Kalman filter

Pavel Sakov

Nansen Environmental and Remote Sensing Center, Norway
Outline

Introduction
  KF and EnKF: historical overview
  Some conventions

Minimisation problem
  Nonlinear case
    No time, no prior information about state
    + prior information about the state
    + time dimension
    + model error
  Linear case

Kalman filter
  Theorem
  Proof
    Analysis
    Propagation
  Terminology
  Conclusions

Extended Kalman filter

Kalman smoother
  Proof
  Conclusions

Appendices
  Appendix 1: To the KF derivation
  Appendix 2: Matrix inversion lemma
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ∼1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ~1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ~1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance
  and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ~1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
KF and EnKF: historical overview

- Kalman filter (Kalman, 1960)
  propagation and update of state error covariance and mean for a linear stochastic system

- Serial square root filter (Potter, ~1963)
  propagation and update of state error covariance in a square root form

- Extended Kalman Filter (Smith et al., 1962)
  Propagation of state error covariance with linearised model

- Parallel square root filter (Andrews, 1968)
  formally equivalent to the ETKF solution

- Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)
  deterministic representation and update of state error covariance in ensemble form

- Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ~1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
KF and EnKF: historical overview

- **Kalman filter (Kalman, 1960)**
  propagation and update of state error covariance and mean for a linear stochastic system

- **Serial square root filter (Potter, ~1963)**
  propagation and update of state error covariance in a square root form

- **Extended Kalman Filter (Smith et al., 1962)**
  Propagation of state error covariance with linearised model

- **Parallel square root filter (Andrews, 1968)**
  formally equivalent to the ETKF solution

- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998)**
  Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration

- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001)**
  deterministic representation and update of state error covariance in ensemble form

- **Iterative solutions for strongly nonlinear case (Zupanski, 2005; Gu and Oliver, 2007)**
Some conventions

Fonts

capital, bold, as in $\mathbf{X}$ — a matrix
small, bold, as in $\mathbf{x}$ — a vector
small, normal, as in $x$ — a scalar
capital, fancy, as in $\mathcal{M}$ — a function

Indices

lower, without brackets, as in $X_i, x_2, x_n$ — $i$th object in a sequence
upper, without brackets, as in $X^a, \sigma^{obs}$ — used to mark the object
after a bracketed object, as in $(X)_i, (X)^T, (X)^2$ — an operation

in particular:

$(X)_i$ — $i$th column of $X$
$(X)_{i,:}$ — $i$th row of $X$
$(X)_{ij}$ — the element in $i$th row and $j$th column of $X$
$(x)_i$ — $i$th element of $x$
$(X)^T, (x)^T$ — transposed matrix or vector
Some conventions

Fonts

capital, bold, as in $\mathbf{X}$ — a matrix
small, bold, as in $\mathbf{x}$ — a vector
small, normal, as in $x$ — a scalar
capital, fancy, as in $\mathcal{M}$ — a function

Indices

lower, without brackets, as in $X_i, x_2, x_n$ — $i$th object in a sequence
upper, without brackets, as in $X^a, \sigma^{obs}$ — used to mark the object
after a bracketed object, as in $(X)_i, (X)^T, (X)^2$ — an operation

in particular:

$(X)_i$ — $i$th column of $X$
$(X)_{i,:}$ — $i$th row of $X$
$(X)_{ij}$ — the element in $i$th row and $j$th column of $X$
$(x)_i$ — $i$th element of $x$
$(X)^T, (x)^T$ — transposed matrix or vector
Nonlinear case

Estimation problem: fit the state to observations with no time dependence and no prior information about state

- Let us have observations \( y \).
- Let \( H \) be the observation function, so that \( H(x) \) gives observations associated with \( x \).
- Let us characterise the uncertainty of \( y \) by covariance \( R \): if \( x \) is the true state, then \( y = H(x) + v \), and \( (R)_{ij} = \text{cov}[(v)_i, (v)_j] \).
- Estimation problem: find \( x^a \) such that

\[
x^a = \arg \min_x \left\{ [y - H(x)]^T (R)^{-1} [y - H(x)] \right\}.
\] (1)

- Statistical interpretation: if PDF of observation error is Gaussian,

\[
\mathcal{P} [y - H(x)] = \exp \left\{ -\frac{1}{2} [y - H(x)]^T (R)^{-1} [y - H(x)] \right\},
\]

then (1) gives the maximal likelihood solution.
Nonlinear case

Estimation problem: fit the state to observations with no time dependence and no prior information about state

- Let us have observations $y$.
- Let $\mathcal{H}$ be the observation function, so that $\mathcal{H}(x)$ gives observations associated with $x$.
- Let us characterise the uncertainty of $y$ by covariance $R$: if $x$ is the true state, then $y = \mathcal{H}(x) + v$, and $(R)_{ij} = \text{cov}[(v)_i, (v)_j]$.
- Estimation problem: find $x^a$ such that

$$
x^a = \arg \min_x \left\{ [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\}.
$$

(1)

- Statistical interpretation: if PDF of observation error is Gaussian,

$$
\mathcal{P} [y - \mathcal{H}(x)] = \exp \left\{ -\frac{1}{2} [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\},
$$

then (1) gives the maximal likelihood solution.
Nonlinear case

Estimation problem: fit the state to observations with no time dependence and no prior information about state

- Let us have observations $y$.
- Let $\mathcal{H}$ be the observation function, so that $\mathcal{H}(x)$ gives observations associated with $x$.
- Let us characterise the uncertainty of $y$ by covariance $R$: if $x$ is the true state, then $y = \mathcal{H}(x) + v$, and $(R)_{ij} = \text{cov}[(v)_i, (v)_j]$.
- Estimation problem: find $x^a$ such that

$$x^a = \arg\min_x \left\{ [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\}.$$  \hspace{1cm} (1)

- Statistical interpretation: if PDF of observation error is Gaussian,

$$\mathcal{P} [y - \mathcal{H}(x)] = \exp \left\{ -\frac{1}{2} [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\},$$

then (1) gives the maximal likelihood solution.
Nonlinear case

Estimation problem: fit the state to observations with no time dependence and no prior information about state

- Let us have observations $y$.
- Let $\mathcal{H}$ be the observation function, so that $\mathcal{H}(x)$ gives observations associated with $x$.
- Let us characterise the uncertainty of $y$ by covariance $R$: if $x$ is the true state, then $y = \mathcal{H}(x) + v$, and $(R)_{ij} = \text{cov}[(v)_i, (v)_j]$.
- Estimation problem: find $x^a$ such that

$$
x^a = \arg \min_x \left\{ [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\}. \tag{1}
$$

- Statistical interpretation: if PDF of observation error is Gaussian,

$$
\mathcal{P}[y - \mathcal{H}(x)] = \exp \left\{ -\frac{1}{2} [y - \mathcal{H}(x)]^T (R)^{-1} [y - \mathcal{H}(x)] \right\},
$$

then (1) gives the maximal likelihood solution.
Nonlinear case

Estimation problem: fit the state to observations with no time dependence and no prior information about state

- Let us have observations $y$.
- Let $H$ be the observation function, so that $H(x)$ gives observations associated with $x$.
- Let us characterise the uncertainty of $y$ by covariance $R$: if $x$ is the true state, then $y = H(x) + v$, and $(R)_{ij} = \text{cov}[(v)_i, (v)_j]$.
- Estimation problem: find $x^a$ such that

$$
x^a = \arg\min_x \left\{ [y - H(x)]^T (R)^{-1} [y - H(x)] \right\}.
$$

(1)

- Statistical interpretation: if PDF of observation error is Gaussian,

$$
P[y - H(x)] = \exp \left\{ -\frac{1}{2} [y - H(x)]^T (R)^{-1} [y - H(x)] \right\},
$$

then (1) gives the maximal likelihood solution.
prior information about the state

- Let us have a prior estimate of the state $x^f$ (“forecast”), with the uncertainty characterised by covariance $P^f$.
- Estimation problem: find $x^a$ such that

$$x^a = \arg \min_x \left\{ (x - x^f)^T (P^f)^{-1} (x - x^f) + [y - H(x)]^T (R)^{-1} [y - H(x)] \right\}.$$  \quad (2)

time dimension

- Let us have a sequence of observations $y_1, \ldots, y_k$ obtained at time $t_1, \ldots, t_k$, with error covariances $R_1, \ldots, R_k$.
- Estimation problem: find $\{x^a_i\}, i = 1, \ldots, k$ such that

$$\{x^a_i\}_{i=1}^k = \arg \min_{\{x_i\}_{i=1}^k} \left\{ (x_1 - x_1^f)^T (P_1^f)^{-1} (x_1 - x_1^f) + \sum_{i=1}^k [y_i - H_i(x_i)]^T (R_i)^{-1} [y_i - H_i(x_i)] \right\}.$$  \quad (3)
Let $x_{i+1} = M_i(x_i) + w_i$, with $w_i$ characterised by covariance $Q_i$: 
$(Q_i)_{lm} = \text{cov}[(w_i)_l, (w_i)_m]$.

The estimation problem becomes:

$$
\{x^a_i\}_{i=1}^k = \arg \min_{\{x_i\}_{i=1}^k} \mathcal{L}(x_1, \ldots, x_k),
$$

where

$$
\mathcal{L}(x_1, \ldots, x_k) = \left\{ (x_1 - x_1^f)^T (P_1^f)^{-1} (x_1 - x_1^f) 
\right.$$

$$
+ \sum_{i=1}^k [y_i - \mathcal{H}_i(x_i)]^T (R_i)^{-1} [y_i - \mathcal{H}_i(x_i)]
\left. \right\} + \sum_{i=2}^k [x_i - \mathcal{M}_{i-1}(x_{i-1})]^T (Q_{i-1})^{-1} [x_i - \mathcal{M}_{i-1}(x_{i-1})].
$$

(4)
Linear case

Standard assumptions of linearity:

\[ x_{i+1} \equiv M_i(x_i) = M_i x_i, \]
\[ H_i(x_i) = H_i x_i. \]

Sufficient assumptions – affine \( M \) and \( H \):

\[ M_i(x_i^2) = M_i(x_i^1) + M_i (x_i^2 - x_i^1), \]
\[ H_i(x_i^2) = H_i(x_i^1) + H_i (x_i^2 - x_i^1), \]
\[ M_i = \nabla_x M_i(x_i), \]
\[ H_i = \nabla_x H_i(x_i). \]
Linear case

Standard assumptions of linearity:

\[ x_{i+1} \equiv M_i(x_i) = M_i x_i, \]
\[ H_i(x_i) = H_i x_i. \]

Sufficient assumptions – affine \( M \) and \( H \):

\[ M_i(x_i^2) = M_i(x_i^1) + M_i(x_i^2 - x_i^1), \quad M_i = \nabla_x M_i(x_i), \]
\[ H_i(x_i^2) = H_i(x_i^1) + H_i(x_i^2 - x_i^1), \quad H_i = \nabla_x H_i(x_i). \]
Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{x_i\}_{i=1}^{k-1}} \mathcal{L}_k(x_1, \ldots, x_k) = (x_k - x_{k}^f)^T (P_k^f)^{-1} (x_k - x_{k}^f) + \text{Const.}$$

- Then

$$\min_{\{x_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(x_1, \ldots, x_k, x_{k+1}) = (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + \text{Const},$$

where

$$x_{k+1}^f = \mathcal{M}_k(x_{k}^a),$$

$$P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k,$$

$$x_{k}^a = x_{k}^f + K_k [y_k - \mathcal{H}(x_{k}^f)],$$

$$K_k = P_k^f (H_k)^T [H_k P_k^f (H_k)^T + R_k]^{-1},$$

$$P_{k}^a = (I - K_k H_k) P_{k}^f.$$
Kalman filter

Consider the minimisation problem for the linear case.

Let us assume that

$$\min_{\{x_i\}_{i=1}^{k-1}} L_k(x_1, \ldots, x_k) = (x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + \text{Const.}$$

Then

$$\min_{\{x_i\}_{i=1}^{k}} L_{k+1}(x_1, \ldots, x_k, x_{k+1}) = (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + \text{Const},$$

where

$$x_{k+1}^f = M_k(x_k^a),$$

$$P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k,$$

$$x_k^a = x_k^f + K_k[y_k - \mathcal{H}(x_k^f)],$$

$$K_k = P_k^f (H_k)^T [H_k P_k^f (H_k)^T + R_k]^{-1},$$

$$P_k^a = (I - K_k H_k) P_k^f.$$
Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

\[
\min_{\{x_i\}^k_{i=1}} L_k(x_1, \ldots, x_k) = (x_k - x_f^k)^T (P_k^f)^{-1} (x_k - x_f^k) + \text{Const.}
\]

- Then

\[
\min_{\{x_i\}^k_{i=1}} L_{k+1}(x_1, \ldots, x_k, x_{k+1}) = (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + \text{Const},
\]

where

\[
x_{k+1}^f = M_k(x_k^a),
\]

\[
P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k,
\]

\[
x_k^a = x_k^f + K_k[y_k - H(x_k^f)],
\]

\[
K_k = P_k^f (H_k)^T [H_k P_k^f (H_k)^T + R_k]^{-1}
\]

\[
P_k^a = (I - K_k H_k) P_k^f.
\]
Kalman filter

Consider the minimisation problem for the linear case.

Let us assume that
\[
\min \left\{ \text{\(x_i\)} \right\}_{i=1}^{k-1} L_k(x_1, \ldots, x_k) = (x_k - x_f^k)^T (P_k^f)^{-1} (x_k - x_f^k) + \text{Const}.
\]

Then
\[
\min \left\{ \text{\(x_i\)} \right\}_{i=1}^{k} L_{k+1}(x_1, \ldots, x_k, x_{k+1}) = (x_{k+1} - x_{f_{k+1}})^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{f_{k+1}}) + \text{Const},
\]

where
\[
x_{f_{k+1}} = M_k(x^a_k),
\]
\[
P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k,
\]
\[
x^a_k = x^f_k + K_k[y_k - \mathcal{H}(x^f_k)],
\]
\[
K_k = P_k^f (H_k)^T [H_k P_k^f (H_k)^T + R_k]^{-1}
\]
\[
P_k^a = (I - K_k H_k) P_k^f.
\]
Consider the minimisation problem for the linear case.

Let us assume that

$$\min_{\{x_i\}_{i=1}^{k-1}} \mathcal{L}_k(x_1, \ldots, x_k) = (x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + \text{Const}.$$ 

Then

$$\min_{\{x_i\}_{i=1}^k} \mathcal{L}_{k+1}(x_1, \ldots, x_k, x_{k+1}) = (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + \text{Const},$$

where

\[
\begin{align*}
  x_{k+1}^f &= \mathcal{M}_k(x_k^a), \\
  P_{k+1}^f &= \mathcal{M}_k P_k^a (\mathcal{M}_k)^T + Q_k, \\
  x_k^a &= x_k^f + K_k [y_k - \mathcal{H}(x_k^f)], \\
  K_k &= P_k^f (H_k)^T [H_k P_k^f (H_k)^T + R_k]^{-1}, \\
  P_k^a &= (I - K_k H_k) P_k^f.
\end{align*}
\]

"propagation"}

"analysis"
Kalman filter, proof (analysis part)

Regroup the terms:

\[(x_k - x^f_k)^T (P^f_k)^{-1} (x_k - x^f_k) + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_k(x_k)]\]  

\[\Rightarrow (x_k - x^a_k)^T (P^a_k)^{-1} (x_k - x^a_k) + C,\]  

and find \(x^a_k\) and \(P^a_k\).

\(\triangleright\) \((x_k)^2:\)

\[\begin{align*}
(P^a_k)^{-1} &= (P^f_k)^{-1} + (H_k)^T (R_k)^{-1} H_k.
\end{align*}\]  

\(\triangleright\) \((x_k)^1:\)

\[\begin{align*}
x^a_k &= x^f_k + P^a_k (H_k)^T (R_k)^{-1} [y_k - \mathcal{H}_k(x^f_k)].
\end{align*}\]
Kalman filter, proof (analysis part)

Regroup the terms:

\[
(x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_k(x_k)]
\]

\[
\Rightarrow (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) + C,
\]

and find \(x_k^a\) and \(P_k^a\).

\(\blacktriangleright\) \((x_k)^2\):

\[
(P_k^a)^{-1} = (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k.
\]

\(\blacktriangleright\) \((x_k)^1\):

\[
x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} [y_k - \mathcal{H}_k(x_k^f)].
\]
Kalman filter, proof (analysis part)

Regroup the terms:

\[(x_k - x_k^f)^T(P_k^f)^{-1}(x_k - x_k^f) + [y_k - \mathcal{H}_k(x_k)]^T(R_k)^{-1}[y_k - \mathcal{H}_k(x_k)]\]

\[\Rightarrow (x_k - x_k^a)^T(P_k^a)^{-1}(x_k - x_k^a) + C,\]  

and find \(x_k^a\) and \(P_k^a\).

\[\text{(x_k)^2:\quad (P_k^a)^{-1} = (P_k^f)^{-1} + (H_k)^T(R_k)^{-1}H_k.}\]  

\[\text{(x_k)^1:\quad (x_k)^T(P_k^f)^{-1}(-x_k^f) - (H_k x_k)^T(R_k)^{-1}[y_k - \mathcal{H}_k(x_k^f) + H_k x_k^f]}\]

\[= (x_k)^T(P_k^a)^{-1}(-x_k^a)\]

\[x_k^a = P_k^a \left\{ (P_k^f)^{-1}x_k^f + (H_k)^T(R_k)^{-1}[y_k - \mathcal{H}_k(x_k^f) + H_k x_k^f] \right\}\]

\[x_k^a = P_k^a \left\{ (P_k^a)^{-1}x_k^f - (H_k)^T(R_k)^{-1}H_k x_k^f\right.\]

\[\left. + (H_k)^T(R_k)^{-1}[y_k - \mathcal{H}_k(x_k^f) + H_k x_k^f] \right\}\]

\[x_k^a = x_k^f + P_k^a(H_k)^T(R_k)^{-1}[y_k - \mathcal{H}_k(x_k^f)].\]
Kalman filter, proof (analysis part)

Regroup the terms:

\[
(x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_k(x_k)]
\]  \(5\)

\[
\Rightarrow (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) + C,
\]  \(6\)

and find \(x_k^a\) and \(P_k^a\).

\(\Rightarrow (x_k)^2:\)

\[
(P_k^a)^{-1} = (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k.
\]  \(7\)

\(\Rightarrow (x_k)^1:\)

\[
x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} [y_k - \mathcal{H}_k(x_k^f)].
\]  \(8\)
Rearranging (7):
\[
(P_k^a)^{-1} = (P_k^f)^{-1} + (H_k)^T(R_k)^{-1}H_k \quad \Rightarrow
\]
\[
P_k^f = P_k^a + P_k^f(H_k)^T(R_k)^{-1}HP_k^a \quad \Rightarrow
\]
\[
P_k^a = \left[ I + P_k^f(H_k)^T(R_k)^{-1}H \right]^{-1}P_k^f.
\]
Matrix inversion lemma:
\[
(A + XBY)^{-1} = (A)^{-1} - (A)^{-1}X[(B)^{-1} + Y(A)^{-1}X]^{-1}Y(A)^{-1}
\]
set \( A = I, \quad X = P_k^f(H_k)^T, \quad B = (R_k)^{-1}, \quad Y = H_k \quad \Rightarrow
\]
\[
P_k^a = \left\{ I - P_k^f(H_k)^T[H_kP_k^f(H_k)^T + R_k]^{-1}H_k \right\}P_k^f,
\]
or
\[
P_k^a = (I - K_kH_k)P_k^f, \quad (9)
\]
\[
K_k \equiv P_k^f(H_k)^T[H_kP_k^f(H_k)^T + R_k]^{-1}. \quad (10)
\]
(8) \(\Rightarrow\)
\[
x_k^a = x_k^f + K_k \left[ y_k - H_k(x_k^f) \right]. \quad (11)
\]
We are aiming to show that
\[
\min_{x_k} \left\{ (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) + [x_{k+1} - M_k(x_k)]^T (Q_k)^{-1} [x_{k+1} - M_k(x_k)] \right\}
\]
\[
= (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + C,
\]
and find \( x_{k+1}^f \) and \( P_{k+1}^f \).

This problem is solved in two stages:
1. Find expression for the cost function at minimum over \( x_k \)
2. Factorise the expression and transform to the final form

Solution:
\[
x_{k+1}^f = M_k(x_k^a),
\]
\[
P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k.
\]
Kalman filter, proof (propagation part)

- We are aiming to show that

\[
\min_{x_k} \left\{ (x_k - x^a_k)^T (P_k^a)^{-1} (x_k - x^a_k) + [x_{k+1} - \mathcal{M}_k(x_k)]^T (Q_k)^{-1} [x_{k+1} - \mathcal{M}_k(x_k)] \right\} \\
= (x_{k+1} - x^f_{k+1})^T (P^f_{k+1})^{-1} (x_{k+1} - x^f_{k+1}) + C,
\]

and find \( x^f_{k+1} \) and \( P^f_{k+1} \).

- This problem is solved in two stages:
  1. Find expression for the cost function at minimum over \( x_k \)
  2. Factorise the expression and transform to the final form

- Solution:

\[
x^f_{k+1} = \mathcal{M}_k(x^a_k), \quad (12) \\
P^f_{k+1} = \mathcal{M}_k P^a_k (\mathcal{M}_k)^T + Q_k. \quad (13)
\]
Kalman filter, proof (propagation part)

- We are aiming to show that

\[
\min_{x_k} \left\{ (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) + [x_{k+1} - M_k(x_k)]^T (Q_k)^{-1} [x_{k+1} - M_k(x_k)] \right\}
\]

\[
= (x_{k+1} - x_{k+1}^f)^T (P_{k+1}^f)^{-1} (x_{k+1} - x_{k+1}^f) + C,
\]

and find \( x_{k+1}^f \) and \( P_{k+1}^f \).

- This problem is solved in two stages:
  1. Find expression for the cost function at minimum over \( x_k \)
  2. Factorise the expression and transform to the final form

- Solution:

\[
x_{k+1}^f = M_k(x_k^a),
\]

\[
P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k.
\]
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - H(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K[y - H(x^f)]$. 

![Diagram of time series with assimilation cycle]
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K [y - \mathcal{H}(x^f)]$. 

![Diagram](image-url)
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K[y - \mathcal{H}(x^f)]$. 

![Diagram showing the assimilation cycle and Kalman gain calculation](image)
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K [y - \mathcal{H}(x^f)]$. 

Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K[y - \mathcal{H}(x^f)]$. 

![Diagram showing time steps and assimilation cycle](image)
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast

- Quantity $X$ after propagation = forecast $X$
- Estimate of the state after analysis = analysis
- Quantity $X$ after analysis = analysed $X$ (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $x^a - x^f = K[y - \mathcal{H}(x^f)]$. 

[Diagram showing time steps and assimilation cycles]
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity \( X \) after propagation = forecast \( X \)
- Estimate of the state after analysis = analysis
- Quantity \( X \) after analysis = analysed \( X \) (or analysis \( X \))
- The difference between observations and observation estimate \( [y - \mathcal{H}(x)] \) = innovation
- The sensitivity of analysis to innovation = Kalman gain, \( x^a - x^f = K [y - \mathcal{H}(x^f)] \).

![Diagram showing time intervals and analysis cycles]
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity X after propagation = forecast X
- Estimate of the state after analysis = analysis
- Quantity X after analysis = analysed X (or analysis X)
- The difference between observations and observation estimate \([y - H(x)]\) = innovation
- The sensitivity of analysis to innovation = Kalman gain, \(x^a - x^f = K[y - H(x^f)]\).
**Terminology**

- Advancing in time from the previous analysis to the next analysis = **propagation** (or integration)
- Assimilating observations = **analysis** (or update)
- One cycle of propagation and analysis = **assimilation cycle**
- Time interval covered by an assimilation cycle = **assimilation window**
- Estimate of the state after propagation = **forecast**
- Quantity $X$ after propagation = **forecast $X$**
- Estimate of the state after analysis = **analysis**
- Quantity $X$ after analysis = **analysed $X$** (or analysis $X$)
- The difference between observations and observation estimate $[y - \mathcal{H}(x)] = **innovation**$
- The sensitivity of analysis to innovation = **Kalman gain**, $x^a - x^f = K [y - \mathcal{H}(x^f)]$. 

![Diagram](image.png)
Terminology

- Advancing in time from the previous analysis to the next analysis = propagation (or integration)
- Assimilating observations = analysis (or update)
- One cycle of propagation and analysis = assimilation cycle
- Time interval covered by an assimilation cycle = assimilation window
- Estimate of the state after propagation = forecast
- Quantity X after propagation = forecast X
- Estimate of the state after analysis = analysis
- Quantity X after analysis = analysed X (or analysis X)
- The difference between observations and observation estimate \([y - \mathcal{H}(x)]\) = innovation
- The sensitivity of analysis to innovation = Kalman gain, \(x^a - x^f = K [y - \mathcal{H}(x^f)]\).
Conclusions

- The KF provides a **recursive** solution of the least squares minimisation problem in the **linear** case.
- The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
- The state of the DA system at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
- The assimilation cycle breaks into two stages: **propagation** and **analysis**.
- The observation and propagation operators can be affine, so that
  \[
  H_i(x^1) - H_i(x^2) = H_i(x^1 - x^2),
  \]
  \[
  M_i(x^1) - M_i(x^2) = M_i(x^1 - x^2).
  \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed **state estimate** results from matching **first-order terms**, while the analysed **covariance estimate** is a result of matching **second-order terms**.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

▶ The KF provides a **recursive** solution of the least squares minisation problem in the **linear** case.
▶ The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
▶ The state of the DA system at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
▶ The assimilation cycle breaks into two stages: propagation and analysis.
▶ The observation and propagation operators can be affine, so that
  \[
  H_i(x^1) - H_i(x^2) = H_i(x^1 - x^2),
  \]
  \[
  M_i(x^1) - M_i(x^2) = M_i(x^1 - x^2).
  \]
▶ In the Gaussian case the KF solution coincides with the maximal likelihood solution.
▶ The analysed **state estimate** results from matching **first-order terms**, while the analysed **covariance estimate** is a result of matching **second-order terms**.
▶ The covariance accumulates the prior information
▶ The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
▶ Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a recursive solution of the least squares mininisation problem in the linear case.
- The KF provides optimal solution for the current state of the system given past observations. It does not provide optimal solution for all past states.
- The state of the DA system at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
- The assimilation cycle breaks into two stages: propagation and analysis.
- The observation and propagation operators can be affine, so that
  \[ H_i(x^1) - H_i(x^2) = H_i(x^1 - x^2), \]
  \[ M_i(x^1) - M_i(x^2) = M_i(x^1 - x^2). \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed state estimate results from matching first-order terms, while the analysed covariance estimate is a result of matching second-order terms.
- The covariance accumulates the prior information
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a **recursive** solution of the least squares minimisation problem in the **linear** case.
- The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
- The **state of the DA system** at any stage is given by (i) state estimate $\mathbf{x}$ and (ii) state error covariance estimate $\mathbf{P}$.
- The assimilation cycle breaks into two stages: **propagation** and **analysis**.
- The observation and propagation operators can be **affine**, so that
  \[
  \mathcal{H}_i(\mathbf{x}^1) - \mathcal{H}_i(\mathbf{x}^2) = \mathcal{H}_i (\mathbf{x}^1 - \mathbf{x}^2),
  \]
  \[
  \mathcal{M}_i(\mathbf{x}^1) - \mathcal{M}_i(\mathbf{x}^2) = \mathcal{M}_i (\mathbf{x}^1 - \mathbf{x}^2).
  \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed **state estimate** results from matching **first-order terms**, while the analysed **covariance estimate** is a result of matching **second-order terms**.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a **recursive** solution of the least squares minimisation problem in the **linear** case.
- The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
- The state of the DA system at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
- The assimilation cycle breaks into two stages: **propagation** and **analysis**.
- The observation and propagation operators can be **affine**, so that
  \[
  \mathcal{H}_i(x^1) - \mathcal{H}_i(x^2) = \mathbf{H}_i (x^1 - x^2),
  \]
  \[
  \mathcal{M}_i(x^1) - \mathcal{M}_i(x^2) = \mathbf{M}_i (x^1 - x^2).
  \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed state estimate results from matching **first-order terms**, while the analysed covariance estimate is a result of matching **second-order terms**.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

▶ The KF provides a recursive solution of the least squares minimisation problem in the linear case.
▶ The KF provides optimal solution for the current state of the system given past observations. It does not provide optimal solution for all past states.
▶ The state of the DA system at any stage is given by (i) state estimate $\mathbf{x}$ and (ii) state error covariance estimate $\mathbf{P}$.
▶ The assimilation cycle breaks into two stages: propagation and analysis.
▶ The observation and propagation operators can be affine, so that

$$\mathcal{H}_i(\mathbf{x}^1) - \mathcal{H}_i(\mathbf{x}^2) = \mathbf{H}_i (\mathbf{x}^1 - \mathbf{x}^2),$$
$$\mathcal{M}_i(\mathbf{x}^1) - \mathcal{M}_i(\mathbf{x}^2) = \mathbf{M}_i (\mathbf{x}^1 - \mathbf{x}^2).$$

▶ In the Gaussian case the KF solution coincides with the maximal likelihood solution.
▶ The analysed state estimate results from matching first-order terms, while the analysed covariance estimate is a result of matching second-order terms.
▶ The covariance accumulates the prior information.
▶ The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
▶ Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a recursive solution of the least squares minimisation problem in the linear case.
- The KF provides optimal solution for the current state of the system given past observations. It does not provide optimal solution for all past states.
- The state of the DA system at any stage is given by (i) state estimate $\mathbf{x}$ and (ii) state error covariance estimate $\mathbf{P}$.
- The assimilation cycle breaks into two stages: propagation and analysis.
- The observation and propagation operators can be affine, so that
  \[ \mathcal{H}_i(\mathbf{x}^1) - \mathcal{H}_i(\mathbf{x}^2) = \mathcal{H}_i(\mathbf{x}^1 - \mathbf{x}^2), \]
  \[ \mathcal{M}_i(\mathbf{x}^1) - \mathcal{M}_i(\mathbf{x}^2) = \mathcal{M}_i(\mathbf{x}^1 - \mathbf{x}^2). \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed state estimate results from matching first-order terms, while the analysed covariance estimate is a result of matching second-order terms.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a recursive solution of the least squares minimization problem in the linear case.
- The KF provides optimal solution for the current state of the system given past observations. It does not provide optimal solution for all past states.
- The state of the DA system at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
- The assimilation cycle breaks into two stages: propagation and analysis.
- The observation and propagation operators can be affine, so that
  \[ H_i(x^1) - H_i(x^2) = H_i(x^1 - x^2), \]
  \[ M_i(x^1) - M_i(x^2) = M_i(x^1 - x^2). \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed state estimate results from matching first-order terms, while the analysed covariance estimate is a result of matching second-order terms.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a **recursive** solution of the least squares minimisation problem in the **linear** case.
- The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
- The **state of the DA system** at any stage is given by (i) state estimate $\mathbf{x}$ and (ii) state error covariance estimate $\mathbf{P}$.
- The assimilation cycle breaks into two stages: **propagation** and **analysis**.
- The observation and propagation operators can be affine, so that
  \[
  \mathcal{H}_i(\mathbf{x}^1) - \mathcal{H}_i(\mathbf{x}^2) = \mathbf{H}_i (\mathbf{x}^1 - \mathbf{x}^2),
  \]
  \[
  \mathcal{M}_i(\mathbf{x}^1) - \mathcal{M}_i(\mathbf{x}^2) = \mathbf{M}_i (\mathbf{x}^1 - \mathbf{x}^2).
  \]
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed **state estimate** results from matching **first-order terms**, while the analysed **covariance estimate** is a result of matching **second-order terms**.
- The covariance accumulates the prior information
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- Straightforward implementation is not scalable towards the state size.
Conclusions

- The KF provides a **recursive** solution of the least squares minimisation problem in the **linear** case.
- The KF provides optimal solution for the **current** state of the system given past observations. It does not provide optimal solution for all past states.
- The **state of the DA system** at any stage is given by (i) state estimate $x$ and (ii) state error covariance estimate $P$.
- The assimilation cycle breaks into two stages: **propagation** and **analysis**.
- The observation and propagation operators can be **affine**, so that
  \[
  H_i(x^1) - H_i(x^2) = H_i(x^1 - x^2),
  \]
  \[
  M_i(x^1) - M_i(x^2) = M_i(x^1 - x^2).
  \]
  
- In the Gaussian case the KF solution coincides with the maximal likelihood solution.
- The analysed **state estimate** results from matching **first-order terms**, while the analysed **covariance estimate** is a result of matching **second-order terms**.
- The covariance accumulates the prior information.
- The minimum of the cost function does not depend on the linear transformation of state, so that linear scaling of the state vector elements does not affect the physical solution.
- **Straightforward implementation** is not scalable towards the state size.
Extended Kalman filter

- $\mathcal{H}(x), \mathcal{M}(x)$ are nonlinear
- EKF: uses the estimate

$$M_i = \nabla_x \mathcal{M}_i(x)^a, \quad H_i = \nabla_x \mathcal{H}_i(x)^f,$$

- Requires

$$|\nabla_x M_i(x + \delta x) - \nabla_x M_i(x)| \ll |\nabla_x M_i(x)|,$$
$$|\nabla_x H_i(x + \delta x) - \nabla_x H_i(x)| \ll |\nabla_x H_i(x)|.$$

- Therefore, for EKF to work the state must be “linearly” constrained - that is, constrained to a degree when linearised operators can be applied within the characteristic uncertainty range.
Extended Kalman filter

- $H(x), M(x)$ are nonlinear
- EKF: uses the estimate

$$
M_i = \nabla_x M_i(x^a),
$$
$$
H_i = \nabla_x H_i(x^f),
$$

- Requires

\[
|\nabla_x M_i(x + \delta x) - \nabla_x M_i(x)| \ll |\nabla_x M_i(x)|,
\]
\[
|\nabla_x H_i(x + \delta x) - \nabla_x H_i(x)| \ll |\nabla_x H_i(x)|.
\]

- Therefore, for EKF to work the state must be “linearly” constrained - that is, constrained to a degree when linearised operators can be applied within the characteristic uncertainty range.
Extended Kalman filter

- \( \mathcal{H}(x), \mathcal{M}(x) \) are nonlinear
- EKF: uses the estimate
  \[
  M_i = \nabla_x M_i(x_i^a),
  \]
  \[
  H_i = \nabla_x H_i(x_i^f),
  \]

- Requires
  \[
  |\nabla_x M_i(x + \delta x) - \nabla_x M_i(x)| \ll |\nabla_x M_i(x)|, \\
  |\nabla_x H_i(x + \delta x) - \nabla_x H_i(x)| \ll |\nabla_x H_i(x)|.
  \]

- Therefore, for EKF to work the state must be “linearly” constrained - that is, constrained to a degree when linearised operators can be applied within the characteristic uncertainty range.
Kalman smoother

Let us assume that
\[
\min_{\{x_i\}_{i \neq k+1}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K)
\]
\[
= (x_{k+1} - x_{k+1}^s)^T (P_{k+1}^s)^{-1} (x_{k+1} - x_{k+1}^s) + \text{Const.}
\]

Then
\[
\min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) = (x_k - x_k^s)^T (P_k^s)^{-1} (x_k - x_k^s) + \text{Const},
\]
where
\[
x_k^s = x_k^a + P_k^s (M_k)^T (Q_k)^{-1} [x_{k+1}^s - M_k (x_k^a)],
\]
\[
P_k^s = \left[ (P_k^a)^{-1} + (M_k)^T (Q_k)^{-1} M_k \right]^{-1}.
\]

Compare this with the KF analysis solution:
\[
x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} [x_{k+1}^f - H_k (x_k^f)],
\]
\[
P_k^a = \left[ (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k \right]^{-1}.
\]
Kalman smoother

- Let us assume that

$$\min_{\{x_i\}_{i \neq k+1}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K)$$

$$= (x_{k+1} - x_{k+1}^s)^T (P_{k+1}^s)^{-1} (x_{k+1} - x_{k+1}^s) + \text{Const.}$$

- Then

$$\min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) = (x_k - x_k^s)^T (P_k^s)^{-1} (x_k - x_k^s) + \text{Const},$$

where

$$x_k^s = x_k^a + P_k^s (M_k)^T (Q_k)^{-1} [x_{k+1}^s - M_k(x_k^a)],$$

$$P_k^s = \left[ (P_k^a)^{-1} + (M_k)^T (Q_k)^{-1} M_k \right]^{-1}.$$ 

- Compare this with the KF analysis solution:

$$x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} [x_{k+1}^f - H_k(x_k^f)],$$

$$P_k^a = \left[ (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k \right]^{-1}.$$
Kalman smoother

Let us assume that

\[ \min_{\{x_i\}_{i \neq k+1}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) \]

\[ = (x_{k+1} - x_{k+1}^s)^T (P_{k+1}^s)^{-1} (x_{k+1} - x_{k+1}^s) + \text{Const}. \]

Then

\[ \min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) = (x_k - x_k^s)^T (P_k^s)^{-1} (x_k - x_k^s) + \text{Const}, \]

where

\[ x_k^s = x_k^a + P_k^s (M_k)^T (Q_k)^{-1} [x_{k+1}^s - M_k(x_k^a)], \]

\[ P_k^s = \left[ (P_k^a)^{-1} + (M_k)^T (Q_k)^{-1} M_k \right]^{-1}. \]

Compare this with the KF analysis solution:

\[ x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} \left[ x_{k+1}^f - H_k(x_k^f) \right], \]

\[ P_k^a = \left[ (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k \right]^{-1}. \]
Kalman smoother

Let us assume that

\[
\min_{\{x_i\}_{i \neq k+1}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) = (x_{k+1} - x_{k+1}^s)^T (P_{k+1}^s)^{-1} (x_{k+1} - x_{k+1}^s) + \text{Const}.
\]

Then

\[
\min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, x_{k+1}, \ldots, x_K) = (x_k - x_k^s)^T (P_k^s)^{-1} (x_k - x_k^s) + \text{Const},
\]

where

\[
x_k^s = x_k^a + P_k^s (M_k)^T (Q_k)^{-1} [x_{k+1}^s - M_k(x_k^a)],
\]

\[
P_k^s = \left( (P_k^a)^{-1} + (M_k)^T (Q_k)^{-1} M_k \right)^{-1}.
\]

Compare this with the KF analysis solution:

\[
x_k^a = x_k^f + P_k^a (H_k)^T (R_k)^{-1} [x_{k+1}^f - H_k(x_k^f)],
\]

\[
P_k^a = \left( (P_k^f)^{-1} + (H_k)^T (R_k)^{-1} H_k \right)^{-1}.
\]
Proof

\[ \mathcal{L}_K(x_1, \ldots, x_k, \ldots, x_K) \]
\[ = \mathcal{L}_{k-1}(x_1, \ldots, x_{k-1}) + [M_{k-1}(x_{k-1}) - x_k]^T (Q_{k-1})^{-1} [M_{k-1}(x_{k-1}) - x_k] \]
\[ + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_i(x_k)] \]
\[ + [M_k(x_k) - x_{k+1}]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}] + \text{Terms}(x_{k+1}, \ldots, x_K). \]

\[ \min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, \ldots, x_K) \]
\[ = (x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_i(x_k)] \]
\[ + [M_k(x_k) - x_{k+1}^s]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}^s] + \text{Const} \]
\[ = (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) \]
\[ + [M_k(x_k) - x_{k+1}^s]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}^s] + \text{Const}. \] (14)

Compare (14) with (5) and (6). \( \blacksquare \).
Proof

\[ \mathcal{L}_K(x_1, \ldots, x_k, \ldots, x_K) \]
\[ = \mathcal{L}_{k-1}(x_1, \ldots, x_{k-1}) + [M_{k-1}(x_{k-1}) - x_k]^T (Q_{k-1})^{-1} [M_{k-1}(x_{k-1}) - x_k] \]
\[ + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_i(x_k)] \]
\[ + [M_k(x_k) - x_{k+1}]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}] + \text{Terms}(x_{k+1}, \ldots, x_K). \]

\[
\min_{\{x_i\}_{i \neq k}} \mathcal{L}_K(x_1, \ldots, x_k, \ldots, x_K) \\
= (x_k - x_k^f)^T (P_k^f)^{-1} (x_k - x_k^f) + [y_k - \mathcal{H}_k(x_k)]^T (R_k)^{-1} [y_k - \mathcal{H}_i(x_k)] \]
\[ + [M_k(x_k) - x_{k+1}^s]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}^s] + \text{Const} \]
\[ = (x_k - x_k^a)^T (P_k^a)^{-1} (x_k - x_k^a) \]
\[ + [M_k(x_k) - x_{k+1}^s]^T (Q_k)^{-1} [M_k(x_k) - x_{k+1}^s] + \text{Const.} \quad (14) \]

Compare (14) with (5) and (6). \[\blacksquare\].
Conclusions

- The KS provides a recursive solution of the least squares minimisation problem in the linear case for the states of the DA system at all times.
- These states can be obtained by recursion back in time, starting from the final state given by the KF.
- In the Gaussian case the KS solutions coincide with the maximal likelihood solutions.
Conclusions

- The KS provides a recursive solution of the least squares minimisation problem in the linear case for the states of the DA system at all times.
- These states can be obtained by recursion back in time, starting from the final state given by the KF.
- In the Gaussian case the KS solutions coincide with the maximal likelihood solutions.
Conclusions

- The KS provides a recursive solution of the least squares minimisation problem in the linear case for the states of the DA system at all times.
- These states can be obtained by recursion back in time, starting from the final state given by the KF.
- In the Gaussian case the KS solutions coincide with the maximal likelihood solutions.
Conclusions

- The KS provides a recursive solution of the least squares minimisation problem in the linear case for the states of the DA system at all times.
- These states can be obtained by recursion back in time, starting from the final state given by the KF.
- In the Gaussian case the KS solutions coincide with the maximal likelihood solutions.


Appendix 1: To the KF derivation

- Rearrange the cost function

\[ \mathcal{L}(x_k) \equiv (x_k - x_k^a)^T(P_k^a)^{-1} (\cdot) + [x_{k+1} - \mathcal{M}_k(x_k)]^T(Q_k)^{-1} [\cdot] \]

(use \( \mathcal{M}_k(x_k) = \mathcal{M}_k(x_k^a) + M_k(x_k - x_k^a) \))

\[ = (x_k - x_k^a)^T \left[ (P_k^a)^{-1} + (Q_k)^{-1}M_k \right] (\cdot) \]

\[ - 2 [x_{k+1} - \mathcal{M}_k(x_k^a)]^T (Q_k^{-1}) [M_k(x_k - x_k^a)] + [x_{k+1} - \mathcal{M}_k(x_k^a)]^T (Q_k)^{-1} [\cdot] \]

- Observation:

\[ \min_w (w^T A w - 2b^T w) = -b^T Ab \]

- Apply with

\[ A = (P_k^a)^{-1} + (Q_k)^{-1}M_k \]

\[ b = (M_k)^T(Q_k)^{-1} [x_{k+1} - \mathcal{M}_k(x_k^a)] \]

- Obtain:

\[ \min_{x_k} \mathcal{L}(x_k) = [x_{k+1} - \mathcal{M}_k(x_k^a)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1}M_k (A_k)^{-1} (M_k)^T (Q_k)^{-1} \right] [\cdot] \]

- Apply the Matrix Inversion Lemma

\[ (A - M_k)^{-1} = (A)^{-1} - (A)^{-1}M_k (A_k)^{-1} (M_k)^T (Q_k)^{-1} \]
Appendix 1: To the KF derivation

- Rearrange the cost function

\[ \mathcal{L}(x_k) \equiv (x_k - x^a_k)^T(P^a_k)^{-1}(\cdot) + [x_{k+1} - M_k(x_k)]^T(Q_k)^{-1} [\cdot] \]

( use \( M_k(x_k) = M_k(x^a_k) + M_k(x_k - x^a_k) \) )

\[ = (x_k - x^a_k)^T \left[ (P^a_k)^{-1} + (Q_k)^{-1}M_k \right] (\cdot) \]

\[ - 2 [x_{k+1} - M_k(x^a_k)]^T(Q_k^{-1})[M_k(x_k - x^a_k)] + [x_{k+1} - M_k(x^a_k)]^T(Q_k)^{-1} [\cdot] \]

- Observation:

\[ \min_w (w^T A w - 2b^T w) = -b^T A b \]

- Apply with

\[ A = (P^a_k)^{-1} + (Q_k)^{-1}M_k \]
\[ b = (M_k)^T(Q_k)^{-1} [x_{k+1} - M_k(x^a_k)] \]

- Obtain:

\[ \min_{x_k} \mathcal{L}(x_k) = [x_{k+1} - M_k(x^a_k)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1}M_k (A_k)^{-1}M_k)^T(Q_k)^{-1} \right] [\cdot] \]

- Apply the Matrix Inversion Lemma

\[ (A - XBM)T^{-1} = (A)^{-1} - (A)^{-1}X \left[ (B)^{-1} - M(B)^{-1}M \right]^{-1} M(B)^{-1} \]
Appendix 1: To the KF derivation

- Rearrange the cost function

\[
\mathcal{L}(x_k) \equiv (x_k - x^a_k)^T(P_k^a)^{-1}(\cdot) + [x_{k+1} - \mathcal{M}_k(x_k)]^T(Q_k)^{-1}[\cdot]
\]

( use \( \mathcal{M}_k(x_k) = \mathcal{M}_k(x^a_k) + M_k(x_k - x^a_k) \) )

\[
= (x_k - x^a_k)^T \left[ (P_k^a)^{-1} + (Q_k)^{-1}M_k \right] (\cdot) \\
- 2 [x_{k+1} - \mathcal{M}_k(x^a_k)]^T (Q_k)^{-1} [M_k(x_k - x^a_k)] + [x_{k+1} - \mathcal{M}_k(x^a_k)]^T (Q_k)^{-1} [\cdot]
\]

- Observation:

\[
\min_w (w^T A w - 2b^T w) = -b^T A b
\]

- Apply with

\[
A = (P_k^a)^{-1} + (Q_k)^{-1}M_k \\
b = (M_k)^T(Q_k)^{-1} [x_{k+1} - \mathcal{M}_k(x^a_k)]
\]

- Obtain:

\[
\min_{x_k} \mathcal{L}(x_k) = [x_{k+1} - \mathcal{M}_k(x^a_k)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1}M_k (A_k)^{-1}(M_k)^T(Q_k)^{-1} \right] [\cdot]
\]

- Apply the Matrix Inversion Lemma
Appendix 1: To the KF derivation

- Rearrange the cost function
\[ L(x_k) \equiv (x_k - x^a_k)^T (P^a_k)^{-1}(\cdot) + [x_{k+1} - \mathcal{M}_k(x_k)]^T (Q_k)^{-1} [\cdot] \]
( use \( \mathcal{M}_k(x_k) = \mathcal{M}_k(x^a_k) + M_k(x_k - x^a_k) \) )
\[ = (x_k - x^a_k)^T \left[ (P^a_k)^{-1} + (Q_k)^{-1} M_k \right] (\cdot) \]
\[ - 2 [x_{k+1} - \mathcal{M}_k(x^a_k)]^T (Q_k)^{-1} [M_k(x_k - x^a_k)] + [x_{k+1} - \mathcal{M}_k(x^a_k)]^T (Q_k)^{-1} [\cdot] \]

- Observation:
\[ \min_w (w^T A w - 2b^T w) = -b^T A b \]

- Apply with
\[ A = (P^a_k)^{-1} + (Q_k)^{-1} M_k \]
\[ b = (M_k)^T (Q_k)^{-1} [x_{k+1} - \mathcal{M}_k(x^a_k)] \]

- Obtain:
\[ \min_{x_k} L(x_k) = [x_{k+1} - \mathcal{M}_k(x^a_k)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1} M_k (A_k)^{-1} (M_k)^T (Q_k)^{-1} \right] [\cdot] \]

- Apply the Matrix Inversion Lemma
\[ (A + XBY)^{-1} - 1 = (A)^{-1} - (A)^{-1}X (B)^{-1} - 1 + (A)^{-1}Y (A)^{-1}X (B)^{-1} - 1 \]

\[ \circ \]
Appendix 1: To the KF derivation

- Rearrange the cost function

\[ \mathcal{L}(x_k) \equiv (x_k - x_k^a)^T (P_k^a)^{-1} (\cdot) + [x_{k+1} - \mathcal{M}_k(x_k)]^T (Q_k)^{-1} (\cdot) \]

( use \( \mathcal{M}_k(x_k) = \mathcal{M}_k(x_k^a) + M_k(x_k - x_k^a) \) )

\[ = (x_k - x_k^a)^T \left[ (P_k^a)^{-1} + (Q_k)^{-1} M_k \right] (\cdot) \]

\[ - 2 [x_{k+1} - \mathcal{M}_k(x_k^a)]^T (Q_k^{-1}) [M_k(x_k - x_k^a)] + [x_{k+1} - \mathcal{M}_k(x_k^a)]^T (Q_k)^{-1} (\cdot) \]

- Obtain:

\[ \min_{x_k} \mathcal{L}(x_k) = [x_{k+1} - \mathcal{M}_k(x_k^a)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1} M_k (A_k)^{-1} (M_k)^T (Q_k)^{-1} \right] (\cdot) \]

- Apply the Matrix Inversion Lemma

\[ (A + XBY)^{-1} = (A)^{-1} - (A)^{-1}X [(B)^{-1} + Y (A)^{-1}X]^{-1} Y (A)^{-1} \]

with

\[ A = Q_k, \quad X = M_k, \quad Y = P_k^a, \quad B = P_k^a \]

- Obtain:

\[ P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k, \quad x_{k+1}^f = \mathcal{M}_k(x_k^a) \]
Appendix 1: To the KF derivation

- Rearrange the cost function
  \[ L(x_k) \equiv (x_k - x^a_k)^T (P_k^a)^{-1} (\cdot) + [x_{k+1} - M_k(x_k)]^T (Q_k)^{-1} [\cdot] \]
  
  (use \( M_k(x_k) = M_k(x^a_k) + M_k(x_k - x^a_k) \))

  \[ = (x_k - x^a_k)^T \left[ (P_k^a)^{-1} + (Q_k)^{-1} M_k \right] (\cdot) \]

  \[ - 2 [x_{k+1} - M_k(x^a_k)]^T (Q_k)^{-1} [M_k(x_k - x^a_k)] + [x_{k+1} - M_k(x^a_k)]^T (Q_k)^{-1} [\cdot] \]

- Obtain:
  \[ \min_{x_k} L(x_k) = [x_{k+1} - M_k(x^a_k)]^T \left[ (Q_k)^{-1} - (Q_k)^{-1} M_k (A_k)^{-1} (M_k)^T (Q_k)^{-1} \right] [\cdot] \]

- Apply the Matrix Inversion Lemma
  \[ (A + XBY)^{-1} = (A)^{-1} - (A)^{-1} X \left[ (B)^{-1} + Y (A)^{-1} X \right]^{-1} Y (A)^{-1} \]

  with
  \[ A = Q_k, \quad X = M_k, \quad Y = P_k^a, \quad B = P_k^a \]

- Obtain:
  \[ P_{k+1}^f = M_k P_k^a (M_k)^T + Q_k, \quad x_{k+1}^f = M_k(x_k^a) \]
or Woodbury matrix identity, or Sherman-Morrison-Woodbury formula, or inverse of a small-rank adjustment ...

\[(A + XBY)^{-1} = (A)^{-1} - (A)^{-1}X \left[ (B)^{-1} + Y/(A)^{-1}X \right]^{-1} Y (A)^{-1}\]