

Autonomous and Mobile Robotics

Prof. Giuseppe Oriolo

Localization 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**

basic concepts

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

$$E(\mathbf{X}) = \bar{\mathbf{X}} = \int_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

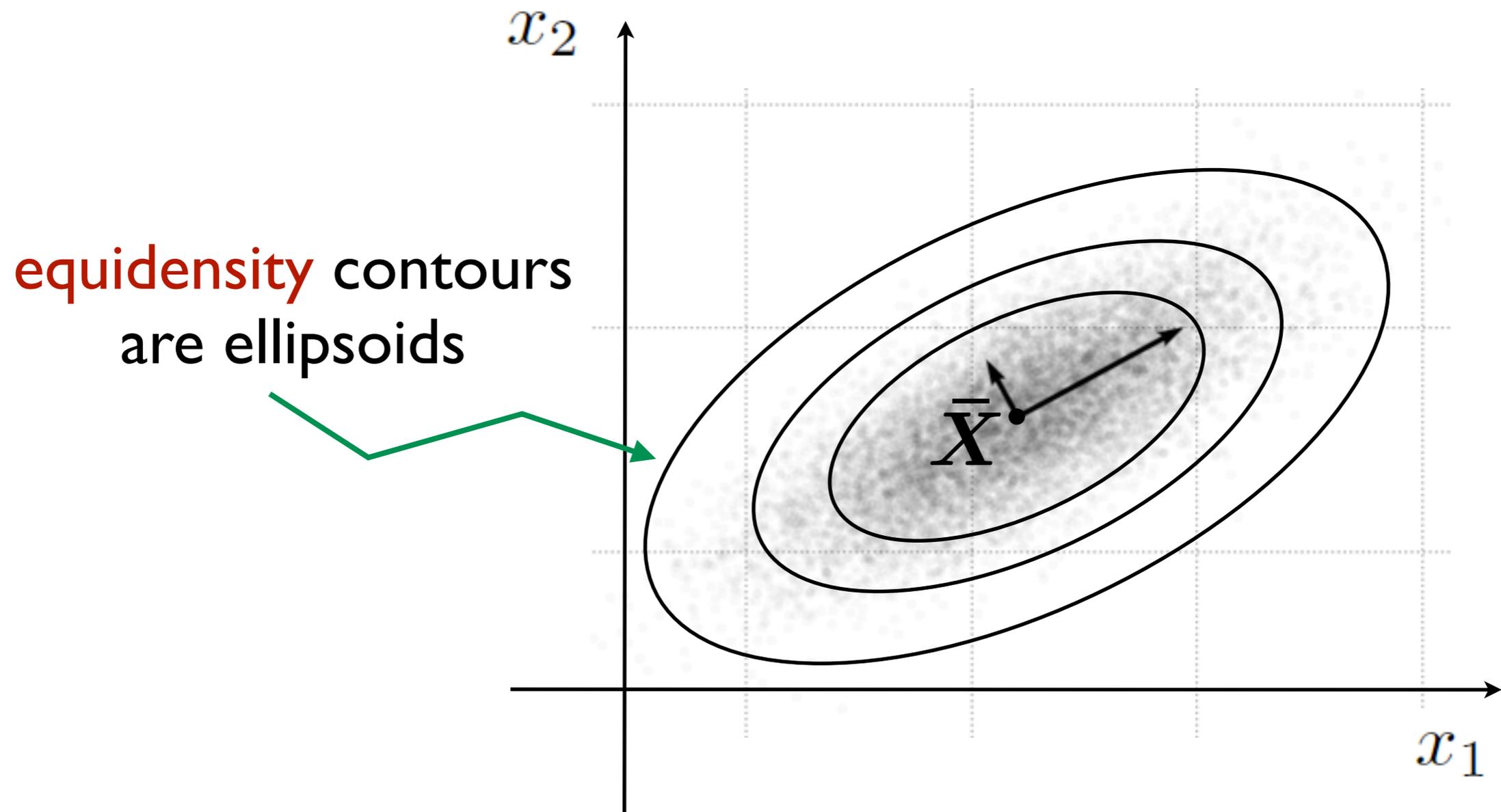
- its **covariance matrix** is

$$P_{\mathbf{X}} = E \left((\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T \right)$$

- \mathbf{X} has a **multivariate gaussian distribution** if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |P_{\mathbf{X}}|}} e^{-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{X}})^T P_{\mathbf{X}}^{-1}(\mathbf{x} - \bar{\mathbf{X}})}$$

- **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**

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k$$

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

- two steps:

1. **prediction**: generate an intermediate estimate $\hat{\mathbf{x}}_{k+1|k}$ by propagating $\hat{\mathbf{x}}_k$ 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

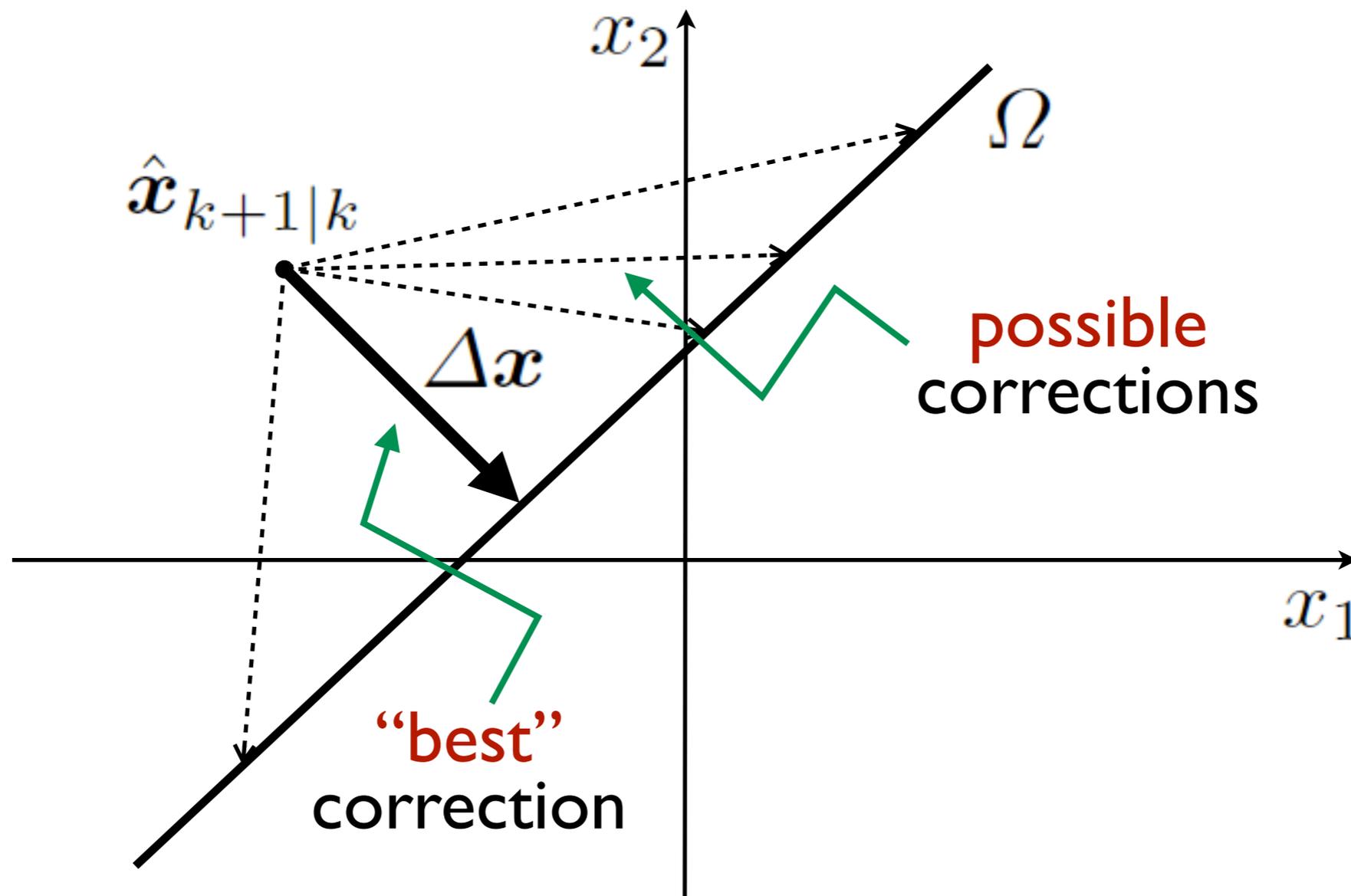
- **correction**: to be consistent with the measured value of the output, \boldsymbol{x}_{k+1} must belong to the **hyperplane**

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

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

- $\Delta \mathbf{x}$ is then the solution of an **optimization** problem

$$\min \|\Delta \mathbf{x}\|$$

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

- it is well known that

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

where

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

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

note that we have assumed \mathbf{C}_{k+1} to be full row rank

- wrapping up, the resulting **two-step observer** is

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

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

- in general, the estimate $\hat{\mathbf{x}}_{k+1}$ will **not** converge to the true value \mathbf{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**

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{v}_k$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k$$

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

- since this is now a **random process**, we estimate both the state \mathbf{x}_{k+1} and the associated covariance \mathbf{P}_{k+1}
- we keep the **prediction/correction** structure

- **state prediction:** as before

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

because \mathbf{v}_k has zero mean

- **covariance prediction:** by definition

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

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

- finally the covariance prediction is

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

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

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

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

- this leads to the (squared) **Mahalanobis distance**, which penalizes deviations along directions where precision is higher

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

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

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

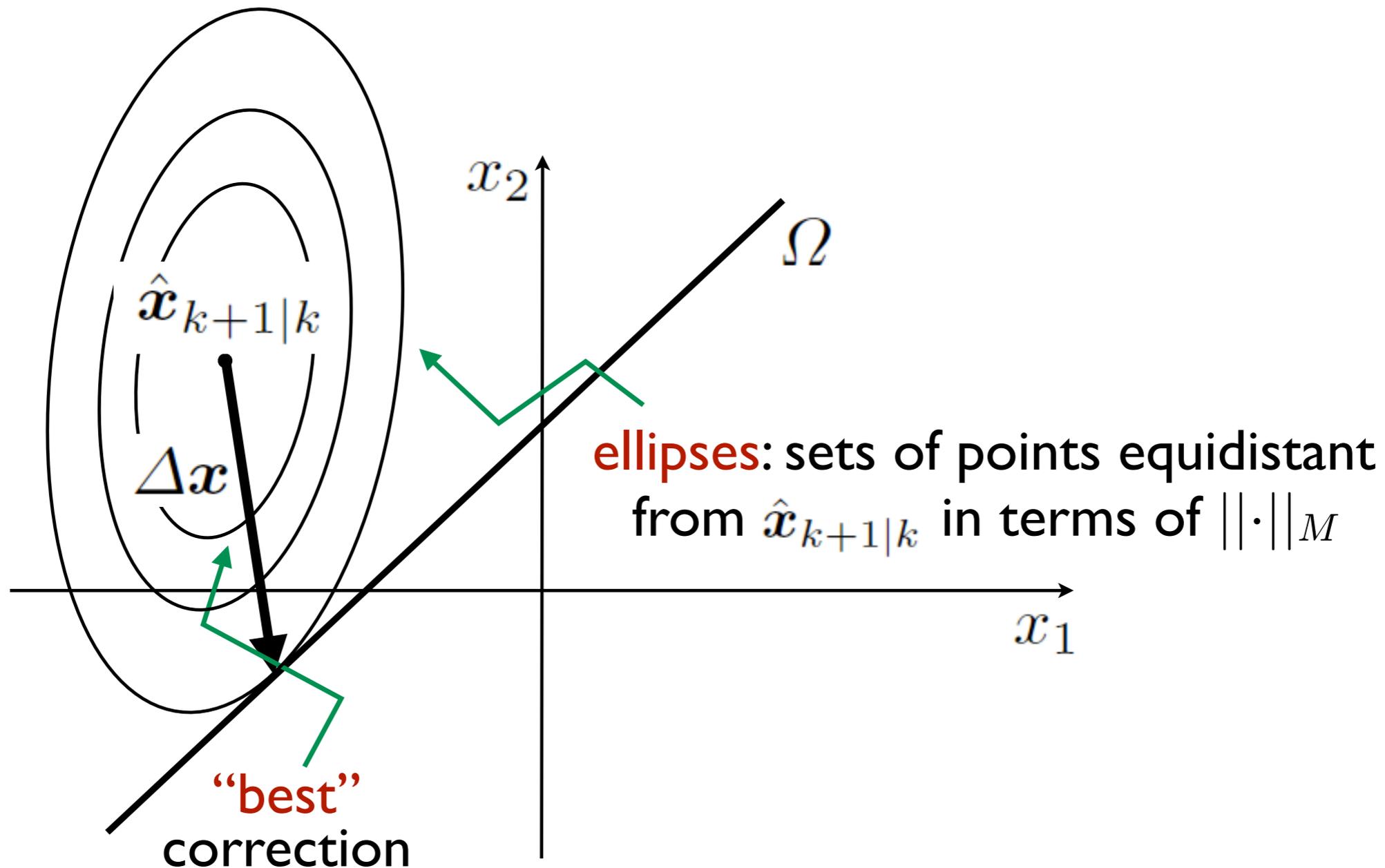
- the optimal solution is

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

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

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

- 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

$$P_{k+1} = P_{k+1|k} - C_{k+1,M}^\dagger C_{k+1} P_{k+1|k}$$

- wrapping up, the resulting **two-step filter** is

$$\hat{\mathbf{x}}_{k+1|k} = A_k \hat{\mathbf{x}}_k + B_k \mathbf{u}_k$$

$$P_{k+1|k} = A_k P_k A_k^T + V_k$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1|k} + C_{k+1,M}^\dagger \boldsymbol{\nu}_{k+1}$$

$$P_{k+1} = P_{k+1|k} - C_{k+1,M}^\dagger C_{k+1} P_{k+1|k}$$

- 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**

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{v}_k$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{w}_k$$

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

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

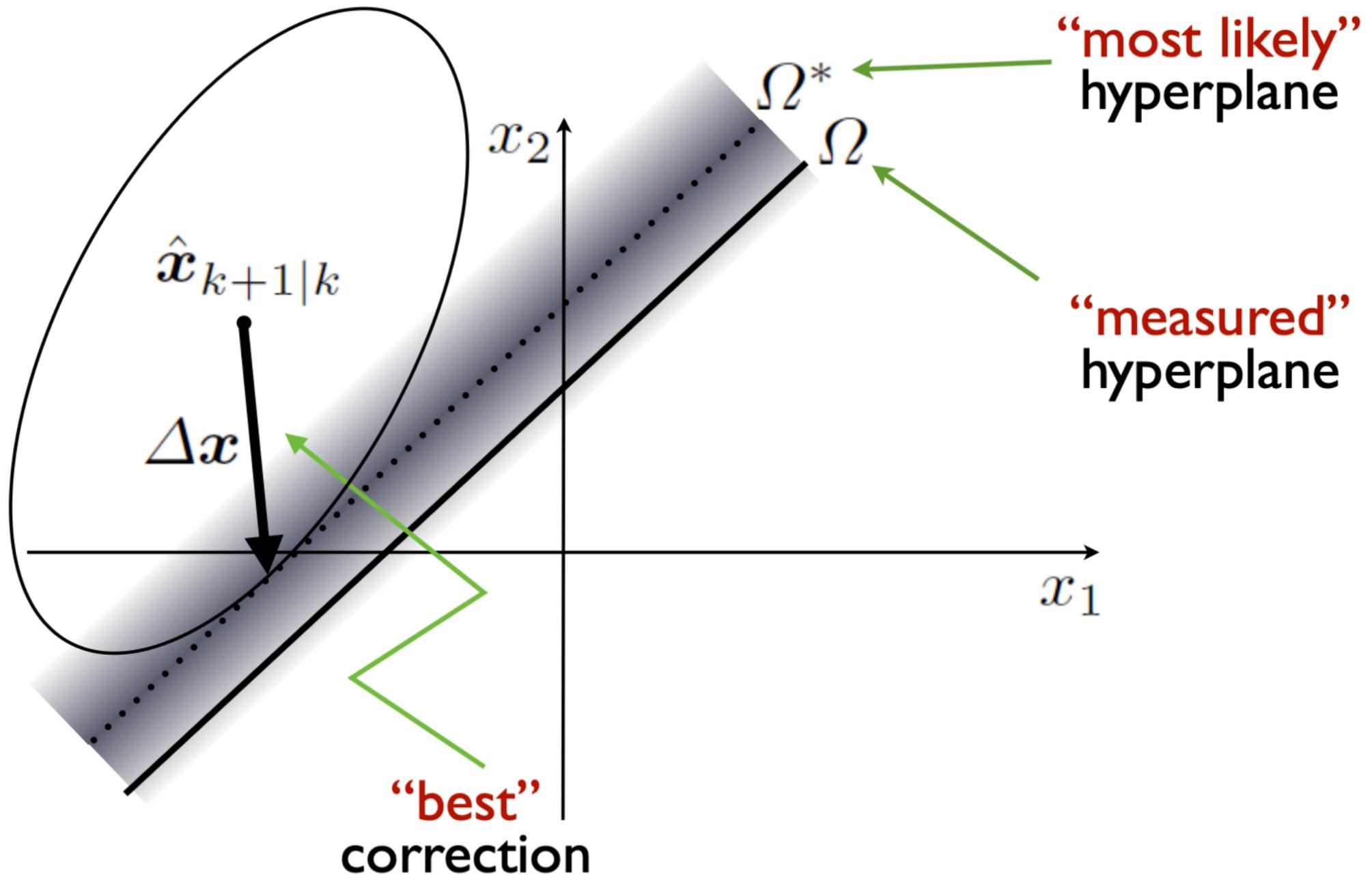
$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{B}_k \mathbf{u}_k$$

$$\mathbf{P}_{k+1|k} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{V}_k$$

- **state correction**: due to the sensor noise, the output value is no more certain; we only know that \mathbf{y}_{k+1} is drawn from a gaussian distribution with mean value $\mathbf{C}_{k+1} \mathbf{x}_{k+1}$ and covariance matrix \mathbf{W}_{k+1}
- first we compute the **most likely** output value \mathbf{y}_{k+1}^* given the predictions and the measured output \mathbf{y}_{k+1}
- then compute the associated **most likely** hyperplane

$$\Omega^* = \{ \mathbf{x} : \mathbf{C}_{k+1} \mathbf{x} = \mathbf{y}_{k+1}^* \}$$
- finally compute the correction $\Delta \mathbf{x}$ as before **but** using Ω^* in place of Ω

- geometric interpretation



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

- the resulting **Kalman Filter (KF)** is

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

$$\mathbf{P}_{k+1|k} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{V}_k$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{R}_{k+1} \boldsymbol{\nu}_{k+1}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1|k} - \mathbf{R}_{k+1} \mathbf{C}_{k+1} \mathbf{P}_{k+1|k}$$

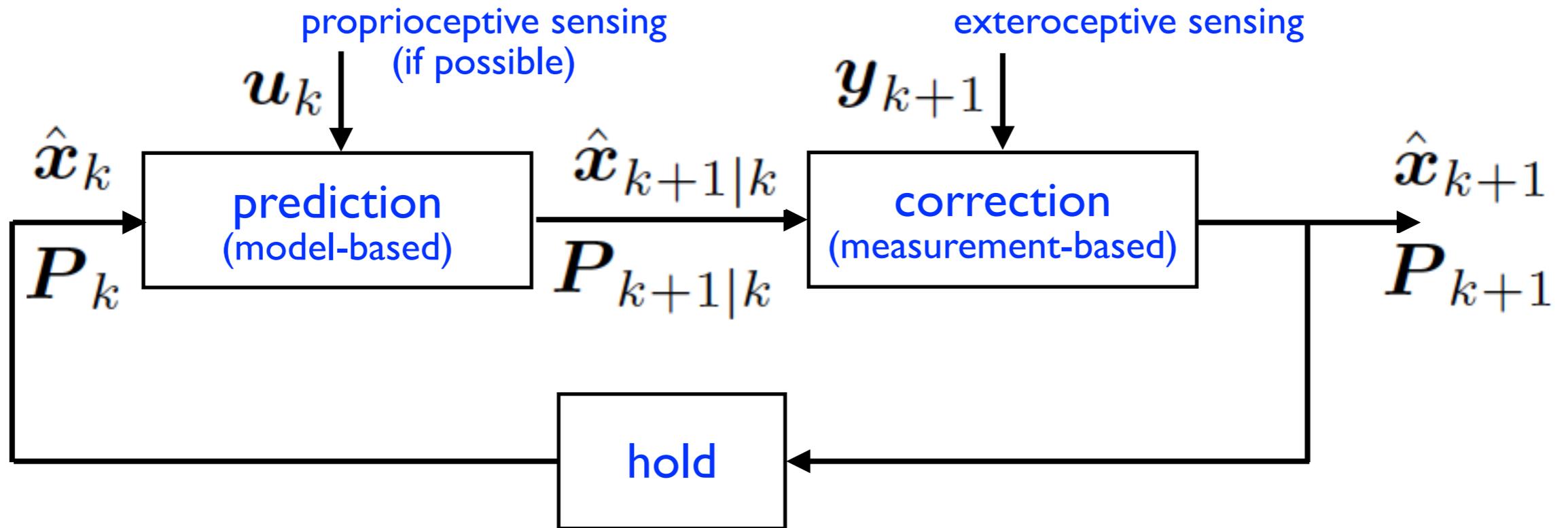
with the **Kalman gain matrix**

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

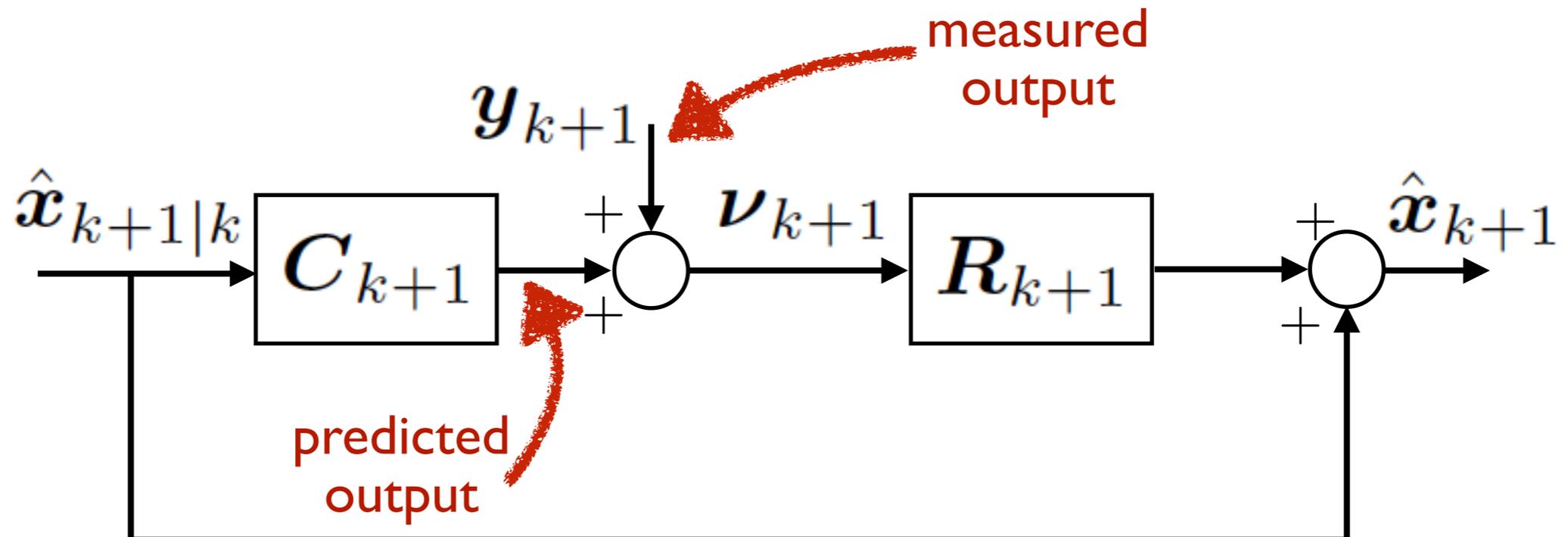
- matrix \mathbf{R} weighs **the accuracy of the prediction vs. that of the measurements**
 - \mathbf{R} “large”: measurements are more reliable
 - \mathbf{R} “small”: prediction is more reliable

- very low computational load, plus a lot of stuff can be precomputed; moreover, no measurements are stored
- the KF provides an **optimal** estimate in the sense that $E(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1})$ 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 $E(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1})$ is bounded for all k

- wrapping up



- correction



Extended Kalman Filter

- consider a **nonlinear** discrete-time system **with noise**

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{w}_k$$

where \mathbf{f}_k and \mathbf{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

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}_k(\hat{\mathbf{x}}_k, \mathbf{u}_k)$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^T + \mathbf{V}_k$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{R}_{k+1} \boldsymbol{\nu}_{k+1}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1|k} - \mathbf{R}_{k+1} \mathbf{H}_{k+1} \mathbf{P}_{k+1|k}$$

with

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

and the gain matrix

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