What does a Kalman Filter do, anyway?

Given the linear dynamical system:

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k)$$

$$y(k) = H(k)x(k) + w(k)$$

$x(k)$ is the n-dimensional state vector (unknown)

$u(k)$ is the m-dimensional input vector (known)

$y(k)$ is the p-dimensional output vector (known, measured)

$F(k), G(k), H(k)$ are appropriately dimensioned system matrices (known)

$v(k), w(k)$ are zero-mean, white Gaussian noise with (known)

covariance matrices $Q(k), R(k)$

the Kalman Filter is a recursion that provides the "best" estimate of the state vector *x*.

What's so great about that?

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k)$$
$$y(k) = H(k)x(k) + w(k)$$

- noise smoothing (improve noisy measurements) [filter]
- state estimation (for state feedback) [observer]
- recursive (computes next estimate using only most recent measurement)

How does it work?

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k)$$

$$y(k) = H(k)x(k) + w(k)$$

1. prediction based on last estimate:

$$\hat{x}(k+1 \mid k) = F(k)\hat{x}(k \mid k) + G(k)u(k)$$

$$\hat{y}(k) = H(k)\hat{x}(k+1 \mid k)$$

2. calculate correction based on prediction and current measurement:

$$\Delta x = f\left(y(k+1), \hat{x}(k+1 \mid k)\right)$$

3. update prediction: $\hat{x}(k+1 \mid k+1) = \hat{x}(k+1 \mid k) + \Delta x$

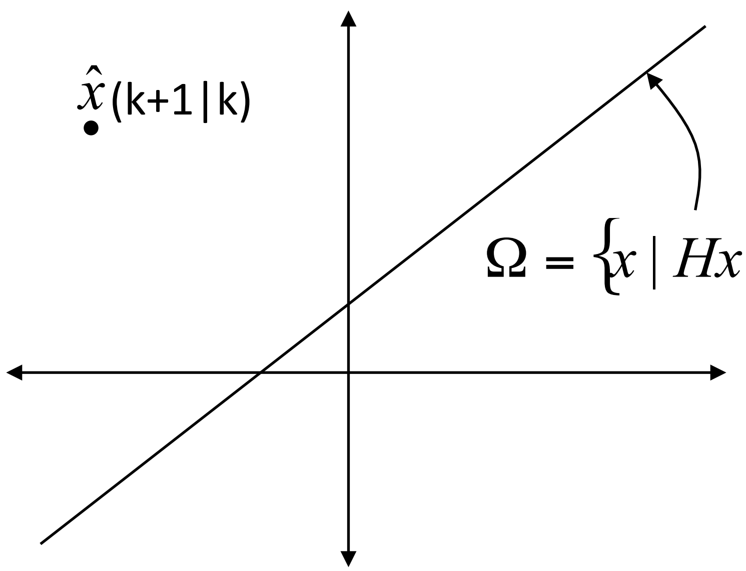
Finding the correction (no noise!)

⇧

Let's be overly naive for a moment

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



Want the best estimate to be consistent with sensor readings

"best" estimate comes from shortest Δx

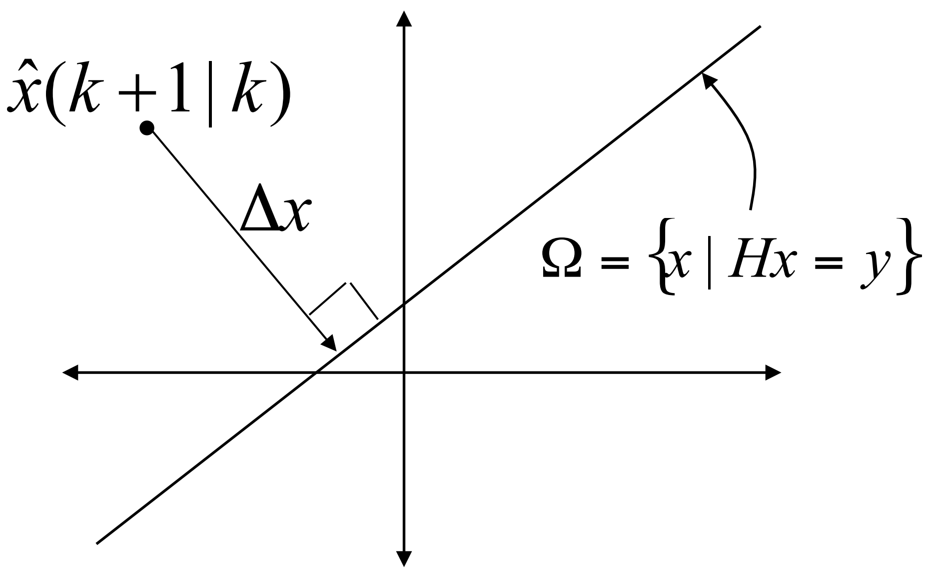
Finding the correction (no noise!)

$$y = Hx$$

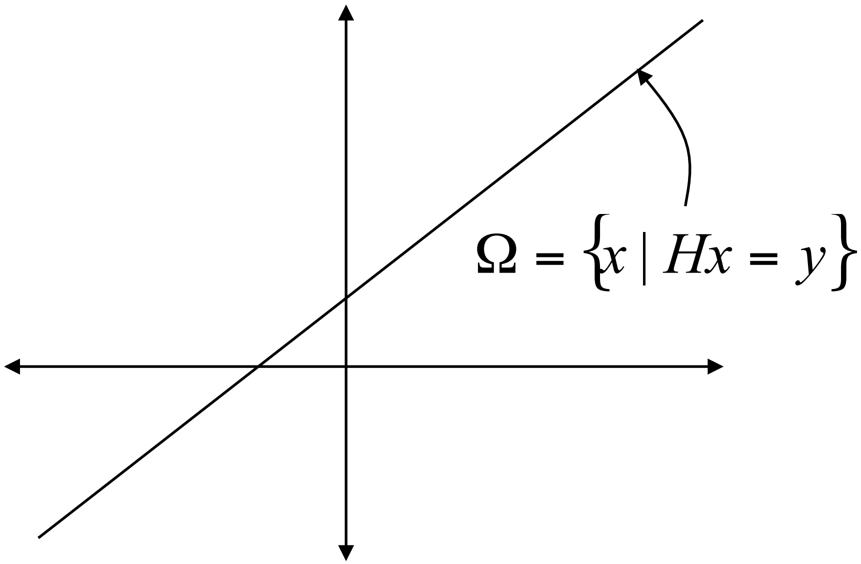
Given prediction $\hat{x}(k+1|1)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|1) + \Delta x$ is the "best" estimate of x.

"best" estimate comes from shortest Δx

shortest Δx is perpendicular to Ω



Some linear algebra



a is parallel to Ω if Ha = 0

$$Null(H) = \{a \neq 0 \mid Ha = 0\}$$

a is parallel to Ω if it lies in the null space of H

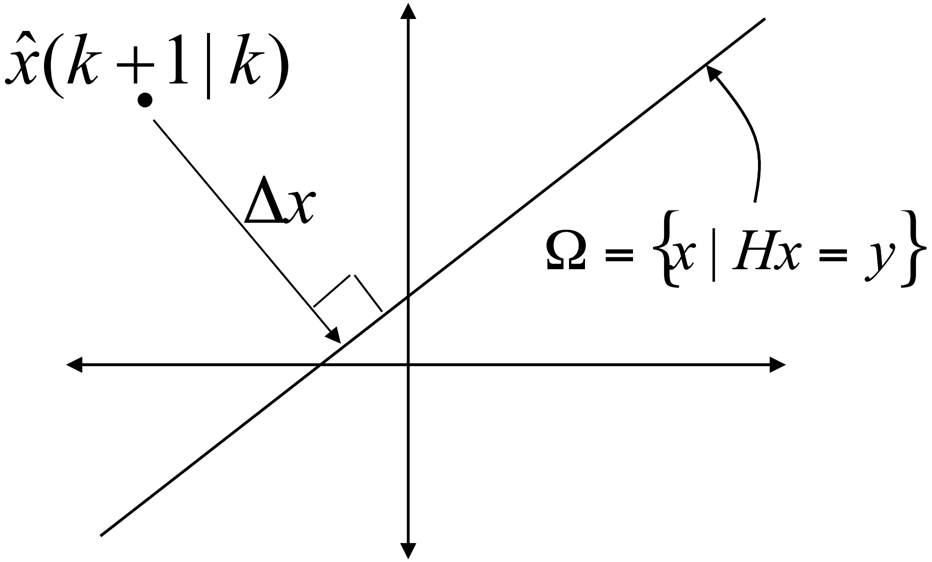
$$\text{for all } v \in Null(H), v \perp b \text{ if } b \in column(H^T)$$

Weighted sum of columns means $b = H\gamma$, the weighted sum of columns

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



"best" estimate comes from shortest Δx

shortest Δx is perpendicular to Ω

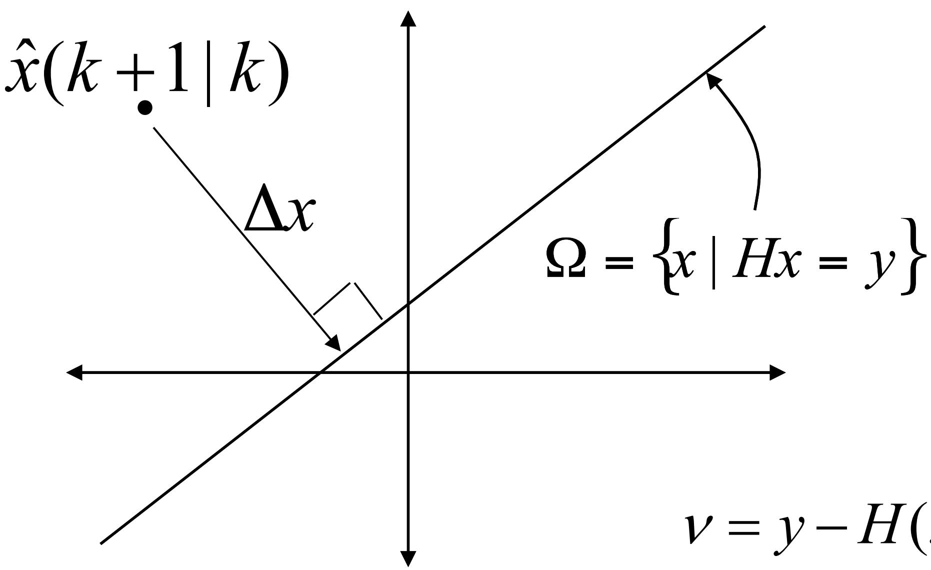
$$\Rightarrow \Delta x \in \text{null}(H)^{\perp} \Rightarrow \Delta x \in \text{column}(H^T)$$

$$\Rightarrow \Delta x = H^T \gamma$$

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



"best" estimate comes from shortest Δx

shortest Δx is perpendicular to Ω

$$\Rightarrow \Delta x \in \text{null}(H)^{\perp} \Rightarrow \Delta x \in \text{column}(H^T)$$

$$\Rightarrow \Delta x = H^T \gamma$$

Real output – estimated output

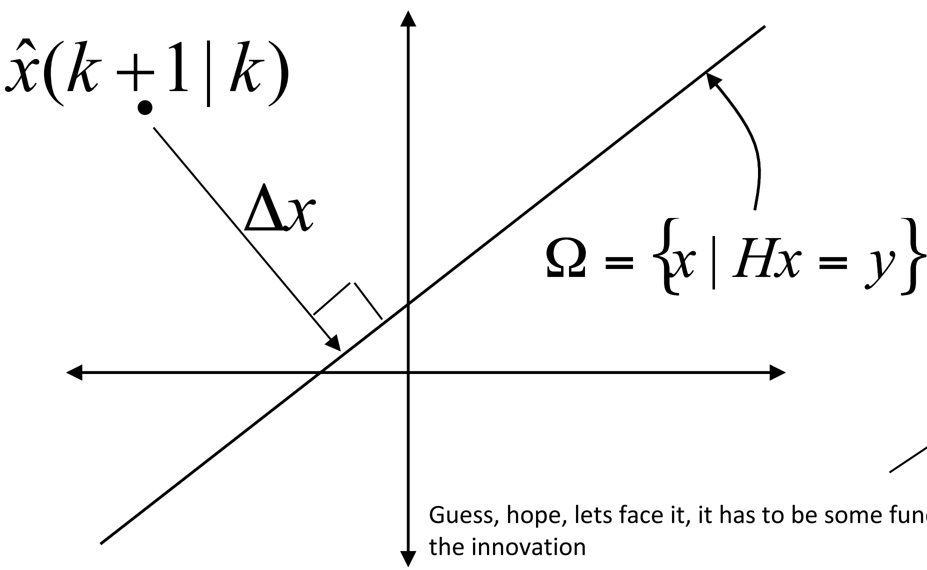
$$\nu = y - H(\hat{x}(k+1|k)) = H(x(k+1) - \hat{x}(k+1|k))$$

innovation

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1\,|\,k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1\,|\,k) + \Delta x$ is the "best" estimate of x.



"best" estimate comes from shortest Δx

shortest Δx is perpendicular to Ω

$$\Rightarrow \Delta x \in \operatorname{null}(H)^{\perp} \Rightarrow \Delta x \in \operatorname{column}(H^T)$$

$$\Rightarrow \Delta x = H^T \gamma$$

assume γ is a linear function of ν

$$\Rightarrow \Delta x = H^T K \nu$$

for some $m \times m$ matrix K

Guess, hope, lets face it, it has to be some function of the innovation

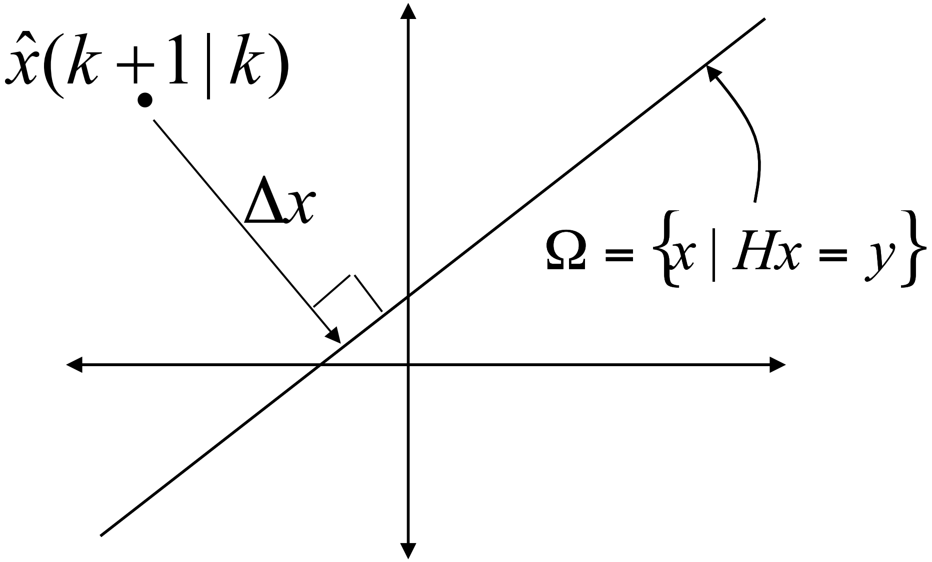
Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.

we require

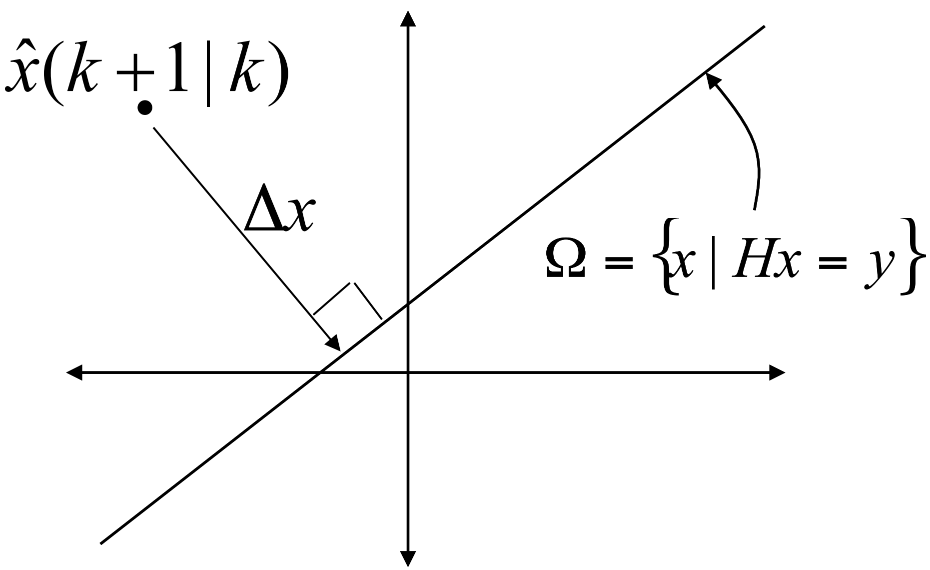
$$H(\hat{x}(k+1|k) + \Delta x) = y$$



Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



we require

$$H(\hat{x}(k+1|k) + \Delta x) = y$$

$$\Rightarrow H\Delta x = y - H\hat{x}(k+1|k) = H(x - \hat{x}(k+1|k)) = \nu$$

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1\,|\,k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1\,|\,k) + \Delta x$ is the "best" estimate of x.

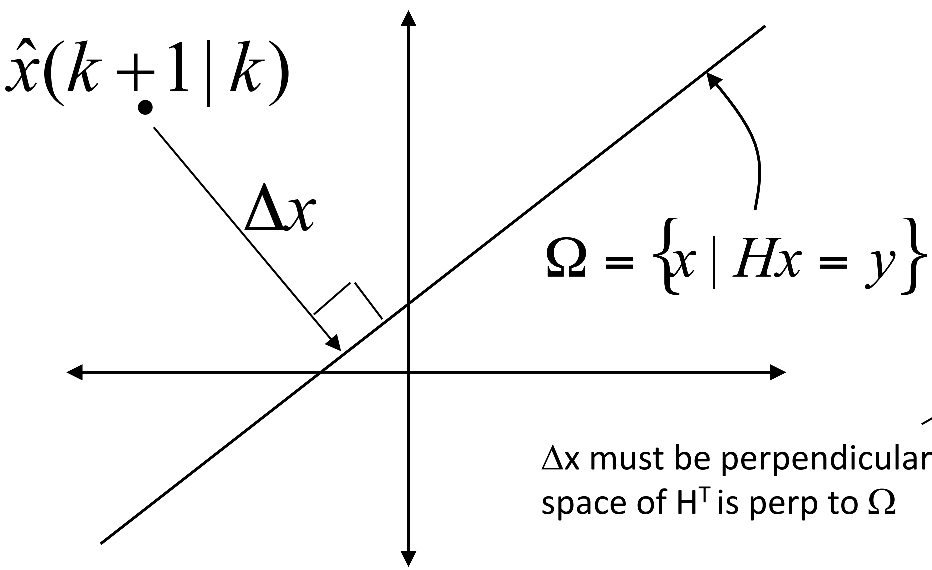
we require

$$H(\hat{x}(k+1\,|\,k) + \Delta x) = y$$

$$\Rightarrow H\Delta x = y - H\hat{x}(k+1\,|\,k) = H(x - \hat{x}(k+1\,|\,k)) = \nu$$

substituting $\Delta x = H^T K \nu$ yields

$$HH^T K\nu = \nu$$

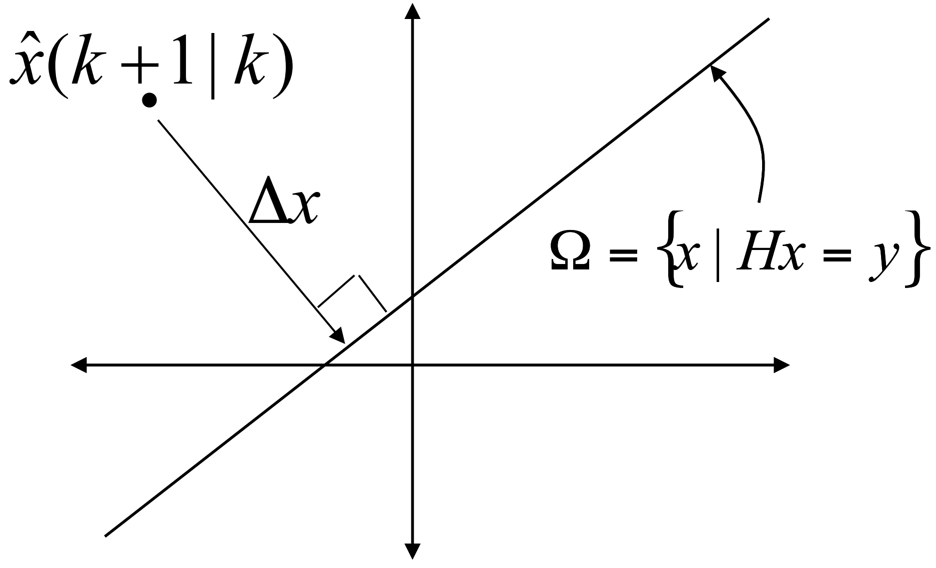


Δx must be perpendicular to Ω b/c anything in the range space of H^T is perp to Ω

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



we require

$$H(\hat{x}(k+1|k) + \Delta x) = y$$

$$\Rightarrow H\Delta x = y - H\hat{x}(k+1|k) = H(x - \hat{x}(k+1|k)) = \nu$$

substituting $\Delta x = H^T K \nu$ yields

$$HH^T K \nu = \nu$$

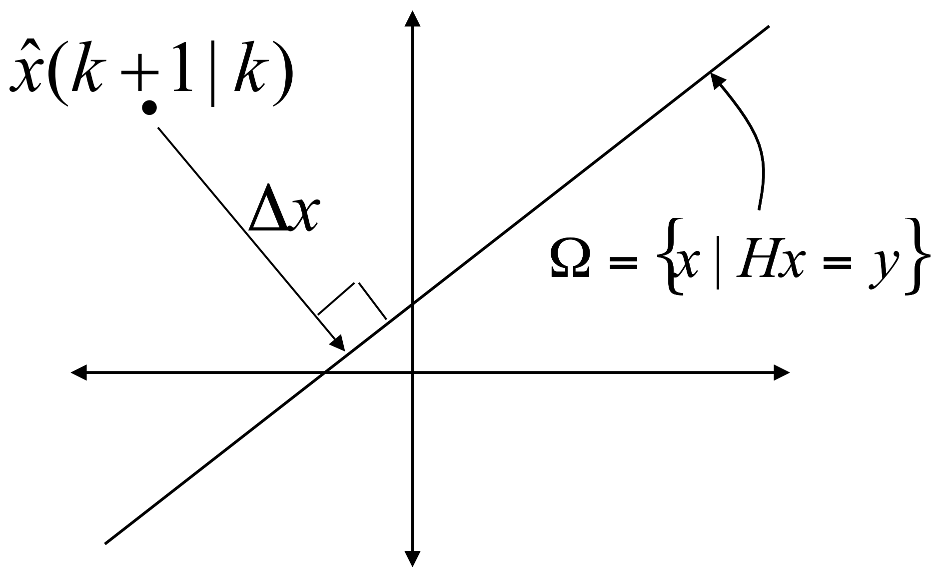
$$\Rightarrow K = \left(HH^T\right)^{-1}$$

The fact that the linear solution solves the equation makes assuming K is linear a kosher guess

Finding the correction (no noise!)

$$y = Hx$$

Given prediction $\hat{x}(k+1|k)$ and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" estimate of x.



we require

$$H(\hat{x}(k+1|k) + \Delta x) = y$$

$$\Rightarrow H\Delta x = y - H\hat{x}(k+1|k) = H(x - \hat{x}(k+1|k)) = \nu$$

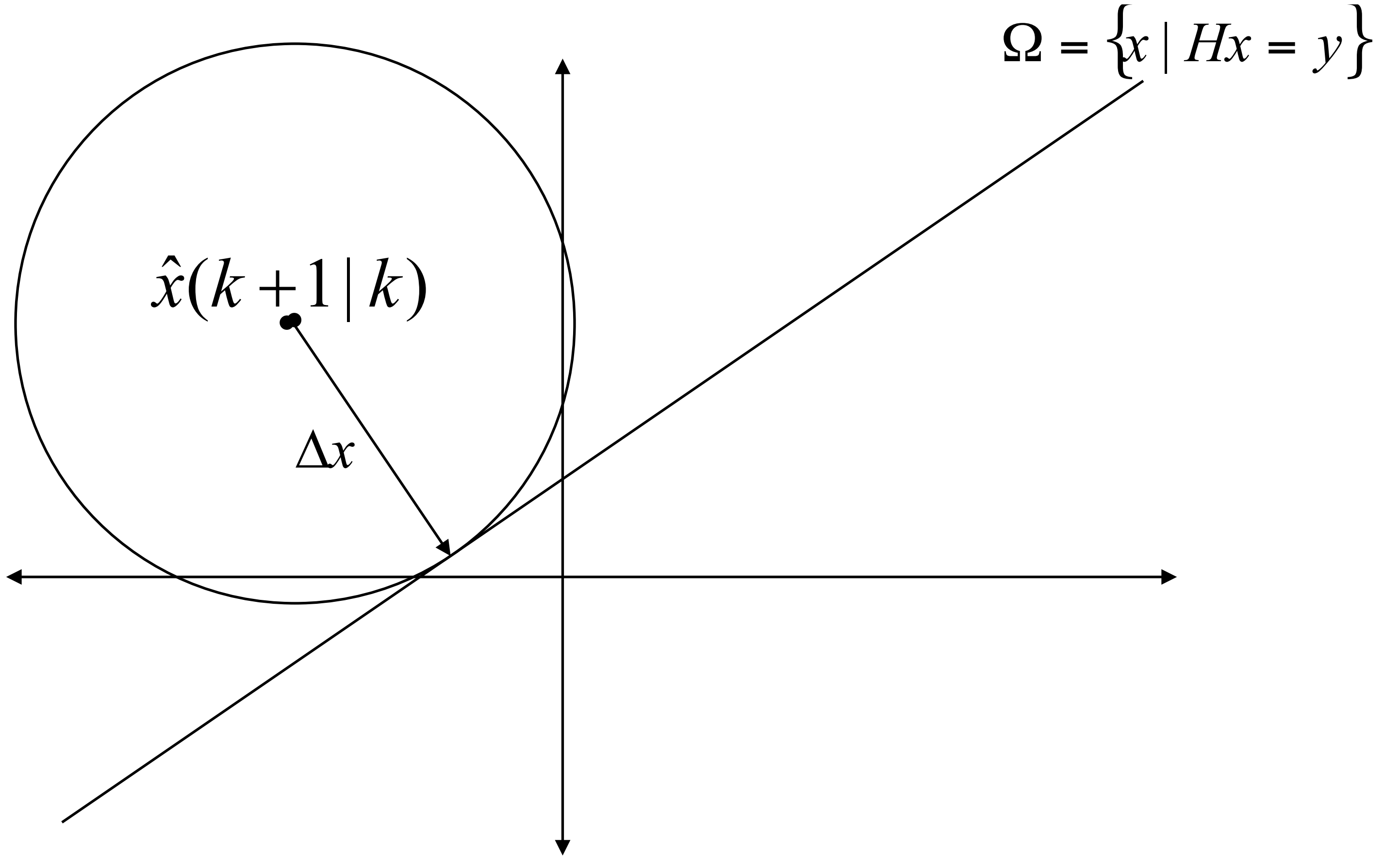
substituting $\Delta x = H^T K \nu$ yields

$$HH^T K \nu = \nu$$

$$\Rightarrow K = \left(HH^T\right)^{-1}$$

$$\boxed{\Delta x = H^T \left(HH^T\right)^{-1} \nu}$$

A Geometric Interpretation



A Simple State Observer

System:

$$x(k+1) = Fx(k) + Gu(k)$$
$$y(k) = Hx(k)$$

Observer:

1. prediction:

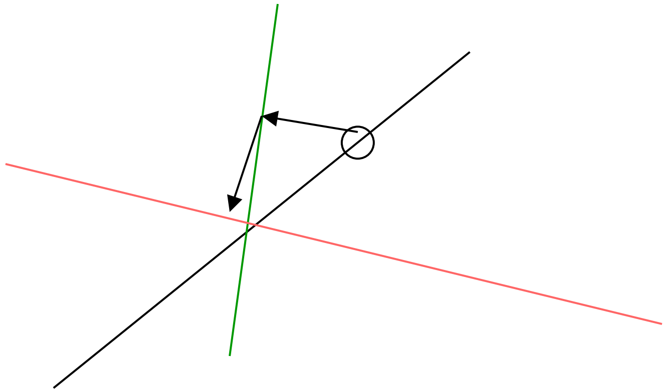
$$\hat{x}(k+1 \mid k) = F\hat{x}(k \mid k) + Gu(k)$$

2. compute correction:

$$\Delta x = H^T \left(HH^T\right)^{-1} \left(y(k+1) - H\hat{x}(k+1 \mid k)\right)$$

3. update:

$$\hat{x}(k+1 \mid k+1) = \hat{x}(k+1 \mid k) + \Delta x$$



Caveat #1

Note: The observer presented here is not a very good observer. Specifically, it is not guaranteed to converge for all systems. Still the intuition behind this observer is the same as the intuition behind the Kalman filter, and the problems will be fixed in the following slides.

It really corrects only to the current sensor information, so if you are on the hyperplane but not at right place, you have no correction..Estimate errors parallel to hyperplane are not corrected….

And we are still assuming no noise

Estimating a *distribution* for *x*

Distribution for state x is parameterized by mean and covariance

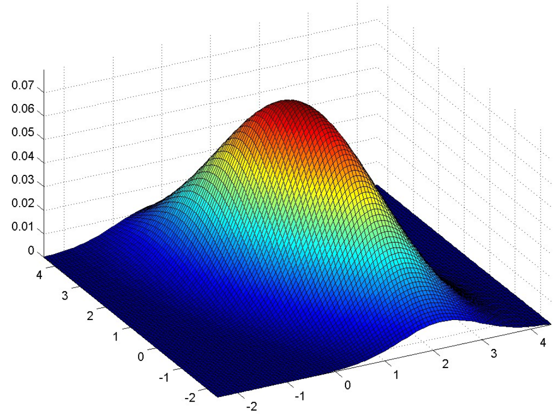
We are estimating the mean of x, hence the mean and estimate are the same

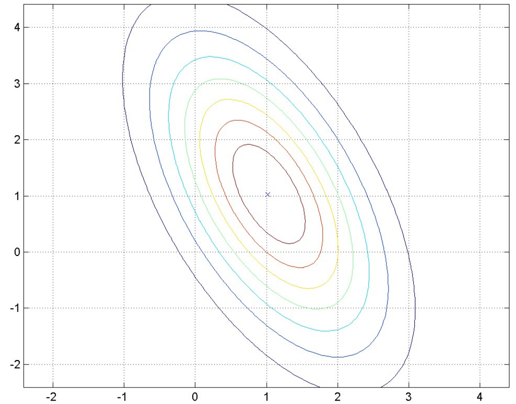
Our estimate of *x* is not exact!

We can do better by estimating a joint Gaussian distribution *p(x)*.

$$p(x) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} e^{\frac{-1}{2}\left((x-\hat{x})^T P^{-1}(x-\hat{x})\right)}$$

where $P = E\left((x-\hat{x})(x-\hat{x})^T\right)$ is the *covariance matrix*





A Better State Observer

$$x(k+1) = Fx(k) + Gu(k) + v(k)$$

$$y(k) = Hx(k)$$

Sample of Guassian Dist. w/ COV Q

We can create a better state observer following the same 3. steps, but now we must also estimate the covariance matrix P.

We start with $x(k|k)$ and $P(k|k)$

Step 1: Prediction

Where did noise go? Expected value…

$$\boxed{\hat{x}(k+1 \mid k) = F\hat{x}(k \mid k) + Gu(k)}$$

What about P? From the definition:

$$P(k \mid k) = E\left((x(k) - \hat{x}(k \mid k))(x(k) - \hat{x}(k \mid k))^T\right)$$

and

$$P(k+1 \mid k) = E\left((x(k+1) - \hat{x}(k+1 \mid k))(x(k+1) - \hat{x}(k+1 \mid k))^T\right)$$

Continuing Step 1

To make life a little easier, lets shift notation slightly:

$$
\begin{aligned}
P_{k+1}^{-} &= E\left((x_{k+1}-\hat{x}_{k+1}^{-})(x_{k+1}-\hat{x}_{k+1}^{-})^{T}\right) \\
&= E\left(\left(Fx_k+Gu_k+v_k-(F\hat{x}_k+Gu_k)\right)\left(Fx_k+Gu_k+v_k-(F\hat{x}_k+Gu_k)\right)^{T}\right) \\
&= E\left(\left(F\left(x_k-\hat{x}_k\right)+v_k\right)\left(F\left(x_k-\hat{x}_k\right)+v_k\right)^{T}\right) \\
&= E\left(F\left(x_k-\hat{x}_k\right)\left(x_k-\hat{x}_k\right)^{T}F^{T}+2F\left(x_k-\hat{x}_k\right)v_k^{T}+v_kv_k^{T}\right) \\
&= FE\left(\left(x_k-\hat{x}_k\right)\left(x_k-\hat{x}_k\right)^{T}\right)F^{T}+E\left(v_kv_k^{T}\right) \\
&= FP_kF^{T}+Q
\end{aligned}
$$

$$\boxed{P(k+1\,|\,k)=FP(k\,|\,k)F^{T}+Q}$$

Continuing Step 1

To make life a little easier, lets shift notation slightly:

$$
\begin{aligned}
P_{k+1}^{-} &= E\left((x_{k+1} - \hat{x}_{k+1}^{-})(x_{k+1} - \hat{x}_{k+1}^{-})^T\right) \\
&= E\left(\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)^T\right) \\
&= E\left(\left(F\left(x_k - \hat{x}_k\right) + v_k\right)\left(F\left(x_k - \hat{x}_k\right) + v_k\right)^T\right) \\
&= E\left(F\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T F^T + 2F\left(x_k - \hat{x}_k\right)v_k^T + v_k v_k^T\right) \\
&= FE\left(\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right)F^T + E\left(v_k v_k^T\right) \\
&= FP_k F^T + Q
\end{aligned}
$$

$$\boxed{P(k+1 \mid k) = FP(k \mid k)F^T + Q}$$

Continuing Step 1

To make life a little easier, lets shift notation slightly:

$$P_{k+1}^{-} = E\left((x_{k+1} - \hat{x}_{k+1}^{-})(x_{k+1} - \hat{x}_{k+1}^{-})^{T}\right)$$

$$= E\left(\left(\underline{Fx_k} + \cancel{Gu_k} + \underline{v_k} - (\underline{F\hat{x}_k} + \cancel{Gu_k})\right)\left(\underline{Fx_k} + \cancel{Gu_k} + \underline{v_k} - (\underline{F\hat{x}_k} + \cancel{Gu_k})\right)^{T}\right)$$

$$= E\left(\left(\underline{F(x_k - \hat{x}_k)} + \underline{v_k}\right)\left(\underline{F(x_k - \hat{x}_k)} + \underline{v_k}\right)^{T}\right)$$

$$= E\left(F(x_k - \hat{x}_k)(x_k - \hat{x}_k)^{T} F^{T} + 2F(x_k - \hat{x}_k)v_k^{T} + v_k v_k^{T}\right)$$

$$= FE\left((x_k - \hat{x}_k)(x_k - \hat{x}_k)^{T}\right)F^{T} + E\left(v_k v_k^{T}\right)$$

$$= FP_k F^{T} + Q$$

$$\boxed{P(k+1 \mid k) = FP(k \mid k)F^{T} + Q}$$

Continuing Step 1

To make life a little easier, lets shift notation slightly:

$$P_{k+1}^{-} = E\left((x_{k+1} - \hat{x}_{k+1}^{-})(x_{k+1} - \hat{x}_{k+1}^{-})^T\right)$$

$$= E\left(\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)^T\right)$$

$v(k)$ is independent of both $x(k)$ and $\hat{x}(k|k)$

$$E\left((x_k - \hat{x}_k)v_k^T\right) = E\left((x_k - \hat{x}_k)\right)E(v_k^T) = 0$$

$$= E\left(\left(F(x_k - \hat{x}_k) + v_k\right)\left(F(x_k - \hat{x}_k) + v_k\right)^T\right)$$

$$= E\left(\underline{F(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T F^T} + 2F(x_k - \hat{x}_k)v_k^T + \underline{v_k v_k^T}\right)$$

$$= \underline{FE\left((x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right)F^T} + \underline{E\left(v_k v_k^T\right)}$$

$$= FP_k F^T + Q$$

$$\boxed{P(k+1|k) = FP(k|k)F^T + Q}$$

Continuing Step 1

To make life a little easier, lets shift notation slightly:

$$P_{k+1}^{-} = E\left((x_{k+1} - \hat{x}_{k+1}^{-})(x_{k+1} - \hat{x}_{k+1}^{-})^T\right)$$

$$= E\left(\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)\left(Fx_k + Gu_k + v_k - (F\hat{x}_k + Gu_k)\right)^T\right)$$

$$= E\left(\left(F\left(x_k - \hat{x}_k\right) + v_k\right)\left(F\left(x_k - \hat{x}_k\right) + v_k\right)^T\right)$$

$$= E\left(F\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T F^T + 2F\left(x_k - \hat{x}_k\right)v_k^T + v_k v_k^T\right)$$

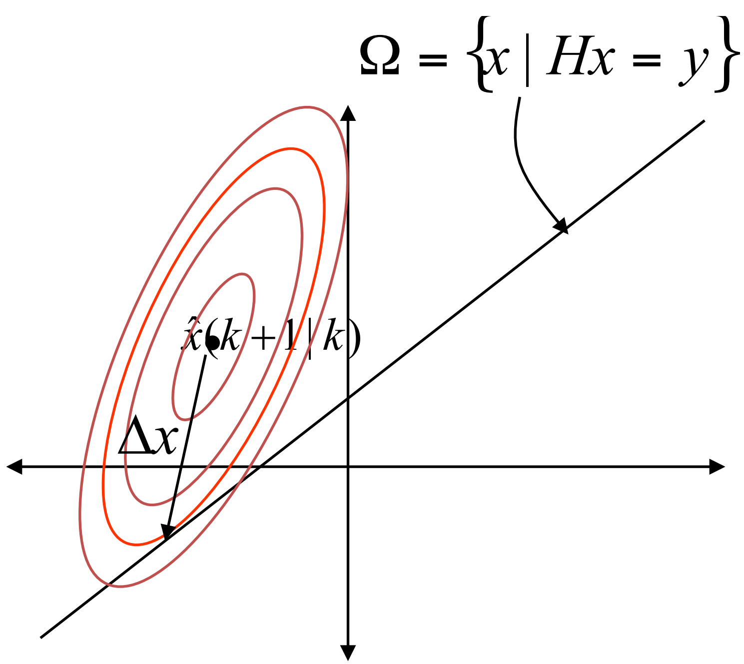
$$= FE\left(\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right)F^T + E\left(v_k v_k^T\right)$$

$$= FP_k F^T + Q$$

$$\boxed{P(k+1 \mid k) = FP(k \mid k)F^T + Q}$$

Finding the correction (geometric intuition)

Given prediction $\hat{x}(k+1|k)$, covariance P, and output y, find Δx so that $\hat{x} = \hat{x}(k+1|k) + \Delta x$ is the "best" (i.e. most probable) estimate of x.



$$p(x) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} e^{\frac{-1}{2}\left((x-\hat{x})^T P^{-1}(x-\hat{x})\right)}$$

The most probable Δx is the one that:

1. satisfies $\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + \Delta x$
2. maximumize p(x) $\rightarrow$ minimizes $\Delta x^T P^{-1} \Delta x$

A new kind of distance

Suppose we define a new inner product on R^n to be :

$$\langle x_1, x_2 \rangle = x_1^T P^{-1} x_2 \qquad \text{(this replaces the old inner product } x_1^T x_2\text{)}$$

Then we can define a new norm $\|x\|^2 = \langle x, x \rangle = x^T P^{-1} x$

The $\hat{x}$ in Ω that minimizes $\|\Delta x\|$ is the orthogonal projection of $\hat{x}(k+1|k)$ onto Ω, so Δx is orthogonal to Ω.

$$\Rightarrow \langle \omega, \Delta x \rangle = 0 \text{ for } \omega \text{ in } \Omega = null(H)$$

$$\boxed{\langle \omega, \Delta x \rangle = \omega^T P^{-1} \Delta x = 0 \text{ iff } \Delta x \in column(PH^T)}$$

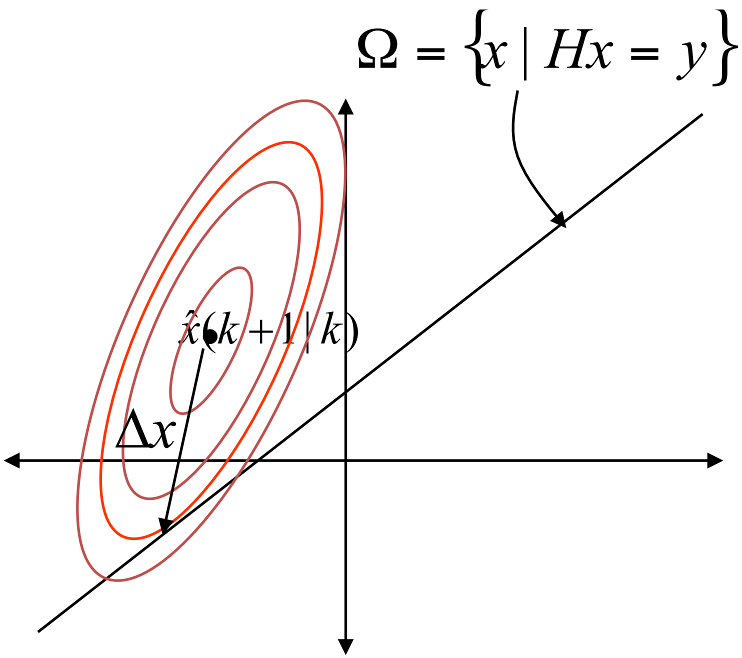
Recall that for all $\omega \in Null(H), \omega \perp b$ if $b \in column(H^T)$

Finding the correction (for real this time!)

Assuming that Δx is linear in $\nu = y - H\hat{x}(k+1 \mid k)$

$$\Delta x = PH^T K \nu$$

The condition $y = H(\hat{x}(k+1 \mid k) + \Delta x) \;\Rightarrow\; H\Delta x = y - H\hat{x}(k+1 \mid k) = \nu$



Substitution yields:

$$H\Delta x = \nu = HPH^T K \nu$$

$$\Rightarrow K = \left(HPH^T\right)^{-1}$$

$$\boxed{\therefore \quad \Delta x = PH^T\left(HPH^T\right)^{-1}\nu}$$

Step 2: Computing the correction

From step 1 we get $\hat{x}(k+1 \mid k)$ and $P(k+1 \mid k)$.

Now we use these to compute Δx :

$$\boxed{\Delta x = P(k+1 \mid k) H^T \left(H P(k+1 \mid k) H^T \right)^{-1} \left(y(k+1) - H \hat{x}(k+1 \mid k) \right)}$$

For ease of notation, define W so that

$$\Delta x = W \nu$$

Step 3: Update

$$\boxed{\hat{x}(k+1 \mid k+1) = \hat{x}(k+1 \mid k) + W\nu}$$

$$P_{k+1} = E\Big((x_{k+1} - \hat{x}_{k+1})(x_{k+1} - \hat{x}_{k+1})^T\Big)$$

$$= E\Big((x_{k+1} - \hat{x}_{k+1}^{-} - W\nu)(x_{k+1} - \hat{x}_{k+1}^{-} - W\nu)^T\Big)$$

(just take my word for it…)

$$\boxed{P(k+1 \mid k+1) = P(k+1 \mid k) - WHP(k+1 \mid k)H^T W^T}$$

Just take my word for it…

$$
\begin{aligned}
P_{k+1} &= E\left((x_{k+1}-\hat{x}_{k+1})(x_{k+1}-\hat{x}_{k+1})^T\right) \\
&= E\left((x_{k+1}-\hat{x}_{k+1}^- - W\nu)(x_{k+1}-\hat{x}_{k+1}^- - W\nu)^T\right) \\
&= E\left(\left((x_{k+1}-\hat{x}_{k+1}^-)-W\nu\right)\left((x_{k+1}-\hat{x}_{k+1}^-)-W\nu\right)^T\right) \\
&= E\left((x_{k+1}-\hat{x}_{k+1}^-)(x_{k+1}-\hat{x}_{k+1}^-)^T - 2W\nu(x_{k+1}-\hat{x}_{k+1}^-)^T + W\nu\left(W\nu\right)^T\right) \\
&= P_{k+1}^- + E\left(-2WH(x_{k+1}-\hat{x}_{k+1}^-)(x_{k+1}-\hat{x}_{k+1}^-)^T + WH(x_{k+1}-\hat{x}_{k+1}^-)(x_{k+1}-\hat{x}_{k+1}^-)^T H^T W^T\right) \\
&= P_{k+1}^- - 2WHP_{k+1}^- + WHP_{k+1}^- H^T W^T \\
&= P_{k+1}^- - 2P_{k+1}^- H^T\left(HP_{k+1}^- H^T\right)^{-1} HP_{k+1}^- + WHP_{k+1}^- H^T W^T \\
&= P_{k+1}^- - 2P_{k+1}^- H^T\left(HP_{k+1}^- H^T\right)^{-1}\left(HP_{k+1}^- H^T\right)\left(HP_{k+1}^- H^T\right)^{-1} HP_{k+1}^- + WHP_{k+1}^- H^T W^T \\
&= P_{k+1}^- - 2WHP_{k+1}^- H^T W^T + WHP_{k+1}^- H^T W^T
\end{aligned}
$$

Better State Observer Summary

System:

$$x(k+1) = Fx(k) + Gu(k) + v(k)$$
$$y(k) = Hx(k)$$

Observer

1. Predict

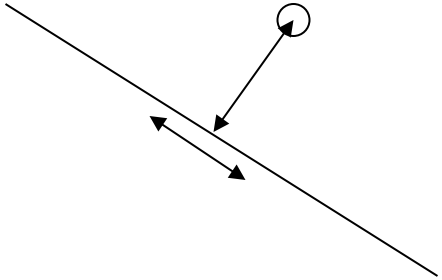
$$\boxed{\begin{aligned}\hat{x}(k+1|k) &= F\hat{x}(k|k) + Gu(k)\\ P(k+1|k) &= FP(k|k)F^T + Q\end{aligned}}$$

2. Correction

$$\boxed{\begin{aligned}W &= P(k+1|k)H\left(HP(k+1|k)H^T\right)^{-1}\\ \Delta x &= W\left(y(k+1) - H\hat{x}(k+1|k)\right)\end{aligned}}$$

3. Update

$$\boxed{\begin{aligned}\hat{x}(k+1|k+1) &= \hat{x}(k+1|k) + W\nu\\ P(k+1|k+1) &= P(k+1|k) - WHP(k+1|k)H^T W^T\end{aligned}}$$



$$\boxed{\begin{aligned} W &= P(k+1\,|\,k)H\left(HP(k+1\,|\,k)H^{T}\right)^{-1} \\ \Delta x &= W\left(y(k+1) - H\hat{x}(k+1\,|\,k)\right) \end{aligned}}$$

•Note: there is a problem with the previous slide, namely the covariance matrix of the estimate P will be singular. This makes sense because with perfect sensor measurements the uncertainty in some directions will be zero. There is no uncertainty in the directions perpendicular to Ω

P lives in the state space and directions associated with sensor noise are zero. In the step when you do the update, if you have a zero noise measurement, you end up squeezing P down.

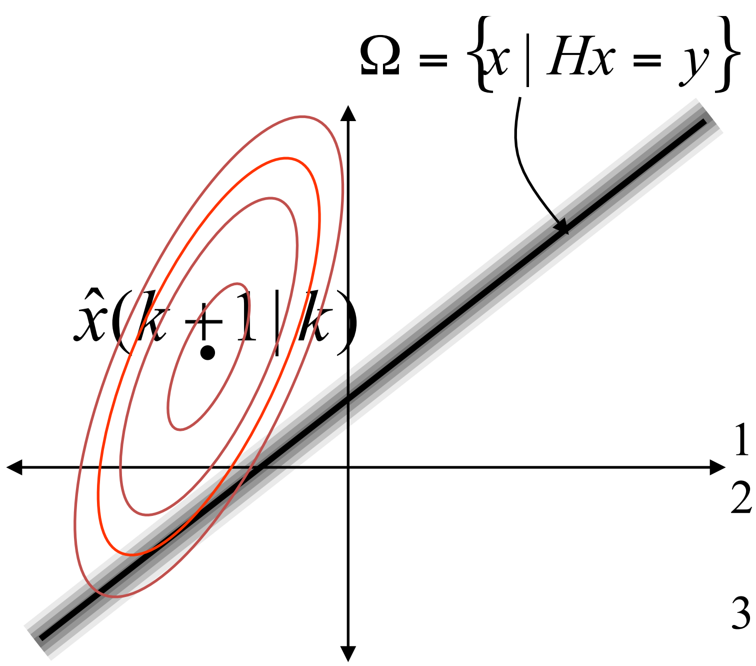
In most cases, when you do the next prediction step, the process covariance matrix Q gets added to the P(k|k), the result will be nonsingular, and everything is ok again.

There's actually not anything wrong with this, except that you can't really call the result a "covariance" matrix because "sometimes" it is not a covariance matrix

Plus, lets not be ridiculous, all sensors have noise.

Finding the correction (with output noise)

$$y = Hx + w$$



The previous results require that you know which hyperplane to aim for. Because there is now sensor noise, we don't know where to aim, so we can't directly use our method.

If we can determine which hyperplane aim for, then the previous result would apply.

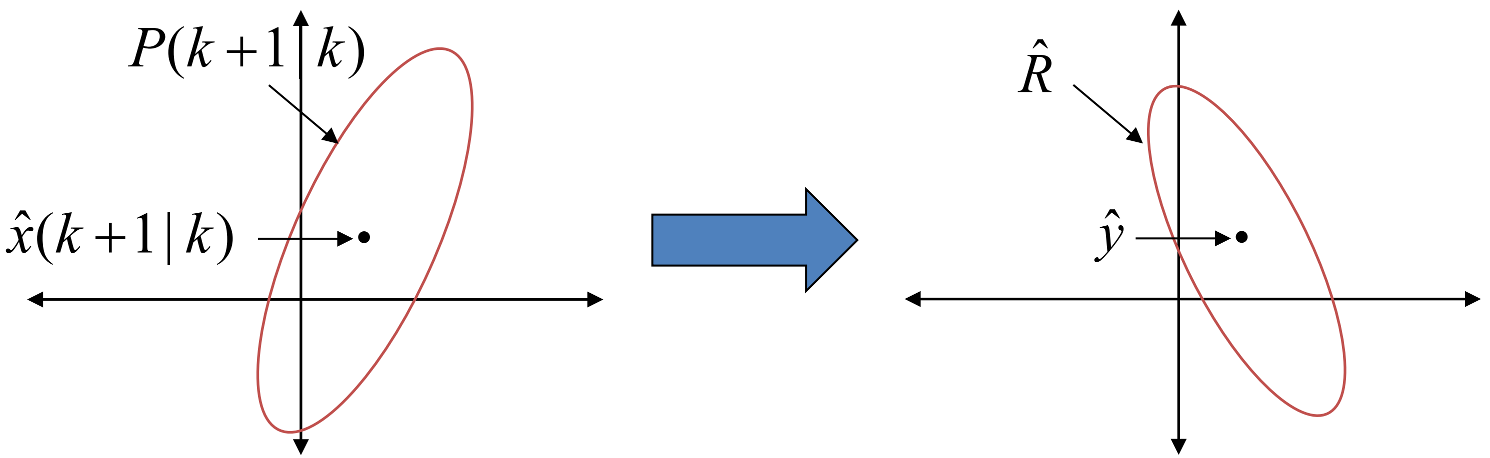
We find the hyperplane in question as follows:

1. project estimate into output space
2. find most likely point in output space based on measurement and projected prediction
3. the desired hyperplane is the preimage of this point

Projecting the prediction
(putting current state estimates into sensor space)

$$\hat{x}(k+1|k) \rightarrow \hat{y} = H\hat{x}(k+1|k)$$

$$P(k+1|k) \rightarrow \hat{R} = HP(k+1|k)H^T$$



state space (n-dimensional) output space (p-dimensional)

Finding most likely output

Y hat has nothing to do with y, hence independent, so multiply them because we want them both to be true at the same time

The objective is to find the most likely output that results from the independent Gaussian distributions

(y, R) measurement and associate covariance

$(\hat{y}, \hat{R})$ projected prediction (what you think your measurements should be and how confident you are)

Fact (Kalman Gains): The product of two Gaussian distributions given by mean/covariance pairs (x_1, C_1) and (x_2, C_2) is proportional to a third Gaussian with mean

$$x_3 = x_1 + K(x_2 - x_1)$$

and covariance

$$C_3 = C_1 - KC_1$$

where

$$K = C_1\left(C_1 + C_2\right)^{-1}$$

Strange, but true, this is symmetric

Most likely output (cont.)

Using the Kalman gains, the most likely output is

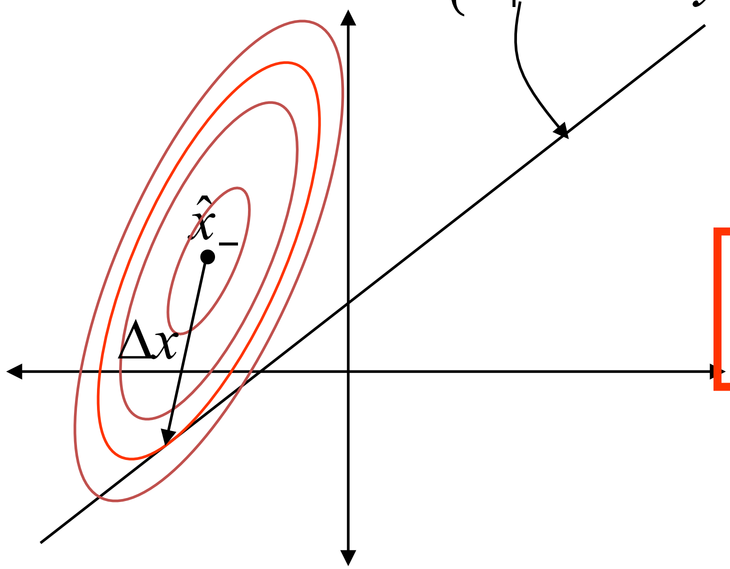
$$y^* = \hat{y} + \left(\hat{R}\left(\hat{R} + R\right)^{-1} \right)(\hat{y} - y)$$

$$= H\hat{x}(k+1|k) + \left(HPH^T\left(HPH^T + R\right)^{-1}\right)\left(H\hat{x}(k+1|k) - y\right)$$

Finding the Correction

Now we can compute the correction as we did in the noiseless case, this time using y* instead of y. In other words, y* tells us which hyperplane to aim for.

$$\Omega = \left\{ x \mid Hx = y^* \right\}$$



The result is:

$$\Delta x = PH^T \left(HPH^T \right)^{-1} \left(y^* - H\hat{x}(k+1 \mid k) \right)$$

Not going all the way to y, but splitting the difference between how confident you are with your Sensor and process noise

Finding the Correction (cont.)

$$\Delta x = PH^T \left(HPH^T \right)^{-1} \left(\mathrm{y}^* - H\hat{x}(k+1 \mid k) \right)$$

$$= PH^T \left(HPH^T \right)^{-1} \left(H\hat{x} + HPH^T \left(HPH^T + R \right)^{-1} \left(y - H\hat{x}(k+1 \mid k) \right) - H\hat{x}(k+1 \mid k) \right)$$

$$= PH^T \left(HPH^T + R \right)^{-1} \left(y - H\hat{x}(k+1 \mid k) \right)$$

For convenience, we define

$$W = PH^T \left(HPH^T + R \right)^{-1}$$

So that

$$\boxed{\Delta x = W \left(y - H\hat{x}(k+1 \mid k) \right)}$$

Correcting the Covariance Estimate

The covariance error estimate correction is computed from the definition of the covariance matrix, in much the same way that we computed the correction for the "better observer". The answer turns out to be:

$$P(k+1 \mid k+1) = P(k+1 \mid k) - W\left(HP(k+1 \mid k)H^T\right)W^T$$

LTI Kalman Filter Summary

System:

$$x(k+1) = Fx(k) + Gu(k) + v(k)$$
$$y(k) = Hx(k) + w(k)$$

Kalman Filter

1. Predict

$$\boxed{\begin{aligned} \hat{x}(k+1|k) &= F\hat{x}(k|k) + Gu(k) \\ P(k+1|k) &= FP(k|k)F^T + Q \end{aligned}}$$

2. Correction

$$\boxed{\begin{aligned} S &= HP(k+1|k)H^T + \mathrm{R} \\ W &= P(k+1|k)H^T S^{-1} \\ \Delta x &= W\left(y(k+1) - H\hat{x}(k+1|k)\right) \end{aligned}}$$

3. Update

$$\boxed{\begin{aligned} \hat{x}(k+1|k+1) &= \hat{x}(k+1|k) + W\nu \\ P(k+1|k+1) &= P(k+1|k) - WSW^T \end{aligned}}$$