MSc Course: Theory and Techniques of Data Assimilation

Two Lectures on "3d-Var."

Ross Bannister, room 1U11, Dept. of Meteorology, Univ. of Reading,
r.n.bannister@reading.ac.uk

Version 2007

Section A: List Of Topics And References

A.1: List Of Topics

A. References.
B. Introduction - why do data assimilation?
C. 3-dimensional variational assimilation and operational data assimilation.
D. The gradient and Hessian of the cost function.
E. Example observation operators.
F. Minimization algorithms.
G. Preconditioning.

A.2: Further Reading

- Kalnay E., Atmospheric Modelling, *Data Assimilation and Predictability*, Ch. 5.
- Daley, *Atmospheric Data Analysis*, Ch.13.
- ECMWF, Data assimilation course handouts, `http://www.ecmwf.int/newsevents/training/lecture_notes/LN_DA.html`.
- Schlatter T.W., *Variational assimilation of meteorological observations in the lower atmosphere: a tutorial on how it works*, Journal of atmospheric and solar-terrestrial physics 62, pp. 1057-1070 (2000).
- Lorenc et al., *The Met Office global 3-dimensional variational assimilation scheme*, QJRMS 126, pp. 2991-3012 (2000).
- This handout and other notes, `http://www.met.rdg.ac.uk/~ross/DARC/MSc/MSc.html`.

Section B: The Need To Do Data Assimilation

B.1: Why do we need to do data assimilation (DA)?

- Bjerknes, 1911: The "*ultimate problem in meteorology*".
- 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*".

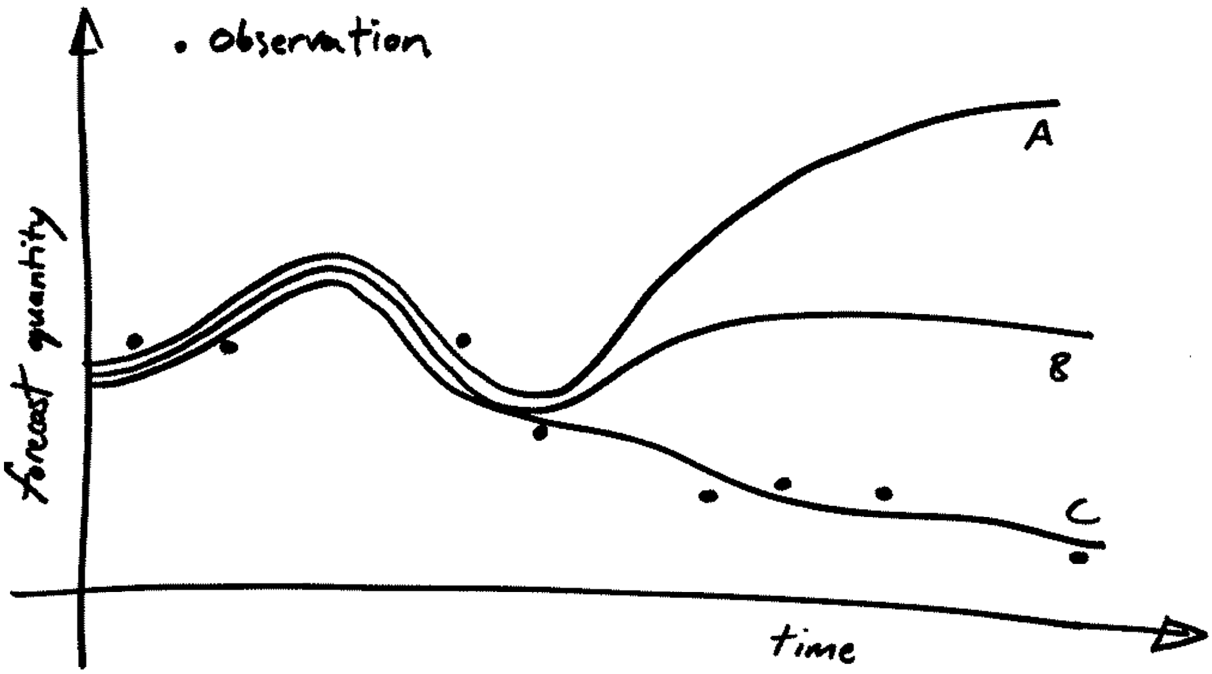


Fig. 1a: Two initially similar free-running forecasts (trajectories A and B) showing sensitive dependence on initial conditions ('chaos'). After a point in time the trajectories diverge. After this point, it might be found that neither is close to the true trajectory. Feeding-in observations (dots) using DA (trajectory C) can help keep the model close to the 'truth'.

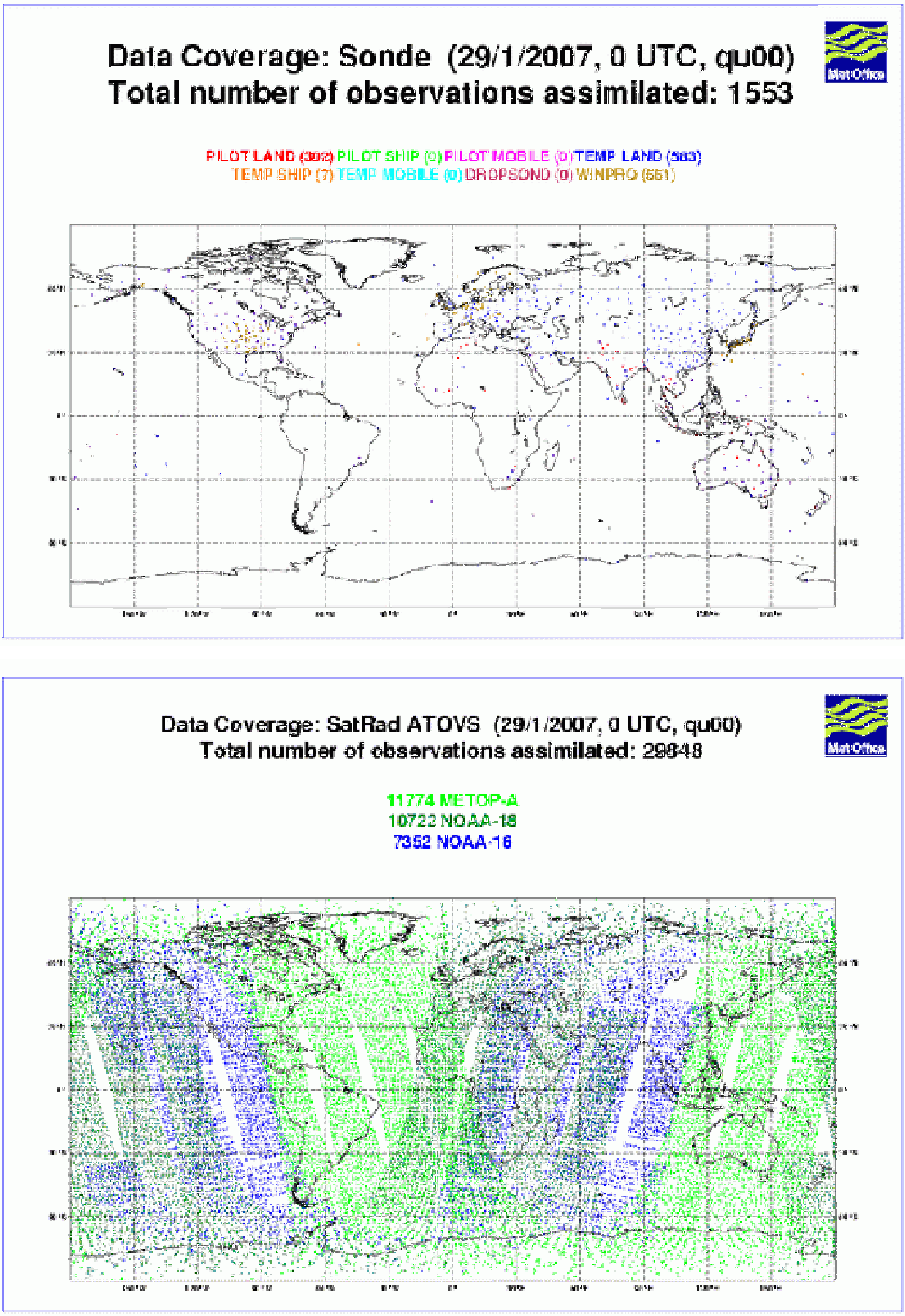


Fig 1b: Example coverage of radiosonde measurements and ATOVS satellite observation locations.

Section C: 3-d Var. And Operational Data Assimilation

C.1: How is data assimilation used in weather forecasting?

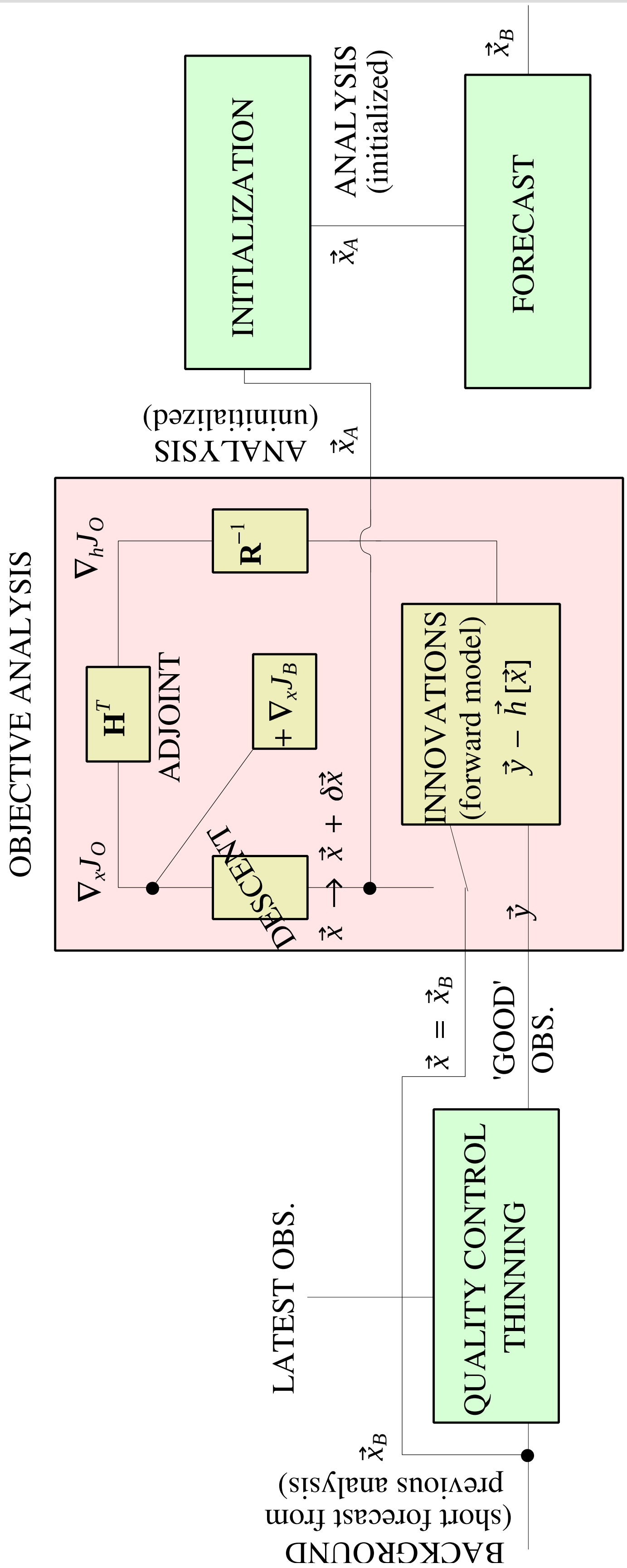


Fig. 2: The intermittent 'data assimilation cycle' showing use of a variational scheme as the data assimilation method.

C.3: What is the 3d-Var. cost function?

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \frac{1}{2}(\vec{y} - \vec{h}[\vec{x}])^T \mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}]), \quad (1b)$$

J is minimized for $J[\vec{x} = \vec{x}_A]$.

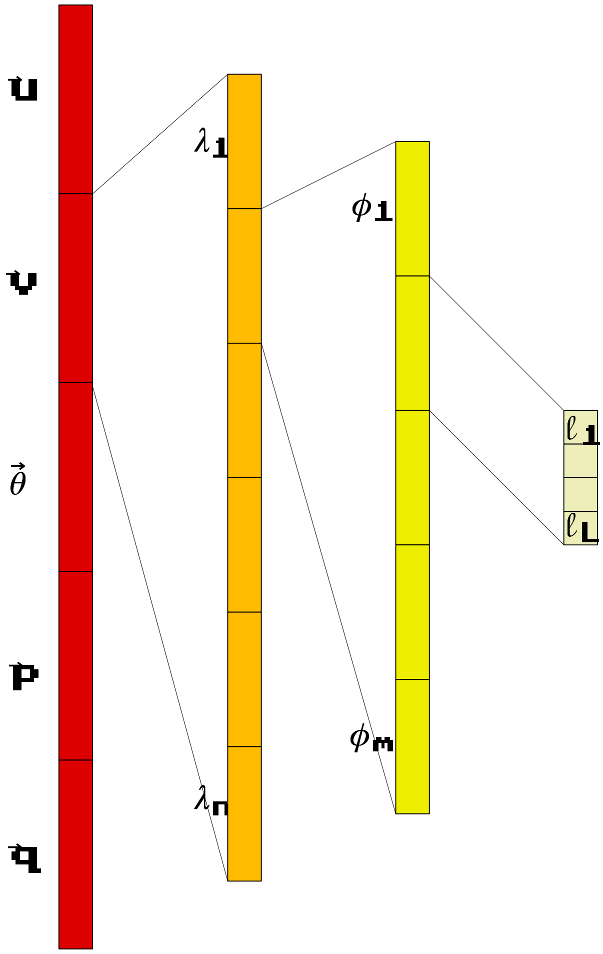


Fig. 3: The meaning of the state vector. The vector has n elements.

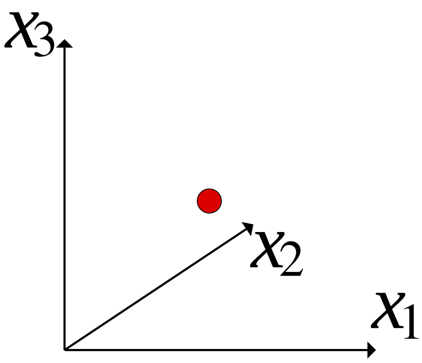


Fig. 4: State space schematic for $n = 3$.

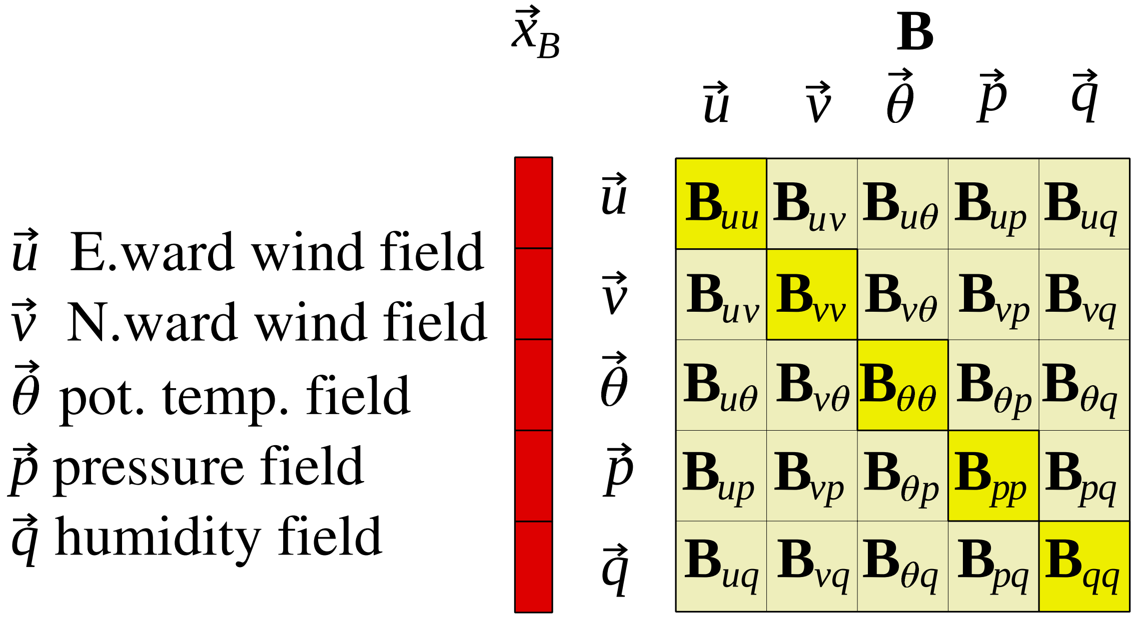


Fig. 5: The background error covariance matrix for a forecast given in the state space of Fig. 3. Each square is itself a matrix. Sub-matrices along the diagonal (deep yellow) are called 'self-covariances' and off-diagonal sub-matrices are called 'multivariate covariances'.

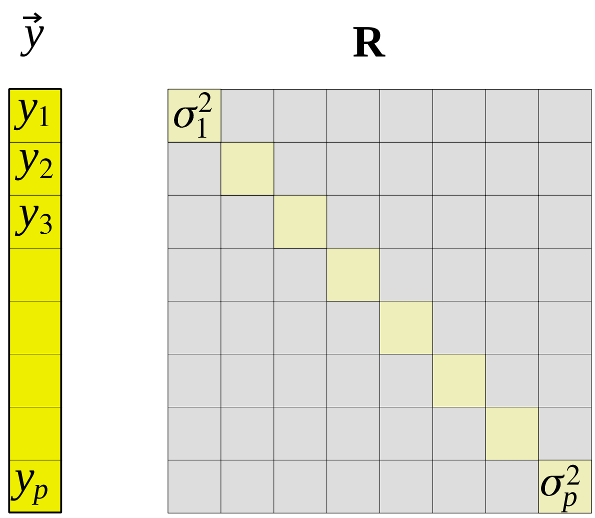


Fig. 6: The observation error covariance matrix (right) shown against the observation vector (left). Often observation errors are taken to be uncorrelated with each other and so **R** is diagonal. The diagonal matrix elements are the respective observation variances (equal to the square of the standard deviations) and the off-diagonal elements are zero. There are p observations.

C.4: What is '3d' about 3d-Var.?

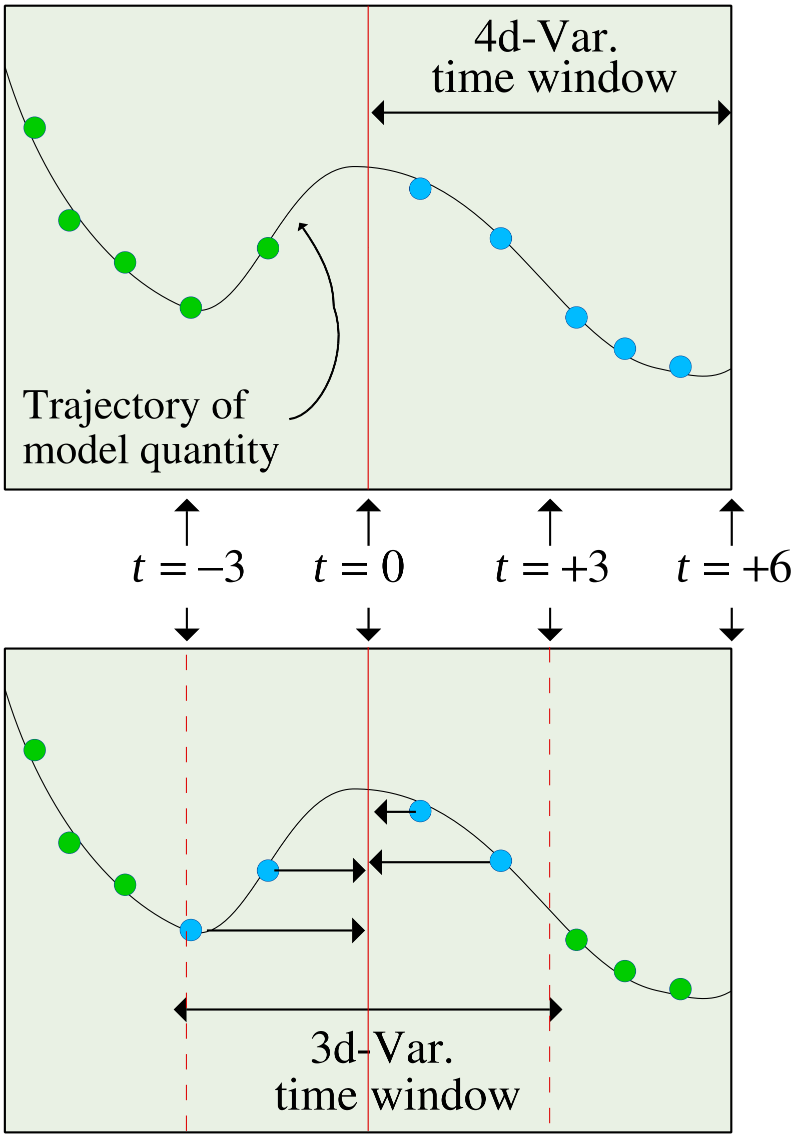


Fig. 7: Under the formulation of 4d-Var. (top), observations are used at their correct time. In 3d-Var. (bottom), the observations within a six-hour cycle are taken as though they have been made at the same time. In each case, the analysis time is at $t = 0$.

C.5: How 'large' is an operational 3d-Var. system?

Observations used in global data assimilation in October 2000.

Obs Group	Sub-group	Items used	Daily	% used
Ground-based Vertical profiles	TEMP	T, V, RH processed to model layer average	1200	97
	PILOT	As TEMP but V only	900	99
	PROFILER	As TEMP but V only (used from Feb 2001)	300	0 (65)
Satellite-based Vertical profiles	TOVS	Radiances directly assimilated with channel selection dependent on surface, instrument and cloudiness	54000	11
	ATOVS		700000	4
Aircraft (manual & automated)	AIREPS	T, V as reported with duplicate checking and blacklist	14000	21
	ACARS AMDAR ASDAR		67000	60
Satellite atmospheric motion vectors	GOES 8,10	High res. 'BUFR' IR winds	55000	24
	Meteosat 5,7	IR, VIS and WV winds	9200	98
	GMS 5	IR, VIS and WV winds	5200	93
Satellite-based surface	ERS-2	wind vector retrievals (ambiguous winds from Feb 2001)	170000	0
	SSMI-13	In-house 1DVAR wind speed retrieval (no moisture yet)	1450000	1
Ground-based surface	Land Synop	Pressure only (processed to model surface)	27000	80
	Ship Synop	Pressure and Wind	6000	90 95
	Buoy	Pressure	9000	75

Typical coverage maps are available at
http://www.metoffice.com/research/nwp/observations/data_coverage/index.html
or http://www.ecmwf.int/services/dcover/index_map.html

Only a small selection of high-density satellite observations are currently used; about 1.6E5 data are presented to the variational analysis. For comparison, the new Met Office model has about 4.3E7 degrees of freedom, which we reduce (using hydrostatic, cloud-moisture and smoothness assumptions) to 1.4E6 independent control variables.

Fig. 8: Typical observations assimilated in Met Office Var. (A. Lorenc, Oxford RAL Spring School Lecture, 2001.)

1993-1999 VAR coding took 42 person-years from 35 different people.

March 2001

	Subroutines, modules etc.	*Lines*
3D-Var	976	338973
PF & adjoint models (converting 3D-Var to 4D-Var)	156	87412
Obs processing & general utilities	1085	277690
Unified Model (vn5.1)	2037	522624

The current global 3D-Var system uses ~8 times more computer resources to assimilate 1 days' data, than to do a 1 day forecast.
~60% scales with resolution, ~40% scales with number of observations.

For the ECMWF 4D-Var system the ratio is ~20-40.

Fig. 9: The amount of computer code written for the Met Office Var. system is comparable to that of the Met Office forecast model. (A. Lorenc, Oxford RAL Spring School Lecture, 2001.)

C.6: How many iterations are required to minimize J?

$$J[\vec{x} = \vec{x}_A] \sim \frac{p}{2}. \quad (2)$$

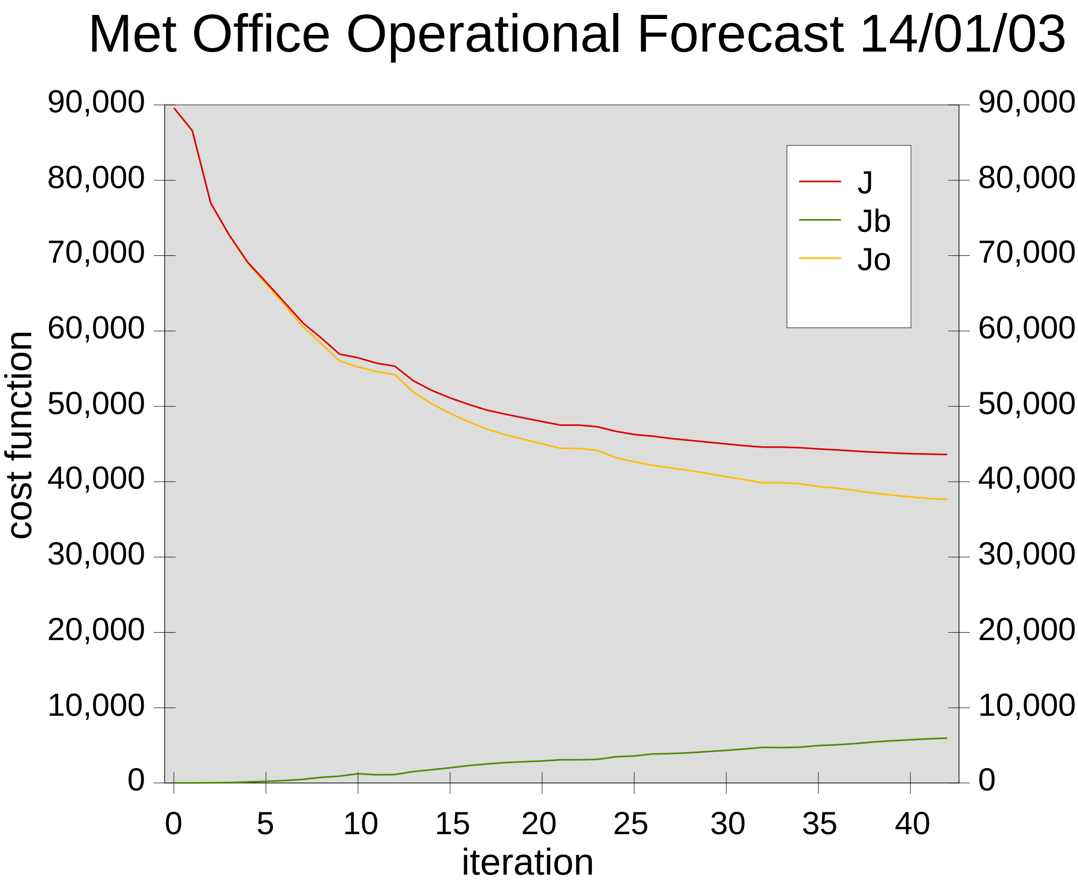


Fig 10: Value of the cost function and its components as a function of iteration for Met Office 3d-Var.

C.7: How is Var. related to the optimal interpolation formula?

$$\text{Let} \quad \vec{x} = \vec{x}_B + \delta\vec{x}, \tag{3}$$

$$\text{then} \quad \vec{h}[\vec{x}_B + \delta\vec{x}] \approx \vec{h}[\vec{x}_B] + \mathbf{H}\delta\vec{x}. \tag{4}$$

$$\mathbf{H} = \left.\frac{\partial\vec{h}}{\partial\vec{x}}\right|_{\vec{x}_B},$$

$$\mathbf{H}_{ij} = \frac{\partial h_i}{\partial x_j} \quad (1 \leqslant i \leqslant p, \quad 1 \leqslant j \leqslant n). \qquad (5)\,(6)$$

$$\begin{aligned} J &= \frac{1}{2}\delta\vec{x}^{\,T}\mathbf{B}^{-1}\delta\vec{x} + \frac{1}{2}(\vec{y} - \vec{h}[\vec{x}_B] - \mathbf{H}\delta\vec{x})^T\mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B] - \mathbf{H}\delta\vec{x}), \\ &= \frac{1}{2}\delta\vec{x}^T\mathbf{B}^{-1}\delta\vec{x} + \frac{1}{2}(\mathbf{H}\delta\vec{x} - \{\vec{y} - \vec{h}[\vec{x}_B]\})^T\mathbf{R}^{-1}(\mathbf{H}\delta\vec{x} - \{\vec{y} - \vec{h}[\vec{x}_B]\}). \end{aligned}$$

$$\nabla_x J[\delta\vec{x} = \delta\vec{x}_A] = \mathbf{B}^{-1}\delta\vec{x}_A + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\delta\vec{x}_A - \{\vec{y} - \vec{h}[\vec{x}_B]\}) = 0,$$

$$(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\,\delta\vec{x}_A = \mathbf{H}^T\mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B]),$$

$$\delta\vec{x}_A = \vec{x}_A - \vec{x}_B = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B]). \tag{7}$$

$$(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\,\mathbf{B}\mathbf{H}^T = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T), \tag{8}$$

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

$$\vec{x}_A - \vec{x}_B = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\vec{y} - \vec{h}[\vec{x}_B]). \tag{9}$$

C.8: Why is 3d-Var. favoured over optimal interpolation?

TABLE 1. % REDUCTION IN RMS FIT OF OBSERVATIONS TO ANALYSIS (T+0) AND BACKGROUND (T+6) IN 3DVAR COMPARED WITH AC SCHEME IN THE NORTHERN HEMISPHERE IN JULY 98 TRIAL

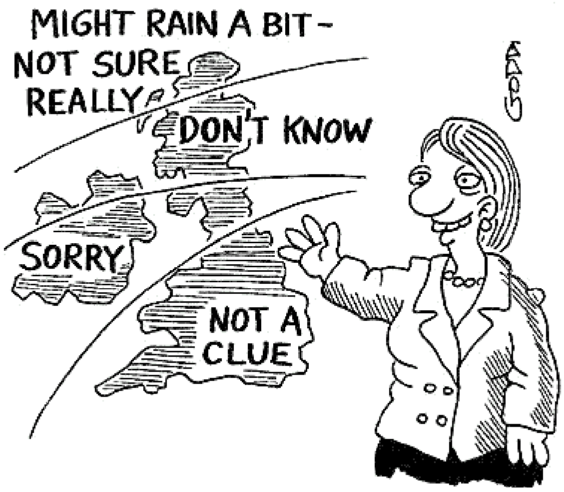
	Temperature		Height or PMSL		Vector Wind		Relative Humidity	
Level	T+0	T+6	T+0	T+6	T+0	T+6	T+0	T+6
100hPa	-5.5	-3.3	-0.1	-3.2	15.2	5.3		
250hPa	0.8	0.0	4.8	2.5	16.8	4.9		
500hPa	5.5	2.7	3.5	5.4	14.4	3.7	5.6	2.9
700hPa	7.2	3.4	2.1	5.2	15.4	2.8	5.7	2.5
850hPa	6.6	1.4	1.4	3.7	9.1	1.8	2.9	1.5
Surface	-1.5	-0.7	6.8	-0.2	1.2	0.6		

Radiosondes TEMP reports used for Upper levels and land SYNOP reports used for Surface

Fig. 11: Performance of the Met Office 3d-Var. scheme for operational weather forecasting vs. the old Analysis Correction (AC) scheme. The AC scheme is a flavour of OI. Taken from Lorenc et al., 2000.

C.9: Why do we need to worry about the error covariance matrices?

Errors are a fundamental consideration in data assimilation: *all models are wrong and all observations are inaccurate*.



$$
\begin{aligned}
\vec{x}_A &= \vec{x}_t + \vec{\varepsilon}_A, & \mathbf{P}_A &= \langle \vec{\varepsilon}_A \vec{\varepsilon}_A^T \rangle, \\
\vec{x}_B &= \vec{x}_t + \vec{\varepsilon}_B, & \mathbf{B} &= \langle \vec{\varepsilon}_B \vec{\varepsilon}_B^T \rangle, \\
\vec{y} &= \vec{h}[\vec{x}_t] + \vec{\varepsilon}_y, & \mathbf{R} &= \langle \vec{\varepsilon}_y \vec{\varepsilon}_y^T \rangle,
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}_A &= (\mathbf{I} - \mathbf{KH})\,\mathbf{B} && (15) \\
&= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} && (11) \\
&= \mathbf{A}^{-1}
\end{aligned}
$$

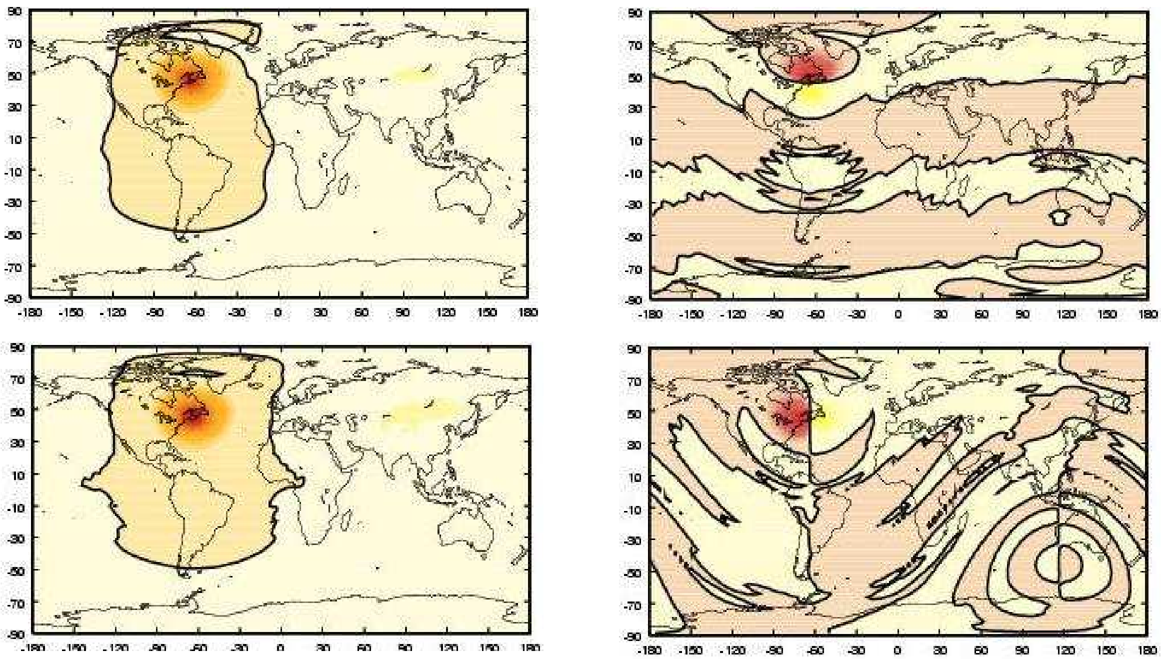


Fig. 12: Analysis increments in Var., $\delta\vec{x}_A$.

Section D. The Gradient And Hessian Of The Cost Function

D.1: What is the gradient vector?

$$\nabla_x J = \frac{\partial J}{\partial \vec{x}} = \begin{pmatrix} \partial J / \partial x_1 \\ \partial J / \partial x_2 \\ \dots \\ \partial J / \partial x_n \end{pmatrix}. \quad (16)$$

D.2 How can the gradient vector be calculated?

$$\begin{aligned} \nabla_x J &\approx \begin{pmatrix} (J[x_1 + \delta_1] - J[x_1 - \delta_1]) / 2\delta_1 \\ (J[x_2 + \delta_2] - J[x_2 - \delta_2]) / 2\delta_2 \\ \dots \\ (J[x_n + \delta_n] - J[x_n - \delta_n]) / 2\delta_n \end{pmatrix}. \\ &= \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) - \mathbf{H}^T\mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}]). \end{aligned} \quad (22)$$

D.3: What is the Hessian matrix?

$$\mathbf{A} = \frac{\partial^2 J}{\partial \vec{x}^2} = \begin{pmatrix} \partial^2 J / \partial x_1^2 & \partial^2 J / \partial x_1 \partial x_2 & \dots & \partial^2 J / \partial x_1 \partial x_n \\ \partial^2 J / \partial x_2 \partial x_1 & \partial^2 J / \partial x_2^2 & \dots & \partial^2 J / \partial x_2 \partial x_n \\ \dots & \dots & \dots & \dots \\ \partial^2 J / \partial x_n \partial x_1 & \partial^2 J / \partial x_n \partial x_2 & \dots & \partial^2 J / \partial x_n^2 \end{pmatrix}. \quad (23)$$

D.4: How can the Hessian be derived?

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

Why is this useful?

How is the gradient vector used in the Var. algorithm?

Section E: Example Observation Operators

E.1: Interpolation of temperature in a single column

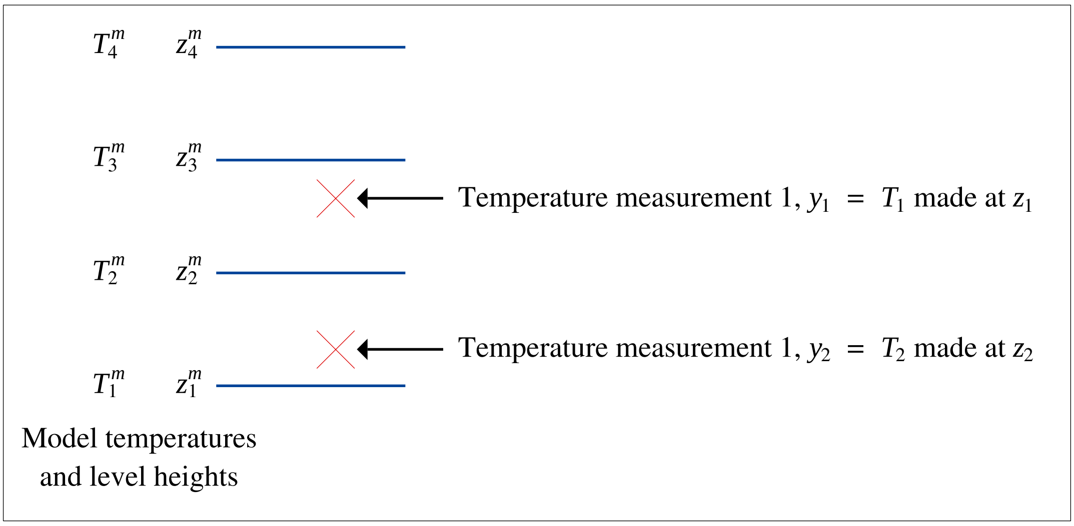


Fig. 13: The model levels and the observations.

E.2: Non-linear forward operator (radiative emission)

$$F = \kappa T^4,$$

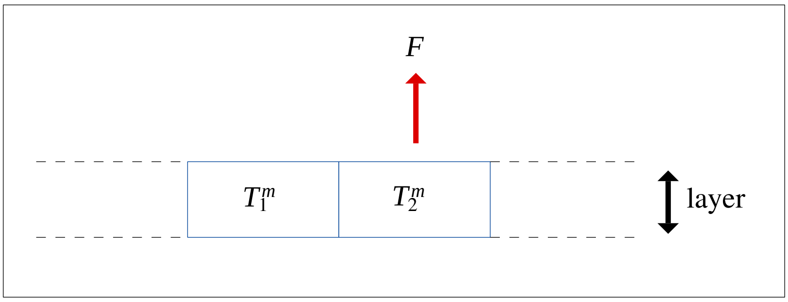


Fig. 14: Two grid boxes making up a layer of the atmosphere whose thermal radiation is being monitored by a satellite instrument.

Section F: Minimization (or Descent) Algorithms

F.1: What is a minimization (or descent) algorithm and what is the geometric interpretation of the gradient vector?

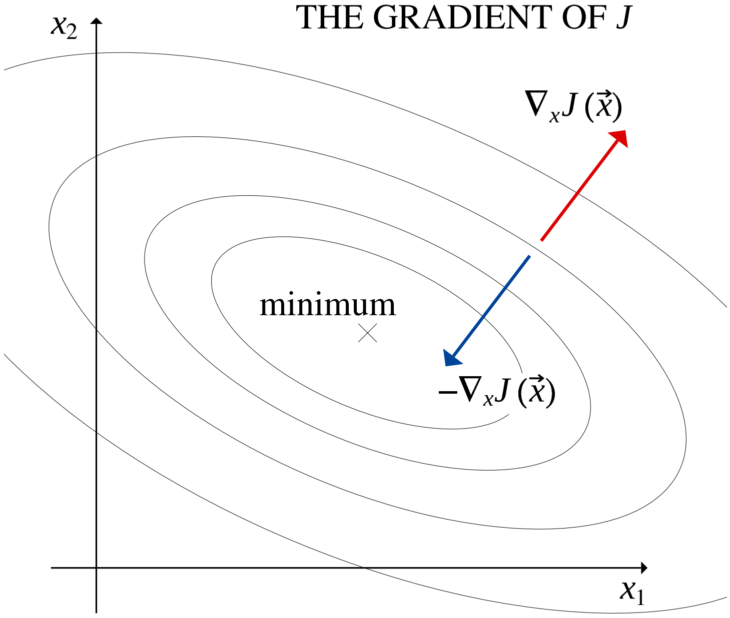


Fig. 15: The gradient vector (red), and its negative (blue) in state space.

F.2: What is the method of steepest descent?

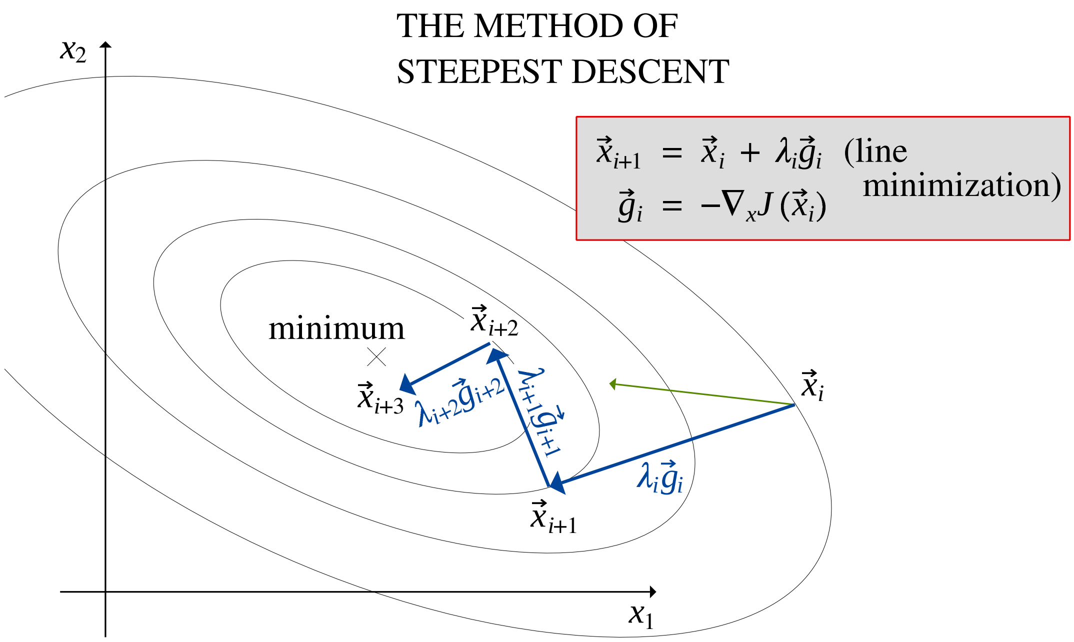


Fig. 16: Schematic of the method of steepest descent.

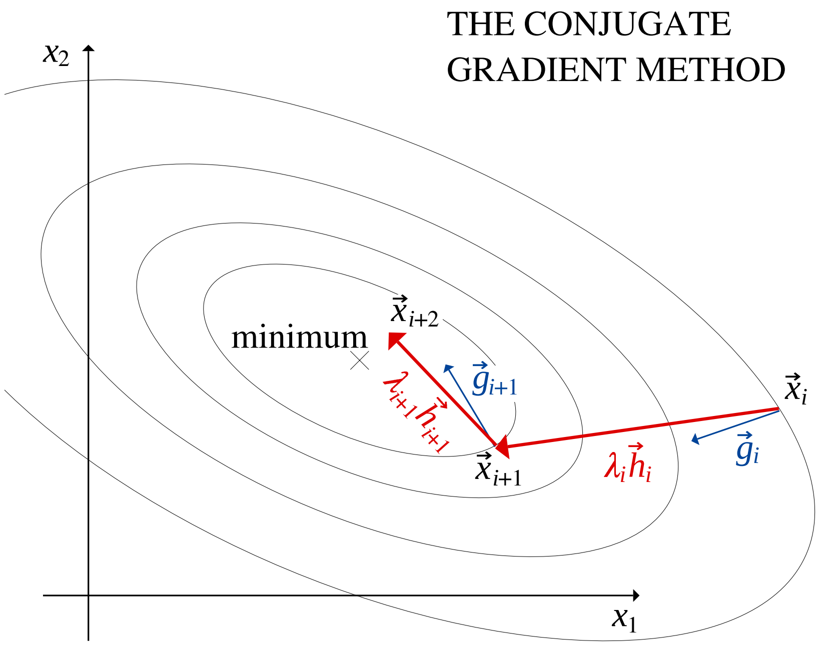
F.3: What is the Newton algorithm?

$$J[\vec{x}_{i+1}] = J[\vec{x}_i] + (\nabla_x J[\vec{x}_i])^T (\vec{x}_{i+1} - \vec{x}_i) + \frac{1}{2}(\vec{x}_{i+1} - \vec{x}_i)^T \mathbf{A} (\vec{x}_{i+1} - \vec{x}_i) . \quad (27)$$

$$\nabla_x J[\vec{x}_{i+1}] = \nabla_x J[\vec{x}_i] + \mathbf{A}(\vec{x}_{i+1} - \vec{x}_i) , \quad (28)$$

$$\vec{x}_{i+1} = \vec{x}_i - \mathbf{A}^{-1} \nabla_x J[\vec{x}_i] . \quad (29)$$

F.4: What is the conjugate gradient algorithm?



$$\vec{x}_{i+1} = \vec{x}_i + \lambda_i \vec{h}_i \quad \text{(line minimization)}$$

$$\vec{g}_{i+1} = -\nabla_x J(\vec{x}_{i+1})$$

$$\vec{h}_{i+1} = \vec{g}_{i+1} + \gamma_i \vec{h}_i$$

$$\gamma_i = \frac{\vec{g}_{i+1}^T \vec{g}_{i+1}}{\vec{g}_i^T \vec{g}_i}$$

$$\left.\begin{array}{l} \vec{g}_i^T \vec{g}_j = 0 \\ \vec{h}_i^T \mathbf{A} \vec{h}_j = 0 \end{array}\right\} \quad i \neq j, \ \mathbf{A} \text{ Hessian}$$

Fig. 17: Schematic of the conjugate gradient algorithm.

Section G: Preconditioning And Control Variable Transforms

G.3: What is meant by 'better conditioned'?

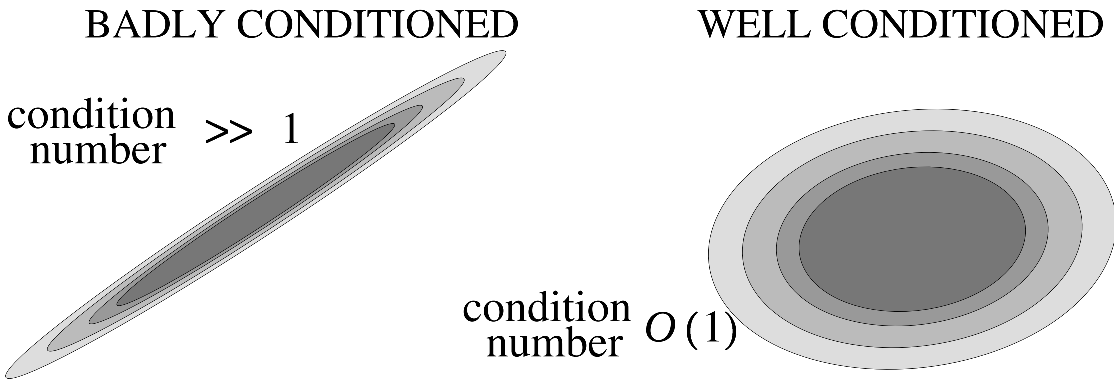


Fig. 18: Contours of *J* illustrating a high conditioning number (left) and a low conditioning number (right).