Scalable, Adaptive Methods for Forward and Inverse Modeling of Continental-Scale Ice Sheet Flow

Tobin Isaac ⟨tisaac@ices.utexas.edu⟩
Center for Computational Geosciences & Optimization, The University of Texas at Austin
SC14, New Orleans

(A version of this poster is available at http://users.ices.utexas.edu/~tisaac/.)

The Bayesian Inverse Problem

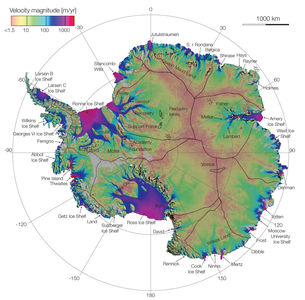
Climate predictions depend on ice sheet dynamics

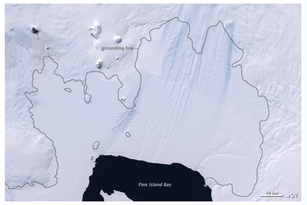
From the IPCC Fourth Assessment Report:

"Dynamical processes related to ice flow not included in current models but suggested by recent observations could increase the vulnerability of the ice sheets to warming, increasing future sea level rise. Understanding of these processes is limited and there is no consensus on their magnitude." [Summary for Policymakers: Projections of Future Changes in Climate]

"[T]he uncertainty in the projections of the land ice contributions [to sea level rise] is dominated by the various uncertainties in the land ice models themselves ... rather than in the temperature projections." [MSC+07]

Ice sheets exhibit complex, localized flows





Landsat mosaic, Pine Island Glacier, Antarctica (NASA Earth Observatory)

[RMS11]

Ice stream formation depends on basal conditions: topography, sediment, till, and subglacial hydrology.

Bayesian inversion for basal parameters

State equations: $\boldsymbol{S}(\boldsymbol{u}, p; m) = \boldsymbol{0}$

Equations of motion (Stokes)

$$-\nabla \cdot [2\mu(\boldsymbol{u})\,\mathbf{D}(\boldsymbol{u}) - Ip] = \rho \boldsymbol{g}, \qquad [\mathbf{D}(\boldsymbol{u}) = \tfrac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)]$$
$$\nabla \cdot \boldsymbol{u} = 0,$$

Nonlinear constitutive relation

$$\mu(\boldsymbol{u}) = B\,\mathbf{D}(\boldsymbol{u})_{\mathrm{II}}^{\frac{1-n}{2n}}, \quad [\mathbf{D}(\boldsymbol{u})_{\mathrm{II}} = \tfrac{1}{2}\mathrm{tr}(\mathbf{D}(\boldsymbol{u})^2)]$$

Boundary conditions

$$\boldsymbol{\sigma n}|_{\Gamma_t} = \boldsymbol{0}, \quad \boldsymbol{u} \cdot \boldsymbol{n}|_{\Gamma_b} = 0,$$
$$\mathrm{T}_{\parallel}(\boldsymbol{\sigma n} + \beta \boldsymbol{u})|_{\Gamma_b} = \boldsymbol{0}. \qquad [\beta = \exp\{m\}]$$

The state variables (velocity and pressure) are implicitly defined by a PDE.

- Solution-dependent effective viscosity
- Shear-thinning flow (typically $n = 3$)
- The parameter fields determines the coefficient of a Robin-type boundary condition in the tangential directions at the base of the ice sheet
- An implicit system of PDEs that is 3D, nonlinear, indefinite, with heterogeneous coefficients, a high-aspect-ratio domain, complex geometry and boundary conditions.

Observations: $\boldsymbol{bu}$

$$\boldsymbol{d}_{\mathrm{obs}} = \{\boldsymbol{u}_{\mathrm{obs}}(\boldsymbol{x}_0), \ldots, \boldsymbol{u}_{\mathrm{obs}}(\boldsymbol{x}_{m-1})\},$$
$$\boldsymbol{bu} = \{\boldsymbol{u}(\boldsymbol{x}_0), \ldots \boldsymbol{u}(\boldsymbol{x}_{m-1})\}.$$

The observation operator restricts the velocity to the points where the observations are made.

Prior distributions and Bayes' rule

$$\pi_{\mathrm{obs}}(\boldsymbol{d}_{\mathrm{obs}}|\boldsymbol{u}) = \mathcal{N}(\boldsymbol{bu}, \mathcal{C}_{\mathrm{obs}})$$
$$\pi_{\mathrm{prior}}(m) = \mathcal{N}(m_{\mathrm{prior}}, \mathcal{C}_{\mathrm{prior}})$$
$$\pi_{\mathrm{post}}(m|\boldsymbol{d}_{\mathrm{obs}}) \propto \pi_{\mathrm{obs}}(\boldsymbol{d}_{\mathrm{obs}}|\boldsymbol{u}(m))\,\pi_{\mathrm{prior}}(m)$$

We assume Gaussian prior distributions of the parameter and the observations. The posterior distribution of parameters given observations is determined by Bayes' rule.

The MAP optimization problem

$$\min_m \mathcal{J}(m) := \tfrac{1}{2}\|\boldsymbol{bu}(m) - \boldsymbol{d}_{\mathrm{obs}}\|^2_{\mathcal{C}_{\mathrm{obs}}^{-1}} + \tfrac{1}{2}\|m - m_{\mathrm{prior}}\|^2_{\mathcal{C}_{\mathrm{prior}}^{-1}},$$
where $\boldsymbol{u}(m)$ solves
$$\boldsymbol{S}(\boldsymbol{u}, p; m) = \boldsymbol{0}.$$

The maximum a posteriori estimate m_{MAP} solves a PDE-constrained optimization problem.

- Parameter space $O(10^5 - 10^6)$
- State space $O(10^7 - 10^9)$
- Evaluating $\mathcal{J}(m)$ requires non-linear PDE solve
- Evaluating $\mathcal{J}'(m)$ requires additional adjoint linear PDE solve
- Applying $\mathcal{J}''(m)\hat{m}$ requires two sequential linear PDE solves

A Scalable Stokes Solver

Discretization/solver co-design [ISG14]

High-order, inf-sup stable discretization

Velocity:

- continuous piecewise tensor polynomials $[\mathbf{Q}_k]^3$, $k > 1$;
- Gauss-Lobatto basis, Gauss-Legendre quadrature;
- fast matrix-free calculations, high flop/memop ratio [Bro10].

Pressure:

- discontinuous, piecewise $\mathbf{Q}_{k-2}$, Gauss-Legendre basis;
- inf-sup constant is $o(k^{-1})$;
- local, element-wise conservation of mass;
- robust at high aspect-ratios [TS03];

Nonlinear Solver

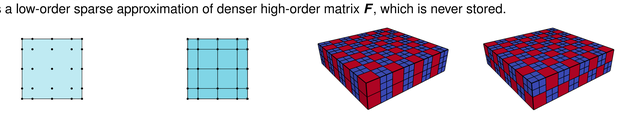
We solve $\boldsymbol{S}(\boldsymbol{u}, p; m) = \boldsymbol{0}$ using an inexact Newton-Krylov method [EW96] with backtracking.

- Newton update system is
$$\frac{\partial \boldsymbol{S}}{\partial\{\boldsymbol{u}, p\}}(\boldsymbol{u}, p; m)[\tilde{\boldsymbol{u}}, \tilde{p}] = -\boldsymbol{S}(\boldsymbol{u}, p; m),$$
which is a Stokes system with a fourth-order tensor for effective viscosity,
$$\mu'(\boldsymbol{u}) = \left\{ I - \frac{1-n}{2n} \frac{\mathbf{D}(\boldsymbol{u}) \otimes \mathbf{D}(\boldsymbol{u})}{\mathbf{D}(\boldsymbol{u})_{\mathrm{II}}} \right\}.$$
- Inexact Newton stresses preconditioner construction/reconstruction because early Krylov solves are inaccurate.

Linear Solver

The Newton step Jacobian is a symmetric saddle point operator
$$\mathcal{S}(\boldsymbol{u}, \beta) = \begin{pmatrix} \boldsymbol{F}(\boldsymbol{u}; m) & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{pmatrix}.$$

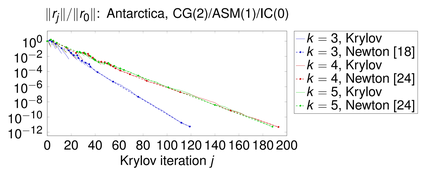
- Solve using right-preconditioned (F)GMRES with block preconditioner
$$\tilde{\mathcal{S}} = \begin{pmatrix} \tilde{\boldsymbol{F}} & \nabla \\ \boldsymbol{0} & \tilde{\boldsymbol{S}} \end{pmatrix},$$
where $\tilde{\boldsymbol{S}}$ is a diagonal approximation to the Schur complement $\boldsymbol{S} = -\boldsymbol{B}\boldsymbol{F}^{-1}\boldsymbol{B}^T$.
- $\tilde{\boldsymbol{F}}$ is a low-order sparse approximation of denser high-order matrix $\boldsymbol{F}$, which is never stored.



- Sparse approximation at non-conforming element boundaries introduces additional complexity.
- $\tilde{\boldsymbol{F}}$ (second-order elliptic system with rigid-body motions in near-nullspace) is preconditioned with smoothed-aggregation AMG v-cycle.
  - Pointwise smoothers (Jacobi, Gauss-Seidel) are less effective on high-aspect-ratio finite elements: we use Chebyshev(1)/ASM(i)/IC(j) or CG(2)/ASM(i)/IC(j) smoothers. [CG = conjugate gradient, ASM = additive Schwarz method, IC = incomplete Cholesky factorization].

Performance and Scalability

Antarctic ice sheet mesh, multiple values of k



Antarctic ice sheet problem: convergence of the incomplete Newton-Krylov solver for regularized geometry. The mesh is distributed over 512 processes. The bracketed numbers are the total number of nonlinear iterations. Problem sizes are 22M, 52M, and 101M dofs for $k = 3, 4, 5$.

Weak scalability, linear solver

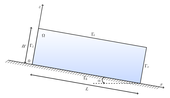
P	N	solve	eff.	#K	matvec	eff.	apply	eff.	setup	eff.
128	11.8M	14.6	1.00	47	0.0223	1.00	0.246	1.00	3.78	1.00
1024	82.6M	15.9	0.92	51	0.0217	1.03	0.245	1.00	4.32	0.88
8192	615M	28.7	0.51	53	0.0234	0.95	0.258	0.95	11.7	0.32

Antarctic ice sheet problem: weak scaling of the linear solver for $k = 3$ for regularized geometry. The smoother for the (1,1)-block AMG v-cycle uses ASM(0): all other solver parameters are the same as above. We report the convergence of the linear Stokes problem with constant μ, which we solve to a relative tolerance of 10^{-15}.

Large Scale Bayesian Inversion and UQ

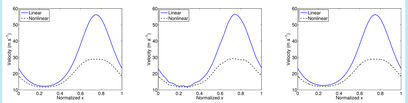
Deterministic model problem [PZS+12]

Setup



- Periodic domain
- Inverted for linear and nonlinear rheology
- Variable domain length L, signal-to-noise ratio
- Regularization parameter α chosen by Morozov's criterion: misfit approximately matches noise in magnitude

Example: $L = 40$, SNR $= 100$



Slice of the horizontal surface velocity from linear and nonlinear rheology (left), noisy data (middle), and reconstructed velocity (right).

Reconstruction table

L	linear rheology: Signal-to-noise ratio 500	100	20	10	nonlinear rheology: Signal-to-noise ratio 500	100	20	10
5	0.031	0.112	*	*	*	*	*	*
10	0.022	0.049	0.157	0.310	0.032	0.165	*	*
20	0.017	0.041	0.109	0.177	0.026	0.045	0.118	0.285
40	0.013	0.039	0.125	0.191	0.015	0.036	0.097	0.149

- Sharper boundary layer in nonlinear rheology weakens response
- Recovery of high frequency (small L) variations in β much worse for nonlinear rheology
- Noise affects recovery more for nonlinear rheology

Gauss-Newton versus Nonlinear CG

mesh	#dof Stokes	#β param	NCG (FR/PR)* #iter	#PDE	Gauss-Newton** #iter	#PDE
$10\times10\times2$	6978	121	54/57	622/676	10 (41)	104
$20\times20\times2$	26538	441	57/58	620/572	10 (43)	108
$40\times40\times2$	103458	1681	99/87	674/648	10 (48)	118
$80\times80\times2$	408498	6561	125/102	664/623	10 (46)	114

* Nonlinear Conjugate Gradient (FR (Fletcher-Reeves) / PR (Polak-Ribière))
** Inexact Gauss-Newton-CG
*** Iteration is terminated when the norm of the gradient drops below a relative tolerance of 10^{-5}.

- Number of Newton inversion iterations and required PDE solves are insensitive to the number of inversion parameters.

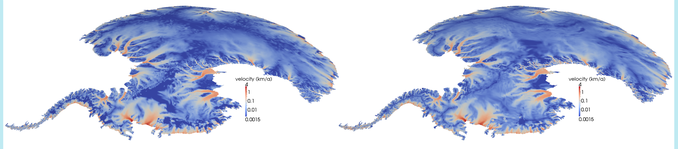
Continental-scale UQ [IPSG14]

Priors

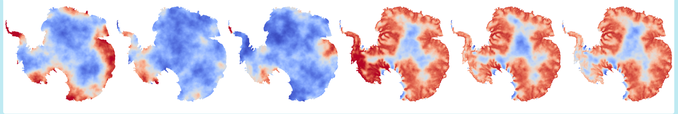
$$\mathcal{C}_{\mathrm{obs}}(\boldsymbol{u}, \boldsymbol{v}) = \int_{\Gamma_{\mathrm{top}}} \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{|\boldsymbol{d}_{\mathrm{obs}}|^2 + \epsilon}\, ds$$
$$\mathcal{C}_{\mathrm{prior}}(m, n) \approx \int_{\Gamma_{\mathrm{base}}} (-\gamma \Delta m + \delta m)(-\gamma \Delta n + \delta n)\, ds$$

- Misfit weighed by observed velocities.
- Higher-order derivatives required so that Gaussian measure converges in the limit as $h \to 0$.

Observations and MAP reconstruction



Prior and posterior samples



Parallel Adaptive Mesh Refinement

p4est: parallel quadtree/octree AMR

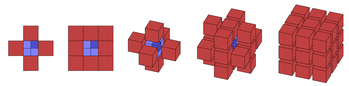
p4est.org

The p4est library provides scalable parallel adaptive mesh refinement using the forest-of-quadtrees/octrees paradigm:

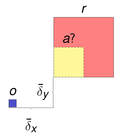
- hierarchically refined quadrants/octants;
- flat storage (leaves only);
- fully-distributed by space-filling-curve partitioning;
- refinement, coarsening, and repartitioning have straight-forward, low-communication, easily scalable algorithms.

The bottlenecks are **mesh-wide local-topology operations**: mesh improvement (enforcing a 2:1 balance condition) and mesh description (converting to adjacency-graph mesh formats).

2:1 Balance [IBG12]



2:1 conditions are *local* requirements: limiting the sizes of neighbors. Scalable enforcement in parallel requires kernels for *action at a distance*:

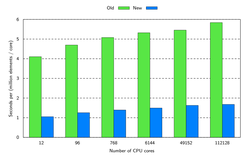


Function $\lambda(\tilde{\delta})$ such that $\mathrm{size}(\boldsymbol{a}) = \lfloor \log_2 \lambda \rfloor$

k	1D	2D	3D
1	$\tilde{\delta}$	$\tilde{\delta}_x + \tilde{\delta}_y$	$\mathtt{Carry3}(\tilde{\delta}_y + \tilde{\delta}_z, \tilde{\delta}_z + \tilde{\delta}_x, \tilde{\delta}_x + \tilde{\delta}_y)$
2		$\max\{\tilde{\delta}_x, \tilde{\delta}_y\}$	$\mathtt{Carry3}(\tilde{\delta}_x, \tilde{\delta}_y, \tilde{\delta}_z)$
3			$\max\{\tilde{\delta}_x, \tilde{\delta}_y, \tilde{\delta}_z\}$

$$\mathtt{Carry3}(\alpha, \beta, \gamma) \equiv \max\{\alpha, \beta, \gamma, \alpha + \beta + \gamma - (\alpha|\beta|\gamma)\}$$

$$\mathtt{Carry3}(\alpha, \beta, \gamma) \equiv \max\{\alpha, \beta, \gamma, \alpha + \beta + \gamma - (\alpha|\beta|\gamma)\}.$$



Deriving and implementing these kernels (and other improvements) led to a 4-5x speedup:

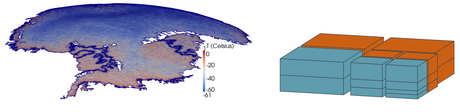
Conversion to adjacency-graph [IBWG14]

Exploiting the hierarchical structure of quadtrees and octrees via recursive algorithms leads to 2-3x speedup on deep-cache architectures.

Old search/hash-based algorithm (top) vs. New recursion-based algorithm (bottom) metrics, conversion to FE mesh (Ivy Bridge)

N	runtime (ms)	instructions	branch misses	cache misses
4.6×10^3	9.5×10^0	4.3×10^7	2.1×10^5	2.2×10^4
	1.0×10^1	3.7×10^7	5.3×10^4	1.1×10^4
3.9×10^4	8.6×10^1	4.2×10^8	1.7×10^6	2.2×10^5
	4.0×10^1	3.1×10^8	3.6×10^5	5.1×10^4
3.2×10^5	8.4×10^2	3.7×10^9	1.3×10^7	4.8×10^6
	3.5×10^2	2.5×10^9	2.7×10^6	4.5×10^5
2.6×10^6	8.0×10^3	3.3×10^{10}	1.0×10^8	6.1×10^7
	2.8×10^3	2.0×10^{10}	2.2×10^7	5.4×10^6

p4est in ice sheet modeling



- (left) Refinement is often highly localized (e.g. grounding lines and ice streams).
- (right) specialized hybrid quadtree/octree scheme decouples horizontal and vertical refinement (unlike isotropic octree refinement). (A presentation on the hybrid scheme is found at http://users.ices.utexas.edu/~tisaac/slides/).

[Bro10] Jed Brown. Efficient nonlinear solvers for nodal high-order finite elements in 3D. *Journal of Scientific Computing*, 45(1-3):48–63, 2010.
[EW96] S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*, 17:16–32, 1996.
[IBG12] Tobin Isaac, Carsten Burstedde, and Omar Ghattas. Low-cost parallel algorithms for 2:1 octree balance. In *Proceedings of the 26th IEEE International Parallel & Distributed Processing Symposium*. IEEE, 2012.
[IBWG14] Tobin Isaac, Carsten Burstedde, Lucas C. Wilcox, and Omar Ghattas. Recursive algorithms for distributed forests of octrees. In preparation, 2014.
[IPSG14] Tobin Isaac, Noemi Petra, Georg Stadler, and Omar Ghattas. From data to prediction with quantified uncertainties: Scalable parallel algorithms and applications to the dynamics of the Antarctic ice sheet. Journal of Computational Physics (submitted), 2014.
[ISG14] Tobin Isaac, Georg Stadler, and Omar Ghattas. Solution of nonlinear Stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics. *SIAM Journal on Scientific Computing (submitted)*, 2014.
[MSC+07] G. A. Meehl, T. F. Stocker, W. D. Collins, A. T. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver, and Z.-C. Zhao. Global climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 747–845. Cambridge University Press, 2007.
[PZS+12] Noemi Petra, Hongyu Zhu, Georg Stadler, Thomas J. R. Hughes, and Omar Ghattas. An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model. *Journal of Glaciology*, 58(211):889–903, 2012.
[RMS11] E Rignot, J Mouginot, and B Scheuchl. Ice flow of the Antarctic ice sheet. *Science*, 333(6048):1427–1430, 2011.
[TS03] A. Toselli and C. Schwab. Mixed hp-finite element approximations on geometric edge and boundary layer meshes in three dimensions. *Numerische Mathematik*, 94(4):771–801, 2003.