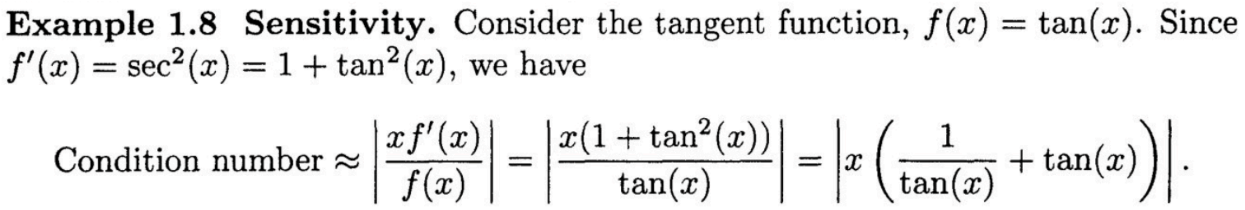
Sample questions from lectures/slides of Chapters 1, 2, 3, & 7

Chapter 1

Q 1.



Example 1.8 Sensitivity. Consider the tangent function, $f(x) = \tan(x)$. Since $f'(x) = \sec^2(x) = 1 + \tan^2(x)$, we have

$$\text{Condition number} \approx \left|\frac{xf'(x)}{f(x)}\right| = \left|\frac{x(1+\tan^2(x))}{\tan(x)}\right| = \left|x\left(\frac{1}{\tan(x)} + \tan(x)\right)\right|.$$

Compute the condition number using the finite-difference approximation formula for x = 0.1 and x = π + 0.1.

Answer:
According to the finite difference approximation formula, the condition number at x = 0.1 is

$|x\, f'(x)\, /\, f(x)\,|$ = $|x\,(1/\tan x\ +\ \tan x)|$ = $|0.1\,(1/\tan 0.1\ +\ \tan 0.1)|$ ≈ 1.01.

The condition number at x = π + 0.1 is

$|(\pi\ +\ 0.1)\,(1/\tan(\pi\ +\ 0.1)\ +\ \tan(\pi\ +\ 0.1))|$ = 32.63.

Q 2.
Now let’s compute the condition number using the ratio between forward and backward error with Δx = 0.01.

Answer:
For x = 0.1 and Δx = 0.01, the condition number is

$$\frac{|\Delta y/y|}{|\Delta x/x|} = \frac{|(\tan 0.11 - \tan 0.1)/\tan 0.1\,|}{0.01/0.1} \approx 1.01.$$

For x = $\pi\ +\ 0.1$ and Δx = 0.01, the condition number is

$$\frac{|\Delta y/y|}{|\Delta x/x|} = \frac{|(\tan\,(\pi{+}0.1{+}0.01) - \tan\,(\pi{+}0.1))/\tan(\pi{+}0.1)\,|}{0.01/(\pi{+}0.1)} \approx 32.67.$$

Q 3.

Which of the following two expressions is better to implement? Explain why. Here, $x - y = \epsilon > 0$ is smaller than the machine precision, but 2ϵ is larger than the machine precision.

a) $fl(x - y) + fl(x - y)$
b) $fl((x - y) + (x - y))$

Answer:

It is better to implement (b). The expression (b) means the number $(x - y) + (x - y)$ is calculated first before being represented in the floating-point number system, which results in 2ϵ . However, in (a), $fl(x - y)$ results in 0.

Extra note: Although ϵ is quite small, it can be a significant problem if such errors accumulate over many such operations.

Chapter 2

Q 1.

If for a square matrix **A** and a non-trivial vector **z**, we have **Az** = **0**, then which of the following are true?

a) **A** is full rank
b) **A** has linearly dependent columns
c) **A** is singular.

Answer:

b and c.

Extra note: A vector being non-trivial means it has at least one non-zero element. Only a singular matrix can annihilate a non-trivial vector, which is equivalent to **A** having linearly dependent columns. It is also equivalent to **A** being rank deficient.

Q 2.

In what case does a square linear system have a unique solution?

Answer:

It is when the square matrix is nonsingular.

Q 3.

For a square linear system, what kind of transformation of the problem leaves the solution unchanged?

Answer:

The solution remains unchanged if both sides of the system are premultiplied by a nonsingular matrix. In other words, both of the following systems have the same solution:

1) **Ax** = **b**,
2) **MAx** = **Mb**, for a nonsingular **M**.

Q 4.

For a square linear system, does the solution change if the nonsingular matrix is post-multiplied by another matrix? Is the solution recoverable?

Answer:

The solution does change. The system **AMx** = **b** has a different solution than **Ax** = **b**.

However, the solution is recoverable by premultiplying the solution by **M**:

The solution to **AMx** = **b** is **x** = $\mathbf{M}^{-1}\mathbf{A}^{-1}\mathbf{b}$, and pre-multiplying it by **M**, we get $\mathbf{A}^{-1}\mathbf{b}$, which is the solution to **Ax** = **b**.

Q 5.

What is the computational complexity of back-substitution?

Answer:

If the nonsingular matrix **A** of the system **Ax** = **b** is n by n, then the computational complexity of back-substitution is $O(n^2)$.

Q 6.

When is the square linear system **A**x = **b** more ill-conditioned? When **A**’s columns are close to becoming linearly dependent or far from becoming linearly dependent. Explain.

Answer:

The square linear system **A**x = **b** is more ill-conditioned when A’s columns are close to becoming linearly dependent, which indicates close to becoming a singular matrix. In this case, **A** has a large condition number, which makes **A**x = **b** more ill-conditioned.

Q 7.

Is it possible to LU factorize the following matrix? How?

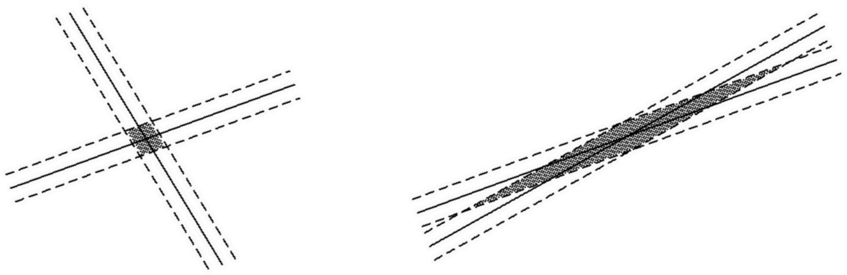
$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Answer:

Yes. By pivoting or pre-multiplying with a permutation matrix to achieve row interchange.

Q 8.

A 2-by-2 square linear system can be represented as two linear equations, each representing a straight line, where the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is the intersection of the two straight lines. The following two represent two different 2-by-2 square linear systems, depicted through the solid straight lines. Which one is more ill-conditioned? Why?



Answer:

The right system is more ill-conditioned. This is because the two straight lines corresponding to its two equations are nearly parallel. If they are exactly parallel, the square matrix will become singular, and its condition number will be infinite. So, as the straight lines become close to parallel, they are approaching singularity with a high condition number.

Q 9.

In the above figures, what do the dashed lines and the gray parallelograms represent? What do they indicate about the solution to their corresponding square linear system?

Answer:

The dashed lines here represent perturbation or error in determining each straight line. Both left and right systems have the same amount of inaccuracy in determining the straight lines, depicted by the dashed lines being away from the solid line by the same amount. This creates inaccuracy in determining the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, which is depicted by the gray parallelograms. Although the inaccuracy in determining the output (the solid lines) is the same in both cases, the inaccuracy in determining the solution (the parallelograms) is much greater for the right system. This is because the right system is more ill-conditioned, causing a larger area of the parallelogram and a higher sensitivity in determining the solution.

Q 10.

Which matrix has a higher condition number?

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Answer:

The matrix condition number is defined by the ratio between the maximum stretching and the minimum shrinking applied by the matrix. For **A**, it is 3/2 = 1.5, and for **B**, it is 2/0.5 = 4. Hence, B has a higher condition number.

Chapter 3

Q 1.

The least squares problem for an overdetermined linear system is given as follows:

$$\min_{\boldsymbol{x}} \|\boldsymbol{r}\|_2^2 = \min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

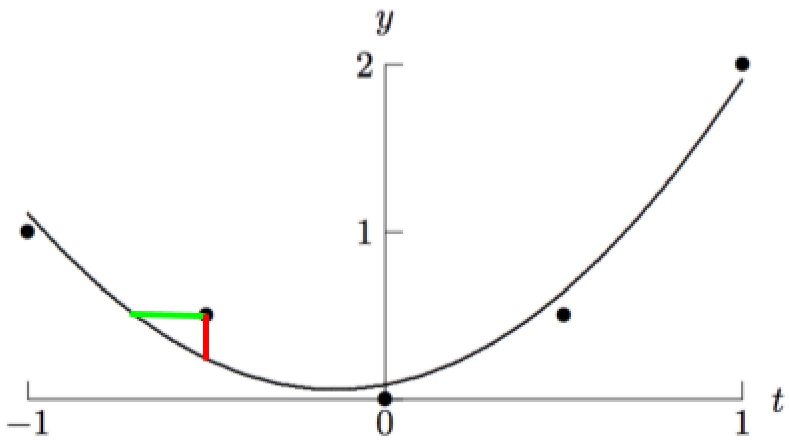
where **r** is the residual vector. Is the minimum of the squared norm of the residual vector zero? Explain your answer.

Answer:

The minimum of the squared norm of the residual vector (given by the LHS) is obtained by solving this minimization problem. But at the solution, the residual vector may not be reduced to a zero vector. It is because **A** is overdetermined; there are more equations than unknowns. The system of equations **b** = **Ax** may not be solved.

Q 2.

The following figure shows five data points (t_i, y_i) and a quadratic polynomial fitting these points. Which of the distances represents an element of the residual vector? The green or the red one? Explain your answer.



Answer:

The residual vector r = b - Ax contains distance in the output, which is given in the y-axis above. Hence, the red distance is an element of the residual vector.

Q 3.

When do we have a unique solution to the least squares problem $\mathbf{Ax} \simeq \mathbf{b}$? How to pick a solution when there are infinitely many solutions?

Answer:

We have a unique solution to the least squares problem when **A** is full rank. If **A** is rank deficient, we have infinitely many solutions. In that case, it is often useful to choose the solution **x** with the minimum norm of **x**. This is because, in real systems, **x** may represent a variable that is costly, such as current. If we can solve the same problem with many different amounts of currents, it is often better to choose the one that saves energy.

Q 4.

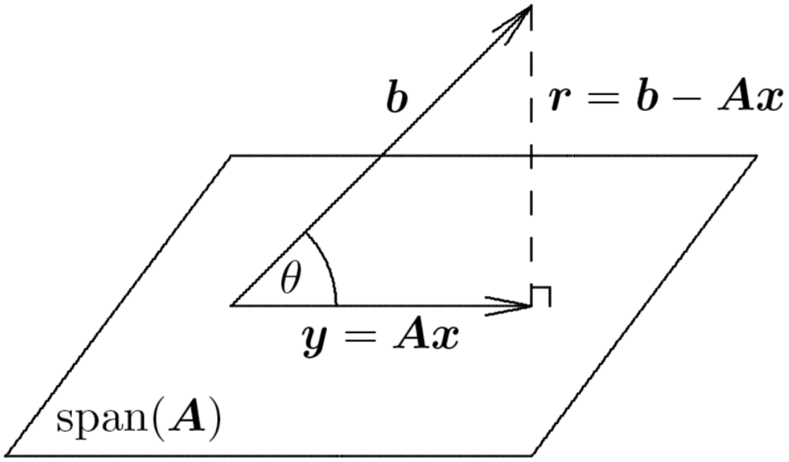
Is the system of normal equations an overdetermined or a square linear system?

Answer:

The system of normal equations $\mathbf{A}^T\mathbf{A}\ \mathbf{x} = \mathbf{A}^T\mathbf{b}$ is a linear system, where the matrix $\mathbf{A}^T\mathbf{A}$ is square. Hence, it is a square linear system.

Q 5.

If, for a least squares problem $\mathbf{Ax} \simeq \mathbf{b}$, we have $\mathbf{A}^T\mathbf{r} = \mathbf{0}$, what does it mean in terms of the following figure? In what case will **ATr = 0** hold?


$\boldsymbol{b}$
$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Ax}$
θ
$\boldsymbol{y} = \boldsymbol{Ax}$
$\mathrm{span}(\boldsymbol{A})$

Answer:

This means that **r** is orthogonal to span(**A**). It will hold only when **x** is the solution to the least squares problem, yielding the shortest residual vector. If **x** is not the solution, that is, if **r** is not of the shortest distance between **b** and span(**A**), then **r** wouldn't be orthogonal to span(**A**).

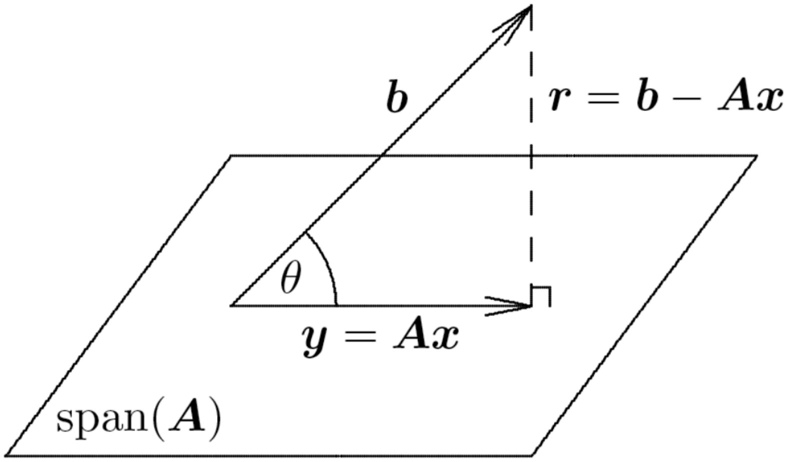
Q 6.

The bound of the condition number of the least squares solution is given as follows:

$$\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \mathrm{cond}(\boldsymbol{A}) \, \frac{1}{\cos(\theta)} \, \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}$$

.

Describe how this bound depends on θ in terms of the following figure.


$\boldsymbol{b}$
$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$
θ
$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
$\mathrm{span}(\boldsymbol{A})$

Answer:

According to the bound, the condition number of **x** depends on the condition number of **A**, which is obvious because if A is nearly rank deficient, it will make **x** difficult to determine. However, the sensitivity of x cannot be solely determined by **A** alone. It also depends on **b**. If be is nearly orthogonal to span(**A**), pointing almost vertically upward, the residual vector will still be quite long, which will create more sensitivity in determining x compared to when the minimum residual is quite short. This is reflected by 1/cos(θ). When the residual vector is quite long, θ is near 90 degrees, making 1/cos(θ) close to infinity.

Q 7.

Why is it preferred that the least squares problem be solved through QR factorization instead of solving normal equations?

Answer:

The condition number of the solution of normal equations is determined by the condition number of matrix $\mathbf{A}^T\mathbf{A}$, which is squared of the condition number of **A**.

On the other hand, QR factorization reduces the least squares problem into the problem of solving a triangular linear system, where the solution depends on the condition number of matrix **R**, which is the same as that of **A**.

Q 8.

Show that **cond**(R) = **cond**(A), where **A** = **QR**.

Answer:

cond(A) = $||A||_2\ ||A^{-1}||_2$. In the following, all norms are Euclidean, that is 2-norm.

Now, $||\mathbf{A}|| = \max_{x\neq 0} \frac{||Ax||}{||x||} = \max_{x\neq 0} \frac{||QRx||}{||x||} = \max_{x\neq 0} \frac{||Rx||}{||x||} = ||\mathbf{R}||$, where the last inequality is because Q is an orthogonal matrix, which preserves the Euclidean norm.

Similarly,

$1/||\mathbf{A}^{-1}|| = \min_{x\neq 0} \frac{||Ax||}{||x||} = \min_{x\neq 0} \frac{||QRx||}{||x||} = \min_{x\neq 0} \frac{||Rx||}{||x||} = 1/||\mathbf{R}^{-1}||$.

Hence, cond(A) = $||A||_2\ ||A^{-1}||_2 = ||R||_2\ ||R^{-1}||_2$ = cond(R).

Q 9.

Show that the least squares problem can be reduced through QR factorization to the problem of solving a triangular linear system.

Answer:

TBD.

Q 10.

The reduced QR factorization is $\mathbf{A} = \mathbf{Q}_1\mathbf{R}$. Then $\mathbf{R} = \mathbf{Q}_1^T\mathbf{A}$. Why the following derivation of R is wrong?

$\mathbf{Q}_1\mathbf{R} = \mathbf{A}$, hence, $\mathbf{R} = \mathbf{Q}_1^{-1}\mathbf{A} = \mathbf{Q}_1^T\mathbf{A}$.

Answer:

It is because $\mathbf{Q}_1$ is not a square matrix. So, $\mathbf{Q}_1^{-1}$ is an invalid expression. We have to find a different way of proving it.

Chapter 7

Q 1.

If I want to interpolate five data points, what polynomial interpolant do I need? Why?

Answer:

Interpolating all points means the linear system should have zero residual, which is achieved in the square linear system. Hence, the number of columns (unknowns) should be the same as the number of rows (data points), that is, five. An order four or quartic polynomial has four unknowns. Hence, we need to use a quartic polynomial interpolant.

Q 2.

If five data points need a quartic polynomial, could I also use a higher-order polynomial? What is the issue with that?

Answer:

Yes. However, it will lead to more unknowns than data points, that is, an underdetermined system, which will have infinitely many solutions. I cannot have a unique interpolant in this case.

Q 3.

What are the main computationally involved steps in polynomial interpolation?

Answer:

There are two main steps: 1) solving the square linear system, and 2) evaluating the interpolant.

Q 4.

For monomial basis functions, what is the computational complexity of the main steps?

Answer:

For monomial basis functions, solving the square linear system $O(n^3)$ where the square matrix is n-by-n. Evaluating the interpolant requires $O(n)$ multiplications.

Q 4.

For Lagrange basis functions, what is the computational complexity of the main steps?

Answer:

For Lagrange basis functions, solving the square linear system O(n) where the square matrix is n-by-n. Evaluating the interpolant requires O(n^2) multiplications.

Q 4.

For Newton basis functions, what is the computational complexity of the main steps?

Answer:

For Newton basis functions, solving the square linear system O(n^2) where the square matrix is n-by-n. Evaluating the interpolant requires O(n) multiplications.

Q 5. (hard)

Newton basis functions can be added to an existing interpolant when a new data point arrives without developing the interpolant for the whole data set from scratch. Why it is not possible for monomial or Lagrange basis functions?

Answer:

The added Newton basis function for the nth data point is such that it is zero for all the existing n-1 data points:

x_n $(t - t_1)(t - t_2)...(t - t_{n-1})$.

Moreover, when this basis function is added, the first n-1 basis functions do not change either. Hence, the new interpolant still goes through all the existing points without re-adjusting the first n-1 unknowns: x_1, …, x_{n-1}.

However, nth monomial basis does not become zero for previous data points: x_n t^{n-1}. Hence, the first n-1 unknowns must be readjusted to interpolate on the first n-1 data points again.

On the other hand, nth Lagrange basis function does become zero for previous data points, but when it is added, the first n-1 basis functions also end up being changed. Hence, the interpolant essentially has to be rebuilt.

Q 6.

For five data points, how many equations do we have with piecewise polynomial interpolation?

Answer:

If we have four data points, we must have four “pieces” or polynomials to join them. Each polynomial must go through two points. Hence, we will have eight equations. Generally, for n points, we will have 2(n-1) equations.

Q 7.

For five data points, how many equations does Hermite cubic interpolation produce? How many unknowns does it have?

Answer:

Hermite cubic interpolation means each polynomial will be cubic, and at the points where two polynomials meet, they will have the same first derivative as well. Each polynomial will have two equations for two points, producing eight equations. For five data points, there are three meeting points of two cubic polynomials, where their first derivative must be equal, producing four more equations. Hence, a total of 11.

Generally, for n points, it will have 2(n-1) + n-2 = 3n - 4 equations.

On the other hand, each cubic polynomial has four unknowns, giving a total of 4n - 4 unknowns.

Q 8.

For five data points, how many equations does cubic spline interpolation produce? How many unknowns does it have?

Answer:

Cubic spline interpolation means each polynomial will be cubic, and at the points where two polynomials meet, they will have the same first and second derivatives. Each polynomial will have two equations for two points, producing eight equations. For five data points, there are three meeting points of two cubic polynomials, where their first and second derivatives must be equal, producing eight more equations. Hence, a total of 19.

Generally, for n points, it will have 2(n-1) + 2*(n-2) = 4n - 6 equations.

On the other hand, each cubic polynomial has four unknowns, giving a total of 4n - 4 unknowns.