Kalman filter

Pavel Sakov

Nansen Environmental and Remote Sensing Center, Norway

Outline

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 $\mathbf{X}_i, \mathbf{x}_2, x_n$	–	ith object in a sequence
upper, without brackets, as in $\mathbf{X}^a, \sigma^{obs}$	–	used to mark the object
after a bracketed object, as in$(\mathbf{X})_i, (\mathbf{X})^{\mathrm{T}}, (\mathbf{X})^2$	–	an operation

in particular:

$(\mathbf{X})_i$	–	ith column of $\mathbf{X}$
$(\mathbf{X})_{i,:}$	–	ith row of $\mathbf{X}$
$(\mathbf{X})_{ij}$	–	the element in ith row and jth column of $\mathbf{X}$
$(\mathbf{x})_i$	–	ith element of $\mathbf{x}$
$(\mathbf{X})^{\mathrm{T}}, (\mathbf{x})^{\mathrm{T}}$	–	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 $\mathbf{X}_i, \mathbf{x}_2, x_n$	–	ith object in a sequence
upper, without brackets, as in $\mathbf{X}^a, \sigma^{obs}$	–	used to mark the object
after a bracketed object, as in$(\mathbf{X})_i, (\mathbf{X})^{\mathrm{T}}, (\mathbf{X})^2$	–	an operation

in particular:

$(\mathbf{X})_i$	–	ith column of $\mathbf{X}$
$(\mathbf{X})_{i,:}$	–	ith row of $\mathbf{X}$
$(\mathbf{X})_{ij}$	–	the element in ith row and jth column of $\mathbf{X}$
$(\mathbf{x})_i$	–	ith element of $\mathbf{x}$
$(\mathbf{X})^{\mathrm{T}}, (\mathbf{x})^{\mathrm{T}}$	–	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 $\mathbf{y}$.
- Let $\mathcal{H}$ be the observation function, so that $\mathcal{H}(\mathbf{x})$ gives observations associated with $\mathbf{x}$.
- Let us characterise the uncertainty of $\mathbf{y}$ by covariance $\mathbf{R}$: if $\mathbf{x}$ is the true state, then $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \mathbf{v}$, and $(\mathbf{R})_{ij} = \mathrm{cov}[(\mathbf{v})_i, (\mathbf{v})_j]$.
- Estimation problem: find $\mathbf{x}^a$ such that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

  then (1) gives the maximal likelihood solution.

+ prior information about the state

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

$$\mathbf{x}^a = \arg\min_{\mathbf{x}} \Big\{ (\mathbf{x} - \mathbf{x}^f)^{\mathrm{T}} (\mathbf{P}^f)^{-1} (\mathbf{x} - \mathbf{x}^f) + [\mathbf{y} - \mathcal{H}(\mathbf{x})]^{\mathrm{T}} (\mathbf{R})^{-1} [\mathbf{y} - \mathcal{H}(\mathbf{x})] \Big\} . \tag{2}$$

+ time dimension

- Let us have a sequence of observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$ obtained at time $t_1, \ldots, t_k$, with error covariances $\mathbf{R}_1, \ldots, \mathbf{R}_k$.
- Estimation problem: find $\{\mathbf{x}_i^a\}$, $i = 1, \ldots, k$ such that

$$\{\mathbf{x}_i^a\}_{i=1}^k = \arg \min_{\{\mathbf{x}_i\}_{i=1}^k} \Big\{ (\mathbf{x}_1 - \mathbf{x}_1^f)^{\mathrm{T}} (\mathbf{P}_1^f)^{-1} (\mathbf{x}_1 - \mathbf{x}_1^f) + \sum_{i=1}^k [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)]^{\mathrm{T}} (\mathbf{R}_i)^{-1} [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)] \Big\} . \tag{3}$$

+ model error

- Let $\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i) + \mathbf{w}_i$, with $\mathbf{w}_i$ characterised by covariance $\mathbf{Q}_i$: $(\mathbf{Q}_i)_{lm} = \mathrm{cov}[(\mathbf{w}_i)_l, (\mathbf{w}_i)_m]$.
- The estimation problem becomes:

$$\{\mathbf{x}_i^a\}_{i=1}^k = \arg \min_{\{\mathbf{x}_i\}_{i=1}^k} \mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_k),$$

$$\begin{aligned}\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \Big\{ & (\mathbf{x}_1 - \mathbf{x}_1^f)^{\mathrm{T}} (\mathbf{P}_1^f)^{-1} (\mathbf{x}_1 - \mathbf{x}_1^f) \\ & + \sum_{i=1}^k [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)]^{\mathrm{T}} (\mathbf{R}_i)^{-1} [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)] \\ & + \sum_{i=2}^k [\mathbf{x}_i - \mathcal{M}_{i-1}(\mathbf{x}_{i-1})]^{\mathrm{T}} (\mathbf{Q}_{i-1})^{-1} [\mathbf{x}_i - \mathcal{M}_{i-1}(\mathbf{x}_{i-1})] \Big\}. \qquad (4)\end{aligned}$$

Linear case

Standard assumptions of linearity:

$$\mathbf{x}_{i+1} \equiv \mathcal{M}_i(\mathbf{x}_i) = \mathbf{M}_i \mathbf{x}_i,$$
$$\mathcal{H}_i(x_i) = \mathbf{H}_i \mathbf{x}_i.$$

Sufficient assumptions – affine $\mathcal{M}$ and $\mathcal{H}$:

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

Linear case

Standard assumptions of linearity:

$$\mathbf{x}_{i+1} \equiv \mathcal{M}_i(\mathbf{x}_i) = \mathbf{M}_i \mathbf{x}_i,$$
$$\mathcal{H}_i(x_i) = \mathbf{H}_i \mathbf{x}_i.$$

Sufficient assumptions – affine $\mathcal{M}$ and $\mathcal{H}$:

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

Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + \mathrm{Const},$$

where

$$\begin{aligned}
\mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a), \\
\mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k, \\
\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k [\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f)], \\
\mathbf{K}_k &= \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} [\mathbf{H}_k \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1} \\
\mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f.
\end{aligned}$$

Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + \mathrm{Const},$$

where

$$\mathbf{x}_{k+1}^f = \mathcal{M}_k(\mathbf{x}_k^a),$$
$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k,$$
$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k[\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f)],$$
$$\mathbf{K}_k = \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} [\mathbf{H}_k \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1}$$
$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f.$$

Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + \mathrm{Const},$$

where

$$\mathbf{x}_{k+1}^f = \mathcal{M}_k(\mathbf{x}_k^a),$$
$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k,$$
$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k[\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f)],$$
$$\mathbf{K}_k = \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} [\mathbf{H}_k \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1}$$
$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f.$$

Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + \mathrm{Const},$$

where

$$\begin{aligned}
\mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a), \\
\mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k, \\
\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k [\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f)], \\
\mathbf{K}_k &= \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} [\mathbf{H}_k \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1} \\
\mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f.
\end{aligned}$$

Kalman filter

- Consider the minimisation problem for the linear case.

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + \mathrm{Const},$$

where

$$\left.\begin{aligned} \mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a), \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k, \end{aligned}\right\} \quad \text{"propagation"}$$

$$\left.\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k[\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f)], \\ \mathbf{K}_k &= \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} [\mathbf{H}_k \mathbf{P}_k^f (\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k]^{-1} \\ \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f. \end{aligned}\right\} \quad \text{"analysis"}$$

Kalman filter, proof (analysis part)

Regroup the terms:

$$(\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}}(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f) + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] \tag{5}$$

$$\Rightarrow (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\mathbf{x}_k - \mathbf{x}_k^a) + C, \tag{6}$$

and find $\mathbf{x}_k^a$ and $\mathbf{P}_k^a$.

- $(\mathbf{x}_k)^2$:

$$\boxed{(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k}. \tag{7}$$

- $(\mathbf{x}_k)^1$:

$$\boxed{\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)\right]}. \tag{8}$$

Kalman filter, proof (analysis part)

Regroup the terms:

$$(\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}}(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f) + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] \tag{5}$$

$$\Rightarrow (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\mathbf{x}_k - \mathbf{x}_k^a) + C, \tag{6}$$

and find $\mathbf{x}_k^a$ and $\mathbf{P}_k^a$.

- $(\mathbf{x}_k)^2$:

$$\boxed{(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k}. \tag{7}$$

- $(\mathbf{x}_k)^1$:

$$\boxed{\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)\right]}. \tag{8}$$

Kalman filter, proof (analysis part)

Regroup the terms:

$$(\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}}(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f) + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] \tag{5}$$

$$\Rightarrow (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\mathbf{x}_k - \mathbf{x}_k^a) + C, \tag{6}$$

and find $\mathbf{x}_k^a$ and $\mathbf{P}_k^a$.

- $(\mathbf{x}_k)^2$:

$$\boxed{(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k}. \tag{7}$$

- $(\mathbf{x}_k)^1$:

$$\cancel{(\mathbf{x}_k)^{\mathrm{T}}}(\mathbf{P}_k^f)^{-1}(-\mathbf{x}_k^f) - (\mathbf{H}_k\cancel{\mathbf{x}_k})^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f) + \mathbf{H}_k\mathbf{x}_k^f\right]$$

$$= \cancel{(\mathbf{x}_k)^{\mathrm{T}}}(\mathbf{P}_k^a)^{-1}(-\mathbf{x}_k^a)$$

$$\mathbf{x}_k^a = \mathbf{P}_k^a\left\{(\mathbf{P}_k^f)^{-1}\mathbf{x}_k^f + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f) + \mathbf{H}_k\mathbf{x}_k^f\right]\right\}$$

$$\mathbf{x}_k^a = \mathbf{P}_k^a\left\{(\mathbf{P}_k^a)^{-1}\mathbf{x}_k^f - \cancel{(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k\mathbf{x}_k^f}\right.$$

$$\left. + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f) + \cancel{\mathbf{H}_k\mathbf{x}_k^f}\right]\right\}$$

$$\boxed{\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)\right]}. \tag{8}$$

Kalman filter, proof (analysis part)

Regroup the terms:

$$(\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}}(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f) + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] \tag{5}$$

$$\Rightarrow (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\mathbf{x}_k - \mathbf{x}_k^a) + C, \tag{6}$$

and find $\mathbf{x}_k^a$ and $\mathbf{P}_k^a$.

- $(\mathbf{x}_k)^2$:

$$\boxed{(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k}. \tag{7}$$

- $(\mathbf{x}_k)^1$:

$$\boxed{\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)\right]}. \tag{8}$$

Kalman filter, proof (analysis part, 2)

- Rearranging (7):

$$(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k \qquad \Rightarrow$$

$$\mathbf{P}_k^f = \mathbf{P}_k^a + \mathbf{P}^f(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}\mathbf{P}_k^a \qquad \Rightarrow$$

$$\mathbf{P}_k^a = \left[\mathbf{I} + \mathbf{P}^f(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}\right]^{-1}\mathbf{P}_k^f.$$

Matrix inversion lemma:

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

set $\quad \mathbf{A} = \mathbf{I}, \quad \mathbf{X} = \mathbf{P}_k^f(\mathbf{H}_k)^{\mathrm{T}}, \quad \mathbf{B} = (\mathbf{R}_k)^{-1}, \quad \mathbf{Y} = \mathbf{H}_k \quad \Rightarrow$

$$\mathbf{P}_k^a = \left\{\mathbf{I} - \mathbf{P}_k^f(\mathbf{H}_k)^{\mathrm{T}}\left[\mathbf{H}_k\mathbf{P}_k^f(\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k\right]^{-1}\mathbf{H}_k\right\}\mathbf{P}_k^f,$$

or

$$\boxed{\begin{aligned} \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f, && (9)\\ \mathbf{K}_k &\equiv \mathbf{P}_k^f(\mathbf{H}_k)^{\mathrm{T}}\left[\mathbf{H}_k\mathbf{P}_k^f(\mathbf{H}_k)^{\mathrm{T}} + \mathbf{R}_k\right]^{-1}. && (10)\end{aligned}}$$

(8) $\Rightarrow$

$$\boxed{\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)\right].} \qquad (11)$$

Kalman filter, proof (propagation part)

- We are aiming to show that

$$\min_{\mathbf{x}_k} \left\{ (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)] \right\}$$
$$= (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + C,$$

  and find $\mathbf{x}_{k+1}^f$ and $\mathbf{P}_{k+1}^f$.

- This problem is solved in two stages:
  1. Find expression for the cost function at minimum over $\mathbf{x}_k$
  2. Factorise the expression and transform to the final form

- Solution:

$$\boxed{\begin{aligned} \mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a), && (12) \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k. && (13) \end{aligned}}$$

Kalman filter, proof (propagation part)

- We are aiming to show that

$$\min_{\mathbf{x}_k} \left\{ (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)] \right\}$$
$$= (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + C,$$

  and find $\mathbf{x}_{k+1}^f$ and $\mathbf{P}_{k+1}^f$.

- This problem is solved in two stages:
  1. Find expression for the cost function at minimum over $\mathbf{x}_k$
  2. Factorise the expression and transform to the final form

- Solution:

$$\mathbf{x}_{k+1}^f = \mathcal{M}_k(\mathbf{x}_k^a), \tag{12}$$
$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k. \tag{13}$$

Kalman filter, proof (propagation part)

- We are aiming to show that

$$\min_{\mathbf{x}_k} \left\{ (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)] \right\}$$
$$= (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f)^{\mathrm{T}} (\mathbf{P}_{k+1}^f)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^f) + C,$$

  and find $\mathbf{x}_{k+1}^f$ and $\mathbf{P}_{k+1}^f$.

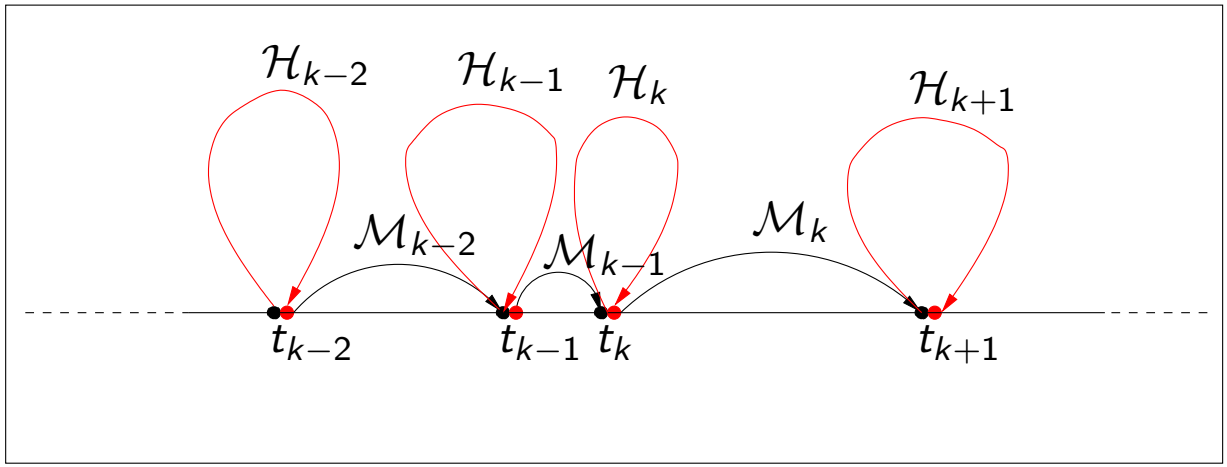
- This problem is solved in two stages:
  1. Find expression for the cost function at minimum over $\mathbf{x}_k$
  2. Factorise the expression and transform to the final form

- Solution:

$$\boxed{\begin{aligned} \mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a), && (12) \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k. && (13) \end{aligned}}$$

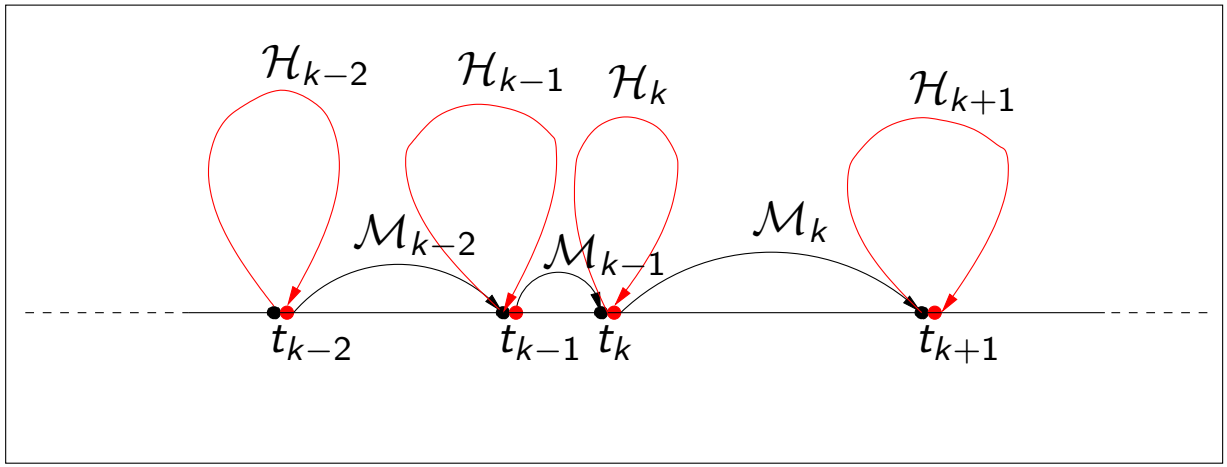
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



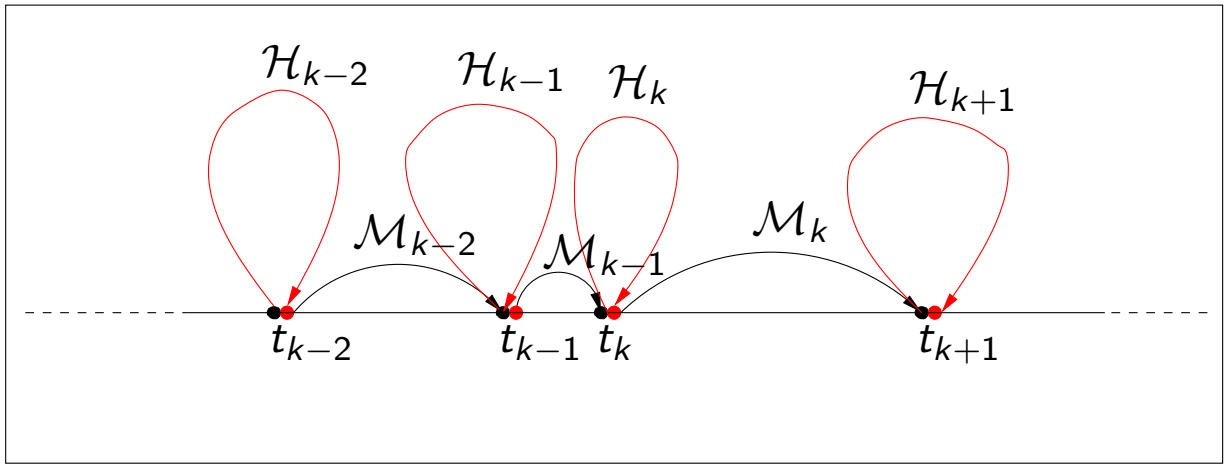
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



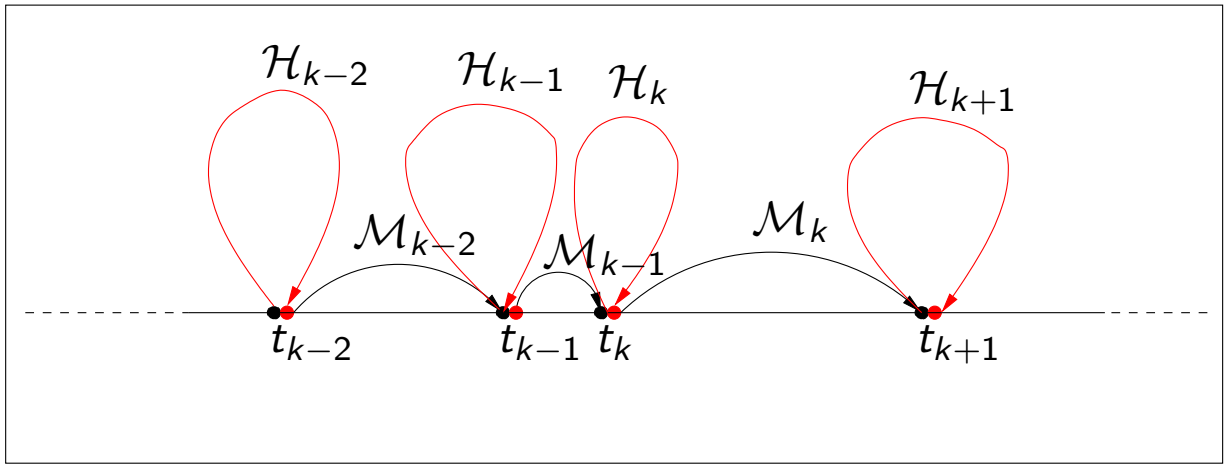
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



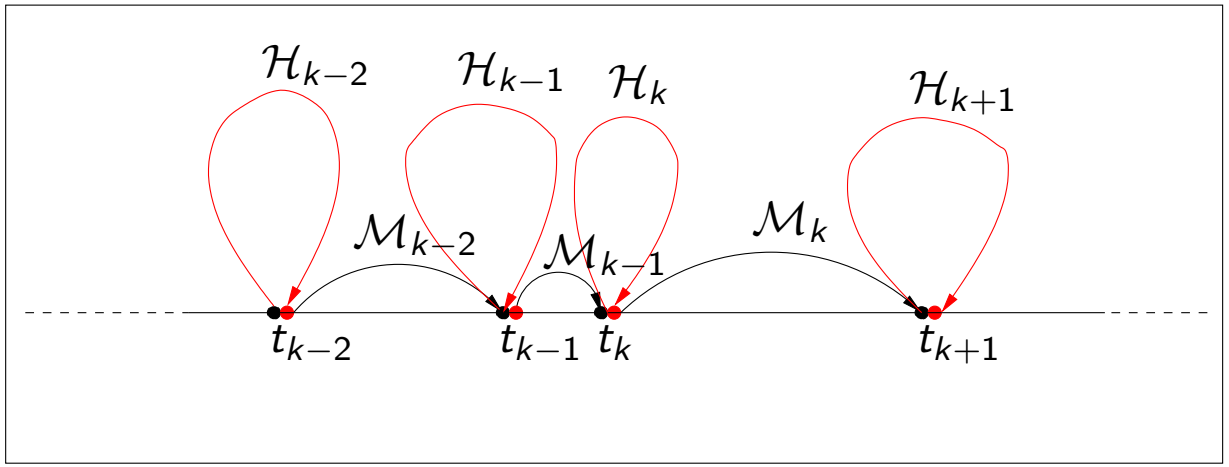
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



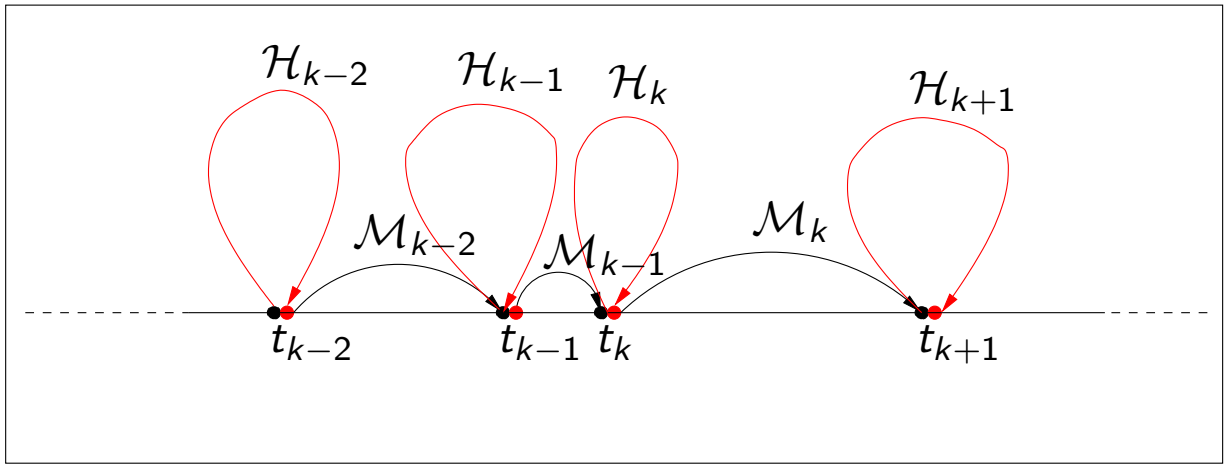
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



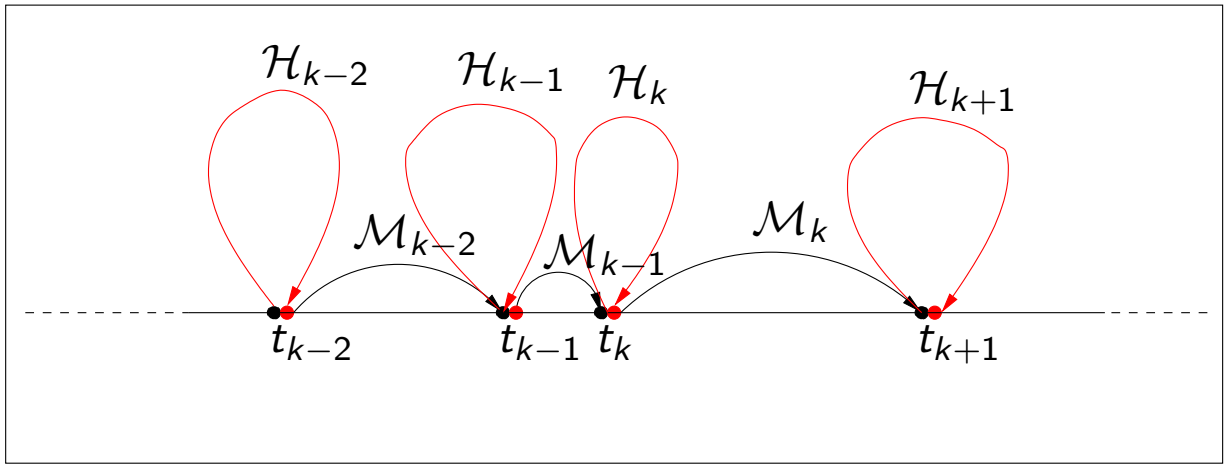
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



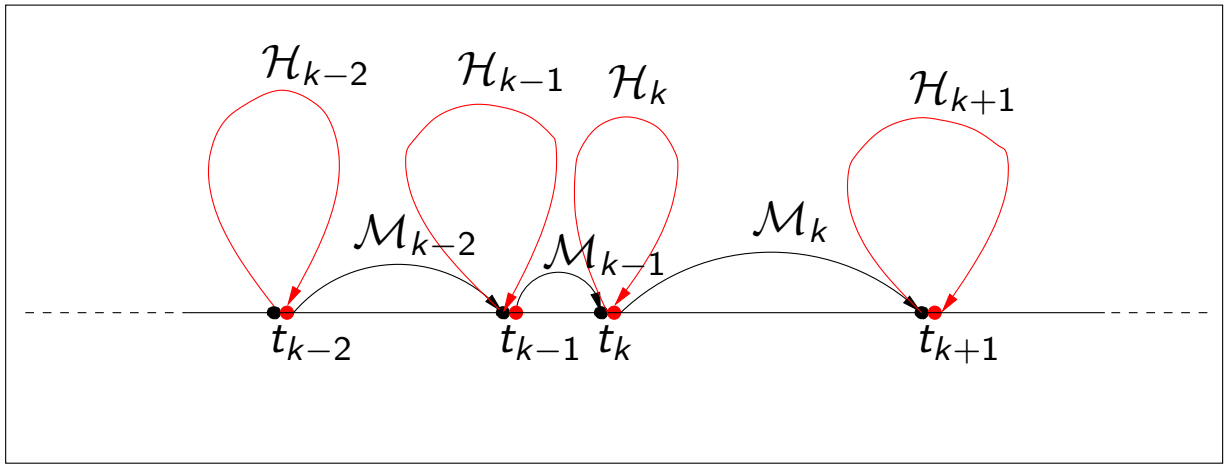
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



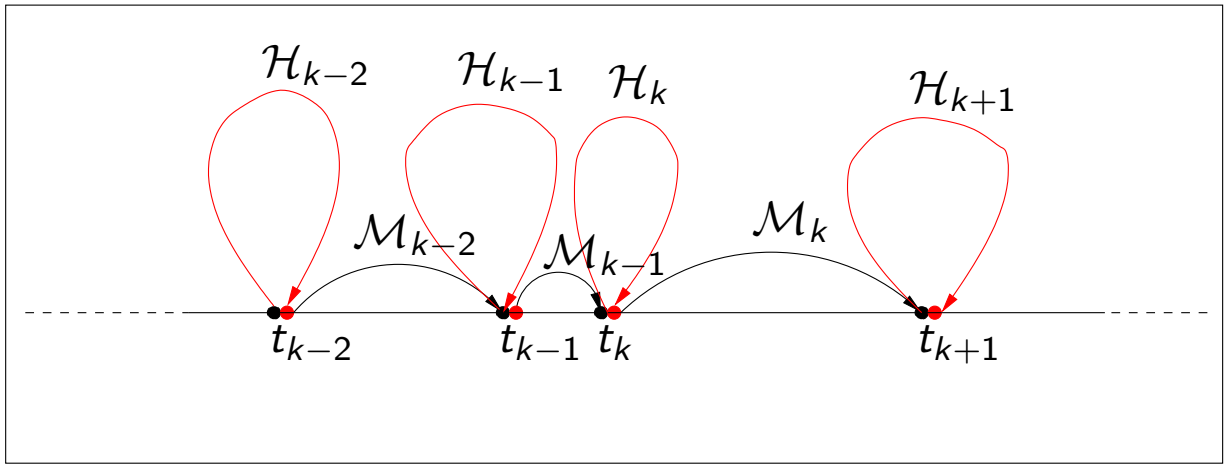
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



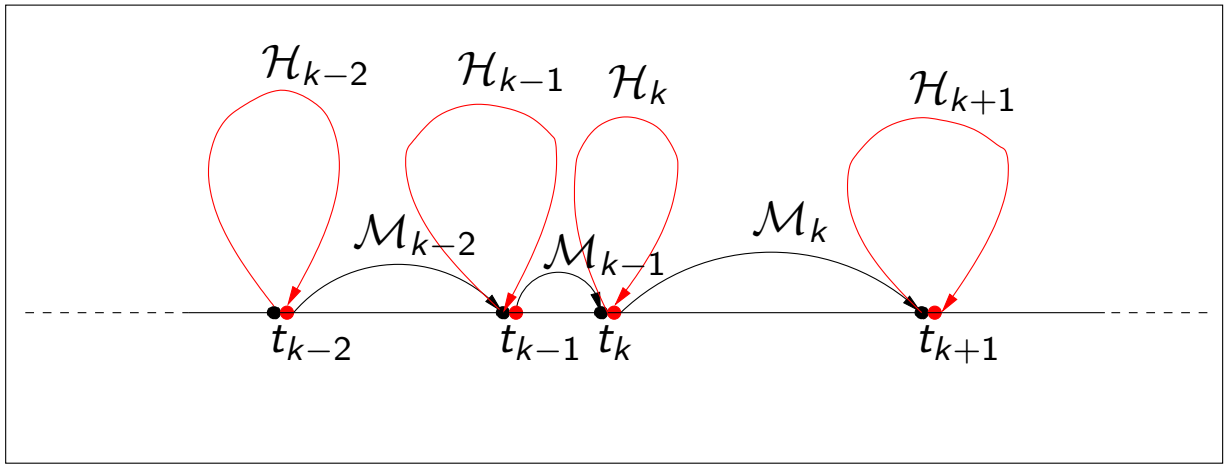
Terminology

- 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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



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 $[\mathbf{y} - \mathcal{H}(\mathbf{x})]$ = innovation
- The sensitivity of analysis to innovation = Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.



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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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 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 $\mathbf{x}$ 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.

Extended Kalman filter

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

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

- Requires

$$|\nabla_x \mathcal{M}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{M}_i(\mathbf{x})| \ll |\nabla_x \mathcal{M}_i(\mathbf{x})|,$$
$$|\nabla_x \mathcal{H}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{H}_i(\mathbf{x})| \ll |\nabla_x \mathcal{H}_i(\mathbf{x})|.$$

- 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

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

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

- Requires

$$|\nabla_x \mathcal{M}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{M}_i(\mathbf{x})| \ll |\nabla_x \mathcal{M}_i(\mathbf{x})|,$$
$$|\nabla_x \mathcal{H}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{H}_i(\mathbf{x})| \ll |\nabla_x \mathcal{H}_i(\mathbf{x})|.$$

- 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

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

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

- Requires

$$|\nabla_x \mathcal{M}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{M}_i(\mathbf{x})| \ll |\nabla_x \mathcal{M}_i(\mathbf{x})|,$$
$$|\nabla_x \mathcal{H}_i(\mathbf{x} + \delta\mathbf{x}) - \nabla_x \mathcal{H}_i(\mathbf{x})| \ll |\nabla_x \mathcal{H}_i(\mathbf{x})|.$$

- 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

$$\min_{\{\mathbf{x}_i\}_{i\neq k+1}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K)$$
$$= (\mathbf{x}_{k+1}-\mathbf{x}^s_{k+1})^{\mathrm{T}}(\mathbf{P}^s_{k+1})^{-1}(\mathbf{x}_{k+1}-\mathbf{x}^s_{k+1}) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i\neq k}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K) = (\mathbf{x}_k-\mathbf{x}^s_k)^{\mathrm{T}}(\mathbf{P}^s_k)^{-1}(\mathbf{x}_k-\mathbf{x}^s_k) + \mathrm{Const},$$

where

$$\mathbf{x}^s_k = \mathbf{x}^a_k + \mathbf{P}^s_k(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\left[\mathbf{x}^s_{k+1} - \mathcal{M}_k(\mathbf{x}^a_k)\right],$$
$$\mathbf{P}^s_k = \left[(\mathbf{P}^a_k)^{-1} + (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\mathbf{M}_k\right]^{-1}.$$

- Compare this with the KF analysis solution:

$$\mathbf{x}^a_k = \mathbf{x}^f_k + \mathbf{P}^a_k(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{x}^f_{k+1} - \mathcal{H}_k(\mathbf{x}^f_k)\right],$$
$$\mathbf{P}^a_k = \left[(\mathbf{P}^f_k)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k\right]^{-1}.$$

Kalman smoother

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i\neq k+1}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K)$$
$$= (\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s)^{\mathrm{T}}(\mathbf{P}_{k+1}^s)^{-1}(\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s)+\mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i\neq k}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K) = (\mathbf{x}_k-\mathbf{x}_k^s)^{\mathrm{T}}(\mathbf{P}_k^s)^{-1}(\mathbf{x}_k-\mathbf{x}_k^s)+\mathrm{Const},$$

where

$$\mathbf{x}_k^s = \mathbf{x}_k^a + \mathbf{P}_k^s(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\left[\mathbf{x}_{k+1}^s - \mathcal{M}_k(\mathbf{x}_k^a)\right],$$
$$\mathbf{P}_k^s = \left[(\mathbf{P}_k^a)^{-1} + (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\mathbf{M}_k\right]^{-1}.$$

- Compare this with the KF analysis solution:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{x}_{k+1}^f - \mathcal{H}_k(\mathbf{x}_k^f)\right],$$
$$\mathbf{P}_k^a = \left[(\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k\right]^{-1}.$$

Kalman smoother

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i\neq k+1}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K)$$
$$= (\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s)^{\mathrm{T}}(\mathbf{P}_{k+1}^s)^{-1}(\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s)+\mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i\neq k}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K) = (\mathbf{x}_k-\mathbf{x}_k^s)^{\mathrm{T}}(\mathbf{P}_k^s)^{-1}(\mathbf{x}_k-\mathbf{x}_k^s)+\mathrm{Const},$$

where

$$\mathbf{x}_k^s = \mathbf{x}_k^a + \mathbf{P}_k^s(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\left[\mathbf{x}_{k+1}^s - \mathcal{M}_k(\mathbf{x}_k^a)\right],$$
$$\mathbf{P}_k^s = \left[(\mathbf{P}_k^a)^{-1} + (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\mathbf{M}_k\right]^{-1}.$$

- Compare this with the KF analysis solution:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{x}_{k+1}^f - \mathcal{H}_k(\mathbf{x}_k^f)\right],$$
$$\mathbf{P}_k^a = \left[(\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k\right]^{-1}.$$

Kalman smoother

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i\neq k+1}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K) = (\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s)^{\mathrm{T}}(\mathbf{P}_{k+1}^s)^{-1}(\mathbf{x}_{k+1}-\mathbf{x}_{k+1}^s) + \mathrm{Const.}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i\neq k}} \mathcal{L}_K(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{x}_{k+1},\ldots,\mathbf{x}_K) = (\mathbf{x}_k-\mathbf{x}_k^s)^{\mathrm{T}}(\mathbf{P}_k^s)^{-1}(\mathbf{x}_k-\mathbf{x}_k^s) + \mathrm{Const},$$

where

$$\mathbf{x}_k^s = \mathbf{x}_k^a + \mathbf{P}_k^s(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\left[\mathbf{x}_{k+1}^s - \mathcal{M}_k(\mathbf{x}_k^a)\right],$$

$$\mathbf{P}_k^s = \left[(\mathbf{P}_k^a)^{-1} + (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\mathbf{M}_k\right]^{-1}.$$

- Compare this with the KF analysis solution:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{P}_k^a(\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\left[\mathbf{x}_{k+1}^f - \mathcal{H}_k(\mathbf{x}_k^f)\right],$$

$$\mathbf{P}_k^a = \left[(\mathbf{P}_k^f)^{-1} + (\mathbf{H}_k)^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k\right]^{-1}.$$

Proof

$$
\begin{aligned}
&\mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K) \\
&\quad = \mathcal{L}_{k-1}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) + [\mathcal{M}_{k-1}(\mathbf{x}_{k-1}) - \mathbf{x}_k]^{\mathrm{T}} (\mathbf{Q}_{k-1})^{-1} [\mathcal{M}_{k-1}(\mathbf{x}_{k-1}) - \mathbf{x}_k] \\
&\qquad + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}} (\mathbf{R}_k)^{-1} [\mathbf{y}_k - \mathcal{H}_i(\mathbf{x}_k)] \\
&\qquad + [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}] + \mathrm{Terms}(\mathbf{x}_{k+1}, \ldots, \mathbf{x}_K).
\end{aligned}
$$

$$
\begin{aligned}
&\min_{\{\mathbf{x}_i\}_{i \neq k}} \mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K) \\
&\quad = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]^{\mathrm{T}} (\mathbf{R}_k)^{-1} [\mathbf{y}_k - \mathcal{H}_i(\mathbf{x}_k)] \\
&\qquad + [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s] + \mathrm{Const} \\
&\quad = (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) \\
&\qquad + [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} [\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s] + \mathrm{Const}. \qquad (14)
\end{aligned}
$$

Compare (14) with (5) and (6). ■.

Proof

$$
\begin{aligned}
&\mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K) \\
&\quad = \mathcal{L}_{k-1}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) + \left[\mathcal{M}_{k-1}(\mathbf{x}_{k-1}) - \mathbf{x}_k\right]^{\mathrm{T}} (\mathbf{Q}_{k-1})^{-1} \left[\mathcal{M}_{k-1}(\mathbf{x}_{k-1}) - \mathbf{x}_k\right] \\
&\qquad + \left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right]^{\mathrm{T}} (\mathbf{R}_k)^{-1} \left[\mathbf{y}_k - \mathcal{H}_i(\mathbf{x}_k)\right] \\
&\qquad + \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}\right]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}\right] + \mathrm{Terms}(\mathbf{x}_{k+1}, \ldots, \mathbf{x}_K).
\end{aligned}
$$

$$
\begin{aligned}
&\min_{\{\mathbf{x}_i\}_{i \neq k}} \mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K) \\
&\quad = (\mathbf{x}_k - \mathbf{x}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) + \left[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right]^{\mathrm{T}} (\mathbf{R}_k)^{-1} \left[\mathbf{y}_k - \mathcal{H}_i(\mathbf{x}_k)\right] \\
&\qquad + \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s\right]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s\right] + \mathrm{Const} \\
&\quad = (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) \\
&\qquad + \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s\right]^{\mathrm{T}} (\mathbf{Q}_k)^{-1} \left[\mathcal{M}_k(\mathbf{x}_k) - \mathbf{x}_{k+1}^s\right] + \mathrm{Const}. \qquad (14)
\end{aligned}
$$

Compare (14) with (5) and (6). ■.

Conclusions

- The KS provides a recursive solution of the least squares mininisation 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 mininisation 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 mininisation 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 mininisation 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.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165–1168.

Bell, B. M., 1994: The iterated Kalman smoother as a Gauss-Newton method. *SIAM J. Optim.*, **4**, 626–636.

Bierman, G. J., 1977: *Factorization Methods for Discrete Sequential Estimation*. Academic Press, 241 pp.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. part I: theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Gu, Y. and D. S. Oliver, 2007: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, **12**, 438–446, doi:10.2118/108438-PA.

Jazwinski, A. H., 1970: *Stochastic processes and filtering theory*. Academic Press, 376 pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic. Eng.*, **82**, 35–45.

Kaminski, P., J. Bryson, A., and S. Schmidt, 1971: Discrete square root filtering: A survey of current techniques. *IEEE T. Automat. Contr.*, **16**, 727–736.

Pham, D. T., 2001: Stochastic methods for sequential data assmilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.

Smith, G. L., S. F. Schmidt, and L. A. McGee, 1962: Application of statistical filtering to the optimal estimation of position and velocity on-board a circumlunar vehicle. Technical report.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Zupanski, M., 2005: Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Observation:

$$\min_{\mathbf{w}} (\mathbf{w}^{\mathrm{T}} A\mathbf{w} - 2\mathbf{b}^{\mathrm{T}}\mathbf{w}) = -\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b}$$

- Apply with

$$\mathbf{A} = (\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k$$

$$\mathbf{b} = (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]$$

- Obtain:

$$\min_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Observation:

$$\min_{\mathbf{w}}\,(\mathbf{w}^{\mathrm{T}}A\mathbf{w} - 2\mathbf{b}^{\mathrm{T}}\mathbf{w}) = -\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b}$$

- Apply with

$$\mathbf{A} = (\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k$$

$$\mathbf{b} = (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]$$

- Obtain:

$$\min_{\mathbf{x}_k}\,\mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Observation:

$$\min_{\mathbf{w}}(\mathbf{w}^{\mathrm{T}}A\mathbf{w} - 2\mathbf{b}^{\mathrm{T}}\mathbf{w}) = -\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b}$$

- Apply with

$$\mathbf{A} = (\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k$$

$$\mathbf{b} = (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]$$

- Obtain:

$$\min_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Observation:

$$\min_{\mathbf{w}} (\mathbf{w}^{\mathrm{T}} A\mathbf{w} - 2\mathbf{b}^{\mathrm{T}}\mathbf{w}) = -\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b}$$

- Apply with

$$\mathbf{A} = (\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k$$

$$\mathbf{b} = (\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]$$

- Obtain:

$$\min_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Obtain:

$$\min_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

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

with

$$\mathbf{A} = \mathbf{Q}_k, \quad \mathbf{X} = \mathbf{M}_k, \quad \mathbf{Y} = \mathbf{P}_k^a, \quad \mathbf{B} = \mathbf{P}_k^a$$

- Obtain:

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k\mathbf{P}_k^a(\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k, \quad \mathbf{x}_{k+1}^f = \mathcal{M}_k(\mathbf{x}_k^a)$$

Appendix 1: To the KF derivation

- Rearrange the cost function

$$\mathcal{L}(\mathbf{x}_k) \equiv (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}(\mathbf{P}_k^a)^{-1}(\cdot) + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

$$\left(\quad \text{use} \quad \mathcal{M}_k(\mathbf{x}_k) = \mathcal{M}_k(\mathbf{x}_k^a) + \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a) \quad \right)$$

$$= (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}}\left[(\mathbf{P}_k^a)^{-1} + (Q_k)^{-1}\mathbf{M}_k\right](\cdot)$$

$$- 2\,[\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k^{-1})\,[\mathbf{M}_k(\mathbf{x}_k - \mathbf{x}_k^a)] + [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}(\mathbf{Q}_k)^{-1}[\cdot]$$

- Obtain:

$$\min_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}_k) = [\mathbf{x}_{k+1} - \mathcal{M}_k(\mathbf{x}_k^a)]^{\mathrm{T}}\left[(\mathbf{Q}_k)^{-1} - (\mathbf{Q}_k)^{-1}\mathbf{M}_k(\mathbf{A}_k)^{-1}(\mathbf{M}_k)^{\mathrm{T}}(\mathbf{Q}_k)^{-1}\right][\cdot]$$

- Apply the Matrix Inversion Lemma

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

with

$$\mathbf{A} = \mathbf{Q}_k, \quad \mathbf{X} = \mathbf{M}_k, \quad \mathbf{Y} = \mathbf{P}_k^a, \quad \mathbf{B} = \mathbf{P}_k^a$$

- Obtain:

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k\mathbf{P}_k^a(\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k, \quad \mathbf{x}_{k+1}^f = \mathcal{M}_k(\mathbf{x}_k^a)$$

Matrix inversion lemma

or Woodbury matrix identity, or Sherman-Morrison-Woodbury formula, or inverse of a small-rank adjustment ...

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