Variable-storage quasi-Newton operators for modeling error covariances

Dr Fabrice Veersé
Projet IDOPT (INRIA/CNRS/UJF/INPG)
LMC-IMAG, Tour IRMA, BP 53
38041 Grenoble cedex 9, FRANCE
E-mail : Fabrice.Veerse@imag.fr
WWW : http://www-lmc.imag.fr/lmc-edp/Fabrice.Veerse/

1 Introduction

A key issue in variational data assimilation (VDA) is the specification of the inverse background (forecast) error covariance matrix appearing in the cost function to be minimized.

Veersé (1999) uses the relation between the Hessian at the minimum and the inverse analysis error covariance matrix to approximate the latter by limited-memory quasi-Newton operators. This enables to cycle variational assimilations using dynamical error covariances, but does not provide any quantitative information on the analysis and forecast errors. The paper focuses on one such quasi-Newton operator, having a Cholesky-factorized form that enables the usual preconditioning of the minimization based on the background term.

Owing to this factorization the transport of the error covariances by the tangent linear model may be performed exactly under the perfect-model assumption, by imbeeding the corresponding tangent-linear-model integration within the inverse change of control variable used to precondition the minimization.

The specification of the inverse background (forecast) error covariance matrix may therefore be done efficiently using limited-memory operators, with a control on both memory and CPU-time requirements.

The paper is organized as follows: using relation between the Hessian of the VDA cost function at the minimum and the analysis error covariance matrix (Thacker, 1989; Rabier and Courtier, 1992; Gunson and Malanotte-Rizzoli, 1996) the next section shows how a Cholesky-factorized limited-memory analysis covariance operator can be built; the following section uses this factorization to form a limited-memory forecast error covariance operator that is suitable for cycling VDAs using the traditional background-term preconditioning; the paper concludes with some perspectives for the future. The standard notations of Ide *et al.* (1997) are used throughout the paper.

Table 1: Memory requirements

a real array of size $2m$	for	Λ
a real array of size $2mN$	for	$\mathbf{V}$ or the $\mathbf{p}_k$s
a real array of size N	for	Δ
a real array of size $2mN$	for	the β_ks
a real array of size N	as	work array

2 Limited-memory analysis error covariance operator

Following Thacker (1989), the analysis error covariance matrix can be approximated by the limited-memory BFGS inverse Hessian approximation provided by the minimization code. This approximation has the form

$$\mathbf{W} = \mathbf{D} + \mathbf{V}\Lambda\mathbf{V}^T \tag{1}$$

where $\mathbf{D}$ is a $N \times N$ positive definite diagonal matrix , Λ is a $2m \times 2m$ diagonal matrix, and $\mathbf{V}$ is an $N \times 2m$ matrix; N and m being respectively the dimension of the control vector and the number of $(\mathbf{s}, \mathbf{y})$ pairs used by the limited-memory minimization code (see Nocedal, 1980; Gilbert and Lemaréchal, 1989; Liu and Nocedal, 1989).

In practice, the minimization code does not provide the matrices but a routine that performs the produt $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$; and the elements of the diagonal matrix $\mathbf{D}$ may be retrieved from the minimization-code work space. Therefore one can write the code for the operator $\mathbf{x} \mapsto (\mathbf{W}-\mathbf{D})\mathbf{x}$ and use a Lanczos-type algorithm to compute $\mathbf{V}$ and Λ.

An operator-oriented Cholesky-factorization update technique (method C1 or C2 of Gill *et al.*, 1974) can then be applied to $\mathbf{D}$ in order to transform (1) into

$$\mathbf{W} = \mathbf{L}_1 \cdots \mathbf{L}_{2m} \Delta \mathbf{L}_{2m}^T \cdots {\mathbf{L}_1}^T$$

where Δ is a $N \times N$ positive definite diagonal matrix and the $\mathbf{L}_k$s are special lower triangular matrices of the form

$$\begin{cases} l_k(i,j) = 1 & \text{if } i = j \\ l_k(i,j) = 0 & \text{if } j > i \\ l_k(i,j) = p_i^k \beta_j^k & \text{if } j < i \end{cases}$$

Each of these triangular matrices is thus given by 2 vectors, $\mathbf{p}^k$ and β^k; and its multiplication by a vector and performing the corresponding forward-substitution operation require only $O(N)$ floating-point operations (flops).

The memory and computational requirements to factorize the operator $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$ are given in tables 1 and 2 respectively.

Table 2: Computational requirements

$O(m^2\ N)$ flops for	computing $\mathbf{V}$ and Λ with a Lanczos method
$O(m^2\ N)$ flops for	computing the Cholesky factorization (*i.e.* the $\mathbf{p}_k$s, the β_ks and Δ)

3 Limited-memory forecast error covariance operator

Under the perfect-model assumption, the forecast error covariance matrix is taken as

$$\begin{aligned}
\mathbf{P}^f(t_f) &= \mathbf{M}(t_f,t_i)\ \mathbf{P}^a(t_i)\ \mathbf{M}^T(t_f,t_i) \\
&\approx \mathbf{M}(t_f,t_i)\ \mathbf{H}\ \mathbf{M}^T(t_f,t_i) \\
&\approx \mathbf{M}(t_f,t_i)\mathbf{L}_1\cdots\mathbf{L}_{2m}\Delta\mathbf{L}_{2m}^T\cdots{\mathbf{L}_1}^T\mathbf{M}^T(t_f,t_i) \\
&\approx \mathbf{U}\ \mathbf{U}^T
\end{aligned} \tag{2}$$

where $\mathbf{M}(t_f,t_i)$ is the resolvant of the tangent-linear model over the assimilation period $]t_i,t_f]$.

The minimization can then be performed with this usual preconditioning, as in practice the minimization is started from the background field $\mathbf{x}=\mathbf{x}^b$ (standard formulation) or $\delta\mathbf{x}=0$ (incremental formulation) and only the inverse change of control variable and its adjoint are needed. More explicitely, one needs a routine corresponding to the operator $\mathbf{x}\mapsto\mathbf{U}\mathbf{x}$ and its adjoint, to be used at each iteration of the minimization. Each such multiplication requires the integration of the tangent linear model over the (previous) assimilation period, plus $2m$ multiplications by special lower-triangular matrices and a diagonal matrix. The latter multiplications can be done in $\mathrm{O}(mN)$ flops.

4 Discussion

A method has been proposed for cycling the error covariances in strong 3D/4D VDAs, using limited-memory quasi-Newton operators. The requirements of the method may be controlled through the number m of $(\mathbf{s},\mathbf{y})$ pairs used during the minimization. It requires $O(m^2N)$ flops once per assimilation period, to Cholesky-factorize the limited-memory inverse Hessian operator, and $O(mN)$ flops at each iteration of the minimization (the latter being the complexity of the inverse-Hessian operator provided by the minimization code). But the main additional cost comes from the integration of the tangent-linear model (TLM) over the previous assimilation period, and the adjoint operation, at each iteration of the minimization process. However, since the number of iterations is usually much smaller than the dimension N of the model phase space, this is small compared to the number of TLM integrations that would be needed to evolve the corresponding $N\times N$ matrices.

It is worth noting that, since the inverse Hessian operator provided by the minimization code is always symmetric positive definite (a nice property for modeling error covariances), the corresponding limited-memory analysis error covariance operator has full rank.

The feasibility of the method depends upon the quality of the inverse Hessian operator provided by the minimization code. Some investigations are being done to assess this quality.

Having a cheap and good-quality limited-memory analysis error covariance operator would also be of interest to reduce the cost of the computation of the "Hessian singular vectors" used in the ECMWF ensemble-prediction system and Simplified Kalman Filter (M. Fisher, this volume).

The cycling strategy proposed here will be studied with a VDA system using a Tropical Pacific ocean model (A. Weaver, this volume). Choosing the assimilation period much shorter than the local and advective time scales of the flow, the error covariances may be assumed constant along the assimilation period (De Mey, 1997) and the expensive additional TLM and adjoint model integrations at each iteration may be omitted.

References

- De Mey, P., 1997: Data assimilation at the oceanic mesoscale: a review. *J. Meteorol. Soc. Japan*, **75-1B,** 415-427
- Fisher, M. and Courtier, P., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Tech. Memo.*, **220,** 28 pp. (for a complementary approach.)
- Gill, P.E., Golub, G.H., Murray, W. and Saunders, M.A., 1974: Methods for modifying matrix factorizations. *Math. Comp.*, **28,** 505-535
- Gilbert, J.-Ch. and Lemaréchal, C., 1989: Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Prog.*, **45,** 407-435
- Gunson, J.R. and Malanotte-Rizzoli, 1996: Assimilation studies of open-ocean flows. 2. Error measures with strongly nonlinear dynamics. *J. Geophys. Res.*, **101,** 28473-28488
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C., 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Japan*, **75-1B,** 181-189
- Liu, D.C. and Nocedal, J., 1989: On the limited memory BFGS method for large scale optimization. *Math. Prog.,* **45,** 503-528.
- Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Math. Comput.*, **35,** 773-782
- Rabier, F. and Courtier, P., 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.*, **118,** 649-672
- Thacker, W.C., 1989: The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.*, **94,** 6177-6196
- Veersé, F., 1999: Variable-storage quasi-Newton operators as inverse forecast/analysis error covariance matrices in variational data assimilation. *INRIA Research Report RR-3685.* (available from http://www.inria.fr/RRRT/RR-3685.html)