Variational Data Assimilation

George C. Craig

Ludwig-Maximilians-Universität, München

Outline

Lecture 1

- Cost function
- Observation operator
- Math aside:
  - Multivariate Gaussian
  - Multivariate Taylor series
- 3DVAR solution!
- Observation errors

Lecture 2

- Background errors
- Time of observations FGAT
- 4DVAR cost function
- Math aside:
  - Tangent linear model
  - Adjoint
- 4DVAR solution

The cost function

- Given: background forecast w^b and observation f
- Goal is to find an analysis w^a, that best matches both, taking into account their (squared) errors B and R
- Least squares estimate – minimize cost function

$$J(w) = \frac{(w^b - w)^2}{B} + \frac{(f - w)^2}{R}$$

The analysis

- *J* has minimum where $\nabla J = 0$
- So analysis w^a is solution of a linear equation:

$$\frac{dJ}{dw} = B^{-1}(w^b - w) + R^{-1}(f - w) = 0$$

- Solution is error-weighted average

$$w^a = \frac{B^{-1}}{B^{-1} + R^{-1}} w^b + \frac{R^{-1}}{B^{-1} + R^{-1}} f$$

- This is Best Linear Unbiased Estimator (BLUE) if the errors in w^b and f are Gaussian distributed, with mean 0 and variance *B* and *R*, resp.

Observation operators

- Problem: *f* does not match *w*
  - Wrong location (in between grid points)
  - Wrong quantity (e.g. radiance)
- No problem: use forward model to estimate what observations should look, given the model state
- *H(w)* are simulated observations that can be compared with *f*
- Cost function is now:

$$J(w) = B^{-1}(w^b - w)^2 + R^{-1}(f - H(w))^2$$

- To minimize, $\left(\nabla J = 0\right)$ need derivatives of *H*

The (new) analysis

- Approximate by Taylor series

$$H(w) = H(w^b) + H'(w^b)(w - w^b) + \ldots$$

- If we keep only the first derivative, grad(J) = 0 is again a linear equation for w^a

$$\frac{dJ}{dw} = B^{-1}(w^b - w) + R^{-1}\left(f - H(w^b) - H'(w^b)(w - w^b)\right) = 0$$

- Solution is still a weighted average, but can also be written:

$$w^a = w^b + \frac{B}{R} H'(w^b)\left(f - H(w^b) - H'(w^b)(w - w^b)\right)$$

- Background forecast is adjusted in proportion to the disagreement with observations

Math aside: Gaussian

- Assume *f* drawn from a Gaussian distribution

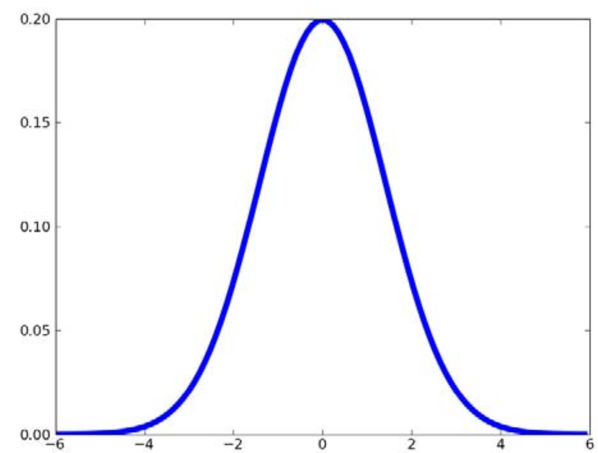
$$p(f) = (2\pi R)^{-1/2} \exp \frac{\left(f - f^{true}\right)^2}{R}$$

$$= (2\pi R)^{-1/2} \exp\left[\left(f - f^{true}\right) R^{-1} \left(f - f^{true}\right)\right]$$

- *B* is the background error variance

$$R = \mathrm{E}\left[\left(f - f^{true}\right)^2\right]$$

- E[…] is expectation, i.e. average over many trials



Math aside: Multivariate Gaussian

- For a vector of observations **f**

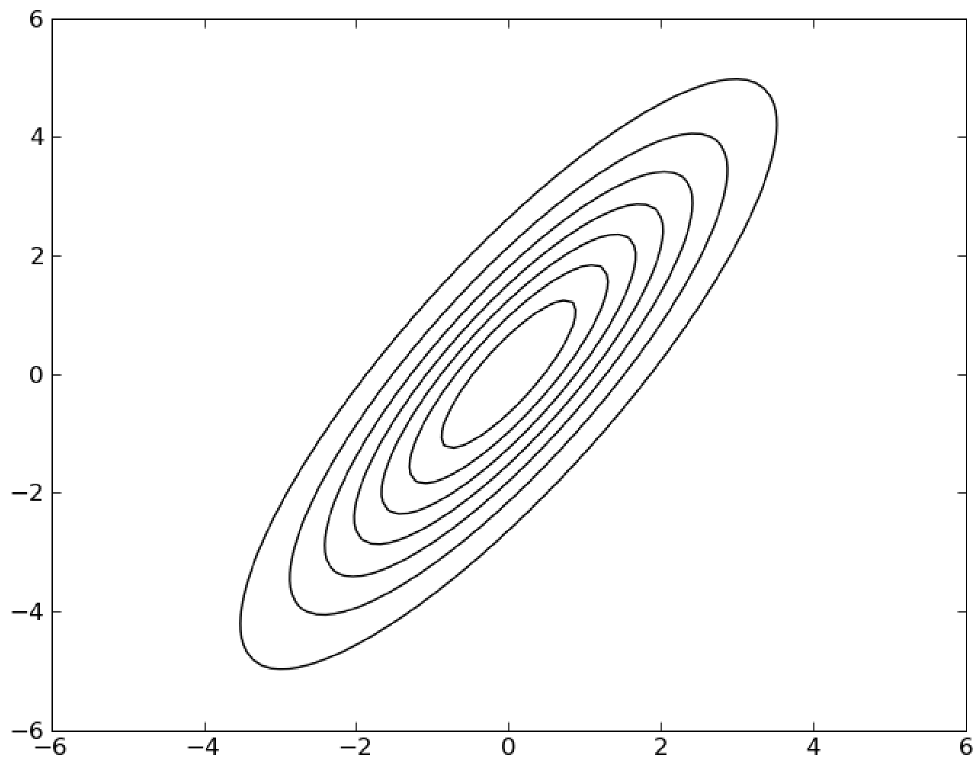
$$p(\mathbf{f}) = \left(2\pi \det(\mathbf{R})\right)^{-1/2} \exp\left[\left(\mathbf{f} - \mathbf{f}^{true}\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{f} - \mathbf{f}^{true}\right)\right]$$

- R is now a covariance matrix

$$\mathbf{R} = \mathrm{E}\left[\left(\mathbf{f} - \mathbf{f}^{true}\right)\left(\mathbf{f} - \mathbf{f}^{true}\right)^{\mathrm{T}}\right] = \mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right] = \mathrm{E}\begin{pmatrix} {\varepsilon_1}^2 & \varepsilon_1\varepsilon_2 & \cdot\cdot & \\ \varepsilon_2\varepsilon_1 & {\varepsilon_2}^2 & & \\ \vdots & & \ddots & \\ & & & {\varepsilon_m}^2 \end{pmatrix}$$

Math aside: A 2D Gaussian function

- 2d Gaussian has errors in e variable
- Here errors in x and y have strong positive correlation



The 3DVAR cost function

- Given: background forecast $\mathbf{w}^b$ and observations $\mathbf{f}$
- Goal is to find an analysis $\boldsymbol{w}^a$, that best matches both, taking into account their error covariances **B** and **R**
- Least squares estimate – minimize cost function

$$J(\mathbf{w}) = (\mathbf{w} - \mathbf{w}^b)^T \mathbf{B}^{-1} (\mathbf{w} - \mathbf{w}^b) + (\mathbf{f} - H(\mathbf{w}))^T \mathbf{R}^{-1} (\mathbf{f} - H(\mathbf{w}))$$

- But, to find grad(J) = 0, need to take derivatives of $H(\mathbf{w}^b)$

Math aside 2: Taylor series and Jacobian matrix

- 1 dimensional Taylor series:

$$H(w) = H(w^b) + \left.\frac{dH}{dw}\right|_{w=w^b} (w - w^b) + \ldots$$

- 2 dimensional version:

$$H_1(\mathbf{w}) = H_1(\mathbf{w}^b) + \left.\frac{\partial H_1}{\partial w_1}\right|_{\mathbf{w}=\mathbf{w}^b} (w_1 - w_1{}^b) + \left.\frac{\partial H_1}{\partial w_2}\right|_{\mathbf{w}=\mathbf{w}^b} (w_2 - w_2{}^b) + \ldots$$

$$H_2(\mathbf{w}) = H_2(\mathbf{w}^b) + \left.\frac{\partial H_2}{\partial w_1}\right|_{\mathbf{w}=\mathbf{w}^b} (w_1 - w_1{}^b) + \left.\frac{\partial H_2}{\partial w_2}\right|_{\mathbf{w}=\mathbf{w}^b} (w_2 - w_2{}^b) + \ldots$$

Math aside 2: Taylor series and Jacobian matrix

- Multi-dimensional Taylor series:

$$H(\mathbf{w}) = H(\mathbf{w}^b) + \mathbf{H}(\mathbf{w}^b)(\mathbf{w} - \mathbf{w}^b) + \ldots$$

- **H** is the matrix of first derivatives (Jacobian)

$$\mathbf{H} = \begin{pmatrix} \frac{\partial H_1}{\partial w_1} & \frac{\partial H_1}{\partial w_2} & .. & \frac{\partial H_1}{\partial w_m} \\ \frac{\partial H_2}{\partial w_1} & \frac{\partial H_2}{\partial w_2} & & \\ \vdots & & \ddots & \vdots \\ \frac{\partial H_m}{\partial w_1} & & .. & \frac{\partial H_m}{\partial w_n} \end{pmatrix}$$

Incremental 3DVAR

- If background forecast is good, i.e. $\mathbf{w}^a$ close to $\mathbf{w}^b$, can approximate H by linear function $H(\mathbf{w}) = H(\mathbf{w}^b) + \mathbf{H}(\mathbf{w} - \mathbf{w}^b)$

$$J(\mathbf{w}) = (\mathbf{w}-\mathbf{w}^b)^T \mathbf{B}^{-1}(\mathbf{w}-\mathbf{w}^b) + (\mathbf{f} - H(\mathbf{w}^b) - \mathbf{H}(\mathbf{w}-\mathbf{w}^b))^T \mathbf{R}^{-1}(\mathbf{f} - H(\mathbf{w}^b) - \mathbf{H}(\mathbf{w}-\mathbf{w}^b))$$

- Rewrite cost function in terms of analysis increments $\delta\mathbf{w} = \mathbf{w} - \mathbf{w}^b$

  and observation increments (innovations) $\delta\mathbf{f} = \mathbf{f} - H(\mathbf{w}^b)$

$$J(\delta\mathbf{w}) = (\delta\mathbf{w})^T \mathbf{B}^{-1}(\delta\mathbf{w}) + (\delta\mathbf{f} - \mathbf{H}\delta\mathbf{w})^T \mathbf{R}^{-1}(\delta\mathbf{f} - \mathbf{H}\delta\mathbf{w})$$

The 3DVAR analysis

- *J* has minimum where $\nabla J = 0$

$$\nabla J = \mathbf{B}^{-1}\delta\mathbf{w} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{w} - \mathbf{H}^T\mathbf{R}^{-1}\delta\mathbf{f} = 0$$

- So analysis $\mathbf{w}^a$ is solution of a system of linear equations:

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

- DONE!
- …but
- Dimension of **w** is $n \sim 10^7$, so **B** has of order 10^{14} elements – 100 TB
- Call your numerical analyst, HPC vendor, and think hard about simplifying
- Iterative methods, Pre-conditioning, Outer loop??

Time of observations

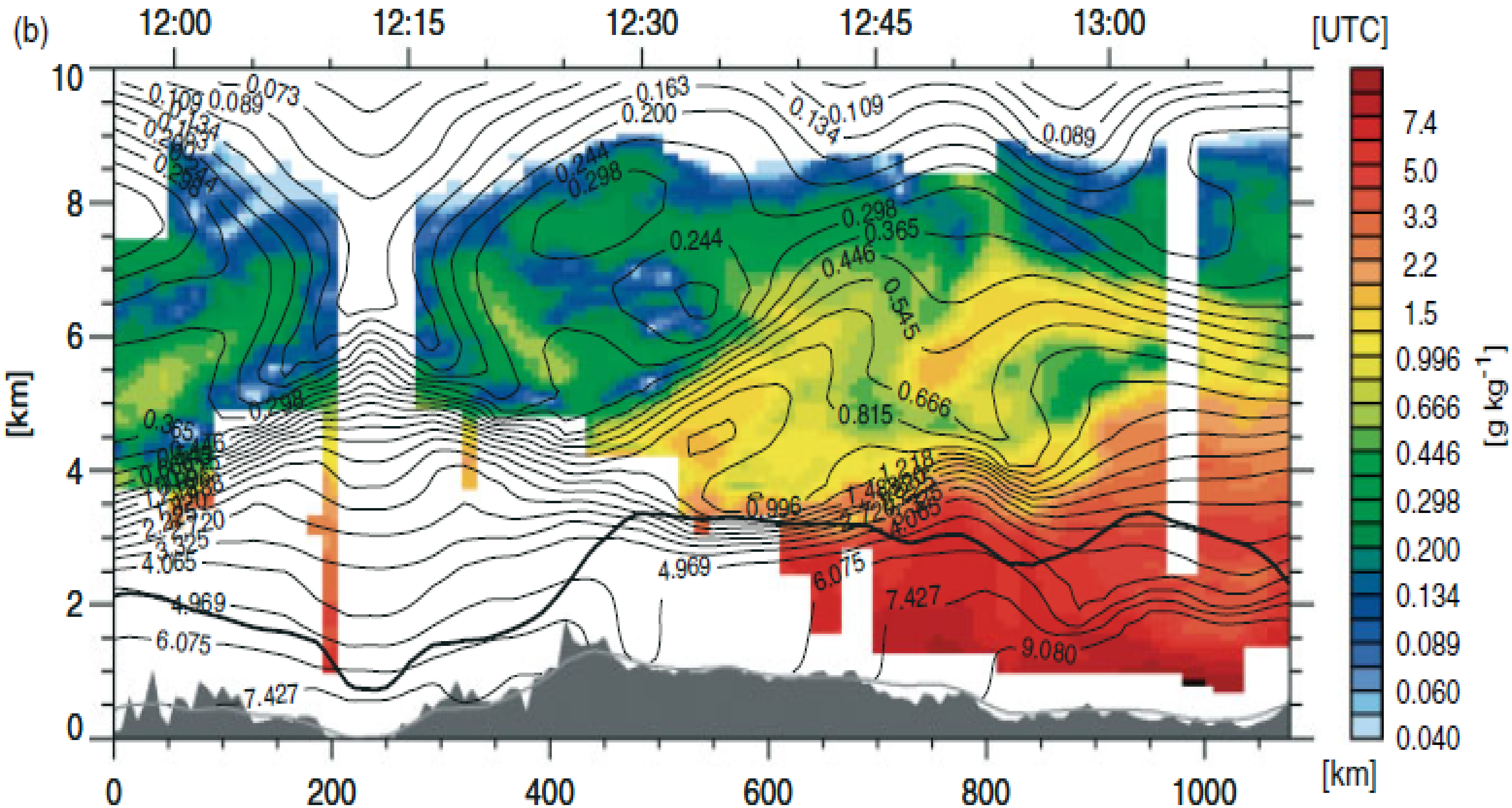
- ➔ In 3DVAR, all observations are assumed to be at the same time
- ➔ Actually collected over a short time window, e.g. plus/minus 1 hour

Observation error covariance matrix R

- 3 sources of error
  - Instrument error
  - Representativity error (e.g. observation is a point value, but model predicts a grid-box average)
  - Forward model error
- Mostly uncorrelated
- Where does it come from? – observation expert (works in DA group)

Example of representativity error

➔ Lidar humidity



Outline

Lecture 1

- Cost function
- Observation operator
- Math aside:
  - Multivariate Gaussian
  - Multivariate Taylor series
- 3DVAR solution!
- Observation errors

Lecture 2

- Background errors
- Time of observations FGAT
- 4DVAR cost function
- Math aside:
  - Tangent linear model
  - Adjoint
- 4DVAR solution