Jet Propulsion Laboratory
California Institute of Technology

State Estimation

Part 2: Methods of State Estimation

Ichiro Fukumori
Jet Propulsion Laboratory, Caltech

ECCO Summer School 2019

Main Points from Lecture 1

1) State estimation (data assimilation) is mathematically an inverse problem,

2) Inverse problems with data are invariably ill-posed and do not have a solution in a strict mathematical sense,

observations

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \mathbf{H} & & & & & \cdots \\ \cdots & & \mathbf{H} & & & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\ \cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

model equations

3) Particular (optimal, best) solutions to inverse problems can be obtained by imposing certain criteria that define what is optimal (best) (e.g., Minimum length, Minimum variance, Least-squares),

4) Minimum variance and least-squares are equivalent,

5) State estimation problems are too large to directly apply basic inverse methods (e.g., SVD);

10-years of Argo T : 3000-flts × 50-depths × 36 × 10=54e6
10-years of Jason : 6700s × 127-rev × 2/3 × 36 × 10=204e6
10-years of ECCO V4 3d UVTS : 2.4e6 gpts × 4 × 36 × 10=3e9

Outline

1. Basic Machinery (yesterday)
   The mathematical problem (inverse problem), Linear inverse methods, Singular value decomposition (SVD), Rank deficiency, Gauss-Markov theorem, Minimum variance estimate, Least-squares,

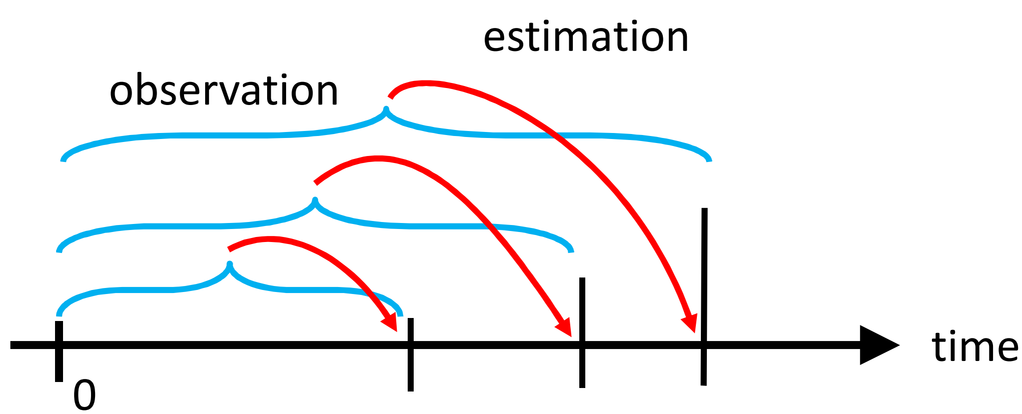
2. Methods of state estimation (this lecture)
   Kalman filter, Rauch-Tung-Striebel smoother, Adjoint method,

3. Practical Matters (Saturday)
   Error estimation, representation error, covariance, approximate Kalman filters, other data assimilation methods (Optimal Interpolation, 3DVAR).

Kalman Filter

Kalman filter is a
minimum variance estimator
that inverts observations
recursively in time.

A ***filter***, in estimation theory, is an estimator that employs observations that are formally in the past.



Derivation of Kalman Filter

Observations $\mathbf{H}(t)\mathbf{x}(t) \approx \hat{\mathbf{y}}(t)$

Model $\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$

Suppose we have an optimal (*minimum variance*) estimate at time $t-1$, $\hat{\mathbf{x}}(t-1)$, that uses all data up to that time with error covariance $\mathbf{P}(t-1)$. How do we get an optimal estimate at the next instance t using $\hat{\mathbf{y}}(t)$?

"–" indicating estimate without use of data at that instant

Model $\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$

Its error $\delta\hat{\mathbf{x}}(t,-) = \mathbf{A}\delta\hat{\mathbf{x}}(t-1) + \mathbf{G}\delta\hat{\mathbf{u}}(t-1)$

process noise/ control error (includes errors in the model equation)

Error covariance $\mathbf{P}(t,-) = \left\langle \delta\hat{\mathbf{x}}(t,-)\delta\hat{\mathbf{x}}(t,-)^T \right\rangle$

$$= \mathbf{A}\left\langle \delta\hat{\mathbf{x}}(t-1)\delta\hat{\mathbf{x}}(t-1)^T \right\rangle \mathbf{A}^T + \mathbf{G}\left\langle \delta\hat{\mathbf{u}}(t-1)\delta\hat{\mathbf{u}}(t-1)^T \right\rangle \mathbf{G}^T$$

$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ etc

$\mathbf{Q} \equiv \left\langle \delta\hat{\mathbf{u}}\delta\hat{\mathbf{u}}^T \right\rangle$

$$= \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Derivation of Kalman Filter

Model prediction $\hat{\mathbf{x}}(t,-)=\mathbf{A}\hat{\mathbf{x}}(t-1)+\mathbf{G}\hat{\mathbf{u}}(t-1)$

$$\mathbf{P}(t,-)=\mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T+\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

"–" indicating estimate without use of data at that instant

Now seek a linear estimate $\hat{\mathbf{x}}(t)=\mathbf{L}\hat{\mathbf{x}}(t,-)+\mathbf{K}\hat{\mathbf{y}}(t)$

true state

$$\mathbf{H}(t)\overline{\mathbf{x}}(t)+\mathbf{n}(t)=\hat{\mathbf{y}}(t)$$

Then $\langle\hat{\mathbf{x}}(t)\rangle=\mathbf{L}\langle\hat{\mathbf{x}}(t,-)\rangle+\mathbf{K}\langle\mathbf{H}\overline{\mathbf{x}}(t)+\mathbf{n}(t)\rangle=(\mathbf{L}+\mathbf{K}\mathbf{H})\langle\overline{\mathbf{x}}(t)\rangle$

Thus, for the estimate to be unbiased, $\mathbf{L}+\mathbf{K}\mathbf{H}=\mathbf{I}$ or $\mathbf{L}=\mathbf{I}-\mathbf{K}\mathbf{H}$

$$\begin{aligned}\hat{\mathbf{x}}(t)&=(\mathbf{I}-\mathbf{K}\mathbf{H})\hat{\mathbf{x}}(t,-)+\mathbf{K}\hat{\mathbf{y}}(t)\\&=\hat{\mathbf{x}}(t,-)+\mathbf{K}\left(\hat{\mathbf{y}}(t)-\mathbf{H}\hat{\mathbf{x}}(t,-)\right)\end{aligned}$$

Derivation of Kalman Filter

Choose $\mathbf{K}$ that minimizes error variance of $\hat{\mathbf{x}}(t)$

$$\hat{\mathbf{x}}(t) = (\mathbf{I} - \mathbf{K}\mathbf{H})\hat{\mathbf{x}}(t,-) + \mathbf{K}\hat{\mathbf{y}}(t)$$

Error $\delta\mathbf{x}(t) = (\mathbf{I} - \mathbf{K}\mathbf{H})\delta\mathbf{x}(t,-) + \mathbf{K}\delta\mathbf{n}(t)$

Error covariance

assuming $\left\langle \delta\mathbf{x}(t,-)\delta\mathbf{n}(t)^T \right\rangle = \mathbf{0}$

$$\mathbf{P}(t) = \left\langle \delta\mathbf{x}(t)\delta\mathbf{x}(t)^T \right\rangle$$

$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ etc

$$= (\mathbf{I} - \mathbf{K}\mathbf{H})\left\langle \delta\mathbf{x}(t,-)\delta\mathbf{x}(t,-)^T \right\rangle(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\left\langle \delta\mathbf{n}(t)\delta\mathbf{n}(t)^T \right\rangle\mathbf{K}^T$$

$$= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}(t,-)(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}(t)\mathbf{K}^T$$

$$= \mathbf{P} - \mathbf{K}\mathbf{H}\mathbf{P} - \mathbf{P}\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\mathbf{H}\mathbf{P}\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\mathbf{R}\mathbf{K}^T$$

$$= \mathbf{K}\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}^T - \mathbf{P}\mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{P} + \mathbf{P}$$

"completing the square"

$$= \left[\mathbf{K} - \mathbf{P}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1}\right]\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)\left[\mathbf{K} - \mathbf{P}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1}\right]^T$$

$$- \mathbf{P}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P} + \mathbf{P}$$

Summary of Kalman Filter Algorithm

1) Integrate model until the next set of observations (forecast step)

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

$$\mathbf{P} \equiv \left\langle \delta\hat{\mathbf{x}}\delta\hat{\mathbf{x}}^T \right\rangle$$

$$\mathbf{Q} \equiv \left\langle \delta\hat{\mathbf{u}}\delta\hat{\mathbf{u}}^T \right\rangle$$

$$\mathbf{R} \equiv \left\langle \delta\left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)\delta\left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)^T \right\rangle$$

"–" indicating estimate without use of data at that instant

2) Update model state with the new observations (correction step)

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

Kalman gain

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}\mathbf{H}(t)\mathbf{P}(t,-)$$

$$= \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

3) Return to 1) to step forward in time.

Properties of Kalman Filter

Forecast

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

1) The Kalman filter defines a recursive relationship in time of model state $\hat{\mathbf{x}}(t)$ and its error covariance $\mathbf{P}(t)$ (i.e., Kalman filter explicitly computes formal uncertainties),

2) The equations besides those of the state $\hat{\mathbf{x}}(t)$ are what's needed to form the Kalman gain $\mathbf{K}(t)$,

Properties of Kalman Filter

Forecast

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

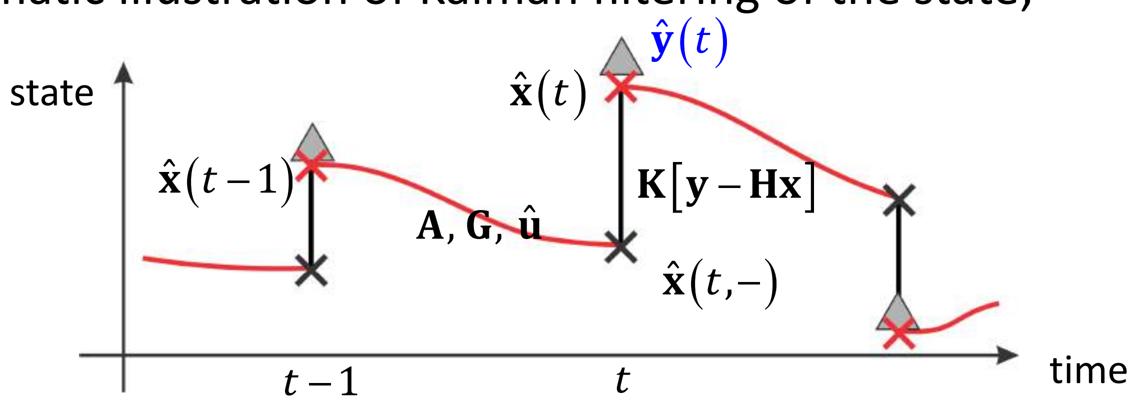
Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

3) Schematic illustration of Kalman filtering of the state;



4) Each observation $\hat{\mathbf{y}}$ is used once, only at the time of measurement,

Properties of Kalman Filter

Forecast

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

5) The Kalman gain $\mathbf{K}(t)$ corresponds to an inversion of observation $\mathbf{H}(t)$

Write $\overline{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \delta\hat{\mathbf{x}}(t,-)$ then $\mathbf{H}(t)\left[\hat{\mathbf{x}}(t,-) + \delta\hat{\mathbf{x}}(t,-)\right] \approx \hat{\mathbf{y}}(t)$

i.e., $\mathbf{H}(t)\delta\hat{\mathbf{x}}(t,-) \approx \hat{\mathbf{y}}'$ where $\hat{\mathbf{y}}' \equiv \hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)$

Minimum variance inversion $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$ $\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1}\hat{\mathbf{y}}$

Therefore, by inverting $\mathbf{H}(t)$

$$\delta\hat{\mathbf{x}}(t) = \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}\hat{\mathbf{y}}'$$

which is the Kalman filter.

Properties of Kalman Filter

Forecast

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

6) As the Kalman filter only inverts observations $\mathbf{H}$, its result is not an optimal solution to the entire state estimation problem.

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \mathbf{H} & & & & & \cdots \\ \cdots & & \mathbf{H} & & & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\ \cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

Properties of Kalman Filter

Forecast

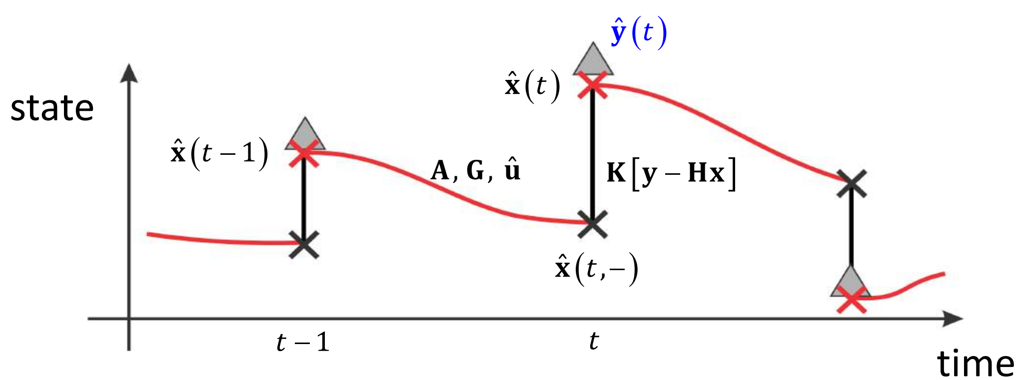
$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$
$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$
$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$
$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

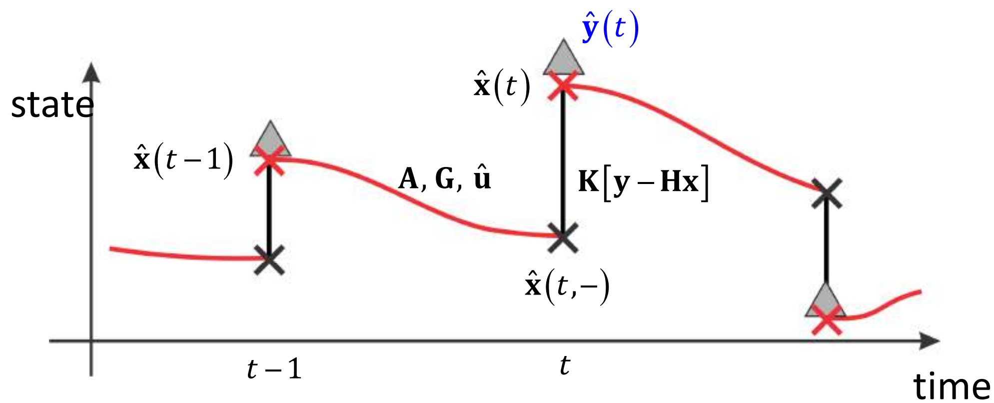
7) In fact, the filtered solution does not satisfy the model;

$$\hat{\mathbf{x}}(t) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\left\{\mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)\right\}\right]$$



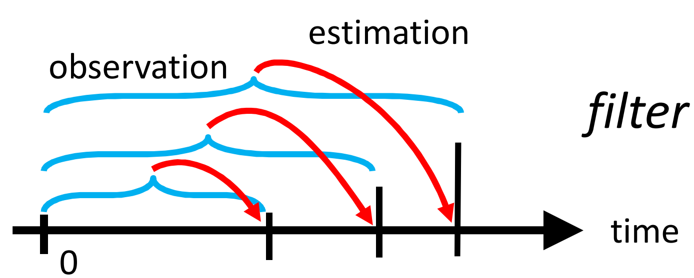
Summary of Kalman Filter

a) Kalman filter is a minimum variance estimator of the model state that inverts observations recursively in time,



b) Filtered solutions do not satisfy the model and are thus not optimal solutions to the full state estimation problem.

A reason for this limitation is the estimate's inconsistent use of data;

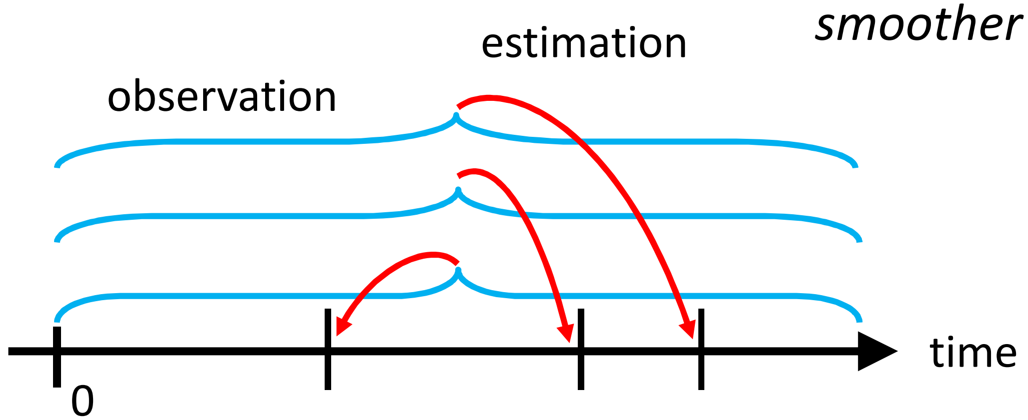


Rauch-Tung-Striebel Smoother

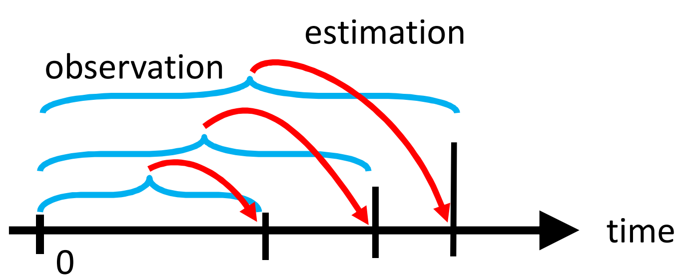
Rauch-Tung-Striebel (RTS) smoother is a
minimum variance estimator
that corrects the Kalman filter estimate
by inverting the model recursively in time
using formally future data.

A ***smoother*** , in estimation theory, is an estimator that employs observations that are both formally in the future as well as the past.

Fixed-interval smoother



filter



Derivation of RTS Smoother

Observations $\mathbf{H}(t)\mathbf{x}(t) \approx \hat{\mathbf{y}}(t)$

Model $\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$

Suppose we have an optimal (*minimum variance*) estimate at time $t-1$, $\hat{\mathbf{x}}(t-1)$, that uses all data up to that time with error covariance $\mathbf{P}(t-1)$. How do we optimize $\hat{\mathbf{x}}(t-1)$ and $\hat{\mathbf{u}}(t-1)$ using the new data $\hat{\mathbf{y}}(t)$?

" + " to indicate estimates using future data

Seek a linear estimates $\hat{\mathbf{x}}(t-1,+) = \mathbf{D}\hat{\mathbf{y}}(t) + \mathbf{E}\hat{\mathbf{x}}(t-1) + \mathbf{F}\hat{\mathbf{u}}(t-1)$

$$\hat{\mathbf{u}}(t-1,+) = \mathbf{D}'\hat{\mathbf{y}}(t) + \mathbf{E}'\hat{\mathbf{x}}(t-1) + \mathbf{F}'\hat{\mathbf{u}}(t-1)$$

An unbiased estimate requires $\mathbf{E} = \mathbf{I} - \mathbf{DHA}$ & $\mathbf{F} = -\mathbf{DHG}$

and likewise $\mathbf{E}' = -\mathbf{D}'\mathbf{HA}$ & $\mathbf{F}' = \mathbf{I} - \mathbf{D}'\mathbf{HG}$

Derivation of RTS Smoother

An unbiased estimate requires certain relationships among the coefficients.

$$\begin{aligned}\left\langle \hat{\mathbf{x}}(t-1,+)\right\rangle &= \mathbf{D}\left\langle \hat{\mathbf{y}}(t)\right\rangle + \mathbf{E}\left\langle \hat{\mathbf{x}}(t-1)\right\rangle + \mathbf{F}\left\langle \hat{\mathbf{u}}(t-1)\right\rangle \\ &= \mathbf{D}\left\langle \mathbf{Hx}(t)+\mathbf{n}(t)\right\rangle + \mathbf{E}\left\langle \hat{\mathbf{x}}(t-1)\right\rangle + \mathbf{F}\left\langle \hat{\mathbf{u}}(t-1)\right\rangle \\ &= \mathbf{D}\left\langle \mathbf{H}\left[\mathbf{Ax}(t-1)+\mathbf{Gu}(t-1)\right]+\mathbf{n}(t)\right\rangle + \mathbf{E}\left\langle \hat{\mathbf{x}}(t-1)\right\rangle + \mathbf{F}\left\langle \hat{\mathbf{u}}(t-1)\right\rangle \\ &= \left(\mathbf{E}+\mathbf{DHA}\right)\left\langle \mathbf{x}(t-1)\right\rangle + \left(\mathbf{F}+\mathbf{DHG}\right)\left\langle \mathbf{u}(t-1)\right\rangle\end{aligned}$$

Then $\mathbf{E}+\mathbf{DHA}=\mathbf{I}$ and $\mathbf{F}+\mathbf{DHG}=\mathbf{0}$

and therefore $\mathbf{E}=\mathbf{I}-\mathbf{DHA}$ & $\mathbf{F}=-\mathbf{DHG}$

$$\begin{aligned}\hat{\mathbf{x}}(t-1,+) &= \mathbf{D}\hat{\mathbf{y}}(t) + \left(\mathbf{I}-\mathbf{DHA}\right)\hat{\mathbf{x}}(t-1) - \mathbf{DHG}\hat{\mathbf{u}}(t-1) \\ &= \hat{\mathbf{x}}(t-1) + \mathbf{D}\left[\hat{\mathbf{y}}(t) - \mathbf{H}\left\{\mathbf{A}\hat{\mathbf{x}}(t-1)+\mathbf{G}\hat{\mathbf{u}}(t-1)\right\}\right]\end{aligned}$$

Derivation of RTS Smoother

$$\hat{\mathbf{x}}(t-1,+) = \mathbf{D}\hat{\mathbf{y}}(t) + (\mathbf{I}-\mathbf{DHA})\hat{\mathbf{x}}(t-1) - \mathbf{DHG}\hat{\mathbf{u}}(t-1)$$

$$\hat{\mathbf{u}}(t-1,+) = \mathbf{D}'\hat{\mathbf{y}}(t) - \mathbf{D}'\mathbf{HA}\hat{\mathbf{x}}(t-1) + (\mathbf{I}-\mathbf{D}'\mathbf{HA})\hat{\mathbf{u}}(t-1)$$

Choose $\mathbf{D}$ and $\mathbf{D}'$ that minimize the error variance of the new estimates.

$$\delta\hat{\mathbf{x}}(t-1,+) = \mathbf{D}\delta\mathbf{n}(t) + (\mathbf{I}-\mathbf{DHA})\delta\hat{\mathbf{x}}(t-1) - \mathbf{DHG}\delta\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{AP}(t-1)\mathbf{A}^T + \mathbf{GQ}(t-1)\mathbf{G}^T$$

$$\mathbf{P}(t-1,+) = \left\langle \delta\mathbf{x}(t-1,+)\delta\mathbf{x}(t-1,+)^T \right\rangle$$

$$= (\mathbf{I}-\mathbf{DHA})\mathbf{P}(t-1)(\mathbf{I}-\mathbf{DHA})^T + \mathbf{DHGQ}(t-1)(\mathbf{DHG})^T + \mathbf{DR}(t)\mathbf{D}^T$$

$$= \mathbf{D}\left[\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right]\mathbf{D}^T - \mathbf{P}(t-1)\mathbf{A}^T\mathbf{H}^T\mathbf{D}^T - \mathbf{DHAP}(t-1) + \mathbf{P}(t-1)$$

"completing the square"

$$= \left[\mathbf{D} - \mathbf{P}(t-1)\mathbf{A}^T\mathbf{H}^T\left(\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right)^{-1}\right]\left(\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right)\left[\mathbf{D} - \cdots\right]^T$$

$$- \mathbf{P}(t-1)\mathbf{A}^T\mathbf{H}^T\left(\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right)^{-1}\mathbf{HAP}(t-1) + \mathbf{P}(t-1)$$

Derivation of RTS Smoother

Thus,

$$\begin{cases} \hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{D}\left[\hat{\mathbf{y}}(t) - \mathbf{H}\left\{\mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)\right\}\right] \\ \mathbf{D} \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{H}^T\left[\mathbf{H}\mathbf{P}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right]^{-1} \\ \mathbf{P}(t-1,+) = \mathbf{P}(t-1) - \mathbf{P}(t-1)\mathbf{A}^T\mathbf{H}^T\left[\mathbf{H}\mathbf{P}(t,-)\mathbf{H}^T + \mathbf{R}(t)\right]^{-1}\mathbf{H}\mathbf{A}\mathbf{P}(t-1) \end{cases}$$

These can be simplified using results of the Kalman filter;

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) = \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

As

$$\begin{cases} \hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t) - \hat{\mathbf{x}}(t,-)\right] \\ \mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1} \\ \mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T \end{cases}$$

RTS Smoother Algorithm

Smoothed estimate indicated with " + " to note use of future data

By recursion

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

similarly,

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{Q}(t-1,+) = \mathbf{Q}(t-1) + \mathbf{M}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

which can be iterated backward in time and is called the Rauch-Tung-Striebel Smoother.

Properties of RTS Smoother

State

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

Control

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{Q}(t-1,+) = \mathbf{Q}(t-1) + \mathbf{M}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

1) RTS smoother algorithm defines a recursive relationship backward in time of smoothed model state $\hat{\mathbf{x}}(+)$ and control $\hat{\mathbf{u}}(+)$, along with their error covariances $\mathbf{P}(+)$ and $\mathbf{Q}(+)$, respectively,

2) Elements of the recursion are results from the Kalman filter; i.e., RTS smoother is correcting results from the Kalman filter,

3) The smoothed error covariances $\mathbf{P}(+)$ and $\mathbf{Q}(+)$ are not used in deriving the smoothed state $\hat{\mathbf{x}}(+)$ and control $\hat{\mathbf{u}}(+)$.

Properties of RTS Smoother

State

$$\hat{\mathbf{x}}(t-1,+)=\hat{\mathbf{x}}(t-1)+\mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+)-\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t)\equiv\mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+)=\mathbf{P}(t-1)+\mathbf{L}(t)\left[\mathbf{P}(t,+)-\mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

Control

$$\hat{\mathbf{u}}(t-1,+)=\hat{\mathbf{u}}(t-1)+\mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+)-\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t)\equiv\mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{Q}(t-1,+)=\mathbf{Q}(t-1)+\mathbf{M}(t)\left[\mathbf{P}(t,+)-\mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

4) Smoother gains $\mathbf{L}(t)$ and $\mathbf{M}(t)$ correspond to inversions of model matrices $\mathbf{A}$ and $\mathbf{G}$, respectively.

As $\mathbf{P}(t,-)=\mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T+\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$

$$\mathbf{L}(t)\equiv\mathbf{P}(t-1)\mathbf{A}^T\left[\mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T+\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T\right]^{-1}$$

$$\mathbf{M}(t)\equiv\mathbf{Q}(t-1)\mathbf{G}^T\left[\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T+\mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T\right]^{-1}$$

Minimum variance inversion $\mathbf{E}\mathbf{x}+\mathbf{n}=\hat{\mathbf{y}}$ $\quad\hat{\mathbf{x}}=\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T+\mathbf{R}_{\mathbf{nn}}\right)^{-1}\hat{\mathbf{y}}$

Model $\delta\mathbf{x}(t)=\mathbf{A}\delta\mathbf{x}(t-1)+\mathbf{G}\delta\mathbf{u}(t-1)$

Properties of RTS Smoother

State

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

Control

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{Q}(t-1,+) = \mathbf{Q}(t-1) + \mathbf{M}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

5) The smoothed solution is consistent with the model.

$$\begin{aligned}\mathbf{A}\hat{\mathbf{x}}(t-1,+) + \mathbf{G}\hat{\mathbf{u}}(t-1,+) &= \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{AL}\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right] + \\ &\qquad \mathbf{G}\hat{\mathbf{u}}(t-1) + \mathbf{GM}\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right] \\ &= \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1) + (\mathbf{AL}+\mathbf{GM})\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right] = \hat{\mathbf{x}}(t,+)\end{aligned}$$

Noting

$$\begin{aligned}\mathbf{AL}+\mathbf{GM} &= \mathbf{AP}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1} + \mathbf{GQ}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1} \\ &= \left[\mathbf{AP}(t-1)\mathbf{A}^T + \mathbf{GQ}(t-1)\mathbf{G}^T\right]\mathbf{P}(t,-)^{-1} \\ &= \mathbf{P}(t,-)\mathbf{P}(t,-)^{-1} = \mathbf{I}\end{aligned}$$

and $\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$

Properties of RTS Smoother

State

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

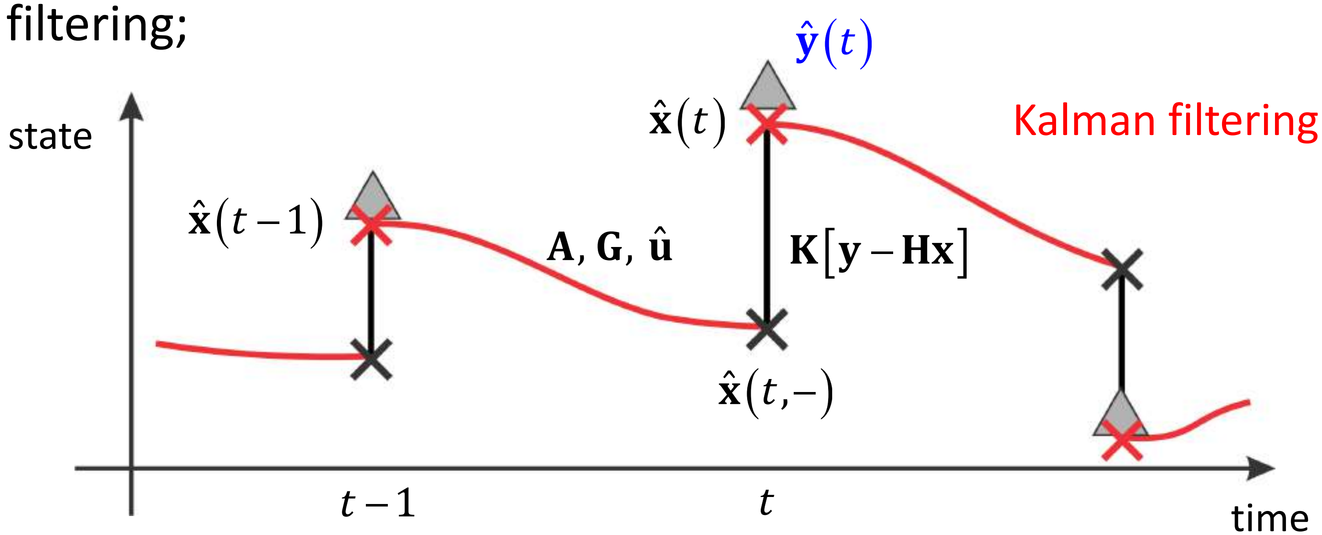
Control

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{Q}(t-1,+) = \mathbf{Q}(t-1) + \mathbf{M}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

6) Schematic illustration of RTS smoothing in relation to Kalman filtering;



Properties of RTS Smoother

State

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P}(t-1,+) = \mathbf{P}(t-1) + \mathbf{L}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{L}(t)^T$$

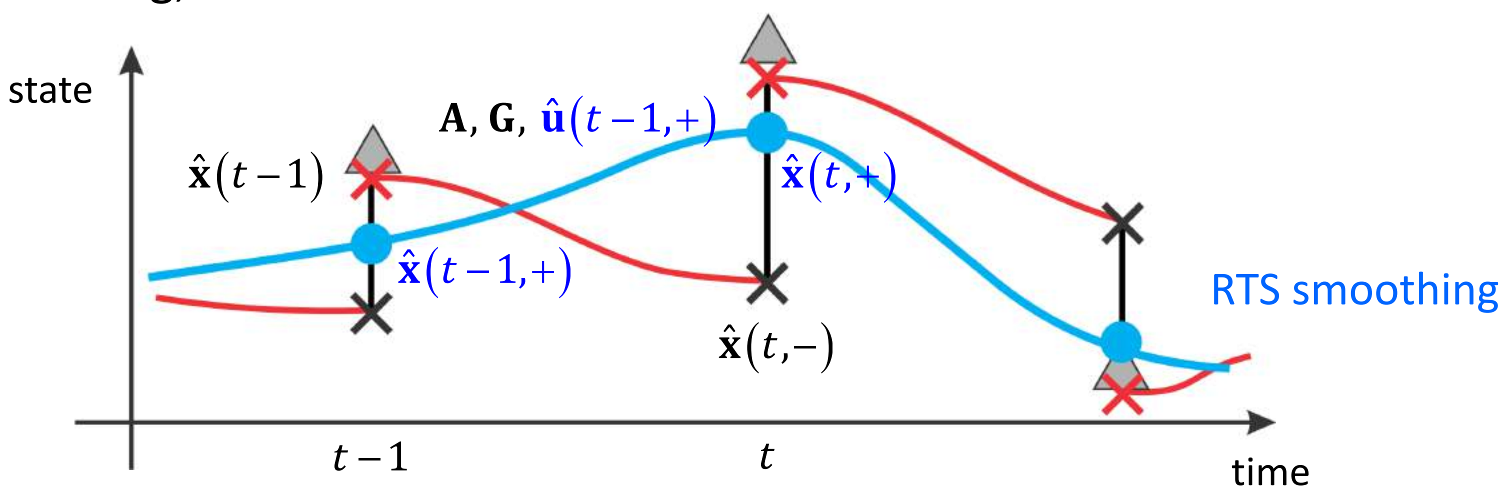
Control

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

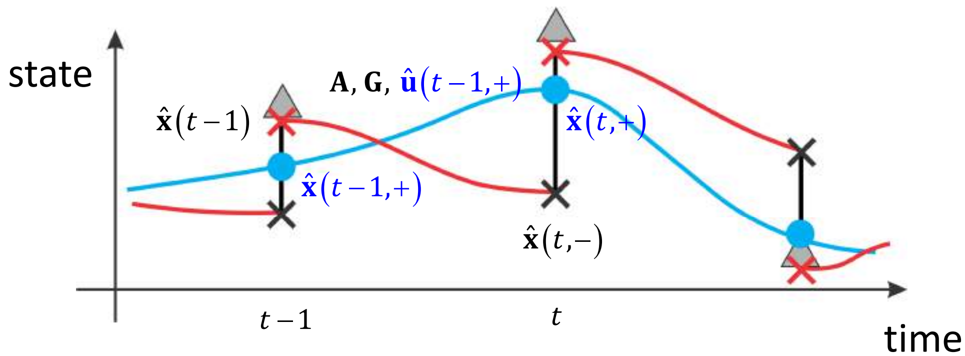
$$\mathbf{Q}(t-1,+) = \mathbf{Q}(t-1) + \mathbf{M}(t)\left[\mathbf{P}(t,+) - \mathbf{P}(t,-)\right]\mathbf{M}(t)^T$$

6) Schematic illustration of RTS smoothing;in relation to Kalman filtering;



Summary of RTS Smoother

a) The Rauch-Tung-Striebel (RTS) smoother is a minimum variance estimator of the state and control that inverts the model recursively backward in time correcting the Kalman filter estimate,



b) Smoothed estimates provide an optimal solution to the entire state estimation problem.

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \mathbf{H} & & & & & \cdots \\
\cdots & & \mathbf{H} & & & & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\
\cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots
\end{pmatrix}
\approx
\begin{pmatrix}
\vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots
\end{pmatrix}
$$

Adjoint Method

Adjoint Method (aka 4DVAR) is a

least-squares method

that solves the entire estimation problem

Iteratively by descent optimization.

analytically

$$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{R}_{\mathbf{nn}}^{-1} \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{x}$$

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}} \mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1} \hat{\mathbf{y}}$$

Description of Adjoint Method

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \mathbf{H} & & & & & \cdots \\
\cdots & & \mathbf{H} & & & & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\
\cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots
\end{pmatrix}
\approx
\begin{pmatrix}
\vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots
\end{pmatrix}
$$

In terms of least-squares, the estimation problem can be cast as finding a state $\hat{\mathbf{x}}(+)$ and control $\hat{\mathbf{u}}(+)$ that minimizes the objective function ,

$$
\begin{aligned}
\mathrm{J} \equiv & \sum_{t=1}^{M} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \\
& + \left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1} \left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] \\
& + \sum_{t=0}^{M-1} \left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1} \left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] \\
& - 2\sum_{t=1}^{M} \boldsymbol{\mu}(t)^T \left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+)\right]
\end{aligned}
$$

Lagrange multipliers

Description of Adjoint Method

Distinction between "soft" and "hard" constraints is artificial.

Solve $\begin{pmatrix} \mathbf{E} \\ \mathbf{A} \end{pmatrix} \mathbf{x} \approx \begin{pmatrix} \mathbf{y} \\ \mathbf{b} \end{pmatrix}$ given $\left\langle (\mathbf{y}-\mathbf{E}\mathbf{x})(\mathbf{y}-\mathbf{E}\mathbf{x})^T \right\rangle = \mathbf{W}$, $\left\langle (\mathbf{b}-\mathbf{A}\mathbf{x})(\mathbf{b}-\mathbf{A}\mathbf{x})^T \right\rangle = \mathbf{Q}$

$$\mathrm{J} \equiv (\mathbf{y}-\mathbf{E}\mathbf{x})\mathbf{W}^{-1}(\mathbf{y}-\mathbf{E}\mathbf{x})^T + (\mathbf{b}-\mathbf{A}\mathbf{x})\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x})^T$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}} = -\mathbf{E}^T\mathbf{W}^{-1}(\mathbf{y}-\mathbf{E}\mathbf{x}) - \mathbf{A}^T\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x})$$
$$= \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E} + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A}\right)\mathbf{x} - \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{y} + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{b}\right)$$

Solving $\partial J / \partial \mathbf{x} = \mathbf{0}$ we get

$$\hat{\mathbf{x}} = \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E} + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{y} + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{b}\right)$$

Description of Adjoint Method

Distinction between "soft" and "hard" constraints is artificial.

Write $\mathbf{Ax}+\mathbf{u}=\mathbf{b}$ then $\begin{pmatrix}\mathbf{E}\\ \mathbf{A}\end{pmatrix}\mathbf{x}\approx\begin{pmatrix}\mathbf{y}\\ \mathbf{b}\end{pmatrix}$ ⇒ $\begin{pmatrix}\mathbf{E} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{u}\end{pmatrix}\approx\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}$

$$J\equiv(\mathbf{y}-\mathbf{Ex})\mathbf{W}^{-1}(\mathbf{y}-\mathbf{Ex})^T+\mathbf{u}\mathbf{Q}^{-1}\mathbf{u}^T-2\boldsymbol{\mu}^T(\mathbf{b}-\mathbf{Ax}-\mathbf{u})$$

Lagrange multipliers

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}}=\mathbf{Q}^{-1}\mathbf{u}+\boldsymbol{\mu}\xrightarrow{=\mathbf{0}}\boldsymbol{\mu}=-\mathbf{Q}^{-1}\mathbf{u}$$

$$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}}=\mathbf{b}-\mathbf{Ax}-\mathbf{u}\xrightarrow{=\mathbf{0}}\mathbf{u}=\mathbf{b}-\mathbf{Ax}$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}}=-\mathbf{E}^T\mathbf{W}^{-1}(\mathbf{y}-\mathbf{Ex})+\mathbf{A}^T\boldsymbol{\mu}\xrightarrow{=\mathbf{0}}-\mathbf{E}^T\mathbf{W}^{-1}(\mathbf{y}-\mathbf{Ex})-\mathbf{A}^T\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{Ax})=\mathbf{0}$$

$$\left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E}+\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A}\right)\mathbf{x}-\left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{y}+\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{b}\right)=\mathbf{0}$$

Therefore,

$$\hat{\mathbf{x}}=\left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E}+\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{y}+\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{b}\right)$$

Description of Adjoint Method

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \mathbf{H} & & & & & \cdots \\ \cdots & & \mathbf{H} & & & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\ \cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

In terms of least-squares, the estimation problem can be cast as finding a state $\hat{\mathbf{x}}(+)$ and control $\hat{\mathbf{u}}(+)$ that minimizes the objective function ,

$$J \equiv \sum_{t=1}^{M} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]$$

$$+\left[\hat{\mathbf{x}}(0,+)-\hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1} \left[\hat{\mathbf{x}}(0,+)-\hat{\mathbf{x}}(0)\right]$$

$$+\sum_{t=0}^{M-1} \left[\hat{\mathbf{u}}(t,+)-\hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1} \left[\hat{\mathbf{u}}(t,+)-\hat{\mathbf{u}}(t)\right]$$

$$-2\sum_{t=1}^{M} \boldsymbol{\mu}(t)^T \left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+)\right]$$

Lagrange multipliers

Description of Adjoint Method

$$
\begin{aligned}
J \equiv & \sum_{t=1}^{M} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \\
& + \left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] \\
& + \sum_{t=0}^{M-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] \\
& - 2\sum_{t=1}^{M} \boldsymbol{\mu}(t)^T \left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+)\right]
\end{aligned}
$$

$$
\left\{
\begin{aligned}
& \frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T = 0 \qquad t = 0,1,\cdots,M-1 \\
& \frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+) = 0 \qquad t = 1,\cdots,M \\
& \frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1) = 0 \\
& \frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(t,+)} = -\mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T\boldsymbol{\mu}(t+1) = 0 \qquad t = 1,\cdots,M-1 \\
& \frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(M,+)} = -\mathbf{H}(M)^T\mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right] - \boldsymbol{\mu}(M) = 0
\end{aligned}
\right.
$$

Description of Adjoint Method

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T \qquad t = 0,1,\cdots,M-1$$

$$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+) \qquad t = 1,\cdots,M$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1)$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(t,+)} = -\mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T\boldsymbol{\mu}(t+1) \qquad t = 1,\cdots,M-1$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(M,+)} = -\mathbf{H}(M)^T\mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right] - \boldsymbol{\mu}(M)$$

1) A solution to $\partial J / \partial \boldsymbol{\mu}(t) = 0$ is obtained by integrating the model from initial condition $t = 0$ to terminal time $t = M$ using a first guess for initial condition $\hat{\mathbf{x}}(0,+)$ and control $\hat{\mathbf{u}}(t,+)$

$$\hat{\mathbf{x}}(t,+) = \mathbf{A}\hat{\mathbf{x}}(t-1,+) + \mathbf{G}\hat{\mathbf{u}}(t-1,+) \qquad t = 1,\cdots,M$$

which yields estimates for $\hat{\mathbf{x}}(t,+) \quad t = 1,\cdots,M$

Description of Adjoint Method

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T \boldsymbol{\mu}(t+1)^T \qquad t = 0,1,\cdots,M-1$$

✓ $$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+) = \mathbf{0} \qquad t = 1,\cdots,M$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T \boldsymbol{\mu}(1)$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(t,+)} = -\mathbf{H}(t)^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T \boldsymbol{\mu}(t+1) \qquad t = 1,\cdots,M-1$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(M,+)} = -\mathbf{H}(M)^T \mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right] - \boldsymbol{\mu}(M)$$

2) A solution to $\partial J / \partial \hat{\mathbf{x}}(M,+) = 0$ is obtained by choosing $\boldsymbol{\mu}(M)$ to be

$$\boldsymbol{\mu}(M) = -\mathbf{H}(M)^T \mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right]$$

Description of Adjoint Method

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T \qquad t = 0,1,\cdots,M-1$$

✓ $$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+) = \mathbf{0} \qquad t = 1,\cdots,M$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1)$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(t,+)} = -\mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T\boldsymbol{\mu}(t+1) \qquad t = 1,\cdots,M-1$$

✓ $$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(M,+)} = -\mathbf{H}(M)^T\mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right] - \boldsymbol{\mu}(M) = \mathbf{0}$$

3) A solution to $\partial J / \partial \hat{\mathbf{x}}(t,+) = 0 \quad t = 1,\cdots,M-1$ is obtained by choosing $\boldsymbol{\mu}(t)$ to satisfy

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \qquad t = M-1,\cdots,1$$

integrating this backward in time from $t = M-1$ to $t = 1$, using $\boldsymbol{\mu}(M)$ from step 2) as the terminal condition. $\mathbf{A}^T$ corresponds to the adjoint of the model.

Description of Adjoint Method

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T \qquad t=0,1,\cdots,M-1$$

✓ $$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+) = \mathbf{0} \qquad t=1,\cdots,M$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1)$$

✓ $$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(t,+)} = -\mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T\boldsymbol{\mu}(t+1) = \mathbf{0} \qquad t=1,\cdots,M-1$$

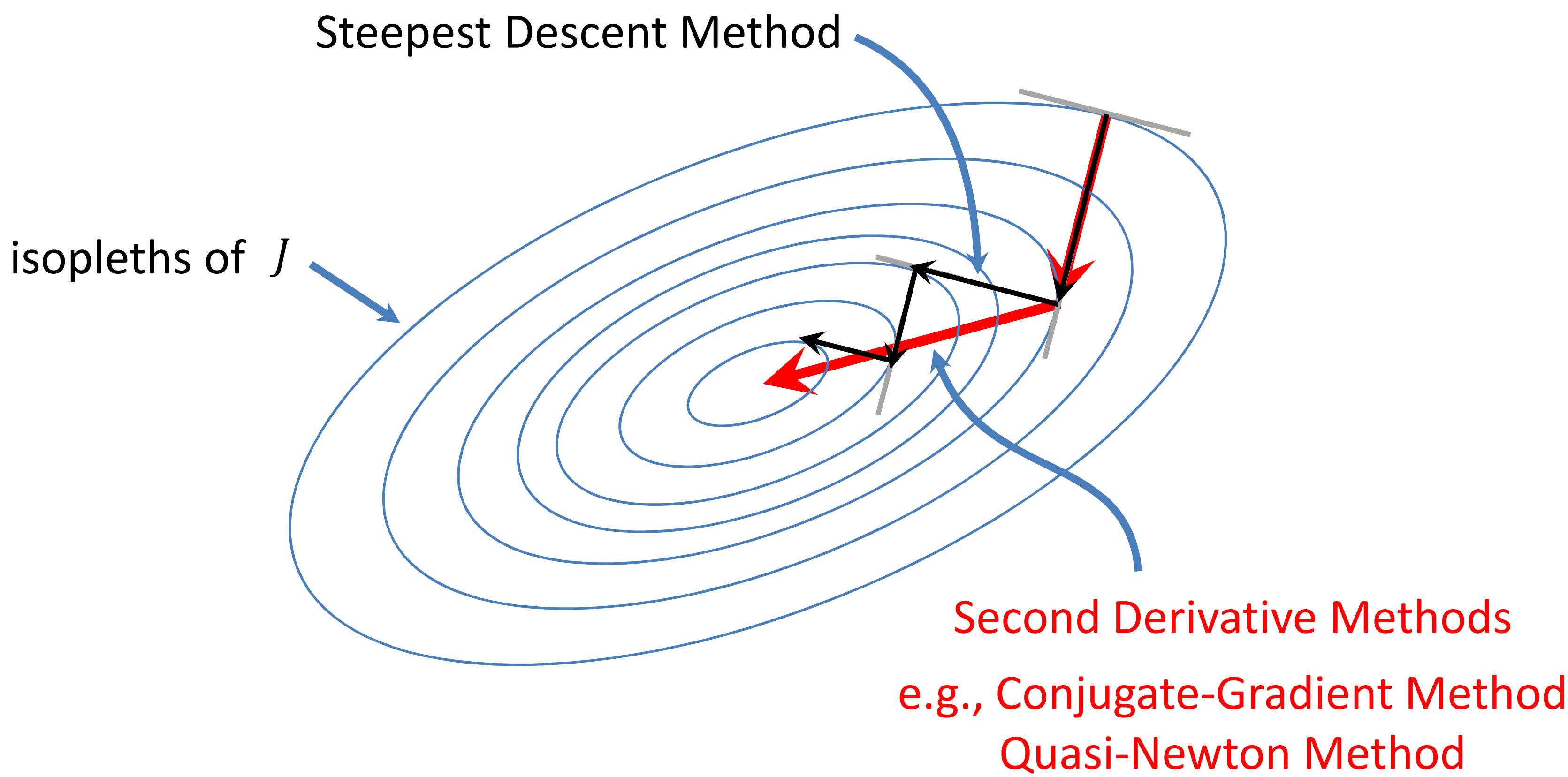
✓ $$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(M,+)} = -\mathbf{H}(M)^T\mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right] - \boldsymbol{\mu}(M) = \mathbf{0}$$

4) The solution obtained by 1) to 3) will, in general, not make the last two gradients zero; i.e., it is not an optimal solution, yet.

However, these remaining gradients provide a direction in which to change the initial guess of initial condition $\hat{\mathbf{x}}(0,+)$ and control $\hat{\mathbf{u}}(t,+)$ $t=0,1,\cdots,M-1$ used in step 1) (i.e., modify the particular solution), so as to decrease the value of J until it is at a minimum and these gradients zero; e.g., steepest descent method.

Description of Adjoint Method

Descent Optimization



Steps of Adjoint Method

1) Start with particular state $\hat{\mathbf{x}}(0,+)$ and control $\hat{\mathbf{u}}(t,+)$, $t=0,\cdots,M-1$
2) Integrate model from initial time $t=0$ to end time $t=M$

$$\hat{\mathbf{x}}(t,+) = \mathbf{A}\hat{\mathbf{x}}(t-1,+)+\mathbf{G}\hat{\mathbf{u}}(t-1,+)$$

3) Compute Lagrange multiplier at end time $t=M$

$$\boldsymbol{\mu}(M) = -\mathbf{H}(M)^T\mathbf{R}(M)^{-1}\left[\hat{\mathbf{y}}(M) - \mathbf{H}(M)\hat{\mathbf{x}}(M,+)\right]$$

4) Integrate <u>*adjoint*</u> model from end time $t=M$ to initial time $t=0$

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1)-\mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]$$

5) Evaluate cost gradients $\partial J/\partial\hat{\mathbf{x}}(0,+)$ and $\partial J/\partial\hat{\mathbf{u}}(t,+)$, $t=M-1,\cdots,0$

$$\frac{1}{2}\frac{\partial J}{\partial\hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+)-\hat{\mathbf{u}}(t)\right]+\mathbf{G}^T\boldsymbol{\mu}(t+1)^T$$

$$\frac{1}{2}\frac{\partial J}{\partial\hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+)-\hat{\mathbf{x}}(0)\right]+\mathbf{A}^T\boldsymbol{\mu}(1)$$

6) Terminate if gradients in 5) (and/or cost J) are negligible. Otherwise, use these gradients in descent optimization to improve the estimates in 1) and repeat the steps.

Properties of Adjoint Method

$$\boldsymbol{\mu}(t) = \mathbf{A}^T \boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right], \quad t = M-1,\cdots,1$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T \boldsymbol{\mu}(t+1)^T, \quad t = M-1,\cdots,0$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T \boldsymbol{\mu}(1)$$

1) What is an adjoint?

An adjoint of a operator $\boldsymbol{\Phi}$, denoted $\boldsymbol{\Phi}^*$, is an operator that satisfies

$$\left\langle \boldsymbol{\Phi}(\mathbf{u}), \mathbf{v} \right\rangle = \left\langle \mathbf{u}, \boldsymbol{\Phi}^*(\mathbf{v}) \right\rangle$$ ← inner product

In linear algebra, adjoint corresponds to matrix transpose (conjugate transpose);

$$\left\langle \boldsymbol{\Phi}\mathbf{u}, \mathbf{v} \right\rangle = \left(\boldsymbol{\Phi}\mathbf{u}\right)^T \mathbf{v} = \mathbf{u}^T \boldsymbol{\Phi}^T \mathbf{v} = \left\langle \mathbf{u}, \boldsymbol{\Phi}^T \mathbf{v} \right\rangle$$

Solutions to adjoints are intimately tied to solutions of the original operators, and its appearance here is no accident. In fact, an adjoint operator also results when the estimation problem is formulated in continuous form, thus the identification here of $\mathbf{A}^T$ as the adjoint.

Properties of Adjoint Method

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right], \quad t = M-1,\cdots,1$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T, \quad t = M-1,\cdots,0$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1)$$

2) The significance of the equations can be understood by recognizing that the adjoint evaluates model's sensitivity backward in time;

$$\frac{\partial F\left[\mathbf{x}(t+N)\right]}{\partial \mathbf{u}(t)} = \frac{\partial \mathbf{x}(t+1)}{\partial \mathbf{u}(t)}\frac{\partial \mathbf{x}(t+2)}{\partial \mathbf{x}(t+1)}\frac{\partial \mathbf{x}(t+3)}{\partial \mathbf{x}(t+2)}\cdots\frac{\partial F\left[\mathbf{x}(t+N)\right]}{\partial \mathbf{x}(t+N)}$$

$$= \mathbf{G}^T\left(\mathbf{A}^T\right)^{N-1}\frac{\partial F\left[\mathbf{x}(t+N)\right]}{\partial \mathbf{x}(t+N)}$$

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$

3) The adjoint equations are "forced" by the model-data misfits. The adjoint variable $\boldsymbol{\mu}(t)$ (Lagrange multiplier) therefore is the sensitivity of these misfits to the model state $\mathbf{x}(t)$ and the "controls" $\mathbf{u}(t)$ at earlier instances.

Summary of Adjoint Method

a) The adjoint method numerically solves the estimation problem by least-squares, in which the model's adjoint provides an efficient means to compute the gradient of the least-squares objective function for use in descent optimization,

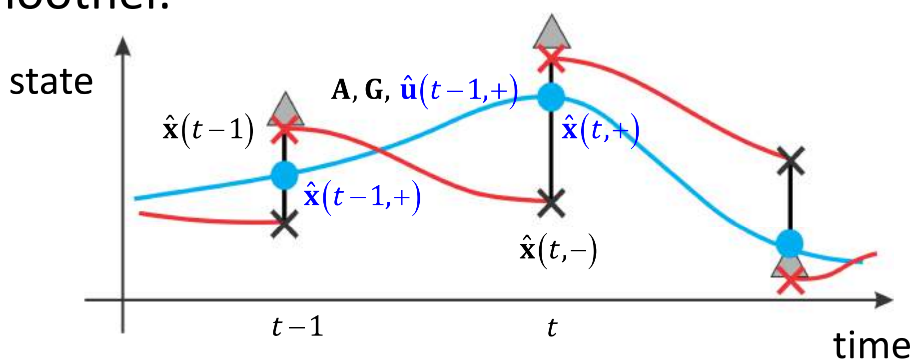
$$
\begin{aligned}
J \equiv & \sum_{t=1}^{M} \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \\
& + \left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] \\
& + \sum_{t=0}^{M-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] \\
& - 2\sum_{t=1}^{M}\boldsymbol{\mu}(t)^T \left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+)\right]
\end{aligned}
$$

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] + \mathbf{G}^T\boldsymbol{\mu}(t+1)^T$$

$$\frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] + \mathbf{A}^T\boldsymbol{\mu}(1)$$

b) By construction, the adjoint method is a smoother equivalent to the RTS smoother.



Concluding Remarks on Estimation Methods

1) Kalman filter is a recursive minimum variance estimator that inverts data up to each instant,

2) Kalman filter does not solve the entire state estimation problem,

3) Rauch-Tung-Striebel smoother is a recursive minimum variance estimator that solves the entire estimation problem by inverting the model to correct the Kalman filter estimate,

4) Adjoint method is a least-squares method that solves the entire estimation problem using iterative descent methods by evaluating the gradient of the least-squares problem with the model's adjoint.

Next Topic

How does one apply these methods in practice?

1) What is data error?
   - Representation error
   - How is it prescribed?

$$\mathbf{R} \equiv \left\langle \left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)\left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)^T \right\rangle \neq \left\langle \left(\hat{\mathbf{y}} - \overline{\mathbf{y}}\right)\left(\hat{\mathbf{y}} - \overline{\mathbf{y}}\right)^T \right\rangle$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T \left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]$$

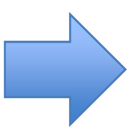
2) How can a Kalman gain (smoother) be derived given a model?

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

- Approximate Kalman filters and smoothers
- Kalman-filter-like methods (e.g., optimal interpolation)

3) How is an adjoint derived given a model?

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) - \mathbf{H}(t)^T\mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]$$



Automatic Differentiation
(Patrick Heimbach)

Jet Propulsion Laboratory
California Institute of Technology