ \mathbf{Q}_v\mathbf{Q}_v^T = (\mathbf{I} - \mathbf{V}_K\mathbf{V}_K^T) \tag{2.290}$$

and which are idempotent. ($\mathbf{V}_K\mathbf{V}_K^T$ is matrix $\mathbf{H}$ of Eq. (2.98)). The two vector sets $\mathbf{Q}_u, \mathbf{Q}_v$ span the data and solution nullspaces respectively. The general solution is,

{34070}
$$\tilde{\mathbf{x}} = \mathbf{V}_K\mathbf{\Lambda}_K^{-1}\mathbf{U}_K\mathbf{y} + \mathbf{Q}_v\boldsymbol{\alpha}, \tag{2.291}$$

where $\boldsymbol{\alpha}$ is now restricted to being the vector of coefficients of the nullspace vectors.

The solution uncertainty (2.271) is,

{34072}
$$\begin{aligned}\mathbf{P} &= \mathbf{V}_K\mathbf{\Lambda}_K^{-1}\mathbf{U}_K^T\left\langle\mathbf{n}\mathbf{n}^T\right\rangle\mathbf{U}_K\mathbf{\Lambda}_K^{-1}\mathbf{V}_K^T \\ &\quad + \mathbf{Q}_v\left\langle\alpha\alpha^T\right\rangle\mathbf{Q}_G^T = \mathbf{C}_{xx} + \mathbf{Q}_v\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^T\right\rangle\mathbf{Q}_v^T\end{aligned} \tag{2.292}$$

or,

$$\mathbf{P} = \sigma_n^2\mathbf{V}_K\mathbf{\Lambda}_K^{-2}\mathbf{V}_K^T + \mathbf{Q}_v\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^T\right\rangle\mathbf{Q}_v^T \tag{2.293}$$

for white noise.

Least-squares solution of simultaneous solutions by SVD has several important advantages. Among other features, we can write down within one algebraic formulation the solution to systems of equations which can be under-, over-, or just-determined. Unlike the eigenvalue/eigenvector solution for an arbitrary square system, the singular values (eigenvalues) are always non-negative and real, and the singular vectors (eigenvectors) can always be made a complete orthonormal set. Furthermore, the relations (2.250) or (2.275) provide a specific, quantitative statement of the connection between a set of orthonormal structures in the data, and the corresponding presence of orthonormal structures in the solution. These relations provide a very powerful diagnostic method for understanding precisely why the solution takes on the form it does.

2.5.5 Some Simple Examples. Algebraic Equations.

Example

The simplest underdetermined system is 1×2. *Suppose* $x_1 - 2x_2 = 3$, *so that*

$$\mathbf{E} = \left\{ \begin{array}{cc} 1 & -2 \end{array} \right\}, \quad \mathbf{U} = \{1\}, \quad \mathbf{V} = \left\{ \begin{array}{cc} .447 & -.894 \\ -.894 & -.447 \end{array} \right\}, \quad \lambda_1 = 2.23,$$

where the second colum of V *is the nullspace of* E. *The general solution is* $\tilde{x} = [0.6, -1.2]^T + \alpha_2 v_2$. *Because* $K = 1$ *is the only possible choice, this solution satisfies the equation exactly, and a data nullspace is not possible.*

Example

The most elementary overdetermined problem is 2×1. *Suppose*

$$\begin{aligned} x_1 &= 1 \\ x_1 &= 3\,. \end{aligned}$$

The appearance of two such equations is possible if there is noise in the observations, and they are properly written as,

$$\begin{aligned} x_1 + n_1 &= 1 \\ x_1 + n_2 &= 3\,. \end{aligned}$$

$\mathbf{E} = \{1, 1\}^T$, $\mathbf{E}^T\mathbf{E}$ *represents the eigenvalue problem of the smaller dimension, again* 1×1 *and,*

$$\mathbf{U} = \left\{ \begin{array}{cc} .707 & -.707 \\ .707 & .707 \end{array} \right\}, \quad \mathbf{V} = \{1\}, \quad \lambda_1 = \sqrt{2}$$

where the second column of $\mathbf{U}$ *lies in the data nullspace, there being no solution nullspace. The general solution is* $\mathbf{x} = x_1 = 2$, *which if substituted back into the original equations produces,*

$$\mathbf{E}\tilde{\mathbf{x}} = \begin{bmatrix} 2 \\ \\ 2 \end{bmatrix} = \tilde{\mathbf{y}}\,,$$

and hence there are residuals $\tilde{\mathbf{n}} = \tilde{\mathbf{y}} - \mathbf{y} = [1, -1]^T$, a*nd which are necessarily proportional to* $\mathbf{u}_2$ *and thus orthogonal to* $\tilde{\mathbf{y}}$. *No other solution can produce a smaller* l_2 *norm residual than this one. The SVD produced a solution which compromised the contradiction between the two original equations.*

Example

The possibility of $K < M$, $K < N$ *simultaneously is also easily seen. Consider the system:*

$$\begin{Bmatrix} 1 & -2 & 1 \\ \\ 3 & 2 & 1 \\ \\ 4 & 0 & 2 \end{Bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ \\ -1 \\ \\ 2 \end{bmatrix}\,,$$

which appears superficially just-determined. But the singular values are $\lambda_1 = 5.67$, $\lambda_2 = 2.80$, $\lambda_3 = 0$. *The vanishing of the third singular value means that the row and column vectors are not linearly independent sets (not spanning sets)—indeed the third row vector is just the sum of the first two (but the third element of* $\mathbf{y}$ *is not the sum of the first two—making the equations inconsistent). Thus there are both solution and data nullspaces, which the reader might wish to find. With a vanishing singular value,* $\mathbf{E}$ *can be written exactly using only two columns of* $\mathbf{U}$, $\mathbf{V}$ *and the linear dependence of the equations is given explicitly as* $\mathbf{u}_3^T\mathbf{E} = 0$.

Example

Consider now the underdetermined system,

$$\begin{aligned} x_1 + x_2 - 2x_3 &= 1 \\ x_1 + x_2 - 2x_3 &= 2\,, \end{aligned}$$

which has no conventional solution at all, being a contradiction, and is thus simultaneously underdetermined and incompatible. If one of the coefficients is modified by a very small quantity, $|\epsilon| > 0$, *to produce,*

$$\begin{aligned} x_1 + x_2 - (2+\epsilon)x_3 &= 1\,, \\ x_1 + x_2 - 2x_3 &= 2\,, \end{aligned} \qquad (2.294) \quad \texttt{\{ex1\}}$$

not only is there a solution, there is an infinite number of them, which the reader should confirm by computing the particular SVD solution and the nullspace. Thus the slightest perturbation in the coefficients has made the system jump from one having no solution to one having an infinite number, an obviously disconcerting situation. The label for such a system is "ill-conditioned." How would we know the system is ill-conditioned? There are several indicators. First, the ratio of the two singular values is determined by ϵ. *If we set* $\epsilon = 10^{-10}$, *the two singular values are* $\lambda_1 = 3.46$, $\lambda_2 = 4.1 \times 10^{-11}$, *an immediate warning that the two equations are nearly linearly dependent. (In a mathematical problem, the non-vanishing of the second singular value is enough to assure a solution. It is the inevitable slight errors in* y *that suggest sufficiently small singular values should be treated as though they were actually zero.)*

Example

A similar problem exists with the system,

$$x_1 + x_2 - 2x_3 = 1$$
$$x_1 + x_2 - 2x_3 = 1\,,$$

which has an infinite number of solutions. But the change to

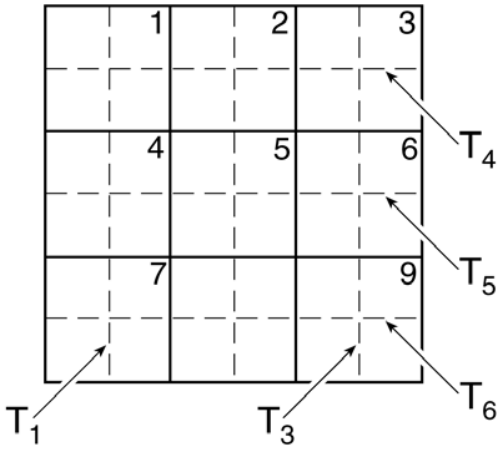
$$x_1 + x_2 - 2x_3 = 1\,,$$
$$x_1 + x_2 - 2x_3 = 1 + \epsilon$$

for arbitrarily small ϵ *produces a system with no solutions in the conventional mathematical sense, although the SVD will handle the system in a sensible way, which the reader should confirm.*

Problems like these are simple examples of the practical issues that arise once one recognizes that unlike textbook problems, observational ones always contain inaccuracies; any discussion of how to handle data in the presence of mathematical relations must account for these inaccuracies as intrinsic—not as something to be regarded as an afterthought. But the SVD itself is sufficiently powerful that it always contains the information to warn of ill-conditioning, and by determination of K to cope with it—producing useful solutions.

Example

The Tomographic Problem from Chapter 1. A square box, is made up of 3×3 *unit dimension*



{tomog3.tif} Figure 2.10: Tomographic problem with 9-unknowns and only 6-integral constraints.

sub-boxes (Fig. 2.10). All rays are in the r_x or r_y directions. So the equations are,

$$\left\{\begin{array}{ccccccccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{array}\right\} \left[\begin{array}{c} x_1 \\ x_2 \\ . \\ . \\ . \\ x_9 \end{array}\right] = \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{array}\right],$$

that is, $\mathbf{Ex} = \mathbf{y}$. *There are six integrals (rays) across the nine boxes in which one seeks the corresponding value of* x_i. $\mathbf{y}$ *was calculated by assuming that the "true" value is* $x_5 = 1$, $x_i = 0$,

$i \neq 5$. *The SVD produces,*

$$\mathbf{U} = \left\{ \begin{array}{cccccc} -0.408 & 0 & 0 & 0.816 & 0 & 0.408 \\ -0.408 & 0.703 & -0.0543 & -0.408 & -0.0549 & 0.408 \\ -0.408 & -0.703 & 0.0543 & -0.408 & 0.0549 & 0.408 \\ -0.408 & -0.0566 & 0.0858 & 0 & -0.81 & -0.408 \\ -0.408 & -0.0313 & -0.744 & 0 & 0.335 & -0.408 \\ -0.408 & 0.0879 & 0.658 & 0 & 0.475 & -0.408 \end{array} \right\},$$

$$\mathbf{\Lambda} = \mathrm{diag}\left(\left[\begin{array}{cccccc} 2.45 & 1.73 & 1.73 & 1.73 & 1.73 & 0 \end{array} \right] \right),$$

$$\mathbf{V} = \left\{ \begin{array}{ccccccccc} -0.333 & -0.0327 & 0.0495 & 0.471 & -0.468 & -0.38 & -0.224 & 0.353 & 0.353 \\ -0.333 & 0.373 & 0.0182 & -0.236 & -0.499 & 0.432 & 0.302 & -0.275 & 0.302 \\ -0.333 & -0.438 & 0.0808 & -0.236 & -0.436 & -0.0515 & -0.0781 & -0.0781 & -0.655 \\ -0.333 & -0.0181 & -0.43 & 0.471 & 0.193 & 0.519 & -0.361 & -0.15 & -0.15 \\ -0.333 & 0.388 & -0.461 & -0.236 & 0.162 & -0.59 & -0.0791 & -0.29 & -0.0791 \\ -0.333 & -0.424 & -0.398 & -0.236 & 0.225 & 0.0704 & 0.44 & 0.44 & 0.229 \\ -0.333 & 0.0507 & 0.38 & 0.471 & 0.274 & -0.139 & 0.585 & -0.204 & -0.204 \\ -0.333 & 0.457 & 0.349 & -0.236 & 0.243 & 0.158 & -0.223 & 0.566 & -0.223 \\ -0.333 & -0.355 & 0.411 & -0.236 & 0.306 & -0.0189 & -0.362 & -0.362 & 0.427 \end{array} \right\}$$

The zeros appearing in $\mathbf{U}$, *and in the last element of* $\mathrm{diag}\,(\Lambda)$ *are actually very small numbers* $(\mathrm{O}(10^{-16})$ or less). *Rank* $K = 5$ *despite there being six equations—a consequence of redundancy in the integrals. Notice that there are four repeated* λ_i, *and the lack of expected simple symmetries in the corresponding* $\mathbf{v}_i$ *is a consequence of a random assignment in the eigenvectors.*

$\mathbf{u}_1$ *just averages the right hand-side values, and the corresponding solution is completely uniform, proportional to* $\mathbf{v}_1$. *The average of* $\mathbf{y}$ *is usually the most robust piece of information.*

The "right" answer is $\mathbf{x} = [0, 0, 0, 0, 1, 0, 0, 0, 0]^T$. *The rank 5 answer by SVD is* $\tilde{\mathbf{x}} =$*[-0.1111, 0.2222, -0.1111, 0.2222, 0.5556, 0.2222, -0.1111, 0.2222, -0.1111]*^{T} *which exactly satisfies the*

same equations. $\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} = 0.556 \; < \mathbf{x}^T\mathbf{x}$. *When mapped into two dimensions,* $\tilde{\mathbf{x}}$ at *rank 5 is,*

$$\begin{matrix} & r_x \rightarrow & & \\ r_y \uparrow & \begin{bmatrix} -.11 & .22 & -.11 \\ .22 & .56 & .22 \\ -.11 & .22 & -.11 \end{bmatrix} \end{matrix}, \tag{2.295}$$

and is the minimum norm solution. The mapped $\mathbf{v}_6$*, which belongs in the null space is,*

$$\begin{bmatrix} -.38 & .43 & -.05 \\ .52 & -.59 & .07 \\ -.14 & .16 & -.02 \end{bmatrix}$$

and along with any remaining null space vectors produces a zero sum along any of the ray paths. $\mathbf{u}_6$ *is in the data nullspace.* $\mathbf{u}_6^T\ \mathbf{E} = 0$ *shows that,*

$$a\,(y_1 + y_2 + y_3) - a\,(y_4 + y_5 + y_6) = 0,$$

if there is to be a solution without a residual, or alternatively, that no solution would permit this sum to be non-zero. This requirement is physically sensible, as it says that the vertical and horizontal rays cover the same territory and must therefore produce the same sum travel times. It shows why the rank is 5, and not 6.

There is no noise in the problem as stated. The correct solution and the SVD solution differ by the null space vectors. One can easily confirm that $\tilde{\mathbf{x}}$ *is column 5 of* $\mathbf{V}_5\mathbf{V}_5^T$*. Least-squares allows one to minimize (or maximize) anything one pleases. Suppose for some reason, we want the solution that minimizes the differences between the value in box 5 and its neighbors, perhaps*

as a way of finding a "smooth" solution. Let

$$\mathbf{W} = \left\{ \begin{array}{ccccccccc} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right\} \tag{2.296}$$

The last row is included to render $\mathbf{W}$ *a full-rank matrix. Then*

$$\mathbf{Wx} = \begin{bmatrix} x_5 - x_1 \\ x_5 - x_2 \\ . \\ x_5 - x_9 \\ x_5 \end{bmatrix} \tag{2.297}$$

and we can minimize

$$J = \mathbf{x}^T \mathbf{W}^T \mathbf{W} \mathbf{x} \tag{2.298}$$

subject to $\mathbf{Ex} = \mathbf{y}$ *by finding the stationary value of*

$$J' = J - 2\boldsymbol{\mu}^T (\mathbf{y} - \mathbf{Ex}) \tag{2.299}$$

The normal equations are then

$$\mathbf{W}^T \mathbf{W} \mathbf{x} \; = \; \mathbf{E}^T \boldsymbol{\mu} \tag{2.300}$$

$$\mathbf{Ex} \; = \; \mathbf{y} \tag{2.301}$$

and

$$\tilde{\mathbf{x}} = \left(\mathbf{W}^T \mathbf{W} \right)^{-1} \mathbf{E}^T \boldsymbol{\mu}$$

and then,

$$\mathbf{E}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{E}^T\boldsymbol{\mu} = \mathbf{y}$$

The rank of $\mathbf{E}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{E}^T$ *is* $K = 5 < M = 6$*, and so we need a generalized inverse,*

$$\tilde{\boldsymbol{\mu}} = \left(\mathbf{E}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{E}^T\right)^{+}\mathbf{y} = \sum_{j=1}^{5}\mathbf{v}_i\frac{\mathbf{v}_i^T\mathbf{y}}{\lambda_i}$$

The null space of $\mathbf{E}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{E}^T$ *is readily confirmed to be the vector,*

$$\begin{bmatrix} -0.408 & -0.408 & -0.408 & 0.408 & 0.408 & 0.408 \end{bmatrix}^T, \tag{2.302}$$

which produces the solvability condition. Here, because $\mathbf{E}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{E}^T$ *is symmetric, the SVD reduces to the symmetric decomposition.*

Finally, the mapped $\tilde{\mathbf{x}}$ *is*

$$\begin{bmatrix} -.20 & .41 & -.20 \\ .41 & .18 & .41 \\ -.20 & .41 & -.21 \end{bmatrix}$$

and one cannot further decrease the sum-squared differences of the solution elements. One can confirm that this solution satisfies the equations. Evidently, it produces a minimum, not a maximum (it suffices to show that the eigenvalues of $\mathbf{W}^T\mathbf{W}$ *are all non-negative). The addition of any of the nullspace vectors of* $\mathbf{E}$ *to* $\tilde{\mathbf{x}}$ *will necessarily increase the value of* J *and hence there is no bounded maximum. In real tomographic problems, the arc lengths making up matrix* $\mathbf{E}$ *are three dimensional curves and depend upon the background index of refraction in the medium, which is usually itself determined from observations.*[45] *There are thus errors in* $\mathbf{E}$ *itself, rendering the problem one of non-linear estimation. Approaches to solving such problems are described in Chapter 3.*

Example

Consider, the flow into a four-sided box with missing integration constant as described in Chapter 1. Total mass conservation and conservation of dye, C_i*. Let the relative areas of each interface be 1, 2, 3, 1 units respectively. Let the corresponding velocities on each side be* $1, 1/2, -2/3, 0$ *respectively, with the minus sign indicating a flow out. That mass is conserved is confirmed by,*

{page:fourside

$$1\,(1) + 2\left(\frac{1}{2}\right) + 3\left(\frac{-2}{3}\right) + 1\,(0) = 0.$$

Now suppose that the total velocity is not in fact known, but an integration constant is missing on each interface, so that

$$1\left(\frac{1}{2}+b_1\right)+2\left(1+b_2\right)+3\left(\frac{1}{3}+b_3\right)+1\left(2+b_4\right)=0$$

where the $b_i=[1/2,-1/2,-1,-2]$*, but are here treated as unknown. Then the above equation becomes*

$$b_1+2b_2+3b_3+b_4=-5.5$$

or one equation in 4 unknowns. Evidently, one linear combination of the unknown b_i *can be determined. We would like more information. Suppose that a tracer of concentration,* $C_i=[2,1,3/2,0]$ *is measured at each side, and is believed conserved. The governing equation is*

$$1\left(\frac{1}{2}+b_1\right)2+2\left(1+b_2\right)1+3\left(\frac{1}{3}+b_3\right)\frac{3}{2}+1\left(2+b_4\right)0=0$$

or

$$2b_1+2b_2+4.5b_3+0b_4=-4.5$$

giving a system of 2 equations in four unknowns

$$\left\{\begin{array}{cccc}1&2&3&1\\2&2&4.5&0\end{array}\right\}\left[\begin{array}{c}b_1\\b_2\\b_3\\b_4\end{array}\right]=\left[\begin{array}{c}-5.5\\-4.5\end{array}\right].$$

The SVD of the coefficient matrix, $\mathbf{E}$*, is* :

$$\mathbf{E}=\left\{\begin{array}{cc}-0.582&-0.813\\0.813&0.582\end{array}\right\}\left\{\begin{array}{cccc}6.50&0&0&0\\0&1.02&0&0\end{array}\right\}\left\{\begin{array}{cccc}-.801&0.179&-0.454&0.347\\0.009&0.832&0.429&0.340\\-0.116&0.479&-0.243&-0.835\\0.581&0.215&-0.742&0.259\end{array}\right\}$$

and the remainder of the solution is left to the reader.

2.5.6 Simple Examples. Differential and Partial Differential Equations

Example

As an example of the use of this machinery with differential equations, consider,

{exponen10}
$$\frac{d^2x(r)}{dr^2} - k^2 x(r) = 0, \tag{2.303}$$

subject to initial and/or boundary condition. Using the simple one-sided, uniform discretization,

{discexponen1}
$$x((m+1)\Delta r) - \left(2 + k^2(\Delta r)^2\right) x(m\Delta r) + x((m-1)\Delta r) = 0, \tag{2.304}$$

at all interior points. Take the specific case, with two-end conditions, $x(\Delta r) = 10, x(51\Delta r) = 1, \Delta r = 0.1$, *the numerical solution is depicted in Fig. 2.11 from the direct (conventional) solution to* $\mathbf{Ax} = \mathbf{y}$. *The first two rows of* $\mathbf{A}$ *were used to impose the boundary conditions on* $x(\Delta r), x(51\Delta r)$. *The singular values of* $\mathbf{A}$ *are also plotted in Fig. 2.11. The range is over about two orders of magnitude, and there is no reason to suspect numerical difficulties. The first and last singular vectors* $\mathbf{u}_1$, $\mathbf{v}_1$, $\mathbf{u}_{51}$, $\mathbf{v}_{51}$, *are plotted too. One infers (by plotting additional such vectors), that the large singular values correspond to singular vectors showing a great deal of small-scale structure, and the smallest singular values correspond to the least structured (largest spatial scales) in both the solution and in the specific corresponding weighted averages of the equations. This result may be counterintuitive. But note that in this problem, all elements of y vanish except the first two, which are being used to set the boundary conditions. We know from the analytical solution that the true solution is large-scale; most of the information contained in the differential equation (2.303) or its numerical counterpart, (2.304) is an assertion that all small scales are absent; this information is the most robust and corresponds to the largest singular values. The remaining information, on the exact nature of the largest scales, is contained in only two of the 51 equations—given by the boundary conditions, is extremely important, but less robust than that concerning the absence of small scales. (Less "robust" is being used in the sense that small changes in the boundary conditions will lead to relatively large changes in the largescale structures in the solution because of the division by relatively small* λ_i.).

Example

Consider now the classical Neumann problem described in Chapter 1. The problem is to be solved on a 10×10 *grid as in Eq. (1.17),* $\mathbf{Ax} = \mathbf{b}$. *The singular values of* $\mathbf{A}$ *are plotted in figure 2.12; the largest one is* $\lambda_1 = 7.8$, *and the smallest non-zero one is* $\lambda_{99} = 0.08$. *As expected,* $\lambda_{100} = 0$. *The singular vector* v_{100} *corresponding to the zero singular value is a constant;* u_{100},*also shown in Fig. 2.12 is not a constant, it has considerable structure—which provides the solvability condition for the Neumann problem,* $u_{100}^T y = 0$. *The physical origin of the solvability condition is readily understood: Neumann boundary conditions prescribe boundary flux rates, and the sum of the interior source strengths plus the boundary flux rates must sum to*

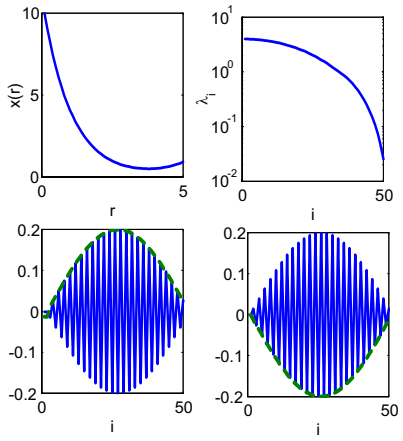


Figure 2.11: Upper left is $\tilde{\mathbf{x}}$ Eq. (2.303) by brute force from the simultaneous equations. Upper right panel displays the corresponding singular values; all are finite (there is no nullspace). Lower left panel displays $\mathbf{u}_1$ (solid curve), and $\mathbf{u}_{51}$(dashed). Lower right panel shows the corresponding $\mathbf{v}_1, \mathbf{v}_{51}$. The most robust information corresponds to the *absence* of small scales in the solution. {exponensvd.ep

zero, otherwise no steady state is possible. If the boundary conditions are homogeneous, then no flow takes place through the boundary, and the interior sources must sum to zero. In particular, the value of u_{100} on the interior grid points is a constant. The Neumann problem is thus a forward one requiring coping with both a solution nullspace and a solvability condition.

2.5.7 Relation of Least-Squares to the SVD

What is the relationship of the SVD solution to the least-squares solutions? To some extent, the answer is already obvious from the orthonormality of the two sets of singular vectors: they *are* the least-squares solution, where it exists. When does the simple least-squares solution will exist? Consider first the formally overdetermined problem, $M > N$. The solution (2.96) exists if and only if the matrix inverse exists. Substituting the SVD for $\mathbf{E}$, one finds

$$(\mathbf{E}^T\mathbf{E})^{-1} = (\mathbf{V}_N\mathbf{\Lambda}_N^T\mathbf{U}_N^T\mathbf{U}_N\mathbf{\Lambda}_N\mathbf{V}_N^T)^{-1} = (\mathbf{V}_N\mathbf{\Lambda}_N^2\mathbf{V}_N^T)^{-1}, \tag{2.305}$$ {34074}

where the semi-orthogonality of $\mathbf{U}_N$ has been used. Suppose that $K = N$, its maximum possible value; then $\mathbf{\Lambda}_N^2$ is $N \times N$ with *all non-zero diagonal elements* λ_i^2. The inverse in (2.305) may be found by inspection, using $\mathbf{V}_N^T\mathbf{V}_N = \mathbf{I}_N$,

$$(\mathbf{E}^T\mathbf{E})^{-1} = \mathbf{V}_N\mathbf{\Lambda}_N^{-2}\mathbf{V}_N^T\,. \tag{2.306}$$ {34075}

Then the solution (2.96) becomes

$$\tilde{\mathbf{x}} = (\mathbf{V}_N\mathbf{\Lambda}_N^{-2}\mathbf{V}_N^T)\mathbf{V}_N\mathbf{\Lambda}_N\mathbf{U}_N^T\mathbf{y} = \mathbf{V}_N\mathbf{\Lambda}_N^{-1}\mathbf{U}_N^T\mathbf{y}\,, \tag{2.307}$$ {34076}

which is identical to the SVD solution (2.285). If $K < N$, $\mathbf{\Lambda}_N^2$ has at least one zero on the diagonal, no matrix inverse exists and the conventional least-squares solution is not defined. The condition for its existence is thus $K = N$, the so-called "full rank overdetermined" case. The condition $K < N$ is called "rank deficient." The dependence of the least-squares solution magnitude upon the possible presence of very small, but non-vanishing, singular values is obvious.

1. That the full-rank overdetermined case is unbiased, as previously asserted (45), can now be seen from

$$\langle\tilde{\mathbf{x}} - \mathbf{x}\rangle = \sum_{i=1}^{N}\frac{(\mathbf{u}_i^T\langle\mathbf{y}\rangle)}{\lambda_i}\mathbf{v}_i - \mathbf{x} = \sum_{i=1}^{N}\frac{\mathbf{u}_i^T\mathbf{y}_0}{\lambda_i}\,\mathbf{v}_i - \mathbf{x} = \mathbf{0}\,,$$

with $\mathbf{y} = \mathbf{y}_0 + \mathbf{n}$, if $\langle\mathbf{n}\rangle = \mathbf{0}$, assuming that the correct $\mathbf{E}$ (model) is being used. {pagemeanbias2

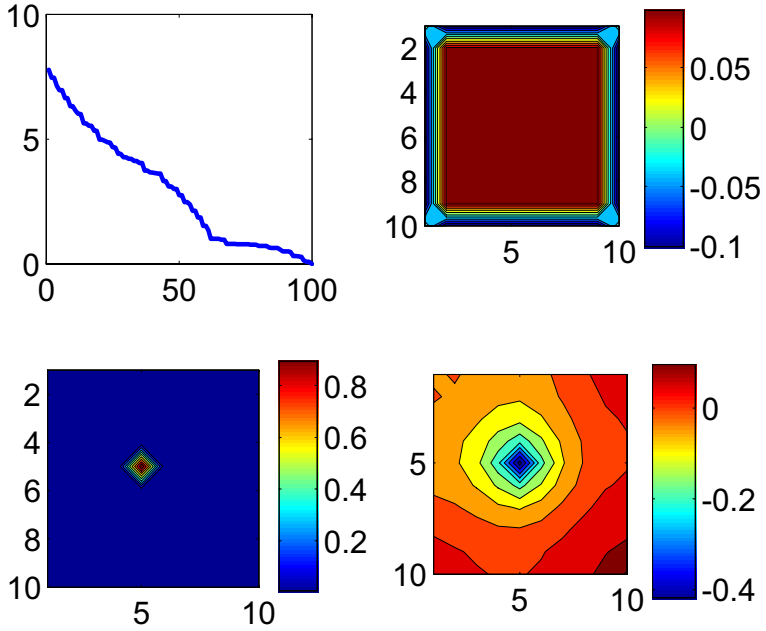


Figure 2.12: Color. (Upper left) Singular values of the coefficient matrix $\mathbf{A}$ of the numerical Neumann problem on a 10×10 grid. All λ_i are non-zero except the last one. (Upper right) $\mathbf{u}_{100,}$ the nullspace vector of $\mathbf{E}^T$ defining the solvability or consistency condition for a solution through $\mathbf{u}_{100}^T\mathbf{y}$ =0. Plotted as mapped onto the two-dimensional spatial grid (r_x, r_y) with $\Delta x = \Delta y = 1$. The interpretation is that the sum of the influx through the boundaries and from interior sources must vanish. Note that corner derivatives differ from other boundary derivatives by $1/\sqrt{2}$. Corresponding $\mathbf{v}_{100}$ is a constant, indeterminate with the information available, and not shown.. (lower left) A source $\mathbf{b}$ (a numerical delta function) is present, not satisfying the solvability condition $\mathbf{u}_{100}^T\mathbf{b}$ =0, because all boundary fluxes were set to vanishing. (Lower right) Particular SVD solution, $\tilde{\mathbf{x}}$, at rank $K = 99$. One confirms that $\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}$ is proportional to $\mathbf{u}_{100}$ as the source is otherwise inconsistent with no flux boundary conditions. With $\mathbf{b}$ a Kronecker delta function at one grid point, this solution is a numerical Green function for the Neumann problem and insulating boundary conditions. {neumann1.eps}

Now consider another problem, the conventional purely underdetermined least-squares one, whose solution is (2.166). When does that exist? Substituting the SVD,

$$\begin{aligned}\tilde{\mathbf{x}} &= \mathbf{V}_M\mathbf{\Lambda}_M\mathbf{U}_M^T(\mathbf{U}_M\mathbf{\Lambda}_M\mathbf{V}_M^T\mathbf{V}_M\mathbf{\Lambda}_M^T\mathbf{U}_M^T)^{-1}\mathbf{y} \\ &= \mathbf{V}_M\mathbf{\Lambda}_M\mathbf{U}_M^T(\mathbf{U}_M\mathbf{\Lambda}_M^2\mathbf{U}_M^T)^{-1}\mathbf{y}\,.\end{aligned} \tag{2.308}$$ {34078a}

Again, the matrix inverse exists if and only if $\mathbf{\Lambda}_M^2$ has all non-zero diagonal elements, which occurs only when $K = M$. Under that specific condition, the inverse is obtained by inspection and,

$$\tilde{\mathbf{x}} = \mathbf{V}_M\mathbf{\Lambda}_M\mathbf{U}_M^T(\mathbf{U}_M\mathbf{\Lambda}_M^{-2}\mathbf{U}_M^T)\mathbf{y} = \mathbf{V}_M\mathbf{\Lambda}_M^{-1}\mathbf{U}_M^T\mathbf{y} \tag{2.309}$$ {34078b}

$$\tilde{\mathbf{n}} = 0\,, \tag{2.310}$$ {34078c}

which is once again the particular-SVD solution (2.285)—with the nullspace coefficients set to zero. This situation is usually referred to as the "full-rank underdetermined case." Again, the possible influence of small singular values is apparent and an arbitrary sum of nullspace vectors can be added to (2.309). The bias of (2.308) is given by the nullspace elements, and its uncertainty arises only from the nullspace contribution, because with $\tilde{\mathbf{n}} = \mathbf{0}$, thenoise variance vanishes, and the particular-SVD solution covariance $\mathbf{C}_{xx}$ would be zero.

The particular-SVD solution thus coincides with the two simplest forms of least-squares solution, and generalizes both of them to the case where the matrix inverses do not exist. *All of the structure imposed by the SVD, in particular the restriction on the residuals in* (2.263)*, is present in the least-squares solution.* If the system is not of full rank, then the simple least-squares solutions do not exist. *The SVD generalizes these results* by determining what it can: the elements of the solution lying in the range of $\mathbf{E}$, and an explicit structure for the resulting nullspace vectors.

The SVD provides a lot of flexibility. For example, it permits one to modify the simplest underdetermined solution (2.166) to remove its greatest shortcoming, the necessity that $\tilde{\mathbf{n}} = \mathbf{0}$. One simply truncates the solution (2.269) at $K = K' < M$, thus assigning all vectors $\mathbf{v}_i$, $K' + 1 \le i \le K$, to an "effective nullspace" (or substitutes K' for K everywhere). The residual is then,

$$\tilde{\mathbf{n}} = \sum_{i=K'+1}^{M} (\mathbf{u}_i^T\mathbf{y})\mathbf{u}_i\,, \tag{2.311}$$ {34079}

with an uncertainty for $\tilde{\mathbf{x}}$ given by (2.292), but with the upper limit being K' rather than K. Such truncation has the effect of reducing the solution covariance contribution to the uncertainty, but increasing the contribution owing to the nullspace (and increasing the bias). In the presence

of singular values small compared to σ_n, the resulting overall reduction in uncertainty may be very great—at the expense of a possibly very small bias.

The solution now consists of three parts,

$$\tilde{\mathbf{x}} = \sum_{i=1}^{K'} \frac{\mathbf{u}_i^T \mathbf{y}}{\lambda_i} \mathbf{v}_i + \sum_{i=K'+1}^{K} \alpha_i \, \mathbf{v}_i + \sum_{i=K+1}^{N} \alpha_i \mathbf{v}_i \,, \tag{2.312}$$ {34080}

where the middle sum contains the terms appearing with singular values too small to be employed—for the given noise—and the third sum is the strict nullspace. Usually, one lumps the two nullspace sums together. The first sum, by itself, represents the particular-SVD solution in the presence of noise. Resolution and covariance matrices are modified by the substitution of K' for K.

This consideration is extremely important—it says that despite the mathematical condition $\lambda_i \neq 0$, some structures in the solution cannot be estimated with sufficient reliability to be useful. The "effective rank" is then not the same as the mathematical rank.

It was already noticed that the simplest form of least-squares does not provide a method to control the ratios of the solution and noise norms. Evidently, truncation of the SVD offers a simple way to do so—by reducing K'. It follows that the solution norm necessarily is reduced, and that the residuals must grow, along with the size of the solution nullspace. The issue of how to choose K', that is, "rank determination," in practice is an interesting one to which we will return (P. 117).

2.5.8 Pseudo-Inverses

Consider an arbitrary $M \times N$ matrix $\mathbf{E} = \mathbf{U}_K \mathbf{\Lambda}_K \mathbf{V}_K^T$ and,

$$\mathbf{E}\mathbf{x} + \mathbf{n} = \mathbf{y}$$

Then if $\mathbf{E}$ is full-rank underdetermined, the minimum norm solution is,

$$\tilde{\mathbf{x}} = \mathbf{E}^T \left(\mathbf{E}\mathbf{E}^T\right)^{-1} \mathbf{y} = \mathbf{V}_K \mathbf{\Lambda}_K^{-1} \mathbf{U}_K^T \mathbf{y}, \quad K = M,$$

and if it is full-rank overdetermined, the minimum noise solution is,

$$\tilde{\mathbf{x}} = \left(\mathbf{E}^T \mathbf{E}\right)^{-1} \mathbf{E}^T \mathbf{y} = \mathbf{V}_K \mathbf{\Lambda}_K^{-1} \mathbf{U}_K^T \mathbf{y}, \quad K = N.$$

The first of these, the Moore-Penrose, or pseudo-inverse, $\mathbf{E}_1^+ = \mathbf{E}^T \left(\mathbf{E}\mathbf{E}^T\right)^{-1}$ is sometimes also known as a "right-inverse," as $\mathbf{E}\mathbf{E}_1^+ = \mathbf{I}_M$. The second pseudo-inverse, $\mathbf{E}_2^+ = \left(\mathbf{E}^T\mathbf{E}\right)^{-1} \mathbf{E}^T$ is a "left-inverse" as $\mathbf{E}_2^+ \mathbf{E} = \mathbf{I}_N$. They can both be represented as $\mathbf{V}_K \mathbf{\Lambda}_K^{-1} \mathbf{U}_K^T$, but with differing values of K. If $K < M, N$ neither of the pseudo-inverses exists, but $\mathbf{V}_K \mathbf{\Lambda}_K^{-1} \mathbf{U}_K^T \mathbf{y}$ still provides the particular SVD solution. When $K = M = N$, one has a demonstration that the left and right inverses are identical; they are then written as $\mathbf{E}^{-1}$.

2.5.9 Row and Column Scaling

The effects on the least-squares solutions of the row and column scaling can now be understood. We discuss them in the context of noise covariances, but as always in least-squares, the weight matrices need no statistical interpretation, and can be chosen by the investigator to suit her convenience or taste.

Suppose we have two equations

$$\begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1.01 & 1 \end{Bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

and there is no information about the noise covariance and so no row scaling is reasonable: $\mathbf{W} = \mathbf{I}$. The SVD of $\mathbf{E}$ is

$$\mathbf{U} = \begin{Bmatrix} 0.7059 & -0.7083 \\ 0.7083 & 0.7059 \end{Bmatrix}, \qquad \mathbf{V} = \begin{Bmatrix} 0.5764 & -0.4096 & 0.7071 \\ 0.5793 & 0.8151 & 0.0000 \\ 0.5764 & -0.4096 & -0.7071 \end{Bmatrix},$$

$$\lambda_1 = 2.4536, \qquad \lambda_2 = .0058\,.$$

The SVD solutions, choosing ranks $K' = 1, 2$ in succession, are very nearly (the numbers having been rounded),

$$\begin{aligned} \tilde{\mathbf{x}} &\approx \left(\frac{y_1 + y_2}{2.45}\right) \begin{bmatrix} 0.58 & 0.58 & 0.58 \end{bmatrix}^T, \\ \tilde{\mathbf{x}} &\approx \left(\frac{y_1 + y_2}{2.45}\right) \begin{bmatrix} 0.58 & 0.58 & 0.58 \end{bmatrix}^T + \left(\frac{y_1 - y_2}{0.0058}\right) \begin{bmatrix} -0.41 & 0.82 & 0.41 \end{bmatrix}^T \end{aligned} \tag{2.313}$$

respectively, so that the first term simply averages the two measurements, y_i, and the difference between them contributes—with great uncertainty—in the second term of the rank 2 solution owing to the very small singular value. The uncertainty is

$$(\mathbf{E}\mathbf{E}^T)^{-1} = \begin{Bmatrix} 1.51 \times 10^4 & -1.50 \times 10^4 \\ -1.50 \times 10^4 & 1.51 \times 10^4 \end{Bmatrix}.$$

Now suppose that the covariance matrix of the noise is known to be

$$\mathbf{R}_{nn} = \begin{Bmatrix} 1 & 0.999999 \\ 0.999999 & 1 \end{Bmatrix}$$

(an extreme case, chosen for illustrative purposes). Then, put $\mathbf{W} = \mathbf{R}_{nn}$,

$$\mathbf{W}^{1/2} = \begin{Bmatrix} 1.0000 & 1.0000 \\ 0 & 0.0014 \end{Bmatrix}, \qquad \mathbf{W}^{-T/2} = \begin{Bmatrix} 1.0000 & 0 \\ -707.1063 & 707.1070 \end{Bmatrix}.$$

The new system to be solved is

$$\begin{Bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.0007 & 7.0718 & 0.0007 \end{Bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ 707.1(-y_1 + y_2) \end{bmatrix}.$$

The SVD is

$$\mathbf{U} = \begin{Bmatrix} 0.1456 & 0.9893 \\ 0.9893 & -0.1456 \end{Bmatrix}, \qquad \mathbf{V} = \begin{Bmatrix} 0.0205 & 0.7068 & 0.7071 \\ 0.9996 & -0.0290 & 0.0000 \\ 0.0205 & 0.7068 & -0.7071 \end{Bmatrix}$$

$$\lambda_1 = 7.1450\,, \qquad \lambda_2 = 1.3996\,.$$

The second singular value is now much larger relative to the first one, and the two solutions are,

$$\begin{aligned} \tilde{\mathbf{x}} &\approx \frac{y_2 - y_1}{7.1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \\ \tilde{\mathbf{x}} &\approx \frac{y_2 - y_1}{7.1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T + \frac{y_1 - 103\,(y_2 - y_1)}{1.4} \begin{bmatrix} 0.71 & 0 & 0.71 \end{bmatrix}^T \end{aligned} \tag{2.314}$$

and the rank 1 solution is given by the difference of the observations, in contrast to the unscaled solution. The result is quite sensible—the noise in the two equations is so nearly perfectly correlated, that it can be removed by subtraction; the difference $y_2 - y_1$ is a nearly noise-free piece of information and accurately defines the appropriate structure in $\tilde{\mathbf{x}}$. In effect, the information provided in the row scaling with $\mathbf{R}$ permits the SVD to nearly eliminate the noise at rank 1 by an effective subtraction, whereas without that information, the noise is reduced in the solution (2.313) at rank 1 only by averaging.

At full rank, that is, $K = 2$, it can be confirmed that the solutions (2.313) and (2.314) are identical, as they must be. But the error covariances are quite different:

$$(\mathbf{E}'\mathbf{E}'^T)^{-1} = \begin{Bmatrix} 0.5001 & -0.707 \\ -0.707 & 0.5001 \end{Bmatrix} .$$

because the imposed covariance permits a large degree of noise suppression.

It was previously asserted (P. 66) that in a full-rank formally underdetermined system, row scaling is irrelevant to $\tilde{\mathbf{x}}$, $\tilde{\mathbf{n}}$, as may be seen as follows,

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{E}'^T(\mathbf{E}'\mathbf{E}'^T)^{-1}\mathbf{y}' \\ &= \mathbf{E}^T\mathbf{W}^{-1/2}(\mathbf{W}^{-T/2}\mathbf{E}\mathbf{E}^T\mathbf{W}^{-1/2})^{-1}\mathbf{W}^{-T/2}\mathbf{y} \\ &= \mathbf{E}^T\mathbf{W}^{-1/2}\mathbf{W}^{1/2}(\mathbf{E}\mathbf{E}^T)^{-1}\mathbf{W}^{T/2}\mathbf{W}^{-T/2}\mathbf{y} \\ &= \mathbf{E}^T(\mathbf{E}\mathbf{E}^T)^{-1}\mathbf{y} \,, \end{aligned} \tag{2.316}$$

but which is true only in the full rank situation. {pagerowscale2

There is a subtlety in row-weighting. Suppose we have two equations of form,

$$\begin{aligned} 10x_1 + 5x_2 + x_3 &= 1 \,, \\ 100x_1 + 50x_2 + 10x_3 &= 2 \,, \end{aligned} \tag{2.317}$$ {34086}

after row scaling to make the expected noise variance in each the same. A rank 1 solution to these equations by SVD is $\tilde{\mathbf{x}} = [.0165, .0083, .0017]^T$, which produces residuals $\tilde{\mathbf{y}} - \mathbf{y} = [-0.79, 0.079]^T$—much smaller in the second equation than in the first one.

Consider that the second equation is 10 times the first one—in effect saying that a measurement of 10 times the values of $10x_1 + 5x_2 + x_3$ has the same noise in it as a measurement of one times this same linear combination. The second equation represents a much more accurate determination of this linear combination and the equation should be given much more weight in determining the unknowns—and both the SVD and ordinary least-squares does precisely that. To the extent that one finds this result undesirable (one should be careful about why it is so found), there is an easy remedy—divide the equations by their row norms $\left(\sum_j E_{ij}^2\right)^{1/2}$. But there will be a contradiction with any assertion that the noise in all equations was the same to begin with. Such row-scaling is best regarded as non-statistical in nature.

An example of this situation is readily apparent in the box balances discussed in Chapter 1. Equations such as (1.32) could have row norms much larger than those (1.31) for the corresponding mass balance, simply because the tracer is measured by convention in its own units.

If the tracer is e.g., oceanic salt, values are, by convention, measured on the Practical Salinity Scale, and are near 35 (but are dimensionless). Because there is nothing fundamental about the choice of units, it seems unreasonable to infer that the requirement of tracer balance has an expected error 35 times smaller than for mass. One usually proceeds in the obvious way by dividing the tracer equations by their row norms as the first step. (This approach need have no underlying statistical validity, but is often done simply on the assumption that salt equations are unlikely to be 35 times more accurate than the mass ones.) The second step is to ask whether anything further can be said about the relative errors of mass and salt balance, which would introduce a second, purely statistical row weight.

Column Scaling

In the least-squares problem, we formally introduced a "column scaling" matrix $\mathbf{S}$. Column scaling operates on the SVD solution exactly as it does in the least-squares solution, to which it reduces in the two special cases already described. That is, we should apply the SVD to sets of equations only where any knowledge of the solution element size has been removed first. If the SVD has been computed for such a column-scaled (and row-scaled) system, the solution is for the scaled unknown $\mathbf{x}'$, and the physical solution is,

{34088}
$$\tilde{\mathbf{x}} = \mathbf{S}^{T/2}\tilde{\mathbf{x}}' \,. \tag{2.318}$$

But there are occasions, with underdetermined systems, where a non-statistical scaling may also be called for, the analogue to the situation considered above where a row-scaling was introduced on the basis of possible non-statistical considerations.

Example

Suppose we have one equation in two unknowns,

{34089}
$$10x_1 + 1x_2 = 3 \,. \tag{2.319}$$

The particular-SVD solution produces $\tilde{\mathbf{x}} = [0.2970, 0.0297]^T$ *in which the magnitude of* x_1 *is much larger than that of* x_2 *and the result is readily understood. As we have seen, the SVD automatically finds the exact solution, subject to making the solution norm as small as possible. Because the coefficient of* x_1 *in* (2.319) *is* 10 *times that of* x_2*, it is obviously more efficient in minimizing the norm to give* x_1 *a larger value than* x_2*—because it contributes more efficiently in producing* y*. Although we have demonstrated this dependence for a trivial example, similar behavior occurs for underdetermined systems in general. In many cases, this distribution of the elements of the solution vector* $\mathbf{x}$ *is desirable, the numerical value* 10 *appearing for good physical reasons. In other problems—the numerical values appearing in the coefficient matrix* $\mathbf{E}$ *are an*

"accident." In the box-balance example of Chapter 1, the distances defining the interfaces of the boxes are a consequence of the spatial distance between measurements. Unless one believed that velocities should be larger where the distances are greater or the fluid depth was greater, then the solutions may behave unphysically.[46] *Indeed, in some situations the velocities are expected to be inverse to the fluid depth and such a prior statistical hypothesis is best imposed after one has removed the structural accidents from the system. (The tendency for the solutions to be proportional to the column norms is not rigid. In particular, the equations themselves may preclude the proportionality.)*

Take a positive definite, diagonal matrix $\mathbf{S}$*, and rewrite* (2.88) *as*

$$\mathbf{E}\mathbf{S}^{T/2}\mathbf{S}^{-T/2}\mathbf{x} \ + \mathbf{n} = \mathbf{y}.$$

Then,

$$\mathbf{E}'\mathbf{x}' + \mathbf{n} = \mathbf{y}, \ \mathbf{E}' = \mathbf{E}\mathbf{S}^{T/2}, \ \mathbf{x}' = \mathbf{S}^{-T/2}\mathbf{x}\,.$$

Solving

$$\tilde{\mathbf{x}}' = \mathbf{E}'^{T}(\mathbf{E}'\mathbf{E}'^{T})^{-1}\mathbf{y}\,, \ \tilde{\mathbf{x}} = \mathbf{S}^{T/2}\tilde{\mathbf{x}}'\,. \tag{2.320}$$ {34090}

How should $\mathbf{S}$ *be chosen? Apply the recipe* (2.320) *for the simple one equation example of* (2.319)*, with*

$$\mathbf{S} = \begin{Bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{Bmatrix}$$

:

$$\mathbf{E}' = \begin{Bmatrix} 10/a & 1/b \end{Bmatrix}, \ \mathbf{E}'\mathbf{E}'^{T} = \frac{100}{a^2} + \frac{1}{b^2} \tag{2.321}$$

$$\left(\mathbf{E}'\mathbf{E}'^{T}\right)^{-1} = \frac{a^2b^2}{100b^2 + a^2} \tag{2.322}$$

$$\tilde{\mathbf{x}}' = \begin{Bmatrix} 10/a \\ 1/b \end{Bmatrix} \frac{a^2b^2}{100b^2 + a^2} 3, \tag{2.323}$$

$$\tilde{\mathbf{x}} = \mathbf{S}^{T/2}\tilde{\mathbf{x}}' = \begin{Bmatrix} 10/a^2 \\ 1/b^2 \end{Bmatrix} \frac{a^2b^2}{100b^2 + a^2} 3. \tag{2.324}$$

The relative magnitudes of the elements of $\tilde{\mathbf{x}}$ *are proportional to* $10/a^2$*,* $1/b^2$*. To make the numerical values identical, choose* $a^2 = 10$*,* $b^2 = 1$*, that is, divide the elements of the first column of* $\mathbf{E}$ *by* $\sqrt{10}$ *and the second column by* $\sqrt{1}$*. The apparent rule (which is correct and*

general) is to divide each column of $\mathbf{E}$ *by the square root of its length. The square root of the length may be surprising, but arises because of the second multiplication by the elements of* $\mathbf{S}^{T/2}$ *in* (2.320). *This form of column scaling should be regarded as "non-statistical," in that it is based upon inferences about the numerical magnitudes of the columns of* $\mathbf{E}$ *and does not employ information about the statistics of the solution. Indeed, its purpose is to prevent the imposition of structure on the solution for which no statistical basis has been anticipated. In general, the elements of* $\tilde{\mathbf{x}}$ *will not prove to be equal—because the equations themselves do not permit it.*

If the system is full-rank overdetermined, the column weights drop out, as claimed for least-squares above. To see this result, consider that in the full-rank case,

$$\begin{aligned}\tilde{\mathbf{x}}' &= (\mathbf{E}'^T\mathbf{E}')^{-1}\mathbf{E}'^T\mathbf{y}\\ \tilde{\mathbf{x}} &= \mathbf{S}^{T/2}(\mathbf{S}^{1/2}\mathbf{E}^T\mathbf{E}\mathbf{S}^{T/2})^{-1}\mathbf{S}^{1/2}\mathbf{E}^T\mathbf{y}\\ &= \mathbf{S}^{T/2}\mathbf{S}^{-T/2}(\mathbf{E}^T\mathbf{E})^{-1}\mathbf{S}^{-1/2}\mathbf{S}^{1/2}\mathbf{E}^T\mathbf{y} = (\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T\mathbf{y}\,.\end{aligned} \tag{2.325}$$

Usually row-scaling is done prior to column scaling so that the row norms have a simple physical interpretation.

2.5.10 Solution and Observation Resolution. Data Ranking

Typically, either or both of the set of vectors $\mathbf{v}_i$, $\mathbf{u}_i$ used to present $\mathbf{x}$, $\mathbf{y}$ will be deficient in the sense of the expansions in (2.186). It follows immediately from Eqs. (2.187) that the particular-SVD solution is,

{34093a}
$$\tilde{\mathbf{x}} = \mathbf{V}_K\mathbf{V}_K^T\mathbf{x} = \mathbf{T}_v\mathbf{x}, \tag{2.326}$$

and the data vector with which both it and the general solution are consistent is,

{34093b}
$$\tilde{\mathbf{y}} = \mathbf{U}_K\mathbf{U}_K^T\mathbf{y} = \mathbf{T}_u\mathbf{y}. \tag{2.327}$$

It is convenient therefore, to define the solution and observation resolution matrices,

{resol1}
$$\mathbf{T}_v = \mathbf{V}_K\mathbf{V}_K^T, \qquad \mathbf{T}_u = \mathbf{U}_K\mathbf{U}_K^T\,. \tag{2.328}$$

The interpretation of the solution resolution matrix is identical to that in the square-symmetric case (P. 77).

Interpretation of the data resolution matrix is slightly subtle. Suppose an element of $\mathbf{y}$ was fully resolved, that is, some row, j_0, of $\mathbf{U}_K\mathbf{U}_K^T$ were all zeros except for diagonal element j_0, which is one. Then a change of unity in y_{j_0} would produce a change in $\tilde{\mathbf{x}}$ which would leave unchanged all other elements of $\tilde{\mathbf{y}}$. If element j_0 is *not* fully resolved, then a change of unity in observation y_{j_0} produces a solution which leads to changes in other elements of $\tilde{\mathbf{y}}$. Stated

slightly differently, if y_i is not fully resolved, the system lacks adequate information to distinguish equation i from a linear dependence on one or more other equations.

One can use these ideas to construct quantitative statements of which observations are the most important ("data ranking"). From (2.189), $\text{trace}(\mathbf{T}_u) = K$ and the relative contribution to the solution of any particular constraint is given by the corresponding diagonal element of $\mathbf{T}_u$.

Consider the example (2.317) without row weighting. At rank 1,

$$\mathbf{T}_u = \begin{Bmatrix} 0.0099 & 0.099 \\ 0.099 & 0.9901 \end{Bmatrix},$$

showing that the second equation has played a much more important role in the solution than the first one—despite the fact that we asserted the expected noise in both to be the same. The reason is that described above, the second equation in effect asserts that the measurement is 10 times more accurate than in the first equation—and the data resolution matrix informs us of that explicitly. The elements of $\mathbf{T}_u$ can be used to rank the data in order of importance to the final solution. All of the statements about the properties of resolution matrices made above apply to both $\mathbf{T}_u$, $\mathbf{T}_v$.

If row and column scaling have been applied to the equations prior to application of the SVD, the covariance, uncertainty, and resolution expressions apply in those new, scaled spaces. The resolution in the original spaces is,

$$\mathbf{T}_v = \mathbf{S}^{T/2}\mathbf{T}_{v'}\mathbf{S}^{-T/2}, \tag{2.329}$$ {34095a}

$$\mathbf{T}_u = \mathbf{W}^{T/2}\mathbf{T}_{u'}\mathbf{W}^{-T/2}, \tag{2.330}$$ {34095b}

so that

$$\tilde{\mathbf{x}} = \mathbf{T}_v\mathbf{x}, \qquad \tilde{\mathbf{y}} = \mathbf{T}_u\mathbf{y} \tag{2.331}$$ {34096}

where $\mathbf{T}_{v'}$, $\mathbf{T}_{u'}$ are the expressions Eq. (2.328) in the scaled space. The uncertainty in the new space is $\mathbf{P} = \mathbf{S}^{1/2}\mathbf{P}'\mathbf{S}^{T/2}$ where $\mathbf{P}'$ is the uncertainty in the scaled space.

We have seen an interpretation of three matrices obtained from the SVD: $\mathbf{V}_K\mathbf{V}_K^T$, $\mathbf{U}_K\mathbf{U}_K^T$, $\mathbf{V}_K\mathbf{\Lambda}_K^{-2}\mathbf{V}_K^T$. The reader may well wonder, on the basis of the symmetries between solution and data spaces, whether there is an interpretation of the remaining matrix $\mathbf{U}_K\mathbf{\Lambda}_K^{-2}\mathbf{U}_K^T$?

To understand its use, recall the normal equations (2.162, 2.163) that emerged from the constrained objective function (2.148). They become, using the SVD for $\mathbf{E}$,

$$\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\boldsymbol{\mu} = \mathbf{x}, \tag{2.332}$$ {35018a}

$$\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T\mathbf{x} = \mathbf{y}. \tag{2.333}$$ {35018b}

No matter what the rank of $\mathbf{E}$, the pair of equations is always square, of dimension $M+N$. These equations show that $\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T\boldsymbol{\mu}=\mathbf{y}$. The particular SVD solution is,

{35019}
$$\tilde{\boldsymbol{\mu}}=\mathbf{U}_K\mathbf{\Lambda}_K^{-2}\mathbf{U}_K^T\mathbf{y}\,, \tag{2.334}$$

involving the "missing" fourth matrix. Thus,

$$\frac{\partial J}{\partial \mathbf{y}}=2\mathbf{U}_K\mathbf{\Lambda}_K^{-2}\mathbf{U}_K^T\mathbf{y}\,,$$

and taking the second derivative,

{35020}
$$\frac{\partial^2 J}{\partial \mathbf{y}^2}=2\mathbf{U}_K\mathbf{\Lambda}_K^{-2}\mathbf{U}_K^T \tag{2.335}$$

is the Hessian of J with respect to the data. If any of the λ_i become very small, the objective function will be extremely sensitive to small perturbations in $\mathbf{y}$—producing an effective nullspace of the problem. Eq. (2.335) supports the suggestion that perfect constraints can lead to difficulties.

2.5.11 Relation to Tapered and Weighted Least-Squares

In using least-squares, a shift was made from the simple objective functions (2.90) and (2.148) to the more complicated ones in (2.115) or (2.126). The change was made to permit a degree of control of the relative norms of $\tilde{\mathbf{x}}$, $\tilde{\mathbf{n}}$, and through the use of $\mathbf{W}$, $\mathbf{S}$ of the individual elements and the resulting uncertainties, and covariances. Application of the weight matrices $\mathbf{W}$, $\mathbf{S}$ through their Cholesky decompositions to the equations prior to the use of the SVD is equally valid—thus providing the same amount of influence over the solution elements. The SVD provides its control over the solution norms, uncertainties and covariances through choice of the effective rank K'. This approach is different from the use of the extended objective functions (2.115), but the SVD is actually useful in understanding the effect of such functions.

Assume any necessary $\mathbf{W}$, $\mathbf{S}$ have been applied. Then, the full SVD, including zero singular values and corresponding singular vectors, is substituted into (2.117),

$$\tilde{\mathbf{x}}=(\gamma^2\mathbf{I}_N+\mathbf{V}\mathbf{\Lambda}^T\mathbf{\Lambda}\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{\Lambda}^T\mathbf{U}^T\mathbf{y}\,,$$

we have

$$\begin{aligned}\tilde{\mathbf{x}} &= \mathbf{V}(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I})^{-1}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}^T\mathbf{U}^T\mathbf{y} \\ &= \mathbf{V}\,\mathrm{diag}\left(\lambda_i^2+\gamma^2\right)^{-1}\mathbf{\Lambda}^T\mathbf{U}^T\mathbf{y},\end{aligned} \tag{2.336}$$

or,

{34097b}
$$\tilde{\mathbf{x}}=\sum_{i=1}^{N}\frac{\lambda_i(\mathbf{u}_i^T\mathbf{y})}{\lambda_i^2+\gamma^2}\mathbf{v}_i\,. \tag{2.337}$$

It is now apparent what the effect of "tapering" has done in least-squares. The word refers to the tapering down of the coefficients of the $\mathbf{v}_i$ by the presence of γ^2 from the values they would have in the "pure" SVD . In particular, the guarantee that matrices like $(\mathbf{E}^T\mathbf{E}+\gamma^2\mathbf{I})$ always have an inverse despite vanishing singular values, is seen to follow because the presence of $\gamma^2 > 0$ assures the inverse of the sum always exists, irrespective of the rank of $\mathbf{E}$. The simple addition of a positive constant to the diagonal of a singular matrix is a well-known ad hoc method for giving it an approximate inverse. Such methods are a form of what is usually known as "regularization," and are procedures for suppressing nullspaces. Note that the coefficients of $\mathbf{v}_i$ vanish with λ_i and a solution nullspace still exists.

The residuals of the tapered least-squares solution can be written in various forms. Eqs. (2.118) are,

$$\begin{aligned}\tilde{\mathbf{n}} &= \gamma^2\mathbf{U}(\gamma^2\mathbf{I}+\mathbf{\Lambda}\mathbf{\Lambda}^T)^{-1}\mathbf{U}^T\mathbf{y} \\ &= \sum_{i=1}^{M}\frac{(\mathbf{u}_i^T\mathbf{y})\gamma^2}{\lambda_i^2+\gamma^2}\mathbf{u}_i\,,\end{aligned} \tag{2.338}$$

that is, the projection of the noise onto the range vectors $\mathbf{u}_i$ no longer vanishes. Some of the structure of the range of $\mathbf{E}^T$ is being attributed to noise and it is no longer true that the residuals are subject to the rigid requirement (2.263) of having zero contribution from the range vectors. An increased noise norm is also deemed acceptable, as the price of keeping the solution norm small, by assuring that none of the coefficients in the sum (2.337) becomes overly large—values we can control by varying γ^2. The covariance of this solution about its mean (Eq. 2.119) is readily rewritten as

$$\begin{aligned}\mathbf{C}_{xx} &= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\lambda_i\lambda_j\mathbf{u}_i^T\mathbf{R}_{nn}\mathbf{u}_j^T}{(\lambda_i^2+\gamma^2)(\lambda_j^2+\gamma^2)}\mathbf{v}_i\mathbf{v}_j^T \\ &= \sigma_n^2\sum_{i=1}^{N}\frac{\lambda_i^2}{(\lambda_i^2+\gamma^2)^2}\,\mathbf{v}_i\mathbf{v}_i^T \\ &= \sigma_n^2\mathbf{V}(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I}_N)^{-1}\mathbf{\Lambda}^T\mathbf{\Lambda}(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I}_N)^{-1}\mathbf{V}^T\end{aligned} \tag{2.339}$$

where the second and third lines are again the special case of white noise. The role of γ^2 in controlling the solution variance, as well as the solution size, should be plain. The tapered least-squares solution is biassed —but the presence of the bias can greatly reduce the solution variance. Study of the solution as a function of γ^2 is known as "ridge regression". Elaborate techniques have been developed for determining the "right" value of γ^2.[47]

The uncertainty, $\mathbf{P}$, is readily found as,

$$\begin{aligned}\mathbf{P}=&\gamma^2\sum_{i=1}^{N}\frac{\mathbf{v}_i\mathbf{v}_i^T}{\left(\lambda_i^2+\gamma^2\right)^2}+\sigma_n^2\sum_{i=1}^{N}\frac{\lambda_i^2\mathbf{v}_i\mathbf{v}_i^T}{\left(\lambda_i^2+\gamma^2\right)^2} \\ =&\ \gamma^2\mathbf{V}\left(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I}\right)^{-2}\mathbf{V}^T+\boldsymbol{\sigma}_n^2\mathbf{V}\left(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I}\right)^{-1}\mathbf{\Lambda}^T\mathbf{\Lambda}\left(\mathbf{\Lambda}^T\mathbf{\Lambda}+\gamma^2\mathbf{I}\right)^{-1}\mathbf{V}^T\end{aligned} \tag{2.340}$$

showing the variance reduction possible for finite γ^2 (reduction of the second term), and the bias error incurred in compensation in the first term.

The truncated SVD and the tapered SVD-tapered least-squares solutions produce the same qualitative effect—it is possible to increase the noise norm while decreasing the solution norm. Although the solutions differ somewhat, they both achieve a purpose stated above—to extend ordinary least-squares in such a way that one can control the relative noise and solution norms. The quantitative difference between them is readily stated—the truncated form makes a clear separation between range and nullspace in both solution and residual spaces: The basic SVD solution contains only range vectors and no nullspace vectors. The residual contains only nullspace vectors and no range vectors. The tapered form permits a merger of the two different sets of vectors: Then both solution and residuals contain some contribution from both formal range and effective nullspaces (for $0 \leq \lambda_i^2 << \gamma^2$).

We have already seen several times that preventing $\tilde{\mathbf{n}}$ from having any contribution from the range of $\mathbf{E}^T$ introduces covariances into the residuals, with a consequent inability to produce values which are strictly white noise in character (although it is only a real issue as the number of degrees of freedom, $M-K$, goes toward zero). In the tapered form of least-squares, or the equivalent tapered SVD, contributions from the range vectors $\mathbf{u}_i,\ i \leq K$, is permitted, and a potentially more realistic residual estimate is obtained. (There is usually no good reason why $\tilde{\mathbf{n}}$ should be expected to be orthogonal to the range vectors.)

2.5.12 Resolution and Variance of Tapered Solutions

The tapered least-squares solutions have an implicit nullspace, arising both from the terms corresponding to zero singular values, or from values small compared to γ^2. To obtain a measure of solution resolution when the $\mathbf{v}_i$ vectors have not been computed, consider a situation in which the true solution were $\mathbf{x}_{j_0} \equiv \delta_{j,j_0}$, that is, unity in the j_0 element and zero elsewhere. Then, in the absence of noise, the correct value of $\mathbf{y}$ would be

{34099}
$$\mathbf{E}\mathbf{x}_{j_0} = \mathbf{y}_{j_0}\,, \tag{2.341}$$

defining $\mathbf{y}_{j_0}$. Suppose we actually knew (had measured) $\mathbf{y}_{j_0}$, what solution $\mathbf{x}_{j_0}$ would be obtained?

Assuming all covariance matrices have been applied and suppressing any primes, tapered least-squares (Eqs. 2.121) produces,

{34100}
$$\tilde{\mathbf{x}}_{j_0} = \mathbf{E}^T(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I})^{-1}\mathbf{y}_{j_0} = \mathbf{E}^T(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I})^{-1}\mathbf{E}\mathbf{x}_{j_0}\,, \tag{2.342}$$

which is row (or column) j_0 of

$$\mathbf{T}_v = \mathbf{E}^T(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I})^{-1}\mathbf{E}\,. \tag{2.343}$$ {34101}

Thus we can interpret any row or column of $\mathbf{T}_v$ as the solution for one in which a Kronecker delta was the underlying correct one. It is an easy matter, using the SVD of $\mathbf{E}$ and letting $\gamma^2 \to 0$ to show that (2.343) reduces to $\mathbf{V}\mathbf{V}^T$. These expressions apply in the row- and column-scaled space and are suitably modified to take account of any $\mathbf{W}, \mathbf{S}$ which may have been applied, as in Eqs. (2.329), (2.330). An obvious variant of (2.343) follows from the alternative least-squares solution (2.128), with $\mathbf{W} = \gamma^2\mathbf{I}$, $\mathbf{S} = \mathbf{I}$,

$$\mathbf{T}_v = \left(\mathbf{E}^T\mathbf{E} + \gamma^2\mathbf{I}\right)^{-1}\mathbf{E}^T\mathbf{E} \tag{2.344}$$ {34142}

Data resolution matrices are obtained similarly. Let $y_j = \delta_{jj_1}$. Eq. (2.136) produces

$$\tilde{\mathbf{x}}_{j_1} = \mathbf{E}^T\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right)^{-1}\mathbf{y}_{j_1}, \tag{2.345}$$

which if substituted into the original equations is,

$$\mathbf{E}\tilde{\mathbf{x}}_{j_1} = \mathbf{E}\mathbf{E}^T\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right)^{-1}\mathbf{y}_{j_1}. \tag{2.346}$$

Thus,

$$\mathbf{T}_u = \mathbf{E}\mathbf{E}^T\left(\mathbf{E}\mathbf{E}^T + \gamma^2\mathbf{I}\right)^{-1} \tag{2.347}$$

The alternate form is,

$$\mathbf{T}_u = \mathbf{E}\left(\mathbf{E}^T\mathbf{E} + \gamma^2\mathbf{I}\right)^{-1}\mathbf{E}^T. \tag{2.348}$$

All of the resolution matrices reduce properly to either $\mathbf{U}\mathbf{U}^T, \mathbf{V}\mathbf{V}^T$ as $\gamma^2 \to 0$ when the SVD for $\mathbf{E}$ is substituted.