Modulation in Communications (cont'd)

Recall that

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{v} = \mathbf{\Phi}\mathbf{D}\mathbf{\Phi}^H\bar{\mathbf{x}} + \mathbf{v}$$

Transceiver scheme #2:

- Transmitter side: $\bar{\mathbf{x}} = \mathbf{\Phi}\tilde{\mathbf{x}}$. Put info. in $\tilde{\mathbf{x}}$ (e.g., $\tilde{\mathbf{x}} \in \{-1, 1\}^T$ for binary signaling)$\Longrightarrow$ 1 IFFT

  $= \Phi D \cancel{\Phi^H \cdot \Phi} \tilde{x} + v$
- Receiver side: $\mathbf{y} = \mathbf{\Phi}\mathbf{D}\tilde{\mathbf{x}} + \mathbf{v}$. Estimate $\tilde{\mathbf{x}}$ via $\mathbf{D}^{-1}\mathbf{\Phi}^H\mathbf{y} \Longrightarrow$ 1 FFT

  LS estimation $\quad \min \| y - \Phi D \tilde{x} \|_2^2$
- Such a transceiver scheme is called orthogonal frequency division multiplexing (OFDM)

ΦD nonsingular

$y = \Phi D \tilde{x}$

$\exists$ solution unique

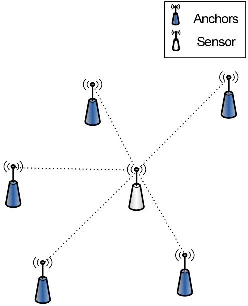
Localization

Aim: Locate the Cartesian coordinate of a sensor or device using distance info.

- Applications: localization in a wireless sensor network, GPS, etc.

- Let $\mathbf{x} \in \mathbb{R}^2$ be the coordinate of the sensor
- The sensor communicates with anchors (i.e., sensors or devices that know their locations)
- Let $\mathbf{a}_i \in \mathbb{R}^2$, $i = 1, \ldots, m$ be the anchors' locations
- The sensor measures the distances

$$d_i = \|\mathbf{x} - \mathbf{a}_i\|_2, \quad i = 1, \ldots, m$$



Localization (cont'd)

$x^Tx - a_i^Tx - x^Ta_i + a_i^Ta_i$

Re-arrange the equations $= (x - a_i)^T(x - a_i) =$

$$d_i^2 = \|\mathbf{x} - \mathbf{a}_i\|_2^2 = \|\mathbf{x}\|_2^2 - 2\mathbf{a}_i^T\mathbf{x} + \|\mathbf{a}_i\|_2^2, \quad i = 1, \ldots, m,$$

$\Rightarrow \|a_i\|_2^2 - d_i^2 = 2a_i^Tx - \|x\|_2^2$

as a matrix equation

$$\begin{bmatrix} \|\mathbf{a}_1\|_2^2 - d_1^2 \\ \vdots \\ \|\mathbf{a}_m\|_2^2 - d_m^2 \end{bmatrix} = \begin{bmatrix} 2\mathbf{a}_1^T & -1 \\ \vdots & \vdots \\ 2\mathbf{a}_m^T & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \|\mathbf{x}\|_2^2 \end{bmatrix}$$

Note that the above matrix equation is nonlinear

Idea: Solve the linear matrix equation

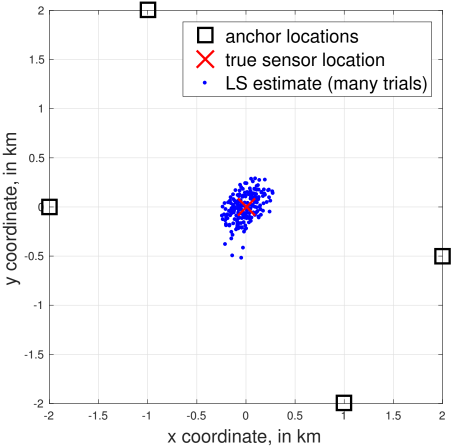
$$\underbrace{\begin{bmatrix} \|\mathbf{a}_1\|_2^2 - d_1^2 \\ \vdots \\ \|\mathbf{a}_m\|_2^2 - d_m^2 \end{bmatrix}}_{=\mathbf{y}} = \underbrace{\begin{bmatrix} 2\mathbf{a}_1^T & -1 \\ \vdots & \vdots \\ 2\mathbf{a}_m^T & -1 \end{bmatrix}}_{=\mathbf{A}} \begin{bmatrix} \mathbf{x} \\ z \end{bmatrix}$$

where $(\mathbf{x}, z)$ is a *free variable* on $\mathbb{R}^3$, i.e., without the constraint $z = \|\mathbf{x}\|_2^2$

Localization (cont'd)

- In practice, the sensor obtains noisy measurements $\hat{d}_i = d_i + v_i$, $i = 1, \ldots, m$, where v_t is noise
- We do the engineers' way:
  - Replace d_i's by $\hat{d}_i$'s, and compute the LS solution
  - Use the first two entries in the LS solution as the location estimate
- Further reading: A. H. Sayed, A. Tarighat, and N. Khajehnouri. "Network-based wireless location," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 24–40, 2005.

Localization (cont'd)



Number of anchors: $m = 4$, noise standard deviation: 0.1581km, number of trials: 200

Matrix Computations
Chapter 3: Least-squares Problems and QR Decomposition

Section 3.2 Least-squares Solution

Jie Lu
ShanghaiTech University

LS Solution

Theorem (LS Optimality Condition)

$\mathbf{x}_{\mathrm{LS}} \in \mathbb{R}^n$ is an optimal solution to the LS problem $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$ if and only if it satisfies the following normal equation:

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_{\mathrm{LS}} = \mathbf{A}^T\mathbf{y}. \qquad (*)$$

- The optimality condition $(*)$ is true for any $\mathbf{A}$, not limited to full-column rank $\mathbf{A}$
- When $\mathbf{A}$ has full-column rank, $\Longleftrightarrow A^TA$ p.d.
  - $\mathbf{A}^T\mathbf{A}$ is nonsingular
  - $\mathbf{x}_{\mathrm{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$ is the *unique* solution to $(*)$
- Same result holds for the complex case

$$\mathbf{A}^H\mathbf{A}\mathbf{x}_{\mathrm{LS}} = \mathbf{A}^H\mathbf{y}$$

$\forall x \in \mathbb{R}^n$,
$x^T(A^TA)x$
$= (Ax)^T(Ax)$
$= \|Ax\|_2^2 \geq 0$

A full column rank

$Ax = 0 \Longleftrightarrow x = 0$

Proof using the Projection Theorem

The above Theorem can be proved using the Projection Theorem

Let $\mathbf{x}_{\text{LS}}$ be an LS solution. Then,

x_{LS} may not be unique

$$\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \arg \min_{\mathbf{z} \in \mathcal{R}(\mathbf{A})} \|\mathbf{z} - \mathbf{y}\|_2^2 = \mathbf{A}\mathbf{x}_{\text{LS}}$$

subspace

unique

From the Projection Theorem (Section 1.2),

$$\begin{aligned}
\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \mathbf{A}\mathbf{x}_{\text{LS}} \quad &\Longleftrightarrow \quad \mathbf{z}^T(\mathbf{A}\mathbf{x}_{\text{LS}} - \mathbf{y}) = 0 \text{ for all } \mathbf{z} \in \mathcal{R}(\mathbf{A}) \\
&\Longleftrightarrow \quad \mathbf{x}^T \left[\mathbf{A}^T(\mathbf{A}\mathbf{x}_{\text{LS}} - \mathbf{y})\right] = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \\
&\Longleftrightarrow \quad \mathbf{A}^T(\mathbf{A}\mathbf{x}_{\text{LS}} - \mathbf{y}) = \mathbf{0}
\end{aligned}$$

Orthogonal Projections

Suppose $\mathbf{A}$ has full column rank

- The projections of $\mathbf{y}$ onto $\mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})^\perp$ are given by

$$\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \mathbf{A}\mathbf{x}_{\mathrm{LS}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$$

$$\Pi_{\mathcal{R}(\mathbf{A})^\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = (\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{y}$$

- The orthogonal projector of $\mathbf{A}$ is defined as

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

- The orthogonal complement projector of $\mathbf{A}$ is defined as

$$\mathbf{P}_{\mathbf{A}}^\perp = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

- $\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \mathbf{P}_{\mathbf{A}}\mathbf{y}$, $\Pi_{\mathcal{R}(\mathbf{A})^\perp}(\mathbf{y}) = \mathbf{P}_{\mathbf{A}}^\perp\mathbf{y}$

Orthogonal Projections

Properties of $\mathbf{P_A}$ (same to $\mathbf{P_A^\perp}$):

- $\mathbf{P_A}$ is idempotent, i.e., $\mathbf{P_A P_A} = \mathbf{P_A}$
- $\mathbf{P_A}$ is symmetric

$$P_A^T = \left[A(A^TA)^{-1}A^T\right]^T = A(A^TA)^{-T}A^T$$
$$= A(A^TA)^{-1}A^T = P_A$$

Some other properties (will be revealed later):

- The eigenvalues of $\mathbf{P_A}$ are either zero or one

$$U_1^T U_1 = I$$

- $\mathbf{P_A}$ can be written as $\mathbf{P_A} = \mathbf{U}_1\mathbf{U}_1^T$ for some semi-orthogonal $\mathbf{U}_1$

Sketch of Proof: There always exists a semi-orthogonal $\mathbf{U}_1$ such that $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U}_1)$, so that $\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \Pi_{\mathcal{R}(\mathbf{U}_1)}(\mathbf{y})$ for all $\mathbf{y}$.
Also note that $\Pi_{\mathcal{R}(\mathbf{U}_1)}(\mathbf{y}) = \mathbf{U}_1\mathbf{U}_1^T\mathbf{y}$.
It follows that $(\mathbf{P_A} - \mathbf{U}_1\mathbf{U}_1^T)\mathbf{y} = \mathbf{0}$ for all $\mathbf{y}$. Therefore, $\mathbf{P_A} = \mathbf{U}_1\mathbf{U}_1^T$.

Pseudo-Inverse

The pseudo-inverse of a full-column-rank $\mathbf{A}$ is defined as

$$\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

- $\mathbf{A}^{\dagger}$ satisfies $\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{I}$, but not necessarily $\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{I}$
- $\mathbf{A}^{\dagger}\mathbf{y}$ is the unique LS solution
- We will study pseudo-inverse for general matrices later

LS by Convex Optimization

The LS optimality condition can also be proved via convex optimization

Definitions:

- The gradient of a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any $\alpha \in [0, 1]$,

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y})$$

Fact: Consider an unconstrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \; f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable

- For convex f, $\mathbf{x}^\star$ is an optimal solution if and only if $\nabla f(\mathbf{x}^\star) = \mathbf{0}$
- For non-convex f, any point $\hat{\mathbf{x}}$ satisfying $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}$ is a stationary point

LS by Convex Optimization (cont'd)

Fact: Consider a quadratic function

$$f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} + \mathbf{q}^T\mathbf{x} + c$$

where $\mathbf{R} = \mathbf{R}^T \in \mathbb{R}^{n\times n}$

- $\nabla f(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{q}$
- f is convex if $\mathbf{R}$ is positive semi-definite

The LS objective function is $(y-Ax)^T(y-Ax)$

$$f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2(\mathbf{A}^T\mathbf{y})^T\mathbf{x} + \|\mathbf{y}\|_2^2$$

A^TA p.s.d.

convex

Using the above fact, $\mathbf{x}_{\rm LS}$ is an LS optimal solution if and only if

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_{\rm LS} - \mathbf{A}^T\mathbf{y} = \mathbf{0}$$

normal equation

$\nabla f(x_{LS}) = 0$

$\| $

$2A^TAx_{LS} - 2A^Ty$

LS by Convex Optimization (cont'd)

Example: Consider a regularized LS problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda\|\mathbf{x}\|_2^2, \quad \text{for some constant (weight) } \lambda > 0$$

$A^TA+\lambda I$ p.s.d. convex

- Solution by optimization:

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{y} + 2\lambda\mathbf{x}$$

  The optimal solution is

$$\mathbf{x}_{\mathrm{RLS}} = (\mathbf{A}^T\mathbf{A}+\lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y}$$

For any two vectors x, y, $\left\|\begin{bmatrix} x \\ y \end{bmatrix}\right\|_2^2 = \|x\|_2^2 + \|y\|_2^2$

- Solution by the Projection Theorem: Rewrite the problem as

RLS $\longrightarrow$ LS

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} - \underbrace{\begin{bmatrix} \mathbf{A} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix}}_{\text{full column rank}} \mathbf{x} \right\|_2^2,$$

  and then use the Projection Theorem to get the same result

Matrix Computations
Chapter 3: Least-squares Problems and QR Decomposition

Section 3.3 QR Decomposition

Jie Lu
ShanghaiTech University

Thin QR Decomposition for Full Column-Rank Matrices

Theorem

Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$ *has full column rank. Then,* $\mathbf{A}$ *admits a decomposition*

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{R}_1 \qquad \textit{(Thin QR Decomposition)}$$

where $\mathbf{Q}_1 \in \mathbb{R}^{m \times n}$ *is semi-orthogonal and* $\mathbf{R}_1 \in \mathbb{R}^{n \times n}$ *is nonsingular and upper triangular.* $\Rightarrow [R_1]_{ii} \neq 0 \quad \forall i = 1, \cdots, n$

In addition, if we restrict $[\mathbf{R}_1]_{ii} > 0$ *for all* $i = 1, \ldots, n$*, then* $(\mathbf{Q}_1, \mathbf{R}_1)$ *is unique.*

Proof:

Since $\mathbf{A}$ has full column rank, $\mathbf{C} := \mathbf{A}^T\mathbf{A}$ is positive definite.

Hence, there exists a unique Cholesky decomposition $\mathbf{C} = \mathbf{R}_1^T\mathbf{R}_1$ where $\mathbf{R}_1$ is upper triangular with positive diagonal entries. $\Rightarrow R_1$ nonsingular

Let $\mathbf{Q}_1 = \mathbf{A}\mathbf{R}_1^{-1}$. It can be verified that $\mathbf{Q}_1^T\mathbf{Q}_1 = \mathbf{I}$ and $\mathbf{Q}_1\mathbf{R}_1 = \mathbf{A}$.

$Q_1^T Q_1 = (AR_1^{-1})^T A R_1^{-1} = R_1^{-T}(A^TA)R_1^{-1} = R_1^{-T}(R_1^TR_1)R_1^{-1} = I$

Note: We don't find QR decomposition via Cholesky decomposition in practice

Gram-Schmidt Procedure

Linearly independent

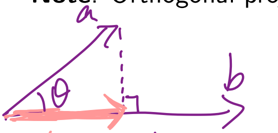
Aim: Given a basis $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ of a subspace $\mathcal{S} \subset \mathbb{R}^m$, find an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$ of $\mathcal{S}$, i.e.,

$= \text{span}\{a_1, \cdots, a_n\}$

1. $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\} = \text{span}\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$
2. $\begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$ is a semi-orthogonal matrix (an orthogonal matrix if $m = n$)

Idea: Let $\mathbf{q}_1$ be normalized $\mathbf{a}_1$
Each $\mathbf{q}_{i+1}$ is obtained by removing $\mathbf{q}_1-, \ldots, \mathbf{q}_i-$component from $\mathbf{a}_{i+1}$, $i = 1, \ldots, n-1$ and then normalizing it

Note: Orthogonal projection of vector $\mathbf{a}$ onto vector $\mathbf{b}$ is given by



$$\frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle}\mathbf{b},$$

$\rightsquigarrow \underbrace{\|a\|_2 \cos\theta}_{\text{length}} \cdot \underbrace{\frac{b}{\|b\|_2}}_{\text{direction}}$

$$\cos\theta = \frac{\langle a, b \rangle}{\|a\|_2 \cdot \|b\|_2}$$

Gram-Schmidt Procedure (cont'd)

$$
\begin{aligned}
\tilde{\mathbf{q}}_1 =& \mathbf{a}_1 \\
\mathbf{q}_1 =& \frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|_2} \\
\tilde{\mathbf{q}}_2 =& \mathbf{a}_2 - (\mathbf{q}_1^T \mathbf{a}_2)\mathbf{q}_1 \\
\mathbf{q}_2 =& \frac{\tilde{\mathbf{q}}_2}{\|\tilde{\mathbf{q}}_2\|_2} \\
& \dots \\
\tilde{\mathbf{q}}_i =& \mathbf{a}_i - (\mathbf{q}_1^T \mathbf{a}_i)\mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{a}_i)\mathbf{q}_2 - \cdots - (\mathbf{q}_{i-1}^T \mathbf{a}_i)\mathbf{q}_{i-1} \\
\mathbf{q}_i =& \frac{\tilde{\mathbf{q}}_i}{\|\tilde{\mathbf{q}}_i\|_2} \\
& \dots \\
\tilde{\mathbf{q}}_n =& \mathbf{a}_n - \sum_{i=1}^{n-1} (\mathbf{q}_i^T \mathbf{a}_n)\mathbf{q}_i \\
\mathbf{q}_n =& \frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|_2}
\end{aligned}
$$

Gram-Schmidt Procedure (cont'd)

$n \leq m$

Algorithm: Gram-Schmidt
input: a collection of linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ $\in \mathbb{R}^m$
$\tilde{\mathbf{q}}_1 = \mathbf{a}_1$, $\mathbf{q}_1 = \tilde{\mathbf{q}}_1 / \|\tilde{\mathbf{q}}_1\|_2$
for $i = 2, \ldots, n$
  $\tilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^T \mathbf{a}_i) \mathbf{q}_j$ $O(m) \times (i-1)$
  $\mathbf{q}_i = \tilde{\mathbf{q}}_i / \|\tilde{\mathbf{q}}_i\|_2$ $O(m)$
end
output: $\mathbf{q}_1, \ldots, \mathbf{q}_n$

The ith iteration : $O(mi)$

Total : $O\left(m \sum_{i=1}^{n} i\right) = O(mn^2)$

- Complexity: $\mathcal{O}(mn^2)$
- $\text{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_i\} = \text{span}\{\mathbf{q}_1, \cdots, \mathbf{q}_i\} = \text{span}\{\tilde{\mathbf{q}}_1, \cdots, \tilde{\mathbf{q}}_i\}$ for all $i = 1, \ldots, n$