Université GRENOBLE ALPES

École Doctorale: Mathématiques, Sciences et Technologies de l'Information, Informatique
Spécialité: Mathématiques appliquées

Habilitation à Diriger des Recherches

presented February 27, 2018 by

Igor GEJADZE [1]

Advanced uncertainty analysis and optimal design in variational estimation (data assimilation)

Jury:

Rapporteurs (Reviewers):

Serge Gratton, Professor of applied mathematics at INPT/ENSEEIHT, France
Adrian Sandu, Professor, Virginia Polytechnic Institute and State University, USA
Andrew Stuart, Bren Professor of Computing and Mathematical Sciences, California Institute of Technology, USA

Examinateurs (Examiners):

Eric Blayo, Professor, University Grenoble-Alpes
Francois-Xavier Le Dimet, Professor Emeritus, University Grenoble-Alpes
Marc Bocquet, Professor, Ecole des Ponts, Paris Tech
Anthony Weaver, Senior Researcher, CERFACS
P.-O. Malaterre, Senior Engineer, IRSTEA-Monpellier

[1]UMR G-EAU, IRSTEA-Montpellier, 361 Rue J.F. Breton, BP 5095, 34196, Montpellier, France. Email address: igor.gejadze@irstea.fr

Summary in French

Analyse d'incertitude et conception optimale avancées en estimation variationnelle (assimilation de données)

Le manuscrit du dossier d'Habilitation à Diriger des Recherches (HDR) s'intitule : "Analyse d'incertitude et conception optimale avancées en estimation variationnelle (assimilation de données)", par Igor Gejadze. Il est basé sur une série d'articles de l'auteur couvrant la période 2008-2017, qui sont sélectionnés à partir de la liste de publication complète sur la base d'un sujet commun, comme en témoigne le titre.

L'estimation d'état et / ou de paramètres pour les systèmes dynamiques à paramètres distribués et à grande échelle est devenue une tâche de routine au cours des deux dernières décennies. C'est un processus qui combine des observations, incomplètes et éventuellement indirectes de variables d'état, avec un modèle mathématique gouverné par des équations aux dérivées partielles, complété par une information a priori. Les applications comprennent l'initialisation de modèles en météorologie et océanographie, la surveillance de la qualité de l'air ou de l'eau, le calage des modèles hydrogéologiques ou de réservoirs pétroliers, l'estimation et la prévision des débits en hydrologie ou hydraulique fluviale, l'estimation et le contrôle des écoulements en conception aéronautique, le contrôle des procédés dans l'industrie chimique ou nucléaire, etc. Dans différentes applications, ces problèmes d'estimation sont également appelés "assimilation de données" (DA), "calage" ou "problèmes inverses". L'estimation variationnelle / DA est une méthode basée sur la théorie du contrôle optimal, qui repose sur le concept de "modèle adjoint". Cette méthode est largement utilisée en géosciences. Par exemple, il s'agit d'une méthode privilégiée pour la prévision météorologique et océanique dans les principaux centres opérationnels du monde entier. La méthode a été récemment appliquée à l'hydraulique fluviale et à l'hydrologie, et a été largement utilisée dans d'autres applications scientifiques et d'ingénierie, telles que l'ingénierie aérospatiale, le traitement d'images astronomiques et biomédicales, etc.

La quantification de l'incertitude (UQ pour Uncertainty Quantification) et la conception optimale sont des sujets importants étroitement associés à l'estimation (assimilation des données). Cependant, dans le cadre de l'approche variationnelle, ces problèmes sont particulièrement difficiles à résoudre. Un aperçu du travail original de l'auteur sur des méthodologies avancées pour l'UQ et la conception optimale est présenté dans les différents chapitres. L'accent est mis sur la faisabilité des méthodes suggérées pour l'UQ et la conception en grande dimension, où les méthodes statistiques (Monte Carlo impliquant des astuces ad hoc, comme par exemple la localisation, l'échantillonnage préférentiel, etc.) peuvent échouer à produire un résultat raisonnable du fait du nombre très réduit d'échantillons disponibles. Certaines des approches pourraient également être utiles pour améliorer les techniques d'estimation elles-mêmes (en termes d'accélération, d'économies de mémoire et de robustesse).

Le Chapitre 1 contient une introduction générale à la théorie de l'estimation variationnelle et à la méthode de quantification de l'incertitude utilisée dans ce cadre. Dans le Chapitre 2 cette méthode

est généralisée au cas essentiellement non linéaire (méthode dite du "effective inverse Hessian") et également considérée du point de vue de l'estimation Bayésienne. Ensuite, le Chapitre 3 contient quatre nouvelles méthodes : une méthode pour évaluer la non-gaussianité des estimations (basée sur le concept de "mesure de coexistence"), une méthode pour calculer le gradient du critère d'optimisation dans le problème de localisation de capteurs, une méthode de contrôle dite "inactive" pour le traitement des erreurs de modèle et de paramétrisation, et une méthode pour la conception optimale du vecteur de contrôle actif. Enfin, le Chapitre 4 présente un nouveau concept de "décomposition des valeurs propres multi-niveaux", utilisé pour la représentation super-compacte de l'inverse du Hessien et pour le préconditionnement des itérations de Gauss-Newton. Pour chaque méthode, d'autres développements sont suggérés.

En outre, ce manuscrit contient le CV de l'auteur et une brève description de ses activités de recherche passées et à venir, qui sont beaucoup plus larges que les sujets reflétés dans le manuscrit.

Contents

General information **6**

Past research in brief **11**

Research plans in brief **14**

Preface **16**

1 Uncertainty quantification (UQ) in variational estimation **17**
1.1 Variational estimation: basic concept 17
1.2 Estimation error covariance and the inverse Hessian 19
1.2.1 Clarification of the existing theory 19
1.2.2 Computing H^{-1} using the LBFGS 20
1.2.3 On the importance of dynamic formulation 21
1.2.4 Illustration 22
1.2.5 On the role of the Hessian and its inverse 23

2 Advanced methods for UQ in variational estimation **26**
2.1 Effective inverse Hessian method 26
2.1.1 Theory of the method 26
2.1.2 Key implementation details 28
2.1.3 Illustration 30
2.2 Estimation error covariance versus Bayesian posterior covariance 32
2.2.1 General theory 32
2.2.2 Implementation 34
2.2.3 Illustration 35
2.3 Future developments 35
2.4 Summary 37

3 Non-trivial applications of the Hessian for UQ and design **38**
3.1 On gauss-verifiability of optimal solutions 38
3.1.1 Introduction 38
3.1.2 Coexistence measure 39
3.1.3 Coexistence measure deconvolution 42
3.1.4 Implementation details 43
3.1.5 Illustration 44
3.1.6 Conclusions and future work 45
3.2 Design of optimal observation schemes 48

3.2.1 Introduction . . . 48
3.2.2 Sensor-location problem statement . . . 49
3.2.3 Gradient via adjoint of the Hessian derivative . . . 50
3.2.4 Computation of the gradient of the design function . . . 51
3.2.5 Key implementation trick . . . 53
3.2.6 Illustration . . . 54
3.2.7 Conclusions and future development . . . 54
3.3 Implicit ('idle') control and model error . . . 56
3.3.1 Introduction . . . 56
3.3.2 Theory of the method . . . 57
3.3.3 Implementation . . . 59
3.3.4 One possible application . . . 59
3.3.5 Illustration . . . 60
3.3.6 Conclusions . . . 60
3.4 Design of the control set . . . 62
3.4.1 Introduction . . . 62
3.4.2 Goal-function error in an observed system . . . 63
3.4.3 Goal-function error covariance for partial control . . . 64
3.4.4 Implementation and performance assessment procedure . . . 66
3.4.5 Illustration . . . 67
3.4.6 Conclusions . . . 69
3.4.7 Future developments . . . 70

4 Advanced numerical approaches for computing inverse Hessian 73
4.1 Computing of the inverse Hessian: multigid approach . . . 73
4.1.1 Introduction . . . 73
4.1.2 Multilevel eigenvalue decomposition algorithm . . . 74
4.1.3 Hessian decomposition . . . 78
4.1.4 Approximating the inverse Hessian . . . 80
4.1.5 Illustration . . . 80
4.1.6 Conclusions . . . 83
4.1.7 Future developments . . . 84

General information

Curriculum vitae

Contact information:

Home address: App.114, 300 Rue des Brusses, Bat. E, Montpellier, 34090
E-mail address: igor.gejadze@irstea.fr
Telephone: +33(0)467166408 (office)
Place of birth: Tbilisi, Georgia/USSR
Nationality: United Kingdom

Post-school education with details of academic qualifications obtained:

PhD in Mechanical Engineering, obtained 23.04.1999 at Moscow Aviation Institute, College №6: Astronautical and Rocket Engineering. Thesis title: Algorithms for close-to-real time monitoring of thermal loads in flight vehicle structures, supervisor - Prof. O.M. Alifanov.

MS in Mechanical Engineering, obtained 04.03.1985 at Moscow State Technical University (Bauman's), Faculty of Mechanical Engineering, specialization M9: Flight Dynamics and Control.

Career steps:

Since 10.2013 - Senior RF (Accueil de Chercheurs de Haut Niveau), IRSTEA Montpellier regional center, France.

06.2012 - 10.2013: NERC Advanced Senior RF, Dept. of Civil Engineering, University of Strathclyde in Glasgow, UK, (open end contract, funded by NERC).

08.2006 - 06.2012: Senior RF, Dept. of Civil Engineering, University of Strathclyde in Glasgow, UK, (open end contract, funded by Glasgow Research Partnership in Engineering).

03.2005 - 08.2006: PDRF, Jean Kuntzmann Laboratory (LJK, former LMC-IMAG), Grenoble, France.

02.2003 - 02.2005: PDRA, Dept. of Civil Engineering, University of Strathclyde in Glasgow, UK.

10.2000 - 02.2003: PDRF, Faculty of Applied Mathematics and Computer Science, the Weizmann Institute of Science, Israel.

10.1999 - 09 .2000: PDRF, Polytechnic School, University of Nantes, France.

10.1994 - 02.1999: PhD student, Moscow Aviation Institute, Russia.

1991-1994: Computer programmer in private company.

1985-1991: Engineer/Senior engineer (Principal Investigator), the State Research Institute of Aviation Systems, Moscow (Aerospace industry of the USSR).

PhD students supervised:

Student: Hind Oubanas, since 10.2014 at IRSTEA-Montpellier.

Thesis title: "Détermination des débits des fleuves et grands canaux d'irrigation à partir de donnés acquises au sol ou par télédétection par des méthodes d'assimilation de données".

Contribution: 50%, joint supervision with P.-O. Malaterre (IRSTEA-Montpellier) and F. Mercier (CNES).

Stage: in progress, to be completed by 10.2017.

Current position: engineer at IRSTEA-Montpellier.

Student: Kirsty L. Brown, 10.2012 - ... , at Dept. of Mathematics and Statistics, University of Strathclyde

Thesis title: "Efficient multigrid computation of the posterior covariance matrix in large-scale variational data assimilation".

Contribution: 50% , joint supervision with A. Ramage (University of Strathclyde).

Stage: This student is currently on a long-term medical leave.

Student: Hossam Mohamed El-Hanafy, 10.2005 - 10.2007 at Dept. of Civil Engineering, University of Strathclyde.

Thesis title: "Sensitivity and uncertainty analysis for flood wave propagation in river channels using the adjoint method".

Contribution: 50%, joint supervision with G.J.M. Copeland (University of Strathclyde).

Stage: defended in 2007.

Current position: lecturer at Egyptian Military College (MTC), Cairo.

Successful funding attempts:

05.2017: TOSCA CNES call 2017: Estimation des débits avec SWOT, FUNDED by CNES, 35000 euro. (with P.-O. Malaterre)

11.2016: LEFE MANU call: Accelerating non-intrusive PC methods using adjoint sensitivities and Hessian computations, FUNDED by INSU, 5000 euro.

01.2016: TOSCA CNES call 2016: Modélisation des débits avec SWOT, FUNDED by CNES, 18240 euro. (with P.-O. Malaterre)

08.2016: PhD proposal: Assimilation de données appliquée à un modèle hydrologique distribué: calage régional et assimilation des débits observées dans la méthode AIGA. (with P.-O. Malaterre, IRSTEA Montpellier, and P. Javelle, IRSTEA Aix-en-Provence), FUNDED by IRSTEA/SHAPI.

10.2014: PhD proposal: Détermination des débits des fleuves et grands canaux d'irrigation à partir de données acquises au sol ou par télédétéction par des méthodes d'assimilation de données. (with P.-O. Malaterre, IRSTEA Montpellier), FUNDED by IRSTEA/CLS.

10.2011: NERC Advanced Fellowship Award: Method for improving forecast statistics in large-scale variational data assimilation problems, FUNDED by NERC, 300000 BP.

12.2010: PhD proposal: Efficient multigrid computation of the posterior covariance matrix in large-scale variational data assimilation (with A.Ramage, Mathematics and Statistics, University of Strathclyde). FUNDED by the University of Strathclyde.

05.2008: Bridging The Gap Grant (with A.Ramage, Mathematics and Statistics, University of Strathclyde), FUNDED, 6000BP.

07.2007: EPSRC Industrial CASE Central Pool PhD Award: Uncertainty analysis in river and estuary flood risk modeling (with G.J.M. Copeland, Civil Engineering, University of Strathclyde), FUNDED.

05.2007: Sir David Anderson Bequest Grant, FUNDED by the University of Strathclyde, 5000BP.

Miscellaneous:

General areas of expertize:
Computational and numerical mathematics, partial differential equations, control theory, inverse problems, applied statistics, fluid mechanics/heat transfer.

Particular areas of expertize:
Computational Fluid Dynamics, numerical experience (finite differences/finite volume methods) with compressible Navier-Stokes equations, incompressible Navier-Stokes equations including a free surface, boundary layer approximation of Navier-Stokes equation, shallow water equations, Saint-Venant equations, Burgers equations, convection-diffusion-reaction equation. Inverse problems in engineering applications and large-scale data assimilation problems in geophysics. Multigrid methods.

Special area of expertize:
Significant experience with Automatic Differentiation (TAPENADE), manual code differentiation and their combinations. Significant experience with large data-sets and high level database programming languages (Clipper5.1), management of data-bases.

Programming skills:
Fortran 77/90 (main experience) including basic MPI, C, C++ (superficial), Clipper 5(main experience), Visual FoxBase (superficial); UNIX/LINUX/Windows environments.

Potential teaching areas:
Calculus of variations, functional analysis, inverse problems, numerical methods, basic statistics and probability;
Analytical mechanics (basic mechanics, fluid mechanics/heat transfer, flight and orbital mechanics, vibrations);
High level programming languages for scientific computing (Fortran, C, Matlab).

Teaching experience:
Support class for undergraduates including mathematics and basic mechanics, selected lectures and practical classes in analytical mechanics. 2008-2012.

Communication skills:
Russian (native), English (excellent), French (superficial).

Collaborations:

Jean Kuntzmann Laboratory (LJK), University of Grenoble, France, visiting researcher (05/2008, 05/2009, 09/2009, 05/2010, 05/2011).

Institute of Numerical Mathematics, Russian Academy of Science.

List of publications

Book chapters

B1. Le-Dimet F.-X., Shutyaev, V.P., Gejadze I., Second Order Methods for Error Propagation in Variational Data Assimilation. Chapter in book: Advanced Data Assimilation for Geosciences, Edition: Oxford University Press, Chapter: Chapter 14, Publisher: Les Houches, Editors: Blayo, Bocquet, Cosme, Cugliandolo, pp.319-348, January 2015, DOI: 10.1093/acprof:oso/9780198723844.003.0014.

Journal papers

For the list of journal papers see Author's publications in the List of References at the end of this manuscript.

International conferences

C1. Oubanas H., Gejadze I., Malaterre P.-O., Mercier F., Estimation of river discharge from in-situ and remote sensing data, using variational data assimilation and a full Saint-Venant hydraulic model. 3rd Space for Hydrology Workshop, Frascati (Rome), Italy, September 15 17, 2015

C2. Oubanas H., Gejadze I., Shutyaev V.P., On the model error treatment in variational DA using the nuisance parameter approach. Paper 3109. The 5th Annual International Symposium on Data Assimilation. University of Reading, UK. 18-22 July 2016. http://www.isda2016.net.

C3. Oubanas H., Gejadze I., Malaterre P.-O., Mercier F., Estimation of river discharge from in-situ and remote sensing data using variational data assimilation and a full Saint-Venant hydraulic model. Paper O-247. Special Session on Inverse problems in hydrodynamics for rivers and estuaries: uncertainty quantification and data assimilation methods. Hydroinformatics, Republic of Korea. 21-26 August 2016. http://www.hic2016.org/.

C4. Oubanas H., Gejadze I., Malaterre P.-O., Adjoint sensitivity analysis of valuable functions involving the full Saint-Venant 1.5D model with application to Garonne river. Analyse de sensibilité par méthode adjointe de fonctions-objectif issues de solutions du modèle de Saint-Venant 1.5D complet avec une application sur la Garonne. Séminaire sur l'analyse de sensibilité dans les modèles hydrauliques, EDF Chatou, 5 Février 2016.

C5. Shutyaev, V., Gejadze I., Le Dimet, F.-X. Optimal solution error covariances in nonlinear problems of variational data assimilation. In: *Geophysical Research Abstracts* (2011), v.13, 1594-1.

C6. Gejadze, I., Shutyaev, V.P. Optimal solution error covariance in highly nonlinear problems of variational data assimilation. In: *6th International Conference "Inverse Problems: Identification, Design and Control". Samara, 6-11 October, 2010. Proceedings.* Moscow: Moscow Aviation Institute, 2010, 8pp.

C7. Shutyaev, V., Le Dimet, F.-X., Gejadze, I. Posterior covariances of optimal solution errors in variational data assimilation. In: *Abstracts of the 5th International Conference "Inverse Problems: Modeling and Simulation", 24-29 May 2010, Antalya, Turkey.* Izmir: Izmir University, 2010, 82–83.

C8. Le Dimet, F.-X., Gejadze I., Shutyaev, V. Error covariances via Hessian in variational data assimilation. In: *Abstracts of the 5th WMO International Symposium on Data Assimilation, 5-9 October 2009, Melbourne, Australia.* Melbourne: WMO (2009).

C9. Shutyaev, V., Le Dimet, F.-X., Gejadze I. A posteriori error covariances in variational data assimilation. In: *Geophysical Research Abstracts* (2009), v.11, 04997.

C10. Le Dimet, F.-X., Shutyaev, V., Gejadze I. The optimality system: the key for variational data assimilation. In: *5th Asia Oceania Geosciences Society Conference, AOGS 2008*, Busan, Korea, June 16-20, 2008.

C11. Le Dimet, F.-X., Shutyaev, V.P., and Gejadze, I. On optimal solution error analysis in variational data assimilation. In: *5th International Conference "Inverse Problems: Identification, Design and Control". Moscow, 11-17 May, 2007. Proceedings.* Moscow: Moscow Aviation Institute, 2007, 8pp.

C12. Le Dimet, F.-X., Shutyaev, V.P., and Gejadze, I. Analysis error via Hessian in variational data assimilation. In: *Textes des Communications CARI2006, 8th African Conference on Research in Computer Science*, Cotonou (Benin), 2006, 8 pp.

C13. Le Dimet, F.-X., Shutyaev, V.P., Gejadze I. On optimal solution error covariances in variational data assimilation. In: *Geophysical Research Abstracts* (2006), v.8, 00285.

Keynote/invited talks (fully or partially sponsored by the host)

D1. Gejadze I. Design of control set in the framework of variational data assimilation. Jean-Kuntzmann Laboratory Seminars, Grenoble, 1 June, 2017.

D2. Malaterre P.-O., Gejadze I., and Oubanas H. Assimilation de donneés, observabilité et quantification d'incertitudes appliqués à l'hydraulique à surface libre. Colloque National sur l'Assimilation de données. Grenoble. November 30 - December 02, 2016.

D3. Malaterre P.-O., Gejadze I., Lauvernet C., and Oubanas H. Introduction à l'assimilation de données et illustration sur des modèles hydrologiques (GR4J) et hydraulique (Saint-Venant 1D). Séminaire Arceau Megève. 30 Mars - 3 Avril 2015.

D4. Gejadze I. On verifiability of optimal solutions in variational data assimilation problems with nonlinear dynamics. Reading-Warwick Data Assimilation Meeting, May 12-14, 2014.

D5. Gejadze I. On the optimal solution error covariance in highly nonlinear variational DA problems. The UK Met Office, 04/2012.

D6. Gejadze I. Error analysis in nonlinear variational data assimilation problems. Department of Meteorology, University of Reading, UK, 03/2011.

D7. Gejadze I. 6ICIP: International Conference on Inverse Problems, Moscow, Russia, 10/2010.

Past research in brief

2013-2016: **Context: variational DA in hydrology**
In this pilot work the potential of variational DA in hydrology has been assessed. In particular, the tangent linear and adjoint counterparts for the well-known hydrological models GR4J and AIGA (in collaboration with HYDRIS Hydrology) were generated; then, different DA problems (parameter estimation, state monitoring and forecasting) has been investigated in the identical twin experiment framework. The results was presented in [C2], and also has been used for obtaining a PhD grant (to be started in December 2017). The paper on this subject is currently in progress.

2005-2016: **Context: control and variational DA in hydraulics**
a) Coupling of 1D and 2D hydraulic models via optimal control, a joint data-assimilation/coupling procedure implementing the idea of a '2D-zoom' over-posed model [A20]. In this work the 2D shallow water model DASSFLOW was used, whereas its tangent linear and adjoint models has been obtained by means of Automatic Differentiation.
b) Flood risk assessment for a 1D hydraulic model (full Saint-Venant equations) using analytically derived adjoint equations for sensitivity analysis [A19].
c) Data assimilation involving the full Saint-Venant hydraulic network model (SIC) in the context of SWOT (Surface Water and Ocean Topography) satellite mission [A1, A2, A5, A6]. Development of the tangent linear and adjoint counterparts for the SIC model using Automatic Differentiation. Development of a special robust procedure for discharge estimation under uncertainty in basic model parameters including bathymetry and distributed Manning/Strickler coefficient. Application to different rivers (Garonne, Poo, Sacramento).

2006-2016: **Context: optimal solution/estimation error quantification**
"In variational DA, the analysis error covariance can be approximated by the inverse Hessian": there has always been some misconception and confusion about the true meaning of this statement. For example, the following issues are not always correctly understood: the Hessian of which cost-function is considered, at what point the Hessian is defined, what the quality of approximation depends on, what is the difference between the 'analysis error covariance' and 'posterior covariance', etc. Papers published during this period clarify most of these issues [A14, A15, A16, A17, A18, A21, A22]. However, in addition to its pedagogical value, this research offers some new ideas. In particular, a useful concept of *effective inverse Hessian* is suggested in [A12, A10], also used in [A9]. The idea of a *generalized particle*, which is the Gaussian pdf (posterior state plus the corresponding inverse Hessian), can be useful in particle methods.

2012-2016: **Context: application of the Hessian in variational DA**
Working around the Hessian issue resulted into the following new applications:
a) Efficient preconditioning for inner Gauss-Newton iterations in the framework of incremental 4D-Var.

This preconditioning is based on the idea of Hessian decomposition in the observation space, combined with a super-compact storage scheme achieved using the multilevel eigenvalue decomposition [A3].
b) Gradient-based design of optimal sensor location schemes. An efficient new method for computing the gradient of the standard (alphabetical) design criteria via *the Hessian derivative* has been suggested in [A11, A13]. This enables the gradient-based optimization methods (conjugate gradients, quasi-Newton) to be used in the context of optimal experimental design.
c) Gauss-verifiability of optimal solutions in variational DA problems with nonlinear dynamics. A new verification method presented in [A8] allows the gaussianity of the optimal solutions to be quantified. An important advantage of this method is that it allows us to assess how the non-gaussian effects are distributed in space. The method is computationally efficient and matrix-free, thus allowing high-dimensional implementation.
d) Optimal control set design. A new method for ranking the active control subsets in terms of a chosen performance measure has been suggested in [A5]. The method is computationally efficient and matrix free. This is an important methodological advance in design of estimators for complex multi-input systems, given most inputs contain significant uncertainties.
e) Design of robust estimators (robust with respect to the parametric uncertainty and the model error) utilizing the 'nuisance parameter' treatment ideas [A3].

2000-2003, 2012-2015: **Context: multigrid/multilevel methods**
Multigrid approach for control of systems governed by parabolic partial differential equations.
This work has been stopped after two years of research due to appearance of papers describing the very similar approach, but at a more advanced stage (Borzi, 2003/2004). However, multigrid skills obtained at this stage has been utilized later [A4]. Multilevel approach for computing the limited-memory Hessian and its inverse in variational DA. In this work the concept of *multilevel eigenvalue decomposition* for a symmetric positive-definite operator has been introduced, which is a new concept in computational mathematics. In particular, it has been applied for computing the limited-memory representation of the inverse Hessian and its square-root. However, it can be used in many other applications where elliptic PDE operators are involved.

2003-2005: **Context: variational DA for the coastal ocean model**
Open-boundary control for Navier-Stokes equations including a free surface (NSFS); with application to coastal circulation modeling [A23, A24, A25].
This was a collaborative work with the Imperial College (London), to enable data assimilation using ICOM (the Imperial College Ocean Model), which is the finite-element adaptive-mesh model implementing the full Navier-Stockes equations (i.e. non-barotropic). Due to its finite-element base, this model is suitable for modeling of flow around very complex coastal lines and bed topographies. The question was how to define the adjoint to the full NSFS model, i.e. how to derive the adjoint to the problem defined in a variable domain dependent on the model variables? Instead of using the topological derivative, the problem has been solved by using a change of variables shifting the problem into a fixed domain. For this re-formulated problem the tangent linear and adjoint model are derived in a conventional way, then the adjoint sensitivities mapped back into original variables. Moreover, it has been shown that for the original NSFS formulation the adjoint does not exist and, therefore, cannot be obtained by means of Automatic Differentiation.

1999-2000: **Context: optimization of polymer extrusion technology**
Heat flux estimation for coupled heat transfer/fluid flow model in extrusion of polymers.
The visco-plastic polymer flow model (2D with radial symmetry) coupled with the convection-diffusion heat transfer model has been developed and used for the heat flux estimation at the extrusion die wall

[A26]. This was a pioneering work which inspired several subsequent PhD's at ISITEM (University of Nantes).

1994-1999: **Context: heat transfer in flight vehicle structures (PhD)**
Design of algorithms for on-line diagnostics of transient heat fluxes and thermal states, with possible application to the thermal protection systems (TPS). Main results of this work are presented in the PhD thesis and published in [A27, A28, A29, A30, A31, A32]. The need for development such specialized algorithms was justified by the fact that for rapidly varying large-magnitude heat fluxes the quality of Kalman estimates was not satisfactory. The suggested algorithms are based on the inverse problems theory and exploit Tikhonov regularization.

1985-1991: **Context: star-guidance for a supersonic jet**
Numerical modeling of optical distortions in shock waves surrounding a supersonic jet and in the atmosphere, i.e. analysis of refraction and dispersion of a light beam passing through: a) boundary layer in the supersonic shear flow; b) turbulent atmosphere (using 1D-turbulent atmosphere model). Development of an observation error model (dependent of the flight regime) and identification of its parameters using data obtained during test flights.

Research plans in brief

Context: advancing the methodology for UQ and control set design
This immediate-future research is described in detail in §2.3 and §3.4.7, where two new methods are outlined. These methods should be implemented and verified numerically. Other plans in this direction include a method for utilizing adjoint sensitivities in the framework of the **global** sensitivity analysis. The purpose is to enable the uncertainty quantification and propagation for essentially nonlinear models with the uncertainty-loaded high-dimensional inputs. The latter can be achieved by reducing computational costs of a non-intrusive Polynomial Chaos method using the adjoint sensitivities (first derivatives) and the limited-memory Hessians (second derivatives).

Context: variational DA in hydraulics
Assimilation of SWOT (Surface Water and Ocean Topography satellite mission) data for discharge and water level estimation using the full Saint-Venant hydraulic model. The ultimate goal of the project is to develop a piece of software which can eventually be included into the SWOT data processing library. Further research in this direction includes the following:
a) Development of the unconditionally robust estimator. The level of robustness can be increased by modifying the direct solver and by considering additional constraints (enforcing positiveness of depth, for example) in the inverse/control problem formulation.
b) Development of a simplified test-case setup procedure, which has to allow for an inexperienced user (both in hydraulics and DA).
c) Modeling the observation error with the SWOT Scientific Simulator (SS). Verification of the adequacy of the error model implemented in the SWOT SS, using statistical analysis of residuals. Development of algorithms for calibrating the SWOT SS error model.
d) Considering discharge estimation problem outside the Gaussian framework.
e) Improving the estimation accuracy. Here, the two difficulties are still presented: relatively large period of SWOT observations and unaccounted time dependent and spatially distributed lateral inflows/offtakes. These issues cannot be resolved solely at the level of hydraulic modeling. Thus, we must consider a joint hydraulic/hydrological models. For example, the characteristic time of a hydraulic model could be smaller than the satellite revisiting period, in which case all events between observation instants are not observable, whereas the characteristic time of a joint hydraulic/hydrological model can become larger than the observation period. Thus, combining the SIC model with the existing hydrological models (GR4J, AIGA) is also a subject of immediate future research. This is a basic idea of the forthcoming PhD proposal to CNES (to be started 10.2017).

Context: variational DA in hydrology
Variational DA for large-scale distributed hydrological models (AIGA) and hydrological forecasting. This is a subject of the forthcoming PhD (to be started 10.2017), where the following research directions will be considered:

a) Estimating uncertainties in the model inputs (precipitation) and outputs (local discharge) by reanalysis. At this step we decide whether or not the Gaussian framework is relevant. If not, we may turn to the MAP estimator. Also, enforcing positiveness of certain model parameters and state variables.
b) Managing the implementation issues in high dimensions. Model reduction, possibly by using the statistical learning approach.
c) Study of the hydrological predictability, given uncertainty in weather forecasts.
d) Design of optimal observation networks.

Preface

State and/or parameter estimation for large-scale distributed parameters dynamical systems has become a routine task in the past two decades. It is a fusion process of incomplete and, possibly, indirect observations of state variables with a mathematical model governed by partial differential equations, complemented by *a priori* information. The applications include model initialization in meteorology and oceanography, air and water quality monitoring, 'calibration' of groundwater and reservoir models, discharge estimation and forecasting in river hydraulics and hydrology, flow estimation and control in aerospace engineering, process control in chemical and nuclear engineering, etc. In different applications these estimation problems are also referred as 'data assimilation'(DA), 'calibration' and 'inverse problems'. Variational estimation/DA is a method based on the optimal control theory ([S32, S30]), which can also be understood as a special case of the maximum a-posteriory probability (MAP) estimator ([S13]). This method is preferred for weather and ocean forecasting in major operational centers around the globe, particularly in the form of the incremental 4D-Var ([S11]), and in the form of the *ensemble 4D-Var* ([S9]). Variational estimation is widely used in other scientific and engineering applications, for example in aerospace engineering [S2], astronomical image processing [S4], etc.

Uncertainty quantification (UQ) and optimal design are important topics closely associated with estimation/DA. An overview of the original author's work on advanced methodology for UQ and optimal design in the framework of **variational estimation** is presented in the coming chapters. The accent is made on feasibility of the suggested methods for the UQ and design in high-dimensions, where the statistical methods (Monte Carlo involving associated tricks, e.g. localization, importance sampling, etc.) may not produce a sensible outcome due to a very small sample being available. Some of the approaches could equally be useful for improving (in terms of acceleration, memory savings and robustness) the estimation/DA technique itself. This overview is based on selected papers from the complete author's bibliography.

Chapter 1

Uncertainty quantification (UQ) in variational estimation

1.1 Variational estimation: basic concept

Let us consider a mathematical (or numerical) model which describes behavior of a natural system in terms of its state variables $X \in \mathcal{X}$. Let $U \in \mathcal{U}$ be the set of the model inputs (controls), then the model can be considered as a 'control-to-state' mapping $\mathcal{M} : \mathcal{U} \to \mathcal{X}$, such that

$$X = \mathcal{M}(U), \tag{1–1}$$

where $\mathcal{U}$ and $\mathcal{X}$ are the input and state spaces, correspondingly. For modeling the system behavior the true input vector $\bar{U}$ must be specified. Under the 'perfect model' assumption the following can be postulated: $\bar{X} = \mathcal{M}(\bar{U})$. In reality, some components of $\bar{U}$ contain uncertainties $\varepsilon \in \mathcal{U}$. Thus, instead of $\bar{U}$ we use its best available approximation (background/prior)

$$U^* = \bar{U} + \varepsilon, \tag{1–2}$$

where ε is also called the background error. Because of the presence of ε, the predicted state $X|U^* = \mathcal{M}(U^*)$, that is, X evaluated (or conditioned) on U^*, also contains an error $\delta X = \mathcal{M}(U^*) - \mathcal{M}(\bar{U})$.

The state observing tools are represented by an observation operator $C : \mathcal{X} \to \mathcal{Y}$ in the form

$$Y = C(X) = C(\mathcal{M}(U)) := G(U), \tag{1–3}$$

where $G : \mathcal{U} \to \mathcal{Y}$ is a generalized input-to-observations mapping and $\mathcal{Y}$ is the 'observation' space. The true observations would be $\bar{Y} = G(\bar{U})$, however the actual observations usually contain noise ξ (observation uncertainty), i.e.

$$Y^* = \bar{Y} + \xi. \tag{1–4}$$

The aim of data assimilation is to obtain $\hat{U} = U|Y^*$, i.e. an estimate of U conditioned on observations Y^*, which should be better than the prior U^* in the sense $\|\hat{U} - \bar{U}\| < \|U^* - \bar{U}\|$. In the Bayesian framework the posterior probability density of U conditioned on observations Y^* is given by the Bayes formula

$$p(U|Y^*) = \frac{p(Y^*|U)p(U)}{p(Y^*)}. \tag{1–5}$$

Looking for the mode of the posterior density $p(U|Y^*)$, i.e. maximizing $p(U|Y^*)$, is the essence of variational data assimilation (estimation). Under the Gaussian assumption on the prior and observation uncertainties, i.e. $\varepsilon \sim N(0, B)$, $\xi \sim N(0, R)$, where B is the background error covariance and R

- the observation error covariance, maximizing $p(U|Y^*)$ is equivalent to minimizing the cost-function

$$J(U) = \frac{1}{2}\|R^{-1/2}(G(U) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(U - U^*)\|^2_{\mathcal{U}}. \tag{1–6}$$

Thus, the estimate $\hat{U}$ is obtained from the necessary optimality condition

$$J'_U(\hat{U}) = 0. \tag{1–7}$$

For the operator $G(U)$ we define the tangent linear operator $G'(U)$ (Gateaux derivative) and its adjoint $(G'(U))^*$ [S34] as follows:

$$G'_U(U)w = \lim_{t\to 0}\frac{G(U + tw) - G(U)}{t}, \quad w \in \mathcal{U}, \tag{1–8}$$

$$\left(w, (G'_U(U))^* w^*\right)_{\mathcal{U}} = \left(G'_U(U)w, w^*\right)_{\mathcal{Y}}, \quad w^* \in \mathcal{Y}. \tag{1–9}$$

Given the above operator definitions, the full gradient of $J(u)$ in (1–7) can be expressed in the form:

$$J'_U(U) = (G'_U(U))^* R^{-1}(G(U) - Y^*) + B^{-1}(U - U^*). \tag{1–10}$$

Thus, the estimate $\hat{U}$ is the solution to the operator equation

$$(G'_U(\hat{U}))^* R^{-1}(G(\hat{U}) - Y^*) + B^{-1}(\hat{U} - U^*) = 0. \tag{1–11}$$

The above presented general operator formulation of DA/estimation problems is common in nonlinear regression and in the inverse problems theory.

In the theory of dynamical systems it is usual to formulate DA/estimation problems in terms of non-stationary partial differential equations [S32]. Let $\mathcal{X}$ be the state space such that $\mathcal{X} = L_2(0, T; \Omega)$, with a norm $\|\cdot\|_{\mathcal{X}} = (\cdot,\cdot)^{1/2}_{\mathcal{X}}$, where Ω is a bounded domain of the natural space R^d ($d = 1, 2$ or 3). For the initial state control problem considered in [A17] one can write

$$\left\{\begin{array}{rcl} \dfrac{\partial\varphi}{\partial t} & = & F(\varphi) + f, \quad t \in (0, T) \\ \varphi\big|_{t=0} & = & u, \end{array}\right. \tag{1–12}$$

where $\varphi = \varphi(x, t)$ (analog of X) is the unknown function belonging for any time instant t to $L_2(\Omega)$, $x \in R^d$ is the spatial variable, $u(x) \in L_2(\Omega)$, $f(x) \in \mathcal{X}$ are the initial state and the source term, respectively, and $F(x)$ is a nonlinear operator mapping $L_2(\Omega)$ into $L_2(\Omega)$. Suppose that for a given pair (u, f) there exists a unique solution $\varphi \in \mathcal{X}$ to (1–12). Next, we introduce the functional

$$J(u) = \frac{1}{2}\|R^{-1/2}(C(\varphi) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(u - u^*)\|^2_{L_2(\Omega)}, \tag{1–13}$$

where u^* is a prior (background) of u. Thus, operator G is defined by the formula $G(u) = C(\varphi)$, where φ is the solution to equation (1–12). Now, consider the following data assimilation problem with the aim to identify the initial condition: find $u \in L_2(\Omega)$ and $\varphi \in \mathcal{X}$ such that they satisfy (1–12), and on the set of solutions to (1–12), the functional $J(u)$ takes the minimum value, i.e.

$$\left\{\begin{array}{rcl} \dfrac{\partial\varphi}{\partial t} & = & F(\varphi) + f, \quad t \in (0, T) \\ \varphi\big|_{t=0} & = & u, \\ J(u) & = & \inf\limits_{v\in L_2(\Omega)} J(v). \end{array}\right. \tag{1–14}$$

The necessary optimality condition reduces the problem (1–14) to the following optimality system [S32], [S34]:

$$\left\{\begin{array}{rcl} \frac{\partial\varphi}{\partial t} & = & F(\varphi)+f, \quad t\in(0,T) \\ \varphi\big|_{t=0} & = & \hat{u}, \end{array}\right. \tag{1–15}$$

$$\left\{\begin{array}{rcl} -\frac{\partial\varphi^*}{\partial t}-(F'(\varphi))^*\varphi^* & = & -(C'(\varphi))^*R^{-1}(C(\varphi)-Y^*), \quad t\in(0,T) \\ \varphi^*\big|_{t=T} & = & 0, \end{array}\right. \tag{1–16}$$

$$B^{-1}(\hat{u}-u^*)-\varphi^*\big|_{t=0}=0 \tag{1–17}$$

with the unknowns $\varphi, \varphi^*, \hat{u}$, where $(F'(\varphi))^*$ is the adjoint to the Frechet derivative of F, and $(C'(\varphi))^*$ is the adjoint to the Frechet derivative of C defined by

$$(C'(\varphi)\eta,\psi)_{\mathcal{Y}}=(\eta,(C'(\varphi))^*\psi)_{\mathcal{X}},\ \eta\in\mathcal{X},\psi\in\mathcal{Y}.$$

We assume that the system (1–15)–(1–17) has a unique solution. It is easy to see that this system is a detalized representation of the estimator equation (1–11) in the case the model $\mathcal{M}$ in (1–1) is a dynamical model and the input $U=\{u\}$ is the initial state $\varphi(x,0)$. Obviously, the general results obtained for (1–11) are valid for the system (1–15)–(1–17). However, the latter is used for better understanding of the structure of adjoint operators, which is sometimes vital for practical implementation of the variational DA method.

1.2 Estimation error covariance and the inverse Hessian

1.2.1 Clarification of the existing theory

Here, we present the main results from [A17] and [A14] using the general operator formulation, whereas in papers the dynamic formulation is used. The latter has its own value (see §1.2.3).

Let us consider an estimation error $\delta U=\hat{U}-\bar{U}$ and equation (1–11). We notice that

$$G(\hat{U})-Y^*=G(\hat{U})-(G(\bar{U})+\xi)=G'_U(\tilde{U})\delta U-\xi,$$

where $\tilde{U}=\bar{U}+\tau\delta U,\ \tau\in[0,1]$, and

$$\hat{U}-U^*=(\hat{U}-\bar{U})-(U^*-\bar{U})=\delta U-\varepsilon.$$

Then, equation (1–11) yields the error equation

$$(G'_U(\hat{U}))^*R^{-1}(G'_U(\tilde{U})\delta U-\xi)+B^{-1}(\delta U-\varepsilon)=0, \tag{1–18}$$

from where the estimation error can be expressed:

$$\delta U=\left((G'_U(\hat{U}))^*R^{-1}G'_U(\tilde{U})+B^{-1}\right)^{-1}\left((G'_U(\hat{U}))^*R^{-1}\xi+B^{-1}\varepsilon\right). \tag{1–19}$$

Using the first order approximations for operators $G'_U(\hat{U})\approx G'_U(\tilde{U})\approx G'_U(\bar{U})$ we express δU as follows:

$$\delta U\simeq H^{-1}(\bar{U})\left((G'_U(\bar{U}))^*R^{-1}\xi+B^{-1}\varepsilon\right), \tag{1–20}$$

where

$$H(\cdot)=(G'_U(\cdot))^*R^{-1}G'_U(\cdot)+B^{-1}. \tag{1–21}$$

Here, $(\cdot)$ denotes a placeholder for the argument of operators G and H, which shall be called the 'origin'.

It has been shown that $H(\cdot)$ represents the first-order term in the Hessian of the cost-function $J(U)$ in (1–6) or, otherwise, is the Hessian of the auxiliary cost-function as follows:

$$\mathcal{J}(\cdot, \delta U) = \frac{1}{2}\|R^{-1/2}(G'_U(\cdot)\delta U - f_1)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(\delta U - f_2)\|^2_{\mathcal{U}}, \tag{1–22}$$

where $f_1 \in \mathcal{Y}$ and $f_2 \in \mathcal{U}$ are trial functions, including $f_1 = 0$ and $f_2 = 0$. We assume that H is positive definite and, hence, invertible. If the errors ξ and ε indeed satisfy the conditions $\varepsilon \sim N(0, B)$, $\xi \sim N(0, R)$, then the estimation error covariance is

$$P = E[\delta U \delta U^T] \simeq H^{-1}(\bar{U}). \tag{1–23}$$

The above relationship is exact for linear G; for nonlinear G it is valid if the tangent linear hypothesis is valid. Since in reality the true value $\bar{U}$ may not be known, one uses the particular estimate (event) $\hat{U}$ instead, i.e.

$$P \simeq H^{-1}(\hat{U}). \tag{1–24}$$

There are several papers presenting essentially the same result for dynamical models [S36, S44]. However, there has also been some confusion. For example, $H(\cdot)$ is often perceived as a Hessian of the cost-function $J(U)$ in (1–6), whereas it is the Hessian of the auxiliary cost-function $\mathcal{J}(U)$ in (1–22). Another important point is that $H^{-1}(\bar{U})$ is the best possible approximation to P using H, whereas $H(\hat{U})$ is only an approximation to $H(\bar{U})$. The latter fact has never been clearly underlined before.

1.2.2 Computing H^{-1} using the LBFGS

First, we define the preconditioned (projected) Hessian in the form

$$\tilde{H} = (B^{1/2})^* H B^{1/2} = (B^{1/2})^* (G'_U(\cdot))^* R^{-1} G'_U(\cdot) B^{1/2} + I,$$

which is usually much better conditioned that H. Having computed $\tilde{H}^{-1}$ in the limited-memory form H^{-1} is recovered as follows:

$$H^{-1} = B^{1/2} \tilde{H}^{-1} (B^{1/2})^*.$$

It is easy to see that $\tilde{H}$ is the Hessian of the following auxiliary cost-function

$$\tilde{\mathcal{J}}(\cdot, \delta U) = \frac{1}{2}\|R^{-1/2}(G'_U(\cdot)B^{1/2}\delta U - f_1)\|^2_{\mathcal{Y}} + \frac{1}{2}\|\delta U - f_2\|^2_{\mathcal{U}}. \tag{1–25}$$

We use the LBFGS method for generating the limited-memory representation of $\tilde{H}^{-1}$ as the by-product of minimization of $\tilde{\mathcal{J}}(\cdot, \delta U)$. A set of **secant** pairs of a given (prescribed) size is accumulated during minimization, then the product $\tilde{H}^{-1}v$ is recovered by the BFGS recursion. The key improvement of the quality of approximation (for a given number of secant pairs) has been achieved by using the exact step search, as described in §5.1 in [A17]. The optimized (in terms of computational expenses) version of this algorithm is presented in our later paper [A12], §6.2.

The Lanczos method is routinely used for computing $\tilde{H}^{-1}$, whereas the BFGS or LBFGS are usually seen as inferior to the Lanczos. Indeed, for the same accuracy of approximation achieved, representation using the BFGS update formula is less compact than the eigenvalue representation. The eigenvalue analysis, however, could be noticeably more expensive to compute (in terms of function calls). Besides, control of computational costs in the LBFGS algorithm is easier than in the Lanczos method.

We can cite only [S50] (preceding to [A17, A14] in time) where the same technique is considered. The main focus in this paper is, however, on methods for computing the inverse of the analysis error covariance P, given the BFGS representation of H^{-1} obtained in the inner loop of the incremental approach. At the same time, matrix H^{-1} itself is not of interest and its accuracy is not assessed.

1.2.3 On the importance of dynamic formulation

Let us note that in papers [A17] and [A14] the dynamic formulation is used. For example in the initial state control problem, the Hessian-vector product is defined by the successive solution of the tangent linear and adjoint models as follows

$$\left\{ \begin{array}{rcl} \frac{\partial \psi}{\partial t} - F'(\varphi)\psi &=& 0,\ t \in (0,T), \\ \psi|_{t=0} &=& v, \end{array} \right. \tag{1–26}$$

$$\left\{ \begin{array}{rcl} -\frac{\partial \psi^*}{\partial t} - (F'(\varphi))^*\psi^* &=& -(C'(\varphi))^* R^{-1} C'(\varphi)\psi,\ \ t \in (0,T) \\ \psi^*\big|_{t=T} &=& 0, \end{array} \right. \tag{1–27}$$

$$H(u)v = B^{-1}v - \psi^*|_{t=0}, \tag{1–28}$$

where $\varphi(u)$ is the solution of equation (1–12). These equations correspond to equation (1–21) with $U = \{u\}$. The dynamic formulation is important because it shows different implementation options. For example, if operator G is considered as a black box, then operators $G'_U(\cdot)$ and $(G'_U(\cdot))^*$ can be obtained by means of the Automatic Differentiation (AD) following the 'discretize-then-optimize' approach. In one hand, the forward, tangent linear and adjoint models obtained this way are mutually 'consistent'. This is a vital property required for minimization of realistic high-dimensional models. On the other hand, the models generated by means of the AD are often computationally cumbersome and should be optimized manually.

For computing H^{-1} one must compute the product Hv several times at the same origin point (model trajectory). Thus, the trajectory may be computed only once. In practice, computing the solution and the perturbation in the AD produced TL model is done 'simultaneously', i.e. one line of the TL solver follows the corresponding line of the direct solver. Moreover, in case of nonlinear iterations (to converge on φ), the TL solver follows step by step the iterative procedure, computing the perturbation variable by increments. This is highly inefficient. For computing the Hessian, however, it might be sufficient to have full consistency between the TL and adjoint models only. The forward and TL models can be partially consistent, i.e. consistent in terms of the discretized operator $F(\cdot)$, however the time integration method could be different. In this case the model's trajectory once computed and stored in memory can be supplied into the TL solver. Since the TL model is linear, the nonlinear treatment involved with the forward solver is not involved with the TLM. Thus, one can obtain a computationally efficient TL model. By applying the AD to the auxiliary cost-function (1–22), one can get the code which computes its gradient. The latter defines the product Hv according to (1–25). The code involves the adjoint model consistent with the modified TLM. Let us note that this adjoint may not be suitable for solving the DA problem itself.

For the parameter estimation (control) problem considered in [A14], the mathematical model is described by the evolution problem

$$\left\{ \begin{array}{rcl} \frac{\partial \varphi}{\partial t} &=& F(\varphi, \lambda) + f, \quad t \in (0,T) \\ \varphi\big|_{t=0} &=& u, \end{array} \right. \tag{1–29}$$

where u is the known initial condition, $\lambda \in \mathcal{X}_p$ is an unknown model parameter and $\mathcal{X}_p$ is the parameter space.

Let us introduce the functional

$$J(\lambda) = \frac{1}{2}\|R^{-1/2}(C(\varphi) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(\lambda - \lambda^*)\|^2_{\mathcal{X}_p}, \tag{1–30}$$

where $\lambda^* \in \mathcal{X}_p$ is a prior (background) of λ, and B is the covariance of $\delta\lambda = \lambda^* - \bar{\lambda}$, and consider the following DA problem with the aim to estimate the parameter λ: for given (u, f), find $\lambda \in \mathcal{X}_p$ and $\varphi \in \mathcal{X}$ such that they satisfy (1–29), and on the set of solutions to (1–29), the functional $J(\lambda)$ takes the minimum value, i.e.

$$\left\{ \begin{array}{rcl} \dfrac{\partial \varphi}{\partial t} & = & F(\varphi, \lambda) + f, \quad t \in (0, T) \\ \varphi\big|_{t=0} & = & u, \\ J(\lambda) & = & \inf\limits_{v \in \mathcal{Z}_p} J(v). \end{array} \right. \tag{1–31}$$

The Hessian-vector product is defined by the successive solution of the tangent linear and adjoint models as follows:

$$\left\{ \begin{array}{rcl} \dfrac{\partial \psi}{\partial t} - F'_\varphi(\bar{\varphi}, \bar{\lambda})\psi & = & F'_\lambda(\bar{\varphi}, \bar{\lambda})v, \ t \in (0, T), \\ \psi|_{t=0} & = & 0, \end{array} \right. \tag{1–32}$$

$$\left\{ \begin{array}{rcl} -\dfrac{\partial \psi^*}{\partial t} - (F'_\varphi(\bar{\varphi}, \bar{\lambda}))^* \psi^* & = & -(C'(\bar{\varphi}))^* R^{-1} C'(\bar{\varphi})\psi, \ \ t \in (0, T) \\ \psi^*\big|_{t=T} & = & 0, \end{array} \right. \tag{1–33}$$

$$H(\hat{\lambda})v = B^{-1}v - (F'_\lambda(\bar{\varphi}, \bar{\lambda}))^* \psi^*. \tag{1–34}$$

These equations correspond to equation (1–21) with $U = \{\lambda\}$. From (1–26)-(1–28) and (1–32)-(1–34) one can understand the difference between $G'_u(\cdot)$ and $G'_\lambda(\cdot)$. That is, there exists a basic TL equation (system of PDE-based equations) which describes the evolution of perturbations through the system. However, the perturbations are applied at different entries: at the initial condition (case of the initial value control), and as a source term (case of parameter control). The adjoint equations coincide, the expression for the gradient expressed via the adjoint variable ψ^* is different. This structure suggests an easy way of modifying the system (1–26)-(1–28) into (1–32)-(1–34). It is important to keep consistency between operator $F'_\lambda(\cdot, \cdot)$ and its adjoint. Since $F'_\lambda(\cdot, \cdot)$ is not acting on ψ, it may not necessarily be consistent with $F'_\varphi(\cdot, \cdot)$. All these details are important to design a computationally efficient TL and adjoint models for computing the Hessian-vector product (possibly not suitable for computing the gradient of $J(U)$).

1.2.4 Illustration

The papers [A17, A14] describe a unique set of numerical tests which illustrate the use of the inverse Hessian H^{-1} for approximating the estimation error covariance P in data assimilation problems involving dynamical models. While the relationship (1–24) has been well known long before our publications, no results similar to those presented in [A17, A14] could be found in the preceding literature.

As an evolution model for the state variable $\varphi(t, x)$ we use the 1D nonlinear convection-diffusion equation

$$\left\{ \begin{array}{rcl} \dfrac{\partial \varphi}{\partial t} & = & -w\dfrac{\partial \varphi}{\partial x} + \dfrac{\partial}{\partial x}\left(k(\varphi)\dfrac{\partial \varphi}{\partial x}\right), \quad t \in (0, T], \quad x \in (0, 1), \\ \varphi(0, x) & = & u, \end{array} \right. \tag{1–35}$$

where $w(t)$ is the convection velocity, $k(\varphi)$ is the diffusion coefficient, with the Neumann boundary conditions

$$\frac{\partial \varphi(t, 0)}{\partial x} = q_1(t), \quad \frac{\partial \varphi(t, 1)}{\partial x} = q_2(t). \tag{1–36}$$

The initial condition, boundary conditions (fluxes) and diffusion coefficient estimation problems have been considered. H^{-1} have been verified against the ensemble-based covariance estimate. The difference between the two shall be called the 'deterministic covariance approximation' or DCA error.

Given the model is perfect and the input error statistics are precise, the DCA error is due to linearization of (1–18) around $\bar{U}$ (linearization DCA error component), and due to the shift of the origin from $\bar{U}$ to $\hat{U}$ (origin DCA error component). As numerical tests show, the latter could be far more dangerous than the former. This can be explained by the fact that the local linearization error $(F(\hat{\varphi}) - F(\bar{\varphi})) - F'_{\varphi}(\bar{\varphi})(\hat{\varphi} - \bar{\varphi})$ can be dumped during time integration of systems (1–26)-(1–28) or (1–32)-(1–34), and further reduced when the expectation operator is applied to $\delta U \delta U^T$. Thus, H^{-1} can serve as a good approximation of P in the nonlinear case, assuming that the origin DCA error component is reasonably small.

To conclude this section we present examples of the estimation error covariance matrices in the diffusion coefficient and the boundary flux estimation problems. All details and conditions of the tests can be found in [A14].

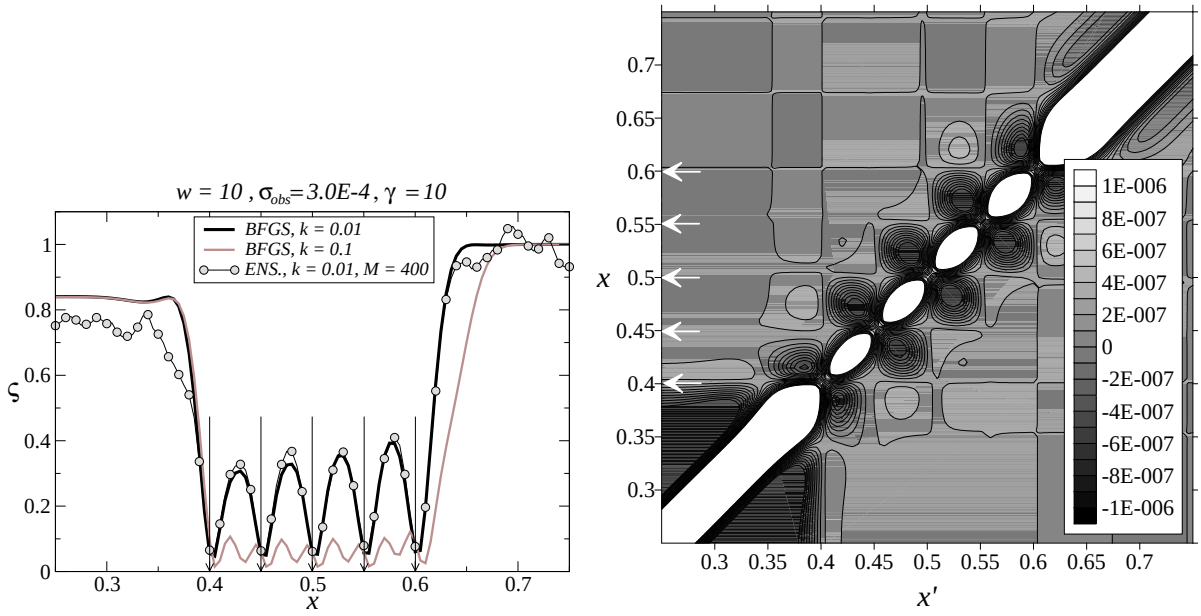

Figure 1.1: Diffusion coefficient estimation problem. Left - Scaled variance ζ and ensemble variance $\hat{\zeta}$ for $k = 0.01$ and ζ for $k = 0.1$. Left - covariance for $k = 0.01$.

1.2.5 On the role of the Hessian and its inverse

The importance of the Hessian matrix and its inverse in variational DA for geophysical applications is underlined in [S44], although this has been a well-known fact in statistics for decades (see, for example, [S42]). The previous section illustrates that for linear and moderately non-linear DA/estimation problems $H^{-1}(\cdot)$ can serve as an approximation of the estimation (analysis) error covariance matrix. In particular, confidence intervals for the components of the estimate $\hat{U}$ can be defined by the corresponding diagonal elements (variance) of $H^{-1}(\hat{U})$. A column c_i of $\mathcal{H}^{-1}(\cdot)$ which includes the i^{th} diagonal element can be obtained by solving the equation $H(\cdot)c_i = e_i$ (where e_i is a Euclidean unit vector). If the number of requested diagonal elements is significant, it would be much less expensive to evaluate $H^{-1}(\cdot)$ once and keep it in some limited-memory form, than to retrieve necessary diagonal elements using the Hessian-vector product rule.

In addition, $H^{-1}(\cdot)$ is involved in several other aspects of statistical pre- and post-processing and design of DA systems. Firstly, as the Hessian $H(\cdot)$ is equivalent to the Fisher information matrix (up

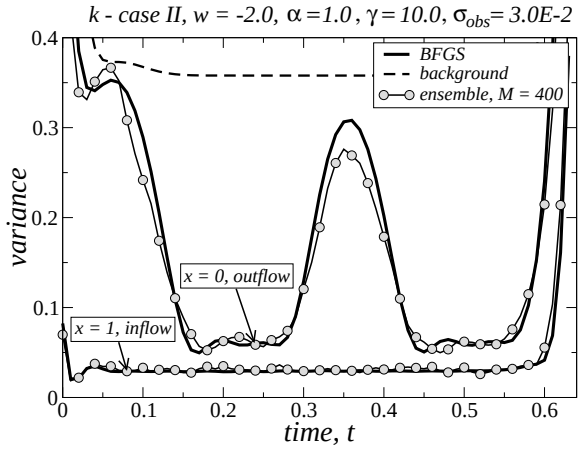

Figure 1.2: Boundary fluxes estimation problem: background, H^{-1}-based and large-sample variances for the inflow and outflow boundaries.

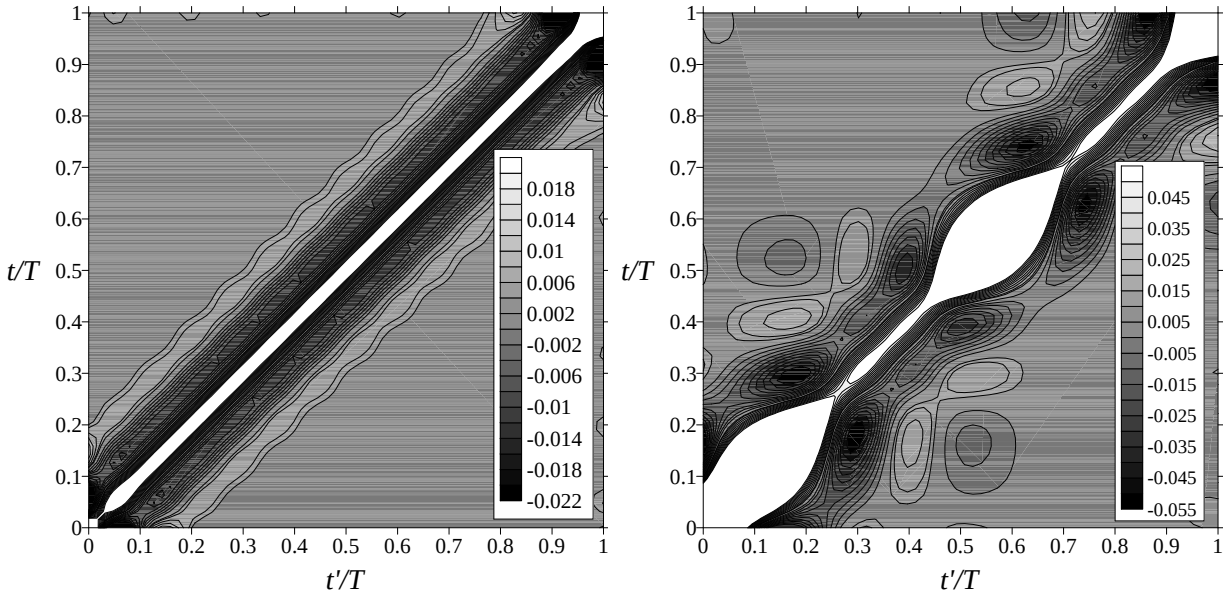

Figure 1.3: Boundary fluxes estimation problem: H^{-1}- based covariance. Left - inflow boundary, right - outflow boundary.

to a constant multiplier), the diagonal elements of the inverse Hessian can also be used in the context of optimal experimental design involving such optimality criteria as l-optimality, for example. Secondly, the analysis probability density function (pdf) is defined by the estimate $\hat{U}$ and the estimation error covariance. Random functions from the Gaussian distribution $\mathcal{N}(\hat{U}, H^{-1}(\hat{U}))$ can therefore be used as 'particles' of the ensemble of initial states, which may be useful for ensemble forecasting [S45]. These functions can be generated using $U = U^* + H^{-1/2}(\hat{U})\xi$, where $\xi \sim \mathcal{N}(0, I)$, or using the eigenvalues of $H(\hat{U})$ [S16]. However, in highly non-linear cases the 'particles' generated using this approach are unlikely to belong to the true posterior distribution, thus one must solve perturbed DA problems. This approach is referred to as the *fully non-linear ensemble method* [A17], or randomised maximum

likelihood method [S8]. In these cases, an approximation of $H^{-1/2}(\hat{U})$ can be used for preconditioning the non-linear minimization process to accelerate convergence, often with impressive results. Thirdly, the analysis error δU and the data errors ε and ξ are related via the approximate error equation (1–20), which can be used as a meta-model for investigating the effects of non-Gaussian data errors on the estimation error pdf. Lastly, if the model depends on parameters $\theta \in \Theta$, where Θ is the parameter space, an important problem is to evaluate the sensitivity of the estimation error to uncertainty ξ_θ in these parameters. This can be done using the relationship

$$H(\cdot)\delta U = (G'_U(\cdot))^* R^{-1} G'_\theta(\cdot)\xi_\theta.$$

Once again, H must be inverted to obtain δU.

Another application is related to the use of H as a coefficient matrix (and preconditioner) in incremental 4D-Var [S11]. Here each step of an outer iterative Gauss-Newton process is of the form $U^{j+1} = U^j + \alpha^j\ \delta U^j$, with U^j a discrete approximation of the unknown initial state at iteration j, descent step α^j and update (descent direction) δU^j. As the update satisfies

$$H(U^j)\delta U^j = -G(U^j), \tag{1–37}$$

where $G(U^j)$ is the gradient of the cost function, a system of linear equations involving H has to be solved at each step. Given a Hessian-vector product evaluation routine, the systems in (1–37) are usually solved iteratively using, for example, the conjugate gradient (CG) algorithm. An approximation of H^{-1} or $H^{-1/2}$, if available at a reasonable cost, can therefore be used to precondition equation (1–37) to accelerate convergence of this inner iteration.

Chapter 2

Advanced methods for UQ in variational estimation

2.1 Effective inverse Hessian method

2.1.1 Theory of the method

Here, we present the main results from [A12] and [A10]. As mentioned above, the DCA error contains two components: linearization error component and the origin error component. First, we will see how the linearization component can be reduced.

Let us first consider a related abstract problem. Let g be a random vector of dimension n defined as follows

$$g = A(u, \eta)\xi,$$

where ξ is a Gaussian random vector of dimension n with zero mean and the covariance matrix B_ξ, $\eta(\xi)$ is a random vector with elements which are linearly or nonlinearly dependent on ξ, u is a given non-random vector and A is a $n \times n$ matrix with elements $a_{i,j}$ nonlinearly dependent on u and η. Let us assume that the matrix A can be presented in the form

$$A(u, \eta) = A(u, 0) + \delta A(u, \eta),$$

where $A(u, 0)$ is the non-random part of A and $\delta A(u, \eta)$ is the random part. The covariance matrix of the vector g is

$$P_g = E\left[(A(u, \eta)\xi - E[g])(A(u, \eta)\xi - E[g])^T\right]. \tag{2–1}$$

One can see that if η does not depend on ξ, i.e. vectors η and ξ are not correlated, then $E[g] = 0$, and $E\left[A(u, \eta)\xi\xi^T A^T(u, \eta)\right] = E\left[A(u, \eta)E[\xi\xi^T]A^T(u, \eta)\right]$. For correlated vectors η and ξ (since we assume that $\eta = \eta(\xi)$), the following proposition can be put forward:

Proposition (intuitive): Matrix $E\left[A(u, \eta)E[\xi\xi^T]A^T(u, \eta)\right] = E\left[A(u, \eta)B_\xi A^T(u, \eta)\right]$ is a better approximation to P_g in (2–1) than $A(u, 0)B_\xi A^T(u, 0)$, both in terms of the Frobenius and Riemann distances. The approximation accuracy increases when the correlation level between η and ξ decreases.

While the justification of this proposition provided in [A12] is rather weak, its validity has been analyzed numerically.

Let us consider equation (1–19) in the form

$$\delta U = \left(\mathcal{H}(\bar{U}, \delta U, \tau)\right)^{-1}\left((G'_U(\bar{U} + \delta U))^* R^{-1}\xi + B^{-1}\varepsilon\right), \tag{2–2}$$

where

$$\mathcal{H}(\bar{U}, \delta U, \tau) = (G'_U(\bar{U} + \delta U))^* R^{-1} G'_U(\bar{U} + \tau\delta U) + B^{-1}. \tag{2–3}$$

Since the nonlinear least-squares estimator is asymptotically unbiased, we assume that $E[\delta U]$ is small and, therefore, for the covariance P we have an expression as follows:

$$\begin{aligned} P := E[\delta U \delta U^T] &= E\left[\mathcal{H}^{-1}(G'_U)^* R^{-1}\ \xi\xi^T\ R^{-1} G'_U(\mathcal{H}^{-1})^*\right] \\ &+ E\left[\mathcal{H}^{-1} B^{-1}\ \varepsilon\varepsilon^T\ B^{-1}(\mathcal{H}^{-1})^*\right] \\ &+ E\left[\mathcal{H}^{-1} B^{-1}\ \varepsilon\xi^T\ R^{-1} G'_U(\mathcal{H}^{-1})^*\right] \\ &+ E\left[\mathcal{H}^{-1}(G'_U)^* R^{-1}\ \xi\varepsilon^T\ B^{-1}(\mathcal{H}^{-1})^*\right]. \end{aligned} \tag{2–4}$$

As discussed in the beginning of this section, we approximate the products $\xi\xi^T$, $\varepsilon\varepsilon^T$, $\xi\varepsilon^T$ and $\varepsilon\xi^T$ in (2–4) by $E[\xi\xi^T] = R$, $E[\varepsilon\varepsilon^T] = B$, $E[\xi\varepsilon^T] = 0$ and $E[\varepsilon\xi^T] = 0$, respectively. Thus, we write an approximation of P as follows:

$$P = E\left[\mathcal{H}^{-1}(\bar{U}, \delta U, \tau)\left((G'_U(\bar{U} + \delta U))^* R^{-1} G'_U(\bar{U} + \delta U) + B^{-1}\right)(\mathcal{H}^{-1}(\bar{U}, \delta U, \tau))^*\right], \tag{2–5}$$

where the expression in the round brackets is the Hessian (1–21) at the point $\bar{U} + \delta U$. Therefore, the equation (2–5) can be written in the form

$$P = E\left[\mathcal{H}^{-1}(\bar{U}, \delta U, \tau)\ H(\bar{U} + \delta U)\ (\mathcal{H}^{-1}(\bar{U}, \delta U, \tau))^*\right]. \tag{2–6}$$

The value $\tau = 1/2$ is the optimal one to achieve the best approximation of a difference $E[G(\bar{U} + \delta U) - G(\bar{U})]$ by $E[G'_U(\bar{U} + \tau\delta U)\delta U]$, for $\delta U \sim N(0, P)$. In this case, however, one must deal with operator $\mathcal{H}(\bar{U}, \delta U, \tau = 1/2)$, which is neither symmetric, nor positive definite, which may seriously complicate its eigenvalue analysis and the subsequent limited-memory representation. Besides, the double product formula is sensitive to the errors which result from the already accepted approximations. By assuming $\tau = 1$ in $\mathcal{H}(\bar{U}, \delta U, \tau)$ and keeping in mind that $\mathcal{H}(\bar{U}, \delta U, \tau = 1) := H(\bar{U} + \delta U)$ this formula can be further simplified by the following expression:

$$P = E\left[H^{-1}(\bar{U} + \delta U)\right]. \tag{2–7}$$

The right-hand side of (2–7) may be called the 'effective inverse Hessian', hence the name of the suggested method. In order to compute P directly using this equation, the expectation must be substituted by the ensemble mean:

$$P = \frac{1}{L}\sum_{l=1}^{L} H^{-1}(\hat{U}_l), \tag{2–8}$$

where $\hat{U}_l$ are elements from the ensemble of estimates $\{\hat{U}_l\}$, $l = 1, \ldots, L$, being obtained using perturbed data. Obviously, having such ensemble evaluated, the covariance matrix can be computed according to its definition as follows:

$$\hat{P} = \frac{1}{L_s}\sum_{l=1}^{L_s}(\hat{U}_l - \bar{U})(\hat{U}_l - \bar{U})^T. \tag{2–9}$$

The advantage of (2–8) is, however, that for $L = L_s$ it gives much better approximation of P than (2–9), particularly for small L_s.

2.1.2 Key implementation details

Preconditioning

Implementation of formula (2–8) implies that L inverse Hessians have to be computed. This looks like an extremely laborious task, however, computational costs can be drastically reduced by preconditioning. We notice that the inverse Hessians are evaluated at the origin points perturbed around the true input value $\bar{U}$, which means that the difference between $H^{-1}(\bar{U}+\delta U_l)$ and $H^{-1}(\bar{U})$ can be described by just a few low-rank updates. Thus, we consider a projected Hessian in the form

$$\tilde{H}(\cdot) = (B^{1/2})^* H(\cdot) B^{1/2} = (B^{1/2})^* (G'_U(\cdot))^* R^{-1} G'_U(\cdot) B^{1/2} + I, \tag{2–10}$$

and use $\tilde{H}^{-1/2}(\bar{U})$ for the second-level preconditioning of $\tilde{H}(\bar{U}+\delta U_l)$ as follows:

$$\tilde{\tilde{H}}(\bar{U}+\delta U_l) = \tilde{H}^{-1/2}(\bar{U}) \tilde{H}(\bar{U}+\delta U_l) \tilde{H}^{-1/2}(\bar{U}). \tag{2–11}$$

One can see that $\tilde{\tilde{H}}(\cdot)$ is the Hessian of an auxiliary cost-function

$$\tilde{\mathcal{J}}(\cdot, \delta U) = \frac{1}{2} \| R^{-1/2} (G'_U(\cdot) \tilde{H}^{-1/2}(\bar{U}) B^{1/2} \delta U - f_1) \|^2_{\mathcal{Y}} + \frac{1}{2} \| \delta U - f_2 \|^2_{\mathcal{U}}. \tag{2–12}$$

The limited memory approximation of $\tilde{\tilde{H}}(\cdot)$ is evaluated as a by-product of minimization of this cost-function by the LBFGS method, then, $H^{-1}(\bar{U}+\delta U_l)$ is recovered by the formula

$$H^{-1}(\bar{U}+\delta U_l) = B^{1/2} \tilde{H}^{-1/2}(\bar{U}) \tilde{\tilde{H}}^{-1}(\bar{U}+\delta U_l) \tilde{H}^{-1/2}(\bar{U}) (B^{1/2})^*. \tag{2–13}$$

Taking into account (2–8) and (2–13), the expression for the optimal solution error covariance P reads

$$P = B^{1/2} \tilde{H}^{-1/2}(\bar{U}) \left(\frac{1}{L} \sum_{l=1}^{L} \tilde{\tilde{H}}^{-1}(\bar{U}+\delta U_l) \right) \tilde{H}^{-1/2}(\bar{U}) (B^{1/2})^*. \tag{2–14}$$

The first step is, therefore, to find a limited-memory approximation of $\tilde{H}^{-1}(\bar{U})$. This is done by minimizing the auxiliary cost-function (1–25) using the LBFGS algorithm. Next, having the product $\tilde{H}^{-1}(\bar{U}) \cdot v$ defined, one can evaluate the leading eigenpairs $(\{\lambda_k^{(0)}, W_k^{(0)}\},\ k = 1, \ldots, K_0)$ of $\tilde{H}^{-1}(\bar{U})$ using the Lanczos algorithm.

Given the leading (maximum magnitude) eigenpairs $(\{\lambda_k, W_k\},\ k = 1, \ldots, K)$ of any symmetric operator A, the limited-memory representation of A in power β can be constructed as follows:

$$A^\beta \cdot v = I \cdot v + \sum_{k=1}^{K} (\lambda_k^\beta - 1) W_k (W_k)^* \cdot v, \tag{2–15}$$

In particular, for the square-root $\tilde{H}^{-1/2}(\bar{U})$ we get:

$$\tilde{H}^{-1/2}(\bar{U}) \cdot v = I \cdot v + \sum_{k=1}^{K_0} ((\lambda_k^{(0)})^{1/2} - 1) W_k^{(0)} (W_k^{(0)})^* \cdot v, \tag{2–16}$$

Let us note that at this factorization stage we do not run PDE models, so the Lanczos method is affordable. The product $\tilde{H}^{-1/2}(\bar{U}) \cdot v$ is involved in (2–12).

Quasi-random approach

The implementation of formulas (2–8) or (2–14) requires a set of estimates $\hat{U}_l = \bar{U} + \delta U_l,\ l = 1, \ldots, L$ to be computed. Each estimate is a solution to the original data assimilation problem for the cost-function (1–6) with perturbed data $U_l^* = \bar{U} + \xi_l$ and $Y_l^* = \bar{Y} + \varepsilon_l$, where ξ_l and ε_l are random events from $\xi \sim N(0, R)$ and $\varepsilon \sim N(0, B)$, respectively. Evaluating such a set could be fairly expensive.

Here, an alternative approach is suggested. If we denote by $f_{\delta U}$ the multivariate probability density of the estimation error δU, then the equation (2–7) can be re-written in the form

$$P = \int_{-\infty}^{+\infty} H^{-1}(\bar{U} + v) f_{\delta U}(v) dv. \tag{2–17}$$

Since we assume that the estimation error is approximately Gaussian with zero mean and covariance P we obtain

$$P = \frac{1}{(2\pi)^{M/2}|P|^{1/2}} \int_{-\infty}^{+\infty} H^{-1}(\bar{U} + v) \exp\left(-\frac{1}{2} v^T P^{-1} v\right) dv. \tag{2–18}$$

In contrast to formula (2–8), the above expression is a nonlinear matrix integral equation (deterministic) with respect to P, while v is a dummy variable. This equation can be solved by the fixed point iterative process

$$P^{k+1} = \frac{1}{(2\pi)^{M/2}|P^k|^{1/2}} \int_{-\infty}^{+\infty} H^{-1}(\bar{U} + v) \exp\left(-\frac{1}{2} v^T (P^k)^{-1} v\right) dv, \tag{2–19}$$

for $k = 0, 1, ...$, starting with $P^0 = H^{-1}(\bar{u})$.

The evaluation of the multi-dimensional integral in (2–19) using the quasi-Monte Carlo method means returning to formula (2–8). Taking into account (2–19), the iterative process takes the form

$$\left\{ \begin{array}{rcl} P^{k+1} & = & \frac{1}{L} \sum_{l=1}^{L} H^{-1}(\bar{U} + \delta U_l^k), \\ P^0 & = & H^{-1}(\bar{u}), \quad k = 0, 1, ... \end{array} \right. \tag{2–20}$$

where $\delta U_l^k \sim N(0, P^k)$.

One can see that for each k the last formula looks similar to (2–8) with one key difference: δU_l^k in (2–20) is not an estimation error itself, but a vector having the statistical properties of this error. It is generated as follows:

$$\delta U_l^k = (P^k)^{1/2} \eta_l, \quad \eta_l \sim N(0, I).$$

where η_l is a quasi-random sequence from $\mathcal{N}(0, I)$. This approach is similar to the inflation approach used in statistical estimation methods (ensemble Kalman and particle filtering). If the preconditioning (2–10)-(2–11) is involved, the process (2–20) reads as follows:

$$\left\{ \begin{array}{rcl} \tilde{\tilde{P}}^{k+1} & = & \frac{1}{L} \sum_{l=1}^{L} \tilde{\tilde{H}}^{-1}(\bar{U} + \delta U_l^k), \\ \tilde{\tilde{P}}^0 & = & I, \quad k = 0, 1, ..., \end{array} \right. \tag{2–21}$$

where $\delta U_l^k \sim N(0, P^k)$, and the perturbation is defined as follows:

$$\delta U_l^k = (\tilde{\tilde{P}}^k)^{1/2} \tilde{H}^{-1/2}(\bar{U}) (B^{1/2})^* \eta_l, \quad \eta_l \sim N(0, I).$$

Finally, the covariance P in the original space is recovered by the formula

$$P = B^{1/2} \tilde{H}^{-1/2}(\bar{U}) \tilde{\tilde{P}} \tilde{H}^{-1/2}(\bar{U}) (B^{1/2})^*. \tag{2–22}$$

2.1.3 Illustration

Numerical experiments which justify the presented theory has been performed for Burgers' equation

$$\left\{ \begin{array}{rcl} \frac{\partial \varphi}{\partial t} & = & -\frac{\partial \varphi^2}{\partial x} + \frac{\partial}{\partial x}\left(k(\varphi)\frac{\partial \varphi}{\partial x}\right), \quad t \in (0,T], \quad x \in (0,1), \\ \varphi(0,x) & = & u. \end{array} \right. \tag{2–23}$$

The solutions to the Burger's equation are known to develop sharp field gradients (shocks in the inviscid case), where the nonlinear effects are particularly strong and, subsequently, the local DCA error may be significant. We define this error as a difference between the deterministic covariance P (either via (1–24) or (2–7)) and the sample covariance $\hat{P}$ obtained by the ensemble method (2–9) using a large sample ($L_s = 2500$). We consider two different initial conditions (cases A and B), the corresponding field evolutions are presented in Fig.2.1. For each case we consider a different sensor configuration: case A - 5 sensors located at $\bar{x} = (0.35,\ 0.4,\ 0.5,\ 0.6,\ 0.65)^T$; case B - 5 sensors located at $\bar{x} = (0.35,\ 0.45,\ 0.5,\ 0.55,\ 0.65)^T$.

The effect of using the 'effective inverse Hessian' (2–7) instead of the inverse Hessian (1–24) is demonstrated in Fig. 2.2. The upper panel of this figure shows the relative error

$$\varepsilon_i = P_{i,i}/\hat{P}_{i,i} - 1, \quad i = 1, ..., M,$$

whereas its lower panel shows the sampling error

$$\hat{\varepsilon}_i = \hat{P}_{i,i}|_{L'_s}/\hat{P}_{i,i}|_{L_s=2500} - 1, \quad i = 1, ..., M,$$

where $L'_s = 25$ and $L'_s = 100$. The latter indicates the level of error which corresponds to the sample covariance matrix computed without 'localization'.

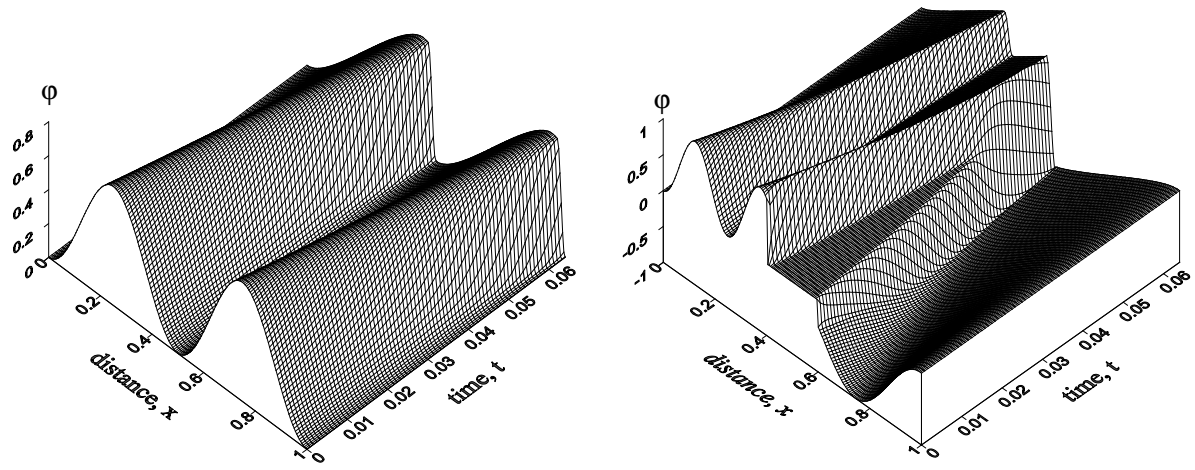

Figure 2.1: - Field evolution for different initial conditions: left - case A, right - case B.

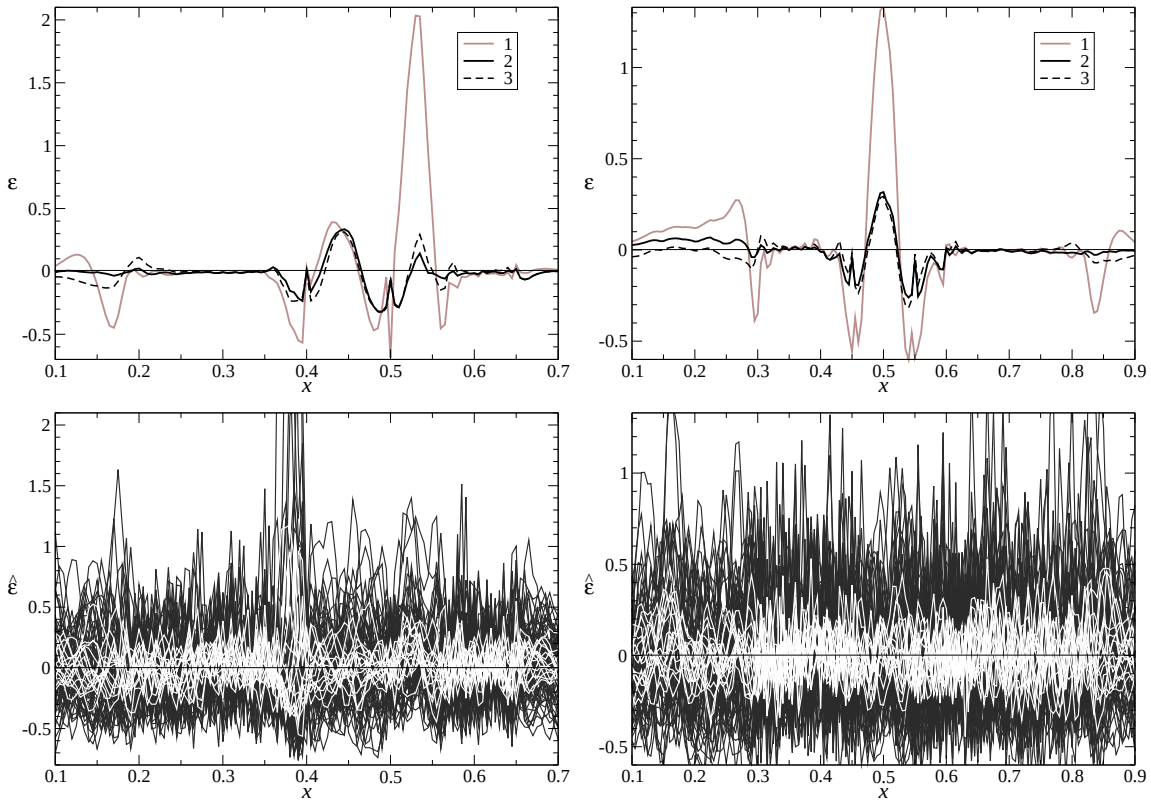

Figure 2.2: Case A. Upper/left: the relative error ε by the inverse Hessian - line 1, and by variants of the EIH method for $L = 25$ - lines 2 and 3. Lower/left: the sample relative error $\hat{\varepsilon}$. Set of $\hat{\varepsilon}$ for $L' = 25$ - dark envelope and set of $\hat{\varepsilon}$ for $L' = 100$ - white envelope. Case B: right panels.

2.2 Estimation error covariance versus Bayesian posterior covariance

2.2.1 General theory

Here, we present the main results from [A9], which point out the difference between the classical estimation (analysis) error covariance and the Bayesian posterior covariances, both in theoretical and computational aspects.

The estimation error covariance is associated with trying to find an approximation around the truth $\bar{U}$, whereas the data is also assumed to come from the truth: $Y^* = G(\bar{U}) + \xi$, $U^* = \bar{U} + \varepsilon$, where $\xi \sim \mathcal{N}(0, R)$ and $\varepsilon \sim N(0, B)$ are the observation and background error, respectively. The estimation error is defined as $\delta U = U - \bar{U}$ and its covariance is given by

$$P = E_a[(U - \bar{U})(U - \bar{U})^T] = E_a[\delta U \delta U^T],$$

where E_a denotes averaging (expectation) with respect to the probability density function (pdf) ρ_a which, taking into account the definitions of the data Y^* and U^*, can be defined as follows:

$$\rho_a(U) = const \cdot exp\left(-\frac{1}{2}\|R^{-1/2}(G(U) - G(\bar{U}))\|^2_{\mathcal{Y}} - \frac{1}{2}\|B^{-1/2}(U - \bar{U})\|^2_{\mathcal{U}}\right). \tag{2–24}$$

The Bayesian posterior covariance is given by

$$\mathcal{P} = E_p\left[(U - E_p[U])(U - E_p[U])^T\right], \tag{2–25}$$

where E_p denotes averaging (expectation) with respect to the pdf

$$\rho_p(U) = const \cdot exp\left(-\frac{1}{2}\|R^{-1/2}(G(U) - Y_0^*)\|^2_{\mathcal{Y}} - \frac{1}{2}\|B^{-1/2}(U - U_0^*)\|^2_{\mathcal{U}}\right). \tag{2–26}$$

In the above formula Y_0^* stands for the actual observations made by an instrument. It could be, in principle, seen in the form $Y_0^* = G(\bar{U}) + \xi_0$, where ξ_0 is no longer a random variable, but a particular event (number) from $\mathcal{N}(0, R)$. However, this representation is not used in the Bayesian approach. Similarly, U_0^* stands for the actual background. Then, the estimate $\hat{U}_0$ satisfies the equation

$$(G'_U(\hat{U}_0))^* R^{-1}(G(\hat{U}_0) - Y_0^*) + B^{-1}(\hat{U}_0 - U_0^*) = 0. \tag{2–27}$$

The posterior covariance is often approximated by trying to find the second moment of ρ_p centered around $\hat{U}$ instead of $E_p[U]$, which is natural because $\hat{U}$ is the output of 4D-Var. In the linear Gaussian setup $E_p[U]$ and $\hat{U}$ coincide. This is not true in general, but can be expected to be a good approximation if the volume of data is large and /or noise is small (due to the asymptotic properties of the nonlinear least-square estimator, see [S42], Vol. 1, Ch.6). Most importantly, due to different centering of Gaussian data, the Bayesian posterior covariance and the estimation error covariance are different objects and should not be confused. However, they are equal in the linear case.

In order to simulate the distribution (2–26) we define perturbed data as follows: $Y^* = Y_0^* + \xi$ and $U^* = U_0^* + \varepsilon$. Then, the perturbed estimate $\hat{U}$ satisfies equation

$$(G'_U(\hat{U}))^* R^{-1}(G(\hat{U}) - Y^*) + B^{-1}(\hat{U} - U^*) = 0. \tag{2–28}$$

One way to derive the error equation is to subtract (2–27) from (2–28). For the difference in the residual terms we obtain

$$(G'_U(\hat{U}))^* R^{-1}(G(\hat{U}) - Y^*) - (G'_U(\hat{U}_0))^* R^{-1}(G(\hat{U}_0) - Y_0^*) =$$

$$= \{(G'_U(\hat{U}))^* R^{-1} G(\hat{U}) - (G'_U(\hat{U}_0))^* R^{-1} G(\hat{U}_0)\} - \{(G'_U(\hat{U}))^* R^{-1} Y^* - (G'_U(\hat{U}_0))^* R^{-1} Y_0^*\}$$
$$= \{(G'_U(\hat{U}))^* R^{-1} G(\hat{U}) - (G'_U(\hat{U}))^* R^{-1} G(\hat{U}_0)\}_1$$
$$+\{(G'_U(\hat{U}))^* R^{-1} (G(\hat{U}_0) - Y_0^*) - (G'_U(\hat{U}_0))^* R^{-1} (G(\hat{U}_0) - Y_0^*)\}_2$$
$$-(G'_U(\hat{U}))^* R^{-1} \xi. \tag{2–29}$$

Using the Taylor-Lagrange formula [S34] the expressions in the brackets take the form

$$\{\cdot\}_1 = (G'_U(\hat{U}))^* R^{-1} G'_U(\tilde{U}_1) \delta U,$$
$$\{\cdot\}_2 = [(G'(\tilde{U}_2))^*]'_U \delta U R^{-1} (G(\hat{U}_0) - Y_0^*),$$

where $\tilde{U}_i = \hat{U}_0 + \tau_i(\hat{U} - \hat{U}_0)$, $\tau_i \in [0,1]$, $\quad i = 1,2$. For the difference in the penalty terms we obtain

$$B^{-1}(\hat{U} - U^*) - B^{-1}(\hat{U}_0 - U_0^*) = B^{-1}(\delta U - \varepsilon). \tag{2–30}$$

Combining (2–29) and (2–30) we obtain the error equation as follows:

$$(G'_U(\hat{U}))^* R^{-1} (G'_U(\tilde{U}_1)\delta U - \xi) + [(G'(\tilde{U}_2))^*]'_U \delta U R^{-1} (G(\hat{U}_0) - Y_0^*) + B^{-1}(\delta U - \varepsilon) = 0. \tag{2–31}$$

In the operator form this equation can be rewritten as

$$\mathcal{H}(\hat{U}, \tilde{U}_1, \tilde{U}_2) \cdot \delta U = (G'_U(\hat{U}))^* R^{-1} \xi + B^{-1} \varepsilon, \tag{2–32}$$

where

$$\mathcal{H}(\hat{U}, \tilde{U}_1, \tilde{U}_2) \cdot v = ((G'_U(\hat{U}))^* R^{-1} G'_U(\tilde{U}_1) + B^{-1}) \cdot v + [(G'(\tilde{U}_2))^*]'_U \cdot v \; R^{-1} (G(\hat{U}_0) - Y_0^*). \tag{2–33}$$

Let us note that, generally, $\mathcal{H}$ is neither symmetric, nor positive definite.

The estimation error is expressed from (2–32) as follows:

$$\delta U = \mathcal{H}^{-1}(\hat{U}, \tilde{U}_1, \tilde{U}_2)((G'_U(\hat{U}))^* R^{-1} \xi + B^{-1} \varepsilon). \tag{2–34}$$

Next, we approximate (2–25) by the formula

$$\mathcal{P} = E_p\,[(\delta U \delta U^T]. \tag{2–35}$$

Since operators $\mathcal{H}(\hat{U}, \tilde{U}_1, \tilde{U}_2)$ and $G'_U(\hat{U})$ nonlinearly depend on δU, it is not possible to consider them as constant multipliers when applying the expectation operator. One way is to fix $\hat{U}, \tilde{U}_1, \tilde{U}_2$ at the value $\hat{U}_0$. In this case $\mathcal{H}(\cdot)$ becomes the Hessian of the cost-function (1–6) at the origin point $\hat{U}_0$. Next, by applying (2–35) we obtain

$$\mathcal{P} = \mathcal{H}^{-1}(\hat{U}_0) H(\hat{U}_0) \mathcal{H}^{-1}(\hat{U}_0), \tag{2–36}$$

(compare this equation to (1–24)).

Since the condition number of the double product $\mathcal{H}^{-1} H \mathcal{H}^{-1}$ is usually larger than the condition number of its components $\mathcal{H}$ and H, the above formula could be quite sensitive to the approximation errors. One way is to simplify this formula assuming $\mathcal{H}$ and H are not very different. These simplification is

$$\mathcal{P} = \mathcal{H}^{-1}(\hat{U}_0), \tag{2–37}$$

and, subsequently,

$$\mathcal{P} = H^{-1}(\hat{U}_0). \tag{2–38}$$

Another approach is to use the 'effective inverse Hessian' method presented in §2.1. In this case equation (2–36) becomes

$$\mathcal{P} = \frac{1}{L} \sum_{l=1}^{L} \mathcal{H}^{-1}(\hat{U}_l) H(\hat{U}_l) \mathcal{H}^{-1}(\hat{U}_l). \tag{2–39}$$

2.2.2 Implementation

As before, the computationally efficient and, therefore, feasible implementation is achieved using preconditioning. The first-level preconditioning yields

$$\tilde{H}(\cdot) = (B^{1/2})^* H(\cdot) B^{1/2}, \quad \tilde{\mathcal{H}}(\cdot) = (B^{1/2})^* \mathcal{H}(\cdot) B^{1/2},$$

and the second-level preconditioning yields

$$\tilde{\tilde{H}}(\cdot) = \tilde{H}^{-1/2}(\hat{U}_0)\ \tilde{H}(\cdot)\ \tilde{H}^{-1/2}(\hat{U}_0), \tag{2–40}$$

$$\tilde{\tilde{\mathcal{H}}}(\cdot) = \tilde{H}^{-1/2}(\cdot)\ \tilde{\mathcal{H}}(\cdot)\ \tilde{H}^{-1/2}(\cdot). \tag{2–41}$$

At the first step we compute the leading eigenpairs $(\{\lambda_k^{(0)}, W_k^{(0)}\},\ k = 1, \ldots, K_0)$ of $\tilde{H}(\hat{U}_0)$. This allows $\tilde{H}^{-1/2}(\hat{U}_0)$ to be defined in the limited-memory form. Then, inside the ensemble loop (index l) we compute:

a) the leading eigenpairs $(\{\lambda_k^{(l)}, W_k^{(l)}\},\ k = 1, \ldots, K_l)$ of $\tilde{\tilde{H}}(\hat{U}_l)$, which allows us to define $\tilde{H}^{-1/2}(\hat{U}_l)$ in (2–41);

b) the leading eigenpairs $(\{\tilde{\lambda}_k^{(l)}, \tilde{W}_k^{(l)}\},\ k = 1, \ldots, \tilde{K}_l)$ of $\tilde{\tilde{\mathcal{H}}}(\hat{U}_l)$.

For a reasonably small $\delta U_l = \hat{U}_l - \hat{U}_0$, the number K_l of eigenpairs describing the Riemann distance between $\tilde{H}(\hat{U}_l)$ and $\tilde{H}(\hat{U}_0)$ has to be small as compared to K_0. Similarly, $\tilde{K}_l$ has to be small as compared to K_l because the difference between $\tilde{\mathcal{H}}(\hat{U}_l)$ and $\tilde{H}(\hat{U}_l)$ is only due to the presence of the second-order term. Thus, the expenses of computing the eigenpairs $\{\lambda_k^{(l)}, W_k^{(l)}\}$ and $\{\tilde{\lambda}_k^{(l)}, \tilde{W}_k^{(l)}\}$ for given l can make only a fraction of those associated with computing $\{\lambda_k^{(0)}, W_k^{(0)}\}$.

Given all eigenpairs are computed, the posterior covariance in (2–39) is evaluated as follows:

$$\mathcal{P} = B^{1/2} \tilde{H}^{-1/2}(\hat{U}_0)\ A(\hat{U}_l)\ \tilde{H}^{-1/2}(\hat{U}_0)(B^{1/2})^*, \tag{2–42}$$

where

$$A(\hat{U}_l) = \frac{1}{L} \sum_{l=1}^{L} \tilde{\tilde{H}}^{-1/2}(\hat{U}_l) \tilde{\tilde{\mathcal{H}}}^{-\alpha}(\hat{U}_l) \tilde{\tilde{H}}^{-1/2}(\hat{U}_l). \tag{2–43}$$

In (2–43), cases $\alpha = 2$, $\alpha = 1$ and $\alpha = 0$ correspond to approximations (2–36), (2–37) and (2–38), respectively. The quasi-random implementation described in §2.1.2 is applied to substitute $\hat{U}_l$ by $\hat{U}_0 + \delta U_l$.

In the dynamic formulation the Hessian-vector product $\mathcal{H}(\cdot)v$ is defined by the successive solutions of the following problems [S28]:

$$\left\{ \begin{array}{rcl} \dfrac{\partial \psi}{\partial t} - F'(\varphi)\psi & = & 0,\ t \in (0, T), \\ \psi|_{t=0} & = & v, \end{array} \right. \tag{2–44}$$

$$\left\{ \begin{array}{rcl} -\dfrac{\partial \psi^*}{\partial t} - (F'(\varphi))^* \psi^* & = & (F''(\varphi)\psi)^* \varphi^* - C^* V_2 C \psi,\ \ t \in (0, T) \\ \psi^*\big|_{t=T} & = & 0, \end{array} \right. \tag{2–45}$$

$$\mathcal{H}(u)v = V_1 v - \psi^*|_{t=0}. \tag{2–46}$$

Here φ and φ^* are involved, being taken from (1–15)–(1–16). The problem (2–45) is the so-called second-order adjoint problem. One can see that it is a standard adjoint problem with a specially defined source term $(F''(\varphi)\psi)^*\varphi^*$. For certain F, this term is relatively expensive to compute as compared to $(F'(\varphi))^*\psi^*$, which could be a difficulty if the explicit time integration scheme is used. This is not an issue, however, when the implicit or semi-implicit schemes are used.

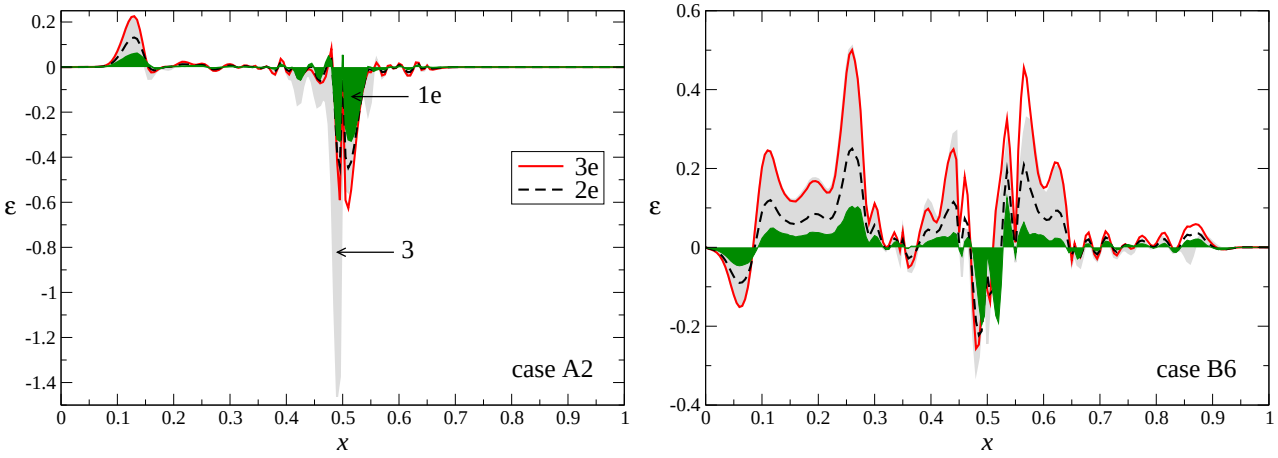

Figure 2.3: The relative errors ε^1, ε^2 and ε^3 in logarithmic scale.

2.2.3 Illustration

As in §2.1, the numerical tests have been performed for Burgers' equation (2–23), involving the same initial conditions (see Fig.2.1) and observation schemes, cases A and B. The effect of using the estimates of $\mathcal{P}$ is demonstrated in Fig.2.3 in terms of the relative error

$$\varepsilon_i^{(n)} = \mathcal{P}_{i,i}^{(n)} / \hat{\mathcal{P}}_{i,i} - 1, \quad i = 1, ..., M,$$

where $\hat{\mathcal{P}}$ is the reference Bayesian covariance (sample Bayesian covariance matrix obtained for a large sample size ($L = 2500$), and $\mathcal{P}^{(n)}$ are the Bayesian covariance approximations obtained by (2–42),2–43). In particular, line $1e$ corresponds to $\alpha = 2$ in (2–43), i.e. to the double-product formula, which is promoted as a correct one; lines $2e$ and $3e$ correspond to $\alpha = 1$ and $\alpha = 0$, which represent the simplifications of the double-product formula. Let us note that the latter is simply the effective inverse Hessian, as presented in §2.1.

2.3 Future developments

It has been shown that the 'effective' approximations of the covariance are noticeably more accurate that the point approximations (1–23) or (2–36). Moreover, the quality of the 'effective' approximations can be further improved.

Let us consider the error equation (2–32)-(2–33). This equation describes the exact relationship between the input errors (ξ, ε) and δU, given the true values of $\tilde{U}_1$ and $\tilde{U}_2$. In reality these values are unknown. It is easy to show, however, that using $\tilde{U}_1 = \tilde{U}_2 = \tilde{U} = (\hat{U} + \hat{U}_0)/2$ guaranties the second-order approximation accuracy of (2–32)-(2–33) with respect to δU. Note that considering $\tilde{U}_1 = \tilde{U}_2 = \hat{U}$ guaranties only the first-order approximation accuracy.

The implementation of this idea is not straightforward, however. The problem is that the resulting $\mathcal{H}(\hat{U}, \tilde{U})$ is not symmetric, which implies the need of computing the truncated singular value decomposition, rather than the eigenvalue decomposition. Thus, it is desirable to derive an error equation which guaranties the second-order approximation accuracy with respect to δU, but such that its $\mathcal{H}$ has only one origin point.

Let us again consider the estimator equation (2–28)

$$J'_U(\hat{U}) = (G'_U(\hat{U}))^* R^{-1}(G(\hat{U}) - Y^*) + B^{-1}(\hat{U} - U^*) = 0. \qquad (2–47)$$

This equation is trivially equivalent to the following:

$$J'_U(\hat{U}) - J'_U(\hat{U}_0) = -J'_U(\hat{U}_0). \tag{2–48}$$

By applying the Taylor-Lagrange formula to (2–48) we obtain

$$(J'_U(\tilde{U}))'_U \cdot \delta U = \mathcal{H}(\tilde{U}) \cdot \delta U = -J'_U(\hat{U}_0). \tag{2–49}$$

Taking into account the expression for $J'_U(\hat{U})$, we can write

$$(J'_U(\tilde{U}))'_U \cdot \delta U = ((G'_U(\tilde{U}))^*)'_U \cdot \delta U \; R^{-1}(G(\tilde{U}) - Y^*) + ((G'_U(\tilde{U}))^* R^{-1} G'_U(\tilde{U}) + B^{-1}) \cdot \delta U. \tag{2–50}$$

For the right-hand side of (2–49) we can also write

$$J'_U(\hat{U}_0) = (G'_U(\hat{U}_0))^* R^{-1}(G(\hat{U}_0) - Y^*) + B^{-1}(\hat{U}_0 - U^*) =$$

$$(G'_U(\hat{U}_0))^* R^{-1}(G(\hat{U}_0) - (Y_0^* + \xi)) + B^{-1}(\hat{U}_0 - (U_0^* + \varepsilon)).$$

Since (2–27) holds, we obtain

$$-J'_U(\hat{U}_0) = (G'_U(\hat{U}_0))^* R^{-1}\xi + B^{-1}\varepsilon. \tag{2–51}$$

Finally, the sought error equation is in the form

$$\mathcal{H}(\tilde{U}) \cdot \delta U = (G'_U(\hat{U}_0))^* R^{-1}\xi + B^{-1}\varepsilon. \tag{2–52}$$

As before, one may express δU from (2–52) and compute $\mathcal{P} = E[\delta U \delta U^T]$ by applying the expectation operator as described in §2.1.1, in which case the posterior covariance is given by expression

$$\mathcal{P} = \frac{1}{L}\sum_{l=1}^{L} \mathcal{H}^{-1}(\tilde{U}_l) H(\hat{U}_0) \mathcal{H}^{-1}(\tilde{U}_l), \quad \tilde{U}_l = \frac{\hat{U}_l + \hat{U}_0}{2}, \tag{2–53}$$

where $\hat{U}_l$, $l = 1, \ldots, L$ are the optimal solutions. Let us note that the quasi-random approach presented in §2.1.2 cannot be used in this formulation because $\mathcal{H}(\cdot)$ contains the second-order term which involves $Y^* = Y_0^* + \xi_l$. Therefore, each $\hat{U}_l$ must be consistent with a given ξ_l, i.e. the former cannot be considered as an independent event from $\mathcal{N}(0, \mathcal{P})$. Unfortunately, using Y_0^* instead of Y^* would mean that the second-order approximation accuracy is no longer asserted. One way to deal with this issue is to move the term $(G'_U(\tilde{U}))^*)'_U \cdot \delta U \; R^{-1}\xi$ to the right-hand side of the error equation where, after applying the expectation operator, ξ is integrated out.

The suggested estimate of $\mathcal{P}$ is expected to be more accurate than the one given in (2–39). However, the bottleneck of the 'effective inverse Hessian' approach is a special use of the expectation operator, which is allowed to be translated through nonlinear operators. From this point of view one may prefer another scheme. First, we notice that if equation (2–52) is satisfied, than the following also holds

$$\mathcal{H}(\tilde{U}_l) \; \delta U \delta U^T \; \mathcal{H}(\tilde{U}_l) = (G'_U(\hat{U}_0))^* R^{-1} \; \xi\xi^T \; R^{-1} G'_U(\hat{U}_0) + B^{-1} \; \varepsilon\varepsilon^T \; B^{-1}. \tag{2–54}$$

Next, the expectation operator is applied directly to equation (2–54) as described in §2.1.1, resulting into

$$\frac{1}{L}\sum_{l=1}^{L} \mathcal{H}(\tilde{U}_l) \; \mathcal{P} \; \mathcal{H}(\tilde{U}_l) = H(\hat{U}_0). \tag{2–55}$$

The motivation to apply such scheme is that $\mathcal{H}(\cdot)$ is less nonlinear than $\mathcal{H}^{-1}(\cdot)$ and, therefore, translating the expectation operator through $\mathcal{H}(\cdot)$ could result into less error.

On the other hand, solving the linear matrix equation (2–54) for $\mathcal{P}$ is not easy. Since $H(\hat{U}_0)$ has to be computed once, one can certainly use preconditioning by $H^{-1/2}(\hat{U}_0)$, then (2–55) becomes

$$\frac{1}{L}\sum_{l=1}^{L}\bar{\mathcal{H}}(\tilde{U}_l)\;\bar{\mathcal{P}}\;\bar{\mathcal{H}}(\tilde{U}_l) = I, \tag{2–56}$$

where $\bar{\mathcal{H}}(\tilde{U}_l) = H^{-1/2}(\hat{U}_0)\;\mathcal{H}(\tilde{U}_l)\;H^{-1/2}(\hat{U}_0)$, and $\mathcal{P} = H^{-1/2}(\hat{U}_0)\;\bar{\mathcal{P}}\;H^{-1/2}(\hat{U}_0)$. Next, one could benefit from the special structure of $\bar{\mathcal{H}}(\cdot)$ (the limited-memory approximation of the type (2–15)). Moreover, $\bar{\mathcal{P}}$ itself can be directly sought in the same limited-memory form by an iterative minimization algorithm. All mentioned issues constitute a subject for an immediate future research.

2.4 Summary

The main issue considered in this Chapter is the relationship between the Hessian of the cost-function and the estimation error covariance matrix in variational DA, whereas the discussion is focused on the essentially nonlinear case. In the classical (frequentist) statistical approach the estimation error δU is considered as a difference between the estimate $\hat{U}$ and the 'truth' $\bar{U}$. The corresponding covariance matrix can be approximated by the inverse of the Hessian H of an auxiliary cost-function computed at the 'truth' point. If the 'truth' is known (the identical twin experiment framework only), this is the first-order approximation with respect to δU. In practice, the Hessian is computed at the estimate $\hat{U}_0$, which is obtained after assimilating the real data. The difference between $\hat{U}_0$ and $\bar{U}$ yields an additional error in the covariance (the origin error), which can not be removed in principle. These are more or less well-known facts in variational DA. However, evaluating the inverse Hessian using the LBFGS algorithm with exact step search has been a novel element.

In the Bayesian approach the estimation error δU is considered as a difference between the estimates $\hat{U}$ and $\hat{U}_0$, conditioned on a perturbed and unperturbed data, respectively. Due to different centering of data, the first-order (with respect to δU) approximations of the estimation (analysis) error covariance and the Bayesian posterior covariance are different and should not be confused. The latter is computed via the Hessian-product formula (2–36), which involves both the Hessian of the original cost-function $\mathcal{H}$, and the Hessian of the auxiliary cost-function H (often called the Jacobian). In the dynamic formulation the Hessian-vector product $\mathcal{H}(\cdot)v$ is defined by the second-order adjoint problem. The formulas of the type (2–36) occasionally appear in the literature on statistics and nonlinear regression in the general context of nonlinear least-squares (see e.g. [S15]).

Introducing the 'effective' Hessian-based covariance approximations (2–8) and (2–39), as well as the relatively feasible methods for their evaluation, is the main author's contribution in terms of novelty. The concept of such approximations stems from the exact error equations (1–18) and (2–31), the idea of which, in turn, has been first presented in [S29]. In practice, evaluating (2–8) or (2–39) requires an ensemble of the inverse Hessians $H^{-1}(\hat{U}_l)$, or the products $\mathcal{H}^{-1}(\hat{U}_l)H(\hat{U}_l)\mathcal{H}^{-1}(\hat{U}_l)$ to be computed, however the size of such ensemble can be small enough ($L = 25 - 50$). The suggested method can be regarded as an alternative to the standard sampling-based covariance evaluation method (involving localization and other possible tricks). The later requires much large ensemble size to achieve the similar approximation quality. The ways of improving the accuracy of the 'effective' Hessian-based covariance approximations are presented in §2.3.

Chapter 3

Non-trivial applications of the Hessian for UQ and design

3.1 On gauss-verifiability of optimal solutions

3.1.1 Introduction

Variational Data Assimilation (DA) is a deterministic approach based on the optimal control theory, suitable for high-dimensional large-scale models arising in geophysical applications. The cost function in variational DA includes the background term (the regularization term), the presence of which usually guarantees that all of three Hadamard conditions are formally met. From this fact, however, very little can be concluded on how close to the truth the optimal solution actually is. That is why variational DA is often considered in a probabilistic context (including the Bayesian context, see e.g. [S41, S43]), where the confidence region for the optimal solution can be constructed on a basis of the estimation error covariance. This implies that the estimation error pdf is reasonably close to the gaussian. We shall say that the optimal solution is *gauss-verifiable* if a sensible covariance-based confidence region for the optimal solution error can be actually constructed. This is the main issue considered in [A8].

The fundamental difficulty here is related to the nonlinearity of the model equations and the observation operator. The nonlinear least squares estimator is *asymptotically normal* [S42], however for a finite number of observations this is not the case. In one hand, the nonlinearity may distort its gaussian properties to the extent when the covariance becomes no longer sufficient for constructing the confidence region (even for the gaussian data errors). On the other hand, this distortion may be localized in certain spatial areas (since it is related to the nonlinearity), whereas outside of these areas the estimator holds its gaussian properties. Assuming the optimal solution is gauss-verifiable, evaluating the estimation error covariance is not an easy task in practical terms. The major difficulty can be attributed to the high-dimensionality of the state vector combined with the complexity of the governing equations. This may result into unaffordable computational costs of a single optimal solution, whereas for computing the sample covariance one needs an ensemble of optimal solutions.

Formally speaking, gauss-verifiability is tied to the approximate gaussianity (normality) of the estimator. An immediate idea would be to use classical test statistics for multivariate normality [S21]. However, there are a few points why this may not be the best option. First, the classical test statistics do not measure the gauss-verifiability (as we understand it) directly, so it is difficult to make a sensible interpretation of results. Secondly, these statistics measure local properties of the nonlinear estimator, which strongly depend on the point of evaluation, whereas we would rather prefer to know its global properties. Thirdly, the sample (ensemble) on a basis of which these statistics are calculated is likely

to be extremely small as compared to the size of the control vector. At the same time almost any invariant test statistic is a function of the Mahalanobis distances and angles, which involve the inverse square root of the sample covariance matrix. The difficulty of evaluating the inverse square root of a matrix of a deficient rank is well known.

The aim of this section is to describe a tool for checking the gauss-verifiability, both total and partial (local). We consider the estimation error pdf defined on the "true" state and its approximation defined on the optimal solution. The basic idea is to quantify our ability to recognize the truth among statistically significant events associated to the analysis pdf, defined by the analysis (the mode), and by the analysis error covariance. First, we introduce the *coexistence principle*. Then, in order to quantify its violation (further referred as *coexistence breach*) we define the *coexistence measure* (CM). The decomposition of the CM into the sum of components, each being associated to the corresponding element of the state vector, is introduced. The distribution of these components in space shows the subsets of the state vector for which the gaussian confidence regions cannot be properly defined. Numerical experiments for the 1D Burgers equation illustrate the developed theory and demonstrate the usefulness of the suggested measure and, especially, of its element-wise decomposition.

3.1.2 Coexistence measure

Let us consider again the cost-function (1–6). In order to underline its dependence on data entries we re-write it as follows

$$J(U, U^*, Y^*) = \frac{1}{2}\|R^{-1/2}(G(U) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(U - U^*)\|^2_{\mathcal{U}}. \tag{3–1}$$

Associated with (3–1) is the estimator (1–11). As discussed in §2.2.1, if data comes from the "truth", i.e. $Y^* = G(\bar{U}) + \xi$ and $U^* = \bar{U} + \varepsilon$, then the estimate U has the pdf

$$\rho_a(U, \bar{U}) = const \cdot \exp\left(-\frac{1}{2}\|R^{-1/2}(G(U) - G(\bar{U}))\|^2_{\mathcal{Y}} - \frac{1}{2}\|B^{-1/2}(U - \bar{U})\|^2_{\mathcal{U}}\right). \tag{3–2}$$

Let us consider an independent variable $w \geq 0$. For each w, the solution $\Gamma(w) \in \mathcal{U}$ to the equation

$$J(\Gamma(w), \bar{U}, G(\bar{U})) = w$$

represents a manifold (locus) of equal likelihood $c_1(\bar{U})exp(-w)$ in the control space $\mathcal{U}$. This manifold bounds the domain $\Omega(w)$ where the cumulative density function $\beta(w)$ is defined as follows:

$$\beta(w) = \int_{\Omega(w)} \rho_a(U, \bar{U})dU.$$

This function shows the probability that an event $U \sim \rho(U, \bar{U})$ falls 'inside' the domain $\Omega(w)$. Let us note that the manifold may consist of a few disconnected sub-surfaces, and the domain - of a few disconnected sub-domains. For a given confidence level γ the corresponding value of w^* satisfying the equation

$$\beta(w^*) = \gamma, \tag{3–3}$$

defines the confidence region $\Omega(w^*)$. All events U falling 'outside' $\Omega(w^*)$ are considered as 'unlikely' or 'statistically insignificant' events to be discarded. It is worth mentioning that for nonlinear G, the confidence region can be topologically very complex, therefore the notions of 'inside' and 'outside' cannot be trivial.

Let us note that $\beta(0) = 0$, $\beta(\infty) = 1$ and $\beta(w)$ is a monotonic increasing function of w. Therefore, $\beta(w) < \gamma$ when $w < w^*$, and the criteria to test whether or not U falls into the confidence region $\Omega(w^*)$ reads

$$J(U, \bar{U}, G(\bar{U})) < w^*. \tag{3–4}$$

A particular optimal solution $\hat{U}_0$ satisfies the estimator equation

$$(G'_U(\hat{U}_0))^* R^{-1}(G(\hat{U}_0) - Y^*) + B^{-1}(\hat{U}_0 - U^*) = 0. \tag{3–5}$$

where $Y^* = \bar{Y} + \xi_0$ and $U^* = \bar{U} + \varepsilon_0$ are the actually observed data defined by the error events ξ_0 and ε_0 which have actually come to pass (see §2.2.1). In this section we use $\hat{U}$ instead of $\hat{U}_0$ to simplify notations.

Given $\hat{U}$ as the best available approximation of $\bar{U}$, the original pdf (3–2) takes the form

$$\rho_a(U, \hat{U}) = c_2(\hat{U}) \cdot \exp\left(-\frac{1}{2}\|R^{-1/2}(G(U) - G(\hat{U}))\|_{\mathcal{Y}} - \frac{1}{2}\|B^{-1/2}(U - \hat{U})\|_{\mathcal{U}}\right) =$$

$$= c_2(\hat{U}) \cdot \exp\left(-J(U, \hat{U}, G(\hat{U})\right),\ c_2(\hat{U}) = const > 0. \tag{3–6}$$

Let us denote $\bullet|_{\bar{U}}$ an "object associated to $\rho_a(U, \bar{U})$", and $\bullet|_{\hat{U}}$ an "object associated to $\rho_a(U, \hat{u})$". We shall say that $\hat{U}$ and $\bar{U}$ coexist if, simultaneously, $\hat{U}$ is a statistically significant event in the distribution $\rho_a(U, \bar{U})$, i.e. $\hat{U} \in \Omega(w^*)|_{\bar{U}}$, and $\bar{U}$ is a statistically significant event in the distribution $\rho_a(U, \hat{U})$, i.e. $\bar{U} \in \Omega(w^*)|_{\hat{U}}$. Since both conditions have probability γ and they are assumed to be statistically independent, the "coexistence" is a random event with probability γ^2 and has to be quantified as a random variable. In terms of the testing criteria (3–4) this reads as follows:

$$Pr[J(\hat{U}, \bar{U}, G(\bar{U})) < w^*|_{\bar{U}},\ J(\bar{U}, \hat{U}, G(\hat{U})) < w^*|_{\hat{U}}] = \gamma^2. \tag{3–7}$$

The *coexistence principle* simply means that, with a given probability, the truth falls within the confidence region defined for the pdf $\rho_a(U, \hat{U})$. The *coexistence breach* occurs when the original non-gaussian pdf is approximated by the gaussian pdf. In this case we shall say that the estimate is not *gauss-verifiable* on the whole.

In the finite-dimensional case ($\mathcal{U} = \mathbf{R^n}$) the gaussian approximation of $\rho_a(U, \hat{U})$ is given by

$$\tilde{\rho}_a(U, \hat{U}) = c(\hat{U}) exp(-\frac{1}{2}\|P^{-1/2}(\hat{U})(U - \hat{U})\|^2_{\mathcal{U}}) \equiv c(\hat{U}) exp(-\tilde{J}(U, \hat{U})), \tag{3–8}$$

where $c(\hat{U}) = (2\pi)^{-n/2}(detP(\hat{U}))^{-1/2}$, n is the dimension of the state space $\mathcal{U}$, $P(\hat{U})$ is the covariance computed from the pdf (3–6), and the function $\hat{U}$ is the *origin*. For $\tilde{\rho}_a(U, \hat{U})$ the cumulative density function $\beta(w)$ is known to be the χ^2_n-cumulative density function $F(w, n)$, and $w^*|_{\hat{U}} = \chi^2_n(\gamma)$ is the critical point of χ^2_n distribution. Taking into account the definition of J in (3–1) we notice that

$$J(\bar{U}, \hat{U}, G(\hat{U})) = J(\hat{U}, \bar{U}, G(\bar{U})). \tag{3–9}$$

This is an important condition for the coexistence principle to hold. However, as a result of using the gaussian $\tilde{\rho}(U, \hat{U})$ instead of $\rho(U, \hat{U})$ this condition may no longer be valid. In addition, the critical point $w^*|_{\hat{U}}$ becomes $\chi^2_n(\gamma)$. Therefore, the coexistence is affected by the two differences:

$$\tilde{J}(\bar{U}, \hat{U}) - J(\hat{U}, \bar{U}, G(\bar{U})) \tag{3–10}$$

and

$$\chi^2_n(\gamma) - w^*|_{\bar{U}}.$$

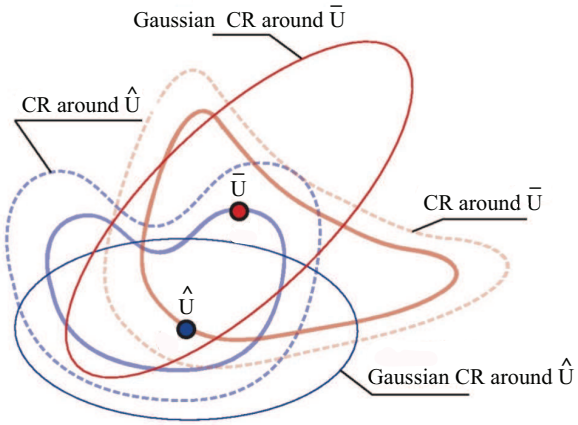

Figure 3.1: Confidence regions.

Let us note that $w^*|_{\bar{U}}$ is an integral quantity of the corresponding pdf and may differ from $\chi^2_n(\gamma)$ not too significantly (unless the specified confidence level γ is too close to 1). In this case the difference (3–10) can be considered as a major cause of the coexistence breach. Therefore, the only condition for our approach to be valid is as follows:

$$|\tilde{J}(\bar{U},\hat{U}) - J(\hat{U},\bar{U},G(\bar{U}))| \gg |\chi^2_n(\gamma) - w^*|_{\bar{U}}|. \tag{3–11}$$

Because the truth is not known, instead of (3–10) we can consider the difference

$$\tilde{J}(\hat{U},\bar{U}) - J(\bar{U},\hat{U},G(\hat{U})) = \frac{1}{2}\|P^{-1/2}(\bar{U})(\bar{U}-\hat{U})\|^2_X - J(\bar{U},\hat{U},G(\hat{U})). \tag{3–12}$$

In the above formulation the truth $\bar{U}$ becomes a random variable which falls inside a neighborhood of $\hat{U}$ consistent with $\rho_a(U,\hat{U})$, i.e. $\bar{U} = \hat{U} + v$, where $v \in \mathcal{U}$ is the estimation error. Thus, we are interested in evaluating the difference

$$\theta(v,\hat{U}) = \frac{1}{2}\|P^{-1/2}(\hat{U}+v)v\|^2_{\mathcal{U}} - J(\hat{U}+v,\hat{U},G(\hat{U})) \tag{3–13}$$

averaged over $\rho_a(U,\hat{U})$:

$$E[\theta(v,\hat{U})] = \int \theta(v,\hat{U})\rho_a(\hat{U}+v,\hat{U})dv. \tag{3–14}$$

We shall call $E[\theta(v,\hat{U})]$ the *coexistence measure*. It can be used to quantify gauss-verifiability of optimal solutions. From another perspective, it is also a global measure of deviation of $\rho_a(U,\hat{U})$ from normality.

The above idea is illustrated in Fig.3.1 for the case of two random variables. Here we present the exact confidence regions associated to $\bar{U}$ and $\hat{U}$ in dashed lines, the gaussian confidence regions - in solid thin lines and the contours of equal likelihood - in solid thick lines. Due to the nonlinearity of operator $G(U)$, the cost-function J is not quadratic and the pdf (3–2) and (3–6) are not gaussian, therefore the shape of the confidence regions differs from ellipsoidal. However, we can see, that $\hat{U}$ belongs to the certain likelihood locus associated to $\bar{U}$, whereas $\bar{U}$ belongs to the equivalent locus associated to $\hat{U}$. However, $\bar{U}$ falls outside the Gaussian confidence region associated to $\hat{U}$.

3.1.3 Coexistence measure deconvolution

It is important to present $E[\theta(v,\hat{U})]$ as a sum of contributions associated to perturbations v_i in the elements of the control vector U. Those could be obtained as an outcome of a global sensitivity analysis applied to $E[\theta(v,\hat{U})]$, but such analysis is hardly feasible for the high-dimensional models. One possible way to achieve the mentioned deconvolution is to consider the relationship

$$E\left[\|P^{-1/2}(\hat{U}+v)v\|_{\mathcal{U}}^2\right] = tr\left\{E[P^{-1}(\hat{U}+v)vv^T]\right\}.$$

There is no guarantee, however, that the elements of the trace are non-negative values, so we consider a modification of $E[\theta(v,\hat{U})]$ allowing us to mitigate this difficulty.

Since $E[P^{-1}(\hat{U}+v)vv^T]$ is the integral with respect to v, under the conditions of the mean value theorem, there exist v_0 such that

$$E[P^{-1}(\hat{U}+v)vv^T] = P^{-1}(\hat{U}+v_0)E[vv^T] = P^{-1}(\hat{U}+v_0)P(\hat{U}), \tag{3–15}$$

where $P(\hat{U})$ is the covariance computed from the pdf (3–6). Since v_0 is not known, instead of $P^{-1}(\hat{U}+v_0)$ we consider in (3–15) its expectation $E[P^{-1}(\hat{U}+v)]$, then $E[P^{-1}(\hat{U}+v)]P(\hat{U}) = E[P^{-1}(\hat{U}+v)P(\hat{U})]$, and instead of $E[\theta(v,\hat{U})]$ we introduce

$$\mathcal{D} = \frac{1}{2}tr\{E[P^{-1}(\hat{U}+v)P(\hat{U})]\} - C_1, \tag{3–16}$$

where

$$C_1 = E[J(\hat{U}+v,\hat{U},G(\hat{U}))]. \tag{3–17}$$

Consider the square-root decomposition $P(\hat{U}) = QQ^T, \quad Q:\mathcal{U}\to\mathcal{U}$. Then

$$tr\{P^{-1}(\hat{U}+v)P(\hat{U})\} = tr\{P^{-1}(\hat{U}+v)QQ^T\} = tr\{Q^TP^{-1}(\hat{U}+v)Q\}$$

and

$$\mathcal{D} = tr\left\{\frac{1}{2}E[Q^TP^{-1}(\hat{U}+v)Q] - \frac{1}{n}C_1I_n\right\}. \tag{3–18}$$

The last formula implies

$$\mathcal{D} = \sum_{i=1}^{n} d_i, \tag{3–19}$$

where

$$d_i = \frac{1}{2}(Ae_i,e_i)_{\mathcal{U}} - \frac{C_1}{n}, \quad A = E[Q^TP^{-1}(\hat{U}+v)Q]. \tag{3–20}$$

The operator A acts from $\mathcal{U}$ to $\mathcal{U}$, therefore formula (3–19) evaluates an individual contribution of each state variable into the integral value $\mathcal{D}$.

Remark 1. Since $P^{-1}(\hat{U}+v)$ is positive definite for each v, it is easily seen that $(Ae_i,e_i)_{\mathcal{U}} \geq 0$, because

$$(Q^TP^{-1}(\hat{U}+v)Qe_i,e_i)_{\mathcal{U}} = (P^{-1}(\hat{U}+v)Qe_i,Qe_i)_{\mathcal{U}} \geq \epsilon\|Qe_i\|^2 \geq 0, \quad \epsilon = const > 0.$$

Let us note that $d_i = (Ae_i,e_i)_{\mathcal{U}}/2 - C_1/n$ may not always be positive in theory. In practice, the coexistence breach mainly occurs when $(Ae_i,e_i)_{\mathcal{U}}/2 \gg C_1/n$.

3.1.4 Implementation details

Further simplifications can be applied to some elements of (3–20). First, one can see that

$$C_1 = E[J(\hat{U}+v, \hat{U}, G(\hat{U}))] \approx \frac{n}{2}. \tag{3–21}$$

Secondly, for practical computations in (3–18) one needs to define the inverse (or pseudo-inverse) of the covariance $P(\hat{U}+v)$ for each integration point v. Computing the invertible $P(\hat{U}+v)$ by the Monte Carlo method requires an ensemble of estimates of a size greater then n, which would be an enormous computational task for large n. However, the covariance $P(\cdot)$ can be approximated by the inverse Hessian, see (1–24), which results into

$$P^{-1}(\hat{U}+v) \approx H(\hat{U}+v). \tag{3–22}$$

The Hessian in (3–22) is defined by (1–21) or, in the dynamic formulation, by the successive solution of the tangent linear and adjoint models (1–26)-(1–28). Using the approximations (3–21) and (3–22), the coexistence measure defined by formula (3–18) can be represented as follows:

$$\mathcal{D} \approx D = tr\left\{\frac{1}{2}E[Q^T H(\hat{U}+v)Q] - \frac{I_n}{2}\right\}. \tag{3–23}$$

In this case formula (3–19) holds with elements

$$d_i = \frac{1}{2}(Ae_i, e_i)_{\mathcal{U}} - \frac{1}{2}, \;\; A = E[Q^T H(\hat{U}+v)Q] \tag{3–24}$$

Remark 2. It has been suggested that the coexistence has been breached locally if

$$d_i > d^* = \frac{\alpha}{\sqrt{2n}}, \tag{3–25}$$

and globally, if

$$D > D^* = \alpha\sqrt{n/2}, \tag{3–26}$$

where $\alpha > 0$ is a real number associated to the confidence level γ. For example, $\alpha = 2$ approximately corresponds to $\chi^2_n(0.05)$, and $\alpha = 3$ - to $\chi^2_n(0.001)$.

Since the factor Q does not depend on v, it must be computed once (outside the expectation operator). It can be either the sample covariance-based or the Hessian-based. In the latter case

$$Q = H^{-1/2}(\hat{U}),$$

and, subsequently

$$A = E[H^{-1/2}(\hat{U})H(\hat{U}+v)H^{-1/2}(\hat{U})]. \tag{3–27}$$

Taking into account the formulas for preconditioning (2–10) and (2–11) (§2.1.2), one can easily see that

$$A = E[\tilde{H}^{-1/2}(\hat{U})\tilde{H}(\hat{U}+v)\tilde{H}^{-1/2}(\hat{U})] = E[\tilde{\tilde{H}}(\hat{U}+v)], \tag{3–28}$$

where $\tilde{H}^{-1/2}(\hat{U})$ is defined by (2–15) using the leading eigenpairs $(\{\lambda_k^{(0)}, W_k^{(0)}\},\ k = 1, \ldots, K_0)$ of $\tilde{H}(\hat{U})$ or $\tilde{H}^{-1}(\hat{U})$. In practice, the expectation operator is, of course, substituted by the mean

$$A = \frac{1}{L}\sum_{l=1}^{L} \tilde{\tilde{H}}(\hat{U}+\delta U_l), \tag{3–29}$$

where $\delta U_l,\ l = 1, \dots L$ could be either the estimation error itself, i.e. $\delta U_l = \hat{U}_l - \bar{U}$, or an arbitrary vector having the statistical properties of the estimation error (see description the quasi-random approach in §2.1.2).

An efficient way of computing A is by performing the eigenvalue analysis of the matrix $\tilde{\tilde{H}}(\hat{U} + v)$. Given $\{\lambda_k^{(l)}, W_k^{(l)}\},\ k = 1, \dots, K_l$ are the eigenpairs of $\tilde{\tilde{H}}(\hat{U} + \delta U_l)$, the expression for D and its components d_i finally takes the form:

$$D = \sum_{i=1}^{n} d_i, \quad d_i = \frac{1}{2L} \sum_{l=1}^{L} \left[\left(\sum_{k=1}^{K_l} \lambda_k^{(l)} - 1 \right) W_{k,i}^{(l)} W_{k,i}^{(l)} \right]. \tag{3–30}$$

Remark 3. The coexistence measure deconvolution in the control space is achieved if operator $Q = H^{-1/2} = B^{1/2} \tilde{H}^{-1/2}$ is a mapping from $\mathcal{U}$ into $\mathcal{U}$. This condition is satisfied if $B^{1/2} : \mathcal{U} \to \mathcal{U}$, which implies $B^{1/2}$ must be a symmetric operator. For example, given the singular value decomposition (SVD) of the prior (background) covariance in the form $B = USU^T$, the appropriate square root is $B^{1/2} = US^{1/2}U^T$, however the Choleski factors of B are not suitable (due to asymmetry). Unfortunately, the former requires the full-rank SVD of B to be computed, which may not be feasible in high-dimensional problems. Thus, instead of defining B or B^{-1} first, then factorizing it, one should directly define a symmetric $B^{1/2}$. For the spatially distributed variables, this can be done by using digital filters or the diffusion equation [S51].

Remark 4. Let us consider formula (2–14) for the covariance approximation P in the 'effective' inverse Hessian method. The expression in round brackets inside this formula is exactly A given by (3–29), i.e. the ensemble average of

$$\tilde{\tilde{H}}^{-1}(\hat{U} + \delta U_l),$$

each being defined by its eigenpairs $\{\lambda_k^{(l)}, W_k^{(l)}\},\ k = 1, \dots, K_l$. This means that the coexistence measure approximation D and its components d_i can be computed at a negligible computational cost in the course of computing the 'effective' inverse Hessian. Therefore, in addition to getting a much better approximation of the covariance P (as compared to $P = H^{-1}(\hat{U})$), one can assess the deviation from gaussianity and, most importantly, its spatial distribution. As numerical experiments have shown, the non-gaussian effects may be localized in certain spatial areas.

3.1.5 Illustration

Here we present some numerical results which illustrate the developed theory. As a dynamical model we again use the Burger's equation (2–23). The field evolution for two different initial states is shown in Fig.3.2. The upper panel in Fig.3.3 shows these initial states, 3σ-envelope (positive margin) for the background and for the estimates ($\bullet|_{H^{-1}}$ - Hessian-based and $\bullet|_{\hat{V}}$ - sample covariance-based). The observation array consists of four sensors located at $x = [0.35,\ 0.45,\ 0.55,\ 0.65]$. The lower panel in Fig.3.3 shows the distributed coexistence measure d_i/D for large $L = 2500$, and the corresponding envelopes (for $L = 50$). The corresponding values of the coexistence measure are: $D = 507.4$ in case A1, and $D = 268.5$ in case B1. In both cases the coexistence has been breached globally (since $D^* = 30$). However, locally, d_i is distributed quite unevenly. In case A1, for example, D accumulates its weight in the central part of the spatial domain, namely in the areas $x \in (0.46, 0.52)$ and $x \in (0.56, 0.58)$. This figure shows that the covariance matrix is useful for building the confidence intervals almost everywhere. However, for these two areas one may need to apply **locally (!)** different estimation and uncertainty quantification methods, such as particle methods.

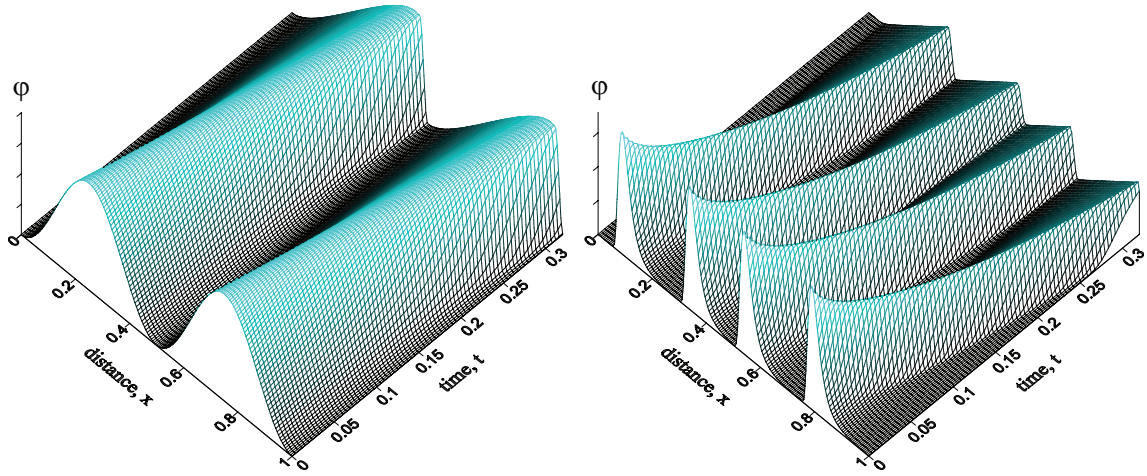

Figure 3.2: - Field evolution for different initial conditions: left/center/right - cases A/B, correspondingly.

3.1.6 Conclusions and future work

In the gaussian context, the estimation errors can be characterized by the confidence regions based on the estimation error covariance matrix. Technical difficulties in computing this covariance are related to the high dimensionality of the state vector, whereas a fundamental difficulty is related to the nonlinearity of the model equations and of the observation operator. Nonlinearity distorts the gaussian properties of the variational DA estimator and, as a result, the constructed gaussian confidence regions may render a totally wrong error characterization. In this case we say that the optimal solution is not gauss-verifiable. Since the distortion of the gaussian properties is due to nonlinearity, one may expect the loss of gauss-verifiability taking place *locally*, in and around the spatial areas where the nonlinear phenomena are particularly strong (such as shock waves, vortexes, etc). In other words, there may exist non-verifiable localized subsets of the state vector, while for the rest of the state vector the gaussian description of the error remains useful.

At a glance, to assess deviation from the gaussianity one should use the classical test statistics for multivariate normality. However, these statistics do not measure gauss-verifiability directly and usually require large ensembles/samples of optimal solutions for practical implementation, which is not feasible for the models in mind. Here we introduce a new statistic called the coexistence measure (CM). This statistic can also be considered as a 'global' test statistic for multivariate normality. We suggest a method for its decomposition into the sum of (predominantly) positive components associated to the elements of the control vector. The subsets contributing the most weight into the total value of the CM can be considered as non-verifiable. The CM and its decomposition is feasible to compute because: a) it can be evaluated on a basis of a very small ensemble of optimal solutions; b) all computations and storage are implemented in a matrix-free form.

In numerical experiments the CM and its decomposition have been tested. We notice that the integral measures not always provide a sufficiently clear insight on the nature and extent of non-gaussianity in variational DA systems. On the contrary, the CM decomposition offers a delicate tool for analysis of gauss-verifiability. Moreover, computing CM is fairly reliable with sample-deficient ensembles because of exploiting information provided by the Hessians. For example, such an ensemble could be a natural outcome of the 'ensemble 4D-Var'.

The method can be easily extended to the Bayesian case. That is, as a basis for defining the

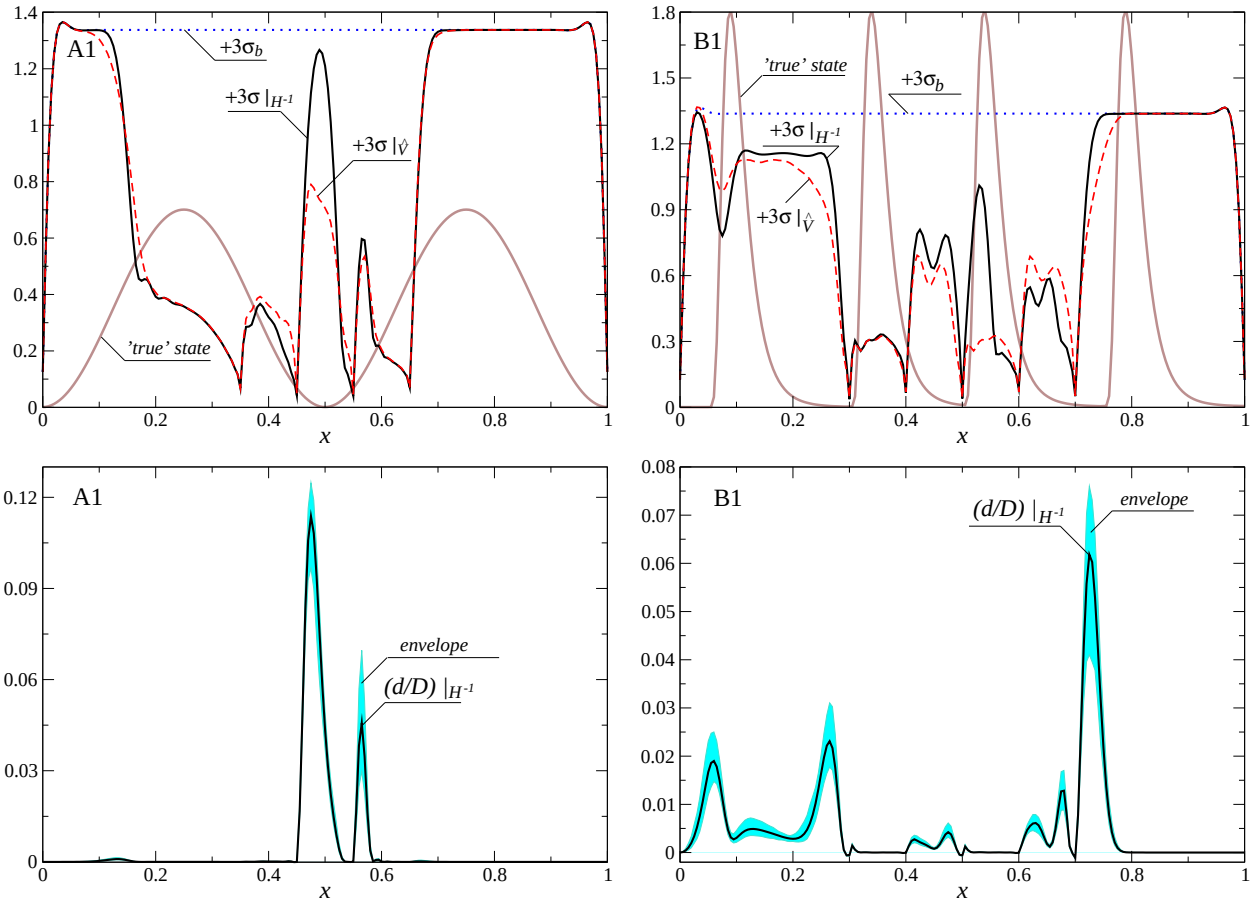

Figure 3.3: Cases A1 and B1.

coexistence measure one must consider the Bayesian pdf (2–26) instead of the 'analysis' pdf (3–2). It has been pointed out that the coexistence measure can be evaluated at a negligible cost in the course of computing the 'effective' inverse Hessian. However, more accurate 'effective' covariance estimates are suggested in §2.3, which implies that new computational procedures will be introduced. Subsequently, the CM evaluation process should stay consistent with these new procedures.

3.2 Design of optimal observation schemes

3.2.1 Introduction

State and/or parameter estimation for large-scale distributed parameters dynamical systems has become a routine task in the past two decades. The applications include model initialization in meteorology and oceanography, air and water quality monitoring, calibration of groundwater and reservoir models, flow estimation and control in aerospace engineering, process control in chemical and nuclear engineering, etc. A fundamental issue in these estimation problems is the selection of optimal sensors locations (stationary sensors) and/or sensor trajectories (mobile sensors). This problem is related to 'optimal sampling' or 'optimal experiment design' considered in statistics [S17, S18, S48]. In geosciences the notions 'adaptive' or 'targeted' observations are usually preferred [S6]. Optimization of a design function defined on the Fisher Information Matrix (FIM) is a classical approach to solve the experiment design problem. In geosciences, however, due to the large-scale nature of the models involved, the FIM based design criteria are not considered to be computationally feasible and, therefore, simplified approaches are used. In the context of variational DA one should mention methods using the total energy singular vectors [S35], the adjoint sensitivity [S12], the sensitivity to observations [S22] and the 'uncertainty field' obtained via the error subspace statistical estimation technique [S31]. The paper [S5] represents a rare example of the FIM based approach (the V-optimality condition) being used in large-scale variational DA.

Let us note that the FIM is actually equivalent (up to a multiplier) to the Hessian of the auxiliary cost-function (1–22). The limited-memory approximation of this Hessian (and of its inverse) can be efficiently evaluated using iterative methods, for example the Lanczos method or the LBFGS method. If the limited-memory approximation of the Hessian is reasonably accurate, then it can also be used to define the design criteria. Since the optimal design problem implies optimization of the design function, its gradient with respect to the design parameters (sensor locations and/or trajectories), if available, can be used in the gradient based optimization methods. For example, in [S47], an explicit formula for the gradient of the FIM-based design function is presented and used for solving the sensor-location problem (which includes some additional inequality constraints) by the linearization method. However, for the classical design criteria (D, A and E-optimality) this formula requires the gradient of *all* elements of the Hessian to be evaluated, thus it is not suitable for large-scale applications. In [S49] a similar sensor-location problem is solved by the global minimization method (the tunneling algorithm) which involves the local minimization stage, where the gradient could be very useful. Instead, this local problem is solved by the gradient-free method, which exhibits a linear convergence rate at best.

It is therefore quite certain that if the computationally inexpensive methods for evaluating the gradient of a design function are suggested, it might facilitate the use of these (well justified) optimality criteria for sensor placement in large-scale estimation problems, such that are common in variational DA. Here, based on the author's previous works [A13] and [A11], special new methods for evaluating the gradient of the design function associated to the limited-memory inverse Hessian are presented. The methods are suitable for large-scale applications. That means no full-storage matrices are involved in computations of the gradient at any stage. The particular design function is defined on the diagonal elements of the inverse Hessian, which approximate the estimation error variance. This function is related to the A-optimality design criterion [S17]; other classical criteria (for example, D and E-optimality design criteria) can be treated in a similar way. An efficient algorithm for evaluation of the gradient is presented. This algorithm exploits the special structure of the limited-memory inverse Hessian defined by a fixed number of its eigenpairs (Ritz pairs). The gradients obtained by the suggested method have been checked against the gradients obtained by the direct finite-difference method in a series of numerical experiments, conducted for the dynamical model governed by Burger's

equation. The methods presented are computationally inexpensive, though may require some additional memory to store the relevant information during the eigenvalue analysis of the Hessian by the Lanczos method.

3.2.2 Sensor-location problem statement

Let $X(x,t)$ be the model state defined for each time instant t in a bounded domain $x \in \Omega$ of the natural space R^d ($d = 1, 2$, or 3). Thus, we define the state space as $\mathcal{X} = L_2(0,T;\Omega)$. Assume that one has a set of L sensors located at coordinates $\bar{x} = (x_1, \ldots, x_L)^T$, $x_i \in \Omega$. We define the observation operator $C : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{Y} = L_2(0,T;R^L)$, by the formula:

$$Y = CX = \left((\phi(x,x_1),X)_{\mathcal{X}}, \ldots, (\phi(x,x_L),X)_{\mathcal{X}}\right)^T, \tag{3–31}$$

where $\phi(x,x_i) \in L_2(\Omega)$ are some support functions of $x \in \Omega$ for fixed x_i. Note that if $\phi(x,x_i)$ is the Dirac delta-function $\delta(x-x_i)$ (which does not belong to $L_2(\Omega)$), then for regular X (3–31) may be treated as

$$CX = (X(x_1,t), \ldots, X(x_L,t))^T, \tag{3–32}$$

so in this case the values of the state $X(x,t)$ at points $x = x_i,\ i = 1, \ldots, L$ are observed. This is a special (linear) case of the nonlinear observation operator defined by (1–3), which is considered for the sake of simplicity.

To define the adjoint operator C^*, we consider the scalar product

$$(CX,q)_{\mathcal{Y}} = \int_0^T (CX(x,t),q)_{R^L}\, dt, \quad X \in \mathcal{X},\ q = (q_1(t), \ldots, q_L(t))^T \in \mathcal{Y}.$$

By definition of C, the scalar product $(CX,q)_{\mathcal{Y}}$ is equal to

$$\int_0^T \sum_{k=1}^{L} (\phi(x,x_k),X(x,t))_{L_2(\Omega)}\, q_k(t)dt = \int_0^T \sum_{k=1}^{L} (\phi(x,x_k)q_k(t),X(x,t))_{L_2(\Omega)}\, dt = (X,C^*q)_{\mathcal{X}},$$

where the adjoint operator $C^* : \mathcal{Y} \to \mathcal{X}$ is defined by

$$C^*q = \sum_{k=1}^{L} \phi(x,x_k)q_k(t). \tag{3–33}$$

Consider the expression for the Hessian (1–21). Since $G(U) = C\mathcal{M}(U)$, see (1–3), one can write:

$$H(\cdot) = (\mathcal{M}'_U(\cdot))^*\, C^*R^{-1}C\, \mathcal{M}'_U(\cdot) + B^{-1}. \tag{3–34}$$

Given definitions for C and C^* in (3–32) and (3–33), the Hessian-vector product is defined as follows:

$$H(\cdot,\bar{x})\, v = (\mathcal{M}'_U(\cdot))^* \left(\sum_{k=1}^{L} \phi(x,x_k) \sum_{j=1}^{L} R^{-1}_{kj} \left(\phi(x,x_j),\, \mathcal{M}'_U(\cdot)v\right)_{\mathcal{X}} \right) + B^{-1}v. \tag{3–35}$$

In the finite-dimensional case $\mathcal{U} = R^m$, and the Hessian H is $m \times m$ matrix. In design of optimal observation scheme we are going to minimize certain diagonal elements of the inverse Hessian (or a combination of these elements) with respect to the position of the sensors. Let us introduce a design function in the form:

$$\Psi(\cdot,\bar{x}) = \sum_{i=1}^{m} p_i \left(H^{-1}(\cdot,\bar{x})e_i, e_i\right)_{R^m},\ p_i = const \geq 0, \tag{3–36}$$

where $\{e_i\}$ is the standard basis in R^m. Then, the *sensor-location problem* may be formulated as follows: find the set of sensor location coordinates $\bar{x}$ such that the function $\Psi(\cdot, \bar{x})$ takes its minimum value.

Since $H_{ii}^{-1} = (H^{-1}e_i, e_i)_{R^m}$ approximately represents the optimal solution error variance in i-th element of the state vector, the function Ψ can be chosen to quantify the accuracy of the optimal solution in a target area defined by p_i. If $p_i = 1,\ i = 1, \ldots, m$, then Ψ is the trace of H^{-1} and $\inf_{\bar{x}} \Psi(\cdot, \bar{x})$ is the classical A-optimality design criterion [S17, S18].

In order to apply the gradient-based optimization methods for solving the stated sensor-location problem one needs the gradient of the function $\Psi(\cdot, \bar{x})$ with respect to $\bar{x}$, that is $\Psi'_{\bar{x}}(\bar{x})$. We notice that Ψ is a linear combination of the following norms:

$$J(\cdot, \bar{x}) = \left(H^{-1}(\cdot, \bar{x})g, h\right)_{R^m} = h^T\, H^{-1}(\cdot, \bar{x})\, g \tag{3–37}$$

with $h = g = e_i$. Hence, the gradient $\Psi'(\cdot, \bar{x})$ can be expressed via the gradients of $J(\cdot, \bar{x})$ with respect to $\bar{x}$, that is $J'_{\bar{x}}(\cdot, \bar{x})$. Furthermore, we show that by specially choosing functions h and g in (3–37) one can express the gradient of eigenvalues and eigenvectors of the inverse Hessian via $J'(\cdot, \bar{x})$. This allows a very efficient algorithm for computing $\Psi'_{\bar{x}}$ to be constructed, with Ψ being defined on the limited-memory inverse Hessian in the form (2–15). Thus, a formula for the gradient $J'_{\bar{x}}(\cdot, \bar{x})$ must be derived first.

3.2.3 Gradient via adjoint of the Hessian derivative

Let $\delta X_v \in \mathcal{X}$ and $\delta X_w \in \mathcal{X}$ are the solutions to the tangent linear model with $\delta U = \{v\}$ and $\delta U = \{w\}$, respectively, i.e.

$$\delta X_v = \mathcal{M}'_U(\cdot)v, \quad \delta X_w = \mathcal{M}'_U(\cdot)w. \tag{3–38}$$

The main theoretical results of [A11] are summarized in the following propositions:

Proposition 1. The gradient of (3–37) is given by the formula

$$J'_{\bar{x}}(\cdot, \bar{x})\delta\bar{x} = -\left(H'\delta\bar{x}\ H^{-1}g,\ H^{-1}h\right)_{\mathcal{U}} = -\left(\delta\bar{x},\ (H'H^{-1}g)^* H^{-1}h\right)_{R^L}, \tag{3–39}$$

where $\delta\bar{x} = (\delta x_1, \ldots, \delta x_L)^T$ is the vector of variations in $\bar{x}$, $H'(\cdot, \bar{x})$ is a 3D-operator (tensor) with two entries, and $*$ denotes the adjoint operator. Thus, in order to find $J'_{\bar{x}}$, one must know the adjoint of the derivative of $H(\cdot, \bar{x})$ with respect to $\bar{x}$.

Proposition 2. The adjoint operator $(H'\delta\ v)^* w$ is defined on the functions $(v, w) \in \mathcal{U}$ by the formula:

$$(H'\ v)^* w = \bar{y}, \tag{3–40}$$

where $\bar{y} = (y_1, \ldots, y_L)^T$ is a vector with elements

$$y_i = \left(\delta X_w, \alpha_i \phi'(x, x_i)\bar{x}\right)_{\mathcal{X}} + \left(\delta X_w, \beta_i \left(\phi'(x, x_i)\bar{x}, \delta X_v\right)_{L_2(\Omega)}\right)_{\mathcal{X}}, \quad x_i \in \Omega, \tag{3–41}$$

$$\alpha_i = \sum_{j=1}^{L} R_{ij}^{-1} \left(\phi(x, x_j), \delta X_v\right)_{L_2(\Omega)}, \quad \beta_i = \sum_{j=1}^{L} R_{ji}^{-1} \phi(x, x_j).$$

In the above formula $\phi'(x, x_i)$ means the derivative of the support function $\phi(x, x_i)$ with respect to x_i:

$$\phi'(x, x_i) = \left\{ \frac{\partial\phi(x_i, x)}{\partial x_i^1}, \ldots, \frac{\partial\phi(x_i, x)}{\partial x_i^d} \right\},$$

where $x_i^k,\ k = 1, ..., d$ are the spatial components of $x_i \in \Omega \subset R^d$. Thus, to find $(H' \bullet v)^* w,\ (v, w) \in L_2(\Omega)$, one needs to solve two tangent linear problems (3–38) and then use (3–40)-(3–41).

Proposition 3. Given the observation error covariance matrix R does not depend on time and $\phi(x, x_j)$ is the Dirac δ-function, formula (3–41) for y_i takes the form

$$y_i = \sum_{j=1}^{L} R_{ij}^{-1} \int_0^T dt \left[\delta X'_w(x_i, t)\ \delta X_v(x_j, t) + \delta X_w(x_j, t) \delta X'_v(x_i, t) \right], \tag{3–42}$$

where $\delta X'(x, t)$ is the derivative of δX with respect to $x^k,\ k = 1, ..., d$. Moreover, if R is a diagonal matrix, then (3–42) can be further simplified as follows:

$$y_i = R_{ii}^{-1} \int_0^T (\delta X_w(x_i, t)\ \delta X_v(x_i, t))'\, dt,\ i = 1, \ldots, L. \tag{3–43}$$

Proposition 4. For a given set of the eigenpairs (Ritz pairs) $\{\lambda_k, U_k\},\ k = 1, \cdots, K$ of the projected Hessian

$$\tilde{H} = (B^{1/2})^* H B^{1/2}, \tag{3–44}$$

where H is given by (3–34), a limited-memory approximation of H^{-1} can be obtained in the form:

$$H^{-1} = B^{1/2} \tilde{H}^{-1} (B^{1/2})^* = B^{1/2} (I + \bar{U}(\bar{S} - \bar{I})\bar{U}^T)(B^{1/2})^*, \tag{3–45}$$

where $\bar{S} = diag\{s_1, \ldots, s_K\}$, $s_k = \lambda_k^{-1}$, and $\bar{U}$ is a rectangular $m \times K$ matrix containing the eigenvectos U_k. Usually, K is much less than the state vector dimension m. By definition

$$U_k^T U_l = \begin{cases} 1,\ l = k \\ 0,\ l \neq k \end{cases}. \tag{3–46}$$

Proposition 5. For the functions $g_k = (B^{-1/2})^* U_k$,

$$H^{-1} g_k = B^{1/2} s_k U_k, \tag{3–47}$$

$$g_l^T B^{1/2}(\cdot)(B^{1/2})^* g_k = U_l^T(\cdot) U_k. \tag{3–48}$$

3.2.4 Computation of the gradient of the design function

When choosing $h = g = e_j$ and $p_i = \delta_{ij}$ (the Kronecker delta) in (3–37), then $\Psi = J = H_{jj}^{-1} = (H^{-1} e_j, e_j)_{R^m}$; thus $\Psi'_{\bar{x}} = J'_{\bar{x}}$ can be computed using (3–39), (3–40), and either (3–41) or (3–42) or (3–43), where δX_v is the solution of the tangent linear model (3–38) with $v = H^{-1} e_j$, and $\delta X_w = \delta X_v$. For large-scale problems this could be a valuable option in the situation when a certain small subset of the state-vector is of particular interest, for example some model parameters in the joint state-parameter estimation problem.

If we are interested in the subsets of the state vector which may include a significant total number of elements, another approach can be used. Let P be a diagonal $m \times m$-matrix with elements p_i and let us consider the design function in the form (3–36). Denoting by $B_i^{1/2}$ the i-th row of $B^{1/2}$ and taking into account (3–45), Ψ can be transformed as follows:

$$\Psi = \sum_{i=1}^{M} p_i e_i^T B^{1/2} (I + \bar{U}(\bar{S} - I)\bar{U}^T)(B^{1/2})^* e_i = \Psi_0 + \sum_{i=1}^{M} p_i B_i^{1/2} \bar{U}(\bar{S} - I)(B_i^{1/2} \bar{U})^T =$$

$$= \Psi_0 + \sum_{i=1}^{M} p_i \sum_{k=1}^{K} (s_k - 1)(B_i^{1/2} U_k)(B_i^{1/2} U_k)^T = \Psi_0 + \sum_{k=1}^{K} (s_k - 1) \sum_{i=1}^{M} p_i (B_i^{1/2} U_k)(B_i^{1/2} U_k)^T =$$

$$= \Psi_0 + \sum_{k=1}^{K} (s_k - 1)(B^{1/2} U_k)^T P (B^{1/2} U_k),$$

where $\Psi_0 = \sum_{i=1}^{M} p_i e_i^T B e_i$ does not depend on $\bar{S}$ or $\bar{U}$. Then, the variation of Ψ is defined by

$$d\Psi = \sum_{k=1}^{K} d\Psi_k, \tag{3–49}$$

where

$$d\Psi_k = ds_k (B^{1/2} U_k)^T P (B^{1/2} U_k) + 2(s_k - 1)(B^{1/2} U_k)^T P B^{1/2} dU_k. \tag{3–50}$$

Since only a limited set of the eigenvectors of the Hessian is available, we assume that the variation dU_k can be spanned using the available orthogonal basis $\bar{U}$, i.e.

$$dU_k = \bar{U} dV_k, \tag{3–51}$$

where $dV_k = (dv_{k,1}, ..., dv_{k,M})^T$ is a vector of coefficients of size K. This is the only approximation accepted in the suggested method. Then equation (3–50) can be written as follows

$$d\Psi_k \approx ds_k (B^{1/2} U_k)^T P (B^{1/2} U_k) + 2(s_k - 1)(B^{1/2} U_k)^T P B^{1/2} \bar{U} dV_k. \tag{3–52}$$

Next, we are going to show how ds_k and dV_k for (3–52) can be evaluated given the rule of computation of the gradient for the cost function (3–37), where g and h are vectors which do not depend on $\bar{x}$ and H^{-1} is given by (3–45). We differentiate (3–37) and obtain

$$dJ = h^T B \left(\bar{U} d\bar{S} \bar{U}^T + d\bar{U} (\bar{S} - I) \bar{U}^T + \bar{U} (\bar{S} - I) d\bar{U}^T \right) (B^{1/2})^* g. \tag{3–53}$$

Let us put $h = g_l = (B^{-1/2})^* U_l, g = g_k = (B^{-1/2})^* U_k,\ l, k = 1, ..., K$. Then, taking into account (3–46), (3–47), (3–48) and (3–51), for each term in (3–53) we derive the following relationships:

$$U_l^T \, \bar{U} \, d\bar{S} \, \bar{U}^T \, U_k = \begin{cases} ds_k, & l = k \\ 0, & l \neq k \end{cases}, \tag{3–54}$$

$$U_l^T \, d\bar{U} (\bar{S} - I) \bar{U}^T \, U_k = \begin{cases} 0, & l = k \\ (s_k - 1) dv_{k,l}, & l \neq k \end{cases}, \tag{3–55}$$

$$U_l^T \, \bar{U} (\bar{S} - I) d\bar{U}^T \, U_k = \begin{cases} 0, & l = k \\ (s_l - 1) dv_{l,k}, & l \neq k \end{cases}. \tag{3–56}$$

Since $dv_{l,k} = -dv_{k,l}$, we finally obtain

$$dJ_{l,k} = \begin{cases} ds_k, & l = k \\ (s_k - s_l) dv_{k,l}, & l \neq k \end{cases}. \tag{3–57}$$

Based on the considerations presented in this section we formulate the following Theorem.

Theorem 1. The gradient of the design function $\Psi(\cdot,\bar{x})$ with respect to $\bar{x}$ can be expressed via the gradients in the eigenvalues and eigenvectors of the projected Hessian (3–44) by the formula

$$\Psi'_{\bar{x}}(\cdot,\bar{x}) = \sum_{k=1}^{K} \Psi'_{\bar{x},k}(\cdot,\bar{x}), \tag{3–58}$$

where

$$\Psi'_{\bar{x},k}(\cdot,\bar{x}) \approx s'_{\bar{x},k}(B^{1/2}U_k)^T P(B^{1/2}U_k) + 2(s_k-1)(B^{1/2}U_k)^T P B^{1/2}\bar{U}V'_{\bar{x},k}, \tag{3–59}$$

with $s'_{\bar{x},k}$ and $V'_{\bar{x},k}$ satisfying

$$J'_{\bar{x},(l,k)} = \begin{cases} s'_{\bar{x},k}, & l = k \\ (s_k - s_l)v'_{\bar{x},(k,l)}, & l \neq k \end{cases}. \tag{3–60}$$

3.2.5 Key implementation trick

It follows from **Propositions 1-5**, that

$$J'_{\bar{x}(l,k)} = -\bar{y}, \tag{3–61}$$

where $\bar{y}$ is defined by (3–41) or (3–42) or (3–43) with

$$\delta X_v := \delta X_l, \;\; \delta X_w := \delta X_k, \;\; \delta X_j = \mathcal{M}'_U(\cdot)B^{1/2}s_jU_j, \; \forall j, \tag{3–62}$$

$$\delta X'_v := \delta X'_l, \;\; \delta X'_w := \delta X'_k, \;\; \delta X'_j = \frac{\partial}{\partial x}\left(\mathcal{M}'_U(\cdot)B^{1/2}s_jU_j\right), \; \forall j. \tag{3–63}$$

Applying the Lanczos method for the preconditioned Hessian matrix $\tilde{H}$ we get the decomposition [S37]

$$W^T\tilde{H}W = T, \tag{3–64}$$

where $W = \{w_1,\ldots,w_K\}$ is the matrix formed by K Lanczos vectors, and T is a tridiagonal $K\times K$ symmetric matrix. The singular value decomposition of T is

$$T = Z\Lambda Z^T,$$

where $\Lambda = diag\{\lambda_1,\ldots,\lambda_K\}$ and $Z = \{z_1,\ldots,z_K\}$ is the matrix of eigenvectors of T. The eigenvalues λ_k of T (or the Ritz values) approximate the first K largest eigenvalues of $\tilde{H}$, and the Ritz vectors

$$U_k = Wz_k, \;\; k = 1,\ldots,K, \tag{3–65}$$

approximate the corresponding eigenvectors of $\tilde{H}$.

In the course of the Lanczos process the product

$$\tilde{H}w_k = w_k + (B^{1/2})^*(\mathcal{M}'_U(\cdot))^*\, C^*R^{-1}\left(C\,\mathcal{M}'_U(\cdot)B^{1/2}w_k\right)$$

is computed for $k = 1,\ldots,K$, which implies that K vectors $\delta\tilde{X}_k(\bar{x},t) = \left(\mathcal{M}'_U(\cdot)B^{1/2}w_k\right)(\bar{x},t)$ and $\delta\tilde{X}'_k(\bar{x},t) = \left(\partial(\mathcal{M}'_U(\cdot)B^{1/2}w_k)/\partial x\right)(\bar{x},t)$ are available after the process. These vectors have to be stored in memory, along with $K\times K$-matrix Z.

Now, consider (3–62). Taking into account (3–64) one can write

$$\delta X_k = C\mathcal{M}'_U(\cdot)B^{1/2}s_kU_k = C\mathcal{M}'_U(\cdot)B^{1/2}s_kWz_k = C\mathcal{M}'_U(\cdot)B^{1/2}s_k\sum_{j=1}^{K} w_jz_{k,j}$$

$$= s_k \sum_{j=1}^{K} z_{k,j}\, C\mathcal{M}'_U(\cdot)B^{1/2}w_j = s_k \sum_{j=1}^{K} z_{k,j}\, \delta\tilde{X}_j \tag{3–66}$$

Similarly we obtain

$$\delta X'_k = s_k \sum_{i=1}^{K} z_{k,j}\, \delta\tilde{X}'_j. \tag{3–67}$$

The above relationships are of a fundamental importance. One can see that $\delta X_k(\bar{x},t)$ and $\delta X'_k(\bar{x},t)$ can be evaluated using $\delta\tilde{X}_j(\bar{x},t)$ and $\delta\tilde{X}'_j(\bar{x},t)$, which have been accumulated during the computation of the eigenpairs of $\tilde{H}$. This means that no additional runs of the tangent linear model are necessary. Therefore, the gradient of the design function $\Psi(\cdot,\bar{x})$ can be obtained at a negligible computational cost (in terms of CPU time) in the process of computing the design function itself.

3.2.6 Illustration

Some results showing the gradients (with respect to sensor location co-ordinates x_i) in the diagonal elements of the inverse Hessian are presented in Fig.3.4. As a model we use the 1D Burger's equation (2–23). The diagonal of H^{-1}, which is an approximation of the estimation error variance, is presented in the left panel. Each j-th element of this diagonal corresponds to the element of the unknown initial state $u(x)$ with the coordinate $x = (j-1)h_x$. The initial location of sensors is given by the vertical lines. The scaled gradients $(H^{-1}_{jj})'_{x_i}/H^{-1}_{jj}$, $j = 1, ..., m$ are presented in the right panel.

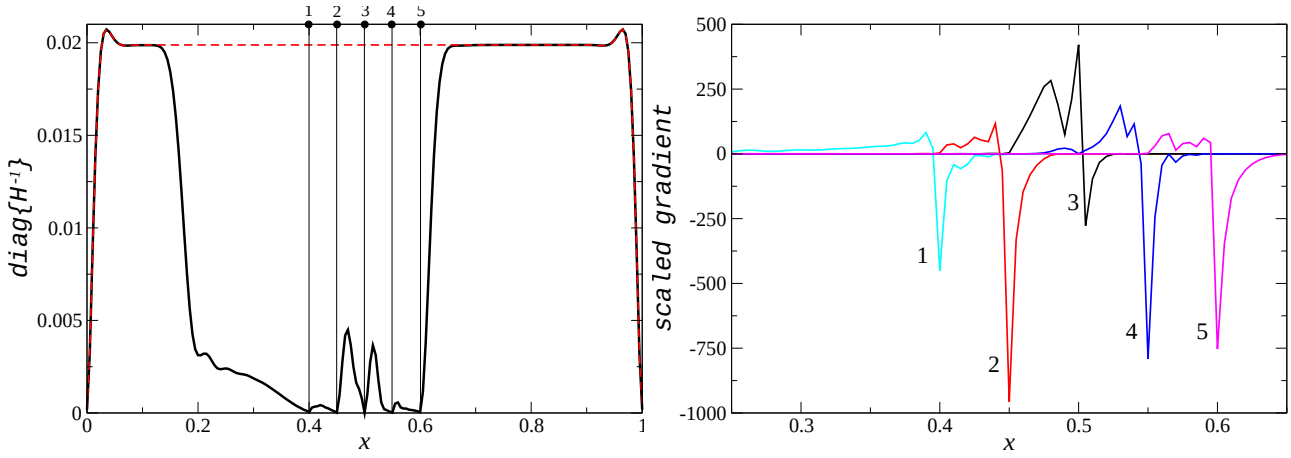

Figure 3.4: Left: the reference 'variance' $diag\{H^{-1}\}$ - solid line, the background variance - dashed line. Right: the scaled gradients of the 'variance' with respect to the sensor location co-ordinates x_i, $i = 1, \ldots, 5$.

3.2.7 Conclusions and future development

Design of optimal observation schemes or trajectories is a fundamental problem in state or/and parameter estimation of distributed parameter dynamical systems. In large-scale applications, such as DA in meteorology and oceanography, the classical design criteria defined on the FIM or, equivalently, on the Hessian of the auxiliary control problem are not computationally feasible. However, the limited-memory approximation of the Hessian (and of its inverse) can be obtained with a reasonable computational cost by iterative methods. We have assumed that this approximation is good, i.e. minimizing design functions defined on the exact (full memory) and approximate (limited-memory)

inverse Hessians produces close solutions. In practice, this assumption is bound to be valid if the eigenspectrum of the projected Hessian (3–44) has a small number of well-separated eigenvalues distant from unity. This can be achieved by using an appropriate preconditioning. The objective has been to derive an inexpensive method for evaluating the gradient of a design function defined on the diagonal elements of the limited-memory inverse Hessian. This gradient can be used for the local optimization of the design function by gradient methods (the conjugate-gradient or quasi-Newton methods, for example) in the framework of a chosen global optimization procedure.

The first approach allows the gradient of a single diagonal element of the inverse Hessian to be evaluated. This requires just a single run of the tangent linear model. The method provides a valuable design option for those variational DA problems where target areas of the state vector (areas of interest) contain a relatively small number of elements. The more general approach allows the gradient of any weighted sum of the diagonal elements of the inverse Hessian (the A-optimality criteria makes a special case) to be evaluated. Moreover, if the inverse Hessian is defined by the Ritz pairs obtained by means of the Lanczos/Arnoldi iterative algorithm and the relevant information is stored in the computer memory during these iterations, then the gradient can be obtained at a negligible computational cost because no additional run of the tangent linear model is required. Given the availability of inexpensive gradients, the benefit of using the gradient-based optimization methods instead of the gradient-free optimization methods is quite obvious. The natural condition for this method to produce accurate gradients is that the limited-memory inverse Hessian must be a good approximation of the exact (full-memory) inverse Hessian.

The presented methodology is well developed in a sense that the major theoretical issues have been answered. If required, it can be easily generalized to the case of nonlinear observation operators, moving sensors, and to different design functions. Since the design problems involving Hessian (or FIM) are usually solved using the gradient-free methods, it is an important task to promote the gradient-based approach for a wider practical use. To this end one could investigate the design of targeted observations for tracer monitoring or the determination of tracer source location in a realistic coastal circulation model governed by the 2D shallow water equations, for example. Another interesting problem is optimal design of observation array for large-scale distributed hydrological models.

3.3 Implicit ('idle') control and model error

3.3.1 Introduction

Variational DA method is preferred for weather and ocean forecasting in major operational centers around the globe, particularly in the form of the incremental 4D-Var ([S11]), and in the form of the *ensemble 4D-Var* ([S9]). Here, the quantity of interest is a forecast: the state of a system at the forecast time which corresponds to the estimated controls. The use of variational DA is not limited to operational forecasting, however. In some problems the quantity of interest can be represented directly by the estimated controls. For example, in [S1] the sea surface heat flux is estimated in order to understand its spatial and temporal variability. Another case is when such estimates serve as inputs to different models. For example, the estimated initial condition from a low resolution model could serve to initialize a higher resolution prognostic model.

A *perfect model* assumption leads to the *strong constraint* formulation of variational DA. However, this assumption is not valid in many realistic applications due to uncertainties, known under the common name of *model error*. Such uncertainties appear due to unresolved scales, unaccounted physical processes, crude discretization of model equations and/or mis-specification of the model parameters. Usually, the model error is defined as a linear source term added to the right-hand side of the model equations. Inclusion of this term into the control vector yields the so-called *weak constraint* variational DA formulation ([S38]). However, the model error can be introduced in many different ways. For example, if we assume that the error magnitude depends on the flow field gradient, then it can be defined as a coefficient inside the convective terms.

In variational data assimilation authors usually try to estimate the model error, which means that all uncertainty-loaded model inputs are included into the control vector. However, this approach suffers from serious implementation difficulties. The first one is an oversizing issue. Indeed, for the model error which depends both on space and time (i.e. belongs to discretized X), the control vector size would be equal to the dimension of the state vector multiplied by the number of time integration steps. Dealing with a vector of that size could be difficult in terms of RAM memory. Furthermore, the conditioning of the extended problem could be far less favorable than that of the original problem, and noticeably more iterations may be needed to converge to a chosen accuracy level. Secondly there are solvability and robustness issues. Let us recall that some models involved in DA are highly nonlinear. For such models, the inclusion of certain additional variables into the control vector may lead to a far more complex cost-function (in terms of its convexity, connectivity, etc.) and, subsequently, to essential difficulties during the minimization process. This is not always the case though, as a weak constraint formulation may actually reduce the nonlinearity level.

Here we describe an alternative approach to the model error treatment in which the control vector does not need to be extended. This description is based on author's work presented in [A7, A3]. The main idea comes from the 'nuisance parameter' concept, well known both in classical and Bayesian statistics. In this concept the uncertainty-bearing model parameters are divided into two groups: parameters of interest, say u, and nuisance parameters, say λ. The estimates of λ, by themselves, have no practical importance; however uncertainty in λ affects the estimates of u. Many methods have been proposed to eliminate the nuisance parameters from the likelihood function in classical statistics. These methods often lead to the construction of a 'pseudo-likelihood', a function of the data and u with properties similar to those of a likelihood function. Commonly used pseudo-likelihood functions include conditional, marginal and profile likelihood functions ([S39]). From the Bayesian point of view, the elimination of λ is achieved by integrating it out, using a conditional prior density for λ given u.

Let us note that the model error affects the quantity of interest in two ways: directly and via the estimate of the model controls. We consider a complete set of model inputs, including the distributed

source term, as a full control set. If uncertainty in some input variables dominates the error in the quantity of interest directly (such variables may even coincide with it), then these variables must be included into the control vector. Such explicitly controlled variables constitute an 'active' control set. In some other input variables, uncertainty could be so small that its influence on the quantities of interest may reasonably be neglected. Such variables constitute a 'passive' set. Finally, if uncertainty in some input variables is significant, but the inclusion of such variables into the 'active' control set entails severe computational difficulties, these variables could be subjected as 'nuisance parameters' to an implicit treatment. A set of such variables shall be called an 'idle' control set. Thus, the basic idea of the proposed method is to consider the model error as an idle control, i.e. to reduce its indirect influence on the quantity of interest by appropriately modifying the cost function. In particular, this is achieved by introducing an 'inflated observation covariance'. Let us note that the method has been established in a purely deterministic framework. The suggested method as it stands should be useful primarily when the defined quantities of interest coincide with the active controls or are totally dominated by them (for example, the state evolution at the beginning of the observation window in the initial condition control problem). The method is feasible with high-dimensional models if the observation space dimension is much smaller than the control space dimension.

3.3.2 Theory of the method

Let us consider the cost function (1–6) and divide vector $U \in \mathcal{U}$ into two parts: U_a and $U_q = U \setminus U_a$ - the 'active' and 'idle' subsets of U, respectively. The input space $\mathcal{U}$ is also devided in two sub-spaces $\mathcal{U}_a$ and $\mathcal{U}_q$, such that $\mathcal{U} = \mathcal{U}_a \times \mathcal{U}_q$. Correspondingly, we consider the 'true' values $\bar{U}_a$ and $\bar{U}_q$, and the background values $U_a^* = \bar{U}_a + \varepsilon_a$ and $U_q^* = \bar{U}_q + \varepsilon_q$, where $\varepsilon_a \sim \mathcal{N}(0, B_a)$ and $\varepsilon_q \sim \mathcal{N}(0, B_q)$. Furthermore we assume that ε_a and ε_q are not correlated (or already de-correlated), and B_a and B_q are symmetric positive definite covariance operators. The 'inputs-to-observations' mapping is now defined as $G(U) \equiv G(U_a, U_q)$, $G : \mathcal{U} \to \mathcal{Y}$, and the observations as $Y^* = G(\bar{U}_a, \bar{U}_q) + \xi$.

By definition, U_a represents the variables which contain significant uncertainties and strongly influence the model forecast. Such variables are usually estimated (controlled) using available observations. For example, in the conventional 4D-Var these are the initial states of the system. Concerning the variables in U_q, we assume that they do not contain significant uncertainties or that these uncertainties do not affect the model forecast. In modeling, such variables are set to their background values. The DA problem is then formulated as the following optimal control problem: for a given $U_a^* \in \mathcal{U}$, $U_q = U_q^* \in \mathcal{U}$ and $Y^* \in \mathcal{Y}$, find $U_a \in \mathcal{U}$ such that

$$J_0(U_a) = \frac{1}{2}\|R^{-1/2}(G(U_a, U_q^*) - Y^*)\|_{\mathcal{Y}}^2 + \frac{1}{2}\|B_a^{-1/2}(U_a - U_a^*)\|_{\mathcal{U}_a}^2 \to \inf_{U_a}. \tag{3–68}$$

However, if uncertainty in U_q can not be ignored, these variables should also be involved in the estimation process. The most obvious and widely used approach is to include U_q into the control vector and consider the following extended estimation problem: for a given $U_a^* \in \mathcal{U}_a$, $U_q^* \in \mathcal{U}_q$ and $Y^* \in \mathcal{Y}$ find a pair (U_a, U_q) such that

$$J(U_a, U_q) = \frac{1}{2}\|R^{-1/2}(G(U_a, U_q) - Y^*)\|_{\mathcal{Y}}^2$$

$$+\frac{1}{2}\|B_a^{-1/2}(U_a - U_a^*)\|_{\mathcal{U}_a}^2 + \frac{1}{2}\|B_q^{-1/2}(U_q - U_q^*)\|_{\mathcal{U}_q}^2 \to \inf_{U_a, U_q}. \tag{3–69}$$

The estimator associated with (3–69) has the form:

$$\left\{ \begin{array}{rcl} B_a^{-1}(\hat{U}_a - U_a^*) + (G'_{U_a}(\hat{U}_a, \hat{U}_q))^* R^{-1}(G(\hat{U}_a, \hat{U}_q) - Y^*) & = & 0, \\ B_q^{-1}(\hat{U}_q - U_q^*) + (G'_{U_q}(\hat{U}_a, \hat{U}_q))^* R^{-1}(G(\hat{U}_a, \hat{U}_q) - Y^*) & = & 0. \end{array} \right. \tag{3–70}$$

Consider a modified cost-function

$$J(U_a|U_q) = \frac{1}{2}\|{R_g}^{-1/2}(G(U_a, U_q^*) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B_a^{-1/2}(U_a - U_a^*)\|^2_{\mathcal{U}_a} \to \inf_{U_a}, \tag{3–71}$$

where

$$R_g = R + G'_{U_q}(\bar{U})B_q(G'_{U_q}(\bar{U}))^*. \tag{3–72}$$

The operator $R_g : \mathcal{Y} \to \mathcal{Y}$ is referred to as the 'inflated observation covariance'. Here, the input subset U_q is involved in the estimation process implicitly via R_g, hence the name 'idle' (as opposite to 'active' or 'passive'). The estimator associated with (3–71) has the form:

$$B_a^{-1}(\hat{U}_a - U_a^*) + (G'_{U_a}(U_a, U_q^*))^* R_g^{-1}(G(U_a, U_q^*) - Y^*) = 0. \tag{3–73}$$

Let $\hat{U}_a^{(1)}$ be the solution to the minimization problem (3–69) and $\hat{U}_a^{(2)}$ - the solution to (3–71)-(3–72). Then, the following Theorem is proved in [A3]:

Theorem 2. The estimators associated to cost-functions $J(U_a^{(1)}, U_q)$ and $J(U_a^{(2)}|U_q)$ are equivalent in the following sense: for a linear mapping G the optimal solution errors $\delta\hat{U}_a^{(1)} = \hat{U}_a^{(1)} - \bar{U}$ and $\delta\hat{U}_a^{(2)} = \hat{U}_a^{(2)} - \bar{U}$ are identical, thus the solutions $\hat{U}_a^{(1)}$ and $\hat{U}_a^{(2)}$ are identical; for a nonlinear G these solutions match approximately if the tangent linear hypothesis is valid.

It follows from the Theorem that $U_a^{(1)}$ from (3–69) can be approximated by the solution $U_a^{(2)}$ to (3–71), using U^* instead of $\bar{U}$ in (3–72). This approximation can be further improved by using the following iterative process:

$$U_{a,k}^{(2)} = \text{argmin } J_k(U_a|U_q),$$

where $J_k(U_a|U_q)$ is defined by (3–71) with $R_g = R + G'_{U_q}(U_{a,\,k-1}^{(2)}, U_q^*)B_q(G'_{U_q}(U_{a,\,k-1}^{(2)}, U_q^*))^*$ and k is the iteration number.

Let us recall that $\xi \sim \mathcal{N}(0, R)$, $\varepsilon_a \sim \mathcal{N}(0, B_a)$ and $\varepsilon_q \sim \mathcal{N}(0, B_q)$. Since non-linear least-square estimators are asymptotically unbiased, we assume that

$$E(\delta\hat{U}_a^{(1)}) = E(\delta\hat{U}_a^{(2)}) = 0.$$

Then, the following is valid:

Corollary. Under conditions of the Theorem 2,

$$E(\delta\hat{U}_a^{(1)}\delta\hat{U}_a^{(1)T}) \approx H_g^{-1}(\bar{U}) \approx H_g^{-1}(\hat{U}_a, U_q^*) \tag{3–74}$$

where

$$H_g(\cdot) = B_a^{-1} + (G'_{U_a}(\cdot))^* {R_g}^{-1}(\cdot)G'_{U_a}(\cdot). \tag{3–75}$$

The Corollary shows the way of computing the estimation error covariance for any chosen subset of the full control vector U. This could a valuable option if the dimension of U is extremely large, both in terms of the memory and the number of Lanczos (or LBFGS) iterations required.

3.3.3 Implementation

In order to calculate the cost-function (3–71) one must define the product $R_g^{-1}v$, $\forall v \in \mathcal{Y}$, whereas R_g is given by (3–72). Let us consider the preconditioned R_g:

$$\tilde{R}_g = R^{-1/2} R_g R^{-1/2} = I + R^{-1/2} G'_{U_q}(\cdot,\cdot) B_q (G'_{U_q}(\cdot,\cdot))^* R^{-1/2}. \tag{3–76}$$

It can be seen from the above formula that all eigenvalues of $\tilde{R}_g$ are greater than or equal to one. Furthermore, it should be anticipated that only a relatively small percentage of the eigenvalues are distinct enough from unity to contribute significantly to $\tilde{R}_g$. This suggests using *limited-memory* representations of $\tilde{R}_g$, where this structure in the spectrum is exploited. Specifically, a few leading eigenvalue/eigenvector pairs $\{\beta_i, z_i\}$, $i = 1, \ldots, N_1$ are computed using the Lanczos method, and $\tilde{R}_g^{-1}$ is replaced by the approximation

$$\tilde{R}_g^{-1} \simeq I + \sum_{i=1}^{N_1} (\beta_i^{-1} - 1) z_i z_i^T.$$

Finally, for $R_g^{-1}v$ we get

$$R_g^{-1} v = R^{-1/2} \left(I + \sum_{i=1}^{N_1} (\beta_i^{-1} - 1) z_i z_i^T \right) R^{-1/2} v. \tag{3–77}$$

Re-evaluating of $\{\beta_i, z_i\}$ has to be done if the entries in $G'_{U_q}(\cdot,\cdot)$ are changed. When using (U_a^*, U_q^*), this procedure has to be performed once before the minimization of (3–71) begins.

The same approach is used for computing the inverse of H_g in (3–75). In particular, the eigenpairs $\{\beta_i, z_i\}$, $i = 1, \ldots, N_2$ of a preconditioned $\tilde{H}_g = B_a^{T/2} H_g B_a^{1/2}$ are evaluated, H_g^{-1} is then recovered using

$$H_g^{-1} = B_a^{1/2} \left(I + \sum_{i=1}^{N_2} (\beta_i^{-1} - 1) z_i z_i^T \right) B_a^{T/2}. \tag{3–78}$$

3.3.4 One possible application

Let us assume that the model error is biased, i.e. $\varepsilon_q \sim \mathcal{N}(\varepsilon_{q,s}, B_q)$. This implies that $\varepsilon_q = \varepsilon_{q,s} + \varepsilon_{q,r}$, where $\varepsilon_{q,s}$ and $\varepsilon_{q,r} \sim \mathcal{N}(0, B_q)$ are the systematic and random components of ε_q, respectively. Let us also consent that the systematic error component('mean field') is not changing in time, whereas the 'random' component is time-dependent, i.e. $\varepsilon_{q,r} = \varepsilon_{q,r}(t)$. Then, the explicit treatment of $\varepsilon_{q,r}$ is not feasible.

Proposition. In treatment of the space-time distributed model error, its systematic component $\varepsilon_{q,r}$ should be considered as 'active' control, whereas the random component $\varepsilon_{q,r}$ - as 'idle' control. This results into the minimization problem for the following cost-function:

$$J(U_a, \varepsilon_{q,s} | U_q) = \frac{1}{2} \| R_g^{-1/2} (G(U_a, U_q^* - \varepsilon_{q,s}) - Y^*) \|_{\mathcal{Y}}^2$$

$$+ \frac{1}{2} \| B_a^{-1/2} (U_a - U^*) \|_{\mathcal{U}_a}^2 + \frac{\alpha}{2} \| W \varepsilon_{q,s} \|_{\mathcal{U}_q}^2 \to \inf_{U_a,\, \varepsilon_{q,s}}, \tag{3–79}$$

where $R_g = R + G'_{U_q}(\bar{U}_a, \bar{U}_q) B_q (G'_{U_q}(\bar{U}_a, \bar{U}_q))^*$, α is a regularization parameter, and W is a weight-matrix.

In practice, instead of $(\bar{U}_a, \bar{U}_q)$ in the expression for R_g one can initially use (U_a^*, U_q^*). The cost-function is interesting in the sense that it combines the elements of the Bayesian and deterministic approaches (the Tikhonov regularization method). This is because no probabilistic description of the systematic model error component $\varepsilon_{q,s}$ is available. Therefore, the parameter α and the weight matrix W must be chosen from additional considerations.

3.3.5 Illustration

Numerical tests has been performed for the model (2–23). The major purpose of these tests has been to assess the accuracy of the method in the case of nonlinear G (bilinear and nonlinear modes of equation (2–23)). The results can be seen in [A3]. Later on, the method has been successfully applied to the discharge estimation problem under uncertainty in bathymetry and the bed roughness (the Strickler coefficient), using the Saint-Venant hydraulic network model.

3.3.6 Conclusions

1. The influence of the model error on the estimates of the variables of interest can be eliminated/reduced by considering this error as an idle control variable, which implies the implicit treatment via the inflated observation error covariance R_g (in essence, the modified likelihood). The theorem has been proved which asserts partial equivalency between the suggested method and the conventional control vector extension method for a linear control-to-observations mapping, and approximate equivalency for a nonlinear mapping. In the latter case, the quality of the approximation seems superior if measured in terms of the estimation error variance. Therefore, an alternative method for the model error treatment is suggested, which has not been considered before, not in the framework of variational DA involving high-dimensional geophysical flow models at least. The only relevant reference is [S23], the paper which has appeared almost simultaneously with [A3].
2. The difficulties associated with the control vector extension method include oversizing, solvability and robustness. The proposed method allows us to alleviate these difficulties. However, the method would only be useful if the active and idle control sets are correctly defined, i.e. if the indirect contribution into the quantities of interest by an idle variable is significantly larger than its direct contribution. By the indirect contribution we mean the contribution via the estimates of the variables from the active control set. Of course, to properly attribute the input variables into the active, idle and passive sets one must perform an uncertainty analysis of the system. For example, it has been shown in the numerical experiments that a spatially distributed mean of the model error should be included in the active control set, whereas time dependent fluctuations around this mean - in the idle control set.
3. The Corollary has also been proved stating that the covariance of the estimation error in the variables of interest can be approximated by the inverse Hessian of the auxiliary cost-function associated to the control problem formulation involving the inflated observation error covariance R_g. This shows a way of computing the estimation error covariance of a chosen part of a control vector.
4. The proposed method is feasible for high-dimensional problems since R_g is represented by a small set of its largest eigenvalues (and corresponding eigenvectors), obtained by means of the Lanczos algorithm. For a linear mapping G the inflated observation error covariance R_g does not depend on the model state, therefore it can be computed once and stored. For a nonlinear mapping G, R_g depends on the state. However, the numerical tests have shown that R_g computed at the background point may suffice, at least for the chosen model (generalized Burgers' equation).
5. The suggested implicit model error treatment method is primarily suited to the case when the quantities of interest coincide with the active controls or are largely dominated by them (reanalysis, as opposed to forecasting).

6. The presented theory can be used as a basis for a control space decomposition approach (work in progress), which might eventually lead to a new variant of the method better suited to forecasting.

3.4 Design of the control set

3.4.1 Introduction

In many applications the choice of the control set seems rather obvious. For example, in short-range forecasting using global atmospheric or ocean models the initial state is controlled, whereas for longer forecasting periods one must also control the forcing term to remove the model bias. When the limited-area models are considered, the boundary conditions at open boundaries are usually controlled. However, there are applications when the control set composition is not so evident, for example, in hydraulic and hydrological modeling. A key role in this modeling plays information about river discharges. A distinctive feature of this problem is the likely presence of significant uncertainty in distributed source terms (lateral inflows and outflows) and in model parameters, such as bathymetry, friction, infiltration rate or in those defining behavior of hydraulic structures. This uncertainty, if not taken into account, could degrade the estimated discharge accuracy very noticeably.

The usual way to tackle the (systematic) uncertainties is to include all uncertainty-bearing model inputs into the control set. An ultimate implementation of this idea results into the model error control concept or the *weak* DA formulation [S38]. Unless the available computational resources are exceeded, working with such control set is not too difficult in the variational DA framework. However, there are clear reasons for limiting the number of control variables included into the control set. First we note that when a certain input is added into the control set, the corresponding constraints should be added to keep well-posedness of the problem formulation. In the framework of unconstrained minimization those are in the form of penalty terms added to the cost function. For some variables constructing such terms is possible, whereas for other variables the inequality constraints must be explicitly introduced, in which case the very nature of the minimization problem would be changed. Solving such problem requires notably more iterations which could be a serious drawback if the time when the results remain usable is limited.

There are even more delicate reasons. For example, if for a certain dynamical model solvability of the initial state control problem has been established, solvability of the joint state-parameter control problem is not warranted. Such problems are nonlinear even for a linear dynamical model, whereas for a nonlinear dynamical model the overall nonlinearity level would grow. This means losing convexity, decreasing the convergence radius around the global minimum, multiplying the local minima, etc. That is, the control problem becomes far more difficult to solve in practice. There is one more reason. In order to use the gradient-based unconstrained minimization we assume that the control-to-observation mapping exists and is continuous everywhere in a vicinity of the reference (true) value, i.e. the operator domain is dense around the truth and the initial guess belongs to this vicinity. In practice, some combinations of the model inputs may arise in the course of minimization such that the control-to-observation mapping does not exist. For example, in hydraulic modeling some combinations of bathymetry, friction and source terms may lead to local super-critical flow conditions. These conditions, however, are not supported by models which utilize the Preissmann discretization scheme. In this case the model execution stops and the minimization process has to be restarted from a different point. This is an additional complication to the minimization procedure which is better to avoid if possible. Moreover, due to nonlinearity of the problem, the unwanted combinations of controls cannot be easily blocked by using inequality constraints.

Taking into account all the above-mentioned reasons one may conclude that for certain DA problems the control set has to be chosen carefully. In one hand, it should provide the model predictions of a reasonable quality, on the other hand - guarantee the robustness and feasibility of the solution procedure. This is the meaning of the the notion '*control set design*'.

In the Gaussian framework, the uncertainty quantification (UQ) method for observed systems

(i.e. systems for which the posterior control estimates are available) includes two basic steps: a) computing the posterior covariance matrix of the control vector; b) computing the variance in chosen quantities of interest (QoI) using the posterior covariance and the control-to-QoI mapping. Under the assumption that the uncertainty propagation is well described by the tangent linear (TL) model (i.e. the nonlinearity of the mappings is mild or perturbations are small), the latter is actually used to represent the control-to-QoI mapping, whereas the posterior covariance is approximated by the inverse Hessian of the cost function (linearized or complete). Since such a UQ method relies on the same principles as variational DA, it seems reasonable calling it the *variational* UQ method. Combining the variational DA and variational UQ methods results into variational filtering [S3]. Recent examples of the variational UQ method being applied to different problems can be found in [S26, S25].

The method described in this section represents a generalization of the variational UQ method in the following respect. We divide the full set of the uncertainty-bearing model inputs in two parts. One part is considered as active controls (the active set), whereas the remaining inputs are fixed at their priors (the passive set). Next, we define a spatially distributed goal-function and its standard deviation (SD) as the *uncertainty measure*. Clearly, the passive set contributes to this measure both directly and via the posterior covariance of the active set. Our method allows both contributions (to the uncertainty measure) to be properly evaluated. We define a *sufficient* control set as a set for which this measure takes a value useful from the practical point of view. All possible active sets have to be examined, then ranked by the associated uncertainty measure level to reveal all sufficient control sets. The choice among these sets should be done in favor of those which will not corrupt the performance of the minimization algorithm.

The implementation of the method is matrix-free, hence it could be suitable for high-dimensional problems. Furthermore, if the Automatic Differentiation is used for producing the tangent linear and adjoint mappings, then the method could be applied to any multi-input black-box system. The method has been implemented with the full Saint-Venant hydraulic network model SIC^2 (Simulation and Integration of Control for Canals) developed at IRSTEA-Montpellier [S33].

3.4.2 Goal-function error in an observed system

Let us consider data assimilation problem formulated in §1.1. In practice, some functionals of the state are of major interest. They are usually called the Quantities of Interest (QoI). Thus, we introduce a vector of QoI, or the goal-function $\Psi = \{\Psi_i,\ i = 1, \dots, K_\Psi\} \in \mathcal{D}$, such that

$$\Psi = D(X), \tag{3–80}$$

where $\mathcal{D}$ is a 'design' space and $D : \mathcal{X} \to \mathcal{D}$ is a linear or nonlinear mapping. Because of the prediction error δX there exists the goal-function error

$$\delta\Psi = D(X) - D(\bar{X}) = D(\mathcal{M}(U)) - D(\mathcal{M}(\bar{U})). \tag{3–81}$$

This error represents uncertainty in X in a practically valuable way.

In many circumstances the level of the goal-function error $\delta\Psi$ which corresponds to the prior guess $U = U^*$ is not acceptable. The aim of data assimilation is to obtain $\hat{U} = U|Y^*$, i.e. an estimate of U conditioned on observations Y^*, which should be better than the prior U^* in the sense $\|\hat{U} - \bar{U}\| < \|U^* - \bar{U}\|$. We shall consider the system as fully/partially identifiable if the goal-function error

$$\delta\Psi = D(\mathcal{M}(\hat{U})) - D(\mathcal{M}(\bar{U})) \tag{3–82}$$

falls (fully/partially) into the margins defined by certain practical requirements.

Let us consider the goal-function error. For small errors equation (3–82) can be linearized as follows:

$$\delta\Psi = D'_X(\bar{X})\mathcal{M}'_U(\bar{U})\delta U. \tag{3–83}$$

The error δU is not known by itself, but we may know its statistical properties. Let us assume, for example, that $E[\delta\Psi] = 0$. Then, the goal-function error covariance is given by

$$P_d := E[\delta\Psi\delta\Psi^T] = D'_X(\bar{X})\mathcal{M}'_U(\bar{U})P_{\delta U}(\mathcal{M}'_U(\bar{U}))^*(D'_X(\bar{X}))^*. \tag{3–84}$$

This covariance quantifies the uncertainty in Ψ. In particular, the square roots of its diagonal elements describe the confidence interval for $\delta\Psi$. For an unobserved system $P_{\delta U}$ in (3–84) is equal to the background (prior) covariance B, whereas for an observed system (i.e. after estimation/data assimilation), the uncertainty in U is given by the estimation error covariance, i.e. $P_{\delta U} = P$.

Remark 1. In the above considerations the perfect model is assumed. This allows us to write $\bar{X} = \mathcal{M}(\bar{U})$ and, subsequently, $\bar{Y} = G(\bar{U})$. What if the model is not perfect? Let us consider, for example, a dynamic system

$$\frac{\partial\varphi}{\partial t} = \mathcal{F}(\varphi), \quad t \in (0,T), \quad \varphi|_{t=0} = u, \tag{3–85}$$

where $\mathcal{F}$ is a true spatial operator. This system is described by a model

$$\frac{\partial\varphi}{\partial t} = F(\varphi), \quad t \in (0,T), \quad \varphi|_{t=0} = u, \tag{3–86}$$

where F is an approximation (both in terms of physics and discretization) to $\mathcal{F}$. It is easy to see that the exact behavior of the system (3–85) can be modeled by equation

$$\frac{\partial\varphi}{\partial t} = F(\varphi) + \mu, \quad t \in (0,T), \quad \varphi|_{t=0} = u,$$

where $\mu = \mathcal{F}(\varphi) - F(\varphi)$ is the model error. Most certainly μ is not available for direct modeling, however, if considered as a part of the extended control vector $U = \{u, \mu\}$, it allows the mapping F to be considered 'perfect'. In this case the approach presented below is also applicable.

3.4.3 Goal-function error covariance for partial control

The theory considered so far is known and can be found in the literature (possibly, in a fragmented form). In what follows we present a new theory and a new implementation approach. That is, we have previously considered DA when the control vector U includes the full set of uncertainty-bearing model inputs. In practice, only a few selected inputs could be included (the *active set*), with the remaining inputs being fixed at their priors (the *passive set*).

Let us define the active set of the full input vector $U_a \in A$, its true value $\bar{U}_a \in A$ and the active set prior $U_a^* = \bar{U}_a + \varepsilon_a \in A$. Then $U_p = U \setminus U_a$, $\bar{U}_p = \bar{U} \setminus \bar{U}_a$ and $U_p^* = \bar{U}_p + \varepsilon_p = U^* \setminus U_a^*$, respectively. Let us assume that the background error covariance B is block-diagonal, i.e. errors ε_a and ε_p are not correlated. This is often the case naturally if the active and passive sets include different variables, otherwise the DA problem can be re-formulated in terms of uncorrelated variables. Thus, the covariance B has the following structure:

$$B = \begin{pmatrix} B_a & 0 \\ 0 & B_p \end{pmatrix},$$

where $B_a = E[\varepsilon_a\varepsilon_a^T]$ and $B_p = E[\varepsilon_p\varepsilon_p^T]$ are the sub-matrices which correspond to the active and passive sets. The goal-function error covariance P_d is given by formula (3–84). It depends on the

estimation error covariance $P_{\delta U}$, which must now correspond to a chosen decomposition of the full input set. Thus, $P_{\delta U}$ is defined in the rest of this section.

The DA problem involving the active control set consists of minimizing the cost function

$$J(U_a) = \frac{1}{2}\|R^{-1/2}(G(U_a, U_p^*) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B_a^{-1/2}(U_a - U_a^*)\|^2_A. \tag{3–87}$$

Thus, the estimate $\hat{U}_a$ is obtained from the optimality condition

$$J'_{U_a}(\hat{U}_a) = 0.$$

Given the above operator definitions in (1–8) and (1–9), the gradient of $J(U_a)$ can be expressed in the form:

$$J'_{U_a}(U_a) = (G'_{U_a}(U_a, U_p^*))^* R^{-1}(G(U_a, U_p^*) - Y^*) + B_a^{-1}(U_a - U_a^*), \tag{3–88}$$

thus the estimate $\hat{U}_a$ must satisfy the operator equation

$$(G'_{U_a}(\hat{U}_a, U_p^*))^* R^{-1}(G(\hat{U}_a, U_p^*) - Y^*) + B^{-1}(\hat{U}_a - U_a^*) = 0. \tag{3–89}$$

Let us consider an estimation error $\delta U_a = \hat{U}_a - \bar{U}_a$. We notice that $\delta U_p = U_p^* - \bar{U}_p = \varepsilon_p$. Then we obtain

$$G(\hat{U}_a, U_p^*) - Y^* = G(\hat{U}_a, U_p^*) - (G(\bar{U}_a, \bar{U}_p) + \xi) = G'_{U_a}(\tilde{U}_a, U_p^*)\delta U_a + G'_{U_p}(\bar{U}_a, \tilde{U}_p^*)\varepsilon_p - \xi,$$

where $\tilde{U}_a = \bar{U}_a + \tau_1 \delta U_a$, $\tilde{U}_p^* = \bar{U}_p + \tau_2 \varepsilon_p$, $\tau_{1,2} \in [0, 1]$ and

$$\hat{U}_a - U_a^* = (\hat{U}_a - \bar{U}_a) - (U_a^* - \bar{U}_a) = \delta U_a - \varepsilon_a.$$

Then equation (3–89) yields the error equation

$$(G'_{U_a}(\hat{U}_a, U_p^*))^* R^{-1}(G'_{U_a}(\tilde{U}_a, U_p^*)\delta U_a + G'_{U_p}(\bar{U}_a, \tilde{U}_p^*)\varepsilon_p - \xi) + B_a^{-1}(\delta U_a - \varepsilon_a) = 0. \tag{3–90}$$

Using the first order approximations $\hat{U}_a = \tilde{U}_a \approx \bar{U}_a$ and $\tilde{U}_p^* = U_p^* \approx \bar{U}_p$ we express δU_a as follows:

$$\delta U_a \simeq H_a^{-1}(\bar{U})((G'_{U_a}(\bar{U}))^* R^{-1}\xi + B_a^{-1}\varepsilon_a - (G'_{U_a}(\bar{U}))^* R^{-1} G'_{U_p}(\bar{U})\varepsilon_p), \tag{3–91}$$

where

$$H_a(\bar{U}) = (G'_{U_a}(\bar{U}))^* R^{-1} G'_{U_a}(\bar{U}) + B_a^{-1} \tag{3–92}$$

is the Hessian of an auxiliary control problem formulated for the active control set.

Since the full input vector error after DA is

$$\delta U = (\delta U_a, \varepsilon_p)^T, \tag{3–93}$$

its covariance takes the form

$$P_{\delta U} = E[\delta U \delta U^T] = \begin{pmatrix} P_{\delta U_a} & P_{\delta U_{ap}} \\ P_{\delta U_{pa}} & B_p \end{pmatrix}, \tag{3–94}$$

where

$$P_{\delta U_a} = E[\delta U_a \delta U_a^T] = H_a^{-1} + H_a^{-1}(G'_{U_a})^* R^{-1} G'_{U_p} B_p (G'_{U_p})^* R^{-1} G'_{U_a} H_a^{-1}, \tag{3–95}$$

$$P_{\delta U_{ap}} = E[\delta U_a \varepsilon_p^T] = -H_a^{-1}(G'_{U_a})^* R^{-1} G'_{U_p} B_p, \tag{3–96}$$

$$P_{\delta U_{pa}} = E[\varepsilon_p \delta U_a^T] = -B_p (G'_{U_p})^* R^{-1} G'_{U_a} H_a^{-1}. \tag{3–97}$$

All operators in (3–95)-(3–97) are taken at the point $\bar{U}$. The error covariance (3–94) must be used in (3–84) for computing the goal-function error covariance in case of partial control. Numerical tests show that using cross-terms $P_{\delta U_{ap}}$ and $P_{\delta U_{pa}}$ is absolutely vital for the method.

3.4.4 Implementation and performance assessment procedure

Let us first consider the formula for computing P_d (3–84). The product $P_d \cdot v,\ \forall v \in \mathcal{D}$ can be used for the eigenvalue analysis of matrix P_d. That is, its K_d largest eigenvalues $\lambda_{d,i}$ and the corresponding eigenvectors $W_{d,i}$ can be computed by the Lanczos method and used for constructing the limited-memory representation of P_d in the form

$$P_d = \sum_{i=1}^{K_d} \lambda_{d,i} W_{d,i} W_{d,i}^T. \tag{3–98}$$

If the elements of vector $\delta\Psi$ are strongly correlated, the number of eigenpairs required for meaningful representation of P_d (and its diagonal elements, in particular) could by surprisingly small as compared to K_Ψ (the dimension of vector Ψ). The same is true for the number of Lanczos iterations needed for evaluating those eigenpairs.

The products $D'_X(\bar{X})\mathcal{M}'_U(\bar{U}) \cdot v,\ \forall v \in \mathcal{U}$ and $(\mathcal{M}'_U(\bar{U}))^*(D'_X(\bar{X}))^* \cdot v,\ \forall v \in \mathcal{D}$ are computed by calling the tangent linear and adjoint models of the corresponding mappings D and $\mathcal{M}$. Thus, we have to define $P_{\delta U} \cdot v,\ \forall v \in \mathcal{U}$. This is done in below.

Proposition 1. Consider the projected Hessian in the form (2–10), which has the following partition:

$$\tilde{H} = (B^{1/2})^* H B^{1/2} = \begin{pmatrix} \tilde{H}_a & \tilde{H}_{ap} \\ \tilde{H}_{pa} & \tilde{H}_p \end{pmatrix}. \tag{3–99}$$

Then, the blocks of matrix $P_{\delta U}$ can be expressed via blocks of $\tilde{H}$

$$P_{\delta U_a} = B_a^{1/2} \tilde{H}_a^{-1/2} (I_a + \tilde{H}_a^{-1/2} \tilde{H}_{ap} \tilde{H}_{pa} \tilde{H}_a^{-1/2}) \tilde{H}_a^{-1/2} (B_a^{1/2})^*, \tag{3–100}$$

$$P_{\delta U_{ap}} = -B_a^{1/2} \tilde{H}_a^{-1} \tilde{H}_{ap} (B_p^{1/2})^*, \tag{3–101}$$

$$P_{\delta U_{pa}} = -B_p^{1/2} \tilde{H}_{pa} \tilde{H}_a^{-1} (B_a^{1/2})^*. \tag{3–102}$$

For $v = (v_a, v_p)^T$, we obtain

$$P_{\delta U} \cdot v = \begin{pmatrix} P_{\delta U_a} \cdot v_a + P_{\delta U_{ap}} \cdot v_p \\ P_{\delta U_{pa}} \cdot v_a + B_p \cdot v_p \end{pmatrix}, \tag{3–103}$$

where the operators $P_{\delta U_a}$, $P_{\delta U_{ap}}$ and $P_{\delta U_{pa}}$ are defined in (3–100)-(3–102). Implementation of the above formulas requires, in turn, the products $\tilde{H}_{pa} \cdot v_a$, $\tilde{H}_{ap} \cdot v_p$ and $\tilde{H}_a^\gamma \cdot v_a$ for $\gamma = -1, -1/2$. These can be defined given the eigenpairs $\{\lambda_i, W_i\},\ i = 1, \ldots, K$ of $\tilde{H}$.

Proposition 2. Let N_d be the total number of all feasible active sets constructed from the full set of model inputs U, and n - a running active set index. Then, the performance assessment procedure consists of the following steps:

Algorithm 1

1. compute by the Lanczos method and store in memory:
$\{\lambda_i, W_i\},\ i = 1, \ldots, K$ of $\tilde{H}(\bar{U})$ in (2--10)
2. for $n = 1, \ldots, N_d$
   a. compute by the Lanczos method and store in memory:
   $\{\lambda_{a,i}, W_{a,i}\},\ i = 1, \ldots, K_a$ of $\tilde{H}_a^{(n)}(\bar{U})$ defined via $\{\lambda_i, W_i\}$
   b. compute by the Lanczos method and store in memory:

$\{\lambda_{d,i}^{(n)}, W_{d,i}^{(n)}\}$, $i = 1, \ldots, K_d$ of $P_d^{(n)}$ `defined in (3--98), using` $P_{\delta U}^{(n)} \cdot v$ `in (3--103)`
`end` n

Remark 2. Step 1 of the Algorithm enables $\tilde{H}_a \cdot v_a$, $\tilde{H}_{pa} \cdot v_a$ and $\tilde{H}_{ap} \cdot v_p$ to be evaluated when necessary. After step 2a, for any chosen active control set we are able to compute $H_a^\gamma \cdot v_a$ and, correspondingly, $P_{\delta U} \cdot v$ using (3–103). At steps 1 and 2b we solve the sequence of the tangent linear and adjoint models, which is, in case of using models based on partial differential equations, the most expensive part in terms of the CPU time. Step 2a requires algebraic computations only. The diagonal elements of $P_d^{(n)}$ can be retrieved on the basis of representation (3–98). Square roots of these elements (standard deviation) are the sought outcome of the algorithm above.

Remark 3. The outcome of the **Algorithm 1** is a set of $P_d^{(n)}$, $n = 1, \ldots, N_d$, each being defined in the limited-memory form via its eigenpairs $\{\lambda_{d,i}^{(n)}, W_{d,i}^{(n)}\}$, $i = 1, \ldots, K_d$. Based on $P_d^{(n)}$ the active sets can be ranked and the sufficient (in terms of the accuracy) sets can be chosen. The choice between sufficient sets has to made considering other criteria.

3.4.5 Illustration

Here we present an example from river hydraulics. Estimating river discharges from *in-situ* and/or remote sensing data is a key component for evaluation of water balance at local and global scales and for water management. A distinctive feature of the river discharge estimation problem is the likely presence of significant uncertainty in parameters defining basic properties of a hydraulic model, such as bathymetry (surface topography), friction, infiltration level, etc. There are also unaccounted lateral tributaries/offtakes and storage areas. Since the discharge estimation problem is considered, the active set must undoubtedly include the inflow discharge at a chosen upstream location (the inlet). Would it be a sufficient control set? If not, what other model inputs should be included into the active set to reduce the impact of uncertainties? Do we have enough data? Indeed, *in-situ* measurements of water elevation and discharge are relatively rare on most rivers because of limited accessibility and associated costs, whereas the satellite data can be sparse in time and far less accurate. Therefore, designing the control set is a key issue for solving the river discharge estimation problem.

Here we consider one reach of a river, which is discretized in space giving a set of nodes with longitudinal abscissas x_i, $i = 1, \ldots, N$. At each node a hydraulic cross-section S_i is defined by a set of points on a plane $\vec{n}_k \cdot (\vec{r} - \vec{r}_k) = 0$ describing the bed profile, which are evaluated from a design sketch or from a topographical survey. For each section this data allows us to compute for any given water level line Z: the wetted area function $A(Z, p_g)$, the wetted perimeter function $P(Z, p_g)$, the hydraulic radius function $R(Z, p_g)$ and the top width function $L(Z, p_g)$, where p_g are geometric parameters of the corresponding cross-section. For a given reach, p_g is a function of x. For a 'regular' section, the shallow water flow in the longitudinal direction x is described by the Saint-Venant equations:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = Q_L, \tag{3–104}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial Q^2/A}{\partial x} + gA\frac{\partial Z}{\partial x} = -gAS_f + C_k Q_L v, \tag{3–105}$$

$$t \in (0, T],$$

where $Q(x,t)$ is the discharge, $Z(x,t)$ is the water level, $v(x,t) = Q/A$ is the mean velocity, $Q_L(x,t)$ is the lateral discharge, $C_k(x)$ is the lateral discharge coefficient and S_f is the friction term dependent

on the Strickler coefficient $C_s(x)$ and on the hydraulic radius $R(Z, p_g)$:

$$S_f = \frac{Q|Q|}{C_s^2 A^2 R^{4/3}}.$$

For the upstream nodes we usually use the inflow discharge $\mathcal{Q}(t)$, whereas for the downstream node it is the rating curve $Q = f(Z, p_{rc})$, where p_{rc} are the rating curve parameters.

Let us assume that we are looking for discharge $\mathcal{Q}(t)$ under uncertainties in geometric parameters p_g and in the Strickler coefficient $C_s(x)$, whereas all other parameters of the model (3–104)-(3–105) are known precisely. The particular geometric parameters include the local bed elevation $z_b(x)$ and the lateral dilation coefficient $b(x)$. The latter scales uniformly a given cross-section geometric dimensions in width. Then, the full control vector $U \in \mathcal{U}$ looks as follows:

$$U = (\mathcal{Q}(t), C_s(\bar{x}), z_b(\bar{x}), b(\bar{x}), U_*)^T,\ \bar{x} = (x_1, \ldots, x_N)^T, \tag{3–106}$$

where U_* stands for all remaining model inputs which contain no uncertainty. The above hydraulic equations are implemented in SIC^2, which is the full nonlinear Saint-Venant hydraulic network model. The routine which maps U into Y represents operator G. The tangent linear model (TLM) and the adjoint model, which represent operators G' and $(G')^*$, respectively, are produced by means of the AD engine TAPENADE [S20] applied to the main computational routine of the SIC^2 package (the forward model).

Let us assume that the water surface elevation $Z(x,t)$ is measured at the specified sections

$$Y = C(Z, Q) = \{Z(x_i, t), i \in I_o\}, \tag{3–107}$$

where I_o is the array of indices of the sections where the measurements are available. The goal-functions (QoI) are usually some functionals of the state trajectory. In hydraulics, certain quantities useful in the flood risk assessment may be of interest, such as the maximum water surface elevation above a given (safe) threshold at some locations, for example. Thus, we consider the goal-function in the form

$$\Psi = D(Q(x,t), Z(x,t)) = (\Psi_Q(x), \Psi_Z(x))^T,$$

where

$$\Psi_Q(x) = \int_{t_1}^{t_2} |Q(x,t) - Q_*(x)| dt,\ \ \Psi_Z(x) = \int_{t_1}^{t_2} |Z(x,t) - Z_*(x)| dt, \tag{3–108}$$

where Q_* and Z_* are some reference discharge and elevation levels, and t_1, t_2 define the time window of interest. Thus, the covariance matrix P_d includes two blocks $P_d[Q]$ and $P_d[Z]$, associated with different goal-functions (QoI) in (3–108); the corresponding standard deviation vectors are denoted $\sigma_d[Q]$ and $\sigma_d[Z]$. All results below are presented in terms of $\sigma_d[\cdot]$.

Here we consider the following active control sets:
*cc*0 - no control case, i.e. $U_a = \{\emptyset\}$;
*cc*1 - full control case: i.e. $U_a = U \setminus U^*$;
*cc*2 - partial control case: inflow discharge only, i.e. $U_a = \mathcal{Q}(t)$;
*cc*3 - partial control case: inflow discharge and bed elevation, i.e. $U_a = (\mathcal{Q}(t), z_b(k))^T$;
*cc*4 - partial control case: inflow discharge and Strickler coefficient, i.e. $U_a = (\mathcal{Q}(t), C_s(k))^T$.
In the 'no control case' $P_{\delta u} = B$.

These figures reveal the following interesting features:
1. In the full control case *cc*1, the posterior standard deviations are strictly smaller than the background values, compare the curves marked *cc*1 to those marked *cc*0. This difference is usually referred as the 'uncertainty reduction'. The curves marked *cc*1 show the minimum level of $\sigma_d[\cdot]$ that can be

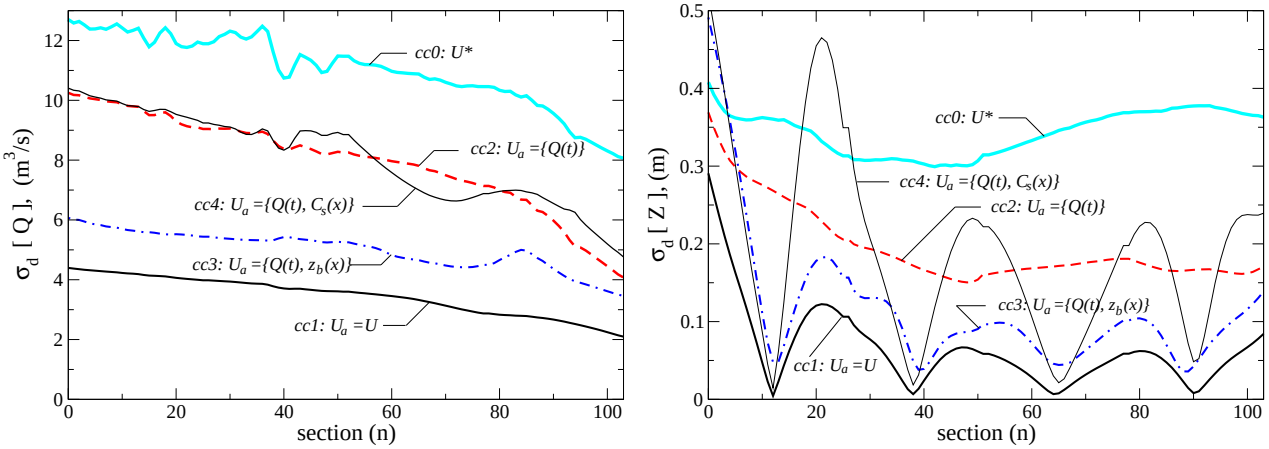

Figure 3.5: Standard deviation of the goal-vector (QoI).

achieved with the available observations. Since estimating the discharge is the major task, let us pay attention to Fig.3.5(left). For the partial control case $cc2$ (the inflow discharge control only), $\sigma_d[\cdot]$ are presented in dashed lines. One can see that the uncertainty reduction achieved in this case makes only a fraction (30-40%) of the one achieved in the full control case. Thus, controlling the discharge only is hardly sufficient.

2. The control set can be extended, however this must be done with caution. It seems reasonable to add such control variable that the resulting uncertainty in the goal-function (QoI) would be most essential. At the same time this extension should not affect the reliability of the minimization process. By comparing results for different active control sets one can see that such component does exist: it is the bed elevation z_b, see the results in dash-dotted line, case $cc3$. For comparison we also present the partial control case $cc4$, where instead of z_b we use C_s. One can notice that, in terms of $\sigma_d[\cdot]$, the effect of inclusion C_s into the active set is negligible. Some improvement can be seen in terms of $\sigma_d[Z]$ in the vicinity of sensors. It has been repeatedly observed that including both z_b and C_s into the active control set leads to unsupported combinations of controls, see [A6]. Thus, the sufficient control set is given by vector $U_a = (\mathcal{Q}(t), z_b(k))^T$.

3.4.6 Conclusions

In this section we describe the *control set design* concept, which offers an additional direction in design of DA systems. Indeed, it is more usual to talk about design of observations. More generally, one could talk about optimal experimental design, which also includes the possibility to influence the system behavior. In hydraulics this can be done using operational devices (gates, weirs, etc). The need for the *control set design* step arises in solving DA problems for models with multiple heterogeneous inputs containing significant uncertainties. In one hand, it looks appealing to include all the uncertainty-bearing inputs into the (active) control set. On the other hand, there are different reasons against such a straightforward approach. Some of them are discussed in the beginning of this section. In order to design the control set one must be able to quantify its performance in terms of the uncertainty level in specially chosen goal-functions (QoI). Those sets for which this level is acceptable from the practical point of view are called 'sufficient'. Technically, the method is a generalization of the standard variational UQ method in a sense that the full set of the uncertainty-bearing model inputs is divided into the active and passive sets, each affecting the goal-function uncertainty covariance in different

ways. The implementation is 'matrix free' in the sense that the limited-memory representations of the operators are used. These are constructed by means of the Lanczos procedure. Because of this, the method may be suitable for high-dimensional problems, though it still depends on the eigenvalue distribution of the operators involved. Let us note that the variational UQ method is only valid for mildly nonlinear models.

3.4.7 Future developments

There are two directions of an immediate future development. First of all, the presented control set design approach can be easily extended to include the 'idle'/implicit control option (see §3.3), i.e. the full set of model inputs U can be divided into the 'active', 'idle' and 'passive' subsets (currently 'active' and 'passive' only). This extension is rather straightforward.

The idea of another development is outlined below. One can conclude from the the results of the previous section that controlling the bed elevation $z_b(x)$ allows the uncertainties in other model parameters (the Strickler coefficient $C_s(x)$ and the dilation parameter $b(x)$) to be successfully absorbed. This shows that the control set design procedure should not be limited to a choice of optimal combinations of the model input parameters. In more general terms, the control design task can be formulated as follows:
find control which guaranties that the model has sufficient degrees of freedom, so its trajectory can be driven, up to a given discrepancy level, toward observations.

Here we present the theory of a generalized control set design procedure. Let us introduce auxiliary inputs $V \in \mathcal{V}$, where $\mathcal{V}$ is an auxiliary input space, and re-define the model input as follows

$$U \to U + S(V). \tag{3–109}$$

Here $S : \mathcal{V} \to \mathcal{U}_\triangle$, where $\mathcal{U}_\triangle \subseteq \mathcal{U}$ is a mapping from the auxiliary space into the chosen subset of the original full input set (inverse to S may not exist). By $\bar{V}$ and V^* we denote the 'true' and the background values of V, respectively. It is most important to realize that, since V does not originally exist in $\mathcal{U}$, then $\bar{V} = V^* = 0$. We shall call V - the '*integrated controls*'.

Let us now define $\mathbf{U} = (U, V)^T$ - the generalized input vector, $Y = G(\mathbf{U}) = G(U, V)$ - the 'input-to-observations' mapping. As before $\bar{Y} = G(\bar{\mathbf{U}}) = G(\bar{U}, 0)$ - 'true'/exact observations and $Y^* = \bar{Y} + \xi$. Then, we can formulate the DA cost-function as follows:

$$J(\mathbf{U}) = J(U, V) = \frac{1}{2}\|R^{-1/2}(G(U, V) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B^{-1/2}(U - U^*)\|^2_{\mathcal{U}} + \frac{1}{2}\|B_V^{-1/2}V\|^2_{\mathcal{V}}, \tag{3–110}$$

where

$$B_V = S'^*_V(0) B_\triangle S'_V(0).$$

Now, we are going to consider U as passive control (i.e. fix U at its background value) and investigate how the error $\delta U = U^* - \bar{U} = \varepsilon$ is absorbed by V. To this end we formulate the following DA problem

$$J(\mathbf{U}) = J(U^*, V) = \frac{1}{2}\|R^{-1/2}(G(U^*, V) - Y^*)\|^2_{\mathcal{Y}} + \frac{1}{2}\|B_V^{-1/2}V\|^2_{\mathcal{V}}, \to \inf_V. \tag{3–111}$$

Our task now is to obtain the expression for the estimation error $\delta V = \hat{V} - \bar{V} = \hat{V}$. Here we follow the logic of §3.4.3. The optimality condition for $\hat{V}$ reads

$$J'_V(U^*, \hat{V}) = 0.$$

The optimality condition yields the estimator equation

$$G_V'^*(U^*,\hat{V})R^{-1}(G(U^*,\hat{V})-Y^*)+B_V^{-1}\hat{V}=0,$$

and, subsequently, the error equation

$$G_V'^*(U^*,\hat{V})R^{-1}\left(G_V'(U^*,\tilde{V})\hat{V}+G_U'(\tilde{U},0)\delta U-\xi\right)+B_V^{-1}\hat{V}=0 \tag{3–112}$$

where $\tilde{V}=\tau_1\hat{V}$, $\tilde{U}=\tau_2(U^*-\bar{U})$, $\tau_{1,2}\in[0,1]$. From (3–112) we obtain

$$\hat{V}={H_V}^{-1}(\bar{U},0)G_V'^*(\bar{U},0)R^{-1}(\xi-G_U'(\bar{U},0)\delta U), \tag{3–113}$$

where

$$H_V(\cdot,\cdot)=G_V'^*(\cdot,\cdot)R^{-1}G_V'(\cdot,\cdot)+B_V^{-1}. \tag{3–114}$$

Consider the goal-function (QoI) in the form (3–81). For the goal-function error one can write

$$\delta\Psi=D(\mathcal{M}(U^*,\hat{V}))-D(\mathcal{M}(\bar{U},0))=D_X'(\bar{X})\left(\mathcal{M}_V'(\bar{U},0)\hat{V}+\mathcal{M}_U'(\bar{U},0)\delta U\right). \tag{3–115}$$

Subsequently, for small $\delta\mathbf{U}=(\delta U,\hat{V})^T=(\varepsilon,\hat{V})^T$, the covariance of $\delta\Psi$ is defined as

$$P_d:=E[\delta\Psi\delta\Psi^T]=D_X'(\bar{X})\mathcal{M}_{\mathbf{U}}'(0,\bar{U})\;P_{\delta\mathbf{U}}\;(\mathcal{M}_{\mathbf{U}}'(0,\bar{U}))^*(D_X'(\bar{X}))^*, \tag{3–116}$$

where

$$P_{\delta\mathbf{U}}=E[\delta\mathbf{U}\delta\mathbf{U}^T]=\begin{pmatrix}E[\varepsilon\varepsilon^T] & E[\varepsilon\hat{V}^T]\\ E[\hat{V}\varepsilon^T] & E[\hat{V}\hat{V}^T]\end{pmatrix}. \tag{3–117}$$

Consider $H_{\mathbf{U}}$ - the Hessian of the auxiliary cost-function associated with (3–110), and $B_{\mathbf{U}}$ - the background covariance of $\delta\mathbf{U}$. These have the following block-structure:

$$H_{\mathbf{U}}=\begin{pmatrix}H_U & H_{UV}\\ H_{VU} & H_V\end{pmatrix},\quad B_{\mathbf{U}}=\begin{pmatrix}B & 0\\ 0 & B_V\end{pmatrix}.$$

Then, the blocks of the covariance matrix (3–117) can be expressed via the blocks of the Hessian $H_{\mathbf{U}}$ as follows:

$$E[\varepsilon\varepsilon^T]=B,$$
$$E[\hat{V}\hat{V}^T]=H_V^{-1}-{H_V}^{-1}\left(B_V^{-1}-H_{UV}BH_{VU}\right){H_V}^{-1},$$
$$E[\varepsilon\hat{V}^T]=-BH_{UV}{H_V}^{-1},$$
$$E[\hat{V}\varepsilon^T]=-{H_V}^{-1}H_{VU}B.$$

This Hessian can be known in the limited-memory form

$$H_{\mathbf{U}}^{\gamma}(\bar{\mathbf{U}})\simeq B_{\mathbf{U}}^{1/2}\left(I+\sum_{i=1}^{L_H}(\lambda_i^{\gamma}-1)W_iW_i^*\right)B_{\mathbf{U}}^{T/2}.$$

As before, the error covariance (3–117) must be used in (3–116) for computing the leading eigenpairs of the goal-function error covariance. Let us mention that our derivation essentially replicates the one described in §3.4.3, but for an extended input vector.

The presented approach provides a general tool for assessing the performance of any chosen control set. It may include those ones which do not match the original model inputs. For example, it can

be a combination of the original control inputs, i.e. invariants, such as characteristic variables at 'open'/'liquid' boundaries of local fluid flow models. It can also be a small subset of a high-dimensional input, such as distributed source term (model error) $\eta(x,t)$. In particular, when considering $\eta(x,\bar{t})$, where $\bar{t}$ represents a few chosen time instants, one replicates the sub-window technique for the model error treatment suggested in [S46]. When considering $\eta(\bar{x},t)$, where $\bar{x}$ represents the observation points, one replicates the optimal nudging technique described in [S52]. Let us note that no sensible error assessment for these two methods have ever been reported, though both methods are currently used in operational forecasting.

Chapter 4

Advanced numerical approaches for computing inverse Hessian

4.1 Computing of the inverse Hessian: multigid approach

4.1.1 Introduction

The importance of the Hessian matrix and its inverse in variational DA for geophysical applications is underlined in [S44], although, for decades, this has been a routine knowledge in statistics (see, for example, [S42]). Some relevant applications of the inverse Hessian are highlighted in §1.2.5. These applications involve either solving multiple systems of linear equations involving H, or having access to the inverse operator H^{-1}. In practice, an explicit discrete representation of H is never required, since the Hessian-vector product can be obtained by successively applying operators in formula (1–21). The development of feasible methods for generation, storage and subsequent use of H^{-1} or $H^{-1/2}$ in this framework are not well understood: this is the prime motivation for our interest in the development of efficient algorithms for computing and managing the inverse Hessian, such as those presented in [A4].

A special feature of working with the Hessian for very high-dimensional problems is that neither the Hessian nor its inverse can be directly accessed in matrix form. While the Hessian-vector product can be computed by solving a sequence of the tangent linear and adjoint problems, no such option exists for defining the inverse Hessian-vector product (or the inverse square root Hessian-vector product, which is also relevant in many applications). One obvious approach is therefore to consider limited-memory schemes for computing and storing the inverse Hessian (or its square root). In this context, the idea of a multilevel framework becomes relevant. This is due to the fact that the inverse Hessian is, in essence, an approximation of the posterior covariance matrix and, if the initial flow field is considered as a spatially distributed control, then the correlations of different lengths between the flow field values can be described at different levels of spatial discretization.

Multigrid methods were initially developed for solving elliptic partial differential equations (PDEs) and have since been extended for solving PDEs of different types. One key modern area of application of multigrid methods is in solving PDE-constrained optimization or inverse problems, see the review paper [S7]. Here the multigrid solver is applied directly to the optimality system, which includes the original model, its adjoint and the optimality condition. Some elements of the multigrid approach have been utilized previously in variational data assimilation algorithms in meteorology and oceanography, but a complete multigrid algorithm has been considered only recently in [S14]. Multigrid methods can also used for solving eigenvalue problems, which is most relevant to the method described in this section. The usual multigrid approach in this context is to treat the eigenvalue problem as a non-linear equation and apply a non-linear multigrid solver [S10, S24]. Alternatively, an outer eigenvalue solver

such as Rayleigh quotient iteration can be employed, which requires the solution of systems of linear equations with a shifted coefficient matrix using multigrid as an inner solver [S40]. A third approach uses a standard eigenvalue solver (such as Lanczos or Arnoldi) with multigrid as a preconditioner. This type of method is reviewed in [S27].

In this section we describe a general multilevel eigenvalue decomposition of a given symmetric operator. Given its spectral decomposition, an operator A, say, can be approximated by a finite number of its eigenvalues and eigenvectors. To achieve a desired approximation quality (in terms of a specified distance between the exact and approximated operators) a certain number of eigenpairs must be used, dependent on the eigenvalue distribution. However, for high-dimensional problems, the computationally feasible number of eigenpairs (in terms of available storage, for example) may be too small to achieve any useful approximation quality. Thus, a single level eigenvalue decomposition approach has its limitations. Our proposed *multilevel eigenvalue decomposition algorithm* involves an outer multilevel loop that provides an incomplete eigenvalue decomposition (using Lanczos) of the operator at each level, resulting in a final approximation involving eigenpairs associated with each discretization level. Note that, if A is not symmetric, the same technique could be applied to $A^T A$.

The multilevel technique described in §4.1.2 allows us to build limited-memory approximations to A^{-1} and $A^{-1/2}$ which, within a fixed memory framework, are much better than their single-level spectral counterparts. These approximations could be used in many situations, for example, as preconditioners for solving systems of linear equations, across multiple application areas. In [A4], we use the technique as an efficient way of approximating the inverse Hessian in variational data assimilation. We also introduce a second idea, namely, decomposition of the Hessian into local sensor-based Hessians. Although this is distinct from the multilevel eigenvalue decomposition, the latter provides a framework for its practical implementation.

4.1.2 Multilevel eigenvalue decomposition algorithm

In this section we describe an algorithm for constructing a multilevel approximation to the inverse (and its square root) of a general symmetric positive definite operator A associated with a model $\mathcal{M}$ based on partial differential equations, e.g. $\tilde{H}(\cdot)$ in (2–10). The key idea can be summarized as follows. Consider a limited-memory approximation to A^β of the form in (2–15) (with $\beta = -1$). If we assume that A is only available in operator-vector product form, that is, we can evaluate Av for some discrete function v on the underlying computational grid, the eigenvalues required can be calculated using the Lanczos method. Given a sequence of nested grids, a conceptual outline of the recursive multilevel process is as follows:

1. represent A on the coarsest grid level;
2. use a local preconditioner to improve the eigenvalue distribution;
3. build a limited memory approximation to its inverse, which forms the basis of the local preconditioner at the next coarsest level;
4. move up one grid level and repeat.

As proof of concept, we describe the ideas in a one-dimensional setting. However, the concept is equally valid for two- and three-dimensional problems.

Multilevel grid structure

We consider the spatial domain $\Omega = [0, 1]$ and construct a sequence of grids for discretizing the model $\mathcal{M}$. We suppose that the base grid on which the problem is defined has a uniform distribution of $m_0 = m + 1$ grid points, and use this as our finest grid (denoted by grid level $k = 0$). We form the next grid by removing a grid point from in between each pair of existing points, to give a grid at level $k = 1$ with $m_1 = m/2 + 1$ uniformly-spaced points. Continuing this refinement leads to a sequence of grids at levels $k = 0, 1, 2, \ldots, k_c$ (where k_c is the coarsest grid level), with the grid at level k containing $m_k = m/2^k + 1$ grid points.

Grid transfer operators

We introduce the prolongation operator $S_{k,k-i}$, $0 \leq i \leq k$, which maps (interpolates) a discrete function v_k defined at grid level k to a finer grid level $k - i$ (or to the same grid for $i = 0$). That is, the operator satisfies

$$v_{k-i} = S_{k,k-i}\, v_k, \qquad S_{k,k} = I_k, \tag{4–1}$$

where I_k is the identity operator at grid level k. Similarly, the restriction operator $S^*_{k,k-i}$ maps a discrete function v_{k-i} defined at grid level $k - i$ to a coarser grid level k. That is,

$$\tilde{v}_k = S^*_{k,k-i}\, v_{k-i}, \qquad S^*_{k,k} = I_k, \tag{4–2}$$

where $S^*_{k,k-i}$ is the adjoint operator to $S_{k,k-i}$. Combining operators (4–1) and (4–2) gives $\tilde{v}_k = S^*_{k,k-i} S_{k,k-i}\, v_k$, so $\tilde{v}_k = v_k$ only if

$$S^*_{k,k-i} S_{k,k-i} = I_k. \tag{4–3}$$

In other words, for $\tilde{v}_k = v_k$ we require $S_{k,k-i}$ to be an orthonormal projection: we will refer to this situation as *perfect interpolation*.

Proposition 1. Let A_k be a discrete representation of A defined at grid level k. Then, the projection of A_k at a finer grid level $k - i$, $0 \leq i \leq k$ is given by

$$\begin{aligned} P_{k-i}(A_k) &= S_{k,k-i}(A_k - I_k)S^*_{k,k-i} + I_{k-i}, \qquad 0 < i \leq k, \\ P_k(A_k) &= A_k, \end{aligned} \tag{4–4}$$

and the projection of an operator A_{k-i} at a coarser grid level $k, 0 \leq i \leq k_c$ is given by

$$\begin{aligned} Q_k(A_{k-i}) &= S^*_{k,k-i}(A_{k-i} - I_{k-i})S_{k,k-i} + I_k, \qquad 0 < i \leq k_c, \\ Q_k(A_k) &= A_k. \end{aligned} \tag{4–5}$$

Proposition 2. Adjoint operators P^* and Q^* satisfy the conditions

$$P^*_{k-i}(A_k) = P_{k-i}(A^*_k), \;\; Q^*_k(A_{k-i}) = Q_k(A^*_{k-i}). \tag{4–6}$$

Proposition 3. In the case of perfect interpolation the operator $P_{k-i}(A_k)$ has the important property that $P_{k-i}(A_k) = P_{k-i}(A_k^{1/2})P_{k-i}((A_k^{1/2})^*)$. Also, the expression (4–5) simplifies to $Q_k(A_{k-i}) = S^*_{k,k-i} A_{k-i} S_{k,k-i}$. However, $Q_k(A_{k-i}) \neq Q_k(A_{k-i}^{1/2})Q_k((A_{k-i}^{1/2})^*)$.

Multilevel algorithm: structure

We now develop an algorithm for constructing a multilevel limited-memory representation of A^{-1} (and $A^{-1/2}$) in operator-vector product form, that is, as $A^{-1}v$ or $A^{-1/2}v$. This is achieved by separating the eigensystem of operator A into the subsystems associated with different representation levels. We assume that the operator-vector product is available at the finest grid level $k = 0$, that is, we have available $A_0 v_0$ for some fine grid function v_0. Here we describe our algorithm based on a sequence of coarser grids $k = 1, 2, \ldots, k_c$, where k_c is the coarsest level. A key ingredient of this algorithm is a sequence of local preconditioners applied at each grid level.

We begin by representing the finest grid operator A_0 on level k as $Q_k(A_0)$ using (4–5), then precondition this to obtain

$$\tilde{Q}_k(A_0) = T^*_{k,k+1} Q_k(A_0) T_{k,k+1}. \tag{4–7}$$

The level k preconditioner $T_{k,k+1}$ will be chosen so that the eigenvalues of $\tilde{Q}_k(A_0)$ are closer to unity than those of $Q_k(A_0)$: details of how this is done follow in §4.1.2. We then use the Lanczos method to compute a specified number, n_k say, of the largest eigenvalues of $\tilde{Q}_k(A_0)$ (measured in a log-squared sense), together with their associated eigenvectors. The resulting n_k eigenpairs $\{\lambda^i_k, W^i_k\}$, $i = 1, \ldots, n_k$, are then used to construct a limited memory approximation to $\tilde{Q}_k(A_0)$, namely,

$$\hat{Q}_k(A_0) = I_k + \sum_{i=1}^{n_k} (\lambda^i_k - 1) W^i_k (W^i_k)^T. \tag{4–8}$$

Note that, as in (2–15), an approximation to $\tilde{Q}_k(A_0)$ raised to any chosen power β is readily available. This means that $\hat{Q}^{-1}_k(A_0)$, $\hat{Q}^{1/2}_k(A_0)$ and, most importantly for the preconditioners defined in the next section, $\hat{Q}^{-1/2}_k(A_0)$, are easily computed.

The accuracy of approximation (4–8) is clearly critically affected by the number of eigenvectors which are calculated and stored at each grid level. To facilitate later investigation of how these values should be chosen, we introduce the notation

$$N_e = (n_0, n_1, \ldots, n_{k_c}), \qquad \widehat{N_e} = \sum_{k=0}^{k_c} n_k \tag{4–9}$$

for the vector containing these values for a particular approximation and the sum of its entries.

Level k preconditioners

The algorithm above involves a preconditioner $T_{k,k+1}$ for $Q_k(A_0)$ local to the current grid level. The motivation for our choice of $T_{k,k+1}$ is the assumption that $Q_{k+1}(A_0)$ is a good approximation to $Q_k(A_0)$, so we can use the projection of the former to grid level k to precondition the latter. That is, we expect that the eigenvalues of the preconditioned operator

$$P_k(Q^{-1/2}_{k+1}(A_0)) Q_k(A_0) P_k(Q^{-1/2}_{k+1}(A_0))$$

to be clustered around 1. Furthermore, it can be seen from (4–7) and (4–8) that

$$Q^{-1}_k(A_0) = T_{k,k+1} \tilde{Q}^{-1}_k(A_0) T^*_{k,k+1} \approx T_{k,k+1} \hat{Q}^{-1}_k(A_0) T^*_{k,k+1} \tag{4–10a}$$

and so

$$Q^{-1/2}_k(A_0) = T_{k,k+1} \tilde{Q}^{-1/2}_k(A_0) \approx T_{k,k+1} \hat{Q}^{-1/2}_k(A_0). \tag{4–10b}$$

Note that with this notation, preconditioner $T_{k,k+1}$ is applied on level k, using information projected from level $k+1$.

The above considerations are valid for grid levels $k = 0, \ldots, k_c - 1$. On the coarsest grid level $k = k_c$, grid level $k_c + 1$ does not exist, so we set $T_{k_c,k_c+1} = I_{k_c}$. Note also that on the finest grid level $k = 0$ we can use A_0 directly, that is, $Q_0(A_0) \equiv A_0$. In practice, moving from the coarsest to the finest grid, we accumulate the eigenpairs $\{\lambda_k, W_k\}$, $k = k_c, \ldots, 0$, in (4–8) which allows us to define the required products $A_0^{-1}v_0$ and $A_0^{-1/2}v_0$ via a recursive algorithm as follows. At a general grid level k, the preconditioner is constructed in a recursive way using information from the previous (coarser) grid levels via

$$T_{k,k+1} = \begin{cases} P_k(T_{k+1,k+2}\hat{Q}_{k+1}^{-1/2}(A_0)) & k = 0, 1, \ldots, k_c - 1; \\ I_{k_c}, & k = k_c; \end{cases}, \tag{4–11a}$$

$$T^*_{k,k+1} = \begin{cases} P_k(\hat{Q}_{k+1}^{-1/2}(A_0)T^*_{k+1,k+2}), & k = 0, 1, \ldots, k_c - 1; \\ I_{k_c}, & k = k_c; \end{cases} \tag{4–11b}$$

where

$$\hat{Q}_k^{-1/2}(A_0) = I_k + \sum_{i=1}^{n_k}((\lambda_k^i)^{-1/2} - 1)W_k^i(W_k^i)^*$$

(cf. (4–8)).

Summary

Using the above definitions, an operator representing the multilevel eigenvalue decomposition algorithm can be constructed as follows:

Multilevel Eigenvalue Decomposition (MLEVD) Algorithm

$[\Lambda, \mathcal{U}] = MLEVD(A_0, N_e)$

```
for   k = k_c, k_c − 1, ..., 0
   compute by the Lanczos method and store in memory:
   {λ_k^i, U_k^i}, i = 1, ..., n_k of Q̃_k(A_0) in (4--7)
   using T_{k,k+1} and T*_{k,k+1} from (4--11)
end
```

The input to this algorithm is A_0 (available in the form of an operator-vector product A_0v_0 at the finest level) together with the vector N_e in (4–9) containing the number of eigenpairs to be calculated at each grid level. At a current level $k < k_c$ the algorithm utilizes the eigenpairs obtained at the previous (coarser) levels. The output is a pair of vectors $[\Lambda_0, \mathcal{W}_0]$ containing the multilevel eigenstructure of A_0. These vectors can be represented as follows:

$$\begin{aligned} \Lambda_0 &= \left[\lambda_{k_c}^1, \ldots, \lambda_{k_c}^{n_{k_c}}, \lambda_{k_c-1}^1, \ldots, \lambda_{k_c-1}^{n_{k_c-1}}, \ldots, \lambda_0^1, \ldots, \lambda_0^{n_0}\right], \\ \mathcal{W}_0 &= \left[W_{k_c}^1, \ldots, W_{k_c}^{n_{k_c}}, W_{k_c-1}^1, \ldots, W_{k_c-1}^{n_{k_c-1}}, \ldots, W_0^1, \ldots, W_0^{n_0}\right]. \end{aligned} \tag{4–12}$$

Given $[\Lambda_0, \mathcal{W}_0]$, for any function v_k the products $Q_k^{-1}(A_0)v_k$ and $Q_k^{-1/2}(A_0)v_k$ can be recovered using (4–10a) and (4–10b), involving (4–11). In particular, for any fine grid function v_0 we can evaluate

$$A_0^{-1}v_0 \approx Q_0^{-1}(A_0) = T_{0,1}\hat{Q}_0^{-1}(A_0)T^*_{0,1}v_0, \tag{4–13a}$$

$$A_0^{-1/2} v_0 \approx Q_0^{-1/2}(A_0) = T_{0,1} \hat{Q}_0^{-1/2}(A_0) v_0, \tag{4–13b}$$

were $T_{0,1}$ and $T_{0,1}^*$ are defined in (4–11). Using the notation in (4–9), we see that the vector Λ in (4–12) contains a total of $\widehat{N_e}$ entries. As the grid at level k contains $m_k = m/2^k + 1$ points, the vector $\mathcal{U}$ has a total of

$$\sum_{k=0}^{k_c} n_k m_k = \left(\sum_{k=0}^{k_c} \frac{n_k}{2^k} \right) m + \widehat{N_e}$$

entries. In what follows, we use the term "memory ratio" for the quantity

$$r = \sum_{k=0}^{k_c} \frac{n_k}{2^k} \tag{4–14}$$

as this gives a useful estimate of the ratio of the amount of storage required to m (where the finest grid has $m+1$ points).

We conclude this section by noting that the MLEVD algorithm can be generalized by using $\tilde{Q}_k(A_{k-i})$ instead of $\tilde{Q}_k(A_0)$ in (4–7), where A_{k-i}, $1 \leq i \leq k$ is a direct representation of A_0 on grid level k. In particular, for the case of interest here (with $A_0 \equiv H(u)$), we can define H directly on a given level k using

$$\tilde{H}_k(U_k) = I_k + Q_k((B^{1/2})^*)(G_k'(U_k))^* Q_k(R^{-1}) G_k'(U_k) Q_k(B^{1/2}), \tag{4–15}$$

where $G_k'(U_k)$ and $(G_k'(U_k))^*$ are the tangent linear and adjoint operators G_U' and $(G_U')^*$ discretized on level k. Note that, as $B_b^{1/2}$ and R^{-1} are defined at the finest grid level, appropriate projections are still required in (4–15). This approach allows the PDE problems defining the Hessian-vector product to be solved at any level $k-i$, while solving the eigenproblem at level k. This may help to reduce computational time, although the multilevel approximation will become less accurate (given the same allocated memory).

4.1.3 Hessian decomposition

Here we focus on a specific application for the MLEVD algorithm, namely, the approximation of the inverse of the projected Hessian matrix (2–10) in variational data assimilation. Before we describe the specific algorithms proposed, we introduce a decomposition of the Hessian into a set of elementary Hessians which will prove to be useful later.

Using the factorization

$$R^{-1} = R^{-1/2}\, \bar{I}\, R^{-1/2},$$

where $\bar{I}$ is the identity on the M-dimensional observation space $\mathcal{Y}$ and $R^{-1/2} : \mathcal{Y} \to \mathcal{Y}$ is a symmetric square-root of R, the preconditioned Hessian can be rewritten in the form

$$\tilde{H}(U) = I + (B^{1/2})^* (G'(U))^* R^{-1/2}\, \bar{I}\, R^{-1/2} G'(U) B^{1/2}. \tag{4–16}$$

Now let $\mathcal{I}$ be a set of indices of the diagonal elements of $\bar{I}$, and suppose that $\mathcal{I}$ is partitioned into L disjoint subsets $\mathcal{I}^l$, $l = 1, \ldots, L$. Defining the diagonal matrices $\bar{I}^l$ such that

$$\bar{I}^l_{i,i} = \begin{cases} 1, & i \in \mathcal{I}^l \\ 0, & i \notin \mathcal{I}^l \end{cases}, \; i = 1, \ldots, M,$$

the identity can be written as

$$\bar{I} = \sum_{l=1}^{L} \bar{I}^l.$$

Combining this with (4–16), after some algebraic manipulation we obtain a Hessian decomposition as follows:

$$\tilde{H}(u) = I + \sum_{l=1}^{L} (\tilde{H}^l(u) - I), \tag{4–17}$$

where

$$\tilde{H}^l(U) = I + (B^{1/2})^*(G'(U))^* R^{-1/2}\, \bar{I}^l\, R^{-1/2} G'(U) B^{1/2}. \tag{4–18}$$

For a specific partition of $\mathcal{I}$, each elementary Hessian $\tilde{H}^l(U)$ can be presented in the limited-memory form with the number of leading eigenpairs n^l required for its accurate representation. In particular, it is possible to define a partition of $\mathcal{I}^l$ such that $n^l \approx n/L$, with the number of Lanczos iterations and, correspondingly, the amount of CPU time needed for computing the eigenpairs of a single H^l proportionally reduced. We also note that the elementary Hessians can be computed in parallel, so that the amount of CPU time required for computing the full limited-memory Hessian should not exceed the largest time spent on computing a single $\tilde{H}^l$. To define an optimal partition, we may use the fact that the influence of the observations made by sensors located within a given spatial subdomain is also spatially localized.

For a Hessian defined directly at a given level k (as per (4–15)), it is easy to see that the local Hessian decomposition is

$$\tilde{H}_k(U_k) = I_k + \sum_{l=1}^{L} (\tilde{H}_k^l(U_k) - I_k), \tag{4–19}$$

where

$$\tilde{H}_k^l(U_k) = I_k + Q_k((B^{1/2})^*)(G_k'(U_k))^* Q_k(R^{-1/2}\, \bar{I}^l\, R^{-1/2}) G_k'(U_k) Q_k(B^{1/2}). \tag{4–20}$$

For a specific partition of $\mathcal{I}$, powers of each elementary Hessian $\tilde{H}_k^l(u_k)$ can be approximated in limited-memory form as

$$(\mathbf{H}_k^l)^\beta \simeq I_k + \sum_{i=1}^{n_k^l} ((\lambda_{k,i}^l)^\beta - 1) W_{k,i}^l (W_{k,i}^l)^* \tag{4–21}$$

using the leading n_k^l eigenpairs $\{\lambda_{k,i}^l, W_{k,i}^l\}$, $i = 1, \ldots, n_k^l$, where the length of each eigenvector $W_{k,i}^l$ is equal to m_k.

The calculation of a single local Hessian $\mathbf{H}_k^l$ is a relatively inexpensive task which requires much less computational time than computing the global projected Hessian because each local Hessian deviates from the identity operator I only within the area of influence surrounding the subdomain Ω^l, with the size of this area depending on the transport mechanisms supported by the dynamical model. The number of eigenpairs used to describe this deviation could be relatively small, therefore a smaller number of Lanczos iterations may also be needed to evaluate the leading eigenvalues of $\tilde{H}_k^l$, with the corresponding eigenvectors being different from zero only within the local area of influence. Thus, a compact storage scheme for the eigenvectors can be utilized. To make further computational savings, the local Hessians $\mathbf{H}_k^l$ can be computed at discretization level $k = k' > 0$, as opposed to on the finest grid $k = 0$. Finally, if the local Hessians are computed in parallel, then a very significant reduction in computing time can be achieved. Specifically, if the eigenvalue analysis for each $\tilde{H}_k^l$ can be carried out on an individual processor, the time needed for computing all L eigenpair sets will be reduced to the maximum time taken to compute the eigenpairs of an individual local Hessian. We also observe that, for computing $\mathbf{H}_k^l$, a local area model rather than the global model has to be run, and it is also possible that only some sensors from the whole observation array, or from only some areas of the computational domain, will be of interest so $\mathbf{H}_k^l$ need not be calculated for every Ω^l: these ideas are not yet investigated further, but represent a subject for future research.

4.1.4 Approximating the inverse Hessian

In this section we use the multilevel eigenvalue decomposition in Algorithm 4.1.2 to build various approximations to the inverse Hessian $\tilde{H}^{-1}$, where $\tilde{H}$ is defined in (4–16). We describe three different possible algorithms, which may be useful depending on the constraints in place in terms of available computing time and memory. Some numerical experiments illustrating their relative accuracy and usefulness in practice are given in §4.1.5.

Algorithm 1

This involves using a straightforward application of the multilevel decomposition in Algorithm 4.1.2 to $\tilde{H}_0$, resulting in an eigenstructure $[\Lambda_0,\ \mathcal{W}_0]$ which can be used to evaluate $\tilde{H}_0^{-1}v_0$ and $\tilde{H}_0^{-1/2}v_0$ via (4–13). For a given optimal solution $\hat{U}_0$, the Hessian-vector product $\tilde{H}_0 v_0$ is defined by (4–15) discretized at the finest grid level $k = 0$. Given the multilevel eigenvalue dimensions N_e, the algorithm can be outlined as follows:

`define` N_e
`compute` $[\Lambda_0, U_0] = \mathrm{MLEVD}(\tilde{H}_0, N_e)$

Algorithm 2

In this approach, we assume that memory restrictions are not important, but that computing time is limited. To this end we utilize the Hessian decomposition idea as described in §4.1.3. This requires the choice of L further parameters, n_k^l, $l = 1, ..., L$, which determine the number of eigenpairs used in $\mathbf{H}_k^l$ (the limited memory approximation to $\tilde{H}_k^l$ in (4–20)). The algorithm can be outlined as follows:

`define` N_e, `level` $k' \geq 0$, `partition` $\mathcal{I}^l$, $n_{k'}^l$, $l = 1, \ldots, L$
for $l = 1, \ldots, L$:
    `compute` $\{\lambda_{k'}^i, U_{k'}^i\}^l,\ i = 1, \ldots, n_{k'}^l$ of $H_{k'}^l$ `in` (4–20)
    `compute` $[\Lambda_0, U_0] =$ `MLEVD`$(P_0(\tilde{H}_{k'}), N_e)$ `where` $\tilde{H}_{k'}$ `is in form` (4–19) `based on`
    $\mathbf{H}_{k'}^l$ `from` (4–21) instead of $\tilde{H}_{k'}^l$
end l-loop.

4.1.5 Illustration

In this section we provide some implementation details and report the results of some numerical experiments from [A4] which illustrate the performance of the algorithms presented in the previous section. Numerical tests has been performed for the model (2–23). The corresponding flow evolution is presented in Figure 3.2(left). We consider two different configurations of the sensors which provide the observations to be assimilated. For this one-dimensional problem, Scheme A has seven stationary sensors, fixed at points $\{0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7\}$ in $[0, 1]$. In Scheme B, there is one moving sensor which traverses the domain $[0, 1]$ from left to right twice during the observation time. This is to emulate satellite observations. The multilevel structure comprises four grids, with 401 grid points on the finest grid level ($k = 0$) and 51 points on the coarsest grid level ($k = 3$). We use cubic splines to implement the prolongation operator $S_{k,k-i}$ in (4–1), with the subroutine for its adjoint operator $S_{k,k-i}^*$ (restriction) again obtained by using automatic differentiation.

Investigating approximation accuracy

In the first set of experiments, we apply Algorithm 1 described in §4.1.4 to $\tilde{H}(\equiv \tilde{H}_0)$ and use the resulting multilevel eigenvalue decomposition (4–12) to build a low-memory approximation to $\tilde{H}^{-1}$ in recursive form (4–13), which we will denote by $\hat{H}^{-1}$. To assess the accuracy of our approximations, we measure the difference between two matrices in terms of the Riemann distance. That is, for two symmetric positive definite $n \times n$ matrices A and B, we define

$$\delta(A,B) = \left\| ln(B^{-1}A) \right\|_F = \left(\sum_{i=1}^{n} ln^2 \lambda_i \right)^{1/2}, \tag{4–22}$$

where λ_i, $i = 1, \ldots, n$ are the eigenvalues of $B^{-1}A$. The Riemann distance can be considered as a symmetric measure of the difference between two Gaussian probability distributions having equal modes. Since the inverse Hessian is an approximation of the analysis error covariance matrix, it is very natural to use this measure in the current context. We will compare matrices after we have applied the first-level preconditioning: it is easily shown that the Riemann distance remains unchanged on applying symmetric preconditioning to A and B, so the distance between two symmetric positive definite matrices in the original and projected spaces is the same.

The accuracy of a given approximation will clearly be dependent on the number of eigenvalues calculated at each grid level (that is, the choice of $N_e = (n_0, n_1, n_2, n_3)$). Ideally, we would like to be able to identify the 'optimal' combination N_e for a given problem but this is non-trivial, particularly as the definition of optimality is itself dependent on the problem constraints. Here we assume that there is a fixed memory ratio r (see (4–14)) allowed for a particular problem. Within this fixed-memory framework, we measure the normalized Riemann distance

$$D = \frac{\delta(\tilde{H}^{-1}, \hat{H}^{-1})}{\delta(\tilde{H}^{-1}, I)} \tag{4–23}$$

where I is an appropriately-sized identity matrix. By evaluating D for approximations constructed using all possible eigenvalue combinations N_e for a given memory ratio r, we can identify the best combinations N_e for this fixed memory problem.

Let us consider a particular case. If $\hat{H}^{-1}$ is constructed from 64 eigenpairs of $\tilde{H}^{-1}$ (corresponding to memory ratio $r = 64$), a very good approximation with $D = 2.98e^{-4}$ is achieved. However, if available memory is restricted to $r = 8$, so only 8 eigenpairs at level 0 can be stored, the resulting approximation is very poor ($D = 7.71e^{-1}$). However, if $\hat{H}^{-1}$ is constructed using the multilevel eigenvalue decomposition, much better approximations can be achieved.

As stated above, it is difficult to characterize a single 'best possible' eigenvalue combination. However, certain traits can be identified. Plots of the minimum distance achieved for various memory ratios r are shown in Figure 4.1. In each subplot, the dashed line shows the minimum distance achieved for a given r, and the solid line shows the average over the 5% of eigenvalue combinations satisfying (4–14) which resulted in the smallest distances. The dotted line shows the equivalent distance achieved using only fine grid vectors. The key observation here is that when there is only room to store a small number of fine grid vectors in memory, using a multilevel approximation clearly gives much better accuracy. Although we did not identify an 'optimal' way of choosing N_e, our experiments did show that eigenvalue combinations which have no or very few eigenpairs on the finest level(s) appear to perform best. Based on this, we suggest the ansatz of doubling the number of eigenpairs calculated at each level (from fine to coarse grids). For example, the combinations which correspond to $r = 8$ are as follows: $N_e = (0, 0, 24, 48)$ and $N_e = (2, 4, 8, 16)$. The distances achieved using this strategy for various values of r are displayed in Figure 4.1 using red crosses. Although these combinations do not

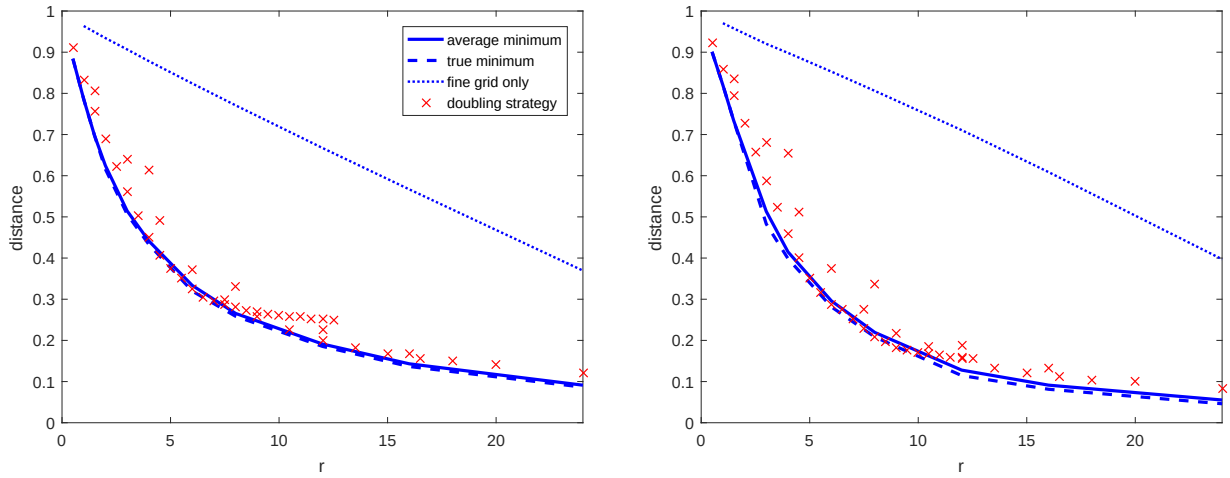

Figure 4.1: Distance D (4–23) plotted against memory ratio r (4–14) for model problems A1 (stationary sensors) and B1 (moving sensor).

always give rise to the minimum distance, they usually give reasonable approximations. This doubling strategy will be adopted in the next section.

Using $\hat{H}^{-1}$ as a preconditioner for Gauss-Newton

Here we illustrate the use of our multilevel approximation to H^{-1} in a typical application. Specifically, we recall the case mentioned in §1.2.5 of incremental 4D-Var, where the solution of a system of linear equations of the form (1–37) with coefficient matrix H has to be approximated at each step of a Gauss-Newton process. This is typically achieved using a few CG iterations. In realistic DA applications, the number of Gauss-Newton (outer) iterations as well as the number of CG (inner) iterations is limited by the time available in the forecast window. More details of approximate Gauss-Newton methods for large-scale data assimilation problems can be found in [S19].

After first-level preconditioning, system (1–37) being discretized at the finest level $k = 0$ takes the form

$$H_0(U_0^j)\delta V_0^j = -(B_0^{1/2})^* G_0(U_0^j), \tag{4–24}$$

where $\delta V_0^j = B_0^{1/2}\delta U_0^j$. As we are assuming that this preconditioning is always applied, we will refer to this case as 'unpreconditioned Gauss-Newton. The convergence of the CG method applied to (4–24) can be further accelerated by preconditioning the system again using $\hat{H}_0^{-1}$ to improve the resulting eigenspectrum, in other words, by solving

$$\hat{H}_0^{-1}(U_0^j)\tilde{H}(U_0^j)\delta V_0^j = -\hat{H}_0^{-1}(U_0^j)(B_0^{1/2})^* G(U_0^j).$$

Here the preconditioner $\hat{H}_0^{-1}(U_k^j)$ is computed once per Gauss-Newton step (before iterating with CG): this is an important difference from the approach presented in [S14], where the multigrid cycle is applied as a preconditioner for each CG iteration. In what follows, we apply incremental 4D-Var to the Burgers' test problem, and investigate the effect of preconditioning using multilevel approximations $\tilde{H}^{-1}$ built using Algorithms 2 (i.e. involving the Hessian decomposition technique). For the test case with stationary sensors (Scheme A), each local Hessian $\mathbf{H}_{k'}^l$ corresponds to one sensor located at $x = x^l$.

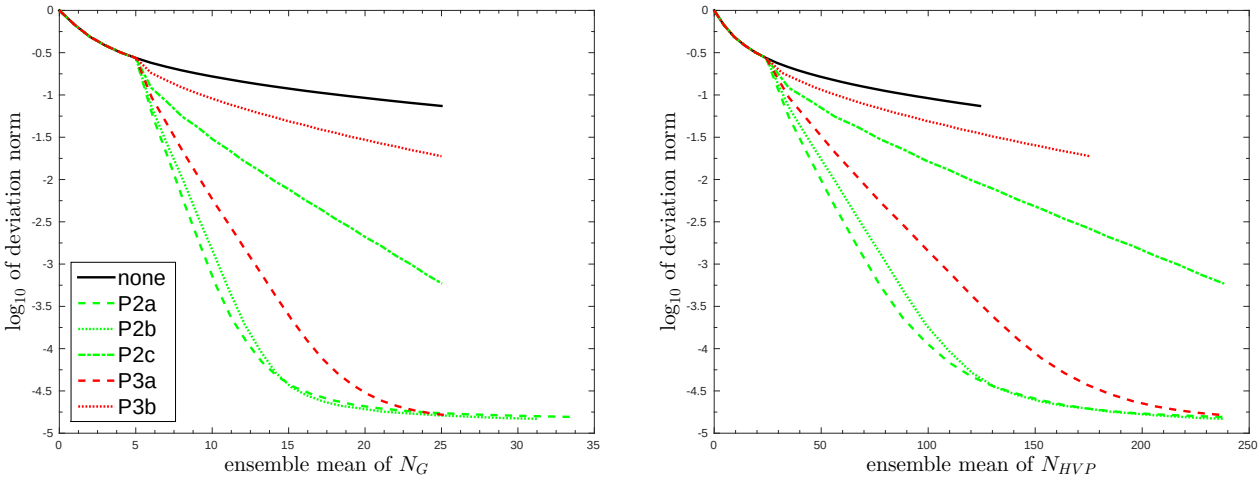

Figure 4.2: Convergence diagram for MP1.

Preconditioners P2a ($N_e = (200, 0, 0, 0)$), P2b ($N_e = (0, 8, 16, 32)$) and P2c ($N_e = (0, 4, 8, 16)$) use Algorithm 2 with different values of N_e in the multilevel representation of the inverse Hessian $\hat{H}_0^{-1}$. For each elementary Hessian $\mathbf{H}^l_{k'}$, the Hessian-vector product is defined at level $k' = 1$, $m = 201$, with $n^l_{k'} = 8$ eigenpairs used for its limited-memory representation. For each test case, the DA problem was solved 25 times with perturbed data, and the resulting ensemble averaged values characterizing the convergence rate and the solution accuracy are presented in convergence diagrams shown in Figure 4.2.

Each subplot of a diagram shows the averaged deviation norm $\|U^j - \hat{U}\|_2$ (where $\hat{U} = U^\infty$ is the final estimate $\bar{U}$), plotted in log_{10}-scale against computational time measured in two different units. These are the number of the cost-function/gradient evaluations and the number of Hessian-vector product evaluations (denoted by N_G and N_{HVP}, respectively) at the finest discretization level needed to carry out the whole solution procedure (that is, the cost of constructing and applying any preconditioning is included in these figures). The former coincides with the actual number of the cost-function/gradient evaluations, the latter also includes the maximum time spent on evaluating an individual $\mathbf{H}^l$ (we assume that those are evaluated in parallel). The figure demonstrates a very significant acceleration of the Gauss-Newton process as compared to the unpreconditioned Gauss-Newton.

4.1.6 Conclusions

The inverse Hessian of the auxiliary cost-function (1–22) plays an important part in different aspects of variational DA. The Hessian-vector product is defined by sequential solution of the tangent linear and adjoint problems; for the inverse Hessian (or its square-root), no such definition is possible. In high dimensions, the requirement to work in a matrix-free environment means that compact representation schemes are of significant interest. The simplest one, based on the eigenvalue decomposition of the projected Hessian after first-level preconditioning, is however not good enough. Here we have introduced the novel concept of a multilevel eigenvalue decomposition, which results in a much more efficient compact representation (supercompact) of the inverse Hessian (and its square root). At a given level, the eigensystem of an operator preconditioned by its own approximation from the next coarser level is computed, ascending from the coarsest to the finest level. The numerical results in §4.1.5 demonstrate that, given a specified memory allowance, the inverse Hessian approximation accuracy is greatly improved as compared to a one-level eigenvalue decomposition scheme. We also believe that a similar algorithm can be utilized beyond the application considered here, for example,

for any symmetric operator resulting from the discretization of a PDE, or for image compression and restoration. In such cases, the resulting low-memory representations of the inverse operator could be particularly useful for preconditioning in problems with multiple right-hand sides.

We have also considered the application of our compact inverse Hessian approximations as preconditioners for a Gauss-Newton minimization procedure. Here we introduced a further novel decomposition principle, namely, allowing the Hessian to be represented by the sum of elementary Hessians, which can be evaluated (and compressed) in parallel. The numerical results in §4.1.5 show that, given a fixed execution time, a much more accurate approximation can be computed as compared to using unpreconditioned Gauss-Newton method. The new multilevel method therefore offers an important parallelisable resource applicable directly to minimization problems.

4.1.7 Future developments

In data science, additional difficulties arise because of the transient and opportunistic nature of the networks that the data represents. Traditional numerical methods often rely on understanding the underlying physical principles involved in the process being modelled but, although the science of complexity and networks is developing rapidly, there is still a long way to go before the understanding of behavior of modern data networks reaches an equivalent stage. There is therefore a pressing need to develop new preconditioners to provide suitable tools for tackling these modern applications.

In order to apply our multilevel eigenvalue decomposition method to investigate large sensor networks, given that many modern datasets are not associated with any sort of underlying physical grid of sensors, the effects of replacing some of the traditional geometric multigrid ideas which have been used to date with ideas from algebraic multigrid should be investigated. These latter methods construct a hierarchy of operators directly from the system matrix (Hessian) without any geometric interpretation, so should be well-suited to more flexible modern data networks. The form of the grid transfer operators used in the geometric multigrid case is such that building equivalent versions for algebraic multilevel methods should work just as effectively, but this idea needs to be tested.

Author's publications

[A1] H. Oubanas, I. Gejadze, P.-O. Malaterre, M. Durand, R. Wei, R.P.M. Frasson, and A. Domeneghetti. Discharge estimation in ungauged basins through variational data assimilation: the potential of the SWOT mission. *Water Resource Research*, pages 1–20, 2017, in review.

[A2] H. Oubanas, I. Gejadze, P.-O. Malaterre, and F. Mercier. River discharge estimation using variational DA involving the full Saint-Venant model and synthetic SWOT-type observations. *Journal of Hydrology*, pages 1–25, 2017, accepted.

[A3] I. Gejadze, H. Oubanas, and V. Shutyaev. Implicit treatment of model error using inflated observation-error covariance. *Q.J.R. Meteorol. Soc.*, 143:1–14, 2017.

[A4] K.L. Brown, I. Gejadze, and A. Ramage. A multilevel approach for computing the limited-memory Hessian and its inverse in variational data assimilation. *SIAM J. Sci. Comput.*, 38(5):A2934–A2963, 2016.

[A5] I. Gejadze and P.-O. Malaterre. Design of the control set in the framework of variational data assimilation. *Journal of Computational Physics*, 325:358–379, 2016.

[A6] I. Gejadze and P.-O. Malaterre. Discharge estimation under uncertainty using variational methods with application to the full Saint-Venant hydraulic network model. *Int. J. Num. Meth. Fluids*, 34:127–147, 2016.

[A7] I. Gejadze, V. Shutyaev, A. Vidard, and F.-X. Le-Dimet. Optimal solution error quantification in variational data assimilation involving imperfect models. *Int. J. Numer. Meth. Fluids*, 83(3):276–290, 2016.

[A8] I. Gejadze and V. Shutyaev. On gauss-verifiability of optimal solutions in variational data assimilation problems with nonlinear dynamics. *Journal of Computational Physics*, 280:439–456, 2015.

[A9] I. Gejadze, V. Shutyaev, and F.-X. Le Dimet. Analysis error covariance versus posterior covariance in variational data assimilation. *Q.J.R. Meteorol. Soc.*, 139:1826–1841, 2013.

[A10] V. Shutyaev, I. Gejadze, G.J.M. Copeland, and F.-X. Le Dimet. Optimal solution error covariance in highly nonlinear problems of variational data assimilation. *Nonlinear Processes in Geophysics*, 19:177–184, 2012.

[A11] I. Gejadze and V. Shutyaev. On computation of the design function gradient for the sensor-location problem in variational data assimilation. *SIAM J. Sci. Comput.*, 34(2):B127–B147, 2012.

[A12] I. Gejadze, G.J.M. Copeland, F.-X. Le Dimet, and V. Shutyaev. Computation of the analysis error covariance in variational data assimilation problems with nonlinear dynamics. *Journal of Computational Physics*, 230(22):7923–7943, 2011.

[A13] V. Shutyaev and I. Gejadze. Adjoint to the Hessian derivative and error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Modelling*, 26(2):179–188, 2011.

[A14] I. Gejadze, F.-X. Le Dimet, and V. Shutyaev. On optimal solution error covariances in variational data assimilation problems. *Journal of Computational Physics*, 229(6):2159–2178, 2010.

[A15] V. Shutyaev, F.-X. Le-Dimet, and I. Gejadze. Reduced-space inverse Hessian for analysis error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Modelling*, 25(2):169–185, 2010.

[A16] V. Shutyaev, F.-X. Le-Dimet, and I. Gejadze. A-posteriori error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Modelling*, 24(2):161–169, 2009.

[A17] I. Gejadze, F.-X. Le Dimet, and V. Shutyaev. On analysis error covariances in variational data assimilation. *SIAM J. Sci. Computing*, 30(4):1847–1874, 2008.

[A18] V. Shutyaev, F.-X. Le-Dimet, and I. Gejadze. On optimal solution error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Modelling*, 23(2):197–206, 2008.

[A19] H. Elhanafy, G.J.M. Copeland, and I. Gejadze. Estimation of predictive uncertainties in flood wave propagation in a river channel using adjoint sensitivity analysis. *Int. J. Numer. Meth. Fluids*, 56(8):1201–1207, 2008.

[A20] I. Gejadze and J. Monnier. On a 2D zoom for the 1D shallow water model: coupling and data assimilation. *Comput. Methods Appl. Mech. Engrg.*, 196:4628–4643, 2007.

[A21] V. Shutyaev, F.-X. Le-Dimet, and I. Gejadze. On error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Modelling*, 22(2):1–12, 2007.

[A22] V. Shutyaev, F.-X. Le-Dimet, and I. Gejadze. On optimal solution error in variational data assimilation: theoretical aspects. *Russ. J. Numer. Anal. Math. Modelling*, 21(2):139–152, 2006.

[A23] I. Gejadze and G.J.M. Copeland. Open boundary control for Navier-Stokes equations including a free surface: adjoint sensitivity analysis. *Comput. Math. Appl.*, 52:1243–1268, 2006.

[A24] I. Gejadze, G.J.M. Copeland, and I.M. Navon. Open boundary control for Navier-Stokes equations including a free surface: data assimilation. *Comput. Math. Appl.*, 52:1269–1288, 2006.

[A25] I. Gejadze and G.J.M. Copeland. Adjoint sensitivity analysis for fluid flow with free surface. *Int. J. Numer. Meth. Fluids*, 47(8-9):1027–1034, 2005.

[A26] I. Gejadze and Y. Jarny. An inverse heat transfer problem for restoring the temperature field in a polymer melt flow through a narrow channel. *Int. J. Therm. Sci.*, 41:528–535, 2002.

[A27] I. Gejadze and V. Shutyaev. An optimal control problem for the initial condition restoration. *Comput. Math. and Math. Physics*, 39(9):1479–1488, 1999.

[A28] I. Gejadze. A fast algorithm for solving a least-squares problem dependent on two regularization parameters. *Comput. Math. and Math. Physics*, 39(3):357–364, 1999.

[A29] I. Gejadze, O.M. Alifanov, and E.A. Artyukhin. Sequential regularized solution of an inverse heat conduction problem. *Doklady Mathematics (Proc. of the Russian Academy of Science)*, 59(1):145–148, 1999.

[A30] I. Gejadze and E.A. Artyukhin. A method for solving an observation problem for the non-stationary temperature field (linear case). *J. of Computer and Systems Sciences Int.*, 37(4):605–614, 1998.

[A31] I. Gejadze and V. Shutyaev. Justification of the perturbation method for a quasi-linear heat conduction problem. *Comput. Math. and Math. Physics*, 38(6):909–915, 1998.

[A32] O.M. Alifanov and I. Gejadze. Thermal loads identification technique for materials and structures in real time. *Acta Astronautica*, 41(4-10):255–265, 1997.

Side references

[S1] V.I. Agoshkov, E.I. Parmuzin, V.B. Zalesny, V.P. Shutyaev, N.B. Zaharova, and A.V Gusev. Variational assimilation of observation data in the mathematical model of the Baltic See dynamics. *Russ. J. Numer. Anal. Math. Modelling*, 30(4):203–212, 2015.

[S2] O.M. Alifanov, E.A. Artyukhin, and S.V. Rumyantsev. *Extreme Methods for Solving Ill-Posed Problems with Applications to Inverse Heat Transfer Problems.* Begel House Publishers, New York - Wallingford (U.K.), 1996.

[S3] H. Auvinen, J.M. Bardsley, H. Haario, and T. Kauranne. The variational Kalman filter and an efficient implementation using limited memory BFGS. *Int. J. Num. Meth. Fluids*, 64(3):314–335, 2010.

[S4] J.M. Bardsley and A. Luttman. Total variation-penalized Poisson likelihood estimation for ill-posed problems. *Advances in Computational Mathematics*, 31(1):35–59, 2009.

[S5] T. Bergot and A. Doerenbecher. A study of the optimization of the deployment of targeted observations using adjoint-based methods. *Quart. J. R. Meteorol. Soc.*, 128:1689–1712, 2002.

[S6] M. Berliner, Z.-Q. Lu, and C. Snyder. Statistical design for adaptive weather observations. *J. Atm. Sciences*, 56:2436–2552, 1999.

[S7] A. Borzi and V. Schulz. Multigrid methods for PDE optimization. *SIAM Review*, 51:361–395, 2009.

[S8] Y. Chen and D. Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.*, 44:1–26, 2012.

[S9] A.M. Clayton, A.C. Lorenc, and D.M. Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, 139:1445–1461, 2013.

[S10] S. Costiner and S. Ta'asan. Adaptive multigrid techniques for large-scale eigenvalue problems: solutions of the Schrodinger problem in two and three dimensions. *Phys. Rev.*, 44(3):3704–3717, 1995.

[S11] P. Courtier, J.-N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, 120:1367–1387, 1994.

[S12] D.N. Daescu and I.M. Navon. Adaptive observations in the context of 4D-Var data assimilation. *Meteorol. Atmos. Phys.*, 85:205–226, 2004.

[S13] M. Dashti, K.J.H. Law, A.M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29:17–44, 2013.

[S14] L. Debreu, E. Neveu, E. Simon, F.-X. Le Dimet, and A. Vidard. Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems. *Quart. J. Roy. Meteor. Soc.*, 142:515–528, 2016.

[S15] J.R. Donaldson and R.B. Schnabel. Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics*, 29(1):67–82, 1987.

[S16] M. Ehrendorfer and J.J. Tribbia. Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, 54:286–313, 1997.

[S17] V.V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

[S18] V.V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments. Lecture Notes in Statistics*. Springer-Verlag, New York, 1997.

[S19] S. Gratton, A.S. Lawless, and N.K. Nichols. Approximate Gauss-Newton methods for nonlinear least squares problems. *SIAM J. Optim.*, 18:106–132, 2007.

[S20] L. Hascoët and V. Pascual. TAPENADE 2.1 user's guide. *INRIA Technical Report*, 0300:1–78, 2004.

[S21] N. Henze. Invariant tests for miltivariate normality: a critical review. *Statistical Papers*, 43:467–506, 2002.

[S22] M.J. Hossen, I.M. Navon, and D.N. Daescu. Effect of random perturbations on adaptive observation techniques. *Int. J. Numer. Meth. Fluids*, 85(1):110–123, 2012.

[S23] K.E. Howes, A.M. Fowler, and A.S. Lawless. Accounting for model error in strong-constraint 4D-Var data assimilation. *Quart. J. Roy. Meteorol. Soc.*, 143:1227–1240, 2017.

[S24] T. Hwang and I.D. Parsons. A multigrid method for the generalized symmetric eigenvalue problem: part I - algorithm and implementation. *Int. J. Numer. Meth. Eng.*, 35:1663–1676, 1992.

[S25] T. Isaac, N. Petra, G. Stadler, and O. Ghattas. Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *J. Comput. Physics*, 296:348–368, 2015.

[S26] A.G. Kalmikov and P. Heimbach. A Hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM J. Sci. Comput.*, 36(5):S267–S295, 2014.

[S27] A.V. Knyazev and K. Neymeyr. Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method. *ETNA: Electronic transactions on Numerical Analysis*, 15:38–55, 2003.

[S28] F.-X. Le Dimet, I.M. Navon, and D. Daescu. Second-order information in data assimilation. *Monthly Weather Review*, 130(3):629–648, 2002.

[S29] F.-X. Le Dimet and V. Shutyaev. On deterministic error analysis in variational data assimilation. *Nonlinear Processes in Geophysics*, 14:1–10, 2005.

[S30] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38:97–110, 1986.

[S31] P.F.L. Lermusiaux. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J. Comput. Physics*, 217:176–199, 2006.

[S32] J.-L. Lions. *Contrôle Optimal des Systèmes Gouvernés par des Équations aux Dérivées Partielles*. Dunod, Paris, 1968.

[S33] P.-O. Malaterre, J.-P. Baume, and D. Dorchies. Simulation and integration of control for canals software (sic^2), for the design and verification of manual or automatic controllers for irrigation canals. In *USCID Conference on Planning, Operation and Automation of Irrigation Delivery Systems*, pages 377–382, Phoenix, Arizona, December 2-5 2014.

[S34] G.I. Marchuk, V.I. Agoshkov, and V.P. Shutyaev. *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*. CRC Press Inc., New York, 1996.

[S35] T.N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, 55:633–653, 1998.

[S36] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteorol. Soc.*, 118:649–672, 1992.

[S37] Y. Saad. *Numerical Methods for Large Eigenvalue Problems. 2nd edition*. SIAM, New York, 2011.

[S38] Y. Sasaki. Some basic formalism in numerical variational analysis. *Month. Wea. Rev.*, 98(12):875–883, 1970.

[S39] T.A. Severini. *Likelihood Methods in Statistics*. Oxford Statistical Science Series, Oxford, 2000.

[S40] G.L.G. Sleijpen and H.A. Van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17:401–425, 1996.

[S41] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[S42] Amemiya T. *Handbook of Econometrics*. North-Holland Publishing Company, Amsterdam, 1983.

[S43] A Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, 2005.

[S44] W.C. Thacker. The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.*, 94(C5):6177–6196, 1989.

[S45] Z. Toth and E. Kalnay. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, pages 2317–2330, 1993.

[S46] Y. Trémolet. Model-error estimation in 4D-Var. *Quart. J. Roy. Meteorol. Soc.*, 133(626):1267–1280, 2007.

[S47] D. Ucinski. Optimal sensor-location for parameter estimation of distributed processes. *Int. J. Control*, 73(13):1235–1248, 2000.

[S48] D Ucinski. *Optimal Measurement Methods for Distributed-Parameter System Identification*. CRC Press, Boca Raton, FL, 2005.

[S49] D. Ucinski and T. Zieba. Mobile sensor routing for parameter estimation of distributed systems using the parallel tunneling method. *Int. J. Appl. Math. Comput. Sci.*, 18(3):307–318, 2008.

[S50] F. Veersé. Variable-storage quasi-Newton operators as inverse forecast/analysis error covariance matrices in variational data assimilation. *INRIA Technical Report 3685*, pages 1–28, 1999.

[S51] A.T. Weaver and I. Mirouze. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Quart. J. Roy. Meteorol. Soc.*, 139:242–260, 2013.

[S52] X. Zou, I.M. Navon, and Le Dimet F.-X. An optimal nudging data assimilation scheme using parameter estimation. *Quart. J. Roy. Meteorol. Soc.*, 118:1163–1186, 1992.