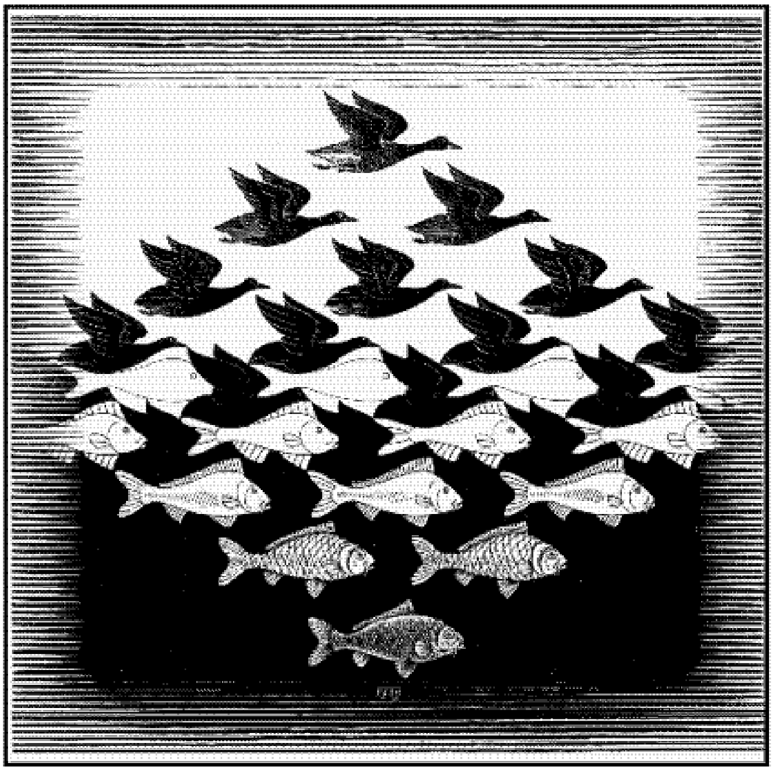
Some Fundamentals of

INVERSE MODELLING

Ross Bannister
Room 2L49

D.A.R.C.

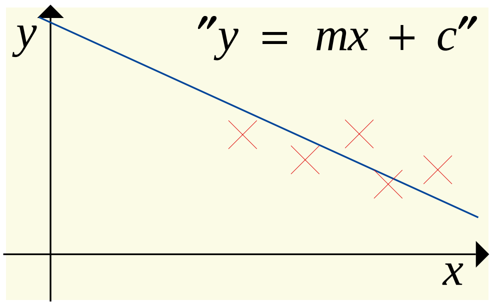
University of Reading



'Sky and Water I', M.C. Escher, 1938

What does an inverse model do?

'Forward' model



Model parameters: *m* and *c*

$$\begin{pmatrix} m \\ c \end{pmatrix} \quad '+' \quad x \rightarrow y$$

'Inverse' model

$$\begin{pmatrix} y_1,\ x_1 \\ y_2,\ x_2 \\ \dots \\ \dots \\ y_n,\ x_n \end{pmatrix} \text{(+ errors)} \rightarrow \begin{pmatrix} m \\ c \end{pmatrix} \text{(+ errors)}$$

Examples of inverse modelling ...

Exact inversion techniques:

- Matrix inversion
- PV inversion
- Abel's integration equation
- Newton-Raphson method

Inexact inversion applications:

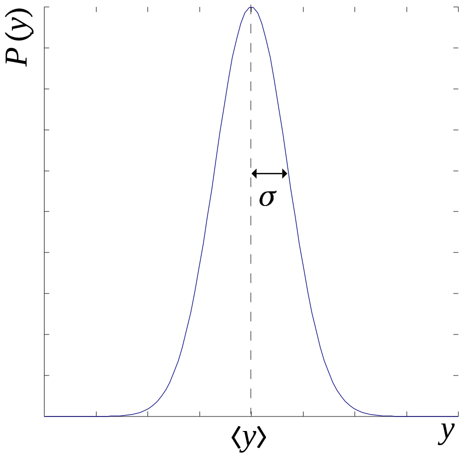
- Data assimilation
- Satellite retrieval
- Medical Imaging
- Geology
- Astronomy
- Solar physics
- Missile interception

... and any situation in where:

- observations are noisy,
- observations are incomplete and irregular,
- parameters cannot be measured directly.

Parameter Estimation by Maximum Likelihood
(Method of Least Squares)

Gaussian error characteristics (one variable)

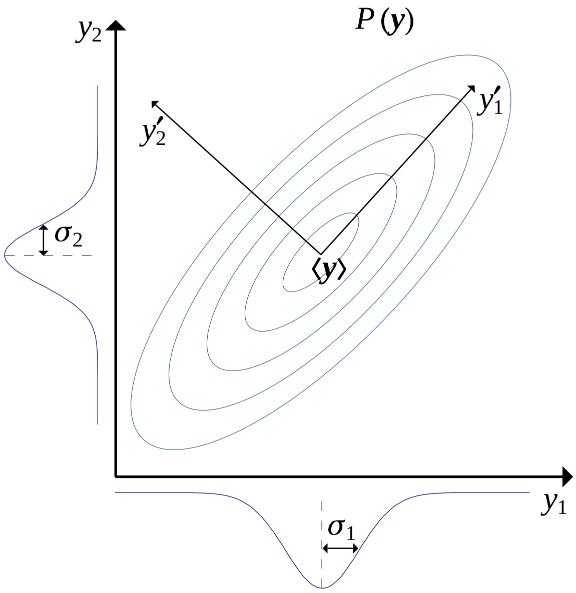


$$P(y) \propto \exp\left(-\frac{(y - \langle y \rangle)^2}{2\sigma^2}\right)$$

Mean : $< y >$

Variance : σ^2

(two or more variables)



$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$P(\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{y} - \langle \boldsymbol{y} \rangle)^T \mathbf{R}^{-1} (\boldsymbol{y} - \langle \boldsymbol{y} \rangle)\right)$$

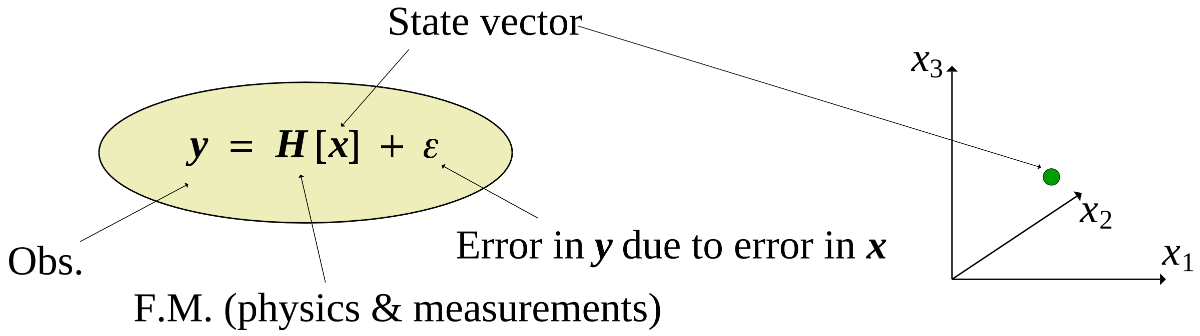
$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}(y_1, y_2) \\ \mathrm{cov}(y_1, y_2) & \sigma_2^2 \end{pmatrix}$$

$$\mathbf{URU}^T = \begin{pmatrix} \lambda'_1 & 0 \\ 0 & \lambda'_2 \end{pmatrix}$$

Ingredients (for an inversion)

1. Observations	$\boldsymbol{y}$	$P(\boldsymbol{y})$	$\mathbf{R}$
2. A-priori (background)	$\boldsymbol{x}_B$	$P(\boldsymbol{x})$	$\mathbf{B}$
3. Constraints ...			

The forward model (strong constraint)



How will 1,2,3 combine to give the most likely set of parameters?

Bayes' Theorem:

$$\boldsymbol{y}) \;\propto\; P(\boldsymbol{x})\,P(\boldsymbol{y} \mid \boldsymbol{x})$$

$$\propto\; \exp\left(-\frac{1}{2}\eta^T\mathbf{B}^{-1}\eta\right)\exp\left(-\frac{1}{2}\varepsilon^T\mathbf{R}^{-1}\varepsilon\right)$$

$$\propto\; \exp-\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_B)^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_B)+\frac{1}{2}(\boldsymbol{H}[\boldsymbol{x}]-\boldsymbol{y})^T\mathbf{R}^{-1}(\boldsymbol{H}[\boldsymbol{x}]-\boldsymbol{y})\right)$$

η : error in a-priori $= \boldsymbol{x} - \boldsymbol{x}_B$

ε : error in model's guess $= \boldsymbol{y} - \boldsymbol{H}[\boldsymbol{x}]$

Maximum likelihood $\Rightarrow$ minimum penalty

$$J[\boldsymbol{x}] \;=\; \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_B)^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_B)+\frac{1}{2}(\boldsymbol{H}[\boldsymbol{x}]-\boldsymbol{y})^T\mathbf{R}^{-1}(\boldsymbol{H}[\boldsymbol{x}]-\boldsymbol{y})$$

$$\boldsymbol{x}\big|_{\min J} \;=\; \boldsymbol{x}_A \text{ (analysed parameters)}$$

Notes on the cost function

$$J[\mathbf{x}] = \frac{1}{2}(\mathbf{x} - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_B) + \frac{1}{2}(\boldsymbol{H}[\mathbf{x}] - \boldsymbol{y})^T \mathbf{R}^{-1} (\boldsymbol{H}[\mathbf{x}] - \boldsymbol{y})$$

For n unknown parameters in $\boldsymbol{x}$, and m observations in $\boldsymbol{y}$,

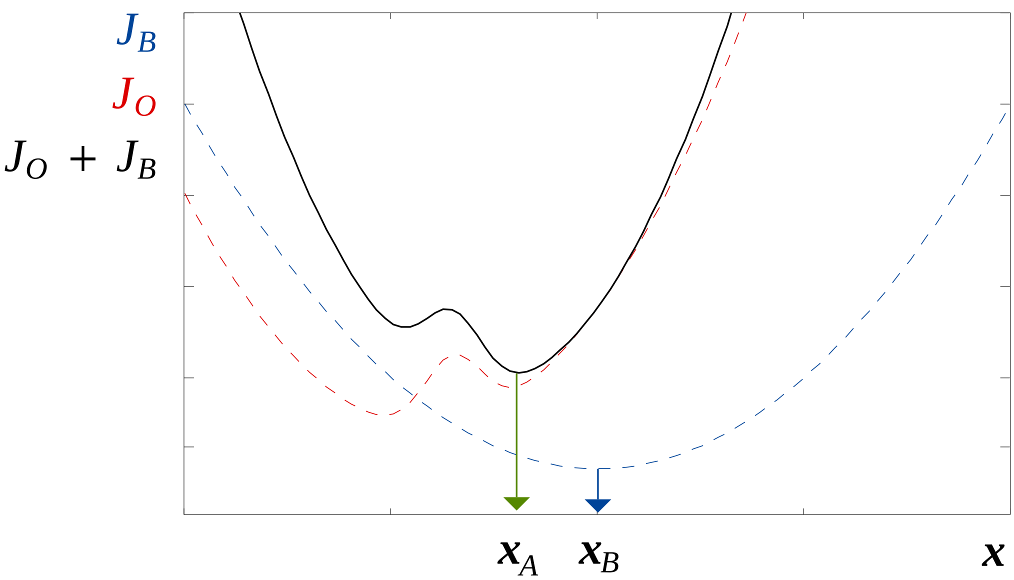
$\mathbf{B}$ $\quad n \times n$

$\mathbf{R}$ $\quad m \times m$

need $\geqslant n$ pieces of independent information

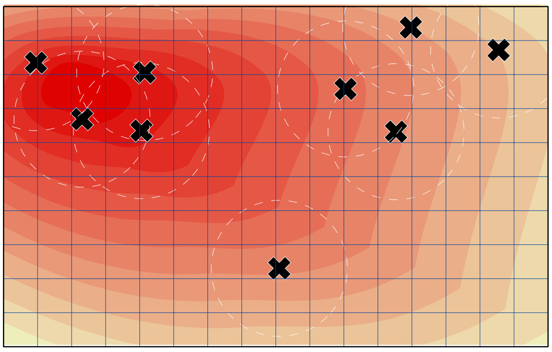
Why "least squares"?

Eg. if $\boldsymbol{H}[\mathbf{x}]$ is non-linear ...



Methods of Inverting

1. Cressman Analysis



- Cheap.
- Easy to implement.
- A-priori in data-poor regions.
- No account of errors.
- Not dynamically consistent.
- Direct observations only.
- No forward model used.

2. Best Linear Unbiased Estimator (BLUE)

$$\nabla J = 0 \ :$$

$$\boldsymbol{x}_A = \boldsymbol{x}_B + \mathbf{K}\,(\boldsymbol{y} - \boldsymbol{H}[\boldsymbol{x}_B])$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\,(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

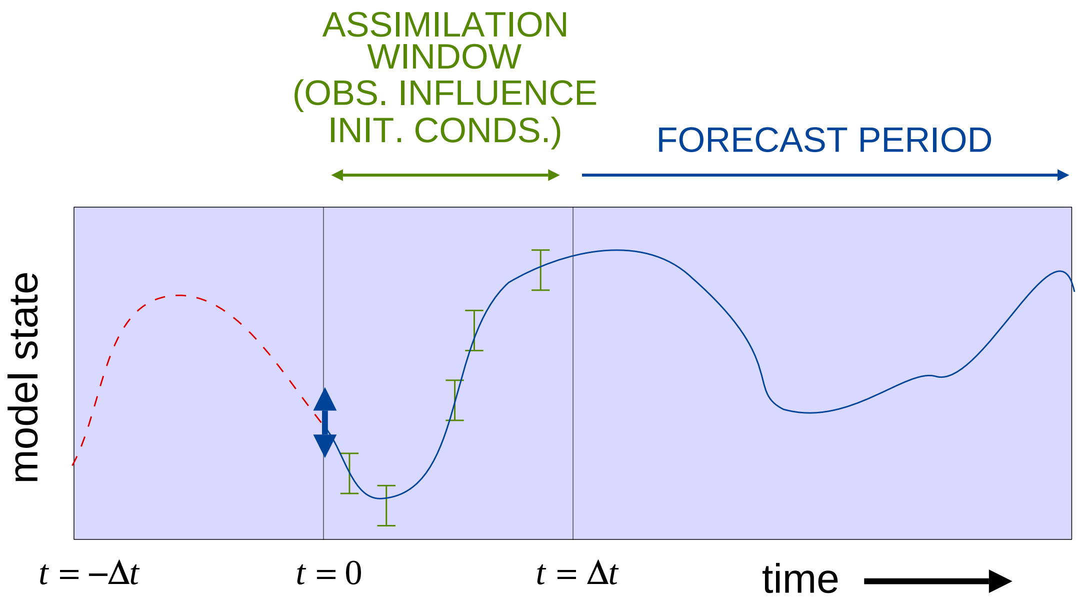
with error covariance

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

- Account taken of errors.
- Can use indirect obs.
- A-priori in data poor regions.
- Difficult to know **B**.
- ☆ Difficult to use with non-linear operators.
- ☆ Expensive for large Nos. of degrees of freedom.

3. Variational Analysis (4d-Var)

$$J[x] = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_B)^T \mathbf{B}^{-1} (\boldsymbol{x} - \boldsymbol{x}_B) + \frac{1}{2}(\boldsymbol{H}[\boldsymbol{x}] - \boldsymbol{y})^T \mathbf{R}^{-1} (\boldsymbol{H}[\boldsymbol{x}] - \boldsymbol{y})$$



- Account taken of errors.
- Can use indirect obs. and non-linear operators.
- Suitable for large Nos. of degrees of freedom.
- 4d-Var is dynamically consistent.

- A-priori in data poor regions.
- Difficult to know **B**.
- Expensive.
- Difficult to implement and use.
- Needs preconditioning.

4. Kalman filter

- Account taken of errors.
- Can use indirect obs.
- A-priori in data poor regions.
- Evolves **B** in time.

- Difficult to use with non-linear operators.
- Very expensive.
- Difficult to use practically.

Example with BLUE

1 unknown parameter, 1 observation, 1 initial estimate

	Value	Uncertainty
Prior estimate	x_B	$\mathbf{B} = \sigma_B^2$
Observation	y	$\mathbf{R} = \sigma_y^2$
Forward model	$\mathbf{H} = \mathbf{I}$	—

BLUE formulae:

$$\boldsymbol{x}_A = \boldsymbol{x}_B + \mathbf{K}\,(\boldsymbol{y} - \boldsymbol{H}[\boldsymbol{x}_B])$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\,(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

$$\mathbf{K} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_y^2}$$

$$\begin{aligned} x_A &= \left(\frac{\sigma_B^2}{\sigma_y^2} + 1\right)^{-1} x_B + \left(\frac{\sigma_y^2}{\sigma_B^2} + 1\right)^{-1} y \\ &= 0 \times x_B + 1 \times \boldsymbol{y} && \lim_{\sigma_B/\sigma_y \to \infty} \\ &= 1 \times x_B + 0 \times \boldsymbol{y} && \lim_{\sigma_y/\sigma_B \to \infty} \end{aligned}$$

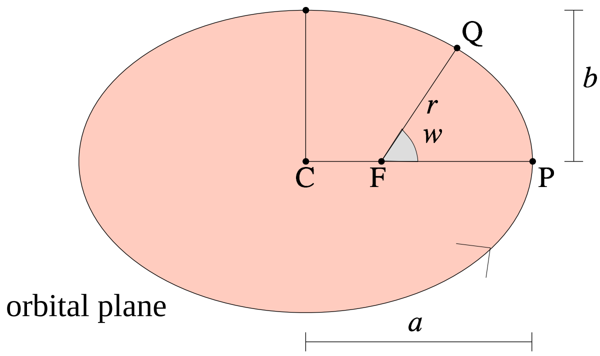
$$\begin{aligned} \mathbf{A} &= \frac{1}{\sigma_B^{-2} + \sigma_y^{-2}} \\ &= \sigma_y^2 && \lim_{\sigma_B/\sigma_y \to \infty} \\ &= \sigma_B^2 && \lim_{\sigma_y/\sigma_B \to \infty} \end{aligned}$$

Example with BLUE
(Astronomy - Inverting Kepler's Equation)

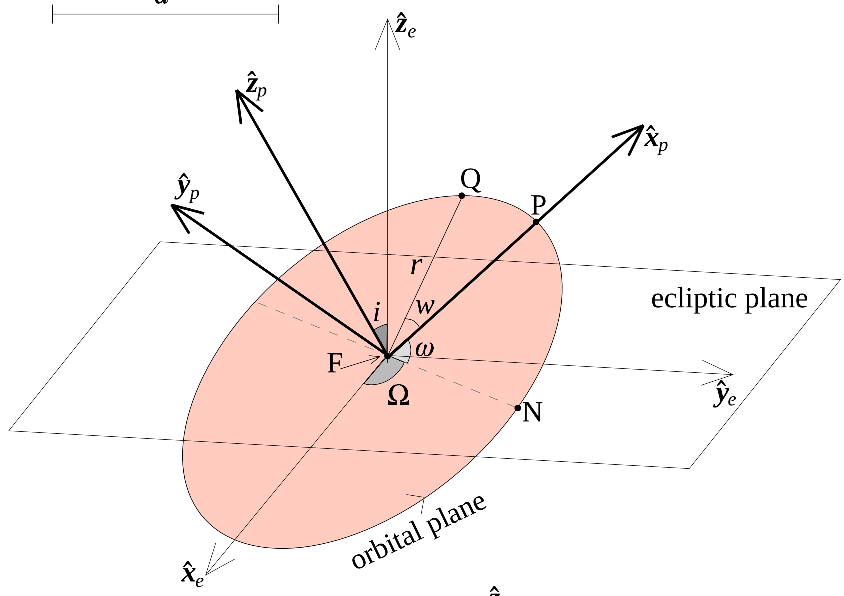
Want to determine orbital parameters:

$$\boldsymbol{x} = (a,\ e,\ i,\ \Omega,\ \varpi,\ \varepsilon)^T$$

Physics of the forward model:



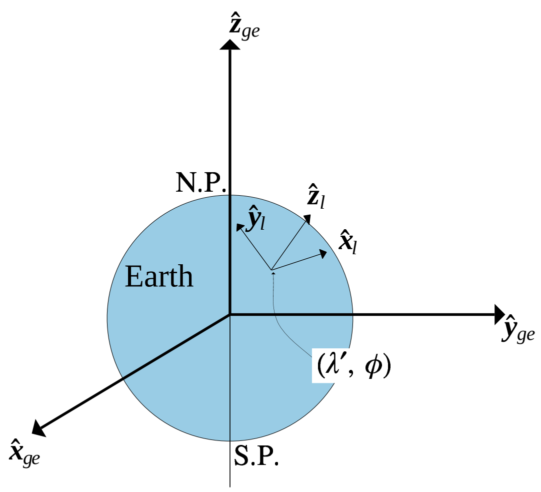
$$e^2 = 1 - \left(\frac{b}{a}\right)^2$$



No a-priori **Observations:**

$$\mathbf{B}^{-1} = 0$$

$$\boldsymbol{y} = \begin{pmatrix} \text{alt}_1 \\ \text{azi}_1 \\ \text{alt}_2 \\ \text{azi}_2 \\ \cdots \\ \cdots \end{pmatrix}$$

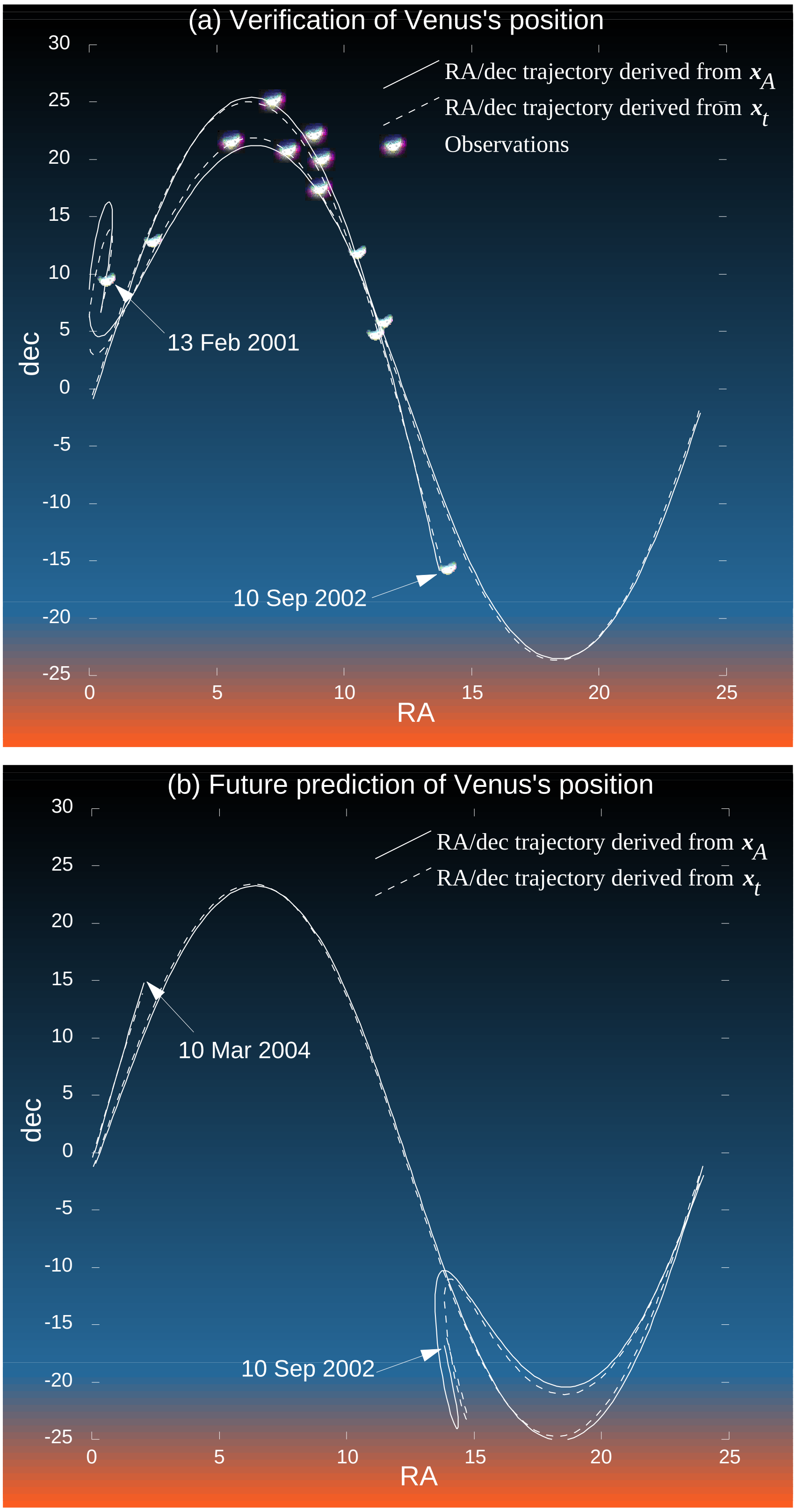


Inversion results

$$\boldsymbol{x}_A = (0{\cdot}7215,\ 0{\cdot}0121,\ 4{\cdot}40,\ 84{\cdot}2,\ 83{\cdot}3,\ 177{\cdot}1)^T$$

$$\text{error (1 s.d.)} = (0{\cdot}0010,\ 0{\cdot}0080,\ 0{\cdot}76,\ 4{\cdot}5,\ 20{\cdot}1,\ 2{\cdot}5)^T$$

$$\boldsymbol{x}_t = (0{\cdot}7233,\ 0{\cdot}0067,\ 3{\cdot}39,\ 76{\cdot}7,\ 131{\cdot}5,\ 182{\cdot}0)^T$$



4-Dimensional Variational Data Assimilation

Leith, 1993:

> ... the atmosphere "is a chaotic system in which errors introduced into the system can grow with time ... As a consequence, data assimilation is a struggle between chaotic destruction of knowledge and its restoration by new observations."

Cost function:

$$J[\boldsymbol{x}] = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_B)^T \mathbf{B}^{-1} (\boldsymbol{x} - \boldsymbol{x}_B) + \frac{1}{2}(\boldsymbol{H}[\boldsymbol{x}] - \boldsymbol{y})^T \mathbf{R}^{-1} (\boldsymbol{H}[\boldsymbol{x}] - \boldsymbol{y})$$

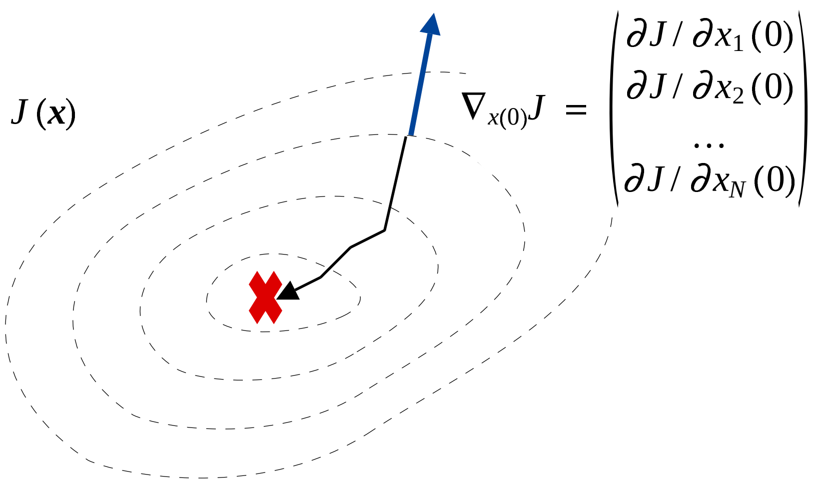
$$= \frac{1}{2}(\boldsymbol{x}(0) - \boldsymbol{x}_B(0))^T \ \mathbf{B}^{-1} \ (\boldsymbol{x}(0) - \boldsymbol{x}_B(0)) +$$

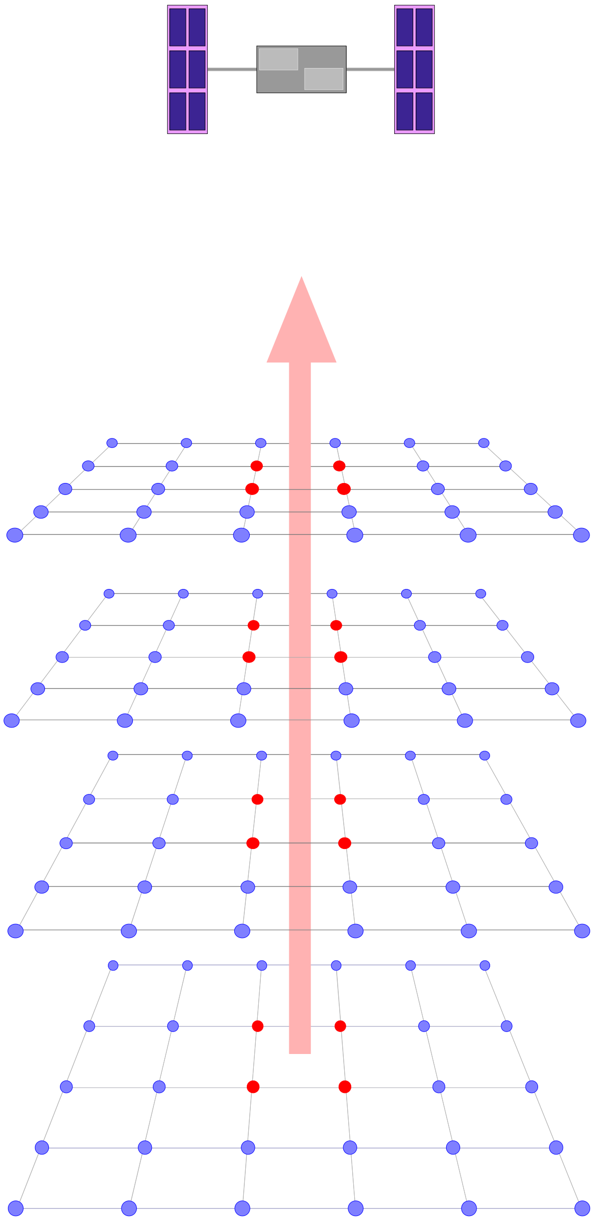
$$\frac{1}{2}\sum_t (\boldsymbol{H}_t[\boldsymbol{M}_t \ldots \boldsymbol{M}_{\delta t}\boldsymbol{x}(0)] - \boldsymbol{y}_t)^T \ \mathbf{R}_t^{-1} \ (\boldsymbol{H}_t[\boldsymbol{M}_t \ldots \boldsymbol{M}_{\delta t}\boldsymbol{x}(0)] - \boldsymbol{y}_t)$$

Gradient vector:

$$\nabla_{x(0)} J = \mathbf{B}^{-1} (\boldsymbol{x}(0) - \boldsymbol{x}_B(0)) +$$

$$\sum_t \mathbf{M}_{\delta t}^T \ldots \mathbf{M}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1} (\boldsymbol{H}_t[\boldsymbol{M}_t \ldots \boldsymbol{M}_{\delta t}\boldsymbol{x}(0)] - \boldsymbol{y}_t)$$





Adjoint Variables and Adjoint Operators

$$\nabla_{x(0)}J = \mathbf{B}^{-1}(\boldsymbol{x}(0) - \boldsymbol{x}_B(0))$$

$$+$$

$$\sum_t \mathbf{M}_{\delta t}^T \ldots \mathbf{M}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1}(\boldsymbol{H}_t[\boldsymbol{M}_t \ldots \boldsymbol{M}_{\delta t}\boldsymbol{x}(0)] - \boldsymbol{y}_t)$$

Forward operator:

$$\boldsymbol{x}(t+\delta t) = \mathbf{M}_{t+\delta t}\,\boldsymbol{x}(t)$$

Adjoint operator:

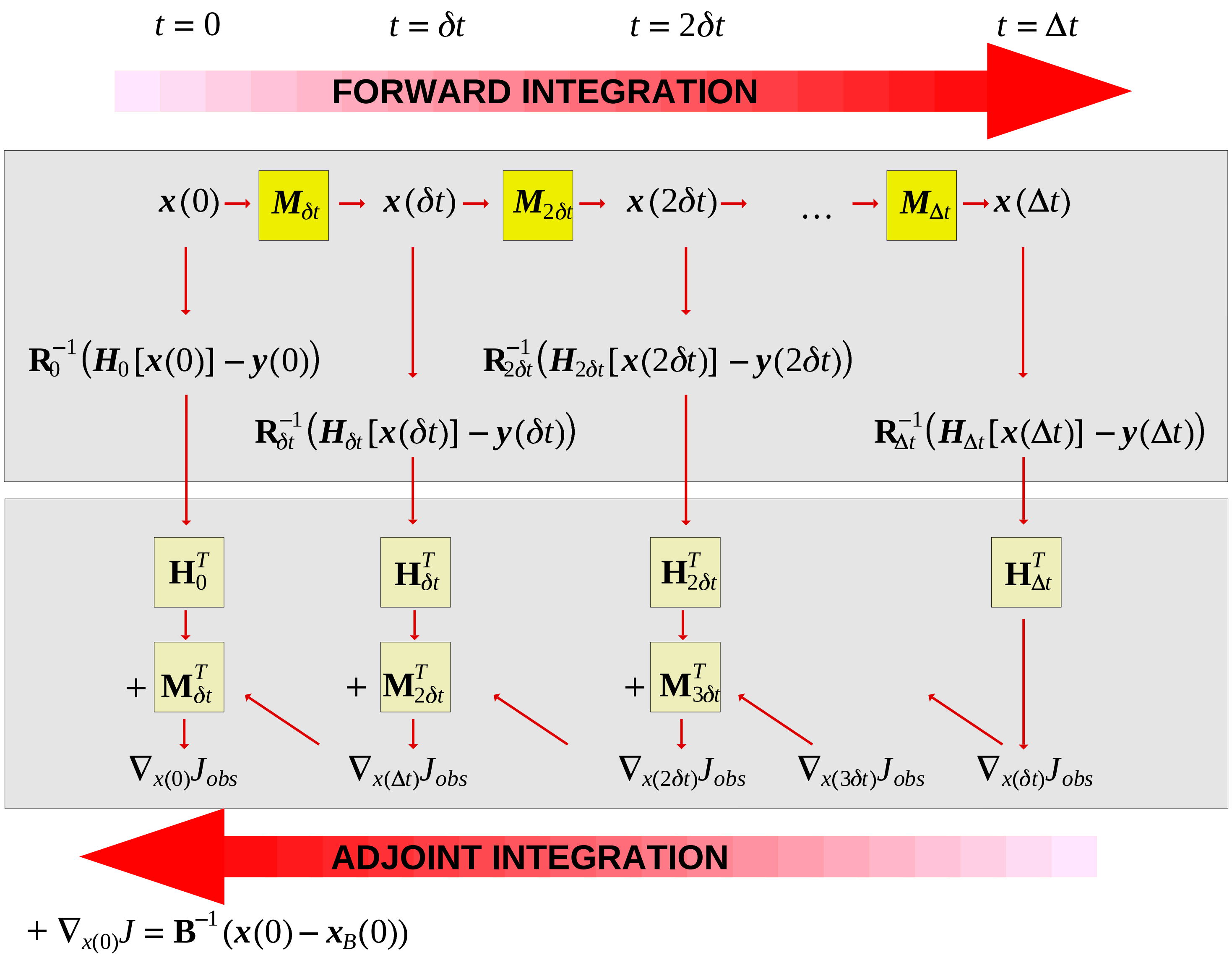
$$\nabla_{x(t)}J = \mathbf{M}_{t+\delta t}^{\dagger}\nabla_{x(t+\delta t)}J$$

Gradient vectors
(adjoint variables)

Adjoint operator
In simplest case,

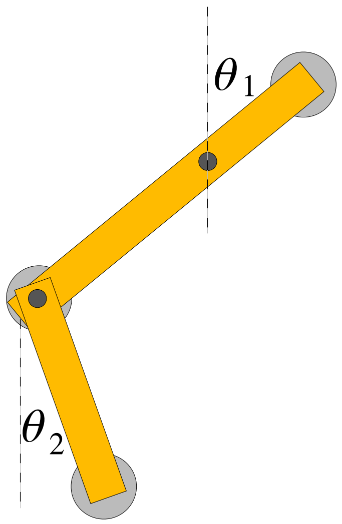
$$\mathbf{M}_{t+\delta}^{\dagger} = \mathbf{M}_{t+\delta}^{T}$$

The adjoint of an operator propagates the adjoint variables in the reverse sense
(this is just the chain rule generalised to many variables)


t = 0
t = δt
t = 2δt
t = Δt
FORWARD INTEGRATION
x(0)
M_δt
x(δt)
M_2δt
x(2δt)
...
M_Δt
x(Δt)
R_0^-1(H_0[x(0)] − y(0))
R_2δt^-1(H_2δt[x(2δt)] − y(2δt))
R_δt^-1(H_δt[x(δt)] − y(δt))
R_Δt^-1(H_Δt[x(Δt)] − y(Δt))
H_0^T
H_δt^T
H_2δt^T
H_Δt^T
+ M_δt^T
+ M_2δt^T
+ M_3δt^T
∇_x(0) J_obs
∇_x(Δt) J_obs
∇_x(2δt) J_obs
∇_x(3δt) J_obs
∇_x(δt) J_obs
ADJOINT INTEGRATION
+ ∇_x(0) J = B^-1 (x(0) − x_B(0))

Example of '4d'-Var. with a simple chaotic system

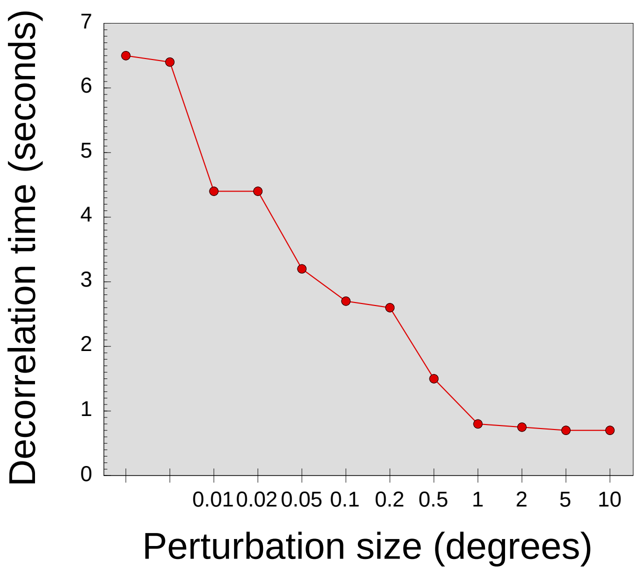
The double pendulum



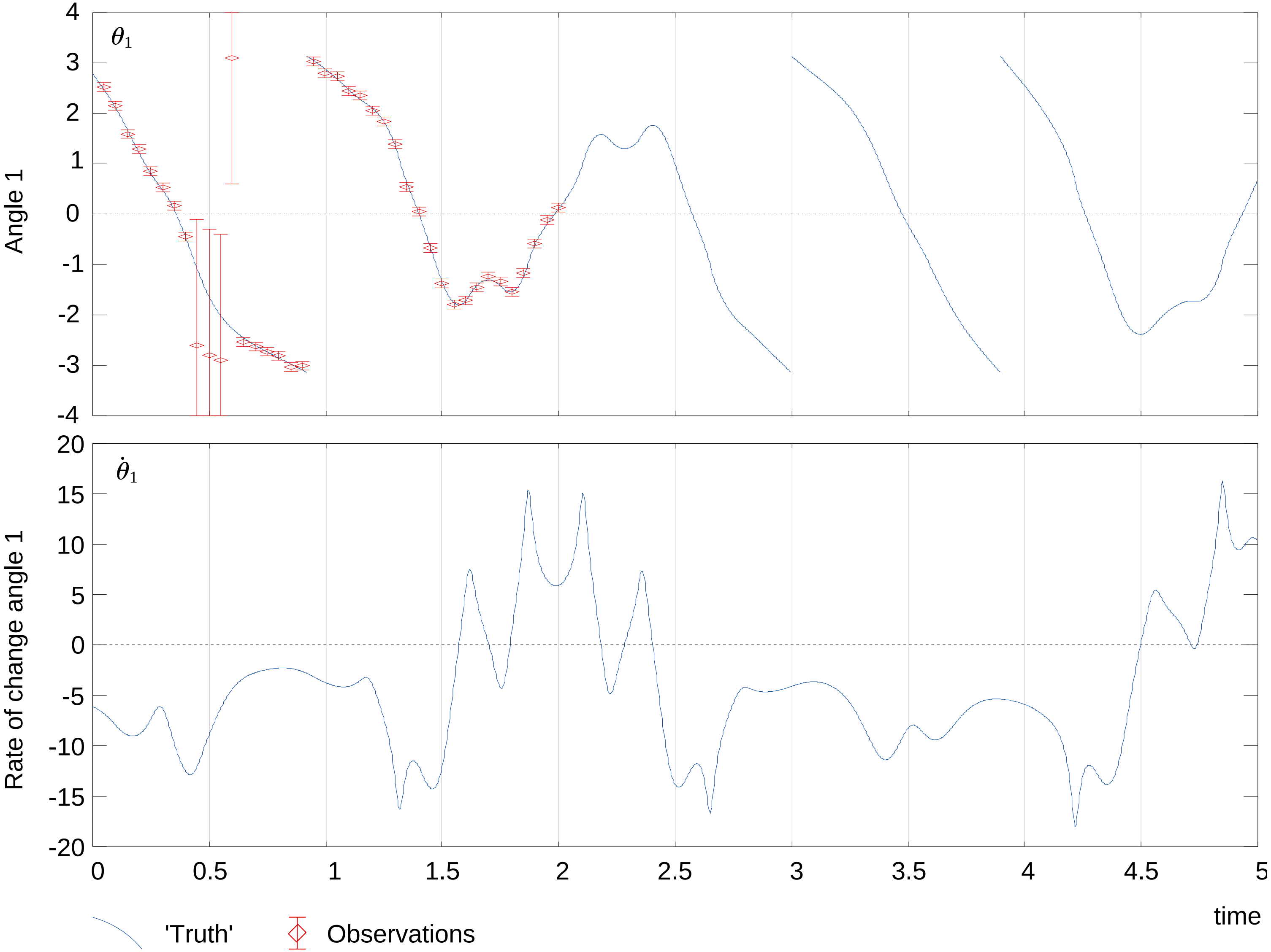
$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\theta}_i}\right) = \frac{\partial L}{\partial \theta_i}$$

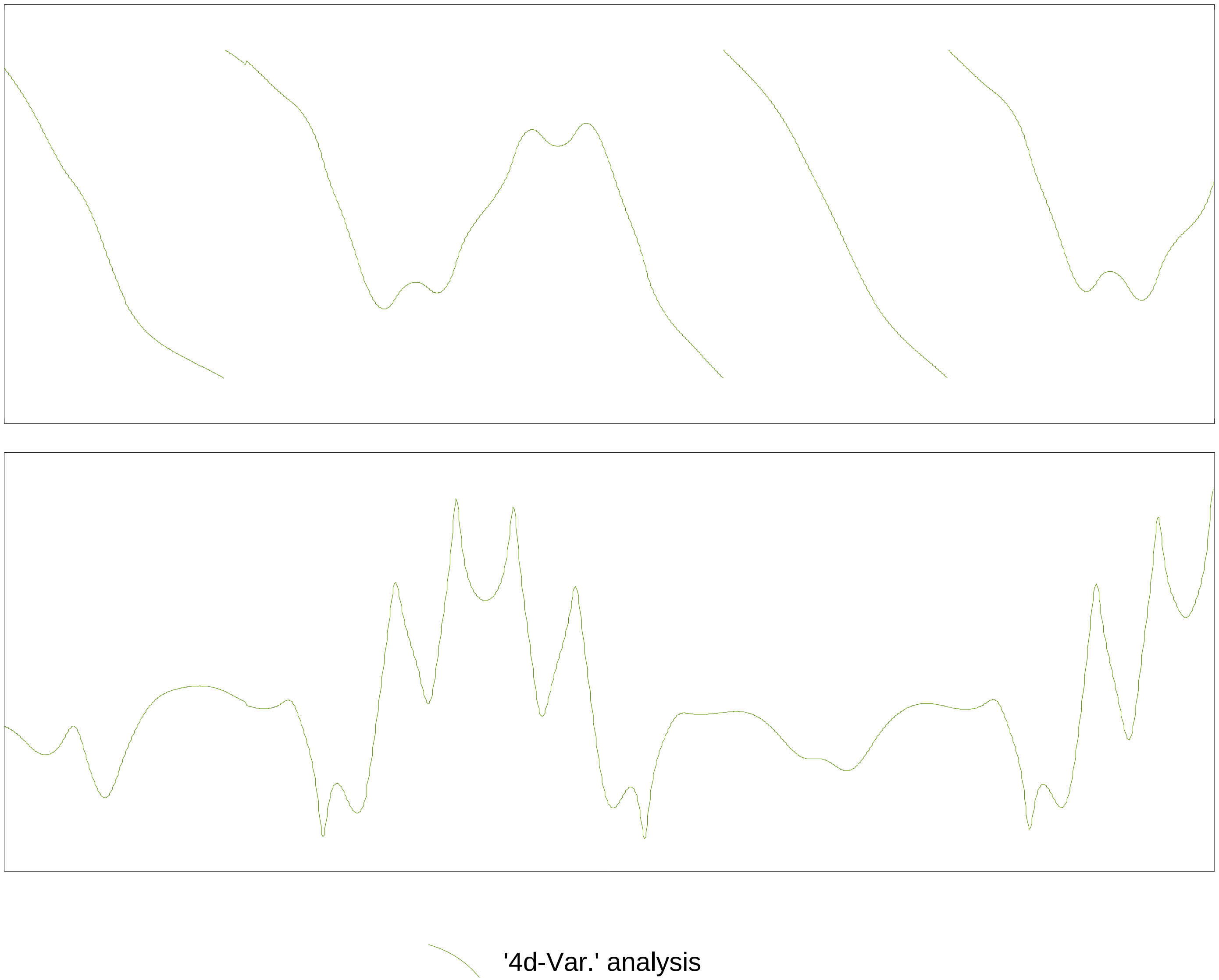
$$L = T - V$$

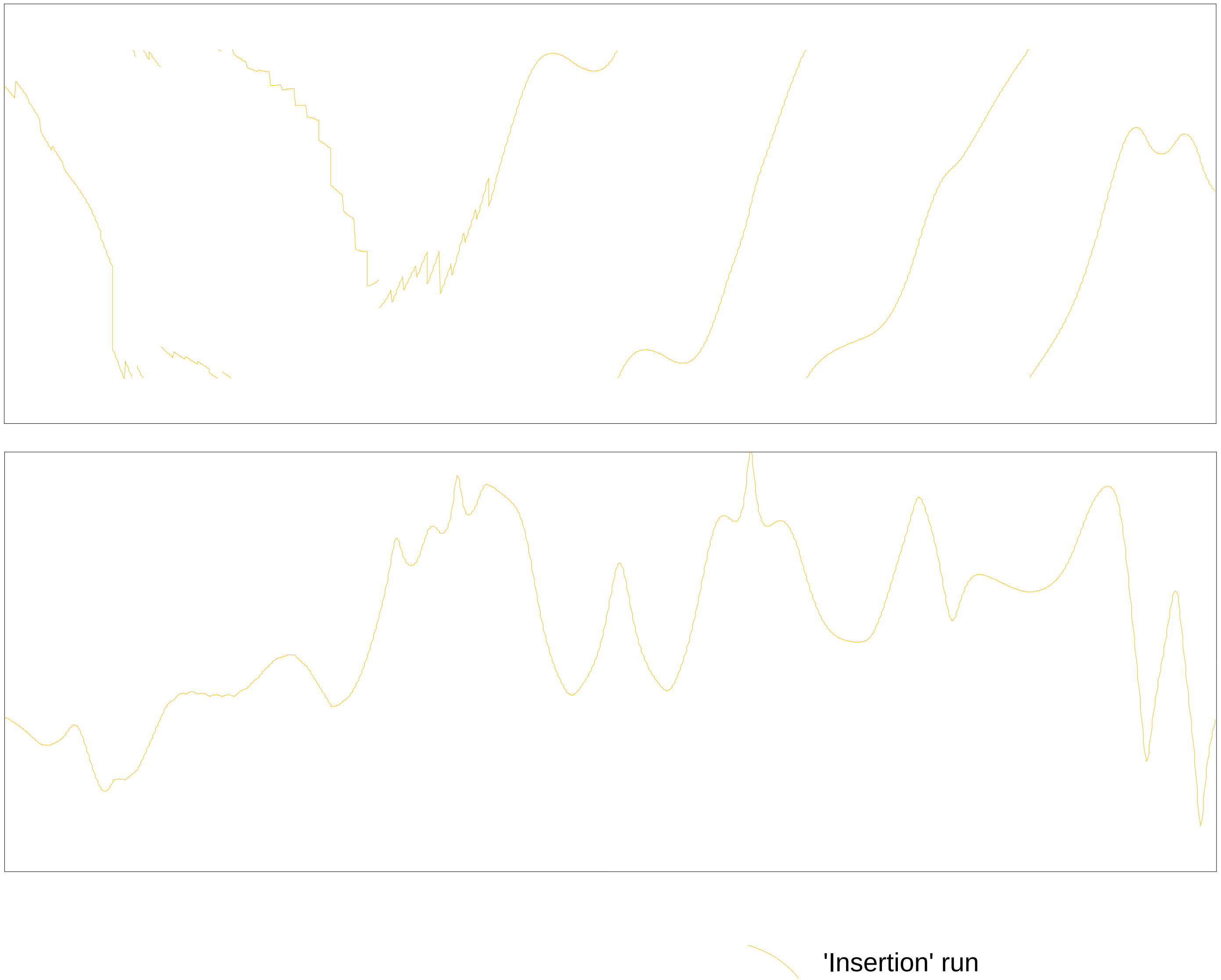
$$\boldsymbol{x} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix}$$



DATA ASSIMILATION WITH THE DOUBLE PENDULUM




'4d-Var.' analysis


'Insertion' run

Other GFD applications

Sources/sinks determination

Forward model (tracer transport Eq.):

$$\frac{\partial q(\mathbf{r},t)}{\partial t} = -\nabla\cdot(q(\mathbf{r},t)\vec{v}(\mathbf{r},t)) + \dot{q}(\mathbf{r})$$

What are the sources/sinks, $\dot{q}(\mathbf{r})$ given observations of $q(\mathbf{r},t)$?

State vector:

$$\dot{\boldsymbol{q}} = \begin{pmatrix} \dot{q}(\boldsymbol{r}_1) \\ \dot{q}(\boldsymbol{r}_2) \\ \dots \\ \dot{q}(\boldsymbol{r}_N) \end{pmatrix}$$

Cost function:

$$J[\dot{\boldsymbol{q}}] = \frac{1}{2}(\dot{\boldsymbol{q}} - \dot{\boldsymbol{q}}_B)^T \mathbf{B}_{\dot{q}}^{-1}(\dot{\boldsymbol{q}} - \dot{\boldsymbol{q}}_B) +$$
$$\frac{1}{2}\sum_t\sum_t (\boldsymbol{H}_t[\boldsymbol{M}_t\dots\boldsymbol{M}_{\delta t}\boldsymbol{q}(0)] - \boldsymbol{y}_t)^T \mathbf{R}_t^{-1}(\boldsymbol{H}_t[\boldsymbol{M}_t\dots\boldsymbol{M}_{\delta t}\boldsymbol{q}(0)] - \boldsymbol{y}_t)$$

Gradient w.r.t. $\dot{\boldsymbol{q}}$:

$$\nabla_{\dot{q}}J = \mathbf{B}_{\dot{q}}^{-1}(\dot{\boldsymbol{q}} - \dot{\boldsymbol{q}}_B) +$$
$$\sum_t\sum_{t'=0}^{t}\left(\frac{\partial\boldsymbol{q}(t')}{\partial\dot{\boldsymbol{q}}}\right)^T \mathbf{M}_{t'+1}^T\dots\mathbf{M}_t^T\mathbf{H}_t^T\mathbf{R}_t^{-1}(\boldsymbol{H}_t[\boldsymbol{M}_t\dots\boldsymbol{M}_{\delta t}\boldsymbol{q}(0)] - \boldsymbol{y}_t)$$

Other bonuses of doing inverse modelling / DA

- Model performance
- Observation quality

Some difficulties with inverse modelling / DA

- Non-linearity (errors, parametrisations)
- Model budget disruption by obs.
- The 'initialization problem'
- Treatment of model error
- Null space
- Error characterization, esp. multivariate **B**
- Artefacts from unrealistic **B**

Summary

Inverse methods:

- are an integral part of science
- infer information about model parameters using:
  - noisy, irregular, and indirect measurements
  - how the system behaves
- require expertise in:
  - forward modelling
  - inverse techniques
  - dealing with large volumes of information
- make use of a number of methods and assumptions, for DA:
  - Gaussian error characteristics
  - method of least squares
  - B.L.U.E. / 3d/4d Var. / Kalman filter
- can help assess:
  - model performance
  - observation quality
- becomes difficult esp:
  - non-linear models
  - large No. of degrees of freedom
  - multivariate
- suffer potential problems:
  - representing B
  - artefacts
  - null space