12 - Recursive estimation

- Recursive estimation
- Conditional independence
- Posterior PDFs
- Example: uniform PDFs
- Recursive estimation for Gaussians
- Conditional PDFs for Gaussians
- Alternative formulae
- Information interpretation
- Example: navigation
- Example: recursive estimation of a scalar

Transition Matrices

Suppose

$$y = f(x, w)$$

We interpret

- y is measured
- x is a quantity we would like to estimate
- w is *noise*

Random variables $x : \Omega \to X$, $y : \Omega \to Y$ and $w : \Omega \to W$, where X, Y, W are finite sets.

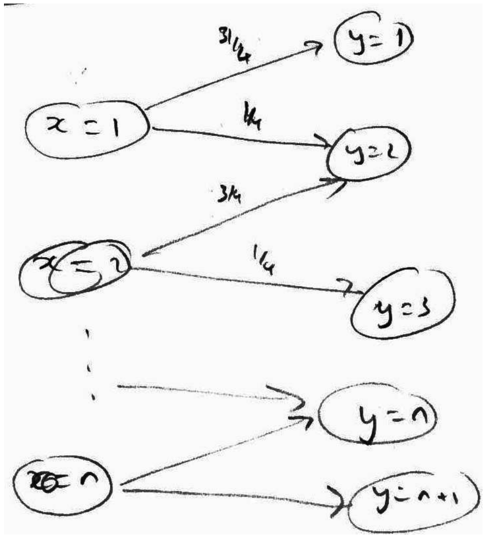
We can represent the *random map* from x to y by the *transition matrix* G given by

$$\boxed{G(q, z) = \mathbf{Prob}(y = q \,|\, x = z)}$$

Example: noisy measurement

Suppose $x : \Omega \to \{1, 2, \dots, n\}$. We measure

$$y = x + w$$



The noise $w : \Omega \to \{0, 1\}$ has pmf

$$\mathbf{Prob}(w = 0) = \frac{3}{4} \qquad \mathbf{Prob}(w = 1) = \frac{1}{4}$$

The *transition matrix* is

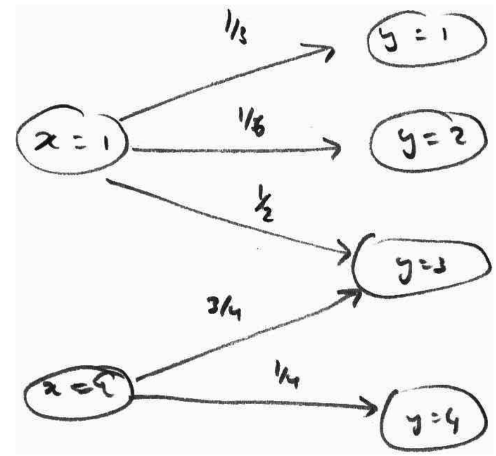
$$G = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & & & \\ & \frac{3}{4} & \frac{1}{4} & & \\ & & \ddots & & \\ & & & \frac{3}{4} & \frac{1}{4} \end{bmatrix}$$

where we use the convention that $G_{ij} = G(j, i)$

Equivalent representations

We can also go the other way, from transition matrix to function. Suppose $x : \Omega \to \{1, 2\}$ and $y : \Omega \to \{1, 4\}$ with transition matrix

$$G = \begin{bmatrix} 1/3 & 1/6 & 1/2 & 0 \\ 0 & 0 & 3/4 & 1/4 \end{bmatrix}$$



We construct a function f and a random variable w so that $y = f(x, w)$. Let $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ where

$$\begin{aligned} \mathbf{Prob}(w_1 = 1) &= 1/3 \\ \mathbf{Prob}(w_1 = 2) &= 1/6 \\ \mathbf{Prob}(w_1 = 3) &= 1/2 \end{aligned} \qquad \begin{aligned} \mathbf{Prob}(w_2 = 3) &= 3/4 \\ \mathbf{Prob}(w_2 = 4) &= 1/2 \end{aligned}$$

Let f be

$$f(x, w) = \begin{cases} w_1 & \text{if } x = 1 \\ w_2 & \text{if } x = 2 \end{cases}$$

For any matrix G we can construct such a function f; it doesn't depend on the prior on x

Transition Matrices

Suppose $y = f(x, w)$ and w has pmf p^w. Suppose

x and w are independent

Then we can find the transition matrix without knowing the prior of x. We have

$$\begin{aligned} G(q, z) &= \mathbf{Prob}(y = q \,|\, x = z) \\ &= \frac{\mathbf{Prob}(f(x, w) = q \text{ and } x = z)}{\mathbf{Prob}(x = z)} \\ &= \frac{\mathbf{Prob}(f(z, w) = q \text{ and } x = z)}{\mathbf{Prob}(x = z)} \\ &= \frac{\mathbf{Prob}(f(z, w) = q)\,\mathbf{Prob}(x = z)}{\mathbf{Prob}(x = z)} \end{aligned}$$

since w and x are independent

$$G(q, z) = \mathbf{Prob}(f(z, w) = q)$$

Continuous random variables

Suppose $x : \Omega \to \mathbb{R}^n$ and $y\Omega \to \mathbb{R}^m$. The transition matrix is replaced by the conditional pdf G defined by

$$\int_A G(q, z)\, dq = \mathbf{Prob}(y \in A \,|\, x = z)$$

for all $A \subset \mathbb{R}^m$.

G is also called a *stochastic kernel*

Linear plus Gaussian

Suppose

$$y = Ax + w \qquad w \sim \mathcal{N}(0, \Sigma)$$

Then the stochastic kernel is

$$\boxed{G(q, z) = f_\Sigma(q - Az)}$$

where f_Σ is the Gaussian pdf for $\mathcal{N}(0, \Sigma)$.

Recursive estimation

Often we have several measurements $y_1, y_2, \ldots, y_m$, and a joint pdf $f(x, y_1, y_2, \ldots, y_m)$

- we receive measurements one at a time
- after measuring y_i, we construct an estimate $\hat{x}_i$
- when we receive y_{i+1}, we would like to *update* $\hat{x}_i$

For example, we often have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix} x + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix}$$

For example, in GPS,

- y_i represents range measurements to satellite i
- When we receive new data, we'd like to update position estimates
- We do not want to have to store old data $y_0, y_1, \ldots, y_{i-1}$

Representation as functiona

More generally

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} f_1(x, w_1) \\ f_2(x, w_2) \\ \vdots \\ f_k(x, w_k) \end{bmatrix}$$

or more succinctly

$$y = f(x, w)$$

where $w = (w_1, w_2, \ldots, w_k)$, etc.

Transition matrix representation

$$G(q_1, q_2, \ldots, q_k, z) = \mathbf{Prob}(y_1 = q_1, \ldots, y_k = q_k \,|\, x = z)$$

or

$$G(q, z) = \mathbf{Prob}(y = q \,|\, x = z)$$

Recursive estimation

We have the following scenario

$$\begin{aligned} y_1 &= f_1(x, w_1) \\ y_2 &= f_2(x, w_2) \\ &\vdots \\ y_k &= f_k(x, w_k) \end{aligned}$$

where $x, w_1, w_2, \dots, w_k$ are *independent*. Then G *factorizes*:

$$\boxed{G(q, z) = G_1(q_1, z) G_2(q_2, z) \dots G_k(q_k, z)}$$

Because

$$\begin{aligned} G(q, z) &= \mathbf{Prob}(f(z, w) = q) \\ &= \mathbf{Prob}(f_1(z, w_1) = q_1, \dots, f_k(z, w_k) = q_k) \\ &= \mathbf{Prob}(f_1(z, w_1) = q_1) \dots \mathbf{Prob}(f_k(z, w_k) = q_k) \end{aligned}$$

Factorization of the pmf

We have

$$G(q, z) = G_1(q_1, z)G_2(q_2, z) \dots G_k(q_k, z)$$

This means

$$\mathbf{Prob}(y = q \,|\, x = z) = \mathbf{Prob}(y_1 = q_1 \,|\, x = z) \dots \mathbf{Prob}(y_k = q_k \,|\, x = z)$$

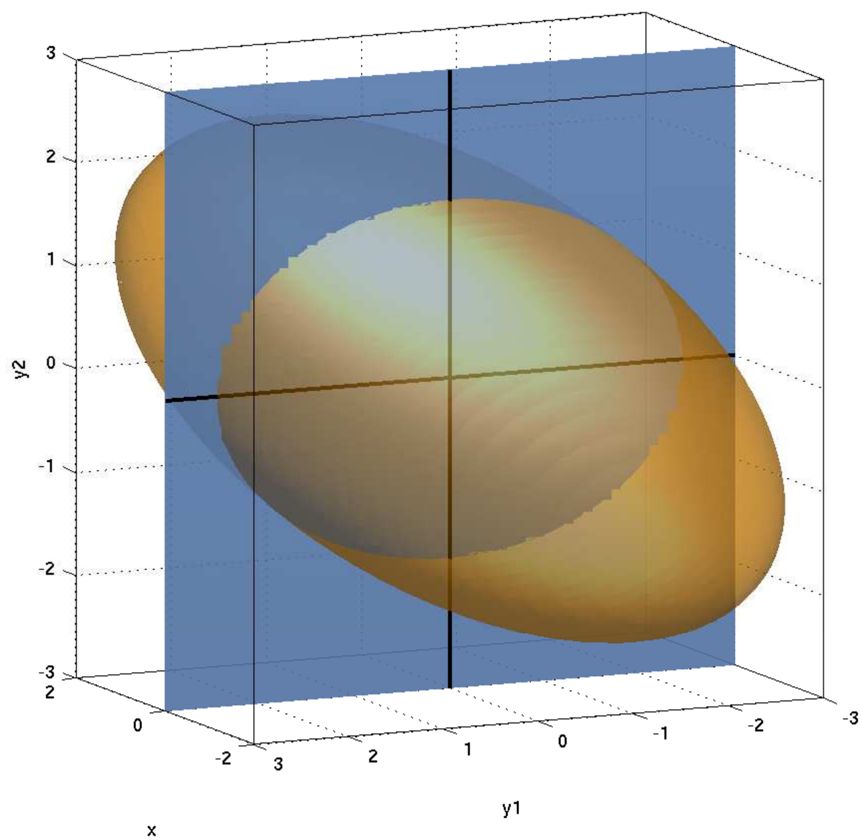
- The random variables $y_1, y_2, \dots, y_k$ are called *conditionally independent*
- This is the key property that allows recursive estimation

Conditional independence

$$y_1 \,|\, x = z \text{ and } y_2 \,|\, x = z \text{ are independent for all } z$$

for example, suppose

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \qquad \text{where} \qquad \mathbf{cov}(w) = \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{cov}(x) = Q = 1$$



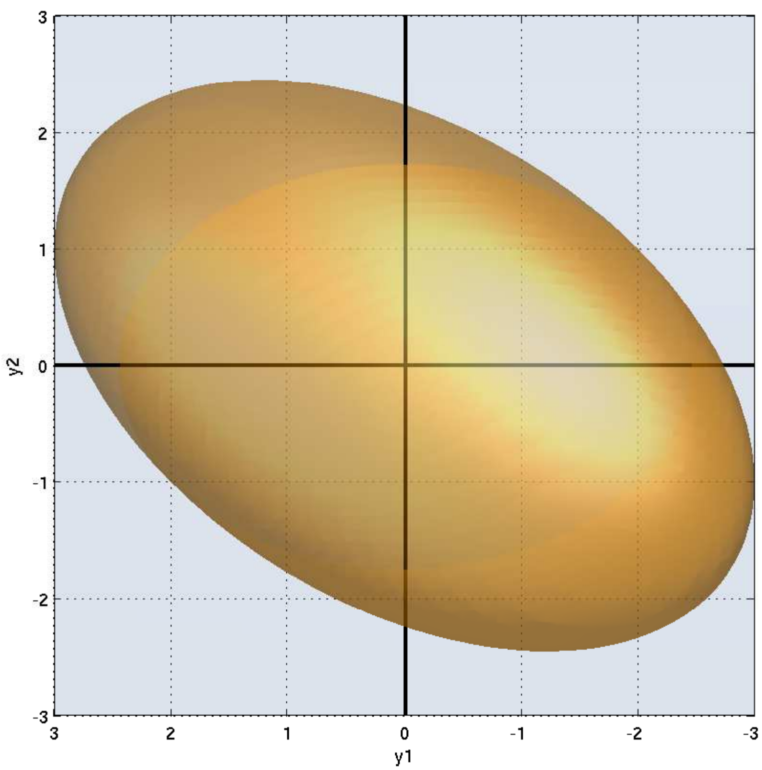
Conditional independence

This does *not* imply that y_1 and y_2 are independent. We have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} A_1 & I & 0 \\ A_2 & 0 & I \end{bmatrix} \begin{bmatrix} x \\ w_1 \\ w_2 \end{bmatrix}$$

hence

$$\mathbf{cov}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} A_1 Q A_1 + \Sigma_1 & A_1 Q A_2 \\ A_2 Q A_1^T & A_2 Q A_2^T + \Sigma_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$



Bayesian estimation review

- Start with
  - prior $p_0(z) = \mathbf{Prob}(x = z)$
  - transition probabilities $G(q, z) = \mathbf{Prob}(y = q \,|\, x = z)$.
- The joint pdf is then

$$\mathbf{Prob}(y = q, x = z) = G(q, z)p_0(z)$$

- Measure $y = y_{\text{meas}}$, and construct posterior $p_1(z, y_{\text{meas}}) = \mathbf{Prob}(x = z \,|\, y = y_{\text{meas}})$

$$p_1(z, y_{\text{meas}}) = \frac{G(y_{\text{meas}}, z)p_0(z)}{\sum_a G(y_{\text{meas}}, a)p_0(a)}$$

- We can then construct an estimate in the usual way; e.g. to minimize a cost function.

Recursive estimation

Let p_t be the *posterior pmf* after measuring $y_1 = q_1, \ldots, y_t = q_t$. By definition

$$p_t(z, q_1, \ldots, q_t) = \frac{G_1(q_1, z) \ldots G_t(q_t, z) p_t(z)}{\sum_a G_1(q_1, a) \ldots G_t(q_t, a) p_t(a)}$$

- We would like to use the posterior pdf p_t after measuring $y_1, \ldots, y_t$ as the prior pdf when we receive measurement y_{t+1}.
- It turns out that this is possible when y_1 and y_2 are conditionally independent.
- And we can forget

  the previous measurements

  where they came from; i.e. the sensors $G_1, \ldots, G_t$

  So we can do *sensor fusion*

Recursive estimation

The main result: if $y_1, \ldots, y_k$ are conditionally independent, then

$$p_{t+1}(z) = \frac{G_{t+1}(q_{t+1}, z) p_t(z)}{\sum_a G_{t+1}(q_{t+1}, a) p_t(a)}$$

- We omit the dependence of p_t on $q_1, \ldots, q_t$.
- If $X = \{1, 2, \ldots, n\}$ then we implement this by storing p_t as a *vector* in $\mathbb{R}^n$.
- p_t is called the *belief state*. It is the only quantity we need to store.
- The history of observations $q_1, \ldots, q_t$ is called the *information state*

Proof

Since p_{t+1} is the posterior given $y_1, \ldots, y_t$, it is by definition

$$p_{t+1}(z) = \frac{p_0(z)G_1(q_1, z) \ldots G_t(q_t, z)G_{t+1}(q_{t+1}, z)}{\sum_a p_0(a)G_1(q_1, a) \ldots G_t(q_t, a)G_{t+1}(q_{t+1}, a)}$$

Now substitute into this expression the definition of p_t to give

$$= \frac{p_t(z)\Big(\sum_b p_0(z)G_1(q_1, b) \ldots G_t(q_t, b)\Big)G_{t+1}(q_{t+1}, z)}{\sum_a p_0(a)G_1(q_1, a) \ldots G_t(q_t, a)G_{t+1}(q_{t+1}, a)}$$

$$= \frac{p_t(z)\Big(\sum_b p_0(b)G_1(q_1, b) \ldots G_t(q_t, b)\Big)G_{t+1}(q_{t+1}, z)}{\sum_a p_t(a)\Big(\sum_c p_0(c)G_1(q_1, c) \ldots G_t(q_t, c)\Big)G_{t+1}(q_{t+1}, a)}$$

$$= \frac{p_t(z)G_{t+1}(q_{t+1}, z)}{\sum_a p_t(a)G_{t+1}(q_{t+1}, a)}$$

as desired.

Continuous case

It's almost the same:

$$p_{t+1}(z) = \frac{p_t(z)G_{t+1}(q_{t+1}, z)}{\displaystyle\int_{a \in \mathbb{R}^n} p_t(a)G_{t+1}(q_{t+1}, a)\, da}$$

The proof is the same as in the discrete case.

Recursive estimation with linear measurements and Gaussian noise

Suppose we have

$$y = Ax + w$$

where x and w are independent, and $x \sim \mathcal{N}(\hat{x}_0, Q_0)$ and $w \sim \mathcal{N}(0, \Sigma)$.

This is equivalent to

- x has prior $x \sim \mathcal{N}(\hat{x}_0, Q_0)$
- $y \,|\, (x = z)$ has pdf $\mathcal{N}(Az, \Sigma)$
  because the joint pdf is $p(x, y) = p^x(x) p^w(y - Ax)$

Then x, y are jointly Gaussian, with

$$\mathbf{E} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \hat{x}_0 \\ A\hat{x}_0 \end{bmatrix} \qquad \mathbf{cov} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Q_0 & Q_0 A^T \\ AQ_0 & AQ_0A^T + \Sigma \end{bmatrix}$$

Recursive estimation with Gaussian noise

Let's consider the problem

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} x + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$$

where

- x has prior pdf $\mathcal{N}(\hat{x}_0, Q_0)$
- w_i has pdf $\mathcal{N}(0, \Sigma_i)$
- w_i and w_j are independent if $i \neq j$

Recursive estimation with Gaussian noise

The conditional covariance of y given $x = z$ is

$$\mathbf{cov}(y \,|\, x = z) = \begin{bmatrix} \Sigma_1 & & & \\ & \Sigma_2 & & \\ & & \ddots & \\ & & & \Sigma_m \end{bmatrix}$$

and hence y_i and y_j are conditionally independent.

Gaussians are special

We *could* just apply the formula

$$p_{t+1}(z) = \frac{p_t(z) G_{t+1}(q_{t+1}, z)}{\displaystyle\int_{a \in \mathbb{R}^n} p_t(a) G_{t+1}(q_{t+1}, a)\, da}$$

because we know $G_t(q_t, z) = f_{\Sigma_t}(q_t - A_t z)$.

But *we don't need to.* Because

- we know p_0 is Gaussian.
- Hence the posterior p_1 will be Gaussian, and we know it's mean and covariance, so we know it completely
- Hence the posterior p_2 will be Gaussian, . . .
- The idea: we don't need to store p_t. Since it's Gaussian, it is characterized completely by its mean and covariance.

Recursive estimation with Gaussian noise

We know how to do Bayesian estimation for Gaussians; that is, if

- x has prior $x \sim \mathcal{N}(\hat{x}_0, Q_0)$
- $y_1 \,|\, (x = z)$ has pdf $\mathcal{N}(A_1 z, \Sigma_1)$

Then the posterior pdf $h_1(x, y_{1\text{meas}})$ of $x \,|\, (y_1 = y_{1\text{meas}})$ is $\mathcal{N}(\hat{x}_1, Q_1)$ where

$$\hat{x}_1 = \hat{x}_0 + Q_0 A_1^T (A_1 Q_0 A_1^T + \Sigma_1)^{-1} (y_{\text{meas}} - A_1 \hat{x}_0)$$

$$Q_1 = Q_0 - Q_0 A_1^T (A_1 Q_0 A_1^T + \Sigma_1)^{-1} A_1 Q_0$$

Recursive estimation with Gaussian noise

Now since y_i and y_j are conditionally independent for $i \neq j$, we can use the posterior pdf of $x \mid (y_1 = y_{1\text{meas}}$ as the prior pdf for the next measurement.

So, after measuring y_1, we have new prior

$$x \mid (y_1 = y_{1\text{meas}}) \sim \mathcal{N}(\hat{x}_1, Q_1)$$

Also the conditional pdf for $y_2 \mid (x = z)$ is $\mathcal{N}(A_2 z, \Sigma_2)$

And so we can apply exactly the same estimator as before.

Summary: recursive estimation with Gaussians noise

Set $k = 0$; repeat

1. *update the covariance*

$$Q_{k+1} = Q_k - Q_k A_{k+1}^T (A_{k+1} Q_k A_{k+1}^T + \Sigma_{k+1})^{-1} A_{k+1} Q_k$$

2. *update the estimate*

$$\hat{x}_{k+1} = \hat{x}_k + Q_k A_{k+1}^T (A_{k+1} Q_k A_{k+1}^T + \Sigma_{k+1})^{-1} (y_{k+1} - A_{k+1} \hat{x}_k)$$

3. $k \mapsto k + 1$

Alternative formulae

Set $k = 0$; repeat

1. *update the covariance*

$$\boxed{Q_{k+1}^{-1} = Q_k^{-1} + A_{k+1}^T \Sigma_{k+1}^{-1} A_{k+1}}$$

2. *update the estimate*

$$\boxed{\hat{x}_{k+1} = \hat{x}_k + Q_{k+1} A_{k+1}^T \Sigma_{k+1}^{-1} \bigl(y_{k+1} - A_{k+1}\hat{x}_k\bigr)}$$

3. $k \mapsto k + 1$

Information interpretation

with each new measurement, we have

$$Q_{k+1}^{-1} = Q_k^{-1} + A_{k+1}^T \Sigma_{k+1}^{-1} A_{k+1}$$

inverse of covariance matrix Q_i is called the *information matrix*

information matrices *add* when combining data

we have $Q_{k+1}^{-1} \geq Q_k^{-1}$, i.e., with each measurement, our information increases

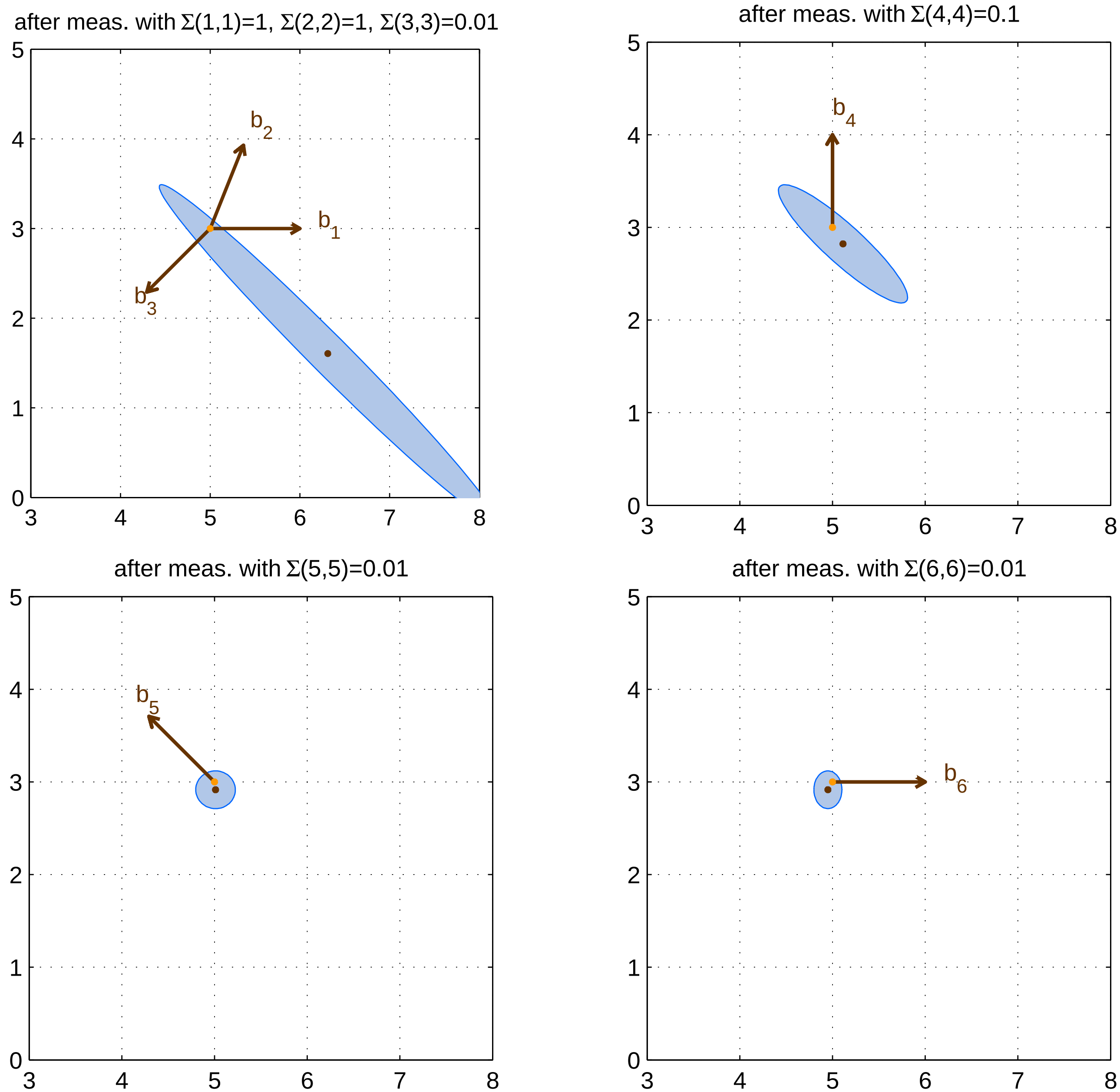
mean-square-error

this is equivalent to $Q_{k+1} \leq Q_k$, and so the mean-square error satisfies

$$\begin{aligned} \mathbf{E}\|x - \hat{x}_{k+1}\|^2 &= \mathbf{trace}\, Q_{k+1} \\ &= \sum_{i=1}^{n} e_i^T Q_{k+1} e_i \\ &\leq \mathbf{trace}\, Q_k \\ &= \mathbf{E}\|x - \hat{x}_k\|^2 \end{aligned}$$

i.e. the mean-square error is non-increasing

Example: navigation



Example: recursive estimation of a scalar

suppose

$$y_i = x + w_i \qquad \text{for } i = 1, \dots, k$$

and $w_i \sim \mathcal{N}(0, 1)$, and w_i, w_j are independent when $i \neq j$

Now assume prior $x \sim \mathcal{N}(0, 1)$. We know

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

and so, for any p

$$\hat{x}_p = \frac{1}{p+1} \sum_{i=1}^{p} y_i$$

This tends to the sample mean of the measurements; as expected it is biased by the prior.

Example: recursive estimation of a scalar

We have

$$Q_{k+1}^{-1} = Q_k^{-1} + 1$$

and therefore $Q_k = \dfrac{1}{k+1}$.

Then the recursive estimator is

$$\begin{aligned} \hat{x}_{k+1} &= \hat{x}_k + Q_{k+1}(y_{k+1} - \hat{x}_k) \\ &= \frac{k+1}{k+2}\hat{x}_k + \frac{1}{k+2}y_{k+1} \end{aligned}$$

so given y_{t+1} and we can update $\hat{x}_t$ find $\hat{x}_{t+1}$; don't need to remember $y_1, \ldots, y_t$

- Notice that the error covariance $Q_k \to 0$
- As time k becomes large, the data has no effect.
- However, if x is changing, we need the estimator to respond to this; as we will see, the *Kalman filter* is a remedy for this problem.