Range-space Krylov methods for nonlinear problems in data assimilation

Philippe Toint (with Serge Gratton and Selime Gürol)

Department of Mathematics, University of Namur, Belgium

(philippe.toint@unamur.be)

Lisbon, July 2013

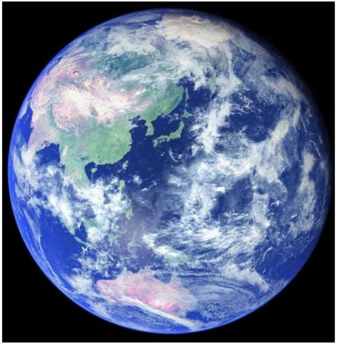
Motivation: data assimilation for weather forecasting

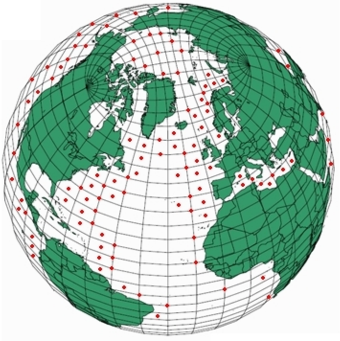
(Attempt to) predict. . .

- tomorrow's weather
- the ocean's average temperature next month
- future gravity field
- future currents in the ionosphere
- . . .

Data assimilation for weather forecasting (2)

Data: temperature, wind, pressure, . . . everywhere and at all times!

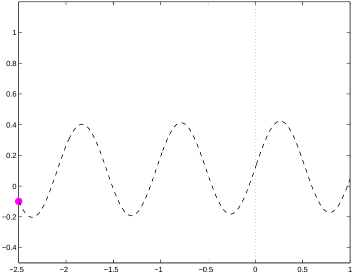




May involve up to 1,000,000,000 variables!

Data assimilation for weather forecasting (3)

The principle:

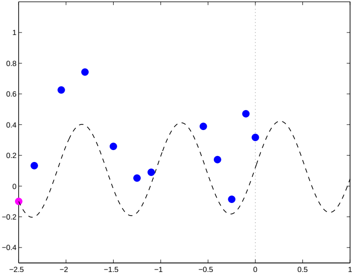


temp. vs. days

- Known situation 2.5 days ago and background prediction

Data assimilation for weather forecasting (3)

The principle:



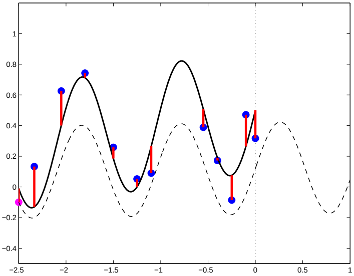
temp. vs. days

- Known situation 2.5 days ago and background prediction
- Record temperature for the past 2.5 days

Data assimilation for weather forecasting (3)

The principle:

Minimize deviation between model and past observations



temp. vs. days

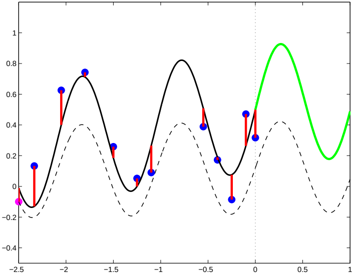
- Known situation 2.5 days ago and background prediction
- Record temperature for the past 2.5 days
- Run the model to minimize difference between model and observations

$$\min_{x_0} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N} \|\mathcal{HM}(t_i, x_0) - b_i\|^2_{R_i^{-1}}.$$

Data assimilation for weather forecasting (3)

The principle:

Minimize deviation between model and past observations



temp. vs. days

- Known situation 2.5 days ago and background prediction
- Record temperature for the past 2.5 days
- Run the model to minimize difference between model and observations
- Predict temperature for the next day

Data assimilation problem: reformulation

Initial formulation (4DVAR):

$$\min_{x_0} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N} \|\mathcal{H}\mathcal{M}(t_i, x_0) - y_i\|^2_{R_i^{-1}}.$$

linearize, concatenate successive times and define $x_0 = x_s + s$:

$$\min_{x_0} J(s) \stackrel{\text{def}}{=} \tfrac{1}{2}(x_s + s - x_b)^T B^{-1}(x_s + s - x_b) + \tfrac{1}{2}(Hs - d)^T R^{-1}(Hs - d)$$

A nonlinear setting

But the 4DVAR formulation

$$\min_{x} \frac{1}{2}\|x - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N}\|\mathcal{H}\mathcal{M}(t_i, x) - b_i\|^2_{R_i^{-1}}.$$

is nonlinear in x!
$\Rightarrow$ linearized model not accurate for long steps
$\Rightarrow$ potential convergence problems
$\Rightarrow$ suggests use of a trust-region algorithm for globalization

Large size
$\Rightarrow$ approximate minimization of the linearized model
$\Rightarrow$ use of Krylov methods in a truncated framework

Trust-region for 4DVAR

1. **Initialization:**
   Set $k = 0$ and calculate the function value at the initial point $f(x_0)$.

2. **Approximate step computation:** Find s_k by solving the subproblem

$$\min_s J_k(s) \quad \text{subject to} \quad \|s\| \leq \Delta_k$$

   (approximately, using truncated conjugate gradients)

3. **Acceptance of the trial point:** Compute the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{J_k(x_k) - J_k(x_k + s_k)},$$

   If $\rho_k \geq \eta_1$, $x_{k+1} = x_k + s_k$; otherwise, $x_{k+1} = x_k$.

4. **Trust-region radius update:** Set

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{if } \rho_k \geq \eta_2, \\ [\gamma_2 \Delta_k, \Delta_k) & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k) & \text{if } \rho_k < \eta_1, \end{cases}$$

   Increment k by 1 and go to step 2.

Range space reformulation (cfr Selime Gürol's talk)

- Use conjugate-gradients (CG) to solve the step computation

$$\min_{x_0} J(s) \stackrel{\rm def}{=} \tfrac{1}{2}(x_s+s-x_b)^T B^{-1}(x_s+s-x_b)+\tfrac{1}{2}(Hs-d)^T R^{-1}(Hs-d)$$

- Reformulate as a constrained problem:

$$\begin{array}{ll} \min_{x_0} J(s) \stackrel{\rm def}{=} \tfrac{1}{2}(x_s+s-x_b)^T B^{-1}(x_s+s-x_b)+\tfrac{1}{2}a^T R^{-1}a \\ \text{such that} \quad a = Hs-d \end{array}$$

- Write KKT conditions of this problem

$$(R+HBH^T)\lambda = d - Hc, \quad s = c + BH^T\lambda, \quad a = -R\lambda$$

- Precondition (1rst level) by R, forget a, assume (wlog) that $c=0$:

$$(I+R^{-1}HBH^T)\lambda = R^{-1}d, \quad s = BH^T\lambda,$$

Solve system using (preconditioned) CG in the H^TBH inner product

$\Rightarrow$ RPCG (Gratton, Tshimanga, 2009)

RPCG and preconditoning (cfr Selime Gürol's talk)

Features of RPCG:

- algorithm form comparable to PCG (additional products)
- only uses vector of size = number of observations
  $\Rightarrow$ suitable for reorthogonalization
- sequence of iterates identical to that generated by PCG on primal
  $\Rightarrow$ good descent properties on $J(s)$!
- numerically stable for range-space perturbations
- any (2nd level) preconditioner F for the primal PCG translates to a preconditioner G for the range-space formulation iff

$$FH^T = BH^T G$$

Note: works for limited memory preconditioners !

Numerical performance on ocean data assimilation ?

1. NEMOVAR: Global 3D-Var system
   $\longrightarrow$ seasonal forecasting $\longrightarrow$ ocean reanalysis
2. ROMS California Current Ocean Modeling: Regional 4D-Var system
   $\longrightarrow$ coastal applications

Implementation details:

- $m \approx 10^5$ and $n \approx 10^6$
- The model variables: temperature, height, salinity and velocity.
- 1 outer loop of Gauss-Newton ($k = 1$), 40 inner loops

Gürol, Weaver, Moore, Piacentini, Arango, Gratton, 2013. *Quarterly Journal of the Royal Meteorological Society*. In press.

- good numerical performance
- reorthogonalization sometimes necessary

In the nonlinear setting

When solving the general 4DVAR problem. . .

Several (outer) Gauss-Newton iterations !

Questions:

1. Can one define the trust-region in the range-space ?

$$s = BH^T\lambda$$

   Ok, provided B an H^T are reasonably behaved!

2. Can one reuse the range-space preconditioner from previous iteration?

$$??? \quad F_{k-1}H_k^T = BH_k^T G_{k-1} \quad ???$$

   In general: No!
   (the old preconditioner is no longer symmetric in the new metric!)

Some fixes

The old preconditioner is no longer symmetric in the new metric!

How to get around this problem:

1. avoid (2nd level) preconditioning ???
   (last resort decision)
2. recompute the preconditioner in the mew metric ??
   (possible with limited memory preconditioners, but costly)
3. ignore the problem and take one's chances ?
   (only reasonable if convergence can still be guaranteed)

Taking measured risks. . .

A simple proposal for computing the step:

1. compute the Cauchy step (1rst step of RPCG)
2. if negative curvature, recompute a complete step without 2nd level preconditioner
3. otherwise, use equivalence with primal to check simple decrease of f at the Cauchy point
4. if no decrease, then unsuccessful TR outer iteration
5. otherwise, continue RPCG with old preconditioner
6. if unsuccessful, go back to Cauchy point

Can be interpreted as a TR “magical step”

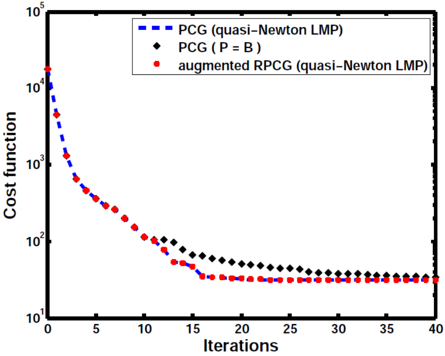
S. Gratton, S. Gürol and Ph. L. Toint, 2013. Preconditioning and globalizing conjugate-gradients in dual space for quadratically penalized nonlinear-least squares problems. *Computational Optimization and Applications* 54: 1-25

S. Gürol, PhD thesis, 2013

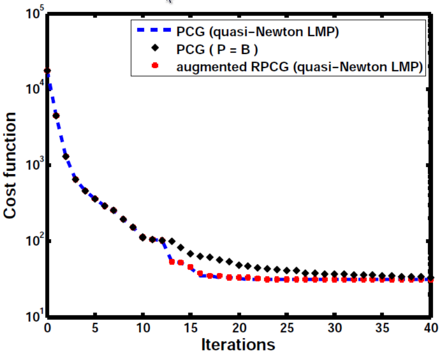
Does it work? (1)

Numerical experiment with a nonlinear heat equation

$f(x) = exp[1x]$

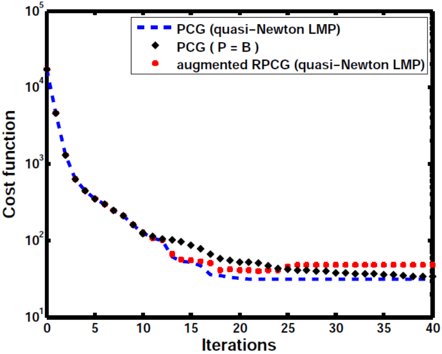


$f(x) = exp[2x]$

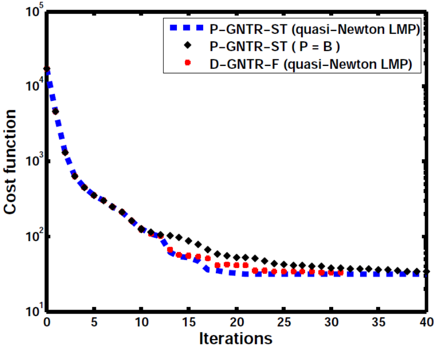


Does it work? (1)

$f(x) = exp[3x]$



$f(x) = exp[3x]$



⟶ The global convergence is ensured by using the **"risky" trust-region algorithm.**
⟶ This algorithm **requires an additional function evaluation** at the Cauchy point for each outer loop.

Conclusions

Range-space approach efficient in some data assimilation problems

Suitable 2nd level preconditioners can be built

Potential symmetry problem solved without compromising convergence

Already in use in real operational systems

Thank you for your attention!