We make further observations.

6.1 Condition numbers

In this section we consider condition numbers, a measure of the sensitivity to perturbations of certain numerical problems. We look in particular at the conditioning of the least-squares problem.

6.1.1 Pseudoinverses

We have seen that the least-squares problem provides a solution concept for overdetermined systems. This leads to the following generalization of the matrix inverse.

**Definition (Pseudoinverse):** Let \( A \in \mathbb{R}^{m \times n} \) be a matrix with SVD \( A = U \Sigma V^T \) and singular values \( \sigma_1 \geq \cdots \geq \sigma_r > 0 \). A pseudoinverse \( A^+ \in \mathbb{R}^{m \times n} \) is defined as

\[
A^+ = V \Sigma^+ U^T
\]

where \( \Sigma^+ \) is the diagonal matrix with diagonal entries \( \sigma_1^{-1}, \ldots, \sigma_r^{-1} \).

While it is not obvious from the definition (Exercise: Why? \(<\)\)), the pseudoinverse is in fact unique (https://en.wikipedia.org/wiki/Proofs_involving_the_Moore%E2%80%93Penrose_inverse#Proof_of_uniqueness). Note further that

\[
AA^+ = U \Sigma V^T V \Sigma^+ U^T = UU^T
\]

and

\[
A^+ A = V \Sigma^+ U^T U \Sigma V^T = VV^T.
\]
Three important cases:

1. In the square nonsingular case, both products are the identity and we necessarily have $A^+ = A^{-1}$ by the Existence of an Inverse Lemma.
2. If $A$ has full column rank $m \leq n$, then $r = m$ and $A^+A = I_{m \times m}$.
3. If $A$ has full row rank $n \leq m$, then $r = n$ and $AA^+ = I_{n \times n}$.

In the second case above, we recover our solution to the least-squares problem in the overdetermined case.

**Lemma (Pseudoinverse and least squares):** Let $A \in \mathbb{R}^{n \times m}$ of full column rank $m$ with $m \leq n$. Then

$$A^+ = (A^T A)^{-1} A^T.$$  

In particular, the solution to the least-square problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|$$

is $x^* = A^+b$.

**Proof idea:** We use the SVD definition and check that the two sides are the same.

**Proof:** We first note that $A^TA$ is nonsingular. Indeed, if $A^TAx = 0$, then $x^T A^TAx = \|Ax\|^2 = 0$ so that $Ax = 0$ by the point-separating property of the vector norm. In turn, since the columns of $A$ are linearly independent by assumption, this implies $x = 0$. Hence the columns of $A^TA$ are also linearly independent.

Now let $A = U\Sigma V^T$ be an SVD of $A$ (with positive singular values). We then note that

$$(A^TA)^{-1} A^T = (V\Sigma U^T U\Sigma V^T)^{-1} V\Sigma U^T = V\Sigma^{-2} V^T V\Sigma U^T = A^+$$

as claimed.

The second claim follows from the normal equations. □

**NUMERICAL CORNER** In Julia, the pseudoinverse of a matrix can be computed using the function `pinv` ([https://docs.julialang.org/en/v1/stdlib/LinearAlgebra/#LinearAlgebra.pinv](https://docs.julialang.org/en/v1/stdlib/LinearAlgebra/#LinearAlgebra.pinv)).

In [1]:  #Julia version: 1.5.1
      using Plots, LinearAlgebra

In [2]:  M = [1.5 1.3; 1.2 1.9; 2.1 0.8]

Out[2]: 3×2 Array{Float64,2}:
  1.5  1.3
  1.2  1.9
  2.1  0.8
6.1.2 Conditioning of matrix-vector multiplication

We define the condition number of a matrix and show that it captures some information about the sensitivity to perturbations of matrix-vector multiplications. First an exercise.

Exercise: Let \( A \in \mathbb{R}^{n \times n} \) be nonsingular with SVD \( A = U \Sigma V^T \) where the singular values satisfy \( \sigma_1 \geq \cdots \geq \sigma_n > 0 \). Show that

\[
\min_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \min_{\mathbf{y} \neq 0} \frac{\|\mathbf{y}\|}{\|A^{-1}\mathbf{y}\|} = \sigma_n = 1/\|A^+\|_2.
\]

Definition (Condition number of a matrix): The condition number (in the induced 2-norm) of a matrix \( A \in \mathbb{R}^{n \times m} \) is defined as

\[
\kappa_2(A) = \|A\|_2 \|A^+\|_2.
\]

In the square nonsingular case, this reduces to

\[
\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}
\]

where we used the exercise above. In words, \( \kappa_2(A) \) is the ratio of the largest to the smallest stretching under \( A \).

Theorem (Relative conditioning of matrix-vector multiplication): Let \( A \in \mathbb{R}^{n \times n} \) be nonsingular. Then, for any \( \mathbf{x} \in \mathbb{R}^n \),

\[
\lim_{\delta \to 0} \sup_{0 < \|\mathbf{d}\| \leq \delta} \frac{\|A(\mathbf{x} + \mathbf{d}) - A\mathbf{x}\|/\|A\mathbf{x}\|}{\|\mathbf{d}\|/\|\mathbf{x}\|} = \max_{\mathbf{d} \neq 0} \frac{\|A(\mathbf{x} + \mathbf{d}) - A\mathbf{x}\|/\|A\mathbf{x}\|}{\|\mathbf{d}\|/\|\mathbf{x}\|} \leq \kappa_2(A)
\]

and the inequality is tight.

The ratio above measures the worst rate of relative change in \( A\mathbf{x} \) under infinitesimal perturbations of \( \mathbf{x} \). The theorem says that when \( \kappa_2(A) \) is large, a case referred to as ill-conditioning, large relative changes in \( A\mathbf{x} \) can be obtained from relatively small perturbations to \( \mathbf{x} \). In words, a matrix-vector product is potentially sensitive to perturbations when the matrix is ill-conditioned.
Proof: Write
\[
\frac{\|A(x + d) - Ax\|/\|Ax\|}{\|d\|/\|x\|} = \frac{\|Ad\|/\|Ax\|}{\|d\|/\|x\|} = \frac{\|A(d/\|d\|)\|}{\|A(x/\|x\|)\|} \leq \frac{\sigma_1}{\sigma_n}
\]
where we used the exercise above.

In particular, we see that the ratio can achieve its maximum by taking \(d\) and \(x\) to be the right singular vectors corresponding to \(\sigma_1\) and \(\sigma_n\) respectively. □

If we apply the theorem to the inverse instead, we get that the relative conditioning of the nonsingular linear system \(Ax = b\) to perturbations in \(b\) is also \(\kappa_2(A)\). The latter can be large in particular when the columns of \(A\) are close to linearly dependent.

**NUMERICAL CORNER** In Julia, the condition number of a matrix can be computed using the function `cond` (https://docs.julialang.org/en/v1/stdlib/LinearAlgebra/#LinearAlgebra.cond).

For example, orthogonal matrices have condition number 1:

```
In [5]: Q = [1/sqrt(2) 1/sqrt(2); 1/sqrt(2) -1/sqrt(2)]
Out[5]: 2×2 Array{Float64,2}:
        0.707107  0.707107
        0.707107 -0.707107

In [6]: cond(Q)
Out[6]: 1.0000000000000002
```

And matrices with nearly linearly dependent columns have large condition numbers:

```
In [2]: eps = 1e-6
A = [1/sqrt(2) 1/sqrt(2); 1/sqrt(2) 1/sqrt(2)+eps]
Out[2]: 2×2 Array{Float64,2}:
        0.707107  0.707107
        0.707107  0.707108

In [3]: cond(A)
Out[3]: 2.8284291245366517e6
```

Let’s look at the SVD of \(A\).
We compute the solution to $Ax = b$ when $b$ is the right singular vector corresponding to the largest singular value.

We make a small perturbation in the direction of the second right singular vector.

The relative change in solution is:
\begin{verbatim}
In [15]: (norm(x-xp)/norm(x))/(norm(b-bp)/norm(b))
Out[15]: 2.828429124665918e6
\end{verbatim}

Note that this is exactly the condition number of $A$.

### 6.1.3 Back to the least-squares problem [optional]

We return to the least-squares problem

$$\min_{x \in \mathbb{R}^m} \|Ax - b\|$$

where

$$A = \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$ 

We showed that the solution satisfies the normal equations

$$A^T Ax = A^T b.$$ 

As we show next, the condition number of $A^T A$ can be much larger than that of $A$ itself.

**Lemma (Condition number of $A^T A$):** Let $A \in \mathbb{R}^{n \times m}$ have full column rank. The

$$\kappa_2(A^T A) = \kappa_2(A)^2.$$ 

**Proof idea:** We use the SVD.

**Proof:** Let $A = U\Sigma V^T$ be an SVD of $A$ with singular values $\sigma_1 \geq \cdots \geq \sigma_m > 0$. Then

$$A^T A = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T.$$ 

In particular the latter expression is an SVD of $A^T A$, and hence the condition number of $A^T A$ is

$$\kappa_2(A^T A) = \frac{\sigma_1^2}{\sigma_m^2} = \kappa_2(A)^2.$$ 

\[\square\]

**NUMERICAL CORNER** We give a quick example.
In [15]:
\[
A = \begin{bmatrix}
1. & 1. & 1. & 1. & 1.\\
101. & 102. & 103. & 104. & 105.
\end{bmatrix}
\]

Out[15]:
5x2 Array{Float64,2}:
1.0  101.0
1.0  102.0
1.0  103.0
1.0  104.0
1.0  105.0

In [16]:
cond(A)

Out[16]:
7503.817028686117

In [17]:
cond(A'*A)

Out[17]:
5.630727000263108e7

This observation - and the resulting increased numerical instability - is one of the reasons we previously developed an alternative approach to the least-squares problem. Quoting [Sol, Section 5.1]:

> Intuitively, a primary reason that \( \text{cond}(A^T A) \) can be large is that columns of \( A \) might look “similar” [...] If two columns \( \mathbf{a}_i \) and \( \mathbf{a}_j \) satisfy \( \mathbf{a}_i \approx \mathbf{a}_j \), then the least-squares residual length \( \| \mathbf{b} - A\mathbf{x} \|_2 \) will not suffer much if we replace multiples of \( \mathbf{a}_i \) with multiples of \( \mathbf{a}_j \) or vice versa. This wide range of nearly—but not completely—equivalent solutions yields poor conditioning. [...] To solve such poorly conditioned problems, we will employ an alternative technique with closer attention to the column space of \( A \) rather than employing row operations as in Gaussian elimination. This strategy identifies and deals with such near-dependencies explicitly, bringing about greater numerical stability.
We quote without proof a theorem from [Ste (https://epubs.siam.org/doi/book/10.1137/1.9781611971408), Theorem 4.2.7] which provides further light on this issue.

**Theorem (Accuracy of computed least-squares solutions):** Let $x^*$ be the solution of the least-squares problem $\min_{x \in \mathbb{R}^m} ||Ax - b||$. Let $x_{NE}$ be the solution obtained by forming and solving the normal equations in floating-point arithmetic (https://en.wikipedia.org/wiki/Floating-point_arithmetic) with rounding unit $\varepsilon_M$. Then $x_{NE}$ satisfies

$$\frac{||x_{NE} - x^*||}{||x^*||} \leq \gamma_{NE} \kappa_2^2(A) \left(1 + \frac{||b||}{||A||_2||x^*||}\right) \varepsilon_M.$$ 

Let $x_{QR}$ be the solution obtained from a QR factorization in the same arithmetic. Then

$$\frac{||x_{QR} - x^*||}{||x^*||} \leq 2\gamma_{QR} \kappa_2(A) \varepsilon_M + \gamma_{NE} \kappa_2^2(A) \frac{||r^*||}{||A||_2||x^*||} \varepsilon_M$$

where $r^* = b - Ax^*$ is the residual vector. The constants $\gamma$ are slowly growing functions of the dimensions of the problem.

To explain, let’s quote [Ste (https://epubs.siam.org/doi/book/10.1137/1.9781611971408), Section 4.2.3] again:

The perturbation theory for the normal equations shows that $\kappa_2^2(A)$ controls the size of the errors we can expect. The bound for the solution computed from the QR equation also has a term multiplied by $\kappa_2^2(A)$, but this term is also multiplied by the scaled residual, which can diminish its effect. However, in many applications the vector $b$ is contaminated with error, and the residual can, in general, be no smaller than the size of that error.

**NUMERICAL CORNER** Here is a numerical example taken from [TB (https://books.google.com/books/about/Numerical_Linear_Algebra.html?id=JaPtxOytY7kC), Lecture 19]. We will approximate the following function with a polynomial.
In [18]:
\[
\begin{align*}
\text{n} & = 100 \\
\text{t} & = (0: n-1) / (n-1) \\
\text{b} & = \exp(\sin(4\times t)) \\
\text{plot}(t, b, \text{legend=false, lw}=2)
\end{align*}
\]

Out[18]:

```
```

In [19]:
\[
\begin{align*}
\text{m} & = 17 \\
\text{A} & = [t[i].^\text{(j-1)} \text{ for i=1:n, j=1:m}];
\end{align*}
\]

The condition numbers of $A$ and $A^T A$ are both high in this case.

In [20]:
\[
\begin{align*}
\text{@show cond(A)} \\
\text{@show cond(A'*A)};
\end{align*}
\]

\[
\begin{align*}
\text{cond(A)} & = 7.558243605585787e11 \\
\text{cond(A' * A)} & = 6.812930320935918e17
\end{align*}
\]

We first use the normal equations and plot the residual vector.
We then use the `qr` function of Julia to compute the QR solution instead. Note that `Matrix` is used to transform the factors obtained from `qr` into regular arrays.
6.2 Proof of Spectral theorem [optional]

When $A$ is symmetric, that is, $a_{ij} = a_{ji}$ for all $i, j$, a remarkable result is that $A$ is similar to a diagonal matrix by an orthogonal transformation. Put differently, there exits an orthonormal basis of $\mathbb{R}^d$ made of eigenvectors of $A$.

**Theorem (Spectral Theorem):** Let $A \in \mathbb{R}^{d \times d}$ be a symmetric matrix, that is, $A^T = A$. Then $A$ has $d$ orthonormal eigenvectors $q_1, \ldots, q_d$ with corresponding (not necessarily distinct) real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. In matrix form, this is written as the matrix factorization

$$A = Q\Lambda Q^T = \sum_{i=1}^{d} \lambda_i q_i q_i^T$$

where $Q$ has columns $q_1, \ldots, q_d$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$. We refer to this factorization as a spectral decomposition of $A$. 
One might hope that the SVD of a symmetric matrix would produce identical left and right singular vectors, thereby providing the desired eigendecomposition. However that is not the case.

Exercise: Compute the SVD of $A = [-1]$. \(<\>

However, the same roadmap that led to the SVD can be made to produce an eigendecomposition.

Exercise: Consider the block matrices

\[
\begin{pmatrix} y \\ z \end{pmatrix}, \quad \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}
\]

where $y \in \mathbb{R}^{d_1}$, $z \in \mathbb{R}^{d_2}$, $X_{11} \in \mathbb{R}^{d_1 \times d_1}$, $X_{12} \in \mathbb{R}^{d_1 \times d_2}$, $X_{21} \in \mathbb{R}^{d_2 \times d_1}$, and $X_{22} \in \mathbb{R}^{d_2 \times d_2}$.

Show that

\[
\begin{pmatrix} y \\ z \end{pmatrix}^T \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = y^T Ay + y^T Bz + z^T Cy + z^T Dz.
\]

\(<\>

Proof idea (Spectral Theorem): Use a greedy sequence, as in the SVD derivation, this time maximizing the quadratic form (https://en.wikipedia.org/wiki/Quadratic_form#Real_quadratic_forms) $\langle v, Av \rangle$. How is this quadratic form related to eigenvalues? Note that, for a unit eigenvector $v$ with eigenvalue $\lambda$, we have $\langle v, Av \rangle = \langle v, \lambda v \rangle = \lambda$.

Proof (Spectral Theorem): We proceed by induction.

A first eigenvector: Let $A_1 = A$ and

\[
v_1 \in \arg \max \{ \langle v, A_1 v \rangle : \|v\| = 1 \}
\]

and

\[
\lambda_1 = \max \{ \langle v, A_1 v \rangle : \|v\| = 1 \}.
\]

Complete $v_1$ into an orthonormal basis of $\mathbb{R}^d$, $v_1$, $\hat{v}_2$, ..., $\hat{v}_d$, and form the block matrix

\[
\hat{W}_1 = \begin{pmatrix} v_1 \\ \hat{V}_1 \end{pmatrix}
\]

where the columns of $\hat{V}_1$ are $\hat{v}_2$, ..., $\hat{v}_d$. Note that $\hat{W}_1$ is orthogonal by construction.
**Getting one step closer to diagonalization:** We show next that $\hat{W}_1$ gets us one step closer to a diagonal matrix by similarity transformation. Note first that

$$\hat{W}_1 A_1 \hat{W}_1^T = \begin{pmatrix} \lambda_1 & w_1^T \\ w_1 & A_2 \end{pmatrix}$$

where $w_1 = \hat{V}_1^T A_1 v_1$ and $A_2 = \hat{V}_1^T A_2 \hat{V}_1$. The key claim is that $w_1 = 0$. This follows from a contradiction argument. Indeed, suppose $w_1 \neq 0$ and consider the unit (Exercise: Why?) vector

$$z = \hat{W}_1^T \times \frac{1}{\sqrt{1 + \delta^2 \|w_1\|^2}} \begin{pmatrix} 1 \\ \delta w_1 \end{pmatrix}$$

which, by the exercise above, achieves objective value

$$z^T A_1 z = \frac{1}{1 + \delta^2 \|w_1\|^2} \left( 1 \right)^T \begin{pmatrix} \lambda_1 & w_1^T \\ w_1 & A_2 \end{pmatrix} \begin{pmatrix} 1 \\ \delta w_1 \end{pmatrix}$$

$$= \frac{1}{1 + \delta^2 \|w_1\|^2} \left( \lambda_1 + 2 \delta \|w_1\|^2 + \delta^2 w_1^T A_2 w_1 \right).$$

In the proof of the *Left Singular Vectors are Orthogonal Lemma*, we showed that for $\epsilon \in (0, 1)$

$$\frac{1}{\sqrt{1 + \epsilon^2}} \geq 1 - \epsilon^2 / 2.$$

So for $\delta$ small enough

$$z^T A_1 z \geq (\lambda_1 + 2 \delta \|w_1\|^2 + \delta^2 w_1^T A_2 w_1)(1 - \delta^2 \|w_1\|^2 / 2)$$

$$\geq \lambda_1 + 2 \delta \|w_1\|^2 + C \delta^2$$

$$> \lambda_1$$

where $C \in \mathbb{R}$ depends on $w_1$ and $A_2$. That gives the desired contradiction. So, letting $\hat{W}_1 = \hat{W}_1^T$, we have

$$W_1 A_1 W_1^T = \begin{pmatrix} \lambda_i & 0 \\ 0 & A_2 \end{pmatrix}.$$

Finally note that $A_2 = \hat{V}_1^T A_1 \hat{V}_1$ is symmetric

$$A_2^T = (\hat{V}_1^T A_1 \hat{V}_1)^T = \hat{V}_1^T A_1^T \hat{V}_1 = \hat{V}_1^T A_1 \hat{V}_1 = A_2$$

by the symmetry of $A_1$ itself.
Next step of the induction: Apply the same argument to the symmetric matrix $A_2 \in \mathbb{R}^{(d-1) \times (d-1)}$, let $\hat{W}_2 \in \mathbb{R}^{(d-1) \times (d-1)}$ be the corresponding orthogonal matrix, and define $\lambda_2$ and $A_3$ through the equation

$$\hat{W}_2 A_2 \hat{W}_2^T = \begin{pmatrix} \lambda_2 & 0 \\ 0 & A_3 \end{pmatrix}.$$ 

Define the block matrix

$$W_2 = \begin{pmatrix} 1 & 0 \\ 0 & \hat{W}_2 \end{pmatrix}$$

and observe that

$$W_2 W_1 A_1 W_1^T W_2^T = W_2 \begin{pmatrix} \lambda_1 & 0 \\ 0 & A_2 \end{pmatrix} W_2^T = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \hat{W}_2 A_2 \hat{W}_2^T \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix}.$$ 

Proceeding similarly by induction gives the claim. $\square$