

PRINCIPLES OF RETRIEVAL THEORY

Clive D Rodgers

Atmospheric, Oceanic and Planetary Physics

University of Oxford

ESA Advanced Atmospheric Training Course

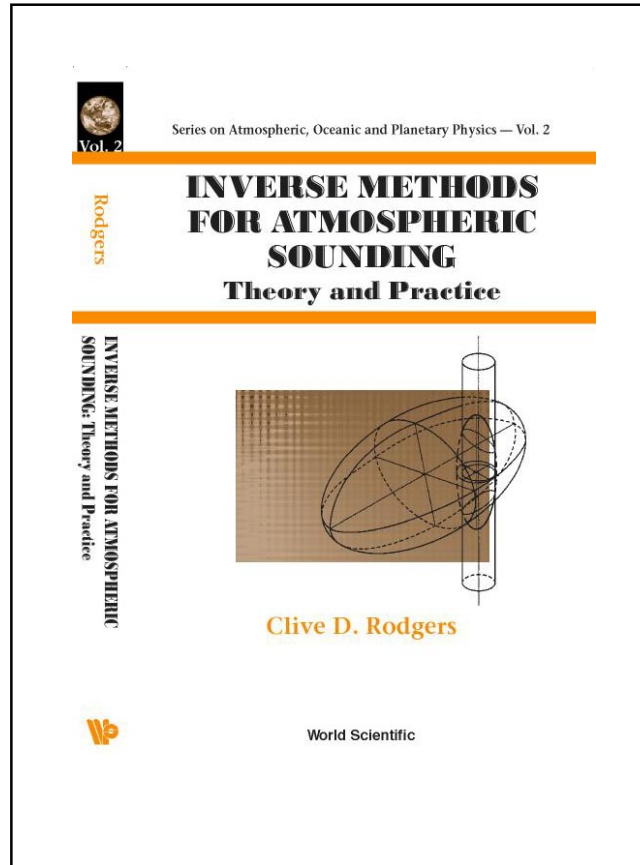
September 15th – 20th, 2008

ATMOSPHERIC REMOTE SENSING: THE INVERSE PROBLEM

Topics

1. Introduction
2. Bayesian approach
3. Information content
4. Error analysis and characterisation

ADVERTISEMENT



C. D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*,
World Scientific Publishing Co., 2000.

WHAT IS AN INVERSE OR RETRIEVAL PROBLEM?

- **Almost any measurement you make...**
 - When you measure some function of the quantity you really want, you have a retrieval problem.
 - Sometimes it's trivial, sometimes it isn't.
- **Various aspects:**
 - *Formulate the problem properly:*
 - Describe the measurement in terms of some Forward Model
 - Don't forget experimental error!
 - *Finding a solution, inverting the forward model*
 - Algebraic
 - Numerical
 - No unique solution
 - No solution at all
 - *Finding the 'best' solution*
 - Uniqueness - a unique solution may not be the best...
 - Accuracy
 - Efficiency
 - *Understanding the answer*

THINGS TO THINK ABOUT

- **Why isn't the problem trivial?**
 - Forward models which are not explicitly invertible
 - Ill-conditioned or ill-posed problems
 - Errors in the measurement (and in the forward model) can map into errors in the solution in a non-trivial way.
- **What to measure?**
 - Does it actually contain the information you want?
- **Updating existing knowledge**
 - You always have *some* prior knowledge of the 'unknown'
 - the measurement improves that knowledge
 - the measurement may not be enough by itself to completely determine the unknown
- **Ill-posed problems**
 - You cannot solve an ill-posed problem. You have to convert it into a well-posed problem.
 - Which of an infinite manifold of solutions do you want?

MATHEMATICAL CONCEPTS I

- **Measurement Vector:** $y = (y_1, y_2, \dots, y_m)$

- Any measurement is of a finite number of quantities.
- Arrange them as a vector for computational purposes

- **State Vector:** $x = (x_1, x_2, \dots, x_n)$

- The desired quantity is often continuous - e.g. a temperature profile
- We can only make a finite number of measurements and calculations
- Express the unknown in terms of a finite number of parameters
- They do not all have to be of the same type
- Arrange them as a vector for computational purposes
- Examples:
 - Temperature on a set of pressure levels, with a specified interpolation rule.
 - Fourier coefficients for a set of waves

MATHEMATICAL CONCEPTS II

Using vectors, it is convenient to think in terms of linear algebra and vector spaces - even if the forward model is not linear.

Measurement Space

– Measurement space is the space of measurement vectors, dimension m .

State Space

– State space is the space of state vectors, dimension n .

Generally the two vector spaces will have different dimensions.

MATHEMATICAL CONCEPTS III

Forward Function and Model

- The Forward *Function* $f(\mathbf{x})$ maps from state space onto measurement space, depending on the physics of the measurement.
- The Forward *Model* $\mathbf{F}(\mathbf{x})$ is the best we can do in the circumstances to model the forward function

Inverse or Retrieval Method

- The inverse problem is one of finding an inverse mapping $\mathbf{R}(\mathbf{y})$:

Given a point in measurement space, which point or set of points in state space could have mapped into it?

NOTATION

I have tried to make it mnemonic as far as possible:

Matrices		Bold upper case
Vectors		Bold lower case
State vectors	\mathbf{x}	
Measurement vectors	\mathbf{y}	
Covariance matrices	\mathbf{S}	(based on σ , and not wanting to use Σ)
Measurement error	ϵ	
Forward model	$\mathbf{F}(\mathbf{x})$	(Really ought to be $\mathbf{f}(\mathbf{x})$ it's a vector)
Jacobian	\mathbf{K}	(originally a K ernel of an integral transform)
Gain matrix	\mathbf{G}	(I've used up \mathbf{K} , which might stand for K alman gain)
Averaging Kernel	\mathbf{A}	

Different vectors, matrices of the same type are distinguished by superscripts, subscripts, etc:

a priori	\mathbf{x}_a	(background)
estimate	$\hat{\mathbf{x}}$	
first guess	\mathbf{x}_0	
'true' value	\mathbf{x}	(no subscript)

STANDARD ILLUSTRATION

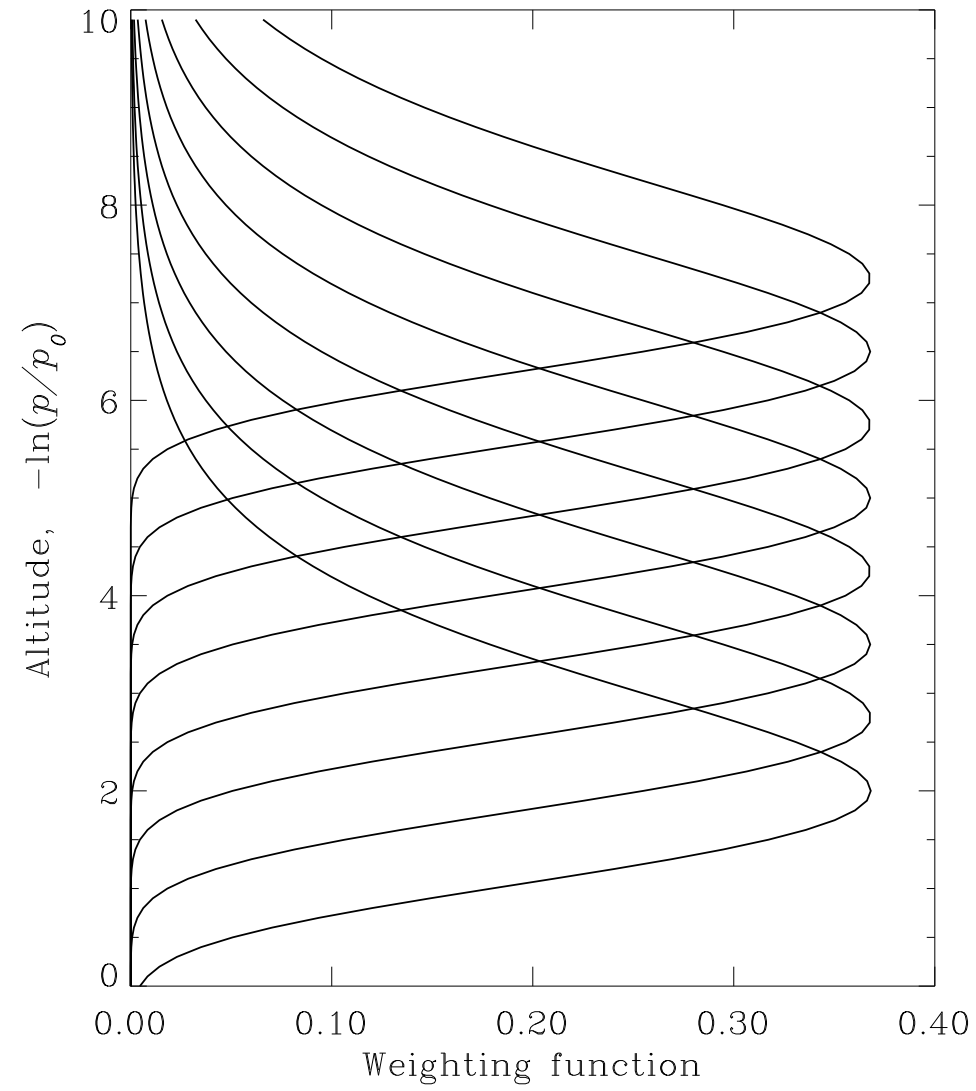
Idealised thermal-emission nadir sounder represented as a linear forward model:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \epsilon$$

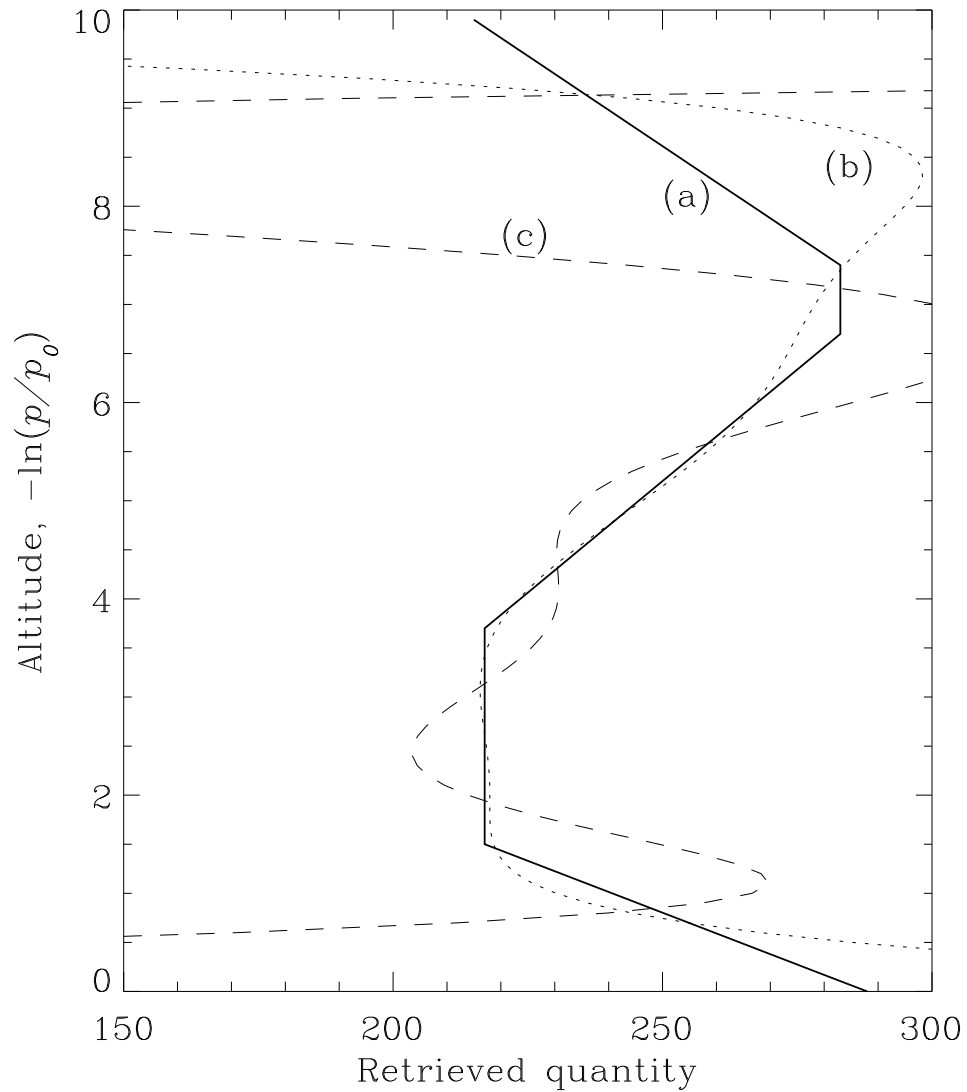
\mathbf{K} is the 'weighting function' matrix, ϵ is measurement error or noise.

- Vertical coordinate is notionally $\ln(p)$, discretised at 100 levels from 0 in steps of 0.1 to 9.9 – around 0 to 70 km.
- Eight channels (elements of \mathbf{y}).
- State vector is notionally temperature at 100 levels.
- Measurement error (when considered) is 0.5 K.

STANDARD WEIGHTING FUNCTIONS



EXACT RETRIEVAL SIMULATION



The state vector is a set of eight coefficients of a degree seven polynomial.

- (a)** Original profile: US standard atmosphere
- (b)** Exact retrieval with no experimental error
- (c)** Exact retrieval with simulated 0.5 K error

NOISE FREE MEASUREMENTS

Row Space and Null Space

Consider an error-free linear measurement, equivalent to solving linear equations:

$$\mathbf{y} = \mathbf{K}\mathbf{x}$$

The *rows* of \mathbf{K} are the weighting functions \mathbf{k}_i :

$$y_i = \mathbf{k}_i^T \mathbf{x}$$

The \mathbf{k}_i are a set of vectors in state space; the measurements are projections of the state \mathbf{x} onto them.

They span a subspace called the *row space*, of dimension equal to the rank of \mathbf{K} , $p \leq \min(n, m)$. If $p < m$ then the weighting functions are not linearly independent.

Only those components of \mathbf{x} in the row space can be measured.

The *null space* is the part of state space which is not in the row space.

ILL-POSED AND WELL-POSED PROBLEMS

Ill or well posed? . . . Under- or over-determined? . . . Under- or over-constrained?

Which is which?

- 1. $p = m = n$. *Well posed.*

The number of unknowns is equal to the number of measurements, and they are all independent.

- 2. $p = m < n$. *Underconstrained, ill-posed*

More unknowns than measurements, but the measurements are all independent.

- 3. $p = n < m$. *Overconstrained, ill-posed*

More measurements than unknowns, so they could be inconsistent, but the unknowns are all in the row space, so there is information about all of them.

ILL-POSED AND WELL-POSED PROBLEMS II

Another category: **Mixed-determined**, the problem is both underconstrained and overconstrained.

1. $p < m = n$.

The number of unknowns is equal to the number of measurements, but the measurements are not independent, so they could be inconsistent, and the number of independent pieces of information is less than the number of unknowns. Simple example:

$$y_1 = x_1 + x_2 + \epsilon_1 \quad (1)$$

$$y_2 = x_1 + x_2 + \epsilon_2 \quad (2)$$

2. $p < m < n$.

More unknowns than measurements, but the measurements are not independent, so they could be inconsistent.

3. $p < n < m$.

More measurements than the rank, so they could be inconsistent, more unknowns than the rank, so not all are defined by the measurement.

ILL-POSED AND WELL-POSED PROBLEMS III

Summary

If $p < n$ then the system is underconstrained; there is a null space.

If $p < m$ then the system is overconstrained in some part of the row space.

How do we identify the row and null spaces?

One straightforward way is Gram-Schmidt orthogonalisation, but . . .

SINGULAR VECTOR DECOMPOSITION [\gg]

Is the neatest way of doing the job.

MATRIX ALGEBRA – EIGENVECTORS

The eigenvalue problem associated with an Arbitrary square matrix \mathbf{A} , of order n , is to find *eigenvectors* \mathbf{l} and scalar *eigenvalues* λ which satisfy

$$\mathbf{A}\mathbf{l} = \lambda\mathbf{l}$$

If \mathbf{A} is a coordinate transformation, then \mathbf{l} has the same representation in the untransformed and transformed coordinates, apart from a factor λ .

This is the same as $(\mathbf{A} - \lambda\mathbf{I})\mathbf{l} = 0$, a homogeneous set of equations, which can only have a solution if $|\mathbf{A} - \lambda\mathbf{I}| = 0$, giving a polynomial equation of degree n , with n solutions for λ . They will be complex in general.

An eigenvector can be scaled by an arbitrary factor. It is conventional to normalise them so that $\mathbf{l}^T\mathbf{l} = 1$ or $\mathbf{l}^\dagger\mathbf{l} = 1$ (Hermitian adjoint)

EIGENVECTORS II

We can assemble the eigenvectors as columns in a matrix \mathbf{L} :

$$\mathbf{A}\mathbf{L} = \mathbf{L}\mathbf{\Lambda}$$

where $\mathbf{\Lambda}$ is a diagonal matrix, with the eigenvalues on the diagonal. ■

Transpose	$\mathbf{L}^T \mathbf{A}^T = \mathbf{\Lambda} \mathbf{L}^T$
Multiply by $\mathbf{R} = (\mathbf{L}^T)^{-1}$	$\mathbf{A}^T = \mathbf{R} \mathbf{\Lambda} \mathbf{L}^T$
Postmultiply by \mathbf{R}	$\mathbf{A}^T \mathbf{R} = \mathbf{R} \mathbf{\Lambda}$

Thus:

\mathbf{R} is the matrix of eigenvectors of \mathbf{A}^T . ■

\mathbf{A}^T has the same eigenvalues as \mathbf{A} . ■

In the case of a Symmetric matrix, $\mathbf{S} = \mathbf{S}^T$ we must have $\mathbf{L} = \mathbf{R}$, so that $\mathbf{L}^T \mathbf{L} = \mathbf{L} \mathbf{L}^T = \mathbf{I}$ or $\mathbf{L}^T = \mathbf{L}^{-1}$, and the eigenvectors are *orthogonal*.

In this case the eigenvalues are real.

EIGENVECTORS – GEOMETRIC INTERPRETATION

Consider the scalar equation:

$$\mathbf{x}^T \mathbf{S} \mathbf{x} = 1$$

where \mathbf{S} is symmetric. This is the equation of a quadratic surface centered on the origin, in n -space. ■

The normal to the surface is the vector $\partial(\mathbf{x}^T \mathbf{S} \mathbf{x}) / \partial \mathbf{x}$, i.e. $\mathbf{S} \mathbf{x}$, and \mathbf{x} is the radius vector, so

$$\mathbf{S} \mathbf{x} = \lambda \mathbf{x}$$

is the problem of finding points where the normal and the radius vector are parallel. These are where the principal axes intersect the surface. ■

At these points, $\mathbf{x}^T \mathbf{S} \mathbf{x} = 1$, so $\mathbf{x}^T \lambda \mathbf{x} = 1$ or:

$$\lambda = \frac{1}{\mathbf{x}^T \mathbf{x}}$$

So the eigenvalues are the reciprocals of the squares of the lengths of the principal axes.

GEOMETRIC INTERPRETATION II

The lengths are independent of the coordinate system, so will also be invariant under an arbitrary orthogonal transformation, i.e. one in which $(distance)^2 = \mathbf{x}^T \mathbf{x}$ is unchanged.

Consider using the eigenvectors of \mathbf{S} to transform the equation for the quadratic surface:

$$\mathbf{x}^T \mathbf{L} \mathbf{\Lambda} \mathbf{L}^T \mathbf{x} = 1 \quad \text{or} \quad \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = 1 \quad \text{or} \quad \sum \lambda_i y_i^2 = 1$$

where $\mathbf{y} = \mathbf{L}^T \mathbf{x}$ or $\mathbf{x} = \mathbf{L} \mathbf{y}$. This transforms the surface into its principal axis representation.

EIGENVECTORS - USEFUL RELATIONSHIPS

Asymmetric Matrices

$$\mathbf{AR} = \mathbf{R}\Lambda$$

$$\mathbf{L}^T \mathbf{A} = \Lambda \mathbf{L}$$

$$\mathbf{L}^T = \mathbf{R}^{-1}, \mathbf{R}^T = \mathbf{L}^{-1}$$

$$\mathbf{LR}^T = \mathbf{L}^T \mathbf{R} = \mathbf{I}$$

$$\mathbf{A} = \mathbf{R}\Lambda \mathbf{L}^T = \sum \lambda_i \mathbf{r}_i \mathbf{l}_i^T$$

$$\mathbf{A}^T = \mathbf{L}\Lambda \mathbf{R}^T = \sum \lambda_i \mathbf{l}_i \mathbf{r}_i^T$$

$$\mathbf{A}^{-1} = \mathbf{R}\Lambda^{-1} \mathbf{L}^T$$

$$\mathbf{A}^n = \mathbf{R}\Lambda^n \mathbf{L}^T$$

$$\mathbf{L}^T \mathbf{A} \mathbf{R} = \Lambda$$

$$\mathbf{L}^T \mathbf{A}^n \mathbf{R} = \Lambda^n$$

$$\mathbf{L}^T \mathbf{A}^{-1} \mathbf{R} = \Lambda^{-1}$$

$$|\mathbf{A}| = \prod_i \lambda_i$$

Symmetric Matrices

$$\mathbf{SL} = \mathbf{L}\Lambda$$

$$\mathbf{L}^T = \mathbf{L}^{-1}$$

$$\mathbf{LL}^T = \mathbf{L}^T \mathbf{L} = \mathbf{I}$$

$$\mathbf{S} = \mathbf{L}\Lambda \mathbf{L}^T = \sum \lambda_i \mathbf{l}_i \mathbf{l}_i^T$$

$$\mathbf{S}^{-1} = \mathbf{L}\Lambda^{-1} \mathbf{L}^T$$

$$\mathbf{S}^n = \mathbf{L}\Lambda^n \mathbf{L}^T$$

$$\mathbf{L}^T \mathbf{S} \mathbf{L} = \Lambda$$

$$\mathbf{L}^T \mathbf{S}^n \mathbf{L} = \Lambda^n$$

$$\mathbf{L}^T \mathbf{S}^{-1} \mathbf{L} = \Lambda^{-1}$$

$$|\mathbf{A}| = \prod_i \lambda_i$$

SINGULAR VECTOR DECOMPOSITION

The standard eigenvalue problem is meaningless for non-square matrices.

A 'shifted' eigenvalue problem associated with an arbitrary non-square matrix \mathbf{K} , m rows and n columns can be constructed:

$$\begin{aligned}\mathbf{K}\mathbf{v} &= \lambda\mathbf{u} \\ \mathbf{K}^T\mathbf{u} &= \lambda\mathbf{v}\end{aligned}\tag{3}$$

where \mathbf{v} , of length n , and \mathbf{u} , of length m , are called the *singular vectors* of \mathbf{K} .

This is equivalent to the symmetric problem:

$$\begin{pmatrix} \mathbf{O} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

From (3) we can get

$$\begin{aligned}\mathbf{K}^T\mathbf{K}\mathbf{v} &= \lambda\mathbf{K}^T\mathbf{u} = \lambda^2\mathbf{v} \\ \mathbf{K}\mathbf{K}^T\mathbf{u} &= \lambda\mathbf{K}\mathbf{v} = \lambda^2\mathbf{u}\end{aligned}\tag{4}$$

so \mathbf{u} and \mathbf{v} are the eigenvectors of $\mathbf{K}\mathbf{K}^T$ ($m \times m$) and $\mathbf{K}^T\mathbf{K}$ ($n \times n$) respectively.

SINGULAR VECTOR DECOMPOSITION II

Care is needed in constructing a matrix of singular vectors, because individual \mathbf{u} and \mathbf{v} vectors correspond to each other, yet there are potentially different numbers of \mathbf{v} and \mathbf{u} vectors. ■

If the rank of \mathbf{K} is p , then there will be p non-zero singular values, and both $\mathbf{K}\mathbf{K}^T$ and $\mathbf{K}^T\mathbf{K}$ will have p non-zero eigenvalues. ■

The surplus eigenvectors will have zero eigenvalues, and can be discarded and we can write:

$$\begin{pmatrix} \mathbf{O} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \mathbf{\Lambda}$$

where $\mathbf{\Lambda}$ is $p \times p$, \mathbf{U} is $m \times p$, and \mathbf{V} is $n \times p$.

There will be $n + m - p$ more eigenvectors of the composite matrix, all with zero eigenvalue.

SINGULAR VECTORS - USEFUL RELATIONSHIPS

$$\begin{aligned}\mathbf{KV} &= \mathbf{U}\Lambda \\ \mathbf{K}^T\mathbf{U} &= \mathbf{V}\Lambda \\ \mathbf{U}^T\mathbf{KV} &= \mathbf{V}^T\mathbf{K}^T\mathbf{U} = \lambda \\ \mathbf{K} &= \mathbf{U}\Lambda\mathbf{V}^T \\ \mathbf{K}^T &= \mathbf{V}\Lambda\mathbf{U}^T \\ \mathbf{V}^T\mathbf{V} &= \mathbf{U}^T\mathbf{U} = \mathbf{I}_p \\ \mathbf{KK}^T\mathbf{U} &= \mathbf{U}\Lambda^2 \\ \mathbf{K}^T\mathbf{KV} &= \mathbf{V}\Lambda^2\end{aligned}$$

(5)

SINGULAR VECTOR DECOMPOSITION [\llcorner]

Is the neatest way of doing the job.

Express \mathbf{K} as

$${}_m\mathbf{K}_n = {}_m\mathbf{U}_p \mathbf{\Lambda}_p \mathbf{V}_n^T$$

where the subscripts indicate the sizes of the matrices. ■

SINGULAR VECTOR DECOMPOSITION [\llcorner]

Is the neatest way of doing the job

Express \mathbf{K} as

$${}_m\mathbf{K}_n = {}_m\mathbf{U}_p\mathbf{\Lambda}_p\mathbf{V}_n^T$$

where the subscripts indicate the sizes of the matrices.

Then the forward model (no noise) becomes:

$${}_m\mathbf{y}_1 = {}_m\mathbf{K}_n\mathbf{x}_1 = {}_m\mathbf{U}_p\mathbf{\Lambda}_p\mathbf{V}_n^T\mathbf{x}_1$$

so that

$${}_p\mathbf{U}_m^T\mathbf{y}_1 = {}_p\mathbf{\Lambda}_p\mathbf{V}_n^T\mathbf{x}_1$$

or

$$\mathbf{y}' = \mathbf{\Lambda}\mathbf{x}'$$

where $\mathbf{y}' = \mathbf{U}^T\mathbf{y}$ and $\mathbf{x}' = \mathbf{V}^T\mathbf{x}$ are both of order p .

The rows of \mathbf{V}^T , or the columns of \mathbf{V} (in state space) are a basis for the row space of \mathbf{K} . Similarly the columns of \mathbf{U} (in measurement space) are a basis of its *column space*.

SINGULAR VECTOR DECOMPOSITION II

We can also see that *an* exact solution is

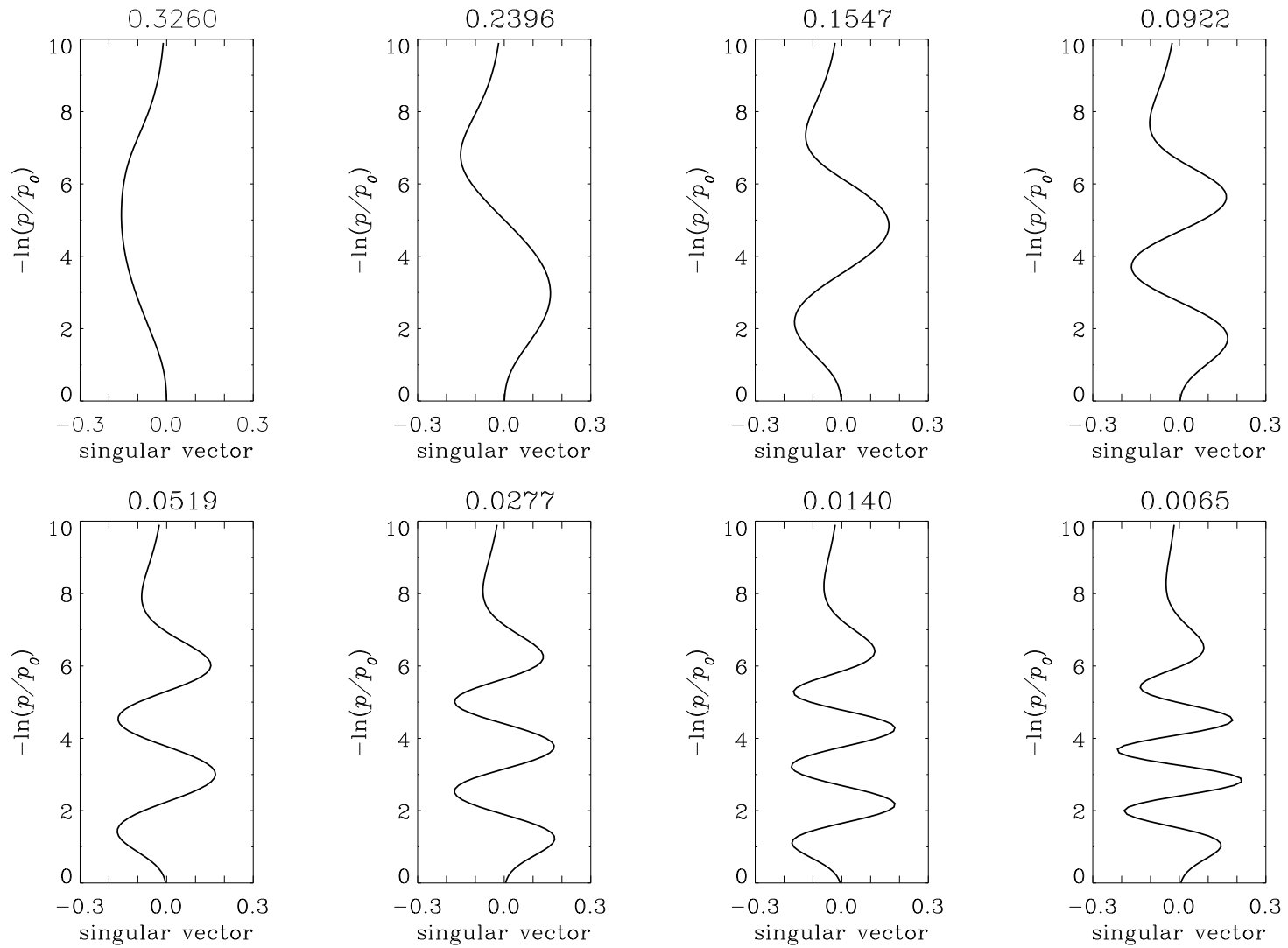
$$\mathbf{x}' = \mathbf{\Lambda}^{-1}\mathbf{y}'$$

or

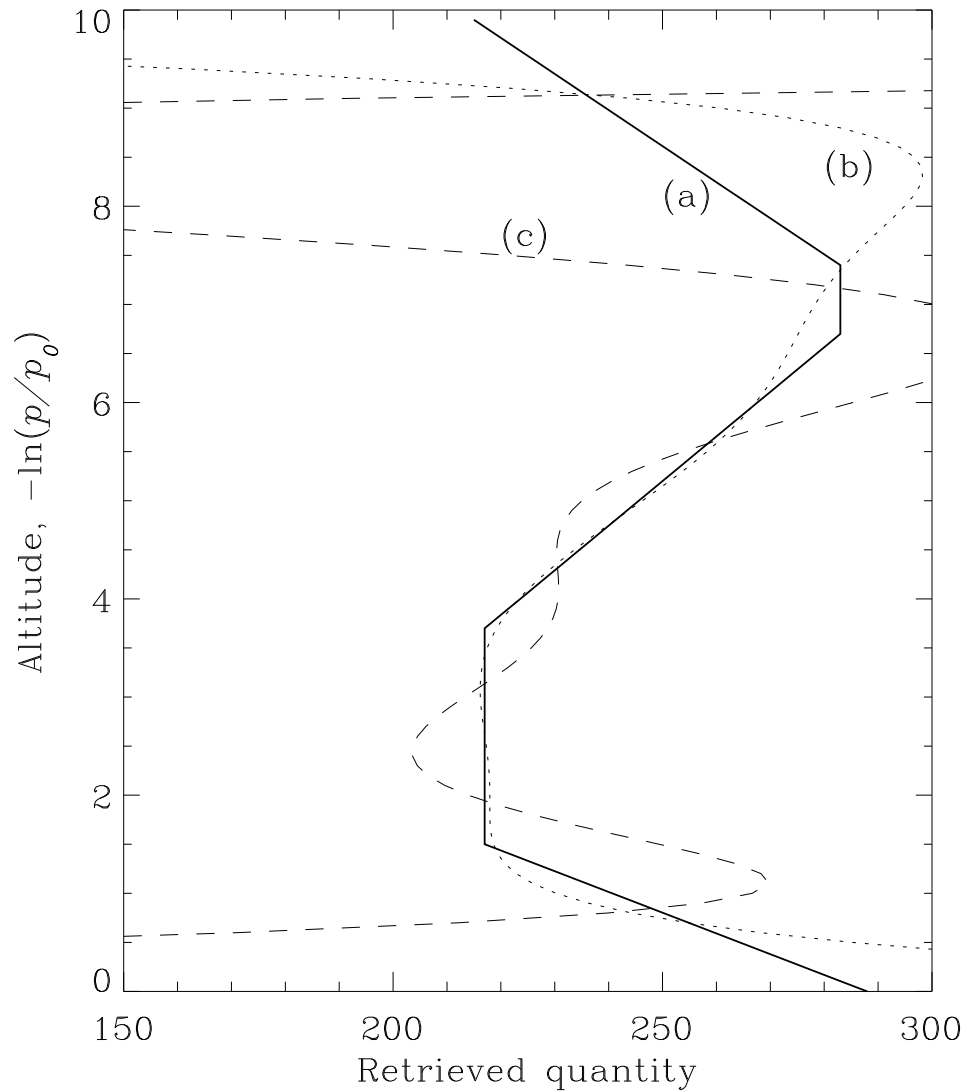
$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{y}$$

This is only a unique solution if $p = n$. If $p < n$ any multiples of vectors with zero singular values can be added, and still satisfy the equations.

SVD OF THE STANDARD WEIGHTING FUNCTIONS



EXACT RETRIEVAL SIMULATION



The state vector is a set of eight coefficients of a degree seven polynomial.

- (a)** Original profile: US standard atmosphere
- (b)** Exact retrieval with no experimental error
- (c)** Exact retrieval with simulated 0.5 K error

APPROACHES TO INVERSE PROBLEMS

- **Bayesian Approach**

- What is the pdf of the state, given the measurement and the *a priori* ?

- **Optimisation Approaches:**

- Maximum Likelihood
- Maximum A Posteriori
- Minimum Variance
- Backus-Gilbert - resolution/noise trade-off

- **Ad hoc Approaches**

- Relaxation
- Exact algebraic solutions

BAYESIAN APPROACH

This is the most general approach to the problem (that I know of).

Knowledge is represented in terms of *probability density functions*:

- $P(\mathbf{x})$ is the *a priori* p.d.f. of the state, describing what we know about the state before we make the measurement. ■
- $P(\mathbf{y})$ is the *a priori* p.d.f. of the measurement. ■
- $P(\mathbf{x}, \mathbf{y})$ is the joint *a priori* p.d.f. of \mathbf{x} and \mathbf{y} . ■
- $P(\mathbf{y}|\mathbf{x})$ is the p.d.f. of the measurement given the state - this depends on experimental error and the forward function. ■
- $P(\mathbf{x}|\mathbf{y})$ is the p.d.f. of the state given the measurement - this is what we want to find.

BAYES THEOREM

The theorem states:

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y})$$

and of course

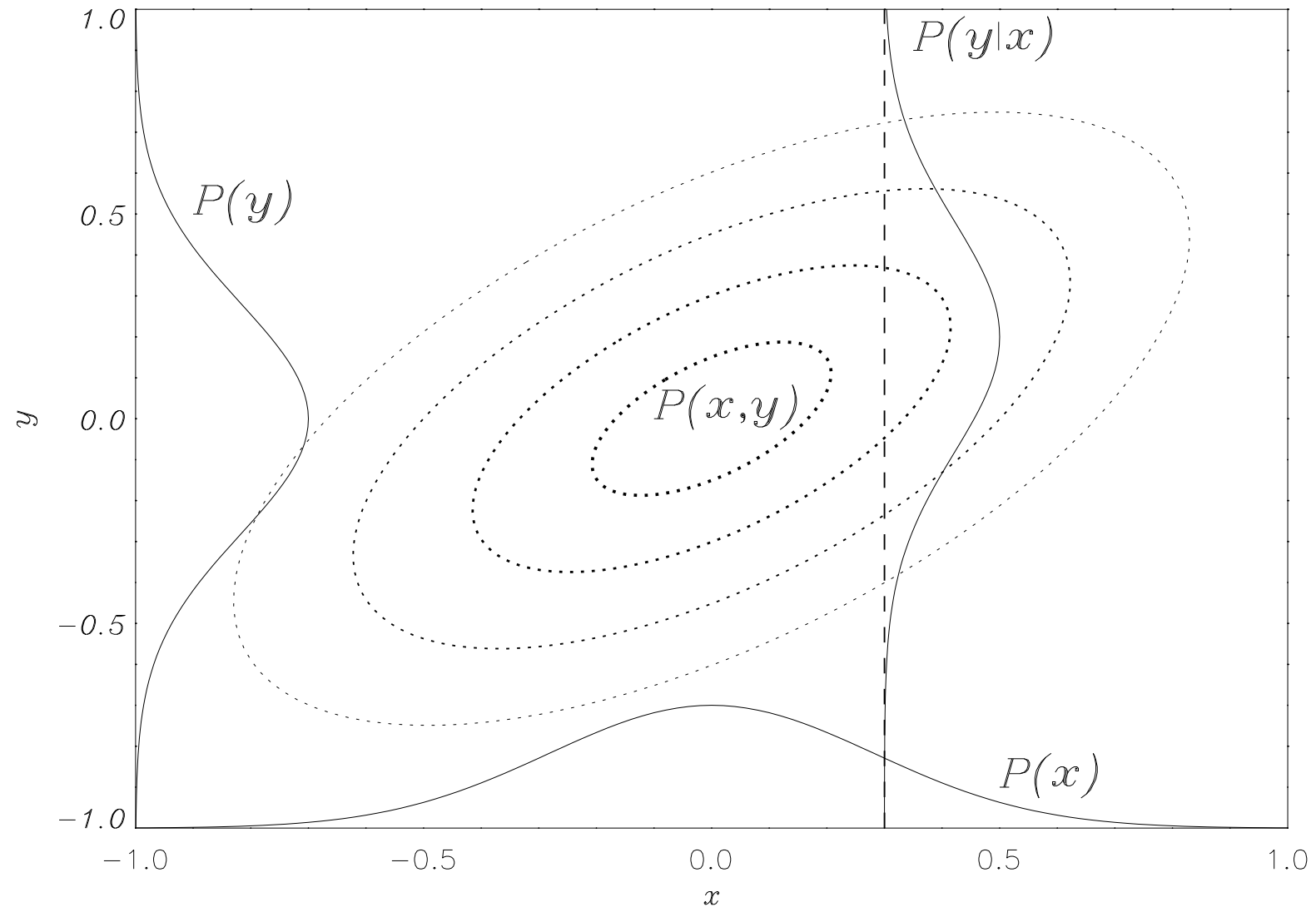
$$P(\mathbf{y}, \mathbf{x}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$$

so that

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}$$

If we have a prior p.d.f. for \mathbf{x} , $\mathbf{P}(\mathbf{x})$, and we know statistically how \mathbf{y} is related to \mathbf{x} *via* $P(\mathbf{y}|\mathbf{x})$, then we can find an un-normalised version of $P(\mathbf{x}|\mathbf{y})$, namely $P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$, which can be normalised if required.

BAYES THEOREM GEOMETRICALLY



APPLICATION OF THE BAYESIAN APPROACH

We need explicit forms for the p.d.f's:

- Assume that experimental error is Gaussian:

$$-\ln P(\mathbf{y}|\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_\epsilon^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x})) + \text{const}$$

where $\mathbf{F}(\mathbf{x})$ is the Forward model:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon$$

and \mathbf{S}_ϵ is the covariance matrix of the experimental error, ϵ :

$$\mathbf{S}_\epsilon = E\{\epsilon\epsilon^T\} = E\{(\mathbf{y} - \mathbf{F}(\mathbf{x}))(\mathbf{y} - \mathbf{F}(\mathbf{x}))^T\}$$

- Assume that the *a priori* p.d.f. is Gaussian (less justifiable):

$$-\ln P(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1}(\mathbf{x} - \mathbf{x}_a) + \text{const}$$

i.e. \mathbf{x} is distributed normally with mean \mathbf{x}_a and covariance \mathbf{S}_a .

APPLICATION OF THE BAYESIAN APPROACH II

- Thus the pdf of the state when the measurements and the *a priori* are given is:

$$-2 \ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a] + \text{const}$$

- If we want a state estimate $\hat{\mathbf{x}}$ rather than a p.d.f., then we must calculate some function of $P(\mathbf{x}|\mathbf{y})$, such as its mean or its maximum

$$\hat{\mathbf{x}} = \int P(\mathbf{x}|\mathbf{y}) \mathbf{x} d\mathbf{x} \quad \text{or} \quad \frac{dP(\hat{\mathbf{x}}|\mathbf{y})}{d\mathbf{x}} = 0$$

BAYESIAN SOLUTION FOR THE LINEAR PROBLEM

The linear problem has a forward model:

$$\mathbf{F}(\mathbf{x}) = \mathbf{K}\mathbf{x}$$

so the p.d.f. $P(\mathbf{x}|\mathbf{y})$ becomes:

$$-2 \ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y} - \mathbf{K}\mathbf{x}]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{K}\mathbf{x}] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a] + c_1$$

This is quadratic in \mathbf{x} , so has to be of the form:

$$-2 \ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{x} - \hat{\mathbf{x}}]^T \hat{\mathbf{S}}^{-1} [\mathbf{x} - \hat{\mathbf{x}}] + c_2$$

Equate the terms that are quadratic in \mathbf{x} :

$$\mathbf{x}^T \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \mathbf{x} + \mathbf{x}^T \mathbf{S}_a^{-1} \mathbf{x} = \mathbf{x}^T \hat{\mathbf{S}}^{-1} \mathbf{x}$$

giving

$$\hat{\mathbf{S}}^{-1} = \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1}$$

- ‘*The Fisher Information Matrix*’.

BAYESIAN SOLUTION FOR THE LINEAR PROBLEM II

Equating the terms linear in \mathbf{x}^T gives:

$$(-\mathbf{K}\mathbf{x})^T \mathbf{S}_\epsilon^{-1}(\mathbf{y}) + (\mathbf{x})^T \mathbf{S}_a^{-1}(-\mathbf{x}_a) = \mathbf{x}^T \hat{\mathbf{S}}^{-1}(-\hat{\mathbf{x}})$$

This must be valid for any \mathbf{x} . Cancel the \mathbf{x}^T 's, and substitute for $\hat{\mathbf{S}}^{-1}$:

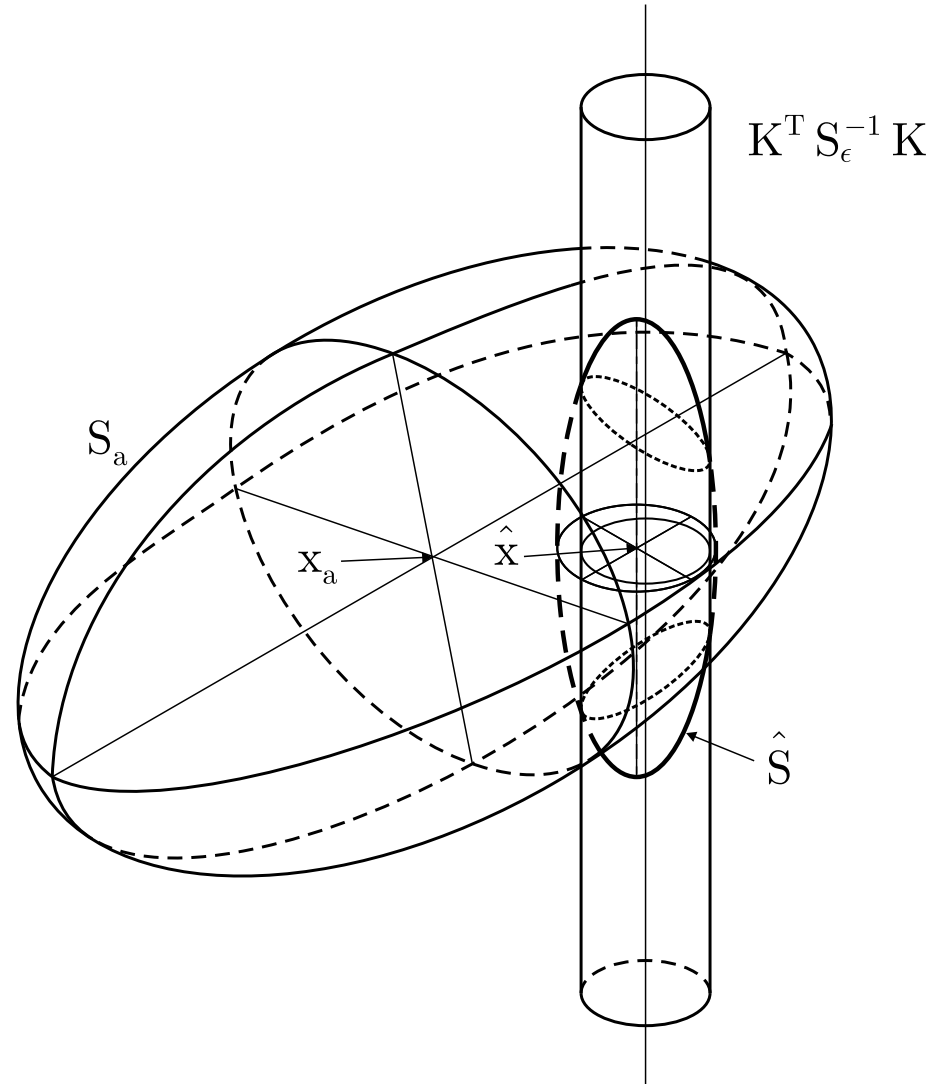
$$\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{y} + \mathbf{S}_a^{-1} \mathbf{x}_a = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1}) \hat{\mathbf{x}}$$

giving:

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{y} + \mathbf{S}_a^{-1} \mathbf{x}_a)$$

The mean $\hat{\mathbf{x}}$ and the covariance $\hat{\mathbf{S}}$ define the full posterior pdf.

A GEOMETRIC INTERPRETATION OF THE SOLUTION



AN ALGEBRAIC INTERPRETATION OF THE SOLUTION

The expected value is:

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{y} + \mathbf{S}_a^{-1} \mathbf{x}_a) \quad (1)$$

Underconstrained case

There must exist at least one 'exact' solution $\mathbf{x}_e = \mathbf{G}\mathbf{y}$ in the sense that $\mathbf{K}\mathbf{x}_e = \mathbf{y}$, i.e. $\mathbf{K}\mathbf{G} = \mathbf{I}$.
For example $\mathbf{G} = \mathbf{K}^T (\mathbf{K}\mathbf{K}^T)^{-1}$.

Replace \mathbf{y} by $\mathbf{K}\mathbf{x}_e$ in (1):

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}\mathbf{x}_e + \mathbf{S}_a^{-1} \mathbf{x}_a)$$

Overconstrained case

The least squares solution \mathbf{x}_l satisfies $\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}\mathbf{x}_l = \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{y}$.

Inserting this in (1) gives:

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}\mathbf{x}_l + \mathbf{S}_a^{-1} \mathbf{x}_a)$$

AN INTERPRETATION OF THE SOLUTION II

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \mathbf{x}_e + \mathbf{S}_a^{-1} \mathbf{x}_a)$$

or

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \mathbf{x}_l + \mathbf{S}_a^{-1} \mathbf{x}_a)$$

- Both represent a weighted mean of a solution (exact, \mathbf{x}_e or least squares \mathbf{x}_l) with \mathbf{x}_a using relative weights $\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}$ and \mathbf{S}_a^{-1} respectively – their Fisher information matrices.
- This is exactly like the familiar combination of scalar measurements x_1 and x_2 of an unknown x , with variances σ_1^2 and σ_2^2 respectively:

$$\hat{x} = (1/\sigma_1^2 + 1/\sigma_2^2)^{-1} (x_1/\sigma_1^2 + x_2/\sigma_2^2)$$

End of Section