Automatic Generation of Efficient Adjoint Code for the Parallel MIT General Circulation Model

Patrick Heimbach[1] **, Chris Hill[1] **, and Ralf Giering[2] **

[1] Massachusetts Institute of Technology, EAPS, Cambridge, MA, USA,
{heimbach, cnh}@mit.edu, http://mitgcm.org
[2] FastOpt, Hamburg, Germany,
ralf.giering@fastopt.de, http://www.fastopt.de

In one of the most complex Earth science inverse modeling initiatives, the *Estimation of the Circulation and Climate of the Ocean* (ECCO) project is developing greatly improved estimates of the three-dimensional, time-evolving state of the global oceans [1, 2]. To this end, the project is applying extimation/optimal control techniques to constrain a state-of-the-art parallel general circulation model (MITgcm) [3, 4] with a diverse mix of observations, ranging from in-situ data to satellite measurements. A central ingredient of this effort is the automated generation of scalable, adjoint codes.

Background Combining, through a formal optimization procedure, the fragmentary observations with a numerical model, produces an ocean state that is spatially and temporally complete and simultaneously consistent with observations and model dynamics. The difference between a model and observation is expressed in terms of a scalar cost,

$$\mathcal{J} = \sum_{i=1}^{n} (M_i - O_i) W_i (M_i - O_i) \tag{1}$$

where M_i refers to a simulated quantity projected onto the i^{th} observation O_i, with W_i the associated a priori error estimate. Denoting certain model parameters, initial state variables and boundary values as adjustable "controls" C, $\mathcal{J}$ is minimized under the side condition of fulfilling the model equations. This leads to a constrained optimization problem for which the gradient

$$\boldsymbol{\nabla}_C \, \mathcal{J}(C, M(C)) \; = \; 0 \tag{2}$$

is used to reduce $\mathcal{J}$ iteratively. It may be transformed into an unconstrained one by incorporating the model equations into the cost function (1) via the method of Lagrange multipliers. Alternatively and equivalently, the gradient may be obtained through application of the chain rule to equation (1).

Automatic differentiation (AD) AD exploits the chain rule in a rigorous manner to produce, from a given model code, corresponding adjoint forms (see e.g. [5]). The adjoint form is extremely efficient for scalar-valued cost functions, for which the gradient (2) can be computed in a single integration. Using AD, we are able to numerically evaluate (2), for any scalar $\mathcal{J}$, in roughly five times the compute cost of evaluating $\mathcal{J}$. Note that, for finite difference perturbation methods, the computation of the full gradient scales with the dimension of the control space. Our prognostic code, MITgcm, is compatible with the Tangent linear and Adjoint Model Compiler (TAMC), developed by Ralf Giering ([6],

** On behalf of the ECCO Consortium, http://www.ecco-group.org

[7]). TAMC is a source-to-source transformation tool. Treating a given forward code as a composition of operations – each line representing a compositional element, the chain rule is rigorously applied to the code, line by line. The resulting adjoint code, then, may be thought of as the composition in reverse order of the Jacobian matrices of the full forward code's compositional elements.

Computational challenges A major challenge of the reverse mode is the fact that the intermediate model state has to be available in reverse sequence. In principle this could be achieved by either storing the intermediate states or by successive recomputing of the forward trajectory. Either approach, in its pure form is prohibitive; storing of the full trajectory is limited by available fast-access, storage media, recomputation is limited by CPU resource requirements which scale as the square of the number of intermediate steps. Through the Efficient Recomputation Algorithm (ERA,[8]) and user-inserted directives into the code, TAMC enables a flexible balance between storing and recomputation (checkpointing) thus reducing both by several orders of magnitudes [9].

A critical requirement of the adjoint code is the preservation of scalability of the forward code's domain decomposition. The integration is split into an extensive computational phase characterized by on-processor operations only for each domain element or virtual processor (tile). Periodically, between computational phases, global arithmetic or communication primitives are invoked which can use MPI, OpenMP, or combination thereof to communicate data between tiles. These performance-critical primitives and their mapping onto specific platforms are part of a commuication layer of the MITgcm in a custom software library called WRAPPER. Corresponding adjoint WRAPPER functions were written by hand which implement transpose forms of the linear operations that are required to support domain decomposition. In adjoint calculations these transpose forms are substituted automatically as appropriate. In this way the adjoint forms of MITgcm automatically inheret the parallel scaling of the base MITgcm prognostic code.

Applications The size of the ECCO optimization problem is formidable. Our "smallest" current configuration is characterised by a cost function that spans nine-years of planetary scale ocean simulation and observation, operating on 10^8 elements, and optimized by corrections to a control vector, C, of size 1.5×10^8. The full Jacobian, $\frac{\partial \mathcal{J}}{\partial C}$, for this system contains more than 10^{16} elements ($10^8 \cdot 1.5 \times 10^8$) which, even allowing for some sparsity, is fundamentally impractical. Therefore, the reverse mode of AD plays a central role. The optimization produces a best estimate of the ocean circulation over the integration period that is both consistent with the model dynamics as well as the observation. Furthermore, it produces corrections of the surface fluxes, in this case the daily NCEP heat, freshwater and momentum fluxes.

Complementary to the estimation problem, the adjoint method has been applied to rigorous process studies of dynamical properties such as the sensitivity of the North Atlantic heat transport to changes in temperature, salinity, surface fluxes or mixing coefficients [10]. Longer time-scale adjoint sensitivity calculations are providing insights into the dynamical balances controlling thermohaline overturning in general circulations models. Applying the adjoint approach to tracer calculations is providing new ways to characterise water mass pathways over centennial to millenial time-scales. The results are comparable to a brute force Green's function perturbations, but are available for a fraction of the computational cost. Work is currently underway to include an atmospheric component. Such studies provide a innovative route to further our understanding of air-sea interaction processes, and, together with quantitative estimates of the oceanic state, provide essential ingredients for climate studies.

Conclusions An invaluable by-product of our adjoint based state-estimation work has been the routine creation of tangent-linear and adjoint forms of the MITgcm ocean model. The capability to numerically evaluate Jacobian elements about the fully non-linear trajectory of a parallel ocean model is proving to be a powerful tool for dynamical analysis. Future HPC climate modeling efforts could benefit greatly if they are designed to accomodate ready maintenance of tangent-linear and adjoint codes. The potential benefits extend far beyond the realm of assimilation and observation synthesis.

Acknowledgements This is paper is a contribution to the ECCO project, supported by NOPP, and with funding from NASA, NSF and ONR.

References

1. Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., Marshall, J.: The global ocean circulation and transports during 1992 – 1997, estimated from ocean observations and a general circulation model. Report 4, (2000a) J. Geophys. Res., in press.
2. Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., Marshall, J.: The global ocean circulation and transports during 1992 –1997, estimated from ocean observations and a general circulation model. part ii: Testing results. Technical Report 5, The ECCO Consortium (2000b) submitted for publication in J. Geophys. Res.
3. Marshall, J., Hill, C., Perelman, L., Adcroft, A.: Hydrostatic, quasi-hydrostatic and nonhydrostatic ocean modeling. J. Geophys. Res. **102, C3** (1997a) 5,733–5,752
4. Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C.: Hydrostatic, quasi-hydrostatic and nonhydrostatic ocean modeling. J. Geophys. Res. **102, C3** (1997b) 5,753–5,766
5. Griewank, A.: Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation. Volume 19 of Frontiers in Applied Mathematics. SIAM, Philadelphia (2000)
6. Giering, R., Kaminski, T.: Recipes for adjoint code construction. ACM Transactions on Mathematical Software **24** (1998) 437–474
7. Giering, R.: Tangent linear and adjoint model compiler. users manual 1.4 (tamc version 5.2). Report , MIT, MIT/EAPS, Cambridge (MA), USA (1999)
8. Giering, R., Kaminski, T.: Generating recomputations in reverse mode AD. In Corliss, G., Faure, C., eds.: Automatic Differentiation 2000: From Simulation to Optimization, Springer (2000c) in press.
9. Griewank, A.: Achieving logarithmic growth of temporal and spatial complexity in reverse AD. Optimization Methods and Software **1** (1992) 35–54
10. Marotzke, J., Giering, R., Zhang, K., Stammer, D., Hill, C., Lee, T.: Construction of the adjoint mit ocean general circulation model and application to atlantic heat transport variability. J. Geophys. Res. **104, C12** (1999) 29,529–29,547