Data Assimilation

Berent og Feda

February 25, 2022

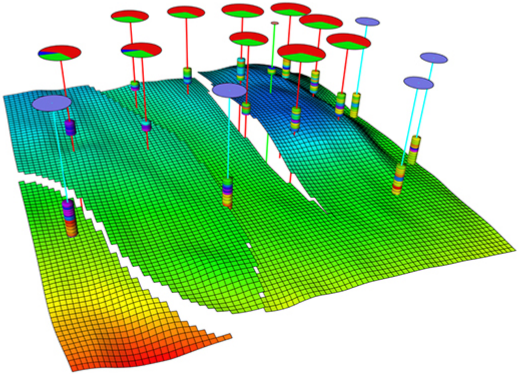
Data assimilation and the Kalman filter

Extended Kalman Filter

Ensemble Kalman Filter

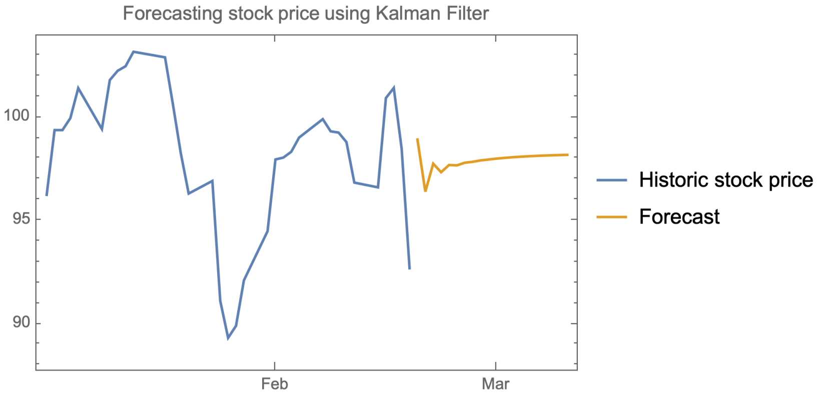
Ensemble Information Filter

Data Assimilation at Equinor



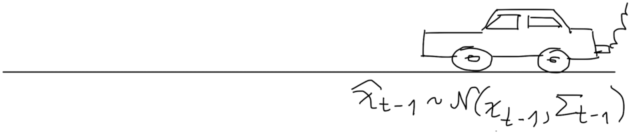
- Improve estimates reservoir properties to improve predictions.

Kalman Filter



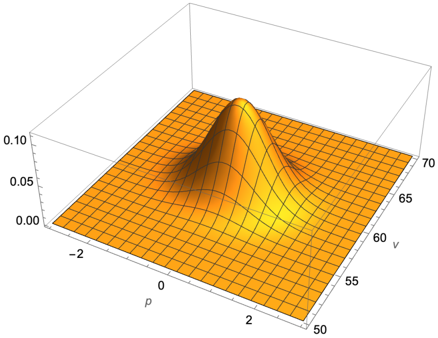
- Blend predictions with measurements.
- Two sensors, even if one is less accurate then the other, is better than one.
- We don't throw away information, no matter how poor it is.
- The behaviour of the physical system we are measuring should influence how we interpret the measurements.

Problem statement Part I



- $\boldsymbol{x}_{t-1} = [p_{t-1} \; v_{t-1}]^T$ is the true state of the system at time $t-1$. We don't know what it is but would like to estimate it.
- We call the estimate $\hat{\boldsymbol{x}}_{t-1}$ and assume normality, i.e., $\hat{\boldsymbol{x}}_{t-1} \sim \mathcal{N}(\boldsymbol{x}_{t-1}, \boldsymbol{\Sigma}_{t-1})$. That is, we assume our estimate / prediction / forecast is centered around the true value $\boldsymbol{x}_{t-1}$ with some uncertainty $\boldsymbol{\Sigma}_{t-1}$.

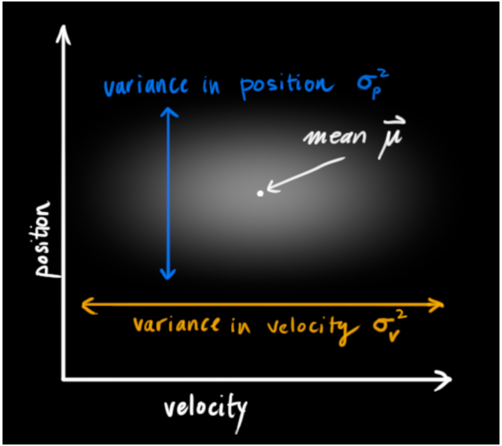
Problem statement Part II



- Our prior belief is that the cars' most likely position and velocity at time $t-1$ are $0[km]$ and $60[km/h]$. We model this belief with a normal distribution with mean $\hat{\boldsymbol{\mu}}_{t-1} = [0\ 60]^T$ and covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{t-1} = \begin{bmatrix} \sigma_p^2 & \rho \\ \rho & \sigma_v^2 \end{bmatrix}$$

Problem statement Part III



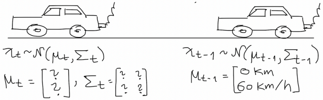
- Here's another picture to help you visualize our prior belief.

Problem statement Part IV

$$x_t \sim \mathcal{N}(\mu_t, \Sigma_t) \qquad x_{t-1} \sim \mathcal{N}(\mu_{t-1}, \Sigma_{t-1})$$

$$\mu_t = \begin{bmatrix} ? \\ ? \end{bmatrix}, \quad \Sigma_t = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \qquad \mu_{t-1} = \begin{bmatrix} 0\ \text{km} \\ 60\ \text{km/h} \end{bmatrix}$$

- **Problem statement**: find the best estimate of the cars' position after one minute.



- A very simple model for predicting state is

$$
\begin{aligned}
p_t &= p_{t-1} + \Delta t v_{t-1} \\
v_t &= v_{t-1}
\end{aligned}
$$

  where Δt is the time elapsed between **time epochs** $t-1$ and t.

- It's useful to re-write this model using matrix notation as

$$
\begin{aligned}
\boldsymbol{x}_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix} &= \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \boldsymbol{w}_t \\
&= \boldsymbol{M}_t \boldsymbol{x}_{t-1} + \boldsymbol{w}_t
\end{aligned}
$$

  where $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$. The evolution matrix $\boldsymbol{M}_t$ determines how the state evolves over time. $\boldsymbol{w}_t$ is the process error that makes this estimate non-perfect.

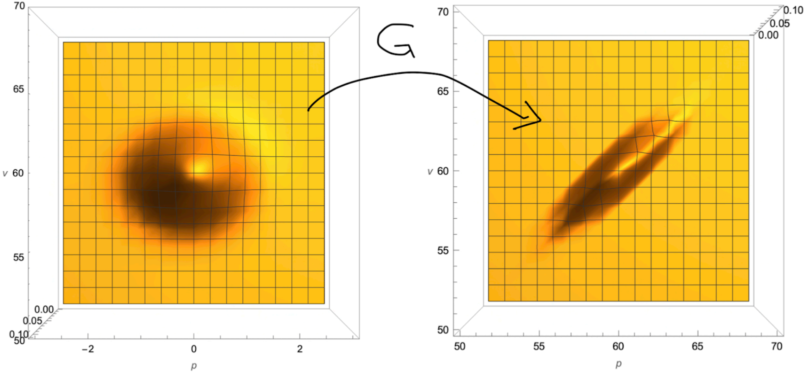
Forecasting Part II

- Recall that estimates $\hat{\boldsymbol{x}}_t$ are Gaussian probability density functions, rather than discrete values. A Gaussian PDF is fully described by its mean and covariance,

$$\begin{aligned}\mathbb{E}[\hat{\boldsymbol{x}}_t] = \boldsymbol{\mu}_{t|t-1} &= \mathbb{E}[\boldsymbol{M}_t\hat{\boldsymbol{x}}_{t-1}] + \mathbb{E}[\boldsymbol{w}_t] \\ &= \boldsymbol{M}_t\boldsymbol{\mu}_{t-1}\end{aligned}$$

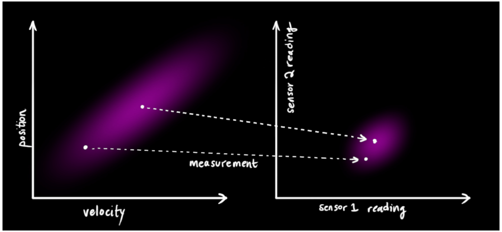
$$\begin{aligned}Cov(\hat{\boldsymbol{x}}_t) = \boldsymbol{\Sigma}_{t|t-1} &= Cov(\boldsymbol{M}_t\hat{\boldsymbol{x}}_{t-1}) + Cov(\boldsymbol{w}_t) \\ &= \boldsymbol{M}_t\boldsymbol{\Sigma}_{t-1}\boldsymbol{M}_t^T\end{aligned}$$

Forecasting Part III - graphical representation



- Notice how the dynamical model induces correlations between position and velocity.

Introducing sensors



- We have two sensors, one measures the position (GPS) while the other measures the velocity (speedometer).
- Measurements of the system can be performed according to the model

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{v}_t, \ \boldsymbol{v}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{y_t})$$

- $\boldsymbol{\Sigma}_{y_t}$ is the sensor noise, since no measurement is perfect.
- $\boldsymbol{H}_t$ is sometimes called the observation matrix and is explained later.

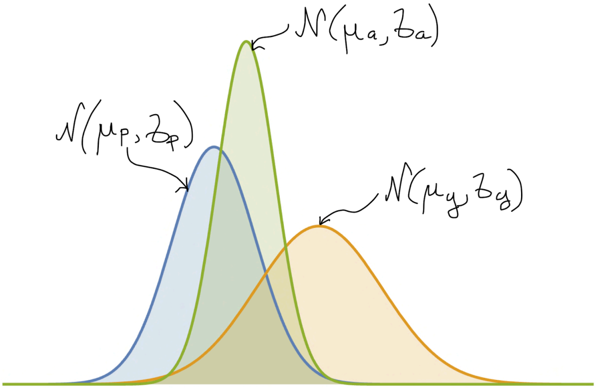
State-space Model (recap)

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{y_t})$$

$$\boldsymbol{x}_t = \boldsymbol{M}_t \boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_t)$$

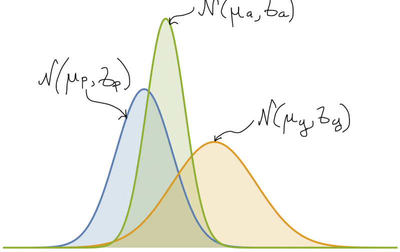
- State-space := Mathematical model of a physical system as a set of input, output and state variables related by first-order differential equations or difference equations.
- Filtering := Obtain **filtering distribution** of the state, that is, $\boldsymbol{x_t}$ given $\boldsymbol{y}_{1:t}$, the data up to time t.
- If the model is Gaussian and linear, we can use the Kalman Filter to obtain the filtering distribution.
- Kalman Filter consists of two steps at every time point; a **forecast step** and an **update step**.

Combining Gaussians (A mathematical de-tour) Part I



- Information is available from two sources: predictions based on the last known position and velocity, and measurements from GPS and speedometer. How to combine predictions with measurements to get the best possible estimate?

Combining Gaussians (A mathematical de-tour) Part II



- The product of two Gaussian distributions is another Gaussian distribution.
- 1D Gaussian is given by $\mathcal{N}(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{x-\mu}{2\sigma^2}}$
- Multiplying two Gaussian PDF's (after some algebra)

$$\mu_a = \mu_p + k(\mu_y - \mu_p)$$
$$\sigma_p^2 = \sigma_p^2 - k\sigma_p^2$$
$$k = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_y^2}$$

Combining Gaussians (A mathematical de-tour) Part III

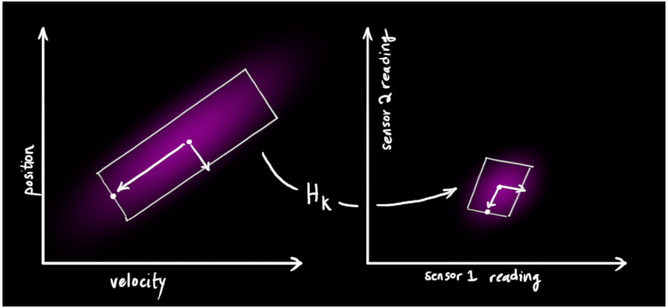
- It's convenient to use matrix notation since we are working with states of multiple dimensions.

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}_p + \boldsymbol{K}(\boldsymbol{\mu}_y - \boldsymbol{\mu}_p)$$
$$\boldsymbol{\Sigma}_a = \boldsymbol{\Sigma}_p - \boldsymbol{K}\boldsymbol{\Sigma}_p$$
$$\boldsymbol{K} = \boldsymbol{\Sigma}_p(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_y)^{-1}$$

Combining Gaussians (A mathematical de-tour) Part IV

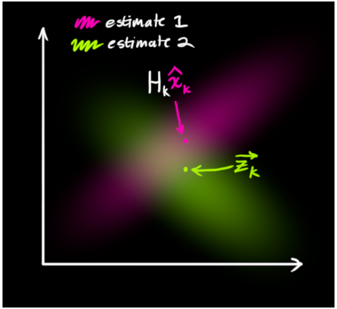
- We've assumed that the predictions and measurements have the same units and scale, which resulted in nice and tidy equations.
- In reality, we usually have to map predictions and measurements into the same domain. We'll use the matrix $\boldsymbol{H}_t$ for this.

Comparing predictions with measurements



- The measurement $\boldsymbol{y}_t$ might have different units and scale than the prediction $\boldsymbol{\mu}_t$, because sensors do not necessarily measure state directly. Think of how height of mercury in thermometers is converted into temperature.
- We therefore model the sensor with the matrix $\boldsymbol{H}_t$, which maps the state vector parameters into the measurement domain.

Comparing predictions with measurements - II

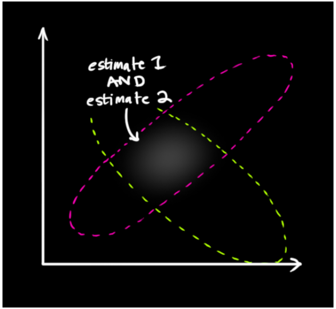


- In order to compare the posterior $\hat{\boldsymbol{x}}_t$ with the measurement $\boldsymbol{y}_t$, we must transform the posterior to the same space as the measurement by,

$$\mathbb{E}[\boldsymbol{H}_t\hat{\boldsymbol{x}}_t] = \boldsymbol{H}_t\boldsymbol{\mu}_t$$

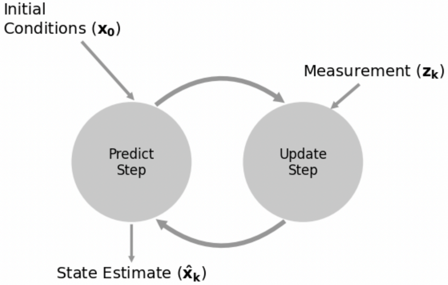
$$Cov(\boldsymbol{H}_t\hat{\boldsymbol{x}}_t) = \hat{\boldsymbol{H}}_t\boldsymbol{\Sigma}_t\hat{\boldsymbol{H}}_t^T$$

Putting it all together



$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_{t|t-1})$$
$$\boldsymbol{\Sigma}_{t|t} = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H}_t)\boldsymbol{\Sigma}_{t|t-1}$$
$$\boldsymbol{K} = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^t(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{\Sigma}_{y_t})^{-1}$$

Predict step + Update step



Predict step

$$\boldsymbol{\mu}_{t|t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1}$$

Update step

$$\boldsymbol{\mu}_{t|t}$$

$$\boldsymbol{\Sigma}_{t|t}$$

References

- https://www.youtube.com/watch?v=IFeCIbljreY
- https://www.bzarg.com/p/how-a-kalman-filter-works-in-pictures/
- https://ieeexplore.ieee.org/document/6279585
- https://www.tandfonline.com/doi/abs/10.1080/00031305.2016.1141709

Exercise 1: Kalman filter

- `https://github.com/Sonat-Consulting/kf-demo`

Predict step

$$\boldsymbol{\mu}_{t|t-1} = \boldsymbol{M}\boldsymbol{\mu}_{t-1|t-1}$$
$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{M}_t\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{M}_t^T$$

Update step

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_{t|t-1})$$
$$\boldsymbol{\Sigma}_{t|t} = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H}_t)\boldsymbol{\Sigma}_{t|t-1}$$
$$\boldsymbol{K} = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^t(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{\Sigma}_{y_t})^{-1}$$

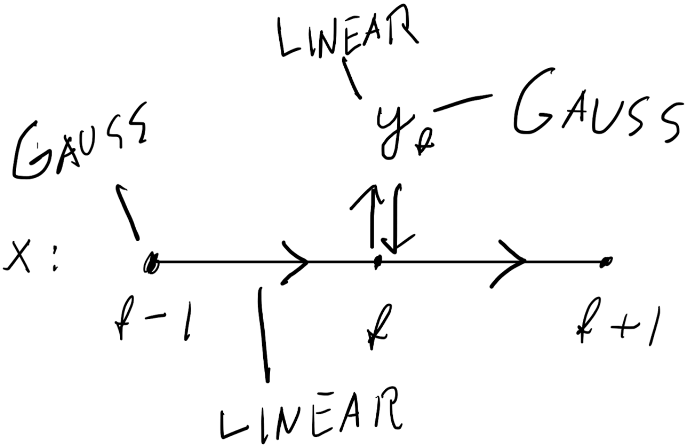
Data assimilation and the Kalman filter

Extended Kalman Filter

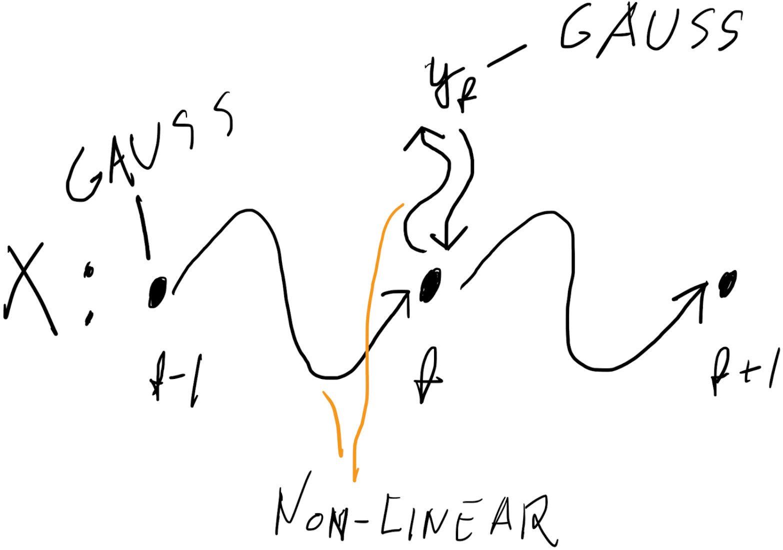
Ensemble Kalman Filter

Ensemble Information Filter

What we have assumed



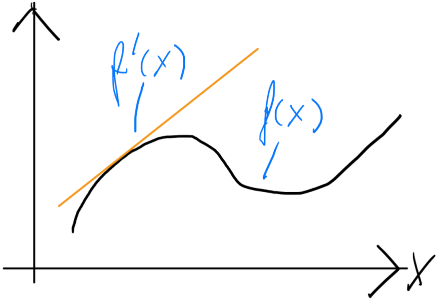
Problem: The world is wiggly



The Extended Kalman Filter solution

Approximate wigglyness! Linearly!

Then the normality of the Kalman filter still holds.
Use *first order Taylor expansions of non-linear functions*



Exercise 2: Extended Kalman filter

- `https://github.com/Sonat-Consulting/kf-demo`

Predict step

$$\boldsymbol{\mu}_{t|t-1} = \boldsymbol{M}_t \boldsymbol{\mu}_{t-1|t-1}$$
$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{M}_t \boldsymbol{\Sigma}_{t-1|t-1} \boldsymbol{M}_t^T$$

Update step

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}(\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{\mu}_{t|t-1})$$
$$\boldsymbol{\Sigma}_{t|t} = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H}_t)\boldsymbol{\Sigma}_{t|t-1}$$
$$\boldsymbol{K} = \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{H}_t^t (\boldsymbol{H}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{H}_t^T + \boldsymbol{\Sigma}_{y_t})^{-1}$$

- Set $\boldsymbol{H}_t$ and/or $\boldsymbol{M}_t$ dynamically
- Use a linear first-order Taylor approximation

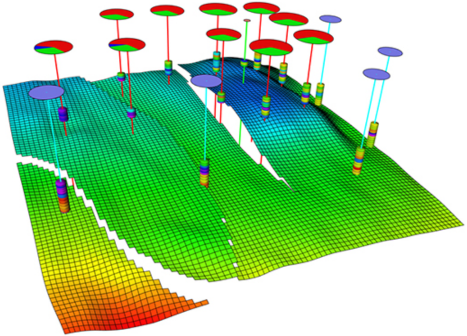
$$\boldsymbol{x}_t = \boldsymbol{m}(\boldsymbol{x}_{t-1}) \longrightarrow \boldsymbol{M}_t = \left.\frac{\partial \boldsymbol{m}}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{x}=\boldsymbol{\mu}_{t-1|t-1}}$$

$$\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t) \longrightarrow \boldsymbol{H}_t = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\mu_{t|t-1}}$$

Problem: Highly non-linear dynamics

Linear approximations of wigglyness is not very non-linear

- Underlying distribution is not preserved



The Ensemble Kalman Filter solution: Just do the following

Using (3.41), (3.50) and (3.51) in (3.45) gives $\boldsymbol{b} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}$. Furthermore, by using (3.23), (3.25), (3.41) and (3.45), in addition to the two definitions above, the second term in (3.49) becomes

$$
\begin{aligned}
&-2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad= -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad= -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\left(\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}])\right)^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad= -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\left(\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}])\right)^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad= -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[\psi_4^{\mathrm{t}} - \psi_4^{\mathrm{f}}]}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 + 0 \\
&\quad= -2\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\psi_4^{\mathrm{t}} - \psi_4^{\mathrm{f}})}]\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= -2\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_4)]\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= -2\boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2.
\end{aligned}
\tag{3.52}
$$

Here we have also used that $\overline{\epsilon} = 0$ from (3.25), and that $\boldsymbol{\mathcal{P}}$ is a symmetrical function of the covariance and can be moved outside the averaging.

Further, using $(\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}} = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}$, the last term becomes

$$
\begin{aligned}
&\boldsymbol{r}_1^{\mathrm{T}}\overline{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{\boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{f}}])(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{f}}])^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{f}}])(\boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_2^{\mathrm{f}}])^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}}] + \boldsymbol{\epsilon})(\boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}}] + \boldsymbol{\epsilon})^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{\mathcal{M}}_{(1)}\boldsymbol{\mathcal{M}}_{(2)}^{\mathrm{T}}[\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}})}] + \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}})\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad= \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2.
\end{aligned}
\tag{3.53}
$$

Thus, an error estimate is given as

$$
\begin{aligned}
C_{\psi\psi}^{\mathrm{a}}(\boldsymbol{x}_1, \boldsymbol{x}_2) &= C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_2) \\
&\quad - \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x}_1)\left(\boldsymbol{\mathcal{M}}_{(3)}\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_4)] + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\boldsymbol{r}(\boldsymbol{x}_2).
\end{aligned}
\tag{3.54}
$$

where the definition for $\boldsymbol{\mathcal{P}}$ has been used.

3.2.6 Uniqueness of the solution

By expressing the solution as in (3.39) not all arbitrary functions can be represented. To show that the solution (3.39) is the unique variance minimizing

linear solution we proceed with the following argumentation using a geometrical formulation, identical to the formulation used for the time dependent problem by *Bennett* (1992). First define the inner product

$$
< f(\boldsymbol{x}_1), g(\boldsymbol{x}_2) >= \iint_{\mathcal{D}} f(\boldsymbol{x}_1) W_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_2) g(\boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2. \tag{3.55}
$$

Note that

$$
\begin{aligned}
&< C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_1), \psi(\boldsymbol{x}_2) > \\
&\qquad= \iint_{\mathcal{D}} C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_1) W_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_2)\psi(\boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\
&\qquad= \psi(\boldsymbol{x}_3),
\end{aligned}
\tag{3.56}
$$

thus, $C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_1)$ is a "reproducing kernel" for the inner product (3.55) and the expression (3.56) is true for every field ψ in any point $\boldsymbol{x}$.

Recalling the definition of the representer (3.41) we get

$$
\begin{aligned}
< \boldsymbol{r}(\boldsymbol{x}_1), \psi(\boldsymbol{x}_2) > &=< \boldsymbol{\mathcal{M}}_{(1)}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_1)], \psi(\boldsymbol{x}_2) > \\
&= \boldsymbol{\mathcal{M}}_{(1)}[< C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_1), \psi(\boldsymbol{x}_2) >] \\
&= \boldsymbol{\mathcal{M}}_{(1)}[\psi(\boldsymbol{x}_1)]
\end{aligned}
\tag{3.57}
$$

Thus, the measurement of a field $\psi(\boldsymbol{x})$ is equivalent to projecting the field onto the representer using the inner product (3.55).

The penalty function (3.26) can now be written entirely in terms of inner products as

$$
\mathcal{J}[\psi] =< \psi^{\mathrm{f}} - \psi, \psi^{\mathrm{f}} - \psi > + (\boldsymbol{d} - < \psi, \boldsymbol{r} >)^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{d} - < \psi, \boldsymbol{r} >). \tag{3.58}
$$

Assume now that the minimizing solution is expressed as

$$
\psi^{\mathrm{a}}(\boldsymbol{x}) = \psi^{\mathrm{f}}(\boldsymbol{x}) + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{3.59}
$$

where $g(\boldsymbol{x})$ is an arbitrary function orthogonal to the representers, i.e.

$$
< g, \boldsymbol{r} >= \boldsymbol{0}. \tag{3.60}
$$

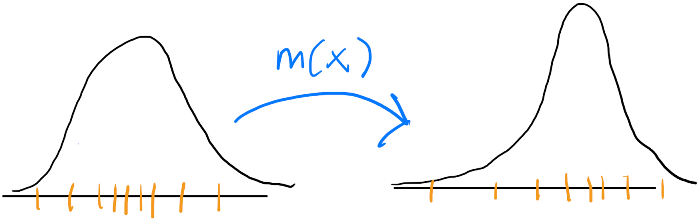
Because of this identity the field g may be regarded as unobservable. Substituting (3.59) into (3.58) gives

$$
\begin{aligned}
\mathcal{J}[\psi^{\mathrm{a}}] &=< \boldsymbol{r}^{\mathrm{T}}\boldsymbol{b} + g, \boldsymbol{r}^{\mathrm{T}}\boldsymbol{b} + g > \\
&\quad + (\boldsymbol{d} - < \psi^{\mathrm{a}}, \boldsymbol{r} >)^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{d} - < \psi^{\mathrm{a}}, \boldsymbol{r} >) \\
&= \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > \boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, g > + < g, \boldsymbol{r}^{\mathrm{T}} > \boldsymbol{b} + < g, g > \\
&\quad + (\boldsymbol{d} - < \psi^{\mathrm{f}}, \boldsymbol{r} > - \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > - < g, \boldsymbol{r} >)^{\mathrm{T}} \\
&\qquad \times \boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{d} - < \psi^{\mathrm{f}}, \boldsymbol{r} > - \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > - < g, \boldsymbol{r} >).
\end{aligned}
\tag{3.61}
$$

JK, kinda: The Ensemble Kalman Filter solution

Drop the Kalman predict equations, and instead sample at time $t-1$

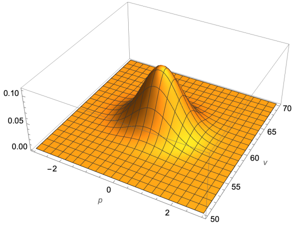
Bring realizations forward in time $\boldsymbol{x}_t = m(\boldsymbol{x}_{t-1})$, have correct sample at time t!



Problem: Highly non-linear dynamics

The mean and covariance are easy to estimate from a sample

- Mean and covariance known: Gaussian is the maximum entropy distribution.
- A Gaussain at $t \longrightarrow$ the KF update equations may be employed.
- Update sample directly or resample from the updated Gaussian with $\boldsymbol{\mu}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}$



Note: There are many numerical tricks that can be employed.

Exercise 3: Ensemble Kalman filter

- `https://github.com/Sonat-Consulting/kf-demo`

Sample from belief

$$\boldsymbol{x}_{t-1|t-1}^{(i)} \sim N\left(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1}\right) \quad i = 1, \dots, n$$

Predict

$$\boldsymbol{x}_{t|t-1}^{(i)} = g(\boldsymbol{x}_{t-1|t-1}^{(i)})$$

Estimate

Using sample $\{\boldsymbol{x}_{t|t-1}^{(i)}\}_{i=1}^{n}$ estimate $\hat{\boldsymbol{\mu}}_{t|t-1}$ and $\hat{\boldsymbol{\Sigma}}_{t|t-1}$

Update

$$\boldsymbol{\mu}_{t|t} = \hat{\boldsymbol{\mu}}_{t|t-1} + \boldsymbol{K}(\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{\mu}_{t|t-1})$$

$$\boldsymbol{\Sigma}_{t|t} = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H}_t)\hat{\boldsymbol{\Sigma}}_{t|t-1}$$

$$\boldsymbol{K} = \hat{\boldsymbol{\Sigma}}_{t|t-1}\boldsymbol{H}_t^t(\boldsymbol{H}_t\hat{\boldsymbol{\Sigma}}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{\Sigma}_{y_t})^{-1}$$

Data assimilation and the Kalman filter

Extended Kalman Filter

Ensemble Kalman Filter

Ensemble Information Filter