Notes on Least Squares Optimization

Delbert Yip

December 29, 2020

Abstract

These notes were taken from Chapter 8 from Strang's "Computational Science and Engineering," [2] and cover ordinary, weighted, and regularized least squares (Sections 8.1-8.3 of [2]). Some background material has been taken from Lay's "Linear algebra and its applications" [1].

1 Least Squares

Consider a matrix A of size mxn with n independent columns, and $m > n$. A^TA is symmetric positive definite, as Rank(A) = n. We wish to solve $Au = b$ for some vector b with m components. The solution $\hat{u}$ has n components.

This gives rise to the **least squares problem**:

$$\min \parallel Au - b \parallel^2 \tag{1}$$

Below are the normal equations, which give the solution $\hat{u}$:

$$\begin{aligned} Au &= b \\ A^TA\hat{u} &= A^Tb \end{aligned} \tag{2}$$

The residual of the cost function is $e = b - Au$, and has dimensions $m \times 1$. Thus,

$$A^Te = A^Tb - A^TAu \tag{3}$$

Therefore, $A^Te = 0$ when $u = \hat{u}$.

e is in the nullspace of A^T, and is therefore perpendicular to the columns (column space) of A (equivalently, e is perpendicular to the rows/rowspace of A^T). Recall that orthogonality means that the dot product of two vectors is 0.

1.1 Refresher on Orthogonality

Before we start on the next section, let's review some concepts from linear algebra. Specifically, the orthogonality of vectors and vector spaces.

Orthogonal Vectors There is an analogue between geometry in Euclidean space and in $\mathbb{R}^n$.

First, *two vectors $\mathbf{u}$ and $\mathbf{v}$ are perpendicular i.f.f. the distance between $\mathbf{u}$ and $-\mathbf{v}$ is the same as the distance between $\mathbf{u}$ and $+\mathbf{v}$.*

First, the distance between $\mathbf{u}$ and $-\mathbf{v}$:

$$\begin{aligned}[\text{dist}(\mathbf{u}, -\mathbf{v})]^2 &= \| \mathbf{u} - (-\mathbf{v}) \|^2 = \| \mathbf{u} + \mathbf{v} \|^2 && (4)\\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) && (5)\\ &= \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2 + 2\mathbf{u} \cdot \mathbf{v} && (6)\end{aligned}$$

Next, the distance between $\mathbf{u}$ and $+\mathbf{v}$:

$$\begin{aligned}[\text{dist}(\mathbf{u}, \mathbf{v})]^2 &= \| \mathbf{u} - \mathbf{v} \|^2 && (7)\\ &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) && (8)\\ &= \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2 - 2\mathbf{u} \cdot \mathbf{v} && (9)\end{aligned}$$

Therefore,

$$\begin{aligned}[\text{dist}(\mathbf{u}, -\mathbf{v})]^2 = [\text{dist}(\mathbf{u}, \mathbf{v})]^2 &\iff -2\mathbf{u} \cdot \mathbf{v} = 2\mathbf{u} \cdot \mathbf{v} && (10)\\ &\implies \mathbf{u} \cdot \mathbf{v} = 0 && (11)\\ &\implies \| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2 && (12)\end{aligned}$$

1.2 Orthogonal Complements in $\mathbb{R}^n$

If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is orthogonal to W. The set of all such vectors $\mathbf{z}$ is the **orthogonal complement** of W and denoted by $W^\perp$. Let's go through the relation between nullspace, orthogonal complements, and the columns and rows of a $m \times n$ matrix A.

We have the folllowing theorem:

$$(\text{Row}(A))^\perp = \text{Nul}(A) \quad \text{and} \quad (\text{Col}(A))^\perp = \text{Nul}(A)^\perp \tag{13}$$

If $Ax = 0$ for some $n \times 1$ vector x, then $x \in \text{Nul}(A)$, which means x is orthogonal to the rows of A. The rows of A span its row space, so x is orthogonal to $\text{Row}(A)$. This is a consequence of how matrix multiplication works, as shown below:

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{14}$$

$$Au = \begin{pmatrix} a \\ d \end{pmatrix} x + \begin{pmatrix} b \\ e \end{pmatrix} y + \begin{pmatrix} c \\ f \end{pmatrix} z, \quad \text{linear combination of the columns} \tag{15}$$

$$= \begin{pmatrix} ax + by + cz \\ dx + ey + fz \end{pmatrix}, \quad \text{dot product of the rows} \tag{16}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{17}$$

The elements of the product Au are dot products between u and the row vectors of A; the product Au can also be thought of as a linear combination of the columns of A. Since $Au = 0$, u is in the nullspace of A, and from the above, we see that u is orthogonal to the rows of A.

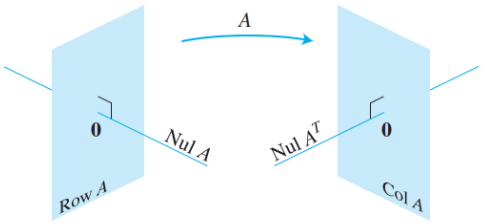

FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Figure 1: Fundamental subspaces of A. From Lay.

1.3 Primal and Dual Problems

In the 'primal' problem, we find the solution $\hat{u}$ by projecting b onto the column space of A, i.e. $A\hat{u}$ is the closest point to b in the column space of A. In general, the product Au yields a linear combination of the columns of A. The column space of A contains all such products.

$$Au = y \quad \Longrightarrow \quad y \in \text{Col}(A), \quad \forall u \tag{18}$$

Where, Col(A) refers to the column space of A.

The 'dual' problem solves $A^T e = 0$ for e, which is the projection of b onto the nullspace of A^T. Recall from above:

$$\text{Nul}(A^T) = (\text{Row}(A^T))^\perp = (\text{Col}(A))^\perp$$

Thus, e is contained in $(\text{Col}(A))^\perp$, which contains *all vectors perpendicular to the columns of A*. The dimension of $(\text{Col}(A))^\perp$ is given by the Rank-Nullity Theorem:

$$\text{Rank}(A) + \text{Nullity}(A) = n, \quad \text{where } n \text{ is the number of columns in } A. \tag{19}$$

$$\text{Rank}(A^T) + \text{Nullity}(A^T) = m \tag{20}$$

$$\text{Rank}(A) = \text{Rank}(A^T) \quad \Longrightarrow \quad \text{Nullity}(A^T) = m - n \tag{21}$$

$$\text{Nullity}(A^T) = \dim(\text{Row}(A^T))^\perp = \dim(\text{Col}(A)^\perp) = m - n \tag{22}$$

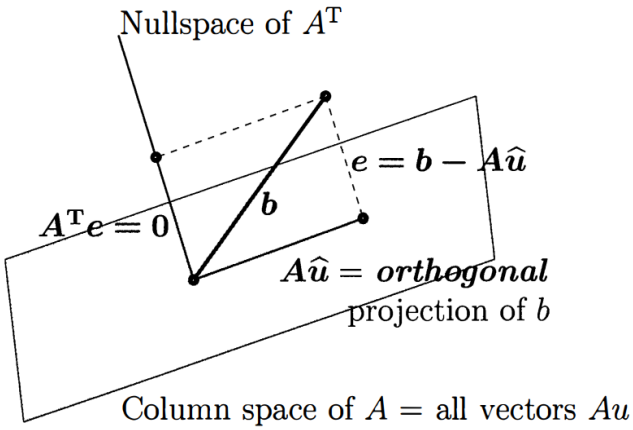

Figure 2: Ordinary least squares. From Strang.

e and $\hat{u}$ solve the two linear equations shown by the figure above:

$$e + A\hat{u} = b \rightarrow m \text{ equations} \tag{23}$$

$$A^T e = 0 \rightarrow n \text{ equations} \tag{24}$$

These are called the "Primal-Dual", "Saddle Point", and "Kuhn-Tucker (KKT)" equations.

1.3.1 "Saddle Point" Equations

The Saddle Point, or KKT, matrix is a block matrix that represents the Primal-Dual equations:

$$S = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \quad \Longrightarrow \quad S \begin{bmatrix} e \\ \hat{u} \end{bmatrix} = \begin{bmatrix} Ie + A\hat{u} \\ A^T e + 0 \end{bmatrix}$$

Where, I is an $m \times m$ identity matrix. Thus, I constitutes the first m pivot columns of S. Thus, to reduce S, we need to make the lower left hand corner of S a $n \times n$ matrix of zeroes. To do this, let's review elimination on block matrices:

Elimination of block matrices:

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A+0 & B+0 \\ -CA^{-1}A + C & -CA^{-1}B + D \end{bmatrix} \tag{25}$$

$$= \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix} \tag{26}$$

The first matrix, $\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$, is also known as the Gaussian elimination matrix, often represented using the symbol E. If A was not an identity matrix, then we can also reduce A into a matrix of pivots:

$$\begin{bmatrix} A^{-1} & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} AA^{-1}+0 & BA^{-1}+0 \\ -CA^{-1}A+C & -CA^{-1}B+D \end{bmatrix} \tag{27}$$

$$= \begin{bmatrix} I & BA^{-1} \\ 0 & D-CA^{-1}B \end{bmatrix} \tag{28}$$

We now reduce our Primal-Dual matrix S with $E = \begin{bmatrix} I & 0 \\ -A^T & I \end{bmatrix}$:

$$ES = \begin{bmatrix} I & 0 \\ -A^T & I \end{bmatrix} \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \tag{29}$$

$$= \begin{bmatrix} I+0 & A+0 \\ -A^T+A^T & -AA^T+0 \end{bmatrix} \tag{30}$$

$$= \begin{bmatrix} I & A \\ 0 & -AA^T \end{bmatrix} \tag{31}$$

By convention, the lower right-hand block remaining after elimination is known as the **Schur complement**. More importantly, the first m pivots are positive (=1), but the final n pivots, determined by $-AA^T$, are negative. When a matrix has pivots of both signs, it is **indefinite**. For such matrices, *there is no maximum or minimum, positive or negative definite.* Instead, there is a **saddle point,** $(\hat{u}, e)$.

1.3.2 "Primal-Dual" Equations

The *primal* problem is to minimize $\frac{1}{2} \parallel Au - b \parallel^2$. The *dual* problem is to minimize $\frac{1}{2} \parallel e - b \parallel^2$ with the constraint $A^T e = 0$. In the dual problem, u acts a Lagrange multiplier to enforce the constraint $A^T e = 0$. Solutions to the primal and dual problems are found simultaneously, and add to b.

Let's do an example!

1.3.3 An Example

$$A = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad A^T A = 5, \quad A^T b = 10 \tag{32}$$

$$\text{Projection: } A\hat{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \tag{33}$$

$$\text{Error: } e = b - A\hat{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \tag{34}$$

Let's verify that $A\hat{u}$ and e are perpendicular. First, $A\hat{u}$ and e sum to b:

$$A\hat{u} + e = A\hat{u} + (b - A\hat{u}) = b \tag{35}$$

$$\| b \|^2 = \| A\hat{u} \|^2 + \| e \|^2 \tag{36}$$
$$= 20 + 5 = 25 \tag{37}$$
$$A\hat{u} \cdot e = \| A\hat{u} \| \| e \| \cos\theta \tag{38}$$
$$0 = \sqrt{20}\sqrt{5}\cos\theta \tag{39}$$
$$\therefore \cos\theta = 0 \quad \Longrightarrow \quad \theta = \frac{\pi}{2} \tag{40}$$

The normal equation is: $A^T A\hat{u} = A^T b$, so $\hat{u} = 2$. This solves the primal problem. The dual minimizes $\frac{1}{2} \| e - b \|^2$ under the constraint $A^T e = 0$. The solution to the dual problem gives e and $\hat{u}$:

The **Lagrange** function is:

$$L = \frac{1}{2} \| e - b \|^2 + u(A^T e) \tag{41}$$
$$L(e_1, e_2, u) = \frac{1}{2}(e_1 - b_1)^2 + \frac{1}{2}(e_2 - b_2)^2 + u(A_1^T e_1 + A_2^T e_2) \tag{42}$$
$$\frac{\partial L}{\partial(e_1, e_2, u)} = \begin{bmatrix} e_1 - b_1 + 2\hat{u} = 0 \\ e_2 - b_2 + \hat{u} = 0 \\ 2e_1 + e_2 = 0 \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \hat{u} \\ 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} \tag{43}$$

The matrix $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix}$ is the saddle point matrix S, whose columns correspond to e_1, e_2, and u, respectively.

2 Weighted Least Squares

The weighted least squares equation is:

$$\min \| WAu - Wb \|^2 \tag{44}$$

with normal equations:

$$(WA)^T(WA)\hat{u}_W = (WA)^T(Wb) \tag{45}$$
$$A^T W^T W A\hat{u}_W = A^T W^T W b \tag{46}$$
$$A^T C A\hat{u}_W = A^T C b, \quad C = W^T W \tag{47}$$
$$\Longrightarrow 0 = A^T C b - A^T C A\hat{u}_W \tag{48}$$
$$= A^T C(b - A\hat{u}_W) = A^T C e, \quad e = (b - A\hat{u}_W) \tag{49}$$

Where, $\hat{u}_W$ is the solution, and changes with the weight matrix W. Let's consider the geometric effects of introducing W.

- First, $A^T e = 0$ becomes $A^T C e = 0$. So, e is perpendicular to $A^T C$ now, not A^T by itself. So, e is in the nullspace of $A^T C$, not that of A^T.
- With W, e is now in the orthogonal complement to the rows of $A^T C$, or equivalently, the columns of $C^T A$. Thus, e is not orthogonal to the column space of A. We say that e is "C-orthogonal" to the columns of A.
- Consequently, solving weighted least squares involves finding a projection of b that is orthogonal to $\text{Col}(C^T A)$, but not $\text{Col}(A)$.
- We still split b into two components: $A\hat{u}_W \in \text{Col}(A)$, and $e \in \text{Nul}(A^T C)$.

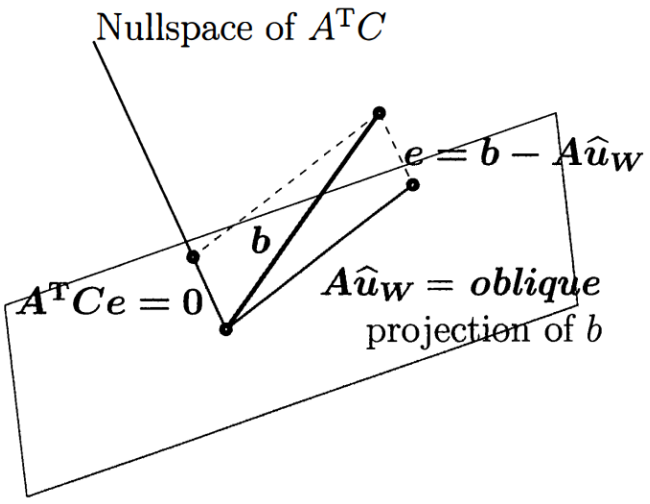

Figure 3: Weighted least squares.

So, we have the following primal-dual equations to solve:

$$e + A\hat{u}_W = b \tag{50}$$

$$A^T C e = 0 \tag{51}$$

The saddle point matrix is:

$$\begin{bmatrix} I & A \\ A^T C & 0 \end{bmatrix}, \tag{52}$$

where the columns represent coefficients of e and $\hat{u}_W$, respectively. Ideally, we'd like that the equations be symmetric. However, we have a factor of C in the bottom-left that is not in the top-right, so our current matrix is not symmetric. To make it symmtric, introduce $w = Ce$ and $e = C^{-1} w$, and let $\hat{u}_W = u$:

$$C^{-1}w + Au = b \tag{53}$$
$$A^T w = 0 \tag{54}$$

This gives rise to a symmetric matrix:

$$S = \begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix}, \tag{55}$$

where columns represent coefficients for w and u, respectively. Since $C = W^T W$, which is the vector analogue to squaring, the elements of C, and thus C^{-1}, are positive. W is a $m \times n$ matrix, so C^{-1} contributes m *positive* pivot columns. The remaining pivots are found by elimination on S:

$$E = \begin{bmatrix} I & 0 \\ -A^T C & I \end{bmatrix} \tag{56}$$
$$ES = \begin{bmatrix} I & 0 \\ -A^T C & I \end{bmatrix} \begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} \tag{57}$$
$$= \begin{bmatrix} C^{-1} + 0 & A + 0 \\ -A^T C C^{-1} + A & -A^T C A + 0 \end{bmatrix} \tag{58}$$
$$= \begin{bmatrix} C^{-1} & A \\ -A^T + A & -A^T C A \end{bmatrix} = \begin{bmatrix} C^{-1} & A \\ 0 & -A^T C A \end{bmatrix} \tag{59}$$

The Schur complement is $-A^T C A$, which is negative like before. So, we have n negative pivots from this block and m positive pivots from C^{-1}, making S indefinite. The equations to solve are:

$$\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{60}$$
$$\begin{bmatrix} C^{-1} & A \\ 0 & -A^T C A \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} b \\ -A^T C b \end{bmatrix} \quad \text{after elimination.} \tag{61}$$

Due to the constraint that $A^T w = 0$, Au and w are perpendicular:

$$(Au)^T w = u^T (A^T w) = 0 \tag{62}$$

2.0.1 Duality and Weak Duality

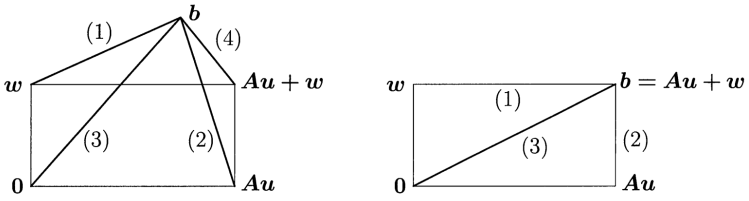

Figure 8.2: Four squares $(1)^2+(2)^2 = (3)^2+(4)^2$ give weak duality. Then $(1)^2+(2)^2 = (3)^2$ is duality. In that case $(4)^2 = \|b - Au - w\|^2 = 0$ splits b into its projections.

Figure 4: Geometric interpretation of weak duality. From Strang.

Imagine we have a rectangle, where the perpendicular sides are Au and w. The remaining corners of this rectangle are 0 and $Au + w$. Lines from b to the four corners of this rectangle are vectors $b-X+a$ where X is a corner of the rectangle and a is some offset (since the lines don't all start from 0).

We can ignore the offset by comparing the vectors' magnitudes. In particular, the diagonal distance starting from either the lower-left or lower-right corner of the rectangle (right side of Fig. 4) is the same, but involve different vectors. The magnitude of each diagonal is the sum of two vectors that meet at b. One diagonal connects points $(0, b, Au + w)$, and has magnitude $\| b \|^2 + \| b - (Au+w) \|^2$. The other diagonal connects points (Au, b, w), and has magnitude $\| b - Au \|^2 + \| b - w \|^2$. Equating these yields the following relationship:

$$\| b - Au \|^2 + \| b - w \|^2 = \| b \|^2 + \| b - Au - w \|^2 \tag{63}$$

At the minimum, we have that $\| b \|^2 = \| b - Au \|^2 + \| b - w \|^2$. This is summarized by the image on the right of Fig. 4, which clearly shows how Au and w are projections of b under this condition:

$$\| b - Au - w \|^2 = 0 \tag{64}$$

$$b - Au - w = 0 \tag{65}$$

$$Au - w = b \tag{66}$$

The solution $u = \hat{u}$ and $w = e$ gives **duality**: $\| b - A\hat{u} \|^2 + \| b - e \|^2 = \| b \|^2$. **Weak duality** is when $\| b - A\hat{u} \|^2 + \| b - e \|^2 > \| b \|^2$. In this scenario, we have a "duality gap" $\| b - Au - w \|^2$. Duality is achieved when the gap is zero. When this is true, $b = Au - w$, and b becomes a corner of the rectangle, as shown in the right side of Fig. 4. Finally, by Pythagoras' rule:

$$\| b - w \|^2 + \| b - Au \|^2 = \| b \|^2 = \| Au + w \|^2, \tag{67}$$

which is exactly what happens when we have perfect duality!

To summarize this section, weighted least squares involves solving two problems:

1. Project b onto the column space of $A\hat{u}$:

$$\min \parallel b - Au \parallel^2$$

2. Project b across to e:

$$\begin{aligned} \min \quad & \parallel b - w \parallel^2 \\ \text{s.t.} \quad & A^T w = 0 \end{aligned}$$

The solutions to these problems converge at optimality, as we showed above. At optimality, $\hat{w} = C(b - A\hat{u})$. C is the bridge between the two dual problems and identiies the solutions $\hat{u}$ and $\hat{w}$. Alternatively, we can see the problem as a maximization problem, simply by attaching a negative sign:

$$\parallel b - e \parallel^2 + \parallel b - A\hat{u} \parallel^2 = \parallel b \parallel^2 \tag{68}$$

$$\parallel b - e \parallel^2 - \parallel b \parallel^2 = - \parallel b - A\hat{u} \parallel^2 \tag{69}$$

$$\therefore \quad \min(\parallel b - e \parallel^2 - \parallel b \parallel^2) \quad \equiv \quad \max(- \parallel b - A\hat{u} \parallel^2) \tag{70}$$

2.1 Minimizing with Constraints using Lagrange's Method

Imagine we have a line of two springs and one mass and that we wish to minimize the energy in the springs. The constraint is $A^T w = f$, where we have forces w_1 and w_2, and a mass that exerts external force f. The springs are vertically arranged with the mass in between, which causes spring 1 (top) to stretch, and spring 2 (bottom) to compress (Fig. 5).

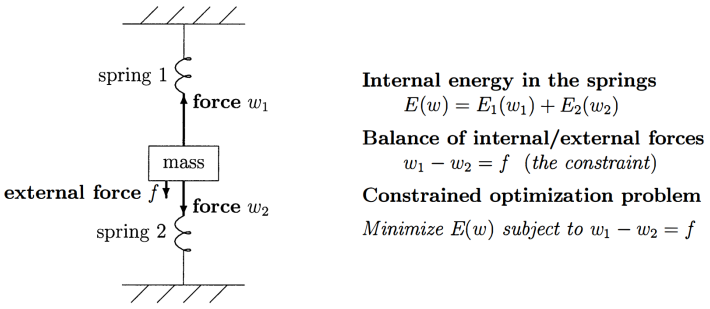

Figure 5: System with two linear springs and one mass.

To solve the force balance equation $w_1 + w_2 = f$, we won't use calculus (setting derivatives to zero), because while we have an energy function $E(w_1, w_2)$, it is unclear which derivatives we should adjust to achieve force balance. We could make the substitution: $w_2 = w_1 - f$, so that we only minimize over one variable, w_1. Here, we will use Lagrange multipliers to *build constraints into the function* by adding an unknown u that enforces our constraint.

With n constraints on m unknowns, this method has $m+n$ unknowns. We will add a Lagrange multiplier for each constraint. The Lagrange function is:

$$L(w_1, w_2, u) = E_1(w_1)+E_2(w_2)-u(w_1-w_2-f) = E_1+E_2-u(w_1-w_2-f) \quad (71)$$

Where, u takes a negative sign by convention. w_2 takes a negative sign as well, to distinguish the compression of spring 2 from the stretching of spring 1. Note that constraints, the term multiplied by u, must equal zero. The Lagrange multiplier u has a special meaning: here, the displacement of the mass. In economics, the selling price for maximal profit. In all problems, u measures the answer's sensitivity (here, the minimum energy $E_{\min}$) to changes in constraints.

Hooke's law tells us that the force in a linear spring is proportional to elongation e, so $w = ce$. Work = force $\times$ displacement, and the integral of work is energy. Thus, Work $= ce$, and energy $E = \int$ Work $= \frac{1}{2}ce^2$. Substituting $e = \frac{w}{c}$, we get $E = \frac{w^2}{2c}$. The resulting minimization problem is then:

$$\begin{aligned} \min \quad & E(w) = \frac{w_1^2}{2c} + \frac{w_2^2}{2c} \\ \text{s.t.} \quad & w_1 - w_2 = f \end{aligned} \quad (72)$$

2.1.1 Geometry of the method

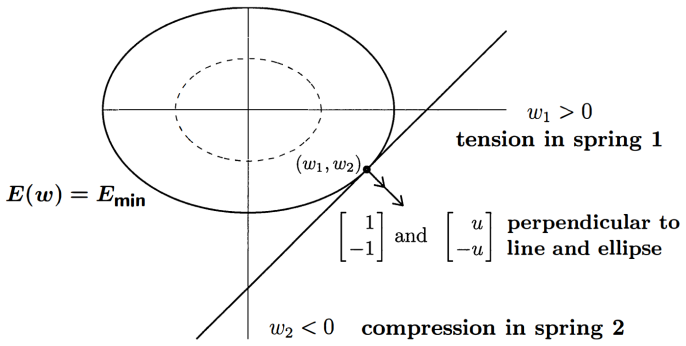

Figure 8.4: The ellipse $E(w) = E_{\min}$ touches $w_1 - w_2 = f$ at the solution (w_1, w_2).

Figure 6: Geometric depiction of $E(w_1, w_2)$ and constraint $w_1 - w_2 = f$.

Notice that $E(w)$ is the equation of an ellipse. We can think of the constraint geometrically in the plane with axes w_1 and w_2: $E(w)$ is some ellipse in this plane, and $w_1 - w_2 = f$ is a line with positive slope and negative y-intercept (Fig. 6). The solution is achieved when the line meets the ellipse at a right-angle. At this point (w_1, w_2), there are two vectors of interest that describe the rate of change of the ellipse and line with respect to w_1 and w_2. One is

perpendicular to the line:

$$\begin{bmatrix}\partial l/\partial w_1 \\ \partial l/\partial w_1\end{bmatrix} = \begin{bmatrix}1 \\ -1\end{bmatrix}, \quad l = w_1 - w_2 - f = 0$$

The second is perpendicular to the ellipse, and arises from the gradient of $E(w)$:

$$\begin{aligned} L(w_1, w_2, u) &= E_1 + E_2 - u(w_1 - w_2 - f) \\ \frac{\partial L}{\partial w_1} &= \frac{\partial E}{\partial w_1} - u = 0 \\ \frac{\partial L}{\partial w_2} &= \frac{\partial E}{\partial w_2} + u = 0 \\ \begin{bmatrix}\partial E/\partial w_1 \\ \partial E/\partial w_2\end{bmatrix} &= \begin{bmatrix}u \\ -u\end{bmatrix} \end{aligned}$$

The above gives us w_1 and w_2 as follows:

$$\frac{\partial E}{\partial w_1} = \frac{\partial}{\partial w_1}\frac{w_1^2}{2c_1} = \frac{w_1}{c_1} \tag{73}$$

$$\therefore \begin{bmatrix}\partial E/\partial w_1 \\ \partial E/\partial w_2\end{bmatrix} = \begin{bmatrix}w_1/c_1 \\ w_2/c_2\end{bmatrix} = \begin{bmatrix}u \\ -u\end{bmatrix} \tag{74}$$

$$w_1 = c_1 u, \quad w_2 = -c_2 u \tag{75}$$

$$w_1 - w_2 = f = c_1 u + c_2 u = (c_1 + c_2)u \tag{76}$$

$c_1 + c_2$ is the stiffness matrix $A^T C A$, which we discussed in Section 2. Since the problem is small, $K = A^T C A$ is just 1×1:

$$A^T = \begin{bmatrix}1 \\ -1\end{bmatrix}, \quad K = A^T C A = \begin{bmatrix}1 \\ -1\end{bmatrix}\begin{bmatrix}c_1 & \\ & c_2\end{bmatrix}\begin{bmatrix}1 \\ -1\end{bmatrix} = \begin{bmatrix}c_1 + c_2\end{bmatrix}$$

We now find the forces w_1 and w_2:

$$\begin{aligned} Ku = f \quad &\Longrightarrow \quad u = \frac{f}{c_1 + c_2} \\ w_1 = c_1 u = \frac{c_1 f}{c_1 + c_2}, \quad &w_2 = -c_2 u = -\frac{c_2 f}{c_1 + c_2} \\ E_{\min} = \frac{w_1^2}{2c} + \frac{w_2^2}{2c} &= \frac{c_1 f^2}{2(c_1 + c_2)^2} + \frac{c_2 f^2}{2(c_1 + c_2)^2} \\ &= \frac{f^2}{2(c_1 + c_2)} \end{aligned}$$

Above, we mentioned that u measures the sensitivity of $E_{\min}$ to changes in the constraint, f. We can show this by computing the derivative $\frac{dE_{\min}}{df}$:

$$\begin{aligned} \frac{dE_{\min}}{df} &= \frac{d}{df}\left(\frac{f^2}{2(c_1 + c_2)}\right) \\ &= \frac{f}{c_1 + c_2} = u \end{aligned} \tag{77}$$

2.2 The Fundamental Problem

This is what Strang describes as the *fundamental problem of scientific computing*: the full linear case with m springs. Each spring experiences a force w_i, which we can represent as $w = (w_1, \cdots, w_m)$. We have n constraints for force balance, $A^T w = (f_1, \cdots, f_n)$, enforced by n Lagrange multipliers $u = (u_1, \cdots, u_n)$.

Let C be a matrix with masses $c_1, \cdots, c_m$ on the diagonal and zeroes everywhere else. Then, the total energy function for m springs is given by: $E(w) = \frac{1}{2} w^T C^{-1} w$. Our Lagrange function is thus:

$$L(w, u) = \frac{1}{2} w^T C^{-1} w - u^T (A^T w - f) \tag{78}$$

The minimizer w is found by setting the $m + n$ partial derivatives of L to zero:

$$\frac{\partial L}{\partial w} = C^{-1} w - Au = 0 \tag{79}$$

$$\frac{\partial L}{\partial u} = -A^T w + f = 0 \tag{80}$$

From Eq. 79, we have that $w = Ce = CAu$, which implies that $e = Au$. Eq. 80 indicates that $f = A^T w$. Together, we have that:

$$A^T C A u = A^T C e = A^T w = f \tag{81}$$

Least squares problems have $e = b - Au$, which appears in E as:

$$E(b, w) = \frac{1}{2} w^T C^{-1} w - b^T w$$

The derivatives of L (with u positive) are:

$$\frac{\partial L}{\partial w} = C^{-1} w + Au - b = 0 \tag{82}$$

$$\frac{\partial L}{\partial u} = -A^T w - f = 0 \tag{83}$$

$$\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} b \\ f \end{bmatrix} \tag{84}$$

The vector $(w, u)^T$ has size $m+n$, since w has m components and u has n. If C is positive definite, and A has full column rank (all columns are independent, $\text{Rank}(A) = n$), then $A^T C A$ and S are invertible. We can factor S into three parts as follows:

$$S = \begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ A^T C & I \end{bmatrix} \begin{bmatrix} C^{-1} & 0 \\ 0 & -A^T C A \end{bmatrix} \begin{bmatrix} I & CA \\ 0 & I \end{bmatrix}$$

S is invertible when both block columns have full rank. Otherwise, it can be singular.

3 Regularized Least Squares

Regularized least squares is a special case of weighted least squares:

$$\begin{aligned}
\textbf{Ordinary LSQ} \quad & \min \parallel Au - b \parallel^2 \\
\text{Solve:} \quad & A^T A\hat{u} = A^T b \\
\textbf{Weighted LSQ} \quad & \min (b - Au)^T C(b - Au) \\
\text{Solve:} \quad & A^T CA\hat{u} = A^T Cb \\
\textbf{Regularized LSQ} \quad & \min \parallel Au - b \parallel^2 + \alpha \parallel Bu - d \parallel^2 \\
\text{Solve:} \quad & (A^T A + \alpha B^T B)\hat{u} = A^T b + \alpha B^T d
\end{aligned}$$

We can rewrite the regularized LSQ equations to resemble those for weighted LSQ by replacing A in the latter by $[A\ B]^T$:

$$\begin{bmatrix} A^T & B^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \alpha I \end{bmatrix} \hat{u} = \begin{bmatrix} A^T & B^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \alpha I \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} \tag{85}$$

Where, the weighting matrix $C = \begin{bmatrix} I & 0 \\ 0 & \alpha I \end{bmatrix}$. There are two important applications that lead to this sum of two squares $(A^T A, B^T B)$:

1. If $A^T A$ is ill-conditioned, e.g. very large ratio of largest to smallest eigenvalues, $m < n$, $A^T A$ is singular.
   Regularization acts to 'reduce the noise.' Normally, as α is increased, $\parallel \hat{u} \parallel$ decreases and $\parallel A\hat{u} - b \parallel$ increases. The **Discrepancy Principle** selects α so that $\parallel \hat{u} \parallel$ decreases and $\parallel A\hat{u} - b \parallel \approx$ expected noise (uncertainty in b).

2. As $\alpha \to \infty$, $Bu \to d$. The limiting $\hat{u}_\infty$ solves

$$\begin{aligned} \min \quad & \parallel Au - b \parallel^2 \\ \text{s.t.} \quad & Bu = d \end{aligned} \tag{86}$$

3.1 Example: Large Penalty Enforced $Bu = d$

$$\begin{aligned} \text{Problem:} \quad \min \quad & \parallel Au \parallel^2 = u_1^2 + u_2^2 \\ \text{s.t.} \quad & Bu = u_1 - u_2 = 8 \end{aligned}$$

Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\quad B = \begin{bmatrix} 1 & -1 \end{bmatrix}$, $\quad d = \begin{bmatrix} 8 \end{bmatrix}$.

$$A^T A = I, \quad B^T B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1+\alpha & -\alpha \\ -\alpha & 1+\alpha \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 8\alpha \\ -8\alpha \end{bmatrix} = \alpha B^T d$$

$$u_1 + u_2 = 0, \quad (1 + 2\alpha)u_1 = 8\alpha$$

$$u_1 = \frac{8\alpha}{1+2\alpha} = \frac{4}{1+\frac{\alpha}{2}} = 4 - \frac{4}{2\alpha} + \cdots \to u_1 = 4$$

The order is of order $1/\alpha$, so we need large α.

3.2 The Pseudoinverse

Suppose A is $m \times n$ and b has m components. The normal equations $A^T A\hat{u} = A^T b$ only give $\hat{u}$ when $A^T A$ is invertible. Otherwise, we can use the pseudoinverse to find the best solution u^+. This works even when A is not full rank and $A^T A$ is singular

3.2.1 What is u^+?

$u^+ = A^+ b$ provides the **shortest** vector that solves $A^T Au^+ = A^T b$.

Other solutions, longer than u^+, have components in the nullspace of A. u^+ is **the particular solution with no nullspace component**. The **pseudoinverse** of A is a matrix A^+, which produces $u^+ = A^+ b$. If A is square and invertible, then $u^+ = u = A^{-1}b$, and $A^+ = A^{-1}$.

If A has independent columns, but $m < n$, then $\hat{u} = (A^T A)^{-1} A^T b$ is the only solution, where $A^+ = (A^T A)^{-1} A^T$. If A has dependent columns and thus a nonzero nullspace: any Au is a linear combination of the columns of A, but if the columns are dependent, then there exists a nonzero vector x such that $Ax = 0$, and so $x \in \text{Nul}(A)$ and $\text{Nul}(A) \neq \{0\}$.

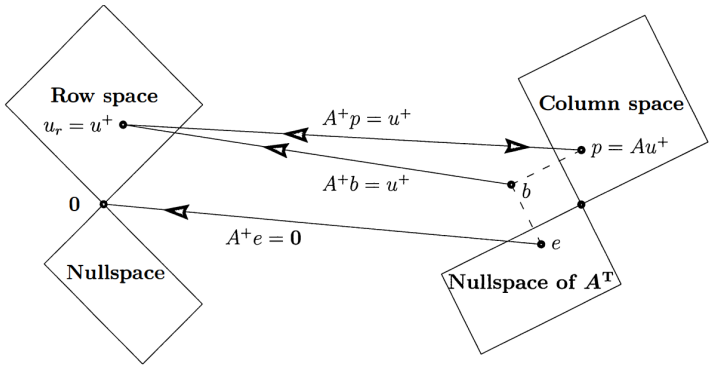

Figure 8.6: The pseudoinverse A^+ inverts A where it can, on the column space.

Figure 7: Au takes vectors $u = u_{\text{row}} + u_{\text{null}}$ to the column space of A. Since u_{row} and u_{null} are orthogonal, a non-zero u_{null} increases $|u|^2$. Thus, $u^+ = u_{\text{row}}$

.

u^+ solves $Au^+ = p$, where p is the projection of b onto the column space of A. The error $|e| = |b - p| = |b - Au^+|$ is minimized by u^+. In summary, u^+ is in the row space (to be as short as possible), and $Au^+ = p$ (to be as close to b as possible); u^+ minimizes e and solves $A^T Au^+ = A^T b$.

3.2.2 Singular Value Decomposition (SVD)

One method to compute u^+ is by SVD. Let's review diagonalization and SVD.

Recall the **Diagonalization Theorem**:

- An $n \times n$ matrix A is diagonalization if and only if A has n linearly independent eigenvectors.
- $A = PDP^{-1}$ with D being a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. Then, the entries of D are the corresponding eigenvalues of A for the eigenvectors in P.

Diagonalization of Symmetric Matrices

Furthermore, if A is symmetric, then:

- $A^T = A$
- All eigenvectors from different eigenspaces (i.e. corresponding to different eigenvalues) are **orthogonal** to each other.
- $A = PDP^{-1}$, but P is square and has orthonormal columns $\implies$ P is an *orthogonal matrix*, and $P^{-1} = P^T$.
- A is orthogonally diagonalizable if and only if A is symmetric.

We can prove that, for symmetric A, two given eigenvectors v_1 and v_2 are orthogonal:

$$\begin{aligned}
\lambda_1 v_1 \cdot v_2 &= (\lambda_1 v_1)^T v_2 = (Av_1)^T v_2 \quad \leftarrow \lambda_1 v_1 = Av_1 \\
&= (v_1^T A^T) v_2 = v_1^T (Av_2) = v_1^T (Av_2) \quad \leftarrow A^T = A \\
&= v_1^T \lambda_2 v_2 = \lambda_2 v_1^T v_2 = \lambda_2 v_1 \cdot v_2 \\
(\lambda_1 - \lambda_2) v_1 \cdot v_2 &= 0 \\
\lambda_1 - \lambda_2 \neq 0 \quad &\implies \quad v_1 \cdot v_2 = 0
\end{aligned}$$

An $n \times n$ matrix A is *orthogonally diagonalizable* if there are an orthogonal matrix P with $P^{-1} = P^T$, and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1} \tag{87}$$

This requires n linearly independent and orthonormal eigenvectors. This is possible when A is symmetric:

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Spectral decomposition of a Symmetric Matrix

Spectral decomposition is possible when A is symmetric:

$$\begin{aligned}
A = PDP^T &= \begin{bmatrix} u_1 \cdots u_n \end{bmatrix} \operatorname{diag}(\lambda_1, \cdots, \lambda_n) \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 u_1 \cdots \lambda_n u_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \\
&= \lambda u_1 u_1^T + \cdots + \lambda_n u_n u_n^T
\end{aligned}$$

Here, each matrix $u_i u_i^T$ is 1×1 and is a **projection matrix** in the sense that, $u_i u_i^T x$, for all $x \in \mathbb{R}^n$, is the orthogonal projection of x onto the subspace spanned by u_i.

SVD of an $m \times n$ matrix A

Let's finally talk about SVD. The key insight is that, even if we can't factorize an $m \times n$ matrix A as $A = PDP^{-1}$, **we can always perform SVD to get $\mathbf{A = QDP^{-1}}$.**

Let A be an $m \times n$ matrix. Then, $A^T A$ is symmetric, and can be orthogonally diagonalized as $A^T A = PDP^T$. Let $\{v_1, \cdots v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$. Let $\lambda_1, \cdots, \lambda_n$ be the corresponding eigenvalues. Then,

$$\begin{aligned} |Av_i|^2 &= (Av_i)^T (Av_i) = v_i^T A^T A v_i \\ &= v_i^T (\lambda_i v_i) \quad \leftarrow Av_i = \lambda_i v_i \\ &= \lambda_i v_i^T v_i = \lambda_i \quad \leftarrow v_i^T v_i = 1 \\ \implies \quad \lambda_i &\geq 0 \end{aligned}$$

The **singular values** of A are the square roots of the eigenvalues of $A^T A$: $\sigma_i = \sqrt{\lambda_i} = \sqrt{|Av_i|^2} = |Av_i|$. Therefore, we can also see that σ_i is the length of the vector Av_i.

Since v_i and $\lambda_j v_j$ are orthogonal while $i \neq j$,

$$(Av_i)^T (Av_j) = v_i^T A^T A v_j = v_i^T (\lambda_j v_j) = 0$$

This shows that $\{Av_1, \cdots, Av_n\}$ is an **orthogonal set**. The lengths of Av_i is given by the corresponding singular values of A. Typically, we arrange the singular values in order of decreasing magnitude, so that:

$$\begin{aligned} \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0 \\ \sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0 \end{aligned}$$

If A has r non-zero singular values, then $\text{Rank}(A) = r$, and $\{Av_1, Av_2, \cdots, Av_r\}$ is an **orthogonal basis** for $\text{Col}(A)$. To show that the basis is $r-$dimensional, rather than $n-$dimensional, we will show that a vector y in $\text{Col}(A)$ can be expressed as a linear combination of $\{Av_1, \cdots, Av_r\}$:

$$\begin{aligned} x &= c^T v = c_1 v_1 + \cdots + c_n v_n \\ y &= Ax = c_1 Av_1 + \cdots c_r Av_r + c_{r+1} Av_{r+1} + \cdots + c_n Av_n \\ &= c_1 Av_1 + \cdots + c_r Av_r + 0 + \cdots + 0 \end{aligned}$$

Thus, y is in $\text{Span}\{Av_1, \cdots, Av_r\}$. Hence, $\text{Rank}(A) = \dim(\text{Col}(A)) = r$.

SVD

Let A be an $m \times n$ matrix of rank r. Then, there is an $m \times n$ matrix Σ,

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

for which the diagonal entries in D are the first r (i.e. the non-zero) singular values of A, where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

Finally, there is an $m \times m$ orthogonal matrix U and an $n \times n$ matrix V such that:

$$A = U\Sigma V^T \tag{88}$$

U and V are not uniquely determined by A, and are referred to as the left- and right-singular vectors of A, respectively. We prove SVD below, starting with the assumption that A is a $m \times n$ matrix with eigenvalues λ_i, corresponding eigenvectors v_i, and r non-zero singular values σ_i. As shown above, $\{Av_1, \cdots, Av_r\}$ will be an orthogonal basis for $\text{Col}(A)$.

Normalize each Av_i to obtain an orthonormal basis U:

$$\begin{aligned} U &= \{u_1, \cdots, u_r\}, \text{ where} \\ u_i &= \frac{Av_i}{|Av_i|} = \frac{Av_i}{\sigma_i} \\ Av_i &= \sigma_i u_i, \quad 1 \leq i \leq r \end{aligned}$$

Extend U to a basis of $\mathbb{R}^m$:

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

Let $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$. Then, U and V are both orthogonal matrices. We also have that

$$AV = \begin{bmatrix} Av_1 & \cdots & Av_r & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \cdots & \sigma_r u_r & 0 & \cdots & 0 \end{bmatrix}$$

Let D be a diagonal matrix with diagonal entries $\sigma_1, \cdots, \sigma_r$. Then,

$$\begin{aligned} U\Sigma &= \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix} D & n-r \text{ columns of } 0 \\ m-r \text{ rows of } 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r u_r & 0 & \cdots & 0 \end{bmatrix} \\ &= AV \end{aligned}$$

$$V^{-1} = V^T \quad \Longrightarrow \quad U\Sigma V^T = AVV^T = A$$

3.2.3 Computing u^+ with SVD

$$\textbf{SVD:} \quad A = U\Sigma V^T = \begin{bmatrix} U_{\text{col}} & U_{\text{null}} \end{bmatrix} \begin{bmatrix} \Sigma_{\text{pos}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{\text{row}} & V_{\text{null}} \end{bmatrix}^T \tag{89}$$

U and V are square matrices that have orthonormal columns: $U^TU = I$ and $V^TV = I$. U_{col} and V_{row} are bases for the column and row spaces of A, respectively, and have dimensions $r = \text{Rank}(A)$. The other 'null' columns are in the nullspaces of A^T (U_{null}) and A (V_{null}).

The pseudoinverse of A is computed as:

$$A^+ = V_{row}(\Sigma_{pos})^{-1}U_{col}^T \tag{90}$$

Where, Σ_{pos} is a diagonal matrix containing the *positive* singular values of A. $A^+b = u^+$ puts u^+ in the row space of A^+.

3.2.4 Computing u^+ with Tychonov Regularization

SVD may not always be the most practical approach for solving the normal equations. When A^TA is singular (the problem here), we can add a small multiple of αI, producing what is known as **Tychonov Regularization**:

$$\begin{aligned} \min \quad & |Au - b|^2 + \alpha|u|^2 \\ \text{s.t.} \quad & (A^TA + \alpha I)\hat{u}_\alpha = A^Tb \end{aligned} \tag{91}$$

This is similar to our original formulation of regularized LSQ, with $B = I$, $d = 0$, and $\alpha|u|^2$ as the regularizing term. When α is small, we prioritize the minimization of $|Au - b|^2$. In the limit that $\alpha \to 0$, we have that

$$\begin{aligned} \lim_{\alpha\to 0} \hat{u}_\alpha &= \lim_{\alpha\to 0}(A^TA + \alpha I)^{-1}A^Tb \\ &= (A^TA)^{-1}A^Tb = A^+b \\ &= u^+ \end{aligned}$$

We don't get u^+ exactly, with small $\alpha > 0$. There is also uncertainty in b. The presence of a small α in $(A^TA + \alpha I)^{-1}A^T$ introduces small, non-zero singular values along the diagonal of our singular values matrix Σ in SVD. Importantly, $\frac{\sigma}{\sigma^2+\alpha} \to \frac{1}{\sigma}$ while $\alpha > 0$, but stays zero if $\sigma = 0$.

3.3 Tychonov Regularization

This section discusses how to choose α. The key ideas are as follows:

1. There is noise in measurement of b, represented by an error e.
2. e is amplified when A is ill-conditioned (e.g. singular A^TA) and $\alpha = 0$ (no regularization). This is why we use regularization.
3. Choosing an appropriate α, we can stabilize the problem by adding αI. However, if α is too small, then e will still grow with A^{-1}. If α is too large, we lose accuracy on $\hat{u}$.

We will refer to multiple versions of the solution using the notation $\hat{u}_i^j$, where i represents our choice of α, and j the expected size of uncertainty in b. For instance, $\hat{u}_0^0$ refers to the solution without regularization or measurement error. On the other hand, $\hat{u}_\alpha^e$ is the solution attained from data with measurement error e, using regularization parameter α.

3.3.1 Error bounds in $\hat{u}$

Error bounds when A is a scalar.

Let's first see how different we expect the solution $\hat{u}$ to be in the two extreme cases: $\hat{u}_0^0$ (no error, no regularization) and $\hat{u}_\alpha^e$ (error and regularization).

$$\hat{u}_0^0 - \hat{u}_\alpha^e = \hat{u}_0^0 - \hat{u}_0^0 + \hat{u}_0^0 - \hat{u}_\alpha^e \tag{92}$$

$$|\hat{u}_0^0 - \hat{u}_0^\alpha| \leq C\alpha|b| \qquad |\hat{u}_\alpha^0 - \hat{u}_\alpha^e| \leq \frac{|e|}{2\sqrt{\alpha}} \tag{93}$$

Where, C is a constant that we will define below. Let's first prove the error bounds in Eq. 93 when A is a scalar s. Then, $A^T A = s^2$. When noise is absent from measurements:

$$\text{Without regularization:} \quad s^2\hat{u}_0^0 = sb \tag{94}$$

$$\text{With regularization:} \quad (s^2+\alpha)\hat{u}_\alpha^0 = sb \tag{95}$$

$$\begin{aligned}\hat{u}_0^0 - \hat{u}_\alpha^0 &= \frac{b}{s} - \frac{sb}{s^2+\alpha} \\ &= \frac{b(s^2+\alpha) - s^2 b}{s(s^2+\alpha)} \\ &= \frac{\alpha b}{s(s^2+\alpha)} \leq C\alpha b, \quad C = \frac{1}{s^3+s\alpha}\end{aligned} \tag{96}$$

Eq. 96 shows that, in the absence of measurement noise, the error associated with regularization is proportional to α, i.e. $O(\alpha)$. C scales steeply with $1/s^3$. We perform the same procedure to compare the effect of measurement noise:

$$\text{Without noise:} \quad (s^2+\alpha)\hat{u}_\alpha^0 = sb \tag{97}$$

$$\text{With noise:} \quad (s^2+\alpha)\hat{u}_\alpha^e = s(b-e) \tag{98}$$

$$\begin{aligned}\hat{u}_\alpha^0 - \hat{u}_\alpha^e &= \frac{sb}{s^2+\alpha} - \frac{s(b-e)}{s^2+\alpha} \\ &= \frac{s}{s^2+\alpha}e\end{aligned} \tag{99}$$

The maximum value of $s/(s^2+\alpha)$ will determine the bound. We can find this value by setting its derivative to zero.

$$\begin{aligned}0 = \frac{d}{ds}\frac{s}{s^2+\alpha} &= (s^2+\alpha)^{-1} - s(s^2+\alpha)^{-2}(2s) \\ &= \frac{1}{s^2+\alpha} - \frac{2s^2}{(s^2+\alpha)^2} \\ \frac{2s^2}{(s^2+\alpha)^2} = \frac{1}{s^2+\alpha} &\implies 2s^2 = s^2+\alpha, \quad \therefore s = \sqrt{\alpha} \\ \therefore \frac{s}{s^2+\alpha} &\leq \frac{1}{2\sqrt{\alpha}}\end{aligned} \tag{100}$$

Error bounds when A is a matrix.

SVD of A produces orthonormal bases $\{u_1, u_2, \cdots\}$ and $\{v_1, v_2, \cdots\}$, with $Av_j = \sigma_j u_j$ and $A^T u_j = \sigma_j v_j$. Note that, unlike before, we are not using v_i to refer to the eigenvectors of A; instead, they are the right-singular vectors of A.

We can apply SVD to rewrite the regularized LSQ problem in terms of u_j and v_j:

$$
\begin{aligned}
(A^TA + \alpha I)\hat{u}_\alpha^0 &= A^T b \\
b = B_1 u_1 + B_2 u_2 + \cdots, \quad \hat{u}_\alpha^0 &= U_1 v_1 + U_2 v_2 + \cdots \\
(A^TA + \alpha I)U_j v_j &= A^T B_j u_j \qquad (101) \\
U_j A^T A v_j + U_j \alpha I v_j &= \sigma_j B_j v_j \\
U_j \sigma_j A^T u_j + U_j \alpha v_j &= \\
U_j \sigma_j (\sigma_j v_j) + U_j \alpha v_j &= \\
U_j v_j (\sigma_j^2 + \alpha) &= \\
U_j (\sigma_j^2 + \alpha) &= \sigma_j B_j \qquad (102) \\
\implies U_j &= \frac{\sigma_j}{\sigma_j(\sigma_j^2 + \alpha)} B_j = \frac{\sigma_j}{\sigma_j^3 + \sigma_j \alpha} B_j \qquad (103)
\end{aligned}
$$

Eq. 103 is identical to what we saw for the scalar case (Eq. 96) if $s = \sigma_j$. Likewise, the results in the presence of noise are also similar to the scalar case. To get the total difference between either solution group, we need to sum each of the U_j coefficients:

$$\hat{u}_0^0 - \hat{u}_\alpha^0 = \sum_{j=1}^{\infty} \frac{\alpha B_j}{\sigma_j(\sigma_j^2 + \alpha)} v_j \qquad (104)$$

$$\hat{u}_\alpha^0 - \hat{u}_\alpha^e = \sum_{j=1}^{\infty} \frac{s_j E_j}{s_j^2 + \alpha} v_j \qquad (105)$$

Where, in the presence of noise, we replace b with $e = E_1 v_1 + E_2 v_2 \cdots$. The norm of either sum is a sum of squares wherein the v_j terms disappear because they are orthonormal: $\sum_{j=1}^{\infty} v_j^2 = 1$. Hence,

$$|\hat{u}_0^0 - \hat{u}_\alpha^0|^2 = \sum_{j=1}^{\infty} \left(\frac{\alpha B_j}{\sigma_j(\sigma_j^2 + \alpha)} \right)^2 \qquad (106)$$

$$|\hat{u}_\alpha^0 - \hat{u}_\alpha^e|^2 = \sum_{j=1}^{\infty} \left(\frac{s_j E_j}{s_j^2 + \alpha} \right)^2 \qquad (107)$$

The u_j terms are also orthonormal, so

$$|b^2| = B_1^2 u_1^2 + B_2^2 u_2^2 + \cdots = \sum_{j=1} |B_j|^2$$
$$|e^2| = E_1^2 u_1^2 + E_2^2 u_2^2 + \cdots = \sum_{j=1} |E_j|^2$$

This gives

$$|\hat{u}_0^0 - \hat{u}_\alpha^0| \leq \left(\sum_{j=1}^{\infty} \frac{1}{\sigma_j(\sigma_j^2 + \alpha)} \right) \alpha |b| \approx \frac{\alpha}{\sum \sigma_j^3} |b| = \frac{\alpha}{s_{min}^3} |b| \tag{108}$$

$$|\hat{u}_\alpha^0 - \hat{u}_\alpha^e| \leq \left(\sum_{j=1}^{\infty} \frac{s_j}{s_j^2 + \alpha} \right)^2 |e| = \frac{|e|}{2\sqrt{\alpha}} \tag{109}$$

In Eq. 109, we made the substitution $s_j = \sqrt{\alpha}$, by analogy with the scalar case (see Eq. 100).

What is a good α?

We can find a possible α by finding the value that equalizes the two parts of overall error (Eq. 93):

$$C\alpha|b| = \frac{|e|}{2\sqrt{\alpha}} \tag{110}$$

$$2\alpha^{3/2} = \frac{|e|}{C|b|} \implies \alpha = \left(\frac{|e|}{2C|b|} \right)^{2/3} \tag{111}$$

Then, the overall error is found by plugging this value into the two parts of error, and summing them:

$$\begin{aligned}
|\hat{u}_0^0 - \hat{u}_\alpha^e| &\leq C\alpha|b| + \frac{|e|}{2\sqrt{\alpha}} \\
&\leq \frac{|e| + 2C\alpha^{3/2}|b|}{2\sqrt{\alpha}} \\
&\leq \frac{|e| + C|b| \left(\frac{|e|}{C|b|} \right)}{2 \left(\frac{|e|}{2C|b|} \right)^{1/3}} \\
&\leq \frac{2|e| \left(2C|b|\right)^{1/3}}{2|e|^{1/3}} = \frac{|e| \left(2C|b|\right)^{1/3}}{|e|^{1/3}} \\
&\leq |e|^{2/3} \left(2C|b|\right)^{1/3} = \left[|e|^2 |b| C \right]^{1/3}
\end{aligned}$$

Strang writes that this rule suggests that "we know more than we really do." I'm not sure what to make of this, but my guess is that this is due to the exponent of $|e|$ being $2/3$, rather than ≥ 1. Perhaps, this means that this choice of α underestimates how much error there really is in our measurements.

References

[1] David C. Lay. *Linear algebra and its applications*. Addison-Wesley, Boston, 4th ed edition, 2012. OCLC: ocn693750928.

[2] Gilbert Strang. *Computational science and engineering*. Wellesley-Cambridge Press, Wellesley, Mass, 2. print edition, 2012. OCLC: 935765561.