Constrained data assimilation

W. Carlisle Thacker
Atlantic Oceanographic and Meteorological Laboratory
Miami, Florida 33149 USA

Plan

Range constraints: : HYCOM layers have minimum thickness.

Optimal interpolation: no inequality constraints.
normally distributed errors.

Graphical interpretation: linear regression.
active constraint $\equiv$ additional data

Algorithms: little knowledge of error statistics.
how precise should solution be?

Simple examples: 4 variables, 1 new observation.

Variational formalism

The objective is to reach a compromise $\mathbf{x}$ between a background estimate $\mathbf{b}$ of the model state and additional information provided by observations $\mathbf{d}$ based on their respective error covariances B and D. The minimum of

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{b})^{\mathsf{T}}\mathsf{B}^{-1}(\mathbf{x}-\mathbf{b}) + \frac{1}{2}(\mathsf{H}\mathbf{x}-\mathbf{d})^{\mathsf{T}}\mathsf{D}^{-1}(\mathsf{H}\mathbf{x}-\mathbf{d})$$

defines this compromise:

$$\mathbf{x}-\mathbf{b} = \mathsf{B}\mathsf{H}^{\mathsf{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathsf{T}}+\mathsf{D})^{-1}(\mathbf{d}-\mathsf{H}\mathbf{b})$$

where H selects model counterparts of $\mathbf{d}$.

Posterior covariance matrix

After assimilation, just as the background state $\mathbf{b}$ is updated to get the analysis $\mathbf{x}$, the background error-covariance matrix B can be updated to get an analysis error-covariance matrix

$$\mathsf{A} = \mathsf{B} - \mathsf{B}\mathsf{H}^{\mathsf{T}}(\mathsf{H}\mathsf{B}\mathsf{H}^{\mathsf{T}} + \mathsf{D})^{-1}\mathsf{H}\mathsf{B}\,.$$

Example: 2 variables and 1 measurement

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \\ x_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \\ b_2 \end{pmatrix} \quad \mathsf{B} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix} \quad \mathbf{d} = d_1 \quad \mathsf{D} = \sigma_d^2$$

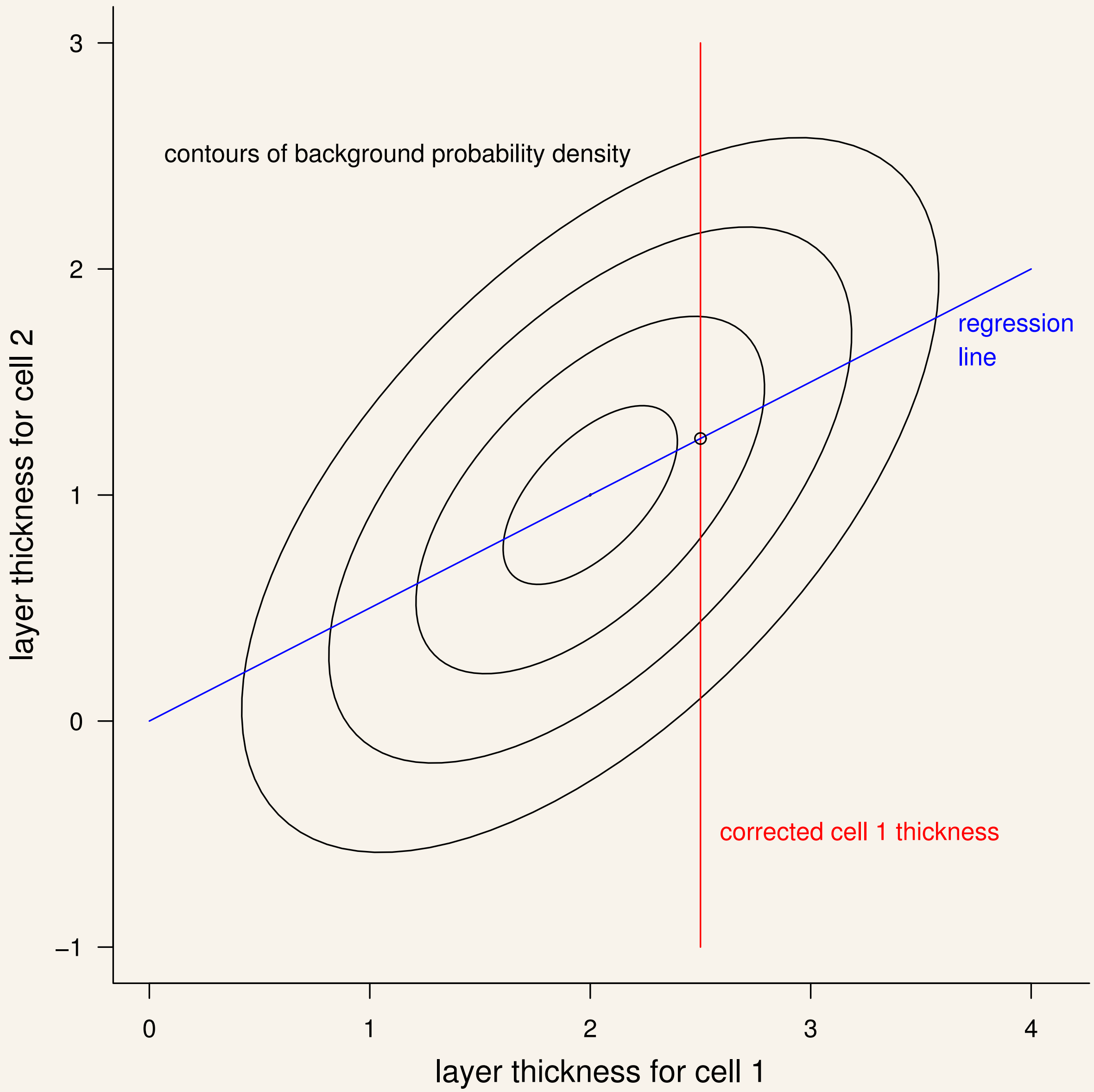
After assimilation x_1 is an information-weighted average of the prior estimate b_1 and the new data d_1:

$$x_1 = \left(\frac{b_1}{\sigma_1^2} + \frac{d_1}{\sigma_d^2} \right) / \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)$$

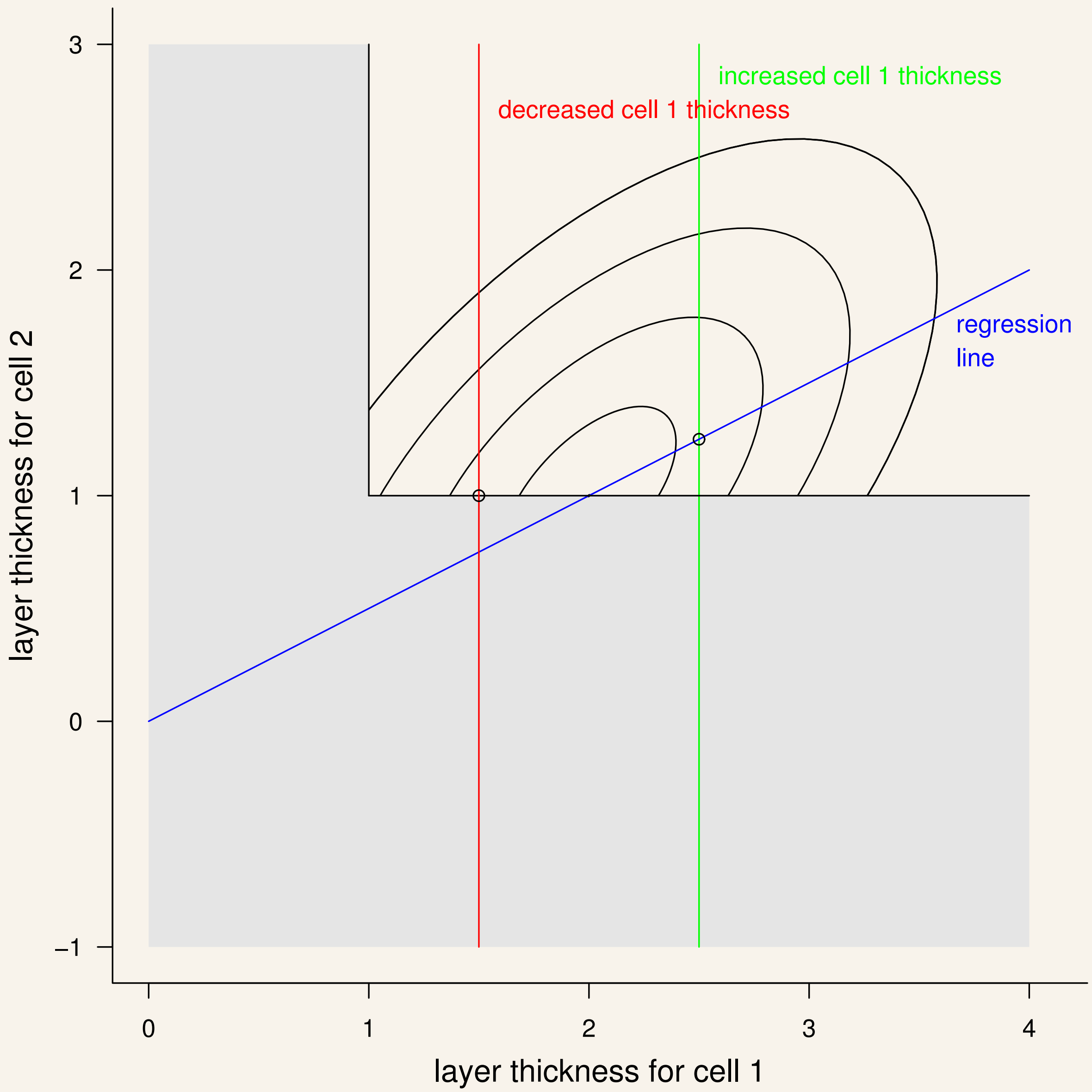
and the correction to x_2 is a regression estimate based on the correction to x_1.

$$x_2 - b_2 = \frac{\sigma_{12}^2}{\sigma_1^2}(x_1 - b_1)$$

Simple example: 2 variables, 1 observed



Constraint: layer thickness $\geq$ 1



Cell 2 initially has minimum layer thickness.

Measurements increase thickness for cell 1. No problem.

What if measurement decreases thickness for cell 1?

Optimum would be at constraint boundary.

Possible prescription

Simply correct by inflating to minimum thickness.

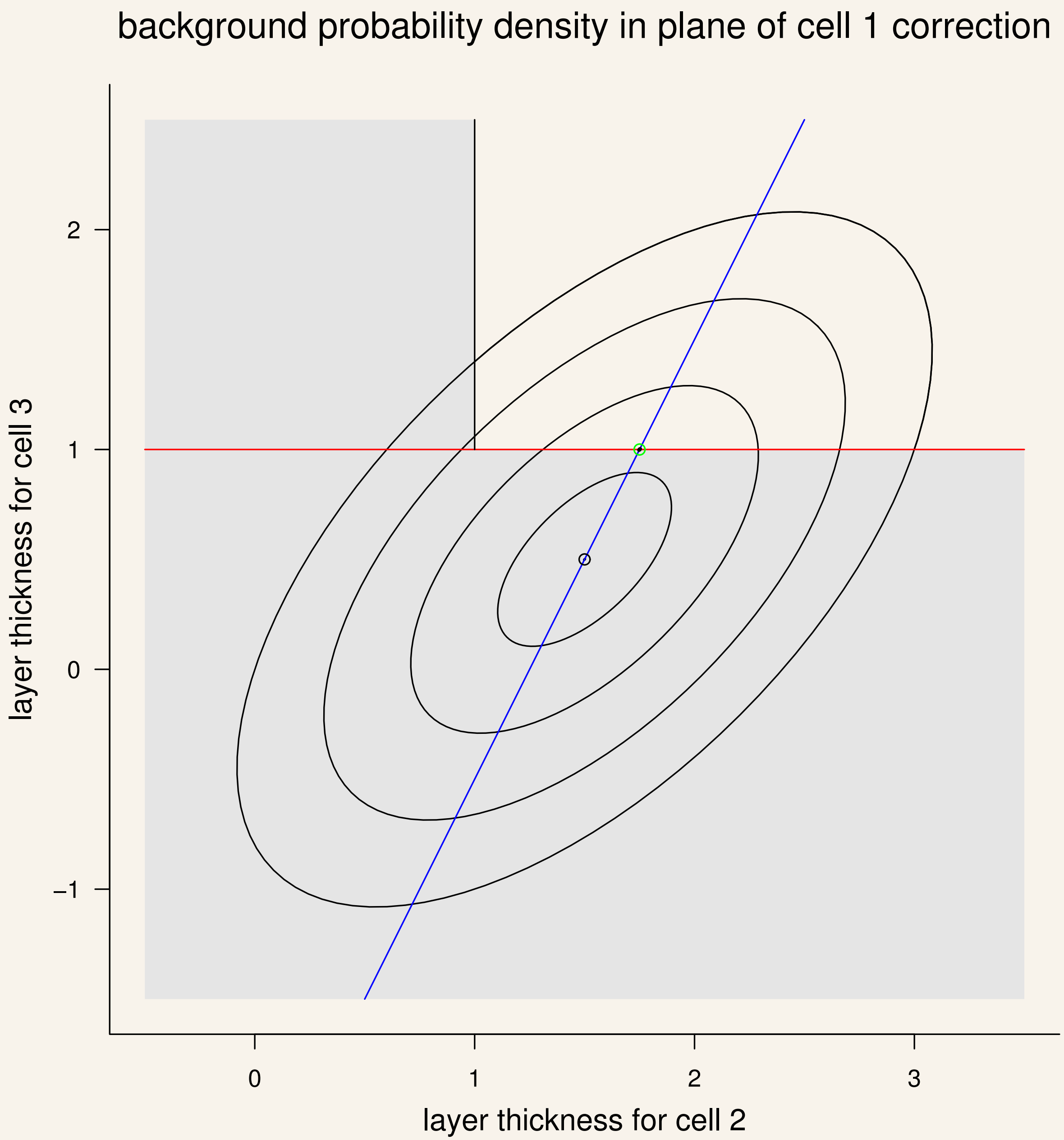
But what if more than two variables?

Should all layers that are too thin be inflated to the minimum thickness?

Should values any of the other variables be adjusted?

Should some constraint-violating thicknesses be inflated beyond the minimum?

Next example: 3 variables, 1 observed



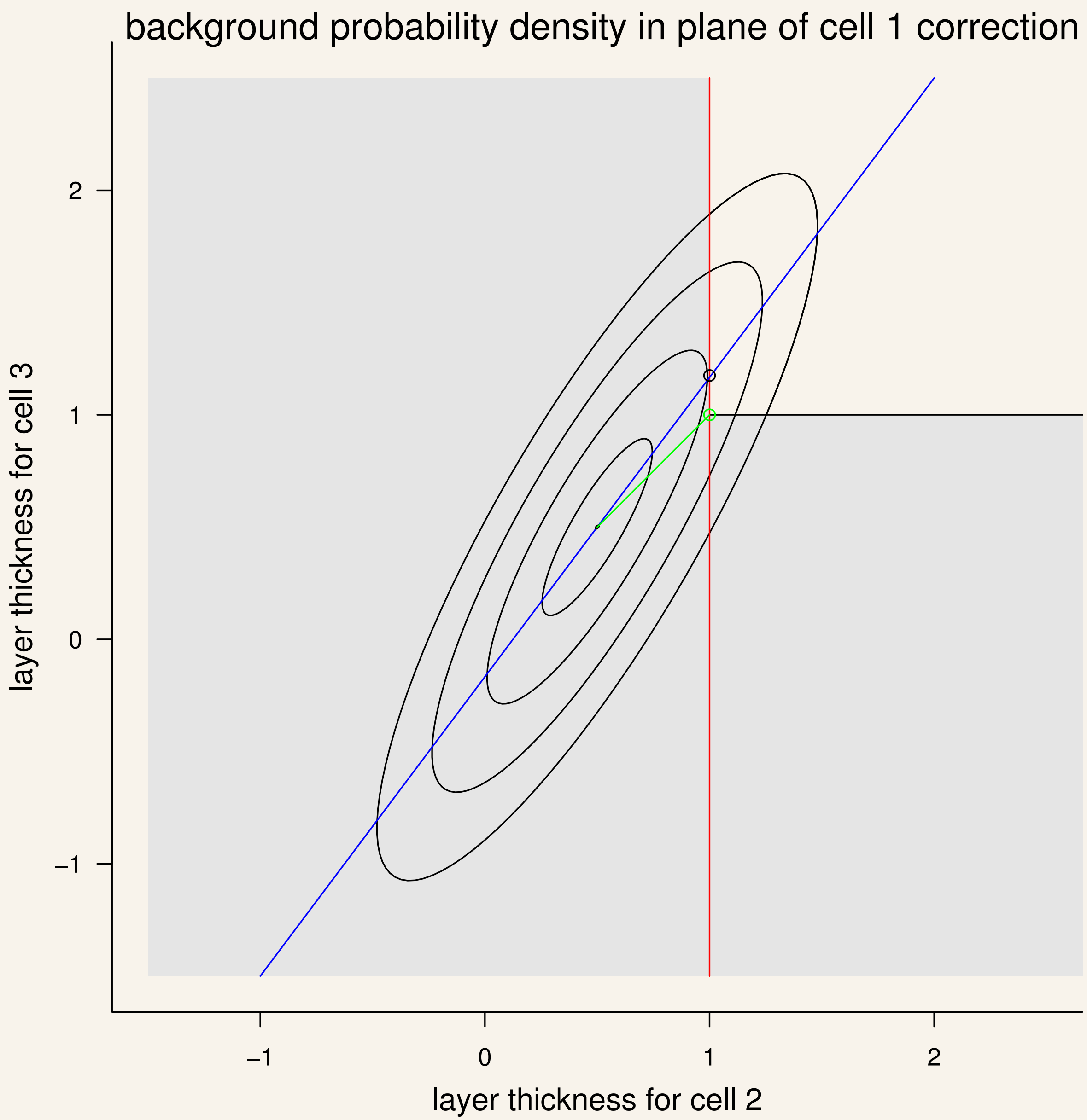
Observe cell 1. Get constraint violation for cell 3

Constraint line provides corrected thickness for cell 3.

Enforcing constraint for cell 3 requires a correction for cell 2

Get another regression problem. Suggests second pass with constraints as data.

Another case: 3 variables, 1 observed



Thicknesses for both cells 2 and 3 are too small after correcting cell 1.

Inflating both to minimal thickness is not optimal.

When cell 2 has minimum thickness, optimum for cell 3 is somewhat larger than minimum.

Two expensive approaches

Quadratic programming considers all possibilities for enforcing range constraints.

Markov Chain Monte Carlo for non-Gaussian integrals.

Any inexpensive approaches?

Error covariances are poorly known, so it doesn't pay to be too picky about finding a precise solution to the resulting math problem.

Maybe enforcing all violated constraints and accounting for the impact on the other variables is good enough.

If we know which constraints to enforce

Inequality constraints that are not violated can be ignored.

Those that are violated can be treated like equality constraints and enforced with Lagrange multipliers.

A practical strategy is to assume that all violations should be enforced.

Variational formalism revisited

Using Lagrange multipliers $\mathbf{y}$, the constrained minimum of J corresponds to a stationary point of:

$$L(\mathbf{x}, \mathbf{y}) = J(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathsf{T}} \mathsf{A}^{-1} (\mathbf{x} - \mathbf{x}_0) - (\mathsf{P}\mathbf{x} - \mathbf{c})^{\mathsf{T}} \mathbf{y}$$

where $\mathbf{x_0}$ is the unconstrained minimum of $J(\mathbf{x})$ and A is the posterior error-covariance matrix, and where the enforced constraints are $\mathsf{P}\mathbf{x} = \mathbf{c}$.

Stationarity requires that $\mathbf{x} - \mathbf{x}_0 = \mathsf{A}\mathsf{P}^{\mathsf{T}}\mathbf{y}$, so the Lagrange multipliers are given by $\mathbf{y} = (\mathsf{P}\mathsf{A}\mathsf{P}^{\mathsf{T}})^{-1}(\mathbf{c} - \mathsf{P}\mathbf{x}_0)$. Thus, the constrained solution is:

$$\mathbf{x} - \mathbf{x}_0 = \mathsf{A}\mathsf{P}^{\mathsf{T}}(\mathsf{P}\mathsf{A}\mathsf{P}^{\mathsf{T}})^{-1}(\mathbf{c} - \mathsf{P}\mathbf{x}_0) \; .$$

Compare: $\mathbf{b} \to \mathbf{x}_0$, $\mathsf{B} \to \mathsf{A}$, $\mathsf{H} \to \mathsf{P}$, $\mathbf{d} \to \mathbf{c}$, and $\mathsf{D} \to 0$.

Algorithm 1

Correct constraint violations using

$$\mathbf{x} - \mathbf{x}_0 = \mathsf{A}\mathsf{P}^\mathsf{T}(\mathsf{P}\mathsf{A}\mathsf{P}^\mathsf{T})^{-1}(\mathbf{c} - \mathsf{P}\mathbf{x}_0) \ .$$

Most difficult part of solving is computing the posterior error covariance matrix

$$\mathsf{A} = \mathsf{B} - \mathsf{B}\mathsf{H}^\mathsf{T}(\mathsf{H}\mathsf{B}\mathsf{H}^\mathsf{T} + \mathsf{D})^{-1}\mathsf{H}\mathsf{B} \ .$$

If D is diagonal, then A can be computed sequentially, one observation at a time.

Approximate algorithm

Following lead of optimal interpolation, where covariances are not updated after assimilation, use B rather than A.

Algorithm 2

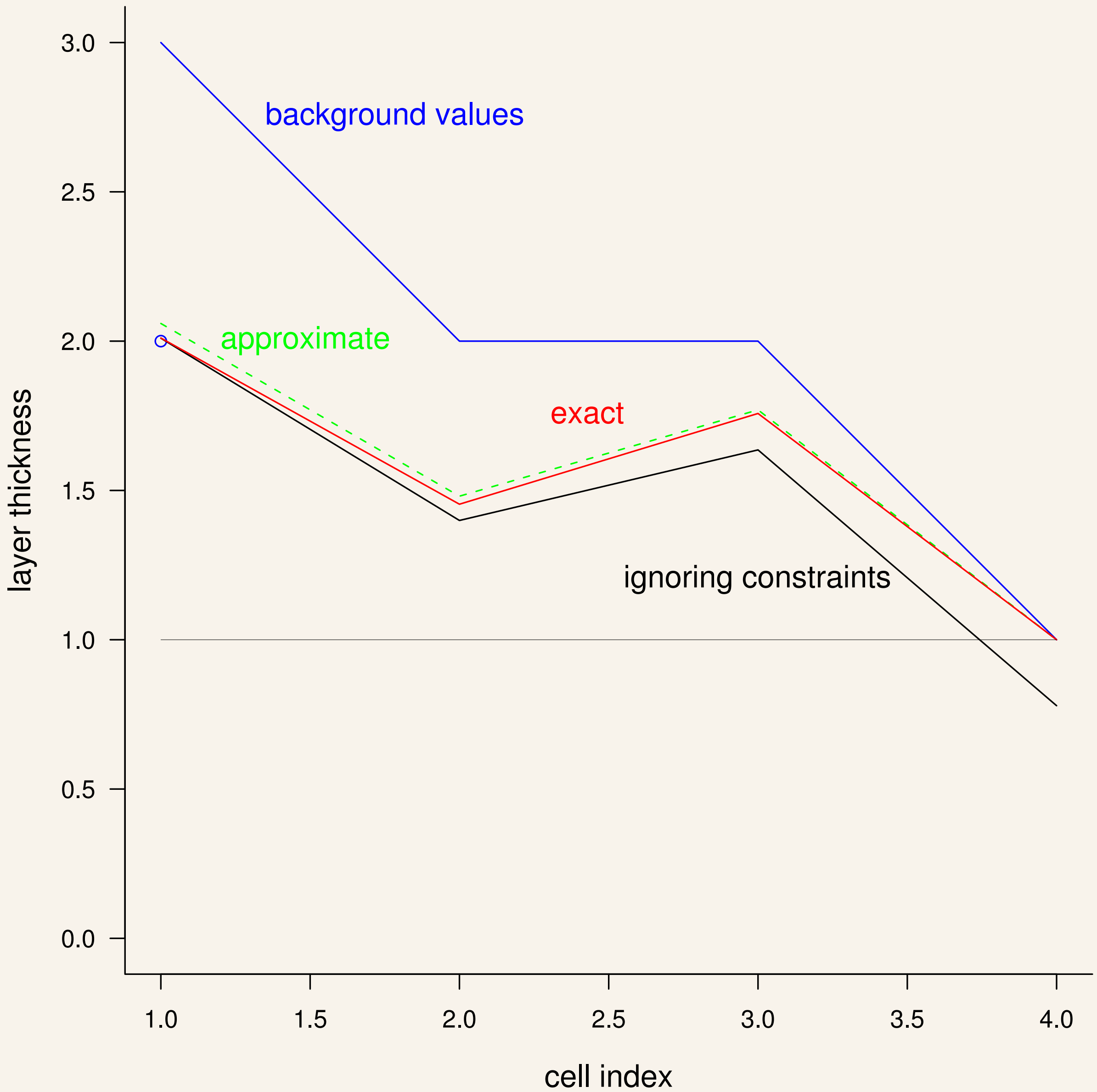
Append constraints $\mathbf{c}$ to $\mathbf{d}$ and treat as error-free data.
Append P to H and expand D with zeros.
Redo assimilation.

Involves solving a larger (rather than smaller) problem on the second pass, but not necessary to know A.

Requires only minimal changes to existing data-assimilation codes.

Gives exactly same results as algorithm 1.

Four-variable computational example



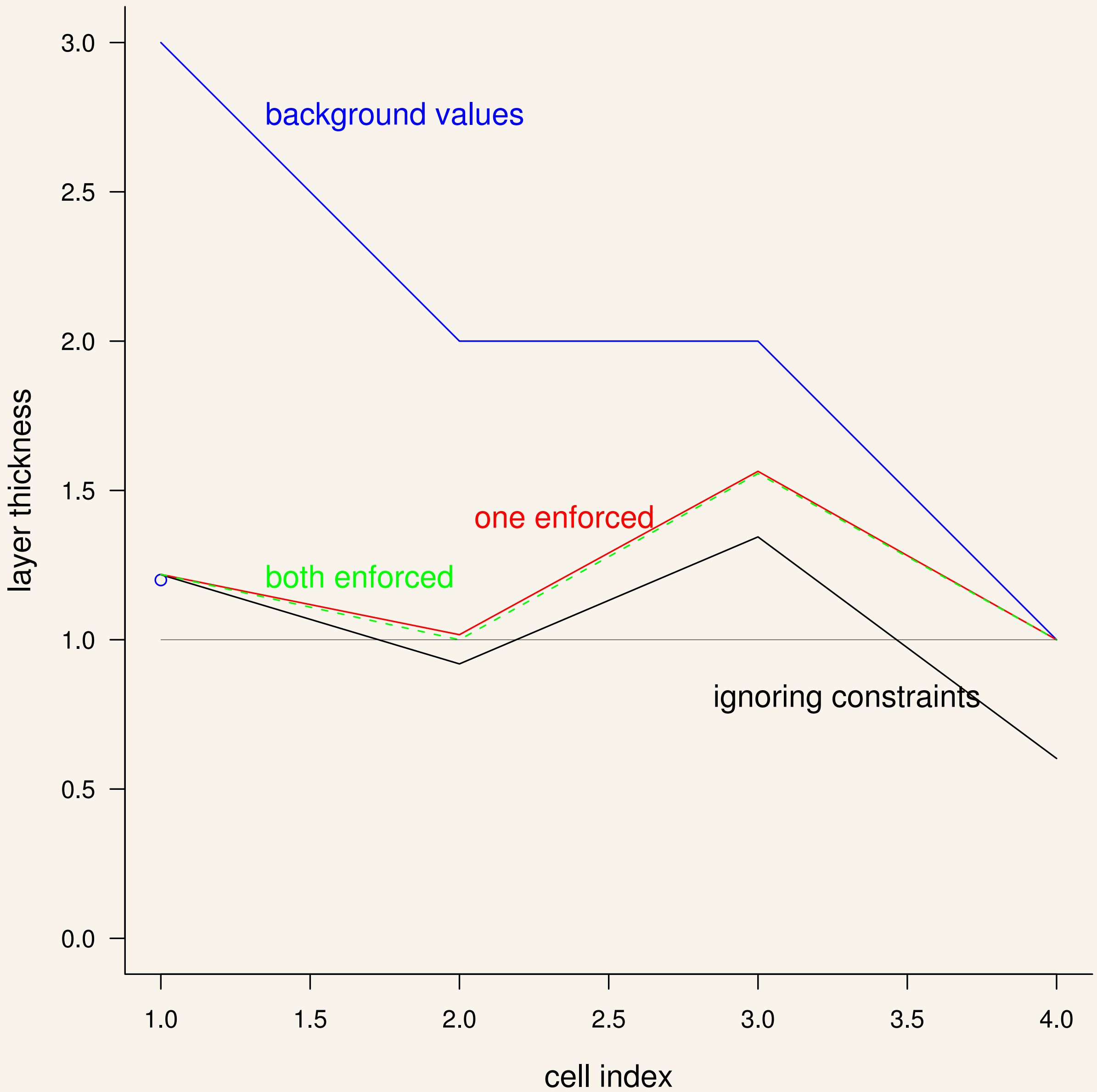
Exponential correlation length of 1-cell width. $\sigma_b^2 = 1$ **and** $\sigma_d^2 = 0.01$.

Assimilating data causes constraint violation three cells away.

Treating constraint as data fixes the problem. Both algorithms give same result.

Not updating B **to** A **before 2nd pass works almost as well.**

Modified example



Same errors as before. Observed thickness is smaller.

Assimilation causes problems for 2 cells.

Correcting only the worst violation fixes the other one.

Correcting both gives about the same answer.

Conclusions

Range constraints can be enforced economically.

Simplest is to enforce constraints without regard for implied corrections.

Implied corrections can be computed by treating enforced constraints like error-free data, which should be assimilated on a second pass.

Few changes are needed for existing assimilation software.

Cost of finding the minimal subset of violations to enforce to get an "optimal" solution, is not justified given lack of precise knowledge about errors.