

Applications of adjoints

As mentioned in the introduction, adjoints (and tangent linear models) have many applications in many different areas of computational science. In this section, we aim to give a very brief overview of each application in which adjoints are used, with the intent of getting the basic idea across.

For each application, a very brief literature review is provided, giving pointers to some key references which may be used to explore the field. I make no claim that these reviews are comprehensive; naturally, I am personally more familiar with some areas than others. Contributions to this section are very welcome.

PDE-constrained optimisation

As discussed in the previous sections, adjoints form the core technique for efficiently computing the gradient $dJ(u, m)/dm$ of a functional J to be minimised. This is usually essential for solving such optimisation problems in practice: gradient-free optimisation algorithms typically take orders of magnitude more iterations to converge; since each iteration involves a PDE solve, minimising the number of iterations taken is crucial.

For an engineering introduction to PDE-constrained optimisation, Gunzburger's book is excellent [1M-Gun03]; for an in-depth mathematical analysis, see the rigorous treatment of Hinze et al. [1M-HPUU09]. PDE-constrained optimisation is also referred to in the literature as optimal control: the book of Lions [5M-Lio71] was a fundamental early contribution.

Sensitivity analysis

Occasionally, the gradient of a functional J with respect to some parameter m is not merely required as an input to an optimisation algorithm, but rather is of scientific interest in its own right. Adjoint computations can tease apart hidden influences and teleconnections; such computations can also inform scientists regarding which variables matter the least, which is often important for deriving approximate models; parameters with little impact on the simulation can be ignored. This process is also often undertaken in advance of solving an optimisation problem: by discarding parameters which do not significantly influence the functional, the dimension of the parameter space may be systematically reduced.

A fundamental early contribution was the work of Cacuci [5M-Cac81]. Much excellent work in applying sensitivity analysis to questions of enormous scientific importance has been done in the areas of oceanography and meteorology: partly because ocean and atmospheric models often have adjoint versions implemented for the purposes of data assimilation, partly because adjoint analysis is often the only practical way of identifying such connections, and partly because practitioners in these fields are aware of adjoints and their potential. Of particular note is the work done by Heimbach and co-workers using the adjoint of the MITgcm ocean model [1M-LH07] [5M-HML+10] [5M-HL12].

Data assimilation

A forward model requires data on which to operate. For example, to start a weather forecast, knowledge of the entire state of the atmosphere at some point in time is required as an initial condition from which to begin the simulation: start from the wrong initial condition, and you will get the wrong weather.

The problem is that, in practice, the initial condition is unknown. Instead, observations of the state of the atmosphere are available, some available at the initial time, and some taken at later times. The goal of data assimilation is to systematically combine observations and computations, usually with the intention of acquiring the best possible estimate of the unknown initial condition, so that a forecast may be run. Indeed, most of the dramatic increase in forecast skill over the past decade has been attributable to improvements in data assimilation algorithms (Prof. Dale Barker, UK Met Office, personal communication). This problem is routinely tackled in every meteorological centre in the world, multiple times a day: your weather forecast relies entirely upon it.

There are two major approaches to data assimilation. In a sequential algorithm, the forward system is timestepped until an observation is available, at which point the model is instantaneously "updated" to incorporate the information contained in the observation. The most popular approach to computing the amplitude of the update is the Kalman filter algorithm [5M-Kal60]. The model is then continued from this updated state until all of the observations are used. A significant drawback of this approach is that an observation only influences the state at times later than the observation time: in other words, its temporal influence only propagates forward, not backwards in time. This is a major disadvantage.

vantage in the case where the initial condition is the main quantity of interest, and most of the observations are made towards the end of the time window, as is the case in studies of mantle convection [5M-BHT03].

The other major approach is referred to as variational data assimilation, which is a special case of PDE-constrained optimisation. In this approach, a functional J is chosen to represent the misfit between the observations and the computations, weighted by the uncertainties in each. The initial condition is treated as a control parameter m , and chosen to minimise the misfit J . The data assimilation community tends to place significant emphasis on modelling the uncertainties in both the computations and observations, as this is key to extracting the maximal amount of information out of both.

Fundamental early works in field of variational data assimilation were undertaken by Le Dimet and Talagrand [1M-LDT86] and Talagrand and Courtier [1M-TC87]. For an excellent introduction, see the book of Kalnay [5M-Kal02]. Information on the data assimilation schemes used in practice by the [European Centre for Medium-range Weather Forecasting](#) and the [UK Met Office](#) is available online.

Inverse problems

Data assimilation can be seen as a particular kind of inverse problem, where the focus is on obtaining the best estimate for the system state at some point in the past. More general inverse problems, where we seek to gain information about unobservable system parameters from observable system outputs, are ubiquitous in science and engineering. Again, the same idea of minimising some functional that measures the misfit between the observations and computed model outputs plays a role. The field also has a heavy emphasis on regularisation of inverse problems (which are generally ill-posed) and on statistical estimates of uncertainty in the obtained results. For an introductory textbook, see the book by Tarantola [5M-Tar05]; for a review of the current state of the art in computational practice, see the compendium of Biegler et al. [5M-BMB+11].

Generalised stability theory

The stability of solutions of physical systems is obviously of key importance in many fields. The traditional approach to stability theory is to linearise the operator about some state, and investigate its eigenvalues: if the real component of every eigenvalue is negative, the state is stable, and the associated eigenmode will vanish in the limit as $t \rightarrow \infty$; while if any eigenvalue has a positive real part, the state is unstable, and the associated eigenmode will grow in amplitude. While this traditional approach works well in many cases, there are many important cases where this analysis predicts stability where in fact the physical system is unstable; in particular, this analysis fails when the operator is *nonnormal* [1M-TTRD93] [5M-Sch07].

In the nonnormal case, the usual stability theory fails. The two main theoretical responses to this development have been the concepts of pseudospectra (by Trefethen et al. [5M-TE05]) and generalised stability theory (by Farrell et al. [1M-FI96]). Instead of focusing on the eigenvalues of the operator linearised about some steady state, generalised stability theory analyses the generalised eigenvalues associated with the *propagator of the system*, which maps perturbations in initial conditions to perturbations in the final state. Essentially, the propagator is the inverse of the tangent linear operator. By examining these values, such an analysis can describe and predict the perturbations that will grow maximally over *finite time windows* [5M-Lor65]. In order to compute these generalised eigenvalues of the system propagator, both the tangent linear and adjoint operators must be repeatedly solved [5M-TB97].

As generalised stability theory yields information about the perturbation directions which grow the most over the period of interest, these vectors are often used to initialise ensemble members to gain the optimal amount of information possible about the variance of the ensemble [5M-IF05] [5M-Bui06]. The growth rates associated with these optimal perturbations have important implications for the timescales of predictability of the physical system. For examples of this analysis, see the work of Zanna et al. [5M-ZHMT12].

We also note in passing that it is possible to use these singular vectors to *guide* the targeting of observations to *maximise* the effectiveness of a data assimilation strategy. For more details, see [5M-PGBB98].

The ECCO2 project

One of the most impressive computational experiments ever attempted is the [ECCO2 project](#), which uses the adjoint of the [MITgcm ocean model](#) [5M-HHG05] to assimilate every observation made of the world's oceans over the past twenty years [5M-Wun96]. With this assimilation, they have produced the most accurate estimation of the state of the world's oceans ever devised. For a stunningly beautiful visualisation of this experiment, see <http://vimeo.com/39333560>.

Nonnormal matrices

A matrix is normal if its eigenvectors form an orthonormal basis. A matrix is nonnormal if the eigenvectors have nonzero projection onto each other. See [the Wikipedia entry](#) for more details.

For more details, see the appendix on [generalised stability theory](#).

Error estimation

Another major application of adjoints is goal-based error estimation, and the related computational technique of goal-based adaptivity. For the purposes of this section, let u and λ denote the *exact* forward and adjoint solutions associated with the PDE $F(u) = 0$, and let u_h and λ_h be some approximations to them computed using a Galerkin finite element discretisation. The fundamental question of goal-based error estimation is: *what impact does the discretisation error $u - u_h$ have on the error in the goal functional $J(u) - J(u_h)$?* One can construct cases where $u - u_h$ is large, but $J(u) - J(u_h)$ is zero; similarly, one can construct cases where $u - u_h$ is small, but $J(u) - J(u_h)$ is large.

The fundamental theorem of error estimation, due to Rannacher and co-workers [1M-BR01] [5M-BR03], states that

$$J(u) - J(u_h) = \frac{1}{2} \langle \lambda - \lambda_h, \rho_u \rangle + \frac{1}{2} \langle u - u_h, \rho_\lambda \rangle + R_h^{(3)},$$

where $u - u_h$ is the discretisation error in the forward solution, $\lambda - \lambda_h$ is the discretisation error in the adjoint solution, ρ_u is the forward residual, ρ_λ is the adjoint residual, and $R_h^{(3)}$ is a remainder term which is *cubic* in the discretisation errors $u - u_h$ and $\lambda - \lambda_h$.

In practice, $u - u_h$ is estimated by approximating u with an extrapolation of u_h into a higher-order function space (and similarly for λ), and the expression for $J(u) - J(u_h)$ is broken up into a sum of element-level *error indicators* that are used to decide which elements should be refined in an adaptive algorithm. For a discussion of how to implement goal-based adaptivity in a general way in the FEniCS framework, see the work of Rognes and Logg [5M-RL10].

The works of Rannacher and co-workers give many examples where a computation that employs goal-based adaptivity is dramatically faster at computing the functional to within a certain tolerance than the corresponding fixed-mesh or heuristically-driven adaptivity. This theorem raises the possibility of *reliable automated computation*: not only can the discretisation of the differential equation be automated with the FEniCS system, it can be automated to *reliably and efficiently compute desired quantities to within a specified accuracy*. The prospect of such a system would dramatically change the social utility of computational science.

References

- [5M-BR03] W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. ETH Zürich Lectures in Mathematics. Birkhäuser, 2003. ISBN 3764370092.
- [5M-BMB+11] L. Biegler, Y. Marzouk, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, editors. *Large-Scale Inverse Problems and Quantification of Uncertainty*. Wiley Series in Computational Statistics. Wiley, 2011. ISBN 978-0-470-69743-6.
- [5M-Bui06] R. Buizza. The ECMWF ensemble prediction system. In T. Palmer, editor, *Predictability of Weather and Climate*, chapter 17, pages 459–488. Cambridge University Press, 2006. doi:10.1017/CBO9780511617652.018.
- [5M-BHT03] H. P. Bunge, C. R. Hagelberg, and B. J. Travis. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography. *Geophysical Journal International*, 152(2):280–301, 2003. doi:10.1046/j.1365-246X.2003.01823.x.

Residuals

To compute the forward residual ρ_u , take the approximate forward solution u_h and plug it in to the forward equation $\rho_u \equiv F(u_h)$. If u_h were the exact solution, $F(u_h)$ would be zero, but since the solution is only approximate $\rho_u \equiv F(u_h)$ will be nonzero.

To compute the adjoint residual, perform the analogous computation: take the approximate adjoint solution λ_h and plug it in to the adjoint equation, and take all terms in the adjoint equation to the left-hand side.

The structure of the error estimator

Notice that the error estimator has a very particular structure: it is the average of the inner product of the adjoint solution error with the forward residual, and the forward solution error with the adjoint residual. As many early results in the field only employed the first term in the error estimator, the approach became known as the “dual-weighted residual” approach (the term “dual” is commonly used to refer to the adjoint in this branch of the literature).

If the averaging is not performed, and only the first term of the error estimator is included, the remainder term is *quadratic* in the forward and adjoint discretisation errors, not cubic.

- [5M-Cac81] D. G. Cacuci. Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. *Journal of Mathematical Physics*, 22(12):2794–2802, 1981. doi:10.1063/1.525186.
- [5M-HHG05] P. Heimbach, C. Hill, and R. Giering. An efficient exact adjoint of the parallel MIT General Circulation Model, generated via automatic differentiation. *Future Generation Computer Systems*, 21(8):1356–1371, 2005. doi:10.1016/j.future.2004.11.010.
- [5M-HL12] P. Heimbach and M. Losch. Adjoint sensitivities of sub-ice shelf melt rates to ocean circulation under Pine Island Ice Shelf, West Antarctica. *Annals of Glaciology*, 53(60):59–69, 2012. doi:10.3189/2012/AoG60A025.
- [5M-HML+10] P. Heimbach, D. Menemenlis, M. Losch, J. M. Campin, and C. Hill. On the formulation of sea-ice models. Part 2: lessons from multi-year adjoint sea-ice export sensitivities through the Canadian Arctic Archipelago. *Ocean Modelling*, 33(1-2):145–158, 2010. doi:10.1016/j.ocemod.2010.02.002.
- [5M-IF05] P. J. Ioannou and B. F. Farrell. Application of generalised stability theory to deterministic and statistical prediction. In T. Palmer, editor, *Predictability of Weather and Climate*, chapter 5, pages 99–123. Cambridge University Press, 2005. doi:10.1017/CBO9780511617652.006.
- [5M-Kal60] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. doi:10.1115/1.3662552.
- [5M-Kal02] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2002. ISBN 978-0521796293.
- [5M-Lio71] J. L. Lions. *Optimal control of systems governed by partial differential equations*. Springer-Verlag, 1971. ISBN 978-0387051154.
- [5M-Lor65] E. N. Lorenz. A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17(3):321–333, 1965. doi:10.1111/j.2153-3490.1965.tb01424.x.
- [5M-PGBB98] T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, 55(4):633–653, 1998. doi:10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2.
- [5M-RL10] M. Rognes and A. Logg. Automated goal-oriented error control. I: stationary variational problems. Submitted to SIAM Journal on Scientific Computing, 2010.
- [5M-Sch07] P. J. Schmid. Nonmodal stability theory. *Annual Review of Fluid Mechanics*, 39(1):129–162, 2007. doi:10.1146/annurev.fluid.38.050304.092139.
- [5M-Tar05] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- [5M-TB97] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. Society for Industrial Mathematics, 1997.
- [5M-TE05] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, 2005.
- [5M-Wun96] C. Wunsch. *The ocean circulation inverse problem*. Cambridge University Press, 1996. ISBN 9780521480901. doi:10.1017/CBO9780511629570.
- [5M-ZHMT12] L. Zanna, P. Heimbach, A. M. Moore, and E. Tziperman. Upper-ocean singular vectors of the North Atlantic climate with implications for linear predictability and variability. *Quarterly Journal of the Royal Meteorological Society*, 138(663):500–513, 2012. doi:10.1002/qj.937.