MATHEMATICAL PROBLEMS ASSOCIATED WITH ATMOSPHERIC DATA ASSIMILATION AND WEATHER PREDICTION

Pierre Gauthier
Department of Earth and Atmospheric Sciences
Université du Québec à Montréal

<u>Work done with</u>
Amal El Akkraoui (McGill DAOS), Simon Pellerin (Environment Canada) and Samuel Buis (CERFACS, France)

Presentation on February 4, 2008
Mathematical Advancement in Geophysical Data Assimilation
Banff Institute Research Station, Banff, Canada

Outline

- **Introduction of the variational data assimilation problem**
- **Lorenc (1986), Courtier (1997)**
  - The dual form of 3D-Var
    - → *Physical Space Statistical Analysis System* (**PSAS**): Cohn *et al.* (1998), GMAO, NASA;
    - → Daley and Barker (2001) (**NAVDAS** at NRL)
  - Equivalence between the primal (3D-Var) and dual (3D-PSAS) formulation of the statistical estimation problem
- **Modular representation of both algorithms**
  - Implementation based on the same operators of the variational assimilation system
- **Convergence of the minimization and preconditioning**
  - Significance of the two cost functions and impact on the results
- **Applications of the dual approach**
  - Sensitivity of observations with respect to observations (Langland, 2004)
  - Information content (Cardinali)
  - The weak-constraint 4D-Var and its dual formulation

The variational problem

Bayes' Theorem: $$p(\mathbf{x}\,|\,\mathbf{y}) = \frac{p(\mathbf{y}\,|\,\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}$$

- **Example:**
  - Observation and background error have Gaussian distributions

$$p(\mathbf{y}\,|\,\mathbf{x}) = \frac{1}{C_3}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}(\mathbf{x}))\right\}$$

$$P(\mathbf{x}) = \frac{1}{C}\exp\left\{-\tfrac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)\right\}$$

  - $p(\mathbf{y}|\mathbf{x})$ is Gaussian only if **H** is linear
  - Maximum likelihood estimate (mode of the distribution):

$$J(\mathbf{x}) = -\ln p(\mathbf{x}\,|\,\mathbf{y})$$
$$= \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \tfrac{1}{2}(\mathbf{H}(\mathbf{x})-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}(\mathbf{x})-\mathbf{y})$$

- Reducing $J(\mathbf{x})$ implies an increase in the probability of **x** being the true value

Incremental approach

Successive linearizations with respect to full model state obtained

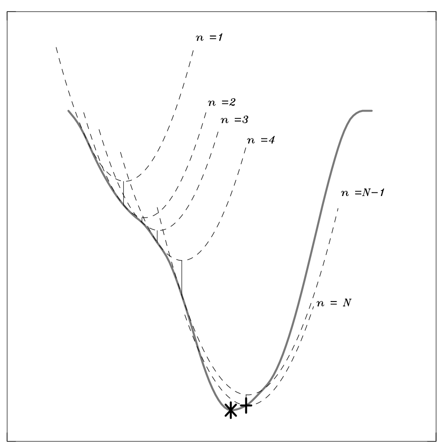
- Minimization of quadratic problems

$$J(\xi) = \tfrac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \tfrac{1}{2}(\mathbf{H}'\delta\mathbf{x}-\mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}'\delta\mathbf{x}-\mathbf{y}')$$
$$\equiv \tfrac{1}{2}\xi^T\xi + \tfrac{1}{2}(\mathbf{H}'\mathbf{G}\xi-\mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}'\mathbf{G}\xi-\mathbf{y}')$$

where

$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$	:	increment
$\mathbf{H}' = \partial\mathbf{H}/\partial\mathbf{x}$	:	tangent-linear of the observation operator

$\mathbf{y}' = \mathbf{y} - \mathbf{H}(\mathbf{x}_b)$:
innovation vector (observation departure with respect to the high resolution background state)
$\mathbf{G} = \mathbf{B}^{1/2}$



From Laroche and Gauthier (1998)

The dual formulation

- Explicit solution when H' is *linear*

$$\delta \mathbf{x}_a = \mathbf{BH}'^T (\mathbf{R} + \mathbf{H}'\mathbf{B}\mathbf{H}'^T)^{-1} \mathbf{y}'$$

- Solving a large matrix problem

$$\mathbf{w}^* = (\mathbf{R} + \mathbf{H}'\mathbf{BH}'^T)^{-1} \mathbf{y}'$$

  - **The PSAS minimization problem**

$$J_P(\mathbf{w}) = \tfrac{1}{2} \mathbf{w}^\mathbf{T} (\mathbf{R} + \mathbf{H}'\mathbf{BH}'^\mathbf{T}) \mathbf{w} - \mathbf{y}'^T \mathbf{w}$$

$$\nabla J_P(\mathbf{w}) = (\mathbf{H}'\mathbf{BH}'^T + \mathbf{R}) \mathbf{w} - \mathbf{y}'$$

$$\delta \mathbf{x}_a = \mathbf{BH}'^\mathbf{T} \mathbf{w}^*$$

- 3D-Var and PSAS should then give the same solution *at convergence*
- Control variable: **x** is in model space for 3D-Var
  w is in observation space for PSAS

3D-Var: variational formulation of the statistical estimation problem

Minimization of the cost function

$$J(\xi) = \frac{1}{2} \xi^T \xi + \frac{1}{2} (\mathbf{H}'\mathbf{G}\xi - \mathbf{y}')^T \mathbf{R}^{-1} (\mathbf{H}'\mathbf{G}\xi - \mathbf{y}')$$

$$\delta x = \mathbf{B}^{1/2} \xi \equiv \mathbf{G}\, \xi$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$: increment
$\mathbf{H}' = \partial\mathbf{H}/\partial\mathbf{x}$: tangent-linear of the observation operator
$\mathbf{y}' = \mathbf{y} - \mathbf{H}(\mathbf{x}_b)$: innovation vector (observation departure) (computed with respect to the high resolution background state)

Preconditioning the minimization

- Condition of the minimization relates to the ratio of the highest to the lowest eigenvalues of the Hessian matrices
- Courtier (1997): preconditioning with respect to $\mathbf{B}^{-1/2}$ for 3D-Var and with respect to $\mathbf{R}^{1/2}$ for PSAS

3D-Var	PSAS
$\xi = \mathbf{B}^{-1/2}\delta\mathbf{x}$	$\mathbf{u} = \mathbf{R}^{1/2}\mathbf{w}$
$J'' = \mathbf{I} + \mathbf{B}^{1/2^T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$	$J''_P = \mathbf{I} + \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1/2^T}$

- The two problems have *exactly* the same condition number (Courtier, 1997) and should therefore have similar convergence rates.
- The argument extends to 4D-Var if **H'** includes the integration of the tangent linear model.

Operations involved in both algorithms

Direct operators

Input	Operator	Output
ξ	$\mathbf{G} = \mathbf{B}^{1/2}$	$\delta\mathbf{x}_0$
$\delta\mathbf{x}_k$	$\boldsymbol{L}(t_k, t_{k+L})$	$\delta\mathbf{x}_{k+L}$
$\delta\mathbf{x}_k$	$\mathbf{H}_k'$	$\mathbf{w}_k$
$\mathbf{w}_k$	$\mathbf{R}^{-1/2}$	$\mathbf{u}_k$

Adjoint operators

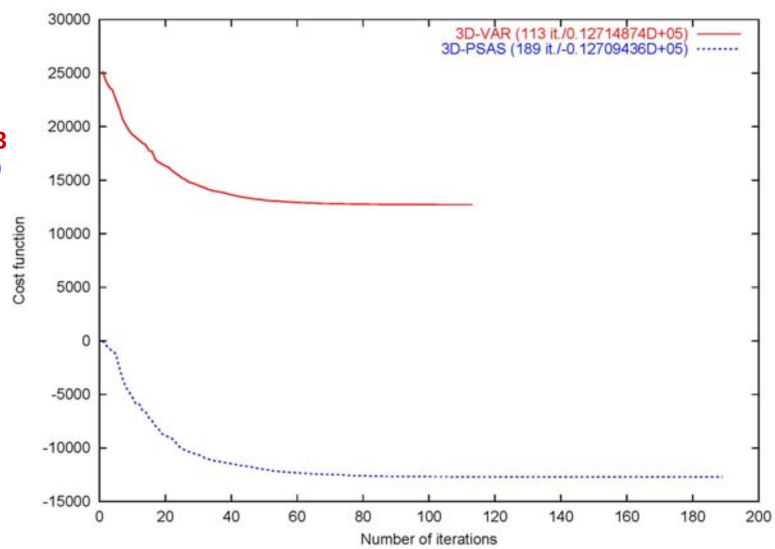
Input	Operator	Output
$\delta\mathbf{x}_0^*$	$\mathbf{G}^* = \mathbf{B}^{1/2^T}$	ξ^*
$\delta\mathbf{x}^*_{k+L}$	$\mathbf{L}^*(t_k, t_{k+L})$	$\delta\mathbf{x}^*_k$
$\mathbf{w}^*_k$	$\mathbf{H}'^*_k$	$\delta\mathbf{x}^*_k$
$\mathbf{u}_k^*$	$\mathbf{R}^{-1/2^T}$	$\mathbf{w}^*_k$

Modularisation of MSC's variational system

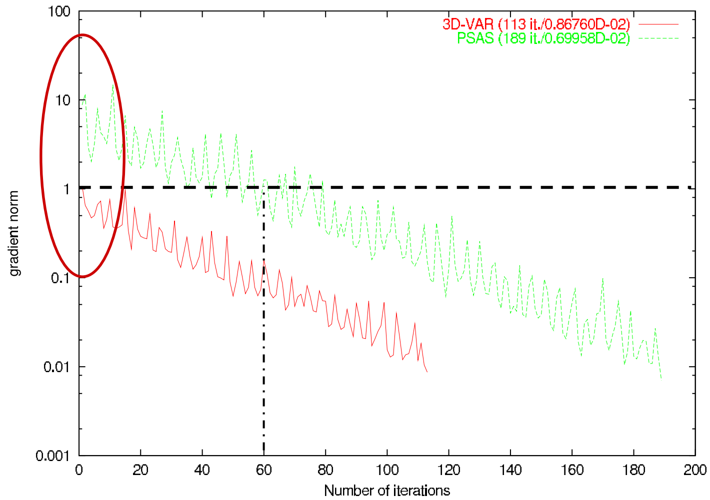
- **Decomposition of the 3D/4D-Var into its basic units**
  - Including the TLM and adjoint operators
  - 3D-Var and PSAS could be built and compared within a quasi-operational framework
    - → Extension to 4D-PSAS has been done also (without the outer iterations)

Cost function of PSAS (blue) and 3D-Var (red)

Number of iterations
3D-Var: 113
PSAS: 189



Norm of gradient as a function of iteration

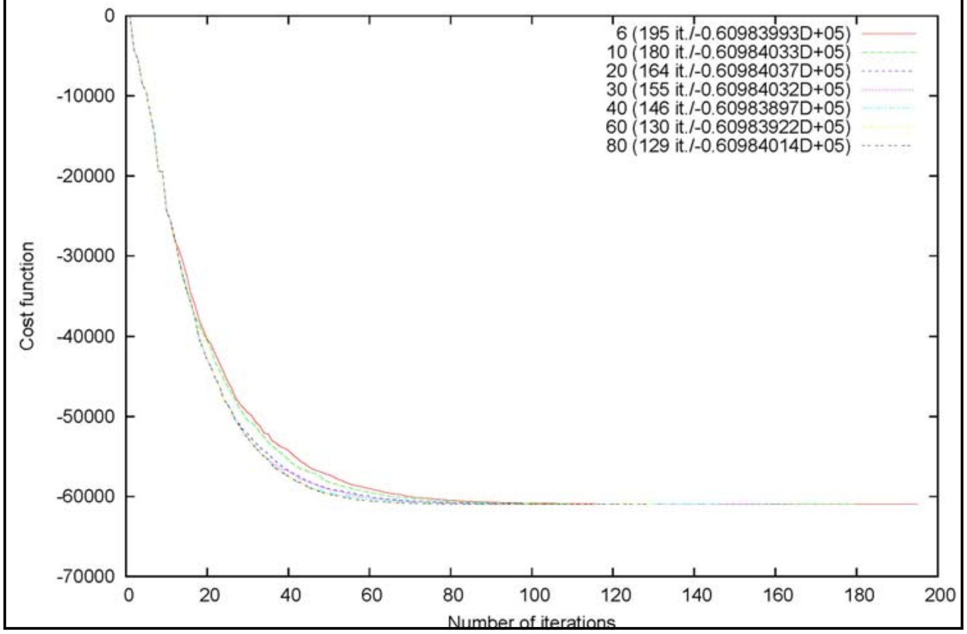


Impact of the dimension of the Hessian used in the BFGS approximation

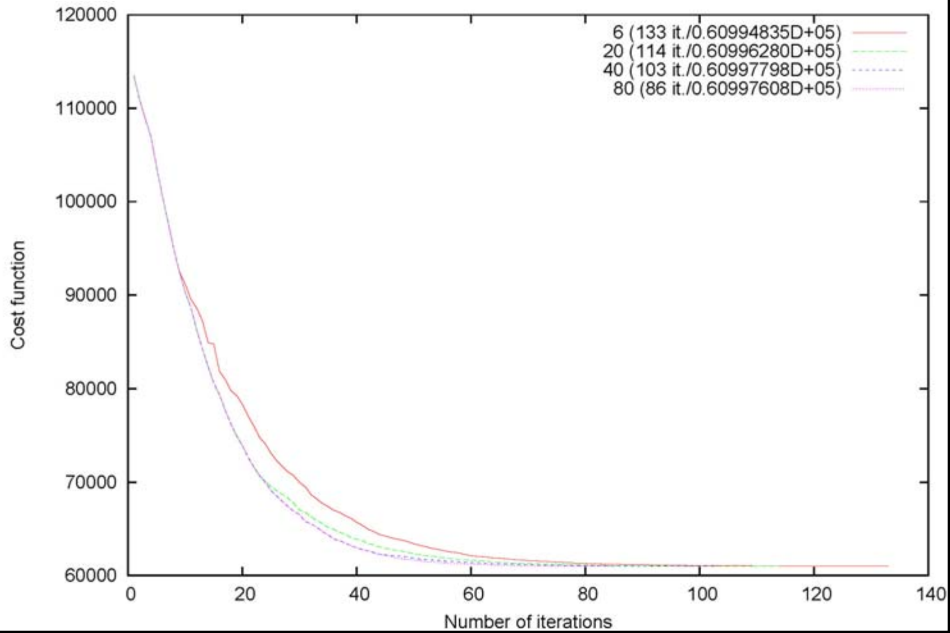
- **Quasi-Newton builds an approximation of the Hessian in the course of the minimization**
  - Limited-memory imposes a limit on the dimension of the approximate of the Hessian
  - Tests in M1QN3 (INRIA) with 6, 10, …, up to 80 pairs of

$$\left(\mathbf{x}_k - \mathbf{x}_{k-1}\right) \; and \; \left(\nabla J\left(\mathbf{x}_k\right) - \nabla J(\mathbf{x}_{k-1}\right)$$

Cost function of PSAS vs. number of iterations



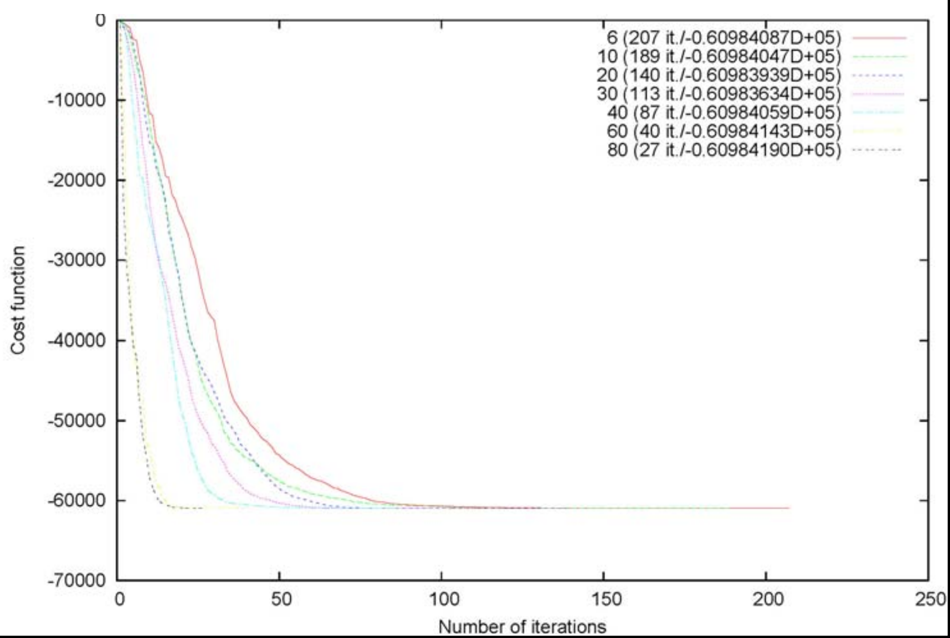
Cost function of 3D-Var vs. number of iterations



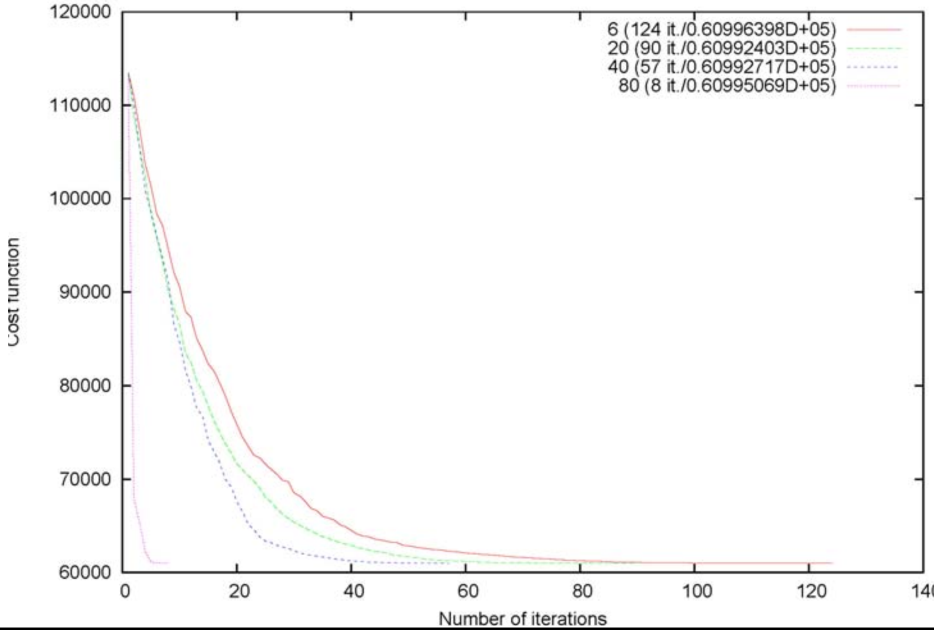
Cycling the Hessian

- **Experiment to test the quality of the approximate Hessian obtained from a limited memory quasi-Newton (M1QN3)**
  - o Use the Hessian of the previous experiment to precondition the *same* problem
  - o Impact of augmenting the dimension used for its representation

Preconditioned PSAS with its own Hessian



Preconditioned 3D-Var with its own Hessian



Estimation of the Hessian

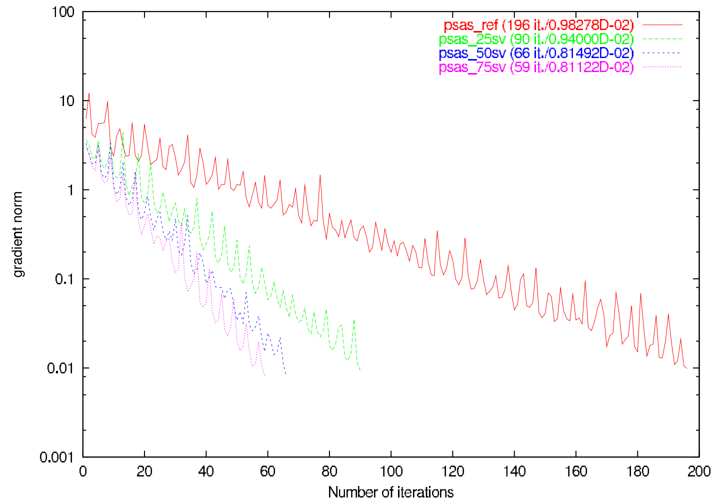
- **Hessian matrices**
  - Let $\mathbf{L} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$
  - 3D-Var: $$J'' = \mathbf{I} + \left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}\right)^T\left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}\right)$$ $$= \mathbf{I}_n + \mathbf{L}^T\mathbf{L}$$
  - PSAS: $$J''_P = \mathbf{I} + \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1/2T} = \mathbf{I}_m + \mathbf{L}\mathbf{L}^T$$
- **Correspondence of eigenvectors**

3D-Var $J''_{3D}\mathbf{v} = \left(\mathbf{I}_n + \mathbf{L}^T\mathbf{L}\right)\mathbf{v} = \lambda\mathbf{v}$

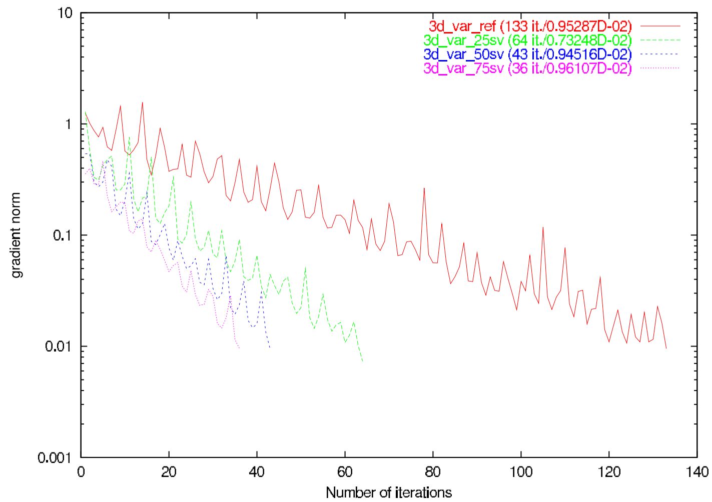
PSAS $J''_P\mathbf{u} = \left(\mathbf{I}_m + \mathbf{L}\mathbf{L}^T\right)\mathbf{u} = \lambda\mathbf{u}$

$$\mathbf{u} = \frac{1}{\sqrt{\lambda - 1}}\mathbf{L}\mathbf{v} \qquad \mathbf{v} = \frac{1}{\sqrt{\lambda - 1}}\mathbf{L}^T\mathbf{u}$$

Preconditioning the PSAS with 25, 50 and 75 singular vectors



Preconditioning the 3D-Var with 25, 50 and 75 singular vectors



Cycling the Hessian

- **3D-Var control variable is the model state and stays the same from one analysis to the next**
  - Approximation of the Hessian from one analysis can be easily used for the next one
- **PSAS control variable depends on the observations used and their distribution (in time and space)**
- **Cycling the Hessian for PSAS uses its mapping in model space**

$$\mathbf{u}_k(t_0) \quad \rightarrow \quad \mathbf{v}_k(t_0) = \frac{1}{\sqrt{\lambda_k - 1}} \mathbf{L}_0^T \mathbf{u}_k(t_0)$$

$$\mathbf{v}_k(t_0) = \frac{1}{\sqrt{\lambda_k - 1}} \mathbf{B}^{1/2^T} \mathbf{H}_0^T \mathbf{R}_0^{-1/2^T} \mathbf{u}_k(t_0)$$

Cycling the Hessian

- **3D-Var control variable is the model state and stays the same from one analysis to the next**
  - Approximation of the Hessian from one analysis can be easily used for the next one
- **PSAS control variable depends on the observations used and their distribution (in time and space)**
- **Cycling the Hessian for PSAS uses its mapping in model space**
  - Remapped according to the distribution of the new set of variables

$$\mathbf{u}_k(t_0 + T) = \frac{1}{\sqrt{\lambda_k - 1}} \mathbf{L}_1 \mathbf{v}_k(t_0)$$

$$= \frac{1}{\sqrt{\lambda_k - 1}} \mathbf{R}_1^{-1/2} \mathbf{H}_1 \mathbf{B}^{1/2} \mathbf{v}_k(t_0)$$

Conclusions

- **Equivalence of 3D/4D-PSAS and 3D/4D-Var**
  - Equivalence of the result at convergence *when the observation operator is linear and the pdfs are Gaussian*
    - → Incremental form extends the validity of this statement to nonlinear observation operators and non-Gaussian p.d.f.
  - Both approaches
    - → have similar convergence rates (Courtier, 1997)
    - → can be built by using the same basic operators common to both algorithms (the PALM approach)
  - PSAS cost function has no immediate significance
    - → Unrealistic values of the a posteriori p.d.f. at the beginning of the minimization
    - → Problematic for applications where the number of iterates is set to a fixed number

Summary (cont'd)

- **Preconditioning of the minimization**
  - o Both algorithms benefits similarly from preconditioning
  - o Cycling is more easily done in 3D/4D-Var as the dimension and topology of the control variable remains the same
    - → Control variable of PSAS changes with the observations used
  - o Equivalence between the eigenvectors of their respective Hessian matrices
    - → Permits cycling of the eigenvectors from the previous analysis to the next
    - → Preconditioning of PSAS is effective but requires solving for the eigenvectors of the Hessian (Lanczos algorithm)
      - – Minimization with a conjugate gradient enables to obtain the singular vectors from the iterates of the minimization

Conclusions

- **Application of the PSAS algorithm is useful for**
  - o the estimation of sensitivity with respect to observations
    - → Relationship between the eigenvectors of both Hessians
    - → Mapping forecast sensitivities into observation space
  - o Weak-constraint 4D-Var
    - → Dimension of the control variable is proportional to the number of observations used
  - o See El Akkraoui *et al.* (QJRMS, 2008) for details