Autonomous and Mobile Robotics

Prof. Giuseppe Oriolo

Localization 2
Kalman Filter

Dipartimento di Ingegneria Informatica
Automatica e Gestionale Antonio Ruberti

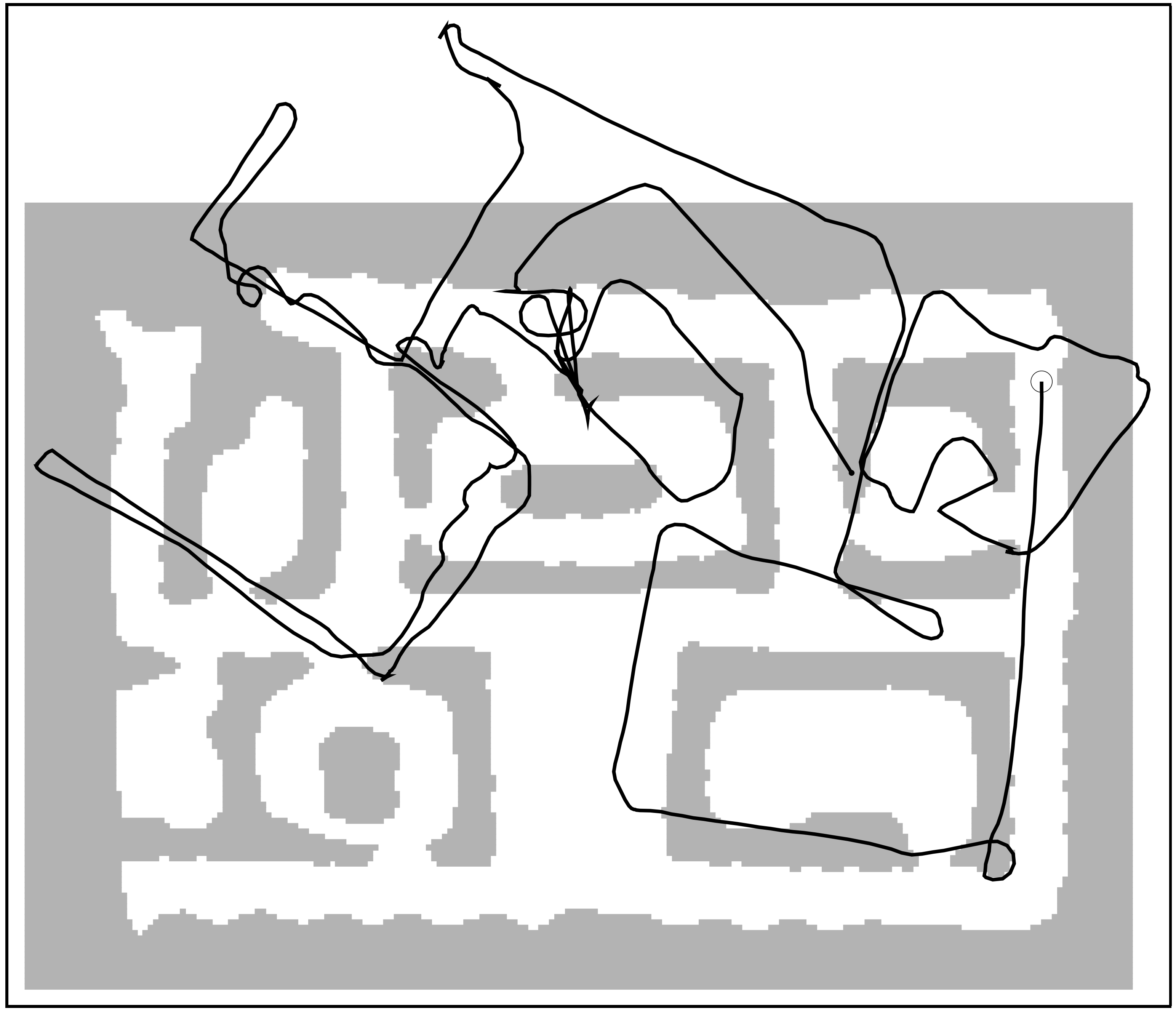
Sapienza
Università di Roma

- recall: estimating the robot configuration by iterative integration of the kinematic model (dead reckoning) is subject to an error that diverges over time

- effective localization methods use proprioceptive as well as exteroceptive sensors: if an environment map is known, compare the actual sensor readings with those predicted using the current estimate

- probabilistic localization: instead of maintaining a single hypothesis on the configuration, maintain a probability distribution over the space of all possible hypotheses

- one possible approach: use a Kalman Filter



a typical dead reckoning result

basic concepts

- given a vector random variable $\boldsymbol{X}$ with probability density function $f_{\boldsymbol{X}}(\boldsymbol{x})$, its expected (or mean) value is

$$E(\boldsymbol{X}) = \bar{\boldsymbol{X}} = \int_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{x} \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$$

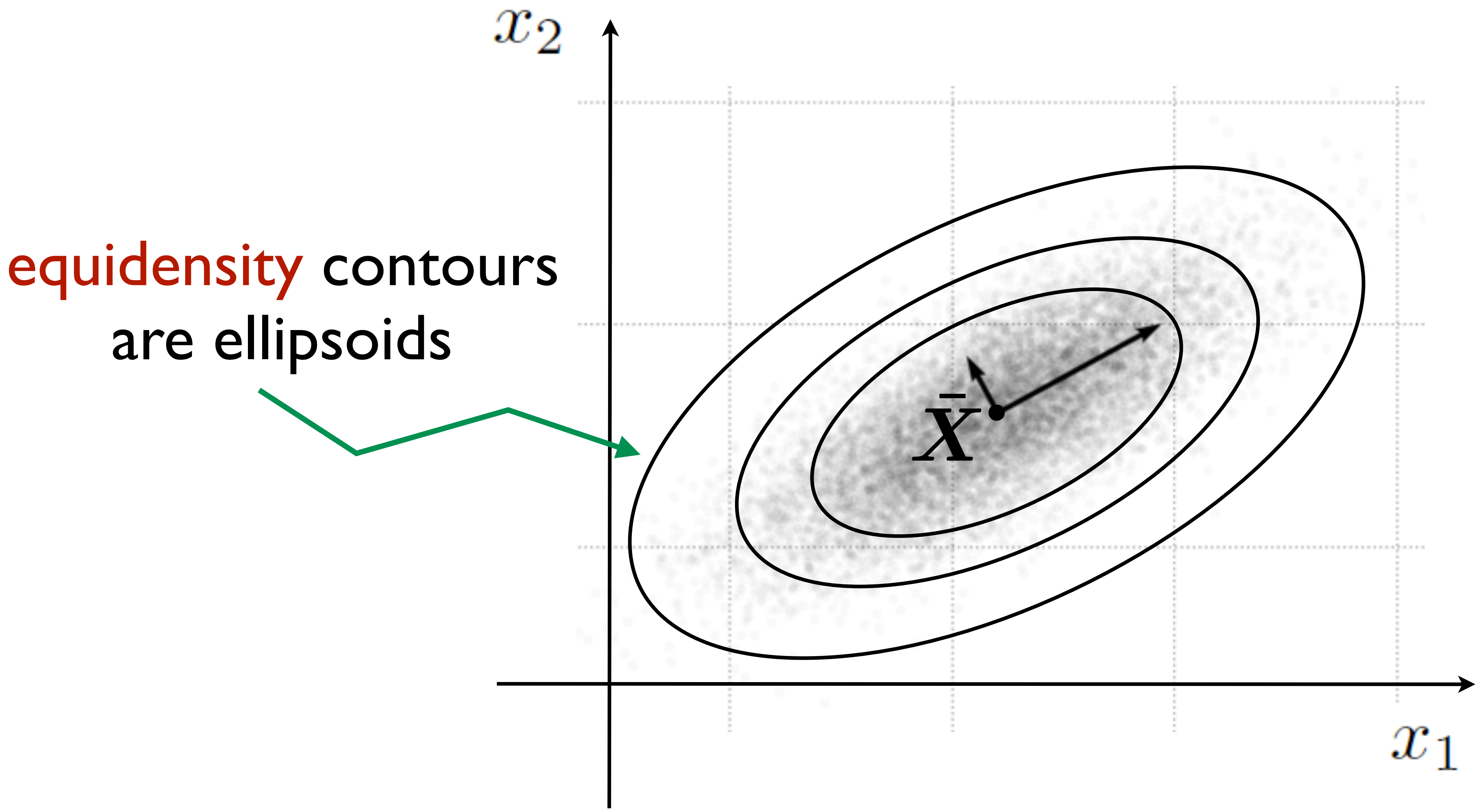
- its covariance matrix is

$$\boldsymbol{P}_{\boldsymbol{X}} = E\left((\boldsymbol{X} - \bar{\boldsymbol{X}})(\boldsymbol{X} - \bar{\boldsymbol{X}})^T\right)$$

- $\boldsymbol{X}$ has a multivariate gaussian distribution if

$$\boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{P}_{\boldsymbol{X}}|}} e^{-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{X}})^T \boldsymbol{P}_{\boldsymbol{X}}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{X}})}$$

- geometric interpretation



- the principal axes are directed as the eigenvectors of P_X
- their squared relative lengths are given by the corresponding eigenvalues

Kalman Filter ...without noise

- consider a linear discrete-time system without noise

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$
$$\boldsymbol{y}_k = \boldsymbol{C}_k \boldsymbol{x}_k$$

- build a recursive observer that computes an estimate $\hat{\boldsymbol{x}}_{k+1}$ of $\boldsymbol{x}_{k+1}$ from $\boldsymbol{u}_k$, $\boldsymbol{y}_{k+1}$ and previous estimate $\hat{\boldsymbol{x}}_k$

- two steps:

  1. prediction: generate an intermediate estimate $\hat{\boldsymbol{x}}_{k+1|k}$ by propagating $\hat{\boldsymbol{x}}_k$ using the process dynamics
  2. correction (update): correct the prediction on the basis of the difference between the measured and the predicted output

- prediction

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$

assuming that we know $\boldsymbol{A}_k$, $\boldsymbol{B}_k$ and $\boldsymbol{u}_k$

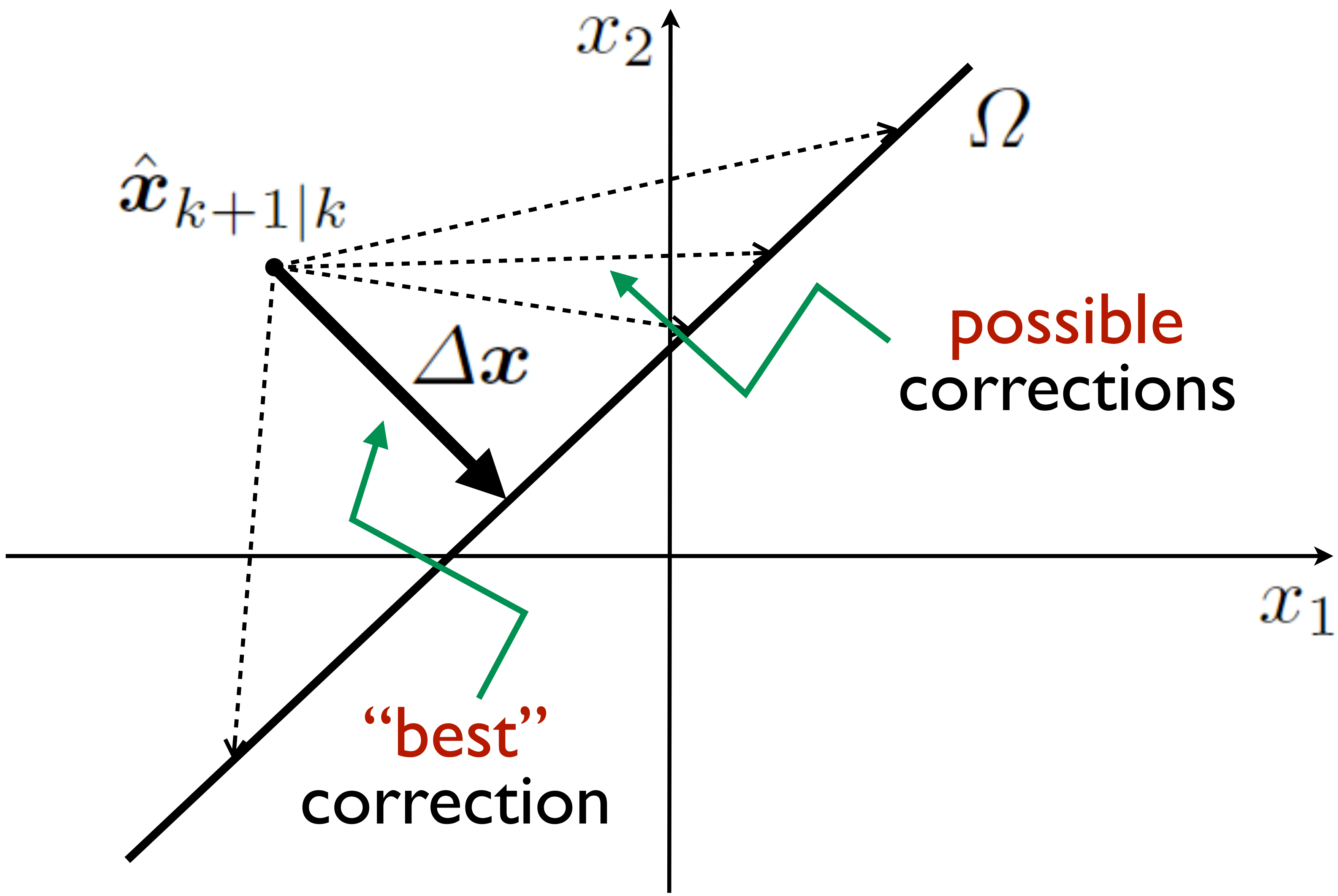
- correction: to be consistent with the measured value of the output, $\boldsymbol{x}_{k+1}$ must belong to the hyperplane

$$\Omega = \{\boldsymbol{x} \, : \, \boldsymbol{C}_{k+1}\boldsymbol{x} = \boldsymbol{y}_{k+1}\}$$

hence the correction $\Delta\boldsymbol{x}$ must satisfy

$$\boldsymbol{C}_{k+1}(\hat{\boldsymbol{x}}_{k+1|k} + \Delta\boldsymbol{x}) = \boldsymbol{y}_{k+1}$$

- geometric interpretation



intuitively, the "best" correction Δx is the closest to the prediction, which we believe is accurate

- Δx is then the solution of an optimization problem

$$\min \|\Delta \boldsymbol{x}\|$$
$$\text{s.t. } \boldsymbol{C}_{k+1}\Delta \boldsymbol{x} = \boldsymbol{y}_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}$$

- it is well known that

$$\Delta \boldsymbol{x} = \boldsymbol{C}_{k+1}^{\dagger}(\boldsymbol{y}_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}) = \boldsymbol{C}_{k+1}^{\dagger}\boldsymbol{\nu}_{k+1}$$

where

$\boldsymbol{C}_{k+1}^{\dagger} = \boldsymbol{C}_{k+1}^{T}(\boldsymbol{C}_{k+1}\boldsymbol{C}_{k+1}^{T})^{-1}$ pseudoinverse of $\boldsymbol{C}_{k+1}$

$\boldsymbol{\nu}_{k+1} = \boldsymbol{y}_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}$ innovation

note that we have assumed C_{k+1} to be full row rank

- wrapping up, the resulting two-step observer is

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{C}_{k+1}^{\dagger} \boldsymbol{\nu}_{k+1}$$

- in general, the estimate $\hat{\boldsymbol{x}}_{k+1}$ will not converge to the true value $\boldsymbol{x}_{k+1}$ because the correction is naive: estimation errors directed as Ω are not corrected

- we need to modify the above structure to take into account the presence of noise; in doing so, we will fix the above problem

Kalman Filter ...with process noise only

- now include process noise

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{B}_k \boldsymbol{u}_k + \boldsymbol{v}_k$$
$$\boldsymbol{y}_k = \boldsymbol{C}_k \boldsymbol{x}_k$$

  where $\boldsymbol{v}_k$ is a white gaussian noise with zero mean and covariance matrix $\boldsymbol{V}_k$

- since this is now a random process, we estimate both the state $\boldsymbol{x}_{k+1}$ and the associated covariance $\boldsymbol{P}_{k+1}$

- we keep the prediction/correction structure

- state prediction: as before

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$

because $\boldsymbol{v}_k$ has zero mean

- covariance prediction: by definition

$$\begin{aligned}
\boldsymbol{P}_{k+1|k} &= E\left((\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1|k})(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1|k})^T\right) \\
&= E\left((\boldsymbol{A}_k(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k) + \boldsymbol{v}_k)(\boldsymbol{A}_k(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k) + \boldsymbol{v}_k)^T\right) \\
&= E\left(\boldsymbol{A}_k(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T \boldsymbol{A}_k^T\right) + \\
&\quad E\left(\boldsymbol{A}_k(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)\boldsymbol{v}_k^T + \boldsymbol{v}_k(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T \boldsymbol{A}_k^T)\right) + E\left(\boldsymbol{v}_k \boldsymbol{v}_k^T\right)
\end{aligned}$$

- now use the linearity of E plus the independence of $\boldsymbol{v}_k$ on $\hat{\boldsymbol{x}}_k$ and $\boldsymbol{x}_k \Rightarrow$ the second term in the rhs is zero

- finally the covariance prediction is

$$\begin{aligned}\boldsymbol{P}_{k+1|k} &= \boldsymbol{A}_k E\left((\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T\right) \boldsymbol{A}_k^T + E\left(\boldsymbol{v}_k \boldsymbol{v}_k^T\right) \\ &= \boldsymbol{A}_k \boldsymbol{P}_k \boldsymbol{A}_k^T + \boldsymbol{V}_k\end{aligned}$$

- state correction: we should choose Δx so as to get the most likely x in Ω, i.e., the x that maximizes the gaussian distribution defined by $\hat{\boldsymbol{x}}_{k+1|k}$ and $\boldsymbol{P}_{k+1|k}$

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{P}_{k+1|k}|}} e^{-\frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{x}}_{k+1|k})^T \boldsymbol{P}_{k+1|k}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{x}}_{k+1|k})}$$

$p(\boldsymbol{x})$ is maximized when the exponent is minimized

- define the (squared) Mahalanobis distance

$$\Delta\boldsymbol{x}^T \boldsymbol{P}_{k+1|k}^{-1} \Delta\boldsymbol{x} = \|\Delta\boldsymbol{x}\|_M^2$$

- $\Delta\boldsymbol{x}$ is the solution of a new optimization problem

$$\begin{aligned} &\min \|\Delta\boldsymbol{x}\|_M \\ &\text{s.t. } \boldsymbol{C}_{k+1}\Delta\boldsymbol{x} = \boldsymbol{y}_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{x}}_{k+1|k} \end{aligned}$$

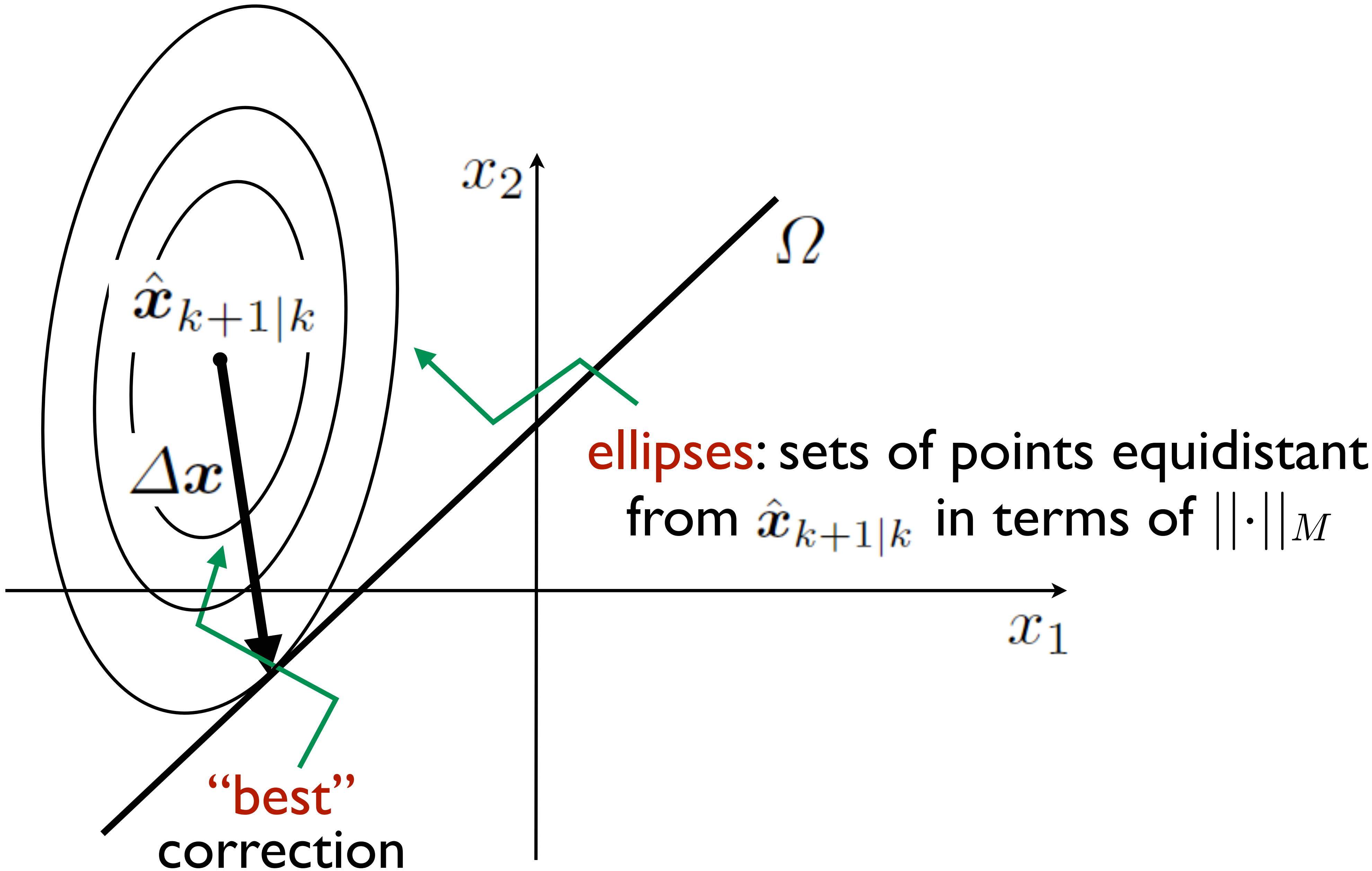
- it is well known that

$$\Delta\boldsymbol{x} = \boldsymbol{C}_{k+1,M}^{\dagger}(\boldsymbol{y}_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}) = \boldsymbol{C}_{k+1,M}^{\dagger}\boldsymbol{\nu}_{k+1}$$

where $\boldsymbol{C}_{k+1,M}^{\dagger}$ is the weighted pseudoinverse of $\boldsymbol{C}_{k+1}$

$$\boldsymbol{C}_{k+1,M}^{\dagger} = \boldsymbol{P}_{k+1|k}\boldsymbol{C}_{k+1}^T(\boldsymbol{C}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{C}_{k+1}^T)^{-1}$$

- geometric interpretation



the "best" correction is the closest to the prediction according to the current covariance estimate

- covariance correction: using the covariance matrix definition and the state correction one obtains

$$\boldsymbol{P}_{k+1} = \boldsymbol{P}_{k+1|k} - \boldsymbol{C}^{\dagger}_{k+1,M} \boldsymbol{C}_{k+1} \boldsymbol{P}_{k+1|k}$$

- wrapping up, the resulting two-step filter is

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$
$$\boldsymbol{P}_{k+1|k} = \boldsymbol{A}_k \boldsymbol{P}_k \boldsymbol{A}_k^T + \boldsymbol{V}_k$$
$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{C}^{\dagger}_{k+1,M} \boldsymbol{\nu}_{k+1}$$
$$\boldsymbol{P}_{k+1} = \boldsymbol{P}_{k+1|k} - \boldsymbol{C}^{\dagger}_{k+1,M} \boldsymbol{C}_{k+1} \boldsymbol{P}_{k+1|k}$$

- problem: no measurement noise $\Rightarrow$ the covariance estimate will become singular (no uncertainty in the normal direction to the measurement hyperplane)

Kalman Filter ...full

- finally include also measurement (sensor) noise

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{B}_k \boldsymbol{u}_k + \boldsymbol{v}_k$$
$$\boldsymbol{y}_k = \boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{w}_k$$

where $\boldsymbol{v}_k$, $\boldsymbol{w}_k$ are white gaussian noises with zero mean and covariance matrices $\boldsymbol{V}_k$, $\boldsymbol{W}_k$

- the dynamic equation is unchanged, therefore the predictions are the same

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$
$$\boldsymbol{P}_{k+1|k} = \boldsymbol{A}_k \boldsymbol{P}_k \boldsymbol{A}_k^T + \boldsymbol{V}_k$$

- state correction: due to the sensor noise, the output value is no more certain; we only know that $\boldsymbol{y}_{k+1}$ is drawn from a gaussian distribution with mean value $\boldsymbol{C}_{k+1}\,\boldsymbol{x}_{k+1}$ and covariance matrix $\boldsymbol{W}_{k+1}$

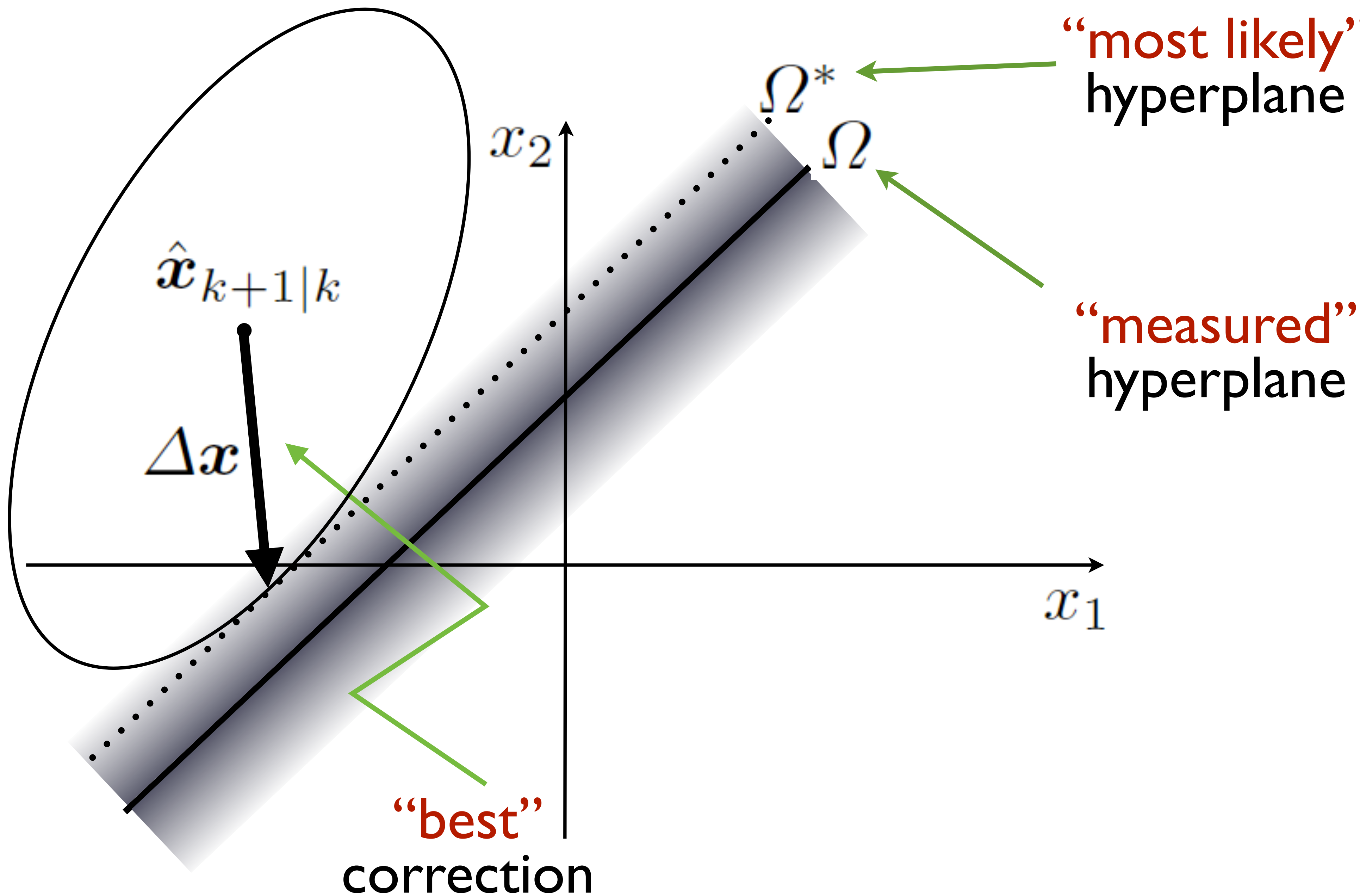
- first we compute the most likely output value $\boldsymbol{y}^*_{k+1}$ given the predictions and the measured output $\boldsymbol{y}_{k+1}$

- then compute the associated most likely hyperplane

$$\Omega^* = \{\boldsymbol{x} \,:\, \boldsymbol{C}_{k+1}\boldsymbol{x} = \boldsymbol{y}^*_{k+1}\}$$

- finally compute the correction $\Delta\boldsymbol{x}$ as before but using Ω^* in place of Ω

- geometric interpretation



the "best" correction is still the closest to $\hat{\boldsymbol{x}}_{k+1|k}$ according to $\boldsymbol{P}_{k+1|k}$, but now it lies on Ω^*

- the resulting Kalman Filter (KF) is

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k$$
$$\boldsymbol{P}_{k+1|k} = \boldsymbol{A}_k \boldsymbol{P}_k \boldsymbol{A}_k^T + \boldsymbol{V}_k$$
$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{R}_{k+1} \boldsymbol{\nu}_{k+1}$$
$$\boldsymbol{P}_{k+1} = \boldsymbol{P}_{k+1|k} - \boldsymbol{R}_{k+1} \boldsymbol{C}_{k+1} \boldsymbol{P}_{k+1|k}$$

  with the Kalman gain matrix

$$\boldsymbol{R}_{k+1} = \boldsymbol{P}_{k+1|k} \boldsymbol{C}_{k+1}^T (\boldsymbol{C}_{k+1} \boldsymbol{P}_{k+1|k} \boldsymbol{C}_{k+1}^T + \boldsymbol{W}_{k+1})^{-1}$$

- matrix $\boldsymbol{R}$ weighs the accuracy of the prediction vs. that of the measurements
  - $\boldsymbol{R}$ "large": measurements are more reliable
  - $\boldsymbol{R}$ "small": prediction is more reliable

- the KF provides an optimal estimate in the sense that $E(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})$ is minimized for each k

- the KF is also correct, i.e., it provides mean value and covariance of the posterior gaussian distribution

- if the noises have non-gaussian distributions, the KF is still the best linear estimator but there might exist more accurate nonlinear filters

- if the process is observable, the estimate produced by the KF converges, in the sense that $E(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})$ is bounded for all k

Extended Kalman Filter

- consider a nonlinear discrete-time system with noise

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_k$$
$$\boldsymbol{y}_k = \boldsymbol{h}_k(\boldsymbol{x}_k) + \boldsymbol{w}_k$$

  where $\boldsymbol{f}_k$ and $\boldsymbol{h}_k$ are continuously differentiable for each k

- one simple way to build a filter is to linearize the system dynamic equations around the current estimate and then apply the KF equations to the resulting linear approximation

- the resulting Extended Kalman Filter (EKF) is

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{f}_k(\hat{\boldsymbol{x}}_k, \boldsymbol{u}_k)$$
$$\boldsymbol{P}_{k+1|k} = \boldsymbol{F}_k \boldsymbol{P}_k \boldsymbol{F}_k^T + \boldsymbol{V}_k$$
$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{R}_{k+1} \boldsymbol{\nu}_{k+1}$$
$$\boldsymbol{P}_{k+1} = \boldsymbol{P}_{k+1|k} - \boldsymbol{R}_{k+1} \boldsymbol{H}_{k+1} \boldsymbol{P}_{k+1|k}$$

with

$$\boldsymbol{F}_k = \left.\frac{\partial \boldsymbol{f}_k}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_k} \qquad \boldsymbol{H}_{k+1} = \left.\frac{\partial \boldsymbol{h}_{k+1}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k+1|k}}$$

and the gain matrix

$$\boldsymbol{R}_{k+1} = \boldsymbol{P}_{k+1|k} \boldsymbol{H}_{k+1}^T (\boldsymbol{H}_{k+1} \boldsymbol{P}_{k+1|k} \boldsymbol{H}_{k+1}^T + \boldsymbol{W}_{k+1})^{-1}$$