Winterschool on Data Assimilation - The Kalman Filter and Regularization Theory

Roland Potthast

Deutscher Wetterdienst / University of Reading / Universität Göttingen

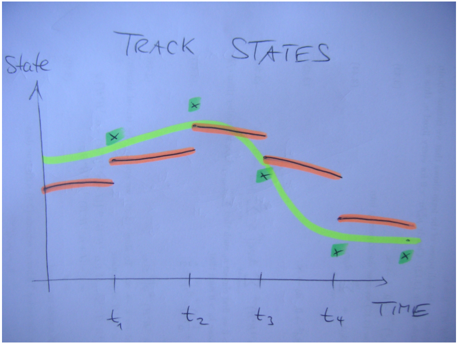
DWD Offenbach
Feb 13-17, 2011

Outline

Outline

Outline

Introduction

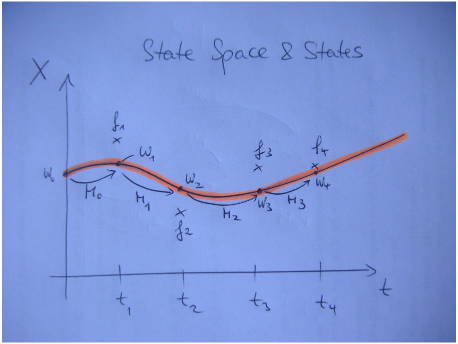


We want to calculate and predict the state of a dynamical system.

We get measurements in at particular points in time.

Our task is to calculate an optimal state estimate.

Some Notation



- State Space X with states w.
- Model M_k mapping the state w_k at time t_k into the state w_{k+1} at time t_{k+1}.
- Measurements f_k at time t_k in the measurement space Y.

Background Material

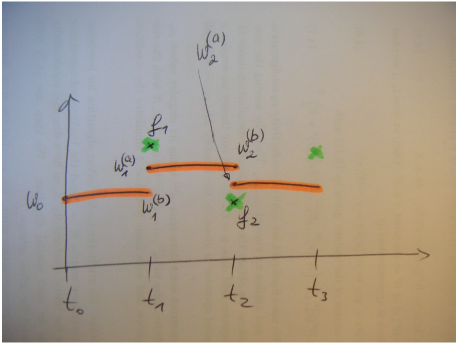
Here we assume that 3dVar and 4dVar are well-known from preceeding lectures. Recall the 3dVar analysis equation

$$w_k^{(a)} = w_k^{(b)} + BH'\left(R + HBH'\right)^{-1}\left(f_k - Hw_k^{(b)}\right) \tag{1}$$

for $k = 1, 2, 3, \ldots$ and the propagation

$$w_k^{(b)} = M_{k-1} w_{k-1}^{(a)}. \tag{2}$$

Background and Analysis



Typical image generated by 3dVar.

Background state, also called first guess

$$w_k^{(b)}$$

Analysis state at time t_k is

$$w_k^{(a)}$$

Recall some notation

$$\|w\|^2 := \sum_{j=1}^{n} w_j^2 \tag{3}$$

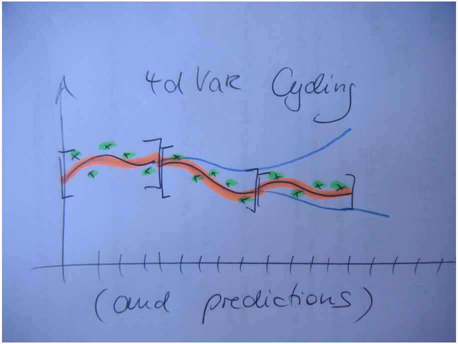
for the **Euclidean norm** or **metric** $d(w, v) = \|w - v\|$ and

$$\langle w, v \rangle := \sum_{j=1}^{n} w_j v_j \tag{4}$$

for the **scalar product**. We might write this in a more *engineering type notation* and a *mathematical notation*

$$\langle w, v \rangle = w^T v = w \cdot v = w^T \circ v \tag{5}$$

4dVar



4dVar is fitting whole trajectories to the measurements.

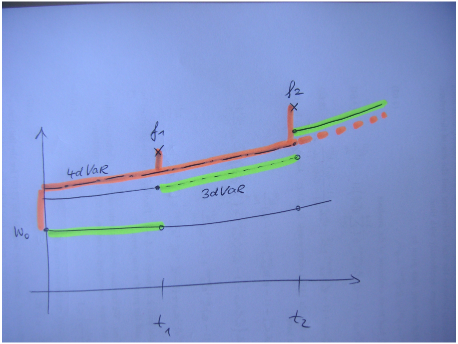
Here, we need some window, over which the fitting is carried out.

Then, the states are propagated to the start time of a new window and another cycle of 4dVar is calculated.

Outline

Outline

4dVar versus 3dVar



Comparison of 4dVar and 3dVar

Our key question.

How do we need to modify 3dVar to make it equivalent to 4dVar (at least for linear systems)?

4dVar Step by Step???

Let us study the 4dVar minimization with

measurements f_1 at t_1 and with f_2 at t_2

$$J_{4dvar}(w) := \|w - w_0^{(b)}\|^2_{B^{-1}} + \|f_1 - HM_0 w\|^2_{R^{-1}} + \|f_2 - HM_1 M_0 w\|^2_{R^{-1}} \quad (6)$$

We assume that M_0, M_1 and H are linear.

Can we do that in two steps?

4dVar Step by Step???

Decompose

$$J_{4dvar}(w) := \|w - w_0^{(b)}\|^2_{B^{-1}} + \|f_1 - HM_0 w\|^2_{R^{-1}} + \|f_2 - HM_1 M_0 w\|^2_{R^{-1}} \quad (7)$$

into

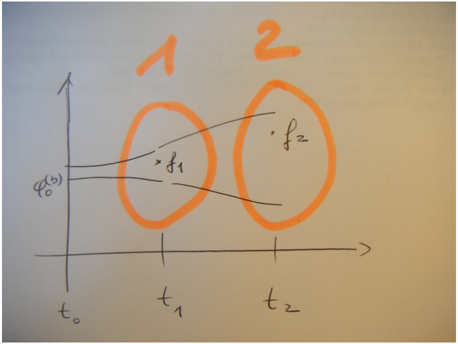
$$J_1(w) := \|w - w_0^{(b)}\|^2_{B^{-1}} + \|f_1 - HM_0 w\|^2_{R^{-1}} \quad (8)$$

and

$$J_2(w) := \|w - \tilde{w}^{(a)}\|^2_{\tilde{B}^{-1}} + \|f_2 - HM_1 M_0 w\|^2_{R^{-1}} \quad (9)$$

where $\tilde{w}^{(a)}$ and $\tilde{B}$ incorporate the information from the the first step!

Decompose it into two steps

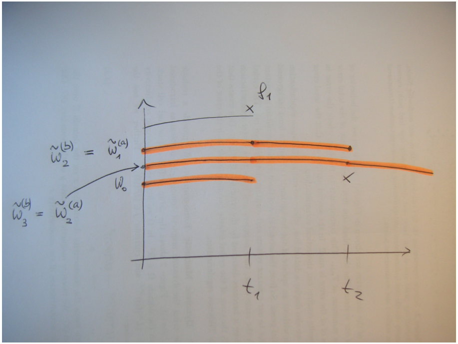


We first want to assimilate f_1 measured at t_1.

Then, in a second step, we want to assimilate f_2 measured at t_2.

But we want to get the same as if we had done it both simultaneously!

Pull it all to the time t_0



First, we need to generate a uniform framework.

We work with states $\tilde{w}_k$ at time t_0 representing states w_k at time t_k which need to assimilate data f_k.

$$w_k = M_{0,k}\tilde{w}_k$$

Simle example in one dimension

$$a(x-b)^2 + c(e-x)^2 + d = g(x-h)^2 + constant \tag{10}$$

We need to determine g and h.

We calculate

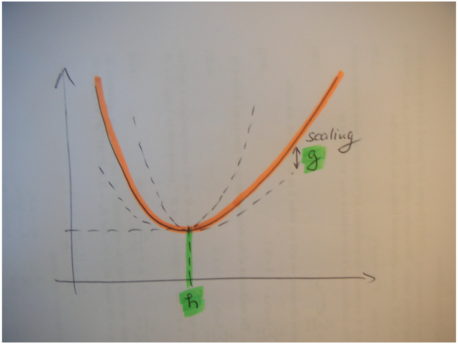
$$a(x-b)^2 + c(x-e)^2 + d = (a+c)x^2 - 2(ab+ec)x + constant$$

and

$$d(x-e)^2 = dx^2 - 2dex + constant$$

This yields $d = a + c$ and $e = d^{-1}(ab + ec)$.

4dVar versus 3dVar



Bring the quadratic function into **vertex form**.

The scaling is given by g, the centre and minimum is located at h.

Collecting all terms quadratic in w, then all terms linear in w etc we obtain

$$\begin{aligned} J_1(w) &= \left\langle w - w_0^{(b)}, B^{-1}\left(w - w_0^{(b)}\right)\right\rangle \\ &\qquad + \left\langle f_1 - HM_0 w, R^{-1}\left(f_1 - HM_0 w\right)\right\rangle \\ &= \underbrace{\left\langle w, \left(B^{-1} + M_0^* H^* R^{-1} HM_0\right) w\right\rangle}_{\text{terms quadratic in } w} \\ &\qquad -2 \underbrace{\left\langle w, B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 \right\rangle}_{\text{terms linear in } w} + c \end{aligned} \tag{11}$$

with some constant c not depending on w.

Collecting all terms quadratic in w, then all terms linear in w etc we obtain

$$J_1(w) = \left\langle w - w_0^{(b)}, B^{-1}\left(w - w_0^{(b)}\right)\right\rangle + \left\langle f_1 - HM_0 w, R^{-1}\left(f_1 - HM_0 w\right)\right\rangle$$

$$= \underbrace{\left\langle w, \left(B^{-1} + M_0^* H^* R^{-1} HM_0\right) w\right\rangle}_{\text{terms quadratic in } w} - 2 \underbrace{\left\langle w, B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1\right\rangle}_{\text{terms linear in } w} + c \tag{11}$$

with some constant c not depending on w.

In the same way we have

$$\|w - \tilde{w}^{(a)}\|^2_{\tilde{B}^{-1}} = \left\langle w, \tilde{B}^{-1} w \right\rangle - 2\left\langle w, \tilde{B}^{-1} \tilde{w}^{(a)} \right\rangle + \tilde{c} \qquad (12)$$

with some constant $\tilde{c}$ not depending on w. From above we had

$$J_1(w) = \left\langle w, \left(B^{-1} + M_0^* H^* R^{-1} H M_0\right) w \right\rangle - 2\left\langle w, B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 \right\rangle + c$$

We now need to determine $\tilde{w}^{(a)}$ and $\tilde{B}$ such that

$$J_1(w) \stackrel{!}{=} \|w - \tilde{w}^{(a)}\|^2_{\tilde{B}^{-1}} \qquad (13)$$

In the same way we have

$$\| w - \tilde{w}^{(a)} \|^2_{\tilde{B}^{-1}} = \left\langle w, \tilde{B}^{-1} w \right\rangle - 2 \left\langle w, \tilde{B}^{-1} \tilde{w}^{(a)} \right\rangle + \tilde{c} \qquad (12)$$

with some constant $\tilde{c}$ not depending on w. From above we had

$$J_1(w) = \left\langle w, (B^{-1} + M_0^* H^* R^{-1} H M_0) w \right\rangle - 2 \left\langle w, B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 \right\rangle + c$$

We now need to determine $\tilde{w}^{(a)}$ and $\tilde{B}$ such that

$$J_1(w) \stackrel{!}{=} \| w - \tilde{w}^{(a)} \|^2_{\tilde{B}^{-1}} \qquad (13)$$

In the same way we have

$$\| w - \tilde{w}^{(a)} \|^2_{\tilde{B}^{-1}} = \left\langle w, \tilde{B}^{-1} w \right\rangle - 2 \left\langle w, \tilde{B}^{-1} \tilde{w}^{(a)} \right\rangle + \tilde{c} \qquad (12)$$

with some constant $\tilde{c}$ not depending on w. From above we had

$$\begin{aligned} J_1(w) \quad = \quad & \left\langle w, \left(B^{-1} + M_0^* H^* R^{-1} H M_0 \right) w \right\rangle \\ & -2 \left\langle w, B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 \right\rangle \; + \; c \end{aligned}$$

We now need to determine $\tilde{w}^{(a)}$ and $\tilde{B}$ such that

$$J_1(w) \stackrel{!}{=} \| w - \tilde{w}^{(a)} \|^2_{\tilde{B}^{-1}} \qquad (13)$$

This yields

$$\tilde{B}^{-1} = B^{-1} + M_0^* H^* R^{-1} H M_0 \tag{14}$$

and

$$\tilde{B}^{-1} \tilde{w}^{(a)} = B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 . \tag{15}$$

leading to

$$\tilde{w}^{(a)} = \tilde{B} \Big(B^{-1} w_0^{(b)} + M_0^* H^* R^{-1} f_1 \Big) . \tag{16}$$

We further note

$$(ab)^{-1} = b^{-1}a^{-1},$$

such that

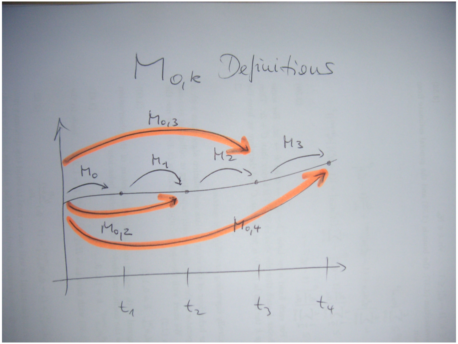
$$\begin{aligned}\tilde{B} &= \Big(B^{-1} + \underbrace{B^{-1}B}_{=I}\ M_0^* H^* R^{-1} H M_0\Big)^{-1} \\ &= \Big(B^{-1}\left\{I + B\ M_0^* H^* R^{-1} H M_0\right\}\Big)^{-1} \\ &= \left(I + B\ M_0^* H^* R^{-1} H M_0\right)^{-1} B\end{aligned}$$

We finally get

$$\begin{aligned}
\tilde{w}^{(a)} &= \tilde{B}\Big(B^{-1}w_0^{(b)} + M_0^* H^* R^{-1} f_1\Big) \\
&= \big(I + BM_0^* H^* R^{-1} HM_0\big)^{-1}\Big(w_0^{(b)} + BM_0^* H^* R^{-1} f_1\Big) \\
&= w_0^{(b)} + \big(I + BM_0^* H^* R^{-1} HM_0\big)^{-1} BM_0^* H^* R^{-1}\Big(f_1 - HM_0 w_0^{(b)}\Big), \\
&= w_0^{(b)} + BM_0^* H^* \big(R + HM_0 BM_0^* H^*\big)^{-1}\Big(f_1 - HM_0 w_0^{(b)}\Big),
\end{aligned} \tag{17}$$

which is the minimizer of J_1 as in 3dVar with the covariance matrix B!

Some Notation



M_0 maps states from t_0 to t_1.

M_1 maps states from t_1 to t_2.

M_2 maps states from t_2 to t_3.

...

$M_1 M_0$ maps from t_0 to t_2.

$M_2 M_1 M_0$ maps from t_0 to t_3.

...

$M_{0,k} := M_{k-1} M_{k-2} \cdots M_1 M_0$

maps from t_0 to t_k

Outline

Definition (Kalman Smoother (KS))

Let H and M_k, $k = 0, 1, 2, \ldots$ be linear and assume that measurements $f_1, f_2, \ldots$ at times $t_1, t_2, \ldots$ are given. Then, we calculate weight matrices

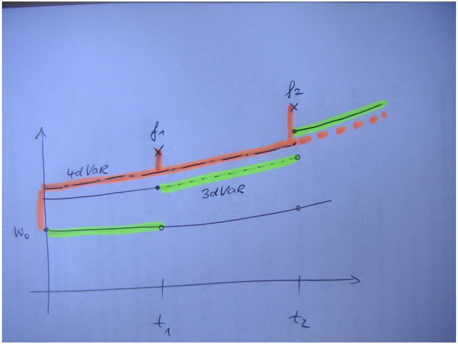
$$\tilde{B}_k^{-1} := \tilde{B}_{k-1}^{-1} + M_{0,k}^* H_k^* R^{-1} H_k M_{0,k}, \quad k = 1, 2, \ldots \tag{18}$$

with $\tilde{B}_0 = B$ and analysis states $\tilde{w}_k^{(a)}$ at time t_0 defined by

$$\begin{aligned} \tilde{w}_k^{(a)} \quad := \quad & \tilde{w}_{k-1}^{(a)} + \tilde{B}_{k-1} M_{0,k}^* H_k^* \\ & (R + H_k M_{0,k} \tilde{B}_{k-1} M_{0,k}^* H_k^*)^{-1} \left(f_k - H_k M_{0,k} \tilde{w}_{k-1}^{(a)} \right) \end{aligned} \tag{19}$$

for $k = 1, 2, \ldots$ with $\tilde{w}_0^{(a)} = w_0^{(b)}$.

Kalman Smoother Adapts the weight



The Kalman Smoother generates the orange dotted result.

For linear systems the Kalman Smoother is equivalent to 4dVar.

4dVar and the Kalman Smoother

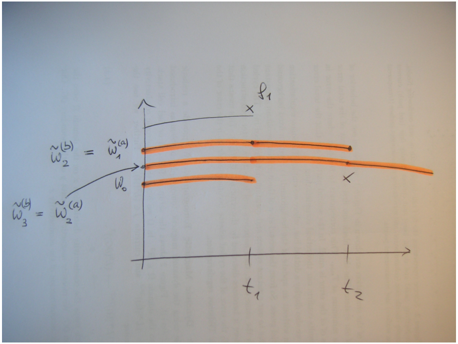
Theorem (Equivalence 4dVar/KS)

Let H and M_k, $k = 0, 1, 2, \ldots$ be linear and data $f_1, f_2, \ldots$ be given. Then, 4dVar carried out with data $f_1, \ldots, f_k$ is equivalent to the Kalman Smoother in the sense that the minimum of the 4dVar functional is given by the analysis $\tilde{w}_k^{(a)}$ for $k = 1, 2, \ldots, K$.

- The Kalman Smoother calculates an analysis at some time t_0.
- Of course, we can now **cycle** the Kalman Smoother. Then, we obtain *smoothened* analysis results for successive analysis times.
- For nonlinear systems we need the **adjoint tangent linear** model for the smoother, which is often challenging to calculate.

Outline

Map it back from t_0 to time t_k



We go back from our uniform framework to carry out our analysis at measurement time.

We work with states w_k at time t_k.

$$w_k = M_{0,k}\tilde{w}_k$$

Transformation of Kalman Smoother to time t_k

For the Kalman Smoother we calculated states $\tilde{w}_k^{(a)}$ at time t_0.

We proparage them to time t_k by

$$w_k^{(a)} = M_{0,k}\tilde{w}_k^{(a)} = M_{k-1}M_{k-2}\cdots M_1 M_0 \tilde{w}_k^{(a)} \tag{20}$$

for $k = 1, 2, 3, \ldots$.

Recall that the background at time t_k is calculated from the analysis at time t_{k-1} by propagation

$$w_k^{(b)} = M_{k-1} w_{k-1}^{(a)} \tag{21}$$

Transformation of Kalman Smoother to time t_k

For the Kalman Smoother we calculated states $\tilde{w}_k^{(a)}$ at time t_0.

We proparage them to time t_k by

$$w_k^{(a)} = M_{0,k}\tilde{w}_k^{(a)} = M_{k-1}M_{k-2}\cdots M_1 M_0 \tilde{w}_k^{(a)} \tag{20}$$

for $k = 1, 2, 3, ...$

Recall that the background at time t_k is calculated from the analysis at time t_{k-1} by propagation

$$w_k^{(b)} = M_{k-1} w_{k-1}^{(a)} \tag{21}$$

Transformation of Kalman Smoother to time t_k

For the Kalman Smoother we calculated states $\tilde{w}_k^{(a)}$ at time t_0.

We proparage them to time t_k by

$$w_k^{(a)} = M_{0,k}\tilde{w}_k^{(a)} = M_{k-1}M_{k-2}\cdots M_1 M_0 \tilde{w}_k^{(a)} \tag{20}$$

for $k = 1, 2, 3, ...$.

Recall that the background at time t_k is calculated from the analysis at time t_{k-1} by propagation

$$w_k^{(b)} = M_{k-1}w_{k-1}^{(a)} \tag{21}$$

Propagation of Covariance Matrices

The matrices $\tilde{B}$ are propagated from t_0 to t_k by

$$B_k^{(b)} = M_{0,k} \tilde{B}_{k-1} M_{0,k}^* \tag{22}$$

and

$$B_k^{(a)} = M_{0,k} \tilde{B}_k M_{0,k}^* \tag{23}$$

for $k = 1, 2, 3, \ldots$, where the *background matrix* at time t_k is obtained by propagating the *analysis matrix* from time t_{k-1} to t_k by

$$B_k^{(b)} = M_{k-1} B_{k-1}^{(a)} M_{k-1}^*. \tag{24}$$

Calculations to carry out the transformation

Multiply the analysis of the Kalman Smoother by $M_{0,k}$ to obtain

$$\begin{aligned} M_{0,k}\tilde{w}_k^{(a)} &= M_{0,k}\tilde{w}_{k-1}^{(a)} + M_{0,k}\tilde{B}_{k-1}M_{0,k}^{*}H_k^{*} \\ &\quad \left(R + H_k M_{0,k}\tilde{B}_{k-1}M_{0,k}^{*}H_k^{*}\right)^{-1}\left(f_k - H_k M_{0,k}\tilde{w}_{k-1}^{(a)}\right) \end{aligned} \tag{25}$$

which by $w_k^{(a)} = M_{0,k}\tilde{w}_k^{(a)}$, $w_k^{(b)} = M_{0,k}\tilde{w}_{k-1}^{(a)}$ and $B_k^{(b)} = M_{k-1}B_{k-1}^{(a)}M_{k-1}^{*}$ is equal to

$$w_k^{(a)} = w_k^{(b)} + B_k^{(b)}H_k^{*}(R + H_k B_k^{(b)}H_k^{*})^{-1}(f_k - H_k w_k^{(b)}) \tag{26}$$

Kalman Filter Equations - Covariance

The calculation for the covariance update

$$\tilde{B}_k^{-1} := \tilde{B}_{k-1}^{-1} + M_{0,k}^* H_k^* R^{-1} H_k M_{0,k}, \quad k = 1, 2, \ldots \tag{27}$$

is multiplied by $M_{0,k}^{-1}$ from the right and by $\left(M_{0,k}^*\right)^{-1}$ from the left to obtain

$$\left(M_{0,k}^*\right)^{-1} \tilde{B}_k^{-1} M_{0,k}^{-1} := \left(M_{0,k}^*\right)^{-1} \tilde{B}_{k-1}^{-1} M_{0,k}^{-1} + H_k^* R^{-1} H_k, \quad k = 1, 2, \ldots \tag{28}$$

which yields

$$\left(B_k^{(a)}\right)^{-1} = \left(B_k^{(b)}\right)^{-1} + H_k^* R^{-1} H_k \tag{29}$$

Kalman Filter Equations - Analysis State

The analysis equation is given by

$$w_k^{(a)} = w_k^{(b)} + B_k^{(b)} H_k^* \big(R + H_k B_k^{(b)} H_k^* \big)^{-1} \big(f_k - H_k w_k^{(b)} \big) \qquad (30)$$

for $k \in \mathbb{N}$, usually written in the form

$$w_k^{(a)} = w_k^{(b)} + K_k \big(f_k - H_k w_k^{(b)} \big) \qquad (31)$$

with the *Kalman gain matrix* (or *Tikhonov Regularization Operator*)

$$K_k := B_k^{(b)} H_k^* \big(R + H_k B_k^{(b)} H_k^* \big)^{-1}. \qquad (32)$$

The weight update is

$$\big(B_k^{(a)} \big)^{-1} = \big(B_k^{(b)} \big)^{-1} + H_k^* R^{-1} H_k. \qquad (33)$$

B matrix analysis equation

Lemma

For $k \in N$ we have

$$B_k^{(a)} = (I - K_k H_k) B_k^{(b)}. \tag{34}$$

Proof. We start from (33) in the form

$$B_k^{(a)} = \left(I + B_k^{(b)} H_k^* R^{-1} H_k \right)^{-1} B_k^{(b)}. \tag{35}$$

We expand

$$\begin{aligned} T &:= \left(I + B_k^{(b)} H_k^* R^{-1} H_k \right) (I - K_k H_k) \\ &= \left(I + B_k^{(b)} H_k^* R^{-1} H_k \right) \left(I - B_k^{(b)} H_k^* (R + H_k B_k^{(b)} H_k^*)^{-1} H_k \right) \\ &= I + B_k^{(b)} H_k^* R^{-1} H_k - B_k^{(b)} H_k^* (R + H_k B_k^{(b)} H_k^*)^{-1} H_k \\ &\quad - B_k^{(b)} H_k^* R^{-1} H_k B_k^{(b)} H_k^* (R + H_k B_k^{(b)} H_k^*)^{-1} H_k \end{aligned} \tag{36}$$

DWD

B matrix analysis equation

$$
\begin{aligned}
T &= I + \underbrace{B_k^{(b)} H_k^* R^{-1} H_k}_{=:S} - \underbrace{B_k^{(b)} H_k^* (R + H_k B_k^{(b)} H_k^*)^{-1} H_k}_{=:S_1} \\
&\quad - \underbrace{B_k^{(b)} H_k^* R^{-1} H_k B_k^{(b)} H_k^* (R + H_k B_k^{(b)} H_k^*)^{-1} H_k}_{:=S_2} .
\end{aligned}
\tag{37}
$$

Remark that

$$
\begin{aligned}
S &= B_k^{(b)} H_k^* R^{-1} (R + H_k B_k^{(b)} H_k^*)(R + H_k B_k^{(b)} H_k^*)^{-1} H_k \\
&= S_1 + S_2 ,
\end{aligned}
\tag{38}
$$

which yields $T = I$. Thus

$$
\left(I + B_k^{(b)} H_k^* R^{-1} H_k \right)^{-1} = (I - K_k H_k)
$$

and the proof is complete. □

Kalman Filter

Definition (Kalman Filter)

Starting with an initial state $w_0^{(b)}$ and an initial matrix $B_0^{(a)} := B$, for $k \in \mathbb{N}$ the Kalman Filter iteratively calculates an analysis $w_k^{(a)}$ at time t_k by

1. propagating the state $w_{k-1}^{(a)}$ from t_{k-1} to t_k via (21), $w_k^{(b)} = M_{k-1} w_{k-1}^{(a)}$
2. propagating $B_{k-1}^{(a)}$ from t_{k-1} to t_k following (24),

$$B_k^{(b)} = M_{k-1} B_{k-1}^{(a)} M_{k-1}^* \tag{39}$$

3. calculating an analysis state by (30)

$$K_k := B_k^{(b)} H_k^* \left(R + H_k B_k^{(b)} H_k^* \right)^{-1} \tag{40}$$

$$w_k^{(a)} = w_k^{(b)} + K_k \left(f_k - H_k w_k^{(b)} \right) \tag{41}$$

4. calculating an analysis weight by (34), $B_k^{(a)} = (I - K_k H_k) B_k^{(b)}$.

Equivalence Result

Theorem (Equivalence 4dVar/KF/KS)

Let H_k for $k \in \mathbb{N}$ and M_k for $k \in \mathbb{N}_0$ be linear. Let $w_k^{(a)}$ be the analysis of the Kalman Filter at time t_k, $\tilde{w}_k^{(a)}$ the analysis of the Kalman smoother with data $f_1, \ldots, f_k$ at time t_0, $\tilde{w}_{4d,k}^{(a)}$ the minimizer of the 4dVar functional at time t_0 and define

$$w_{4d,k}^{(a)} := M_{0,k} \tilde{w}_{4d,k}^{(a)}, \quad k = 1, 2, 3, \ldots \tag{42}$$

Then 4dVar is equivalent to the Kalman Filter and to the Kalman Smoother in the sense that

$$w_{4d,k}^{(a)} = w_k^{(a)} = M_{0,k} \tilde{w}_k^{(a)}. \tag{43}$$

Outline

Outline

Bayes Formula

Let π be some probability density on the state space X.

Then, Bayes Formula for calculating a conditional probability density given some observation is given by

$$\pi(w|f) = \frac{\pi(w)\pi(f|w)}{\pi(f)}. \quad (44)$$

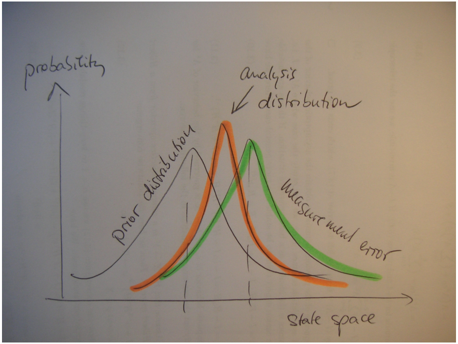
Often, we have

$$\pi(f|w) = \pi_{data}(f - Hw) \quad (45)$$

with the obervation operator H. The value $\pi(f)$ is given by the normalization condition. This leads to

$$\pi(w|f) = c\pi_{prior}(w)\pi_{data}(f - Hw). \quad (46)$$

Bayes Formula Visualized



Prior distribution black.
Measurement error distribution green.

The orange curve shows the analysis density.

Bayes Formula for Data Assimilation

Definition (Bayes Data Assimilation.)

Bayes Data Assimilation determines probability distributions $\pi_k^{(a)}$ at time t_k for the states $w \in X$ by cycling the following propagation and analysis steps.

1. **(Propagation Step.)** Calculate the prior density $\pi_k^{(b)}(w)$ at time t_k by propagating the analysis density $\pi_{k-1}^{(a)}$ from time t_{k-1} to t_k based on the (linear or nonlinear) model dynamics M_{k-1}.
2. **(Analysis Step.)** Calculate the posterior or *analysis density* $\pi_k^{(a)}(w|f_k)$ at time t_k by Bayes formula (44).

Gaussian Densities

Let us consider Gaussian densities

$$\pi(w) = \frac{1}{\sqrt{(2\pi)^n \det(B)}} e^{-\frac{1}{2}(w-\mu)^T B^{-1}(w-\mu)}, \quad w \in \mathbb{R}^n, \qquad (47)$$

around some state $\mu = w^{(b)} \in X$ with some positive define invertible matrix B. Assume that also the error distribution is Gaussian

$$\pi(f|w) = \frac{1}{\sqrt{(2\pi)^n \det(R)}} e^{-\frac{1}{2}(f-H(w))^T R^{-1}(f-H(w))}, \quad f \in \mathbb{R}^m, \qquad (48)$$

Gaussian Posterior Density

Then, according to Bayes formula (44) we obtain

$$\pi(w|f) = c \cdot \exp\Big\{ -\frac{1}{2}\Big((w-\mu)^T B^{-1} (w-\mu) + (f - H(w))^T R^{-1} (f - H(w)) \Big) \Big\} \quad (49)$$

with some constant $c > 0$. If H is linear, this again is a Gaussian density

$$\pi(w|f) = \tilde{c} \exp\Big\{ -\frac{1}{2} (w-\tilde{\mu})^T \tilde{B}^{-1} (w-\tilde{\mu}) \Big\} \quad (50)$$

with some constant $\tilde{c}$.

We obtain the same quadratic transformation problem as above!

Bayes for Gaussian densities

The mean $\tilde{\mu}$ of the posterior Gaussian distribution is given by

$$\tilde{\mu} = w^{(b)} + BH^{*}(R + HBH^{*})^{-1}(f - Hw^{(b)}) \tag{51}$$

and its covariance matrix $\tilde{B}$ is given by

$$\tilde{B}^{-1} = B^{-1} + H^{*}R^{-1}H. \tag{52}$$

Theorem (Bayes for Gaussian Densities and the Kalman Filter.)

For linear operators M_k and H the Bayes data assimilation with Gaussian densities is equivalent to a Kalman Filter.

Outline

Outline

Spectral Theorem

Let A be a real symmetric $n \times n$-matrix. Then, there is a set of n vectors $\psi^{(1)}, \ldots, \psi^{(n)} \in X$ such that

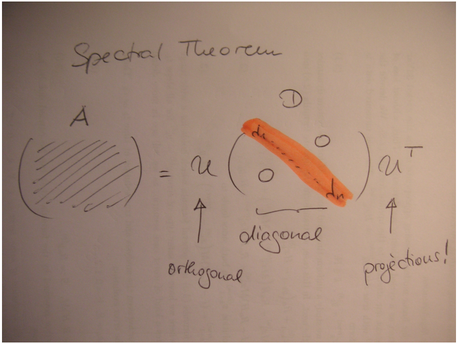
$$A\psi^{(j)} = \lambda_j \psi^{(j)}, \quad j = 1, \ldots, n. \tag{53}$$

Usually, we assume that the eigenvalues and corresponding eigenvectors are ordered according to its size.

The eigenvectors of A are orthonormal, i.e. we have

$$\langle \psi^{(j)}, \psi^{(k)} \rangle = \begin{cases} 1 & j = k \\ 0 & \text{otherwise} \end{cases}$$
$$\|\psi^{(j)}\| = 1, \quad j = 1, \ldots, n. \tag{54}$$

This is called a complete orthonormal set of eigenvectors with real eigenvalues.



A matrix vector multiplication is carried out first by representing the vector in the basis of eigenvectors.

The coefficients are calculated by application of U^T, these are projections onto the eigenvectors which constitute U by

$$U = (\psi^{(1)}, ..., \psi^{(n)}).$$

Then, application of A corresponds to a diagonal matrix, i.e. to multiplication.

Singular Value Decomposition

Study an arbitrary matrix H. Then $A := H^T H$ is symmetric, since

$$A^T = (H^T H)^T = H^T H = A.$$

We have a complete set of orthogonal eigenvectors and eigenvalues ordered according to its size. We define the **singular values** of H

$$\mu_j := \sqrt{\lambda_j} \tag{55}$$

and call the sets $\left\{\psi^{(j)} : j = 1, ..., n\right\}$ and $\left\{g^{(j)} := \mu_j^{-1} H\psi_j : j = 1, ..., n\right\}$ its **singular vectors**. These are two sets of orthonormal vectors, since we have

$$\begin{aligned} \langle g^{(j)}, g^{(k)} \rangle &= \lambda_j^{-1} \langle H\psi^{(j)}, H\psi^{(k)} \rangle \\ &= \lambda_j^{-1} \langle \psi^{(j)}, H^T H\psi^{(k)} \rangle \\ &= \lambda_j^{-1} \lambda_j \delta_{j,k} \\ &= \delta_{j,k}. \end{aligned} \tag{56}$$

We call $(\mu_j, \psi^{(j)}, g^{(j)})$ the **singular system** of H.

Singular Value Decomposition

Study an arbitrary matrix H. Then $A := H^T H$ is symmetric, since

$$A^T = (H^T H)^T = H^T H = A.$$

We have a complete set of orthogonal eigenvectors and eigenvalues ordered according to its size. We define the **singular values** of H

$$\mu_j := \sqrt{\lambda_j} \tag{55}$$

and call the sets $\left\{\psi^{(j)} : j = 1, ..., n\right\}$ and $\left\{g^{(j)} := \mu_j^{-1} H\psi_j : j = 1, ..., n\right\}$ its **singular vectors**. These are two sets of orthonormal vectors, since we have

$$\begin{aligned} \langle g^{(j)}, g^{(k)} \rangle &= \lambda_j^{-1} \langle H\psi^{(j)}, H\psi^{(k)} \rangle \\ &= \lambda_j^{-1} \langle \psi^{(j)}, H^T H\psi^{(k)} \rangle \\ &= \lambda_j^{-1} \lambda_j \delta_{j,k} \\ &= \delta_{j,k}. \end{aligned} \tag{56}$$

We call $\left(\mu_j, \psi^{(j)}, g^{(j)}\right)$ the **singular system** of H.

Singular Value Decomposition

Study an arbitrary matrix H. Then $A := H^T H$ is symmetric, since

$$A^T = (H^T H)^T = H^T H = A.$$

We have a complete set of orthogonal eigenvectors and eigenvalues ordered according to its size. We define the **singular values** of H

$$\mu_j := \sqrt{\lambda_j} \tag{55}$$

and call the sets $\left\{\psi^{(j)} : j = 1, ..., n\right\}$ and $\left\{g^{(j)} := \mu_j^{-1} H \psi_j : j = 1, ..., n\right\}$ its **singular vectors**. These are two sets of orthonormal vectors, since we have

$$\begin{aligned} \langle g^{(j)}, g^{(k)} \rangle &= \lambda_j^{-1} \langle H\psi^{(j)}, H\psi^{(k)} \rangle \\ &= \lambda_j^{-1} \langle \psi^{(j)}, H^T H \psi^{(k)} \rangle \\ &= \lambda_j^{-1} \lambda_j \delta_{j,k} \\ &= \delta_{j,k}. \end{aligned} \tag{56}$$

We call $\left(\mu_j, \psi^{(j)}, g^{(j)}\right)$ the **singular system** of H.

Singular values of observation operator H

Let $(\mu_j, \psi^{(j)}, g^{(j)})$ denote the singular system of the observation operator H. Here, for simplicity we assume H injective and $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ with $n = m$.

Then, we application of H corresponds to a multiplication by μ_j on the particular modes given by the singular vectors $\psi^{(j)}$ of H.

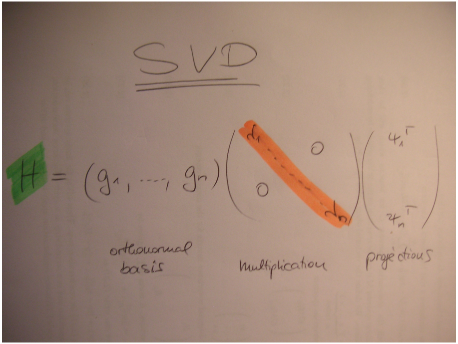
We obtain

$$H\psi^{(j)} = \mu_j g^{(j)} \tag{57}$$

by definition of $g^{(j)}$ and

$$H^T g^{(j)} = \mu_k \psi^{(j)} \tag{58}$$

which is obtained from $H^T g^{(j)} = \mu_j^{-1} H^T H \psi^{(j)} = \mu_j^{-1} \mu_j^2 \psi^{(j)}$.



Application of H corresponds to

1) projection onto the basis of eigenvectors $\psi^{(j)}$,

2) multiplication of the coefficients by μ_j,

3) Set up the result by using the coefficients with respect to the image space basis vectors $g^{(j)}$.

Outline

Spectral resolution of data equation

When we want to solve

$$Hw = f \tag{59}$$

we represent the state by

$$w = \sum_{j=1}^{n} \alpha_j \psi^{(j)} \tag{60}$$

and the measurement f by

$$f = \sum_{k=1}^{m} \beta_k g^{(k)}. \tag{61}$$

such that (59) is reduced to

$$Hw = \sum_{k=1}^{n} H\alpha_k \psi^{(k)} = \sum_{k=1}^{m} \mu_k \alpha_k g^{(k)} = \sum_{k=1}^{m} \beta_k g^{(k)} = f. \tag{62}$$

Picard Theorem

Theorem (Picard Theorem (Simple Version))

The solution of

$$Hw = f \tag{63}$$

with

$$f = \sum_{k=1}^{m} \beta_k g^{(k)} \tag{64}$$

is given by

$$w = \sum_{k=1}^{n} \frac{\beta_k}{\mu_k} \psi^{(k)} \tag{65}$$

If μ_k is small, then there are strong instabilities in the solution, small errors can be strongly amplified!

Regularization

Regularization means that we bound the influence of

$$\frac{1}{\mu_j}$$

when solving $Hw = f$.

A typical bound is achieved by replacing the term $1/\mu_j$ by

$$\frac{\mu_j}{\alpha + \mu_j^2} \tag{66}$$

for $\alpha > 0$, which for $\alpha \to 0$ tends to $1/\mu_j$.
The approach (66) is called **spectral damping**. It is equivalent to an application of the Tikhonov regularization matrix

$$R_\alpha = (\alpha I + H^T H)^{-1} H^T \tag{67}$$

replacing the inverse H^{-1}.

Many Thanks!