

Estimation

- ▶ Gaussian random vectors
- ▶ minimum mean-square estimation (MMSE)
- ▶ MMSE with linear measurements
- ▶ relation to least-squares, pseudo-inverse

Gaussian random vectors

random vector $x \in \mathbb{R}^n$ is *Gaussian* if it has density

$$p_x(v) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}(v - \bar{x})^\top \Sigma^{-1}(v - \bar{x})\right),$$

for some $\Sigma = \Sigma^\top > 0$, $\bar{x} \in \mathbb{R}^n$

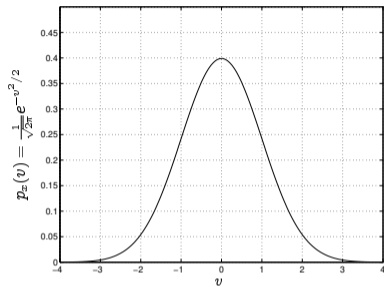
- ▶ denoted $x \sim \mathcal{N}(\bar{x}, \Sigma)$
- ▶ $\bar{x} \in \mathbb{R}^n$ is the *mean* or *expected* value of x , *i.e.*,

$$\bar{x} = \mathbf{E} x = \int v p_x(v) dv$$

- ▶ $\Sigma = \Sigma^\top > 0$ is the *covariance* matrix of x , *i.e.*,

$$\begin{aligned}\Sigma &= \mathbf{E}(x - \bar{x})(x - \bar{x})^\top \\ &= \mathbf{E} x x^\top - \bar{x} \bar{x}^\top \\ &= \int (v - \bar{x})(v - \bar{x})^\top p_x(v) dv\end{aligned}$$

density for $x \sim \mathcal{N}(0, 1)$:



2

- ▶ mean and variance of scalar random variable x_i are

$$\mathbf{E} x_i = \bar{x}_i, \quad \mathbf{E}(x_i - \bar{x}_i)^2 = \Sigma_{ii}$$

hence standard deviation of x_i is $\sqrt{\Sigma_{ii}}$

- ▶ covariance between x_i and x_j is $\mathbf{E}(x_i - \bar{x}_i)(x_j - \bar{x}_j) = \Sigma_{ij}$

- ▶ correlation coefficient between x_i and x_j is $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$

- ▶ mean (norm) square deviation of x from \bar{x} is

$$\mathbf{E}\|x - \bar{x}\|^2 = \mathbf{E} \text{trace}(x - \bar{x})(x - \bar{x})^\top = \text{trace} \Sigma = \sum_{i=1}^n \Sigma_{ii}$$

(using $\text{trace} AB = \text{trace} BA$)

example: $x \sim \mathcal{N}(0, I)$ means x_i are independent identically distributed (IID) $\mathcal{N}(0, 1)$ random variables

Confidence ellipsoids

- ▶ $p_x(v)$ is constant for $(v - \bar{x})^T \Sigma^{-1} (v - \bar{x}) = \alpha$, i.e., on the surface of ellipsoid

$$\mathcal{E}_\alpha = \{v \mid (v - \bar{x})^T \Sigma^{-1} (v - \bar{x}) \leq \alpha\}$$

- ▶ thus \bar{x} and Σ determine shape of density
- ▶ *η -confidence set* for random variable z is smallest volume set S with $\mathbf{Prob}(z \in S) \geq \eta$
 - ▶ in general case confidence set has form $\{v \mid p_z(v) \geq \beta\}$
- ▶ \mathcal{E}_α are the η -confidence sets for Gaussian, called *confidence ellipsoids*
- ▶ α determines confidence level η

Confidence levels

the nonnegative random variable $(x - \bar{x})^T \Sigma^{-1} (x - \bar{x})$ has a χ_n^2 distribution, so $\mathbf{Prob}(x \in \mathcal{E}_\alpha) = F_{\chi_n^2}(\alpha)$ where $F_{\chi_n^2}$ is the CDF

some good approximations:

- ▶ \mathcal{E}_n gives about 50% probability
- ▶ $\mathcal{E}_{n+2\sqrt{n}}$ gives about 90% probability

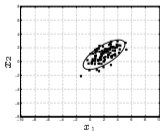
geometrically:

- ▶ mean \bar{x} gives center of ellipsoid
- ▶ semiaxes are $\sqrt{\alpha\lambda_i}u_i$, where u_i are (orthonormal) eigenvectors of Σ with eigenvalues λ_i

example: $x \sim \mathcal{N}(\bar{x}, \Sigma)$ with $\bar{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

- ▶ x_1 has mean 2, std. dev. $\sqrt{2}$
- ▶ x_2 has mean 1, std. dev. 1
- ▶ correlation coefficient between x_1 and x_2 is $\rho = 1/\sqrt{2}$
- ▶ $\mathbf{E}\|x - \bar{x}\|^2 = 3$

90% confidence ellipsoid corresponds to $\alpha = 4.6$:



(here, 91 out of 100 fall in $\mathcal{E}_{4.6}$)

Affine transformation

suppose $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$

consider affine transformation of x :

$$z = Ax + b,$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

then z is Gaussian, with mean

$$\mathbf{E} z = \mathbf{E}(Ax + b) = A \mathbf{E} x + b = A\bar{x} + b$$

and covariance

$$\begin{aligned}\Sigma_z &= \mathbf{E}(z - \bar{z})(z - \bar{z})^\top \\ &= \mathbf{E} A(x - \bar{x})(x - \bar{x})^\top A^\top \\ &= A \Sigma_x A^\top\end{aligned}$$

examples:

- ▶ if $w \sim \mathcal{N}(0, I)$ then $x = \Sigma^{1/2}w + \bar{x}$ is $\mathcal{N}(\bar{x}, \Sigma)$

useful for simulating vectors with given mean and covariance

- ▶ conversely, if $x \sim \mathcal{N}(\bar{x}, \Sigma)$ then $z = \Sigma^{-1/2}(x - \bar{x})$ is $\mathcal{N}(0, I)$
(normalizes & decorrelates; called *whitening* or *normalizing*)

suppose $x \sim \mathcal{N}(\bar{x}, \Sigma)$ and $c \in \mathbb{R}^n$

scalar $c^T x$ has mean $c^T \bar{x}$ and variance $c^T \Sigma c$

thus (unit length) direction of minimum variability for x is u , where

$$\Sigma u = \lambda_{\min} u, \quad \|u\| = 1$$

standard deviation of $u^T x$ is $\sqrt{\lambda_{\min}}$

(similarly for maximum variability)

Degenerate Gaussian vectors

- ▶ it is convenient to allow Σ to be singular (but still $\Sigma = \Sigma^T \geq 0$)
 - ▶ in this case density formula obviously does not hold
 - ▶ meaning: in some directions x is not random at all
 - ▶ random variable x is called a *degenerate Gaussian*

- ▶ write Σ as

$$\Sigma = [Q_+ \quad Q_0] \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} [Q_+ \quad Q_0]^T$$

where $Q = [Q_+ \quad Q_0]$ is orthogonal, $\Sigma_+ > 0$

- ▶ columns of Q_0 are orthonormal basis for **null**(Σ)
 - ▶ columns of Q_+ are orthonormal basis for **range**(Σ)
- ▶ then

$$Q^T x = \begin{bmatrix} z \\ w \end{bmatrix}, \quad x = Q_+ z + Q_0 w$$

- ▶ $z \sim \mathcal{N}(Q_+^T \bar{x}, \Sigma_+)$ is (nondegenerate) Gaussian (hence, density formula holds)
 - ▶ $w = Q_0^T \bar{x} \in \mathbb{R}^n$ is not random, called *deterministic component* of x

Linear measurements

linear measurements with noise:

$$y = Ax + v$$

- ▶ $x \in \mathbb{R}^n$ is what we want to measure or estimate
- ▶ $y \in \mathbb{R}^m$ is measurement
- ▶ $A \in \mathbb{R}^{m \times n}$ characterizes sensors or measurements
- ▶ v is sensor noise

common assumptions:

- ▶ $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$
- ▶ $v \sim \mathcal{N}(\bar{v}, \Sigma_v)$
- ▶ x and v are independent
- ▶ $\mathcal{N}(\bar{x}, \Sigma_x)$ is the *prior distribution* of x (describes initial uncertainty about x)
- ▶ \bar{v} is noise *bias* or *offset* (and is usually 0)
- ▶ Σ_v is noise *covariance*

thus

$$\begin{bmatrix} x \\ v \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \bar{x} \\ \bar{v} \end{bmatrix}, \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_v \end{bmatrix} \right)$$

using

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & 0 \\ A & I \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}$$

we can write

$$\mathbf{E} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \bar{x} \\ A\bar{x} + \bar{v} \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{E} \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T &= \begin{bmatrix} I & 0 \\ A & I \end{bmatrix} \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_v \end{bmatrix} \begin{bmatrix} I & 0 \\ A & I \end{bmatrix}^T \\ &= \begin{bmatrix} \Sigma_x & \Sigma_x A^T \\ A \Sigma_x & A \Sigma_x A^T + \Sigma_v \end{bmatrix} \end{aligned}$$

covariance of measurement y is $A\Sigma_x A^T + \Sigma_v$

- ▶ $A\Sigma_x A^T$ is 'signal covariance'
- ▶ Σ_v is 'noise covariance'

Minimum mean-square estimation

suppose $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are random vectors (not necessarily Gaussian)

we seek to estimate x given y

thus we seek a function $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ such that $\hat{x} = \phi(y)$ is near x

one common measure of nearness: mean-square error,

$$\mathbf{E}\|\phi(y) - x\|^2$$

minimum mean-square estimator (MMSE) ϕ_{mmse} minimizes this quantity

general solution: $\phi_{\text{mmse}}(y) = \mathbf{E}(x|y)$, *i.e.*, the conditional expectation of x given y

MMSE for Gaussian vectors

now suppose $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are jointly Gaussian:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_y \end{bmatrix} \right)$$

(after a lot of algebra) the conditional density is

$$p_{x|y}(v|y) = (2\pi)^{-n/2} (\det \Lambda)^{-1/2} \exp \left(-\frac{1}{2} (v - w)^\top \Lambda^{-1} (v - w) \right),$$

where

$$\Lambda = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^\top, \quad w = \bar{x} + \Sigma_{xy} \Sigma_y^{-1} (y - \bar{y})$$

hence MMSE estimator (*i.e.*, conditional expectation) is

$$\hat{x} = \phi_{\text{mmse}}(y) = \mathbf{E}(x|y) = \bar{x} + \Sigma_{xy} \Sigma_y^{-1} (y - \bar{y})$$

ϕ_{mmse} is an affine function

MMSE estimation error, $\hat{x} - x$, is a Gaussian random vector

$$\hat{x} - x \sim \mathcal{N}(0, \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^\top)$$

note that

$$\Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^T \leq \Sigma_x$$

i.e., covariance of estimation error is always less than prior covariance of x

Best linear unbiased estimator

estimator

$$\hat{x} = \phi_{\text{blu}}(y) = \bar{x} + \Sigma_{xy} \Sigma_y^{-1} (y - \bar{y})$$

makes sense when x , y aren't jointly Gaussian

this estimator

- ▶ is *unbiased*, i.e., $\mathbf{E} \hat{x} = \mathbf{E} x$
- ▶ often works well
- ▶ is widely used
- ▶ has minimum mean square error among all *affine* estimators

sometimes called *best linear unbiased* estimator

MMSE with linear measurements

consider specific case

$$y = Ax + v, \quad x \sim \mathcal{N}(\bar{x}, \Sigma_x), \quad v \sim \mathcal{N}(\bar{v}, \Sigma_v),$$

x, v independent

MMSE of x given y is affine function

$$\hat{x} = \bar{x} + B(y - \bar{y})$$

where $B = \Sigma_x A^\top (A \Sigma_x A^\top + \Sigma_v)^{-1}$, $\bar{y} = A\bar{x} + \bar{v}$

intepretation:

- ▶ \bar{x} is our best prior guess of x (before measurement)
- ▶ $y - \bar{y}$ is the discrepancy between what we actually measure (y) and the expected value of what we measure (\bar{y})
- ▶ estimator modifies prior guess by B times this discrepancy
- ▶ estimator blends prior information with measurement
- ▶ B gives *gain* from *observed discrepancy* to *estimate*
- ▶ B is small if noise term Σ_v in 'denominator' is large

MMSE error with linear measurements

MMSE estimation error, $\tilde{x} = \hat{x} - x$, is Gaussian with zero mean and covariance

$$\Sigma_{\text{est}} = \Sigma_x - \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} A \Sigma_x$$

- ▶ $\Sigma_{\text{est}} \leq \Sigma_x$, *i.e.*, measurement always decreases uncertainty about x
- ▶ difference $\Sigma_x - \Sigma_{\text{est}}$ (or some other comparison) gives *value* of measurement y in estimating x
 - ▶ $(\Sigma_{\text{est } ii} / \Sigma_x ii)^{1/2}$ gives fractional decrease in uncertainty of x_i due to measurement
 - ▶ $(\text{trace } \Sigma_{\text{est}} / \text{trace } \Sigma)^{1/2}$ gives fractional decrease in uncertainty in x , measured by mean-square error

Estimation error covariance

- ▶ error covariance Σ_{est} can be determined *before* measurement y is made!
- ▶ to evaluate Σ_{est} , only need to know
 - ▶ A (which characterizes sensors)
 - ▶ prior covariance of x (*i.e.*, Σ_x)
 - ▶ noise covariance (*i.e.*, Σ_v)
- ▶ you *do not* need to know the measurement y (or the means \bar{x} , \bar{v})
- ▶ useful for *experiment design* or *sensor selection*

Information matrix formulas

we can write estimator gain matrix as

$$\begin{aligned} B &= \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} \\ &= (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1} \end{aligned}$$

- ▶ $n \times n$ inverse instead of $m \times m$
- ▶ Σ_x^{-1} , Σ_v^{-1} sometimes called *information matrices*

corresponding formula for estimator error covariance:

$$\begin{aligned} \Sigma_{\text{est}} &= \Sigma_x - \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} A \Sigma_x \\ &= (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} \end{aligned}$$

can interpret $\Sigma_{\text{est}}^{-1} = \Sigma_x^{-1} + A^T \Sigma_v^{-1} A$ as:

$$\begin{aligned} & \text{posterior information matrix } (\Sigma_{\text{est}}^{-1}) \\ & = \text{prior information matrix } (\Sigma_x^{-1}) \\ & + \text{information added by measurement } (A^T \Sigma_v^{-1} A) \end{aligned}$$

proof: multiply

$$\Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} \stackrel{?}{=} (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1}$$

on left by $(A^T \Sigma_v^{-1} A + \Sigma_x^{-1})$ and on right by $(A \Sigma_x A^T + \Sigma_v)$ to get

$$(A^T \Sigma_v^{-1} A + \Sigma_x^{-1}) \Sigma_x A^T \stackrel{?}{=} A^T \Sigma_v^{-1} (A \Sigma_x A^T + \Sigma_v)$$

which is true

Relation to regularized least-squares

suppose $\bar{x} = 0$, $\bar{v} = 0$, $\Sigma_x = \alpha^2 I$, $\Sigma_v = \beta^2 I$

estimator is $\hat{x} = By$ where

$$\begin{aligned} B &= (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1} \\ &= (A^T A + (\beta/\alpha)^2 I)^{-1} A^T \end{aligned}$$

... which corresponds to regularized least-squares

MMSE estimate \hat{x} minimizes

$$\|Az - y\|^2 + (\beta/\alpha)^2 \|z\|^2$$

over z

Example

navigation using range measurements to distant beacons

$$y = Ax + v$$

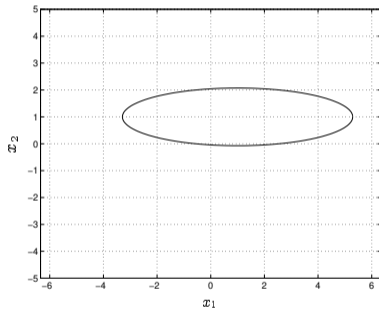
- ▶ $x \in \mathbb{R}^2$ is location
- ▶ y_i is range measurement to i th beacon
- ▶ v_i is range measurement error, IID $\mathcal{N}(0, 1)$
- ▶ i th row of A is unit vector in direction of i th beacon

prior distribution:

$$x \sim \mathcal{N}(\bar{x}, \Sigma_x), \quad \bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma_x = \begin{bmatrix} 2^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}$$

x_1 has std. dev. 2; x_2 has std. dev. 0.5

90% confidence ellipsoid for prior distribution $\{ x \mid (x - \bar{x})^T \Sigma_x^{-1} (x - \bar{x}) \leq 4.6 \}$:



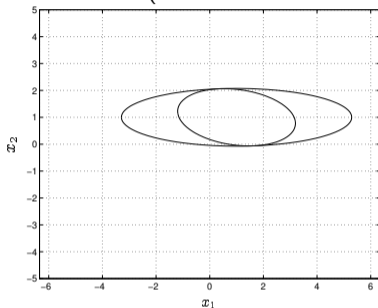
Case 1: one measurement, with beacon at angle 30°

fewer measurements than variables, so combining prior information with measurement is critical

resulting estimation error covariance:

$$\Sigma_{\text{est}} = \begin{bmatrix} 1.046 & -0.107 \\ -0.107 & 0.246 \end{bmatrix}$$

90% confidence ellipsoid for estimate \hat{x} : (and 90% confidence ellipsoid for x)



33

interpretation: measurement

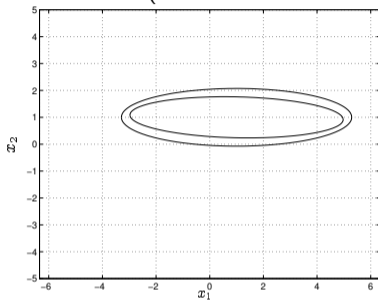
- ▶ yields essentially no reduction in uncertainty in x_2
- ▶ reduces uncertainty in x_1 by a factor about two

Case 2: 4 measurements, with beacon angles 80° , 85° , 90° , 95°

resulting estimation error covariance:

$$\Sigma_{\text{est}} = \begin{bmatrix} 3.429 & -0.074 \\ -0.074 & 0.127 \end{bmatrix}$$

90% confidence ellipsoid for estimate \hat{x} : (and 90% confidence ellipsoid for x)



37

interpretation: measurement yields

- ▶ little reduction in uncertainty in x_1
- ▶ small reduction in uncertainty in x_2