NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Module 7: Least Squares

Learning Objectives

- Learn the normal equations for least squares
- Learn QR decomposition
- Learn nonlinear least squares methods
- Understand mathematics of GPS

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Sources

- Textbook (Chapter 4: Least Squares)

- Wiki: GPS

  https://en.wikipedia.org/wiki/Global_Positioning_System

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

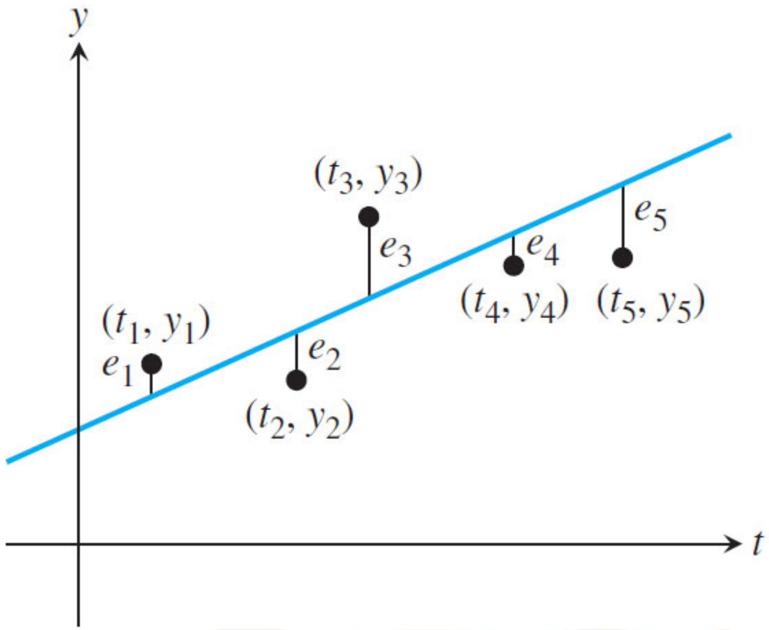
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Introduction

- The need of least squares
  - To find the next best solution if a set of equations are inconsistent.

$$x_1 + x_2 = 2$$
$$x_1 - x_2 = 1$$
$$x_1 + x_2 = 3.$$

  - To fit a model that may only approximate the given data.



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Introduction

- The concept of least squares dates from the work of Gauss and Legendre in the early 19th century.
- Least squares is important in modern statistics and mathematical modeling.
  - The key techniques of regression and parameter estimation have become fundamental tools.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Inconsistent system

- **A simple example:** Let us consider three equations:

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned}$$

  Obviously the first and third equations contradict with each other. There are no common solution (x_1, x_2) satisfying all three equations.

- **Matrix form:** We can rewrite the equations in matrix/vector notation as

$$\underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}}_{b} \quad \Longleftrightarrow \quad Ax = b$$

  The system matrix A has more rows than columns, which means that there are more equations than unknowns. In general, when $A \in \mathbf{R}^{m \times n}$ with $m > n$, we call the corresponding linear system *overdetermined*. Usually, such systems do not have a solution and are then called *inconsistent*.

2.1. Geometric approach

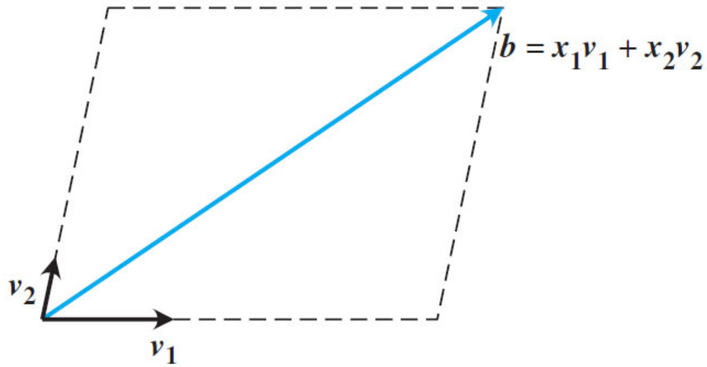
- Introduce the following notations:

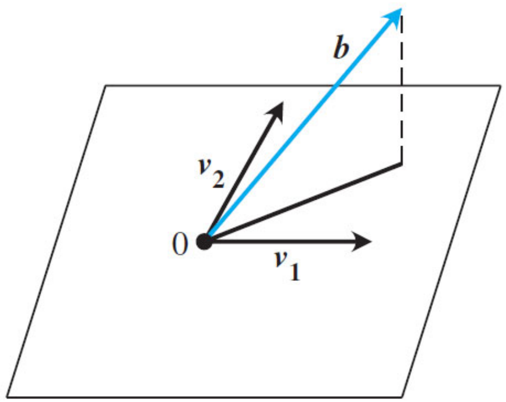
$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

Then $A = (v_1, v_2)$ and the linear system can be rewritten as

$$x_1 v_1 + x_2 v_2 = b$$

- Geometrically, v_1 and v_2 span a plane and the task of solving the linear system can be understood as trying to find a vector on the plane, which "reaches" b. That is, if b lies on the plane, b is the vector. Otherwise, reaching b is impossible. However, we can still get "as close as possible" to b.





NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Geometric approach

- Assume the target vector is v. It is a suitable linear combination of v_1 and v_2 with weights $\bar{x}_1$ and $\bar{x}_2$:

$$v = \bar{x}_1 v_1 + \bar{x}_2 v_2.$$

  Then we say $\bar{x} = (\bar{x}_1, \bar{x}_2)^T$ is the *least squares solution* of the original linear system.

- As v is as close as to b, v should be the orthogonal projection of b into the plane, i.e., $b - v$ is orthogonal to the plane. Thus

$$v_1^T(b - v) = 0, \quad v_2^T(b - v) = 0 \quad \Longrightarrow 0 = A^T(b - v) = A^T(b - A\bar{x}).$$

  This implies that $\bar{x}$ is the solution of the following equations

$$A^T A x = A^T b$$

  which we call the *normal equations*.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Geometric approach

- To sum up,

> **Normal equations for least squares**
>
> Given the inconsistent system
>
> $$Ax = b,$$
>
> solve
>
> $$A^T A\overline{x} = A^T b$$
>
> for the least squares solution $\overline{x}$ that minimizes the Euclidean length of the residual $r = b - Ax$.

- To measure the error, three typical metrics are used:

(1) **2-norm:** $\|r\|_2 = \sqrt{r_1^2 + r_2^2 + \cdots + r_n^2}$

(2) **squared error (SE):** $\text{SE} = r_1^2 + r_2^2 + \cdots + r_n^2$

(3) **root mean squared error (RMSE):** $\text{RMSE} = \sqrt{\frac{r_1^2+r_2^2+\cdots+r_n^2}{n}}$

Example

- Solve the least squares problem Ax=b with

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

The components of the normal equations are

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

and

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

Example

The normal equations

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

can now be solved by Gaussian elimination. The tableau form is

$$\left[\begin{array}{cc|c} 3 & 1 & 6 \\ 1 & 3 & 4 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 3 & 1 & 6 \\ 0 & 8/3 & 2 \end{array}\right],$$

which can be solved to get $\overline{x} = (\overline{x}_1, \overline{x}_2) = (7/4, 3/4)$.

- The residual

$$r = b - A\overline{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.0 \\ 0.5 \end{bmatrix}.$$

$$\|r\|_2 = \sqrt{0.5} \approx 0.707, \qquad \mathrm{SE} = 0.5, \qquad \mathrm{RMSE} = 1/\sqrt{6} \approx 0.408$$

2.2. Differential approach

- **A simple example:** Since the equations

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned}$$

do not have a solution, we seek a solution to the following problem:

$$\min_{(x_1,x_2)} (x_1 + x_2 - 2)^2 + (x_1 - x_2 - 1)^2 + (x_1 + x_2 - 3)^2$$

- **Matrix form:** We can also rewrite the above minimization problem as

$$\min_{(x_1,x_2)} \begin{pmatrix} x_1 + x_2 - 2 & x_1 - x_2 - 1 & x_1 + x_2 - 3 \end{pmatrix} \begin{pmatrix} x_1 + x_2 - 2 \\ x_1 - x_2 - 1 \\ x_1 + x_2 - 3 \end{pmatrix}$$

or

$$\min_{x} (Ax - b)^T (Ax - b)$$

SINGAPORE

Differential approach

- Let $F = (Ax - b)^T(Ax - b)$.

  The solution that minimizes F should satisfy

$$\frac{1}{2}\frac{\partial F}{\partial x} = A^T(Ax - b) = 0$$

  which actually derives the normal equation

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

- This new interpretation of the least squares solution can allow us to easily extend the method to *weighted least squares*. The basic idea is to introduce different weights to the squares of individual errors to reflect different importance.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Differential approach

- Thus the minimization problem could be

$$\min_x \sum_{i=1}^{m} w_i |b_i - (Ax)_i|^2$$

If we let $W = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_m \end{pmatrix}$, then the solution of the weighted least squares problem is the solution of the following equations

$$A^T W A x = A^T W b.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

2.3. Data fitting

One important application of the least squares technique is in data fitting. Suppose we are given a set of m data points $(t_1, y_1), \cdots, (t_m, y_m)$, which

- may result from a measuring process or a numerial simulation,
- and may contain measurement noise or numerial errors.

The data fitting process in general consists of the following four steps:

(1) **Choose a model:** pick the model $y = f(t)$, which best describes the data in theory, e.g., a line $y = c_1 + c_2 t$, or a parabola $y = c_1 + c_2 t + c_3 t^2$, or any other function f that depends *linearly* on some model parameters $c_1, c_2, \cdots, c_n$.

(2) **Force the model to fit the data:** Substitute the data points into the model, resulting in the m equations $y_i = f(t_i), i = 1, \cdots, m$, which can in turn be expressed as a linear system $Ax = b$ where $x = (c_1, \cdots, c_n)$ is the vector of unknown model parameters, $b = (y_1, \cdots, y_m)$ is the vector of ordinates, and A encodes the relation between the abscissa $t_1, \cdots, t_m$ and the chosen model.

Data fitting

(3) **Solve the normal equation:** find the vector $\bar{x}$ that is the solution to $A^T Ax = A^T b$.

(4) **Provide the error statistic:** compute the residual $r = b - A\bar{x}$ and one of the 3 measures. The smaller this measure value, teh better the chosen model fits the data.

- **Example:** Find the best parabola for the four data points (-1,1), (0,0), (1,0) and (2,-2).

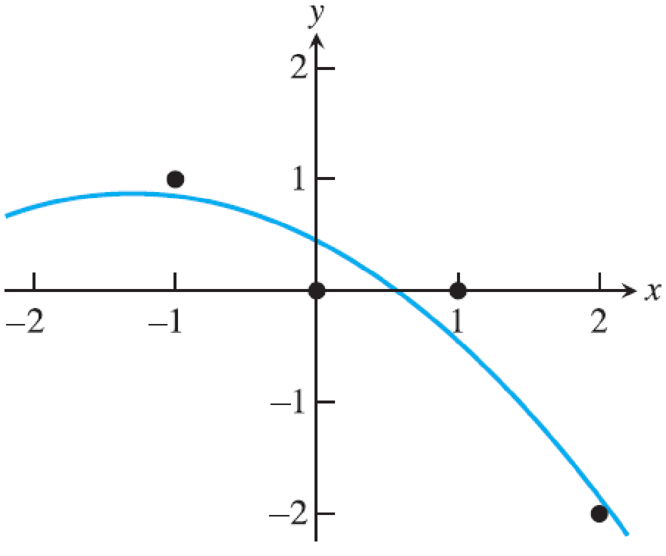
Set $y = c_1 + c_2 t + c_3 t^2$ and substitute the data points to yield

$$
\begin{aligned}
c_1 + c_2(-1) + c_3(-1)^2 &= 1 \\
c_1 + c_2(0) + c_3(0)^2 &= 0 \\
c_1 + c_2(1) + c_3(1)^2 &= 0 \\
c_1 + c_2(2) + c_3(2)^2 &= -2,
\end{aligned}
$$

Data fitting

or, in matrix form,

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix}.$$



(b)

This time, the normal equations are three equations in three unknowns

$$\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ -7 \end{bmatrix}.$$

Solving for the coefficients results in the best parabola $y = c_1 + c_2 t + c_3 t^2 = 0.45 - 0.65t - 0.25t^2$. The residual errors are given in the following table:

t	y	parabola	error
−1	1	0.85	0.15
0	0	0.45	−0.45
1	0	−0.45	0.45
2	−2	−1.85	−0.15

The error statistics are squared error $\text{SE} = (.15)^2 + (-.45)^2 + (.45)^2 + (-.15)^2 = 0.45$ and $\text{RMSE} = \sqrt{.45}/\sqrt{4} \approx 0.335$. ◄

Outline

§1. Introduction

§2. Least squares and the normal equations

§3. QR decomposition

§4. Nonlinear least squares

§5. GPS

§6. Homework

§7. Summary

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

QR decomposition

- One problem with computing the least squares solution via the normal equations: it can be unstable numerically due to a large condition number of the matrix A^TA.
- There is a better way, which is called **QR decomposition** or **QR factorization**, to solve the least squares calculation that is superior to the normal equations.
- QR decomposition has some similarity with the LU factorization that encodes the step of Gaussian elimination and is used to solve the matrix equation.
- QR decomposition can be realized by **Gram-Schmidt orthogonalization** method, or other methods.

Classic Gram-Schmidt orthogonalization

Given a set of n linearly independent vectors $a_1, \cdots, a_n \in \mathbf{R}^m$ with $m \geq n$, the goal of Gram-Schmidt algorithm is to find n mutually orthogonal unit vectors $q_1, \cdots, q_n \in \mathbf{R}^m$, which span the same subspace as the vectors a_i. The algorithm works as follows:

- **step 1:** Turn a_1 into a unit vector:

$$y_1 = a_1, \quad q_1 = \frac{y_1}{\|y_1\|}.$$

- **step 2:** Subtract the projection of a_2 onto q_1 from a_2 and normalize the result:

$$y_2 = a_2 - (a_2^T q_1) q_1, \quad q_2 = \frac{y_2}{\|y_2\|}.$$

  Clearly, this q_2 is orthogonal to q_1 since $q_2^T q_1 = 0$.

- **step 3 to n:** Continue with the process by letting:

$$y_j = a_j - \sum_{k=1}^{j-1} (a_j^T q_k) q_k, \quad q_j = \frac{y_j}{\|y_j\|}, \quad j = 3, \cdots, n..$$

Reduced QR decomposition

- It is easy to verify that $q_j^T q_i = 0$ for all $i = 1, \cdots, j-1$.
- Introducing the notation:

$$r_{j,j} = \|y_j\|, \quad r_{k,j} = a_j^T q_k$$

Then step j can be written:

$$q_j = \frac{y_j}{\|y_j\|} \Longleftrightarrow y_j = r_{j,j} q_j, \quad y_j = a_j - \sum_{k=1}^{j-1} (a_j^T q_k) q_k \Longleftrightarrow a_j = \sum_{k=1}^{j} r_{k,j} q_k$$

- Write the relation between a_i and q_i in matrix form, which is called the *reduced QR decomposition*:

$$\underbrace{(a_1, \ldots, a_n)}_{A \in \mathbb{R}^{m\times n}} = \underbrace{(q_1, \ldots, q_n)}_{\hat{Q} \in \mathbb{R}^{m\times n}} \cdot \underbrace{\begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ & r_{2,2} & \cdots & r_{2,n} \\ & & \ddots & \vdots \\ & & & r_{n,n} \end{pmatrix}}_{\hat{R} \in \mathbb{R}^{n\times n}} \quad \Longleftrightarrow \quad A = \hat{Q}\hat{R}.$$

Example

Find the reduced QR factorization by applying Gram–Schmidt orthogonalization to the columns of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

Set $y_1 = A_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Then $r_{11} = ||y_1||_2 = \sqrt{1^2 + 2^2 + 2^2} = 3$, and the first unit vector is

$$q_1 = \frac{y_1}{||y_1||_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Example (cont)

To find the second unit vector, set

$$y_2 = A_2 - q_1(q_1^T A_2) = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} 2 = \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$$

and

$$q_2 = \frac{y_2}{||y_2||_2} = \frac{1}{5} \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}.$$

Since $r_{12} = q_1^T A_2 = 2$ and $r_{22} = ||y_2||_2 = 5$, the result written in matrix form

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 \\ 2/3 & 1/3 \\ 2/3 & 2/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = QR.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Full QR decomposition

- When the reduced QR decomposition is successful, it is customary to fill out the matrix of orthogonal unit vectors to a complete basis of $\mathbf{R}^m$, to achieve the "full" QR decomposition.

- The full QR decomposition can be computed by

  - first adding $m - n$ vectors $a_{n+1}, \cdots, a_m$ to the given a_i, such that the m vectors $a_1, \cdots, a_m$ span $\mathbf{R}^m$,
  - and then carrying out the Gram-Schmidt algorithm to get a "full set" of vectors $q_1, \cdots, q_m$, forming an orthonormal basis of $\mathbf{R}^m$.
  - This leads to the full QR decompositon of A into the product of an orthogonal matrix Q and an upper triangular matrix R:

$$\underbrace{(a_1, \ldots, a_n)}_{A \in \mathbb{R}^{m\times n}} = \underbrace{(q_1, \ldots, q_m)}_{Q \in \mathbb{R}^{m\times m}} \cdot \underbrace{\begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ & r_{2,2} & \cdots & r_{2,n} \\ & & \ddots & \vdots \\ & & & r_{n,n} \\ \hline 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix}}_{R \in \mathbb{R}^{m\times n}} \qquad \Longleftrightarrow \qquad A = QR$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Example

Find the full QR factorization of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

We found the orthogonal unit vectors $q_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$ and $q_2 = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}$.

Adding a third vector $A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ leads to

$$y_3 = A_3 - q_1 q_1^T A_3 - q_2 q_2^T A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \frac{1}{3} - \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ -\frac{2}{15} \end{bmatrix} \left(-\frac{14}{15}\right) = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix}$$

Example (cont)

and $q_3 = y_3/||y_3|| = \begin{bmatrix} \frac{2}{15} \\ \frac{10}{15} \\ -\frac{11}{15} \end{bmatrix}$. Putting the parts together, we obtain the full QR factorization

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 & 2/15 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & 2/15 & -11/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = QR.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Application

On the one hand, the QR factorization can be applied to solving linear systems, because

$$Ax = b \quad \Longleftrightarrow \quad QRx = b \quad \Longleftrightarrow \quad Rx = Q^T b,$$

where the last step follows from the fact that $Q^{-1} = Q^T$ for any orthogonal matrix Q. The system $Rx = Q^T b$ can then be solved by back substitution. However, this is computationally more expensive than solving the system via the LU factorization of A.

On the other hand, we can use the QR factorization to solve linear least squares problems, because

$$\min_x \|Ax - b\| \quad \Longleftrightarrow \quad \min_x \|QRx - b\| \quad \Longleftrightarrow \quad \min_x \|Rx - Q^T b\|,$$

where the last step follows from the fact that $\|Qx\| = \|x\|$ for any orthogonal matrix Q. Letting $d = Q^T b$ and $e = Rx - d$, minimizing the norm of the residual $r = Ax - b$ (that is, finding the least squares solution of the system $Ax = b$) is therefore equivalent to minimizing the norm of e. Writing out e as

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \\ \hline e_{n+1} \\ \vdots \\ e_m \end{pmatrix} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ & r_{2,2} & \cdots & r_{2,n} \\ & & \ddots & \vdots \\ & & & r_{n,n} \\ \hline 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \\ \hline d_{n+1} \\ \vdots \\ d_m \end{pmatrix},$$

Application (cont)

we notice that the last $m - n$ components of e are always $(e_{n+1}, \ldots, e_m) = -(d_{n+1}, \ldots, e_m)$, *regardless* of the choice of $x_1, \ldots, x_n$. So, all we can do to minimize the norm of e is to make $(e_1, \ldots, e_n)$ as small as possible, but by solving the linear system

$$\hat{R}x^* = \hat{d}, \tag{3}$$

where $\hat{d} = (d_1, \ldots, d_n)$, we can actually get $(e_1, \ldots, e_n) = (0, \ldots, 0)$, hence the solution x^* of (3) turns out to be indeed the least squares solution of $Ax = b$, and the squared error of the residual $r = b - Ax^*$ is

$$\|r\|_2^2 = \|e\|_2^2 = d_{n+1}^2 + \cdots + d_m^2.$$

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Nonlinear least squares

- Consider the system of m equations in n unknowns with $m > n$

$$\begin{aligned} r_1(x_1, \cdots, x_n) &= 0 \\ &\vdots \\ r_m(x_1, \cdots, x_n) &= 0 \end{aligned}$$

where $r_i(x_1, \cdots, x_n)$ are general functions in $x_1, \cdots, x_n$. In general, the system has no solution.

- We construct a new function as the sum of the squares of r_i:

$$E(x_1, \cdots, x_n) = r_1^2 + \cdots + r_n^2 = r^T r$$

where $r = [r_1, \cdots, r_m]^T$.

  - The energy E is clearly non-negative, and zero if and only if there exists some $x = (x_1, \cdots, x_n)$ that solves all m equations.
  - Otherwise, we call the minimum of E the *nonlinear least squares solution* of the system of equations above.

Nonlinear least squares

- To minimize E, we set its gradient to zero

$$\frac{1}{2}\nabla E(x) = \frac{1}{2}\nabla\left(r^T(x)r(x)\right) = r^T(x)Dr(x) = 0$$

where $Dr = \left(\frac{\partial r}{\partial x_1}, \cdots, \frac{\partial r}{\partial x_n}\right)$ is the *Jacobian* of r.

- This is equivalent to finding the zero of the function $F(x)$:

$$F(x) = \frac{1}{2}(\nabla E(x))^T = Dr(x)^T r(x) = 0$$

which in turn can be found iterativaly with the (multivariate version) of Newton's method.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Gauss-Newton method

- The iteration is:

$$x^{(k+1)} = x^{(k)} - (DF(x^{(k)}))^{-1} F(x^{(k)})$$

  To simplify the computation, we actually drop the terms involving the second order partial derivatives, which hence gives

$$DF(x) = D(Dr(x)^T r(x)) \approx Dr(x)^T Dr(x)$$

- If we let $A = Dr(x^{(k)})$, then
  - $F(x^{(k)}) = A^T r(x)$
  - $DF(x^{(k)}) \approx A^T A$
  - $(DF(x^{(k)}))^{-1})F(x^{(k)}) = (A^T A)^{-1} A^T r(x^{(k)})$.
    If we further let $v = (DF(x^{(k)}))^{-1})F(x^{(k)})$, then $A^T A v = A^T r(x)$.
    We can find v by solving the equations.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Gauss-Newton method

- Gauss-Newton method

Gauss–Newton Method

To minimize

$$r_1(x)^2 + \cdots + r_m(x)^2.$$

Set $x^0 =$ initial vector,
for $k = 0, 1, 2, \ldots$

$$\begin{aligned} A &= Dr(x^k) \\ A^T A v^k &= A^T r(x^k) \\ x^{k+1} &= x^k - v^k \end{aligned}$$

end

- If all the functions $r_i(x)$ are linear: $r_i(x) = b_i - a_i^T x$, then we are back into the setting of linear least squares.

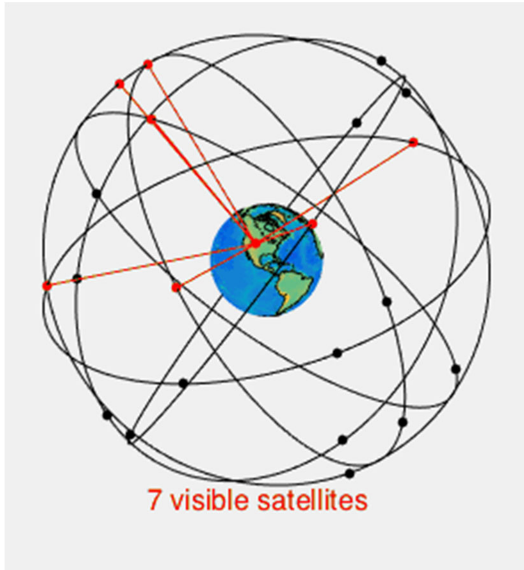
  Indeed, we then find that $r(x) = b - Ax$, where $b = (b_1, \cdots, b_m)$ and A is the matrix with rows $a_1^T, \cdots, a_m^T$ so that $Dr(x) = -A$ and $F(x) = -A^T(b - Ax)$. Therefore, finding the zero of F is equivalent to solving $A^T Ax = A^T b$.

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

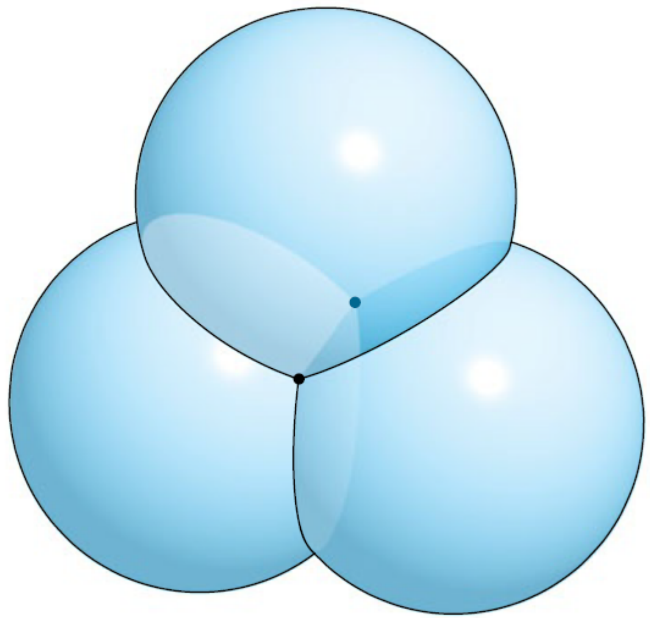
GPS

- The global positioning system (GPS) consists of 24 satellites carrying atomic clocks, orbiting the earth at an altitude of 20,200km. Four satellites in each of six planes, slanted at 55^{o} with respect to the poles, make two revolutions per day. At any time, from any point on earth, five to eight satellites are in the direct line of sight.
- Each satellite has a simple mission: to transmit carefully synchronized signals from predetermined positions in space, to be picked up by GPS receivers on earth. The receivers use the information, to determine accurate (x,y,z) coordinates of the receiver.



GPS math

- At a given instant, the receiver
  - collects the synchronized signal from the i-th satellite
  - Determines its transmission time t_i
- The normal speed of the signal is the speed of light, c, roughly 299792.458 km/s. Multiplying transmission time by c gives the distance between the satellite and the receiver.
- The receiver can be considered to be on a sphere whose center is at the satellite position and radius is $(c\ t_i)$.
- Approach 1: If 3 satellites are available, then 3 spheres are known. Their intersection consists of 2 points, one of which is the receiver's location.



GPS math

- A major problem: the clock in the low-cost receiver has poor accuracy.
- Approach 2: add one unknown and one extra satellite.
  - Define d to be the difference between the synchronized time on the satellite clocks and the earth-bound receiver clock
  - (A_i,B_i,C_i): location of satellites
  - (x,y,z): the true intersection point

$$
\begin{aligned}
(x-A_1)^2+(y-B_1)^2+(z-C_1)^2&=[c(t_1-d)]^2\\
(x-A_2)^2+(y-B_2)^2+(z-C_2)^2&=[c(t_2-d)]^2\\
(x-A_3)^2+(y-B_3)^2+(z-C_3)^2&=[c(t_3-d)]^2\\
(x-A_4)^2+(y-B_4)^2+(z-C_4)^2&=[c(t_4-d)]^2.
\end{aligned}
$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

GPS math

- 2 further problems:
  - Approach 2 is ill-conditioned when the satellites are bunched closely in the sky.
  - Transmission speed of the signals is not precisely c.
- Approach 3:
  - Add information from more satellites
  - Apply Gauss-Newton to solve a least squares problem
- In practice, some other factors are also considered.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Homework

Fit the $m = 4$ data points $P_i = (x_i, y_i)$:

$$P_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad P_4 = \begin{pmatrix} 2 \\ -2 \end{pmatrix},$$

with the parametric circle model $F : [0, 2\pi) \to \mathbf{R}^2$:

$$F(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + c_3 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

with $n = 3$ model parameters c_1, c_2 and c_3, using the initial parameter values

$$t_1 = 1, \quad t_2 = 2, \quad t_3 = 3, \quad t_4 = 4.$$

That is, complete the following:

(1) Construct the $2m$ equations $x_i = x(t_i)$ and $y_i = y(t_i)$ for $i = 1, \cdots, m$, expressed in matrix/vector notation as $Ac = b$ where $c = (c_1, \cdot, c_n)$.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Homework (cont)

(2) Find the least squares solution of this linear system, which in turns gives the best fitting parametric circle model.

(3) Compute the RMSE for this best fit.

(4) Discuss or suggest how to update/refine the parameter values t_i

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Summary

- The normal equations for least squares
- QR decomposition
  - Gram-Schmidt orthogonalization
- Nonlinear least squares
  - Gauss-Newton method

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

End

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE