The ends and means of “data assimilation”

Patrick Heimbach
Oden Institute for Computational Engineering and Sciences
Jackson School of Geosciences
University of Texas at Austin

Estimating the Circulation and Climate of the Ocean
http://ecco-group.org

What is data assimilation?

It's all about …

- making optimal use of,
- consistently extracting,
- or combining

information contained in observations and physical laws expressed through a model, and taking into account all uncertainties.

Outline … = Conclusions

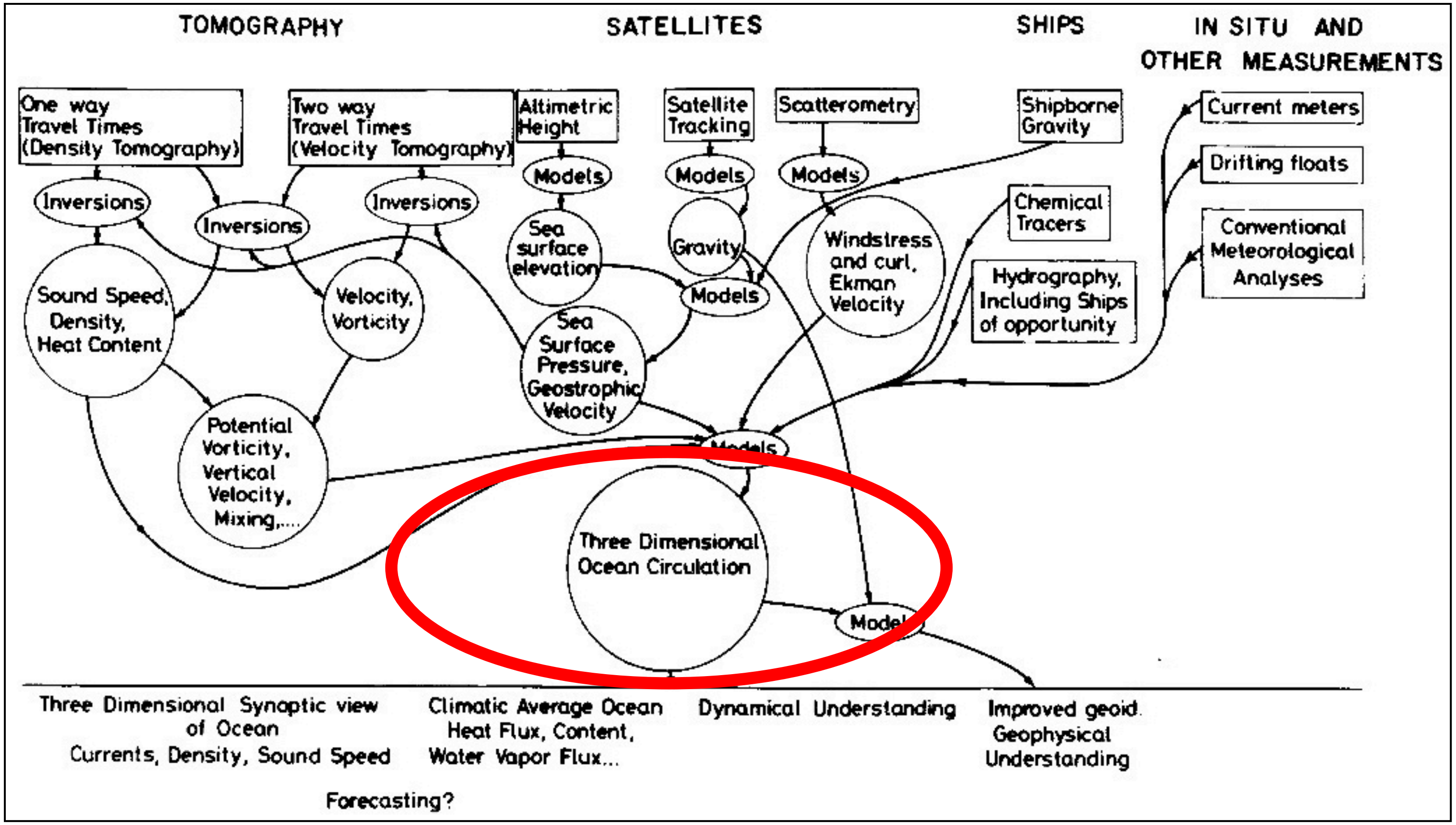
- DA seeks to optimally combine information content in observations, models, and their uncertainties(!)

- DA can mean different things to different people

- Depending on application, different methods:
  - forecasting: *filter methods* (e.g., Kalman filter)
  - reconstruction: *smoother / adjoint methods*

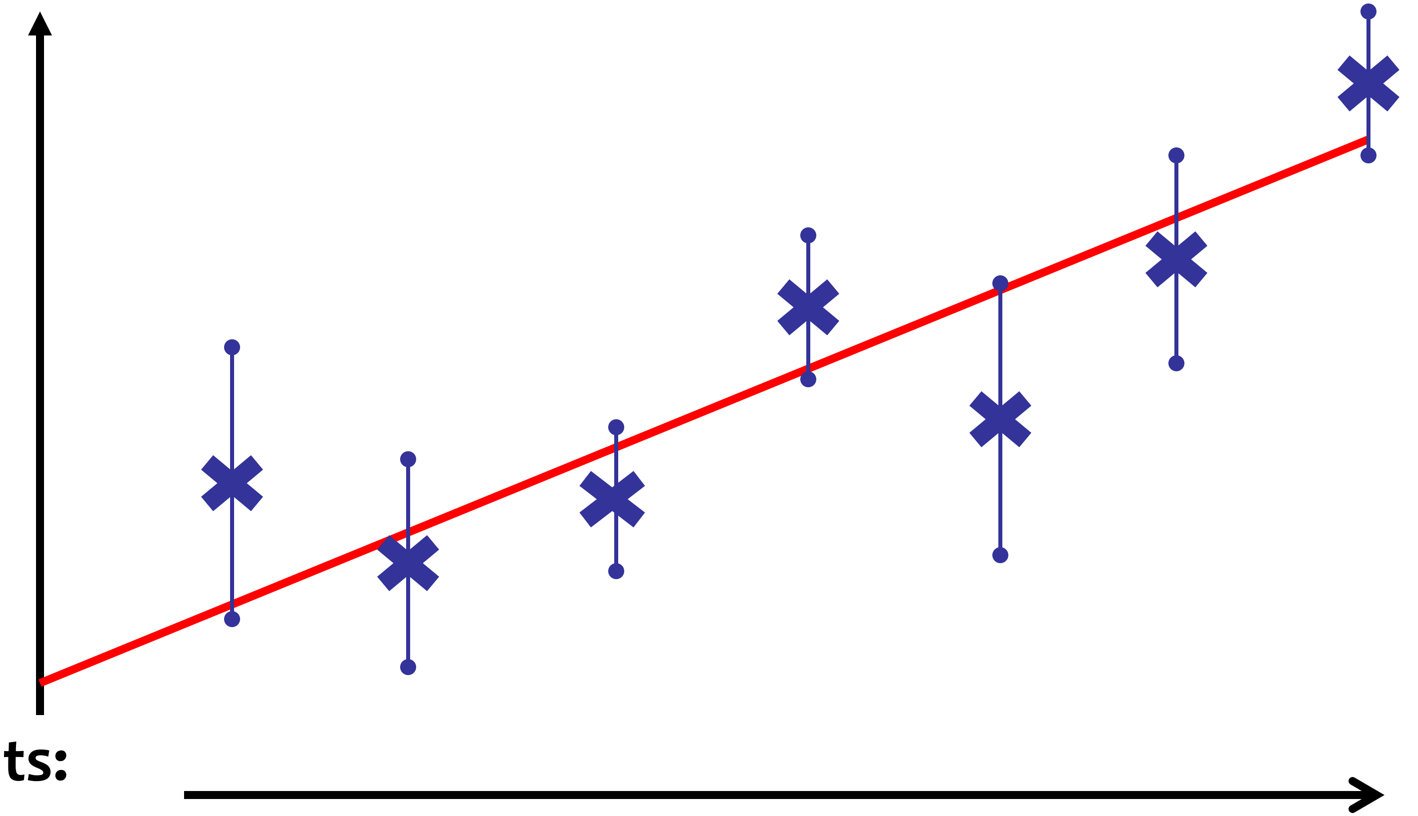
Combine two incomplete information sources

- Combine the heterogeneous streams of measured state variables with simulations of these same variables
- The way of how we combine is influenced by / takes into account the different sources of uncertainties!



***C. Wunsch**, in "A Celebration in Geophysics and Oceanography 1982. In Honor of Walter Munk on his 65th birthday."*

Formal framework – least-squares objective/cost function



Basic ingredients:

- observations (diverse types, sparse, inhomogeneously distributed in space & time)
- model (various levels of complexity)
- errors/uncertainties (of various kinds)

Least-squares objective/cost function

3 basic ingredients:

observations — model — errors

$$J = \sum_{i,j,k} \frac{1}{\sigma^2(i,j,k)} \left[y(i,j,k) - \mathbf{E}x(i,j,k) \right]^2$$

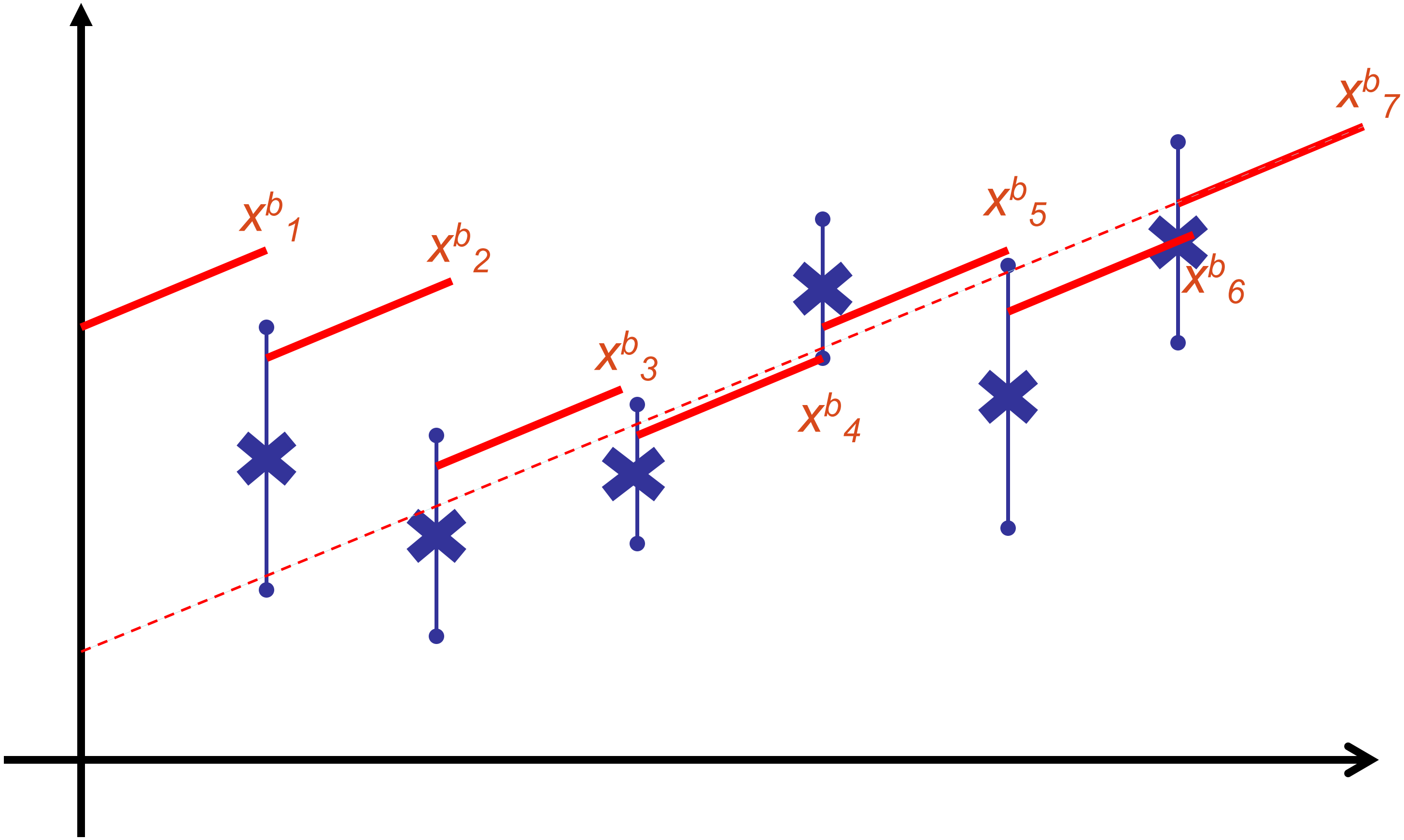
or, more generally (but dropping space indices i, j, k):

$$J = \left[y - \mathbf{E}x \right]^T \mathbf{R}^{-1} \left[y - \mathbf{E}x \right]$$

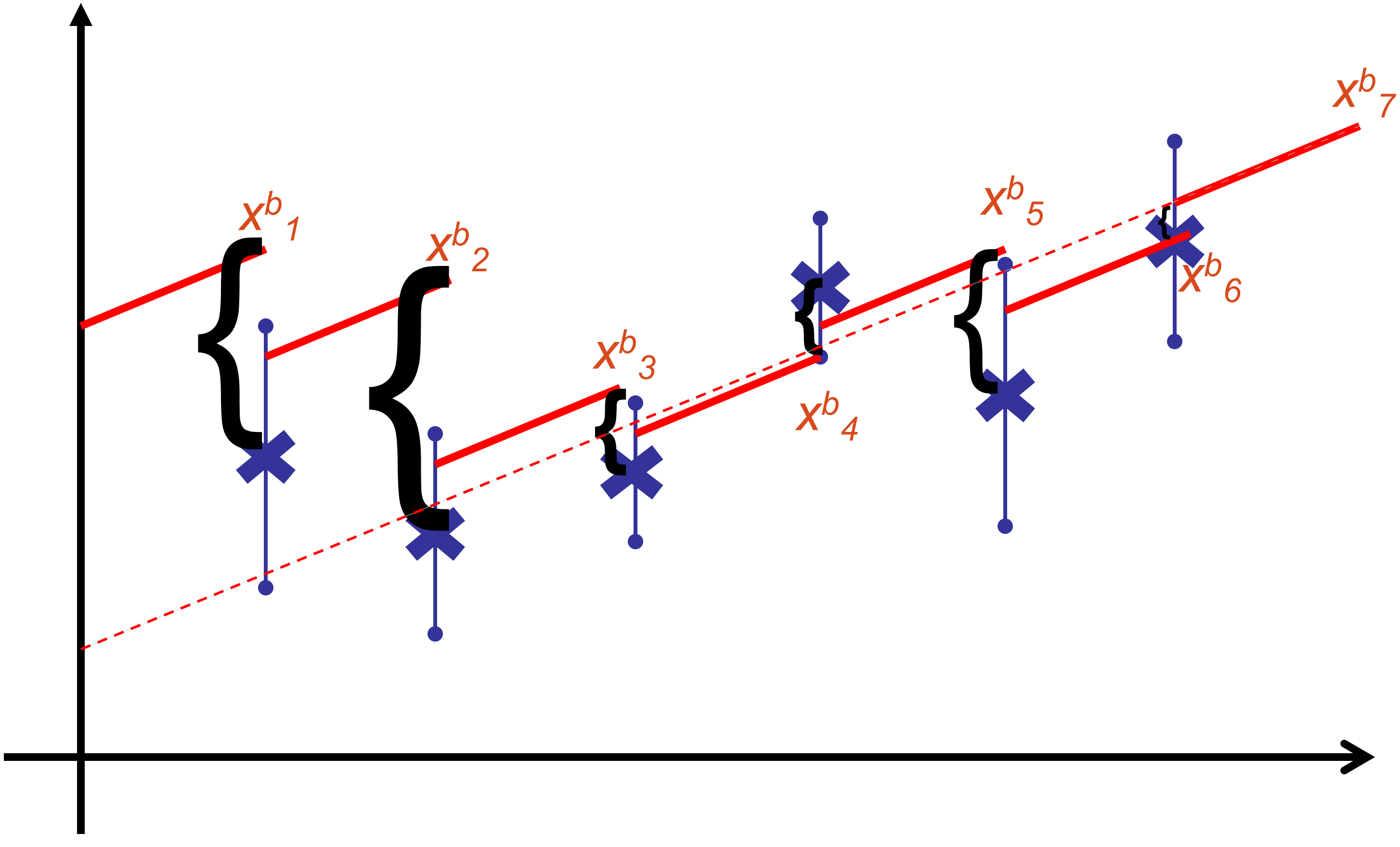
E: operator mapping from model space (x) to obs space (y)

R: error covariance matrix (with error variances σ^2)

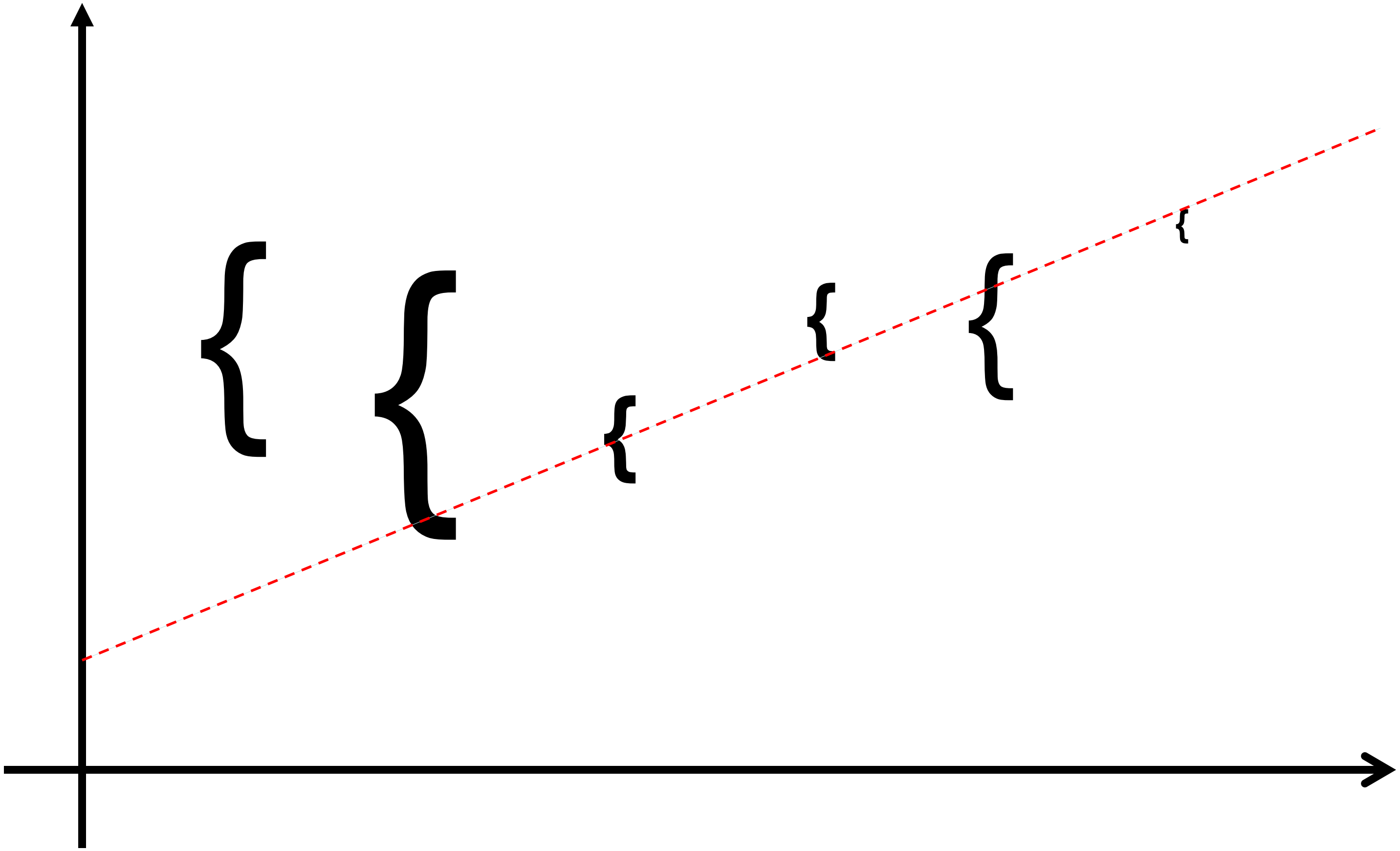
The filtering (forecasting) problem



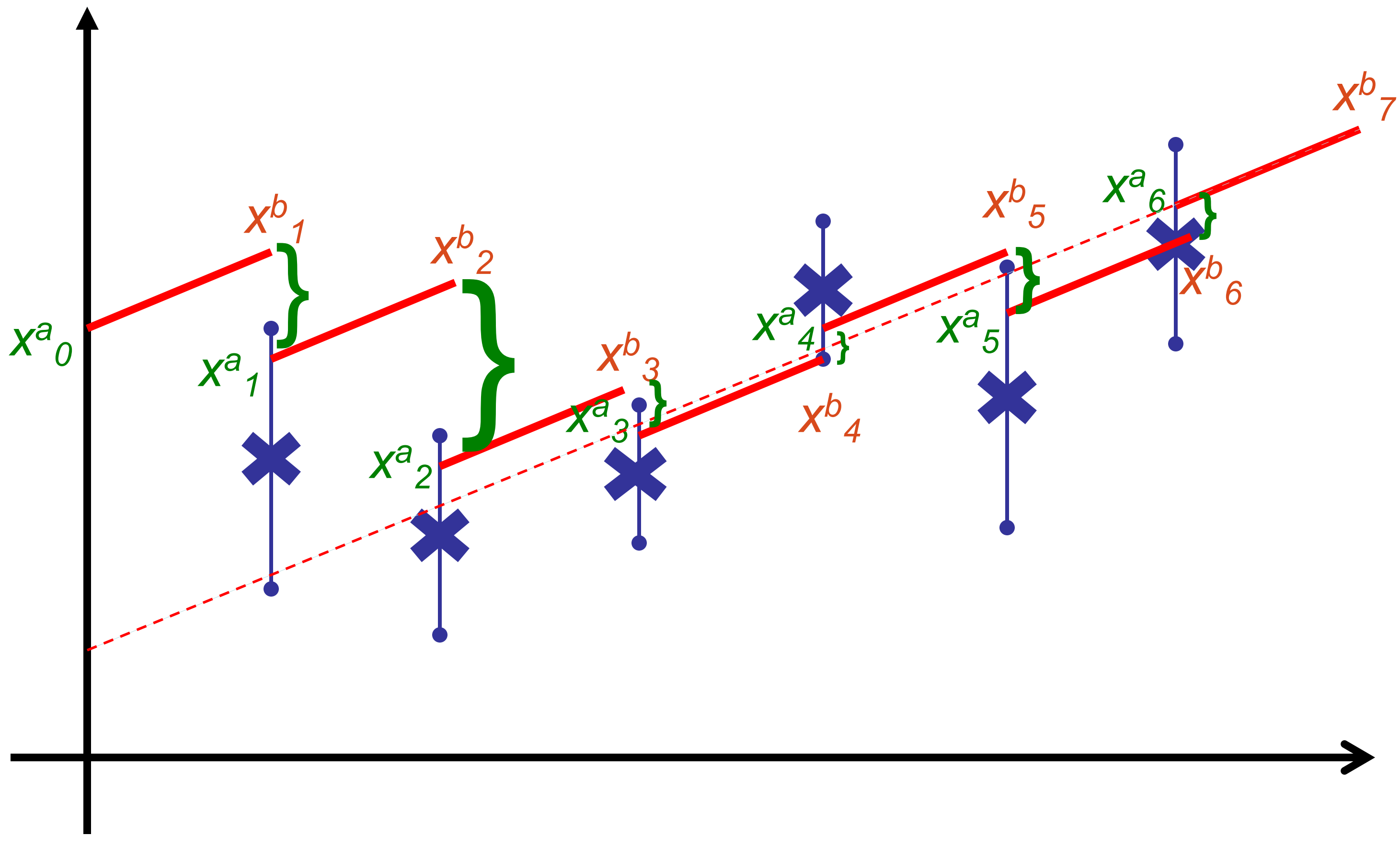
The filtering (forecasting) problem



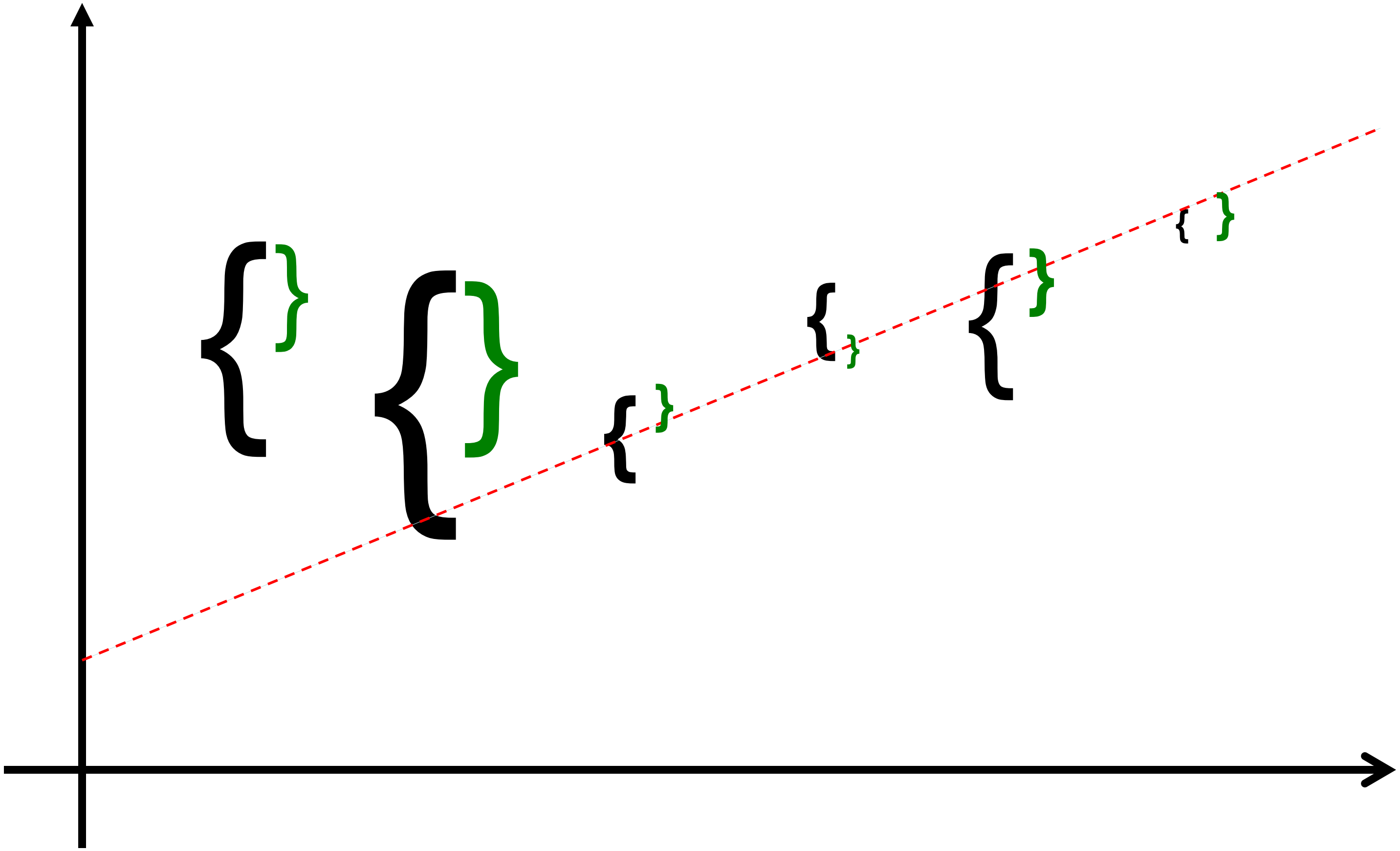
The filtering (forecasting) problem



State estimation via filtering (sequential) methods

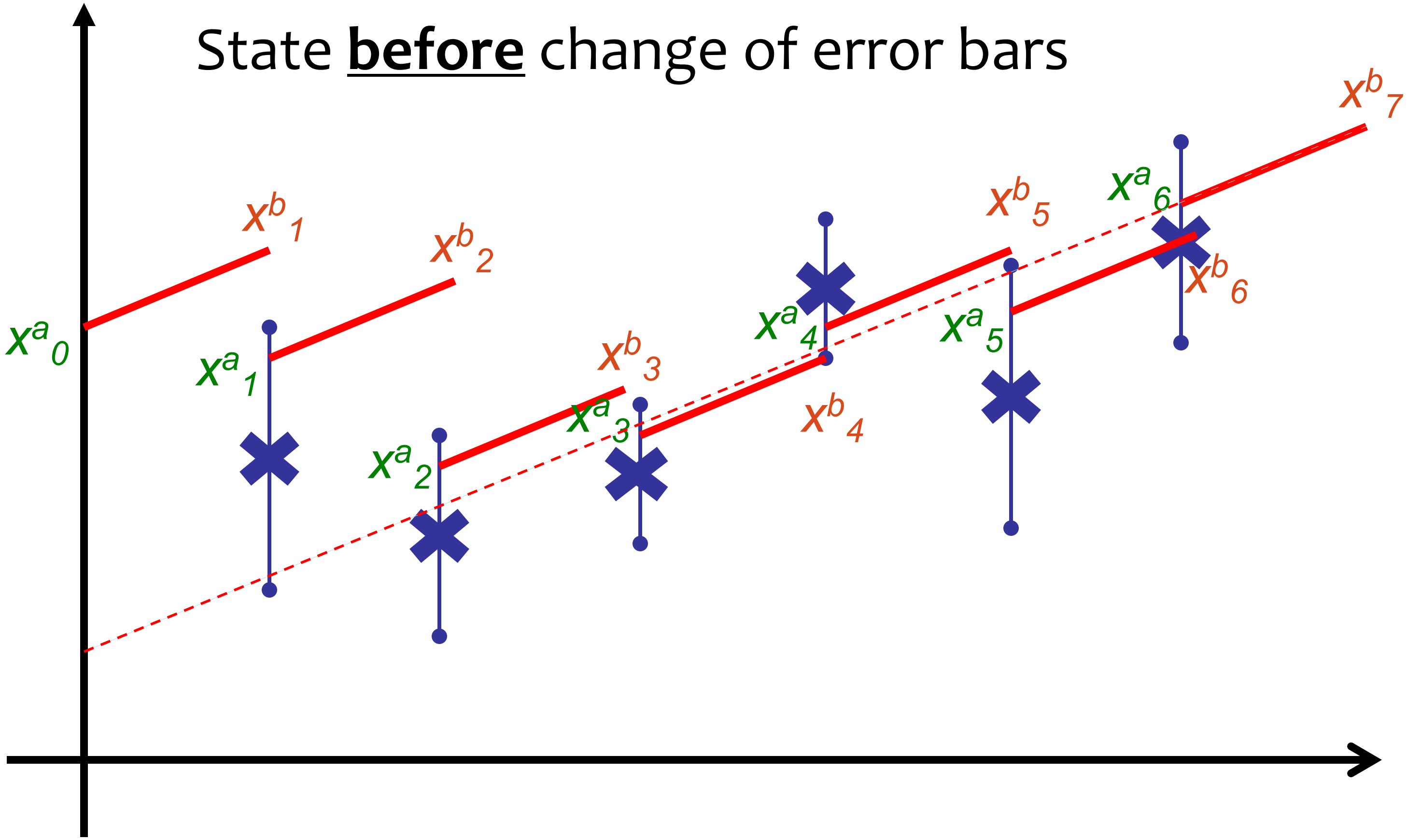


State estimation via filtering (sequential) methods



- Innovation vector (or residual)
- Analysis increment

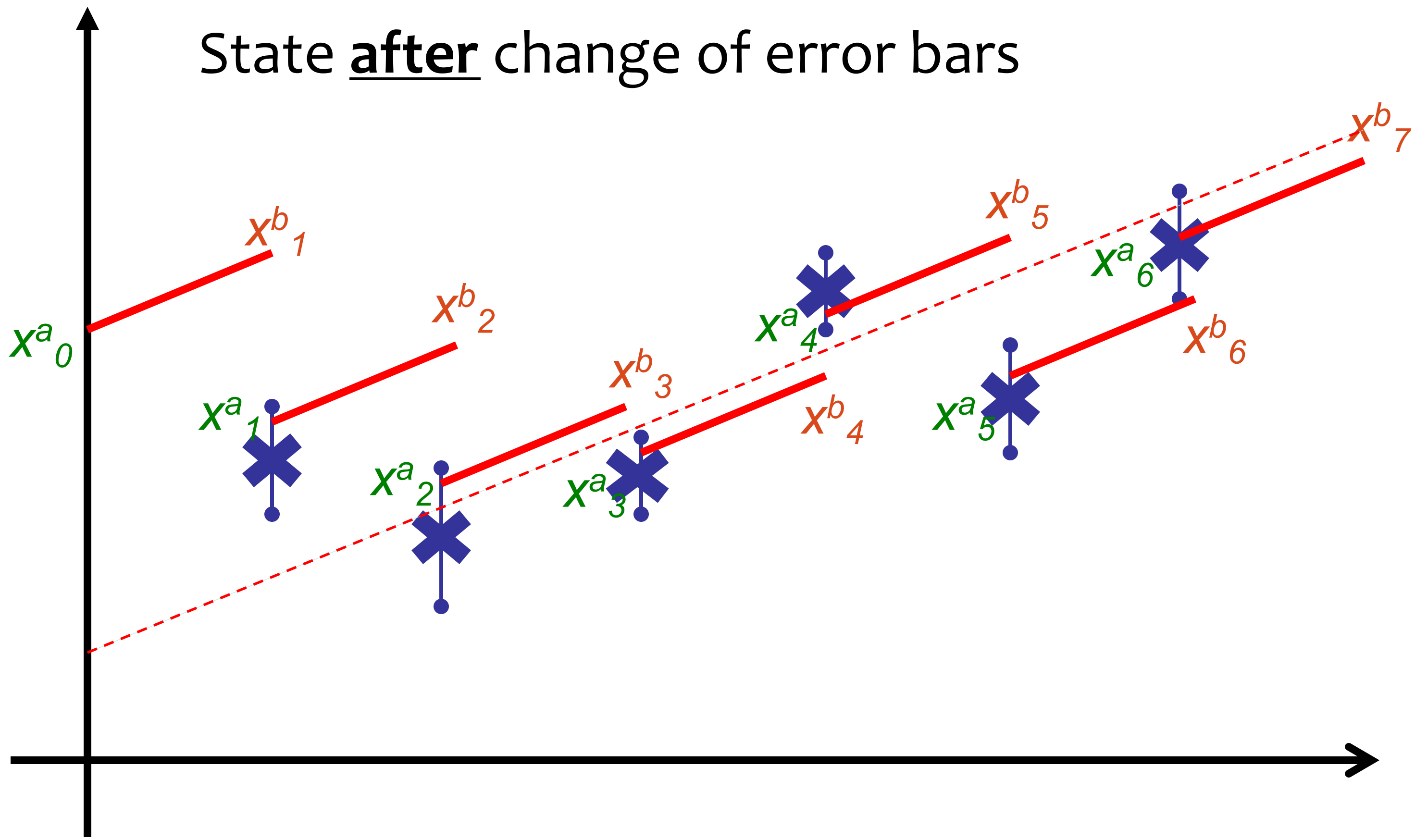
State estimation via filtering (sequential) methods



NOTE:

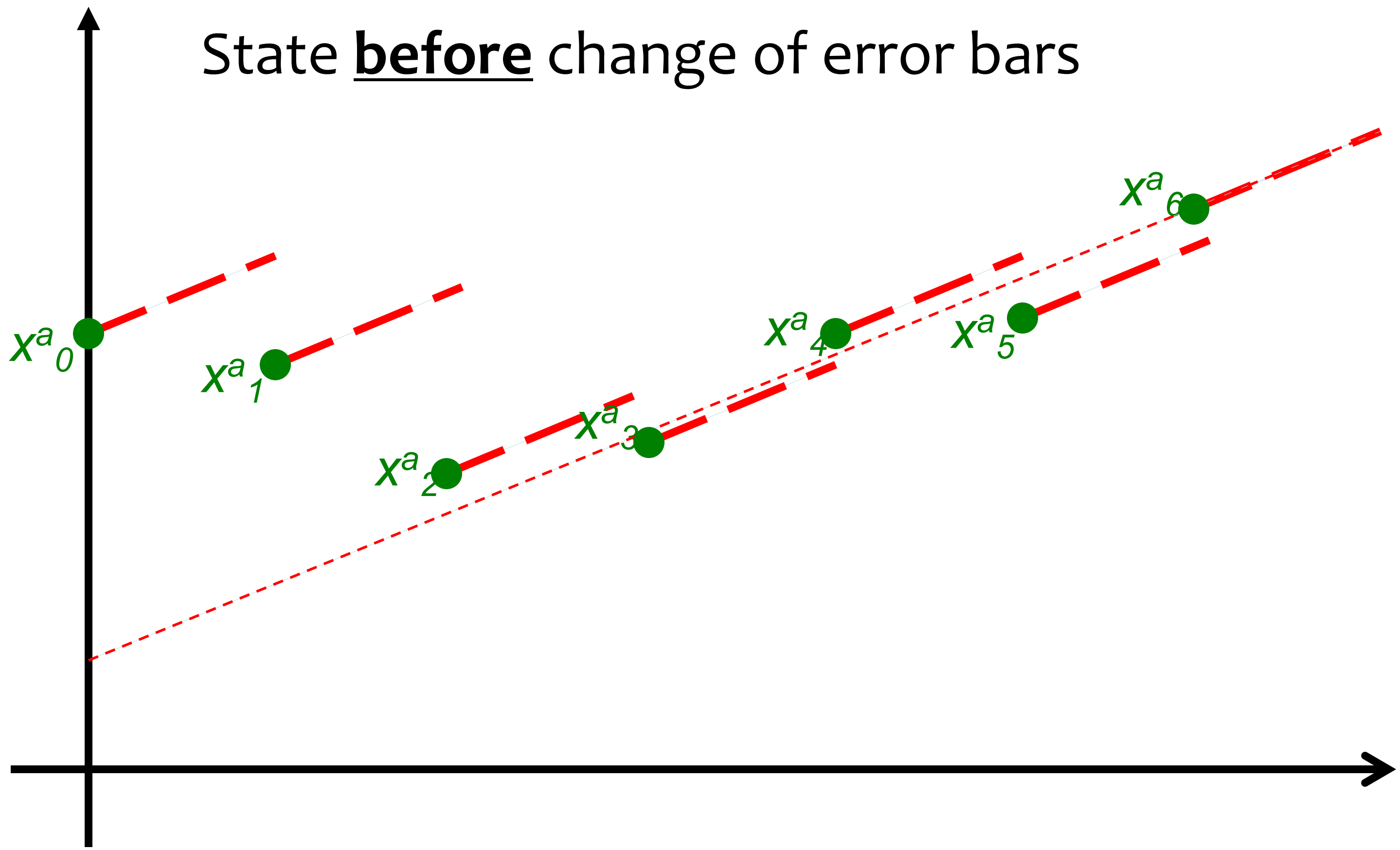
- <u>The role/importance of error/uncertainty estimates</u>!
- What if we reduce observation errors?

State estimation via filtering (sequential) methods



NOTE:

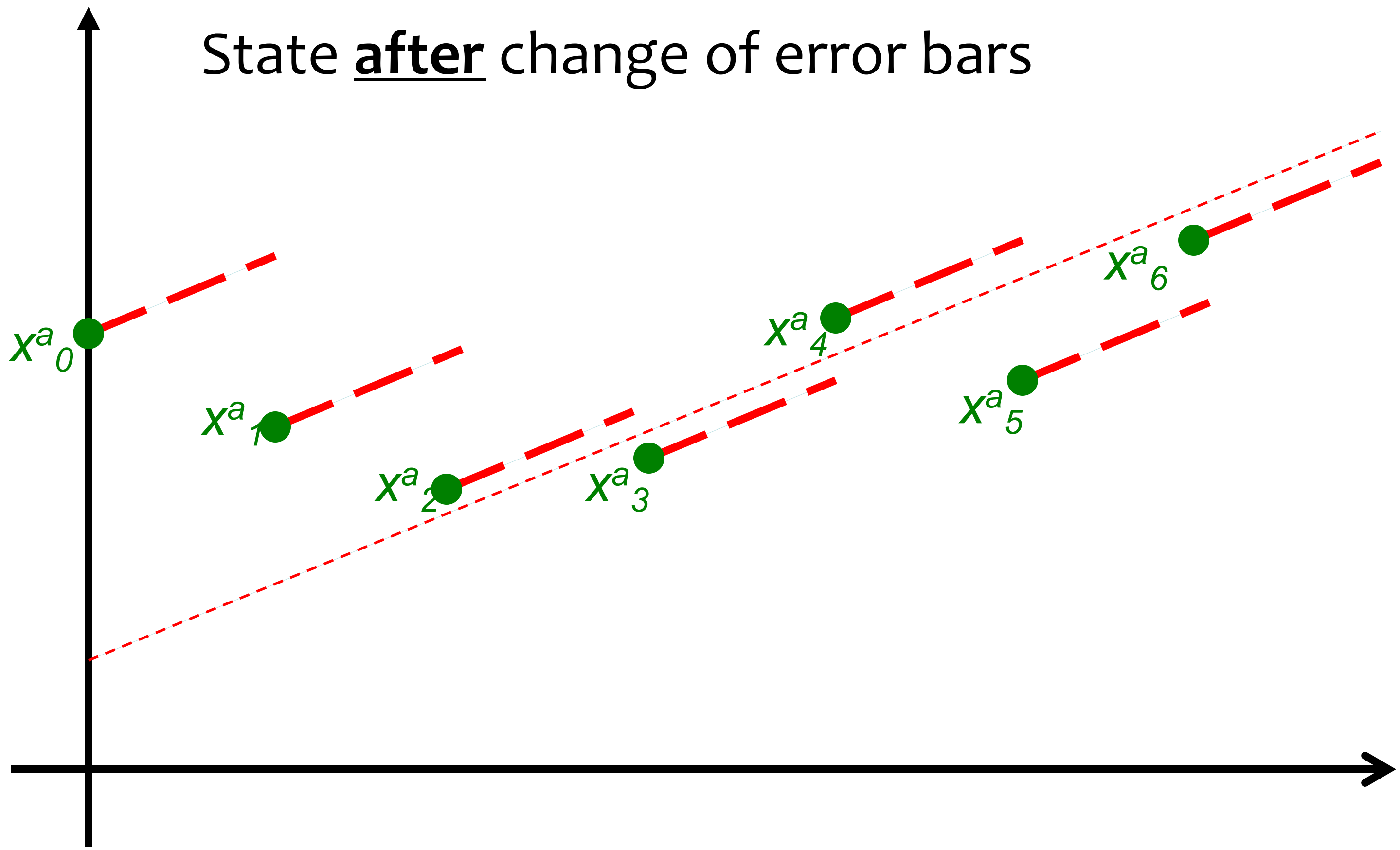
- The role/importance of error/uncertainty estimates!
- What if we reduce observation errors?



NOTE:

- It's *easy* to **over-fit the data** if error bars are unrealistically small!
- But is it *good*?

State estimation via filtering (sequential) methods



NOTE:

- It's *easy* to **over-fit the data** if error bars are unrealistically small!
- But is it *good*?

"Analysis" and "re-analysis" in Numerical Weather Prediction

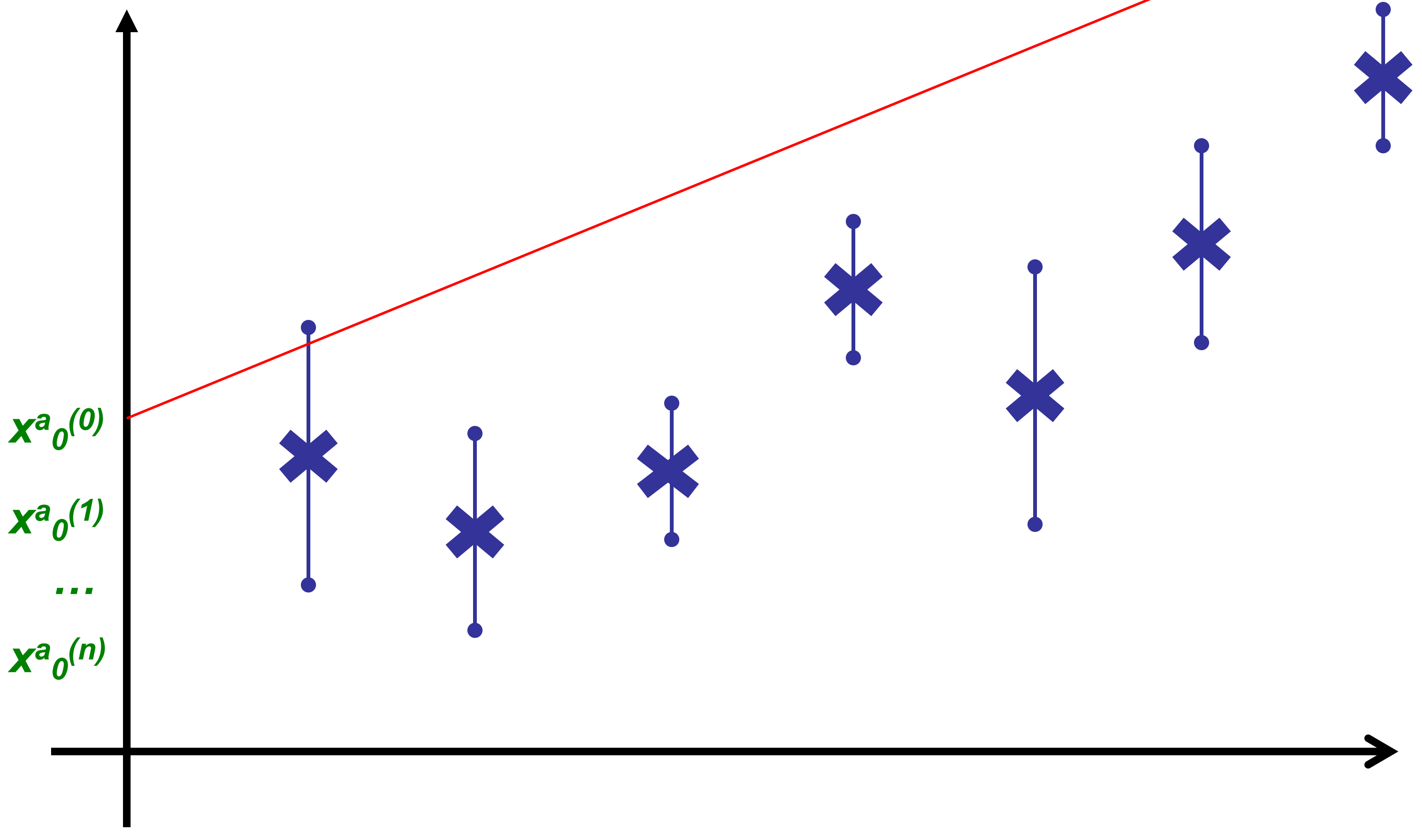
- **<u>"Analysis"</u>** is done in operational (real-time) mode
  - not all observations available in time (< 20%?)
  - forecast model changes over time (e.g., resolution, …)
- **<u>"Re-analysis"</u>** consists of:
  - redoing the forecast/analysis steps over extended period
  - use the same model
  - use all observations (incl. delayed-mode)
- Current global re-analyses:
  - ECMWF/ERA-Interim, NASA/MERRA, JRA-25/55, NCEP-CFSR, CMC-GDPS, NOAA/20CR, …
  - e.g.: *Lindsay et al.; Chaudhuri et al. (both J. Clim., 2014)*
- Regional Arctic high-res. re-analyses:
  - Arctic System Reanalysis (ASR), PIOMAS (ocean/ice), …

Various implementations & approximations:

- Nudging
- Relaxation
- Successive correction
- Optimal (or statistical) interpolation
- 3D-Var
- …

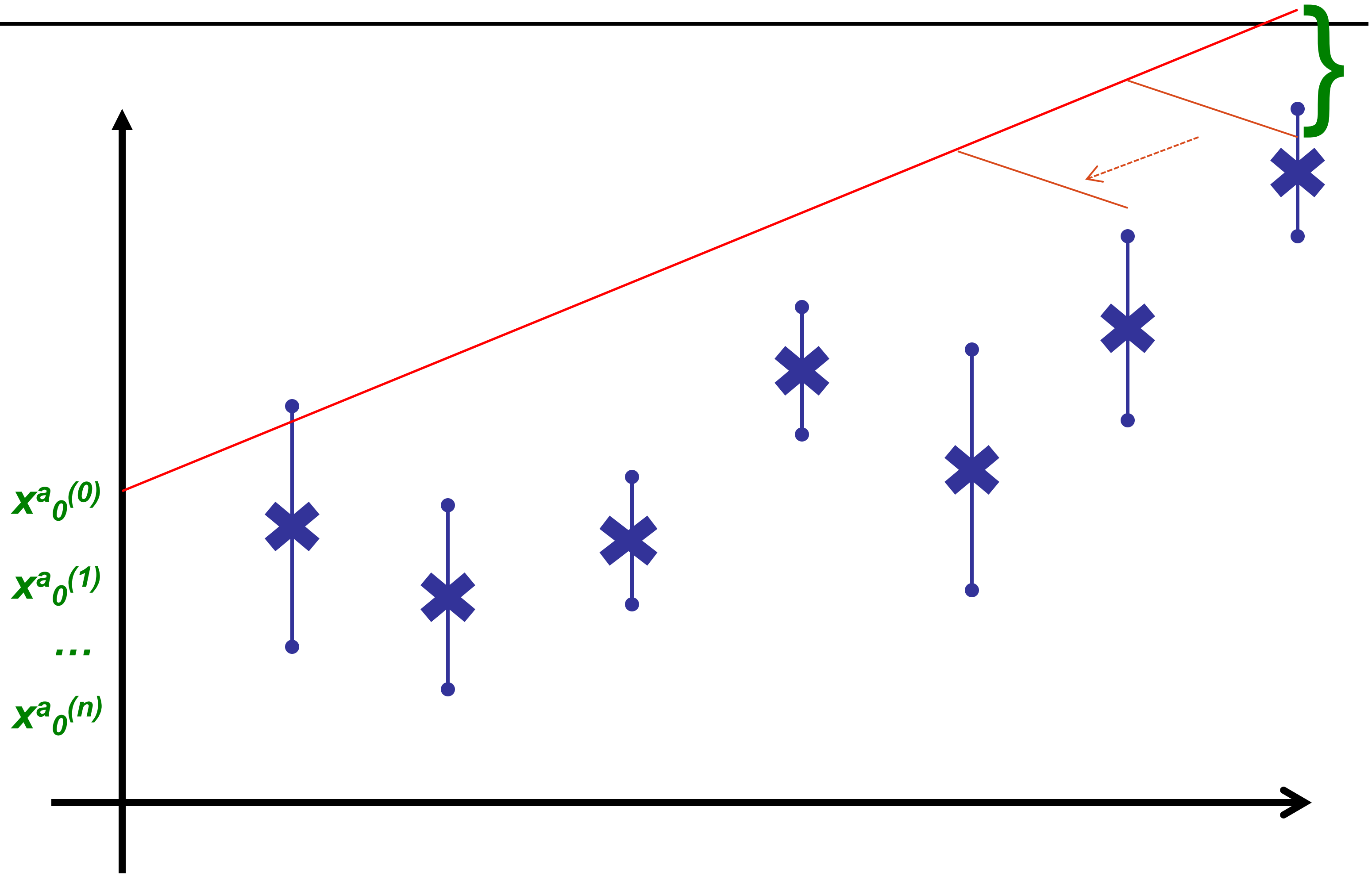
Lots of computational science & engineering involved to make it work

State estimation via smoother (adjoint) method



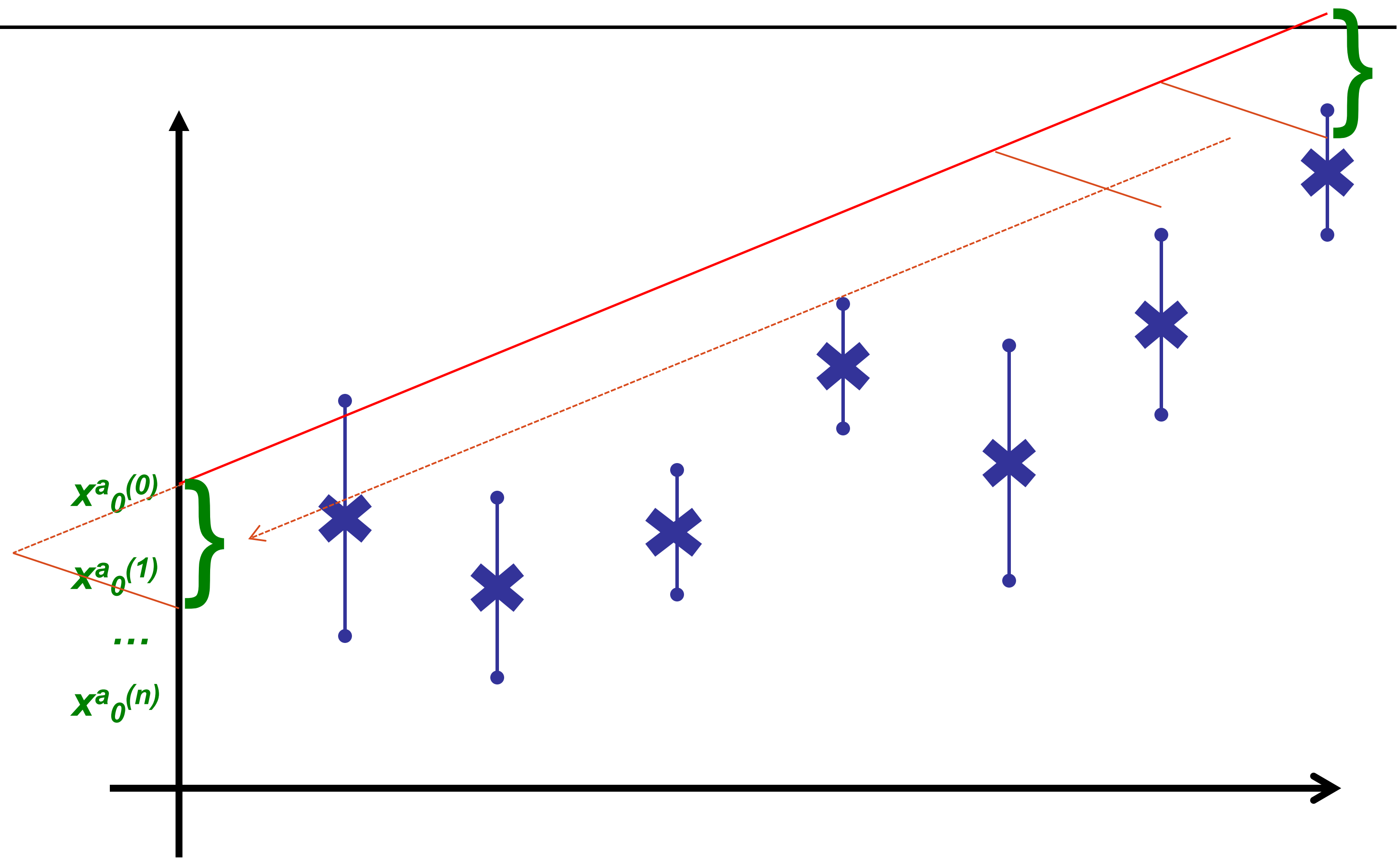
- Starting from first guess solution with initial condition: $x^a{}_0{}^{(0)}$, iteratively vary $x^a{}_0$, such as to minimize model data misfit
- Optimal solution obtained from initial condition $x^a{}_0{}^{(n)}$

State estimation via smoother (adjoint) methods



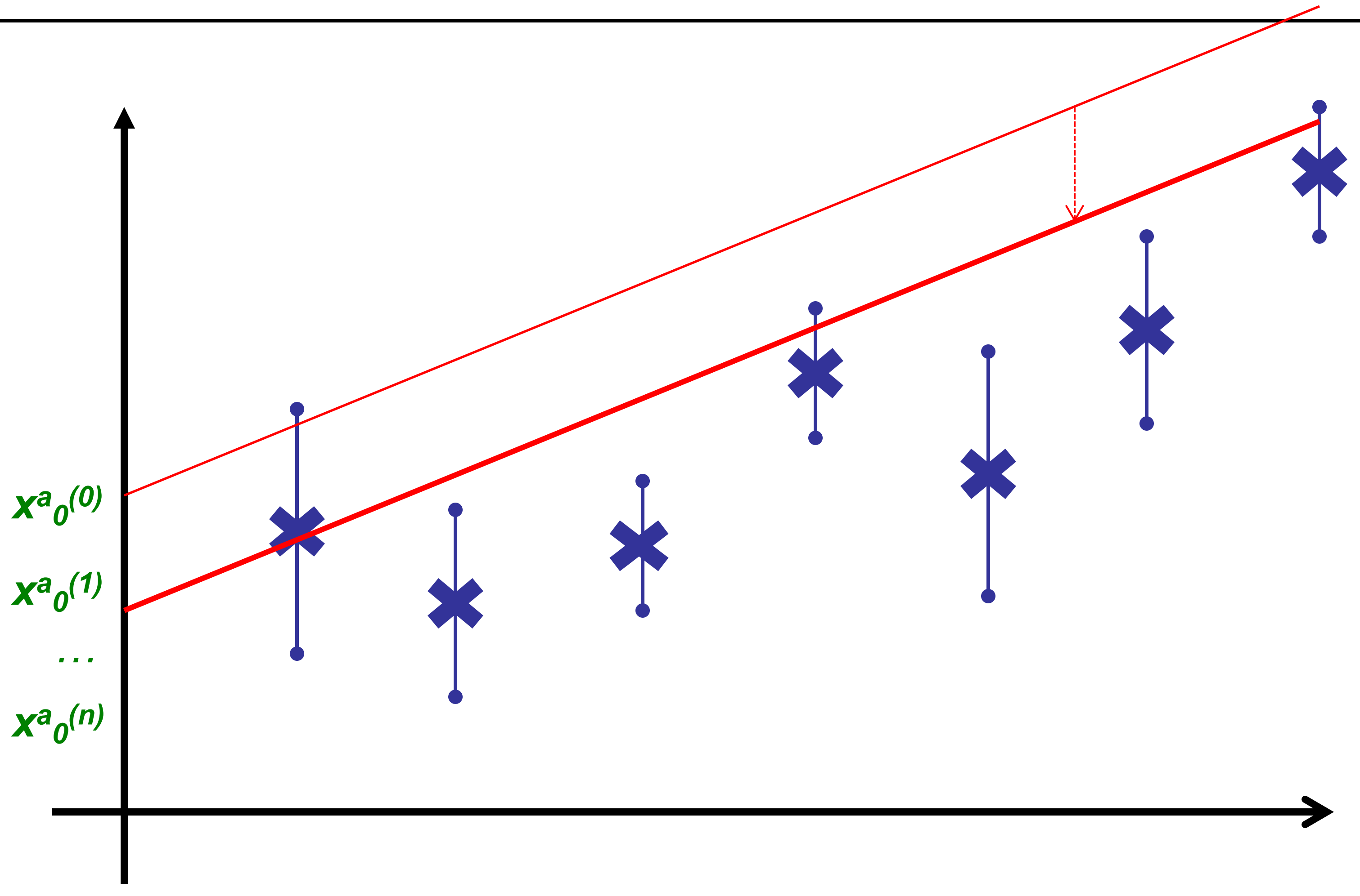
- Entire model trajectory is adjusted simultaneously
- Sensitivity of misfit cost function to previous states is carried (and accumulated) *backward in time* by the **adjoint model**

State estimation via smoother (adjoint) methods



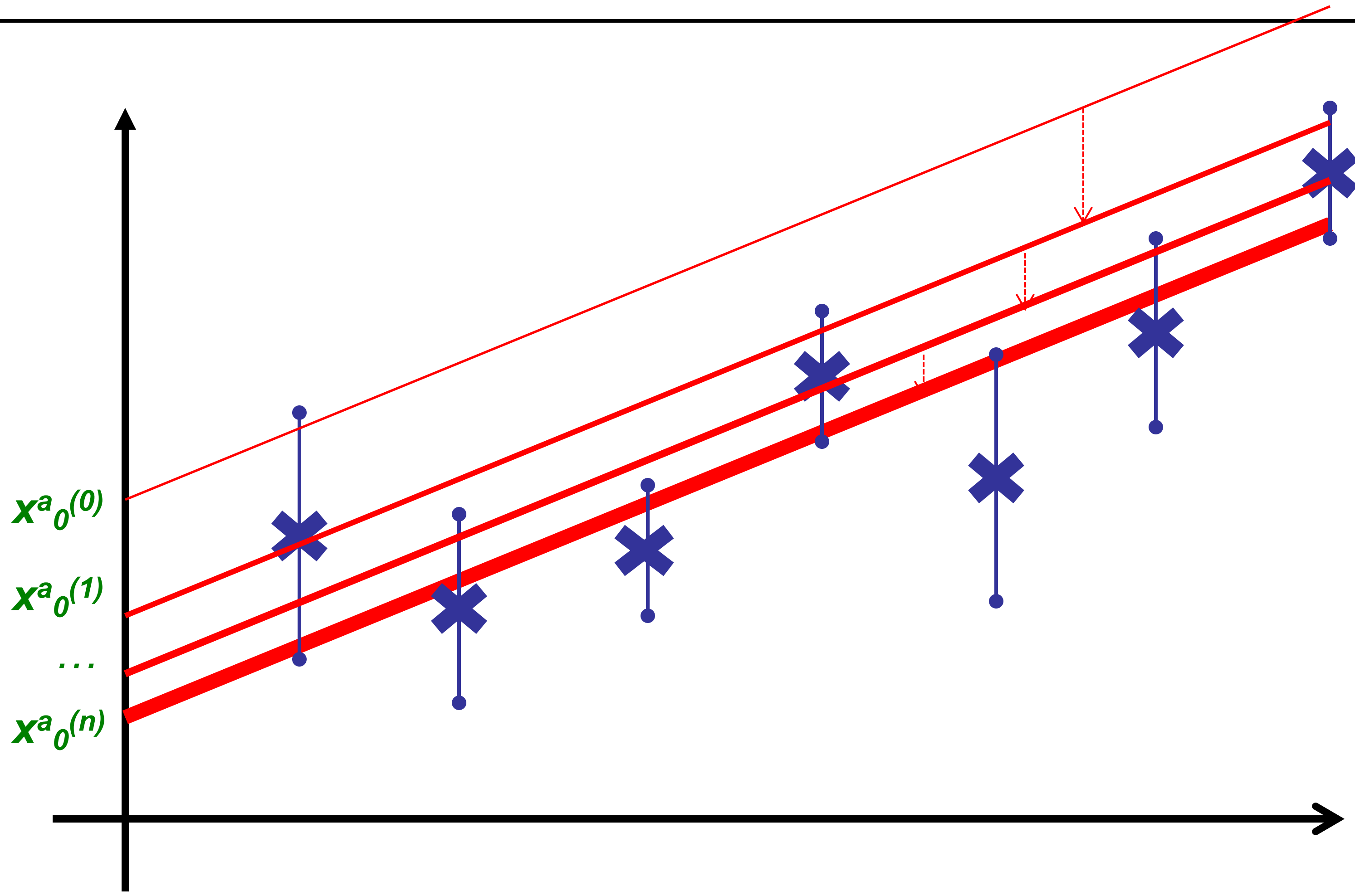
- Entire model trajectory is adjusted simultaneously
- Sensitivity of misfit cost function to previous states is carried (and accumulated) *backward in time* by the **adjoint model**

State estimation via smoother (adjoint) method



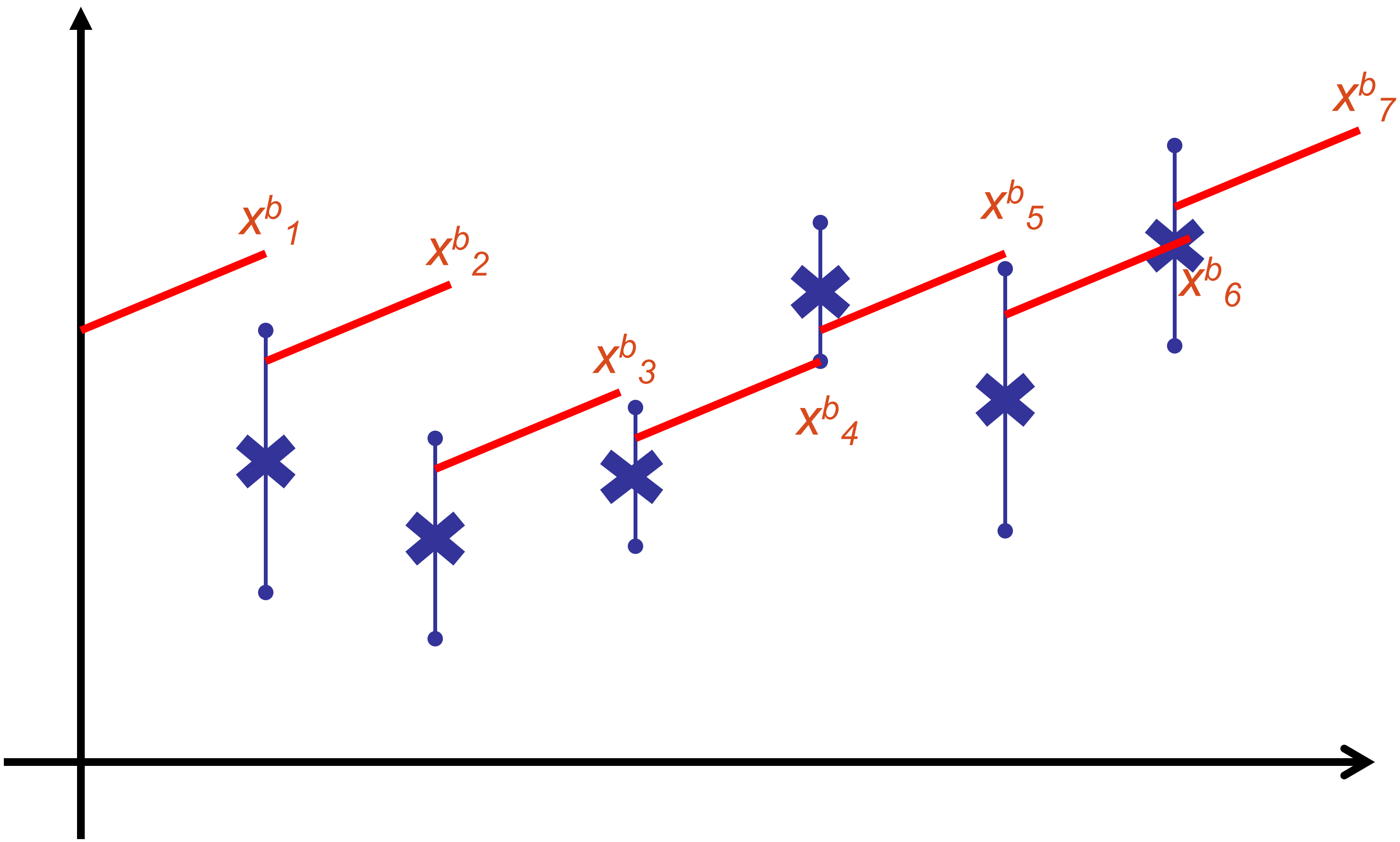
- Optimal initial condition $x^a_0{}^{(n)}$ obtained *iterative*ly
- Need to vary J with respect to x^a_o
  - **gradient-based optimization!**

State estimation via smoother (adjoint) method



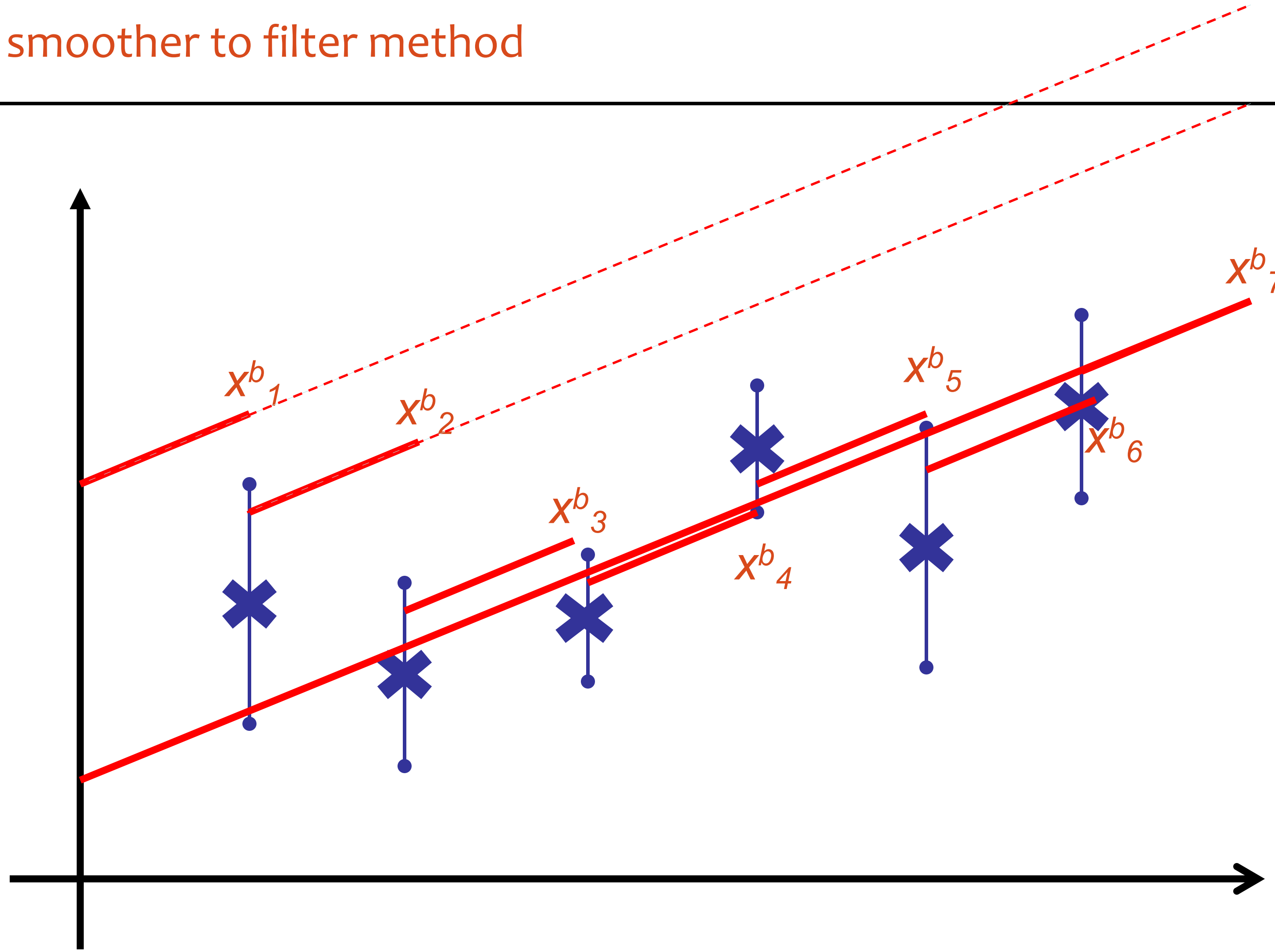
- Optimal initial condition $x^a_0{}^{(n)}$ obtained *iteratively*
- Need to vary J with respect to x^a_0
  - **gradient-based optimization!**

Compare smoother to filter method



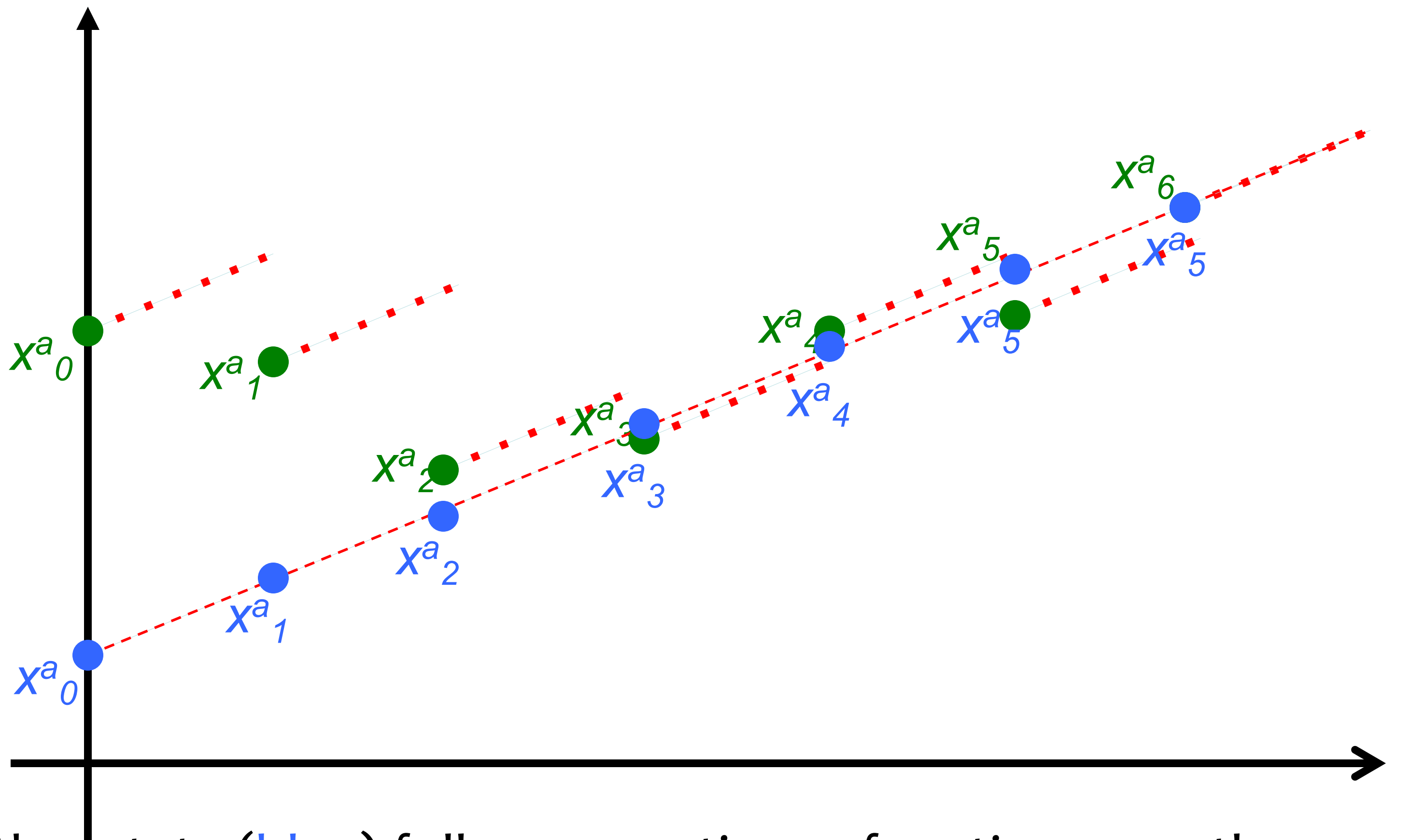
- <u>Filter</u> can only propagate information content from observations *forward in time*
- <u>Smoother</u> uses *past, present, and future* observation combined!

Compare smoother to filter method



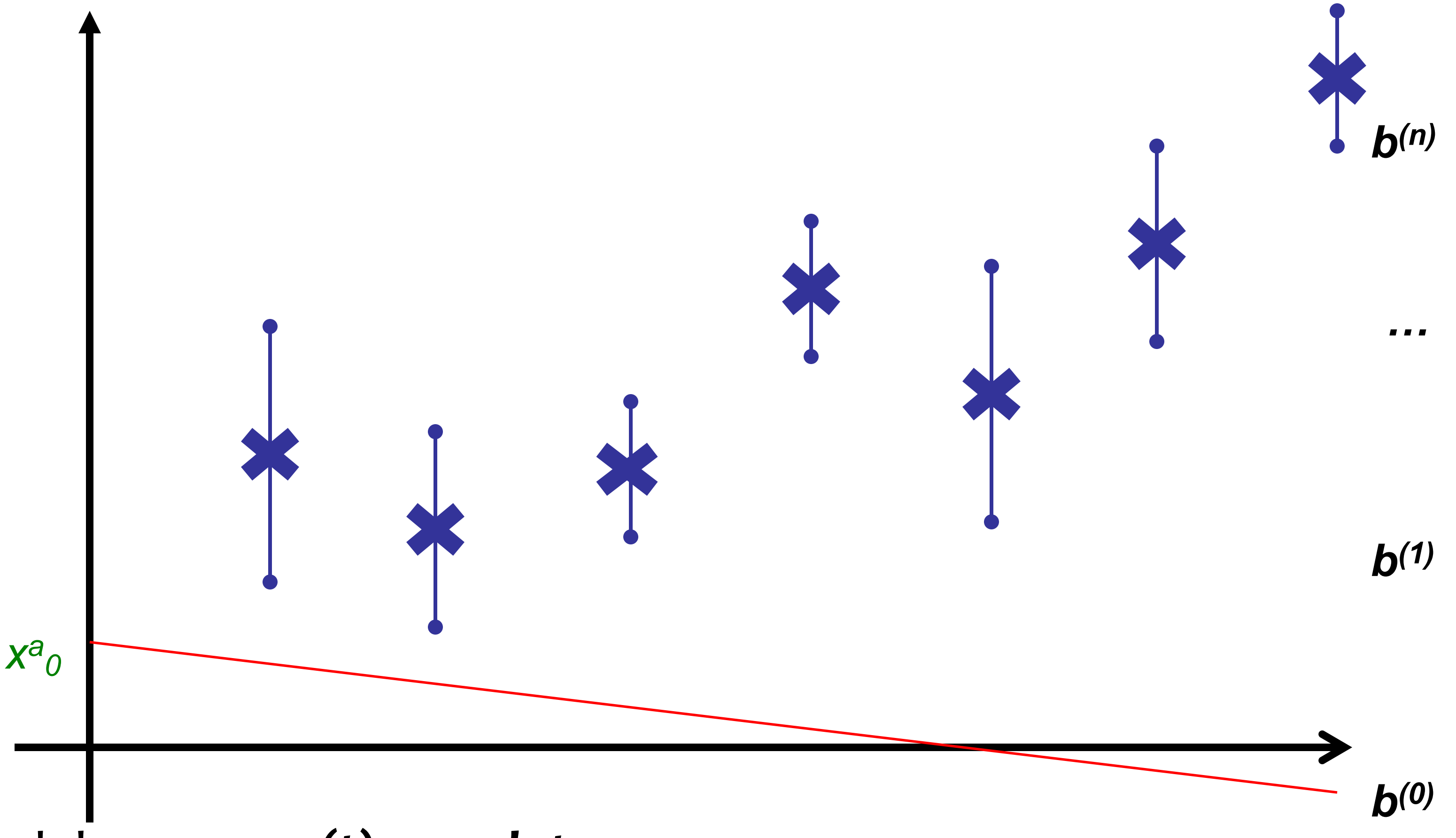
- <u>Filter</u> can only propagate information content from observations *forward in time*
- <u>Smoother</u> uses *past, present, and future* observation combined!

Compare smoother to filter method



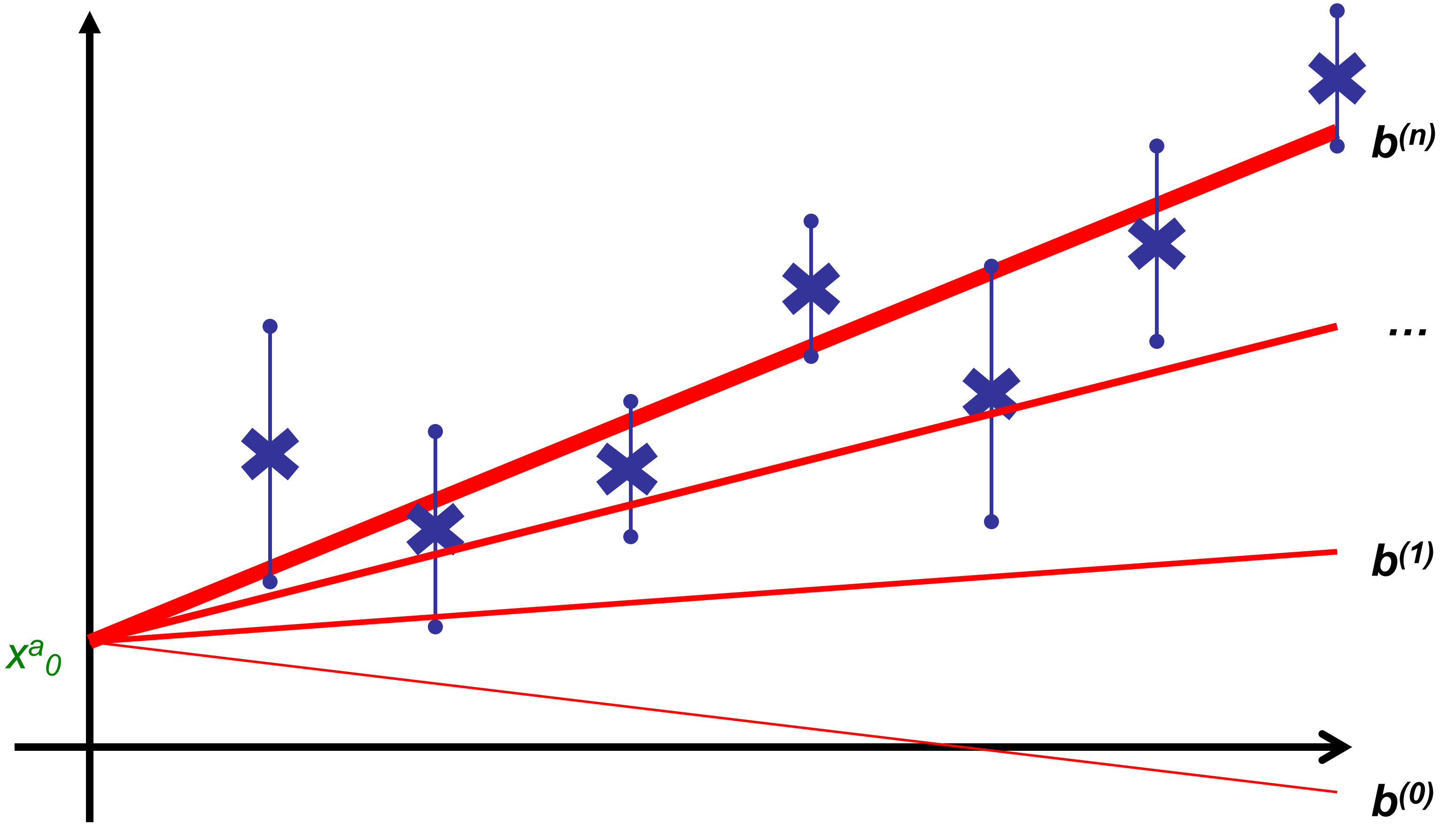
- <u>Smoother state </u>(blue) follows equations of motions exactly
  - tendency/trends (*dx/dt*) physically realistic
- <u>Filter state </u>(green) fits observations better
  - validity of tendency/trends unclear

Parameter estimation via smoother (adjoint) method



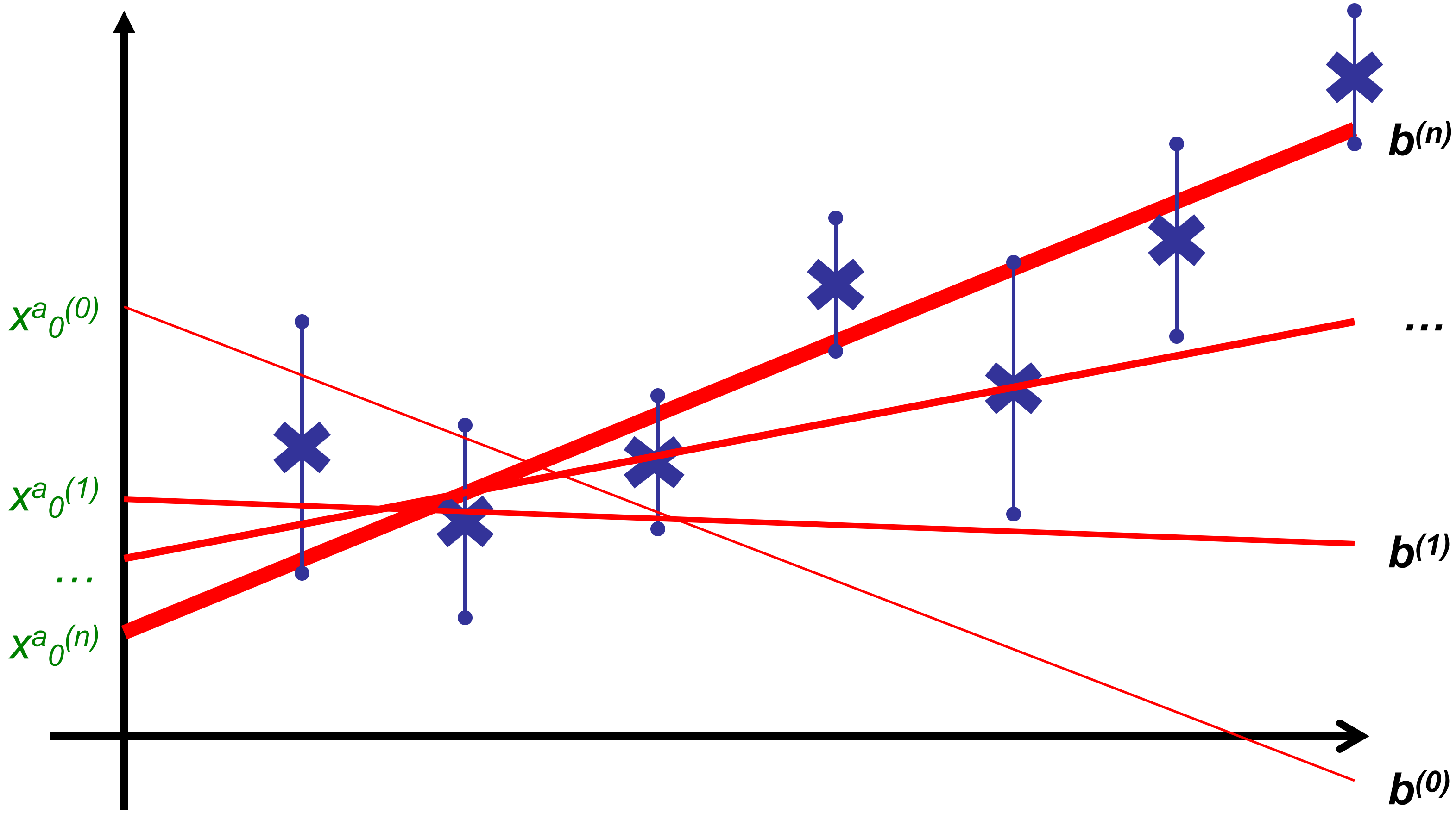
- for model $x(t) = a + b\,t$
  - instead of varying initial condition $x^a(0) = a$,
  - vary slope (i.e. "model parameter") b

<u>Parameter</u> estimation via smoother (adjoint) method



- for model $x(t) = a + b\,t$
  - instead of varying initial condition $x^a(0) = a$,
  - vary slope (i.e. "model parameter") b

<u>**Joint state & parameter**</u> estimation via smoother (adjoint) method

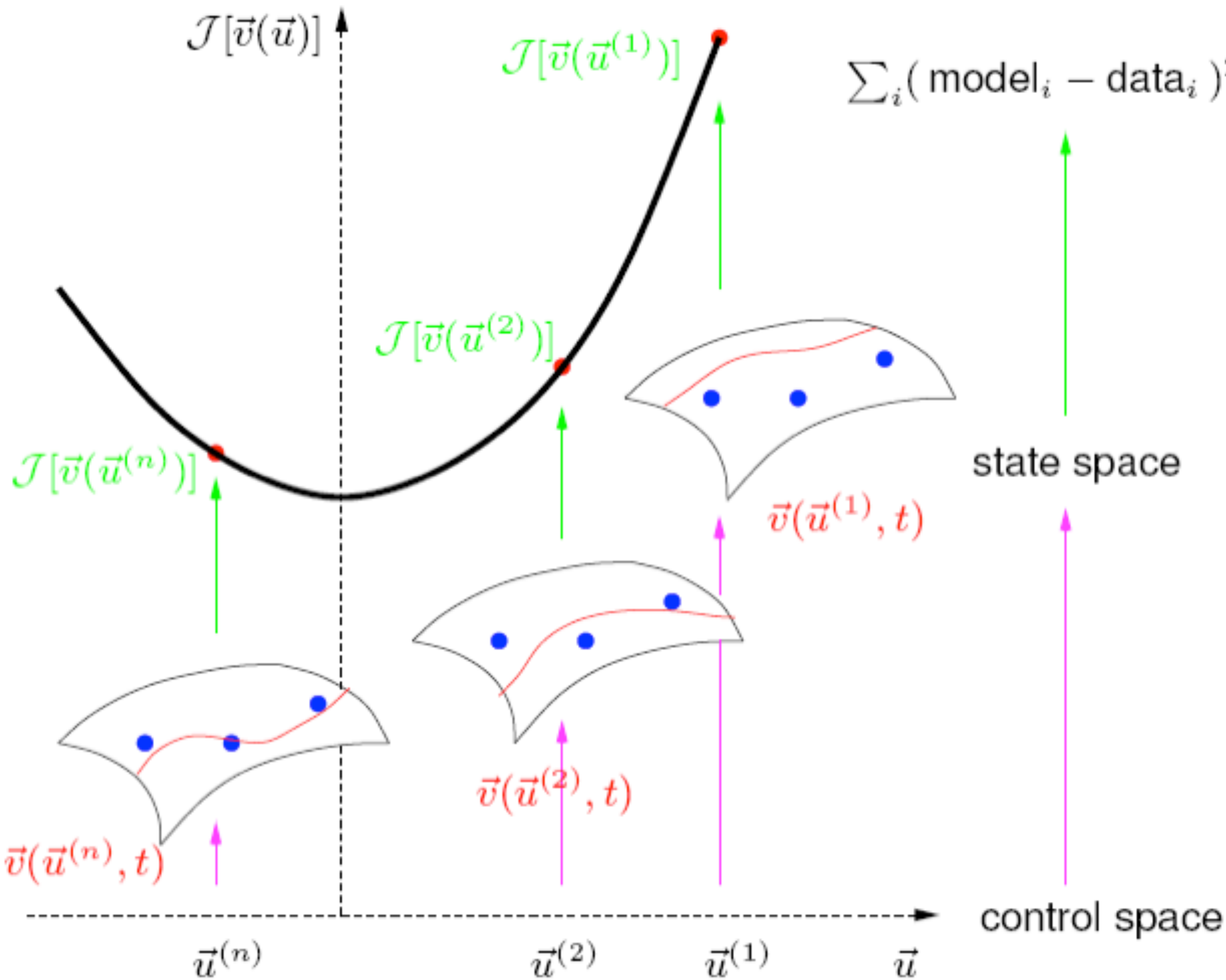


- for model $x(t) = a + b\,t$, <u>*simultaneously vary*</u>
  - initial condition (i.e. "model state") $x^a(0) = a$
  - and slope (i.e. "model parameter") b

END

An optimal estimation/control approach

Iterative optimization via ***gradient*** obtained from ***adjoint*** *model*



The smoother (reconstruction) problem

Consider *perfect* model **L** (i.e., $\eta = 0$), and obs. y with noise ϵ:

$$x_{k+1}^t = \mathbf{L} x_k^t$$
$$y_{k+1} = \mathbf{E} x_{k+1}^t + \epsilon_{k+1}$$

Variational form of least-squares estimation problem:

$$J(x) = \sum_{0 \le k \le N} [\mathbf{E} x_k - y_k]^T \mathbf{R}^{-1} [\mathbf{E} x_k - y_k]$$

Extend to Lagrange function $\mathcal{L}$, Lagrange multipliers μ_k:

$$\mathcal{L}(x, \mu) = J(x) + \sum_{0 \le k \le N} \mu_k^T [x_{k+1} - \mathbf{L} x_k]$$

The smoother (reconstruction) problem

Lagrange multiplier method:

Stationary point of $\mathcal{L}$ leads to set of normal equations:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}(t)} = \mathbf{x}(t) - L\left[\mathbf{x}(t-1)\right] = 0 \qquad 1 \leq t \leq t_f$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)} = \frac{\partial J_0}{\partial \mathbf{x}(t)} - \boldsymbol{\mu}(t) + \left[\frac{\partial L[\mathbf{x}(t)]}{\partial \mathbf{x}(t)}\right]^T \boldsymbol{\mu}(t+1) = 0 \qquad 0 < t < t_f$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t_f)} = \frac{\partial J}{\partial \mathbf{x}(t_f)} - \boldsymbol{\mu}(t_f) = 0 \qquad t = t_f$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}(0)} = \frac{\partial J}{\partial \mathbf{x}(0)} - \left[\frac{\partial L[\mathbf{x}(0)]}{\partial \mathbf{x}(0)}\right]^T \boldsymbol{\mu}(1) \qquad t_0 = 0$$

The smoother (reconstruction) problem

"Variational" hints that we need a gradient:

- gradient of J with respect to *independent* or *control* variables!
- Here: Vary initial conditions, x_0 such as to minimize J

BUT: J depends not just on x_0, but on all $x_k, \ k \geq 0$.

- consider *nonlinear model* $x_{k+1} = L(x_k)$
- linearized version is *state transition matrix* $\mathbf{L}$

$$\delta x_{k+1} = \frac{\partial x_{k+1}}{\partial x_k} \delta x_k = \mathbf{L} \, \delta x_k$$

Need chain rule of differentiation:

$$\begin{aligned} J &= J(x_0, x_1, x_2, \ldots, x_N) \\ &= J\Big(x_0, L(x_0), L(L(x_0)), \ldots, L^N(x_0)\Big) \end{aligned}$$

The smoother (reconstruction) problem

$$\mu_0 = \frac{\partial J}{\partial x_0} = \sum_{1 \le k \le N} \frac{\partial x_k}{\partial x_0} \left(\frac{\partial J}{\partial x_k} \right)$$

$$= \frac{\partial x_1}{\partial x_0} \left(\frac{\partial J}{\partial x_1} \right) + \frac{\partial x_1}{\partial x_0} \frac{\partial x_2}{\partial x_1} \left(\frac{\partial J}{\partial x_2} \right) + \ldots + \frac{\partial x_1}{\partial x_0} \cdots \frac{\partial x_N}{\partial x_{N-1}} \left(\frac{\partial J}{\partial x_N} \right)$$

$$= \mathbf{L}^T \frac{\partial J}{\partial x_1} + \mathbf{L}^T \mathbf{L}^T \frac{\partial J}{\partial x_2} + \ldots + \mathbf{L}^T \cdots \mathbf{L}^T \frac{\partial J}{\partial x_N}$$

$\mathbf{L}^T$: is the **adjoint model** (and $\mathbf{L}$ is the **tangent linear model**)

$\mu_k = \left(\frac{\partial J}{\partial x_k} \right)$: **Lagrange multipliers** or **gradients**

The smoother (reconstruction) problem

For intermediate step of the adjoint model integration one obtains:

$$
\begin{aligned}
\mu_k = \frac{\partial J}{\partial x_k} &= \mathbf{L}^T \frac{\partial J}{\partial x_{k+1}} + \mathbf{E}^T \mathbf{R}^{-1} \left[\mathbf{E} x_k - y_k \right] \\
&= \mathbf{L}^T \left(\mathbf{L}^T \frac{\partial J}{\partial x_{k+2}} + \mathbf{E}^T \mathbf{R}^{-1} \left[\mathbf{E} x_{k+1} - y_{k+1} \right] \right) \\
&\qquad + \mathbf{E}^T \mathbf{R}^{-1} \left[\mathbf{E} x_k - y_k \right]
\end{aligned}
$$

- The adjoint model $\mathbf{L}^T$ propagates μ_k (the sensitivity of J with respect to all earlier states x_k) backward in time to x_0;
- Each model–data misfit (i.e. innovation vector $\mathbf{E} x_k - y_k$) is a *source* of sensitivity;
- The gradient of J with respect to x_0 takes into account (and weighs) the size of *all* misfit terms, *all* (inverse) error covariances, and *all* (linearized) model dynamics.

Conclusions

- DA seeks to optimally combine information content in observations, models, and their uncertainties!
- DA can mean very(!) different things to different people
- Depending on application, different methods warranted:
  - forecasting: *filter methods* (e.g., Kalman filter)
  - reconstruction: *smoother methods* (adjoint method)
- formal estimation methods to synthesize the diverse & sparse observations seems important for climate reconstructions
  - it is feasible,
  - simply copying NWP approaches not always useful,
  - remains a challenge for the time to come