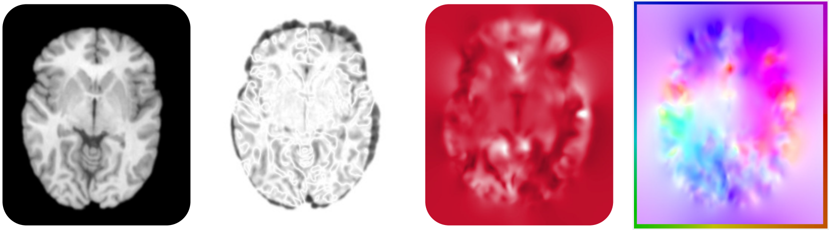
Fast Algorithms for Nonlinear Optimal Control for Diffeomorphic Registration



Andreas Mang

Department of Mathematics, University of Houston

RICAM, New Trends in PDE-Constrained Optimization, 10/17/2019

Teaser: CLAIRE

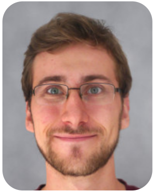
	unknowns	CPUs	GPUs	runtime
50M	(256^3)	512	—	<2 sec
50M	(256^3)	1	1	≈6 sec
200B	(4096^3)	8192	—	≈3.5 min

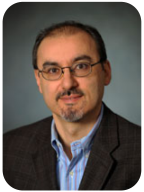
`http://andreasmang.github.io/claire`

[Mang et al., 2016, Gholami et al., 2017, Mang et al., 2019]


R. Azencott
Math UHouston


G. Biros
Oden UTAustin


M. Brunn
CS UStuttgart


C. Davatzikos
CBIA UPenn

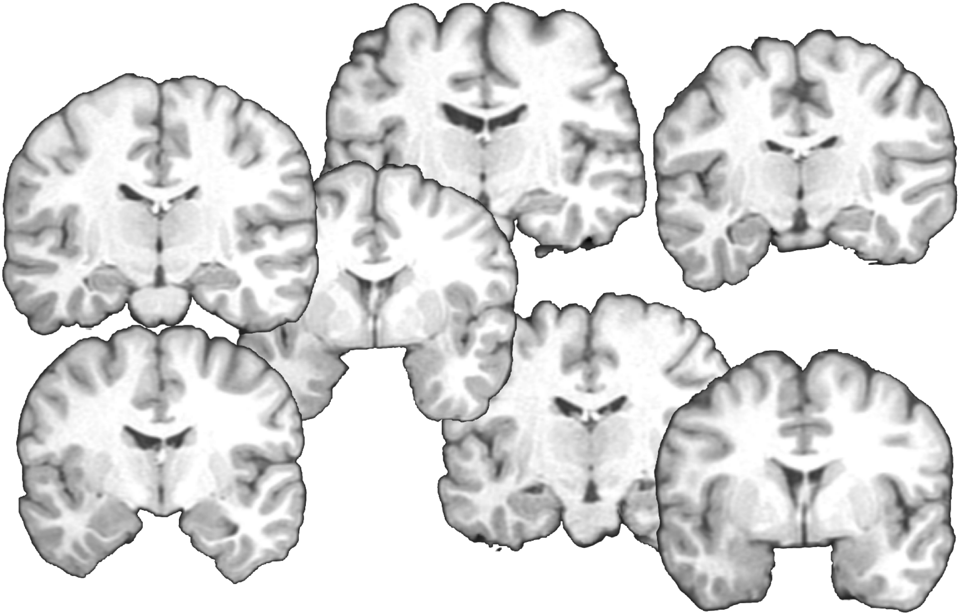

J. He
Math UHouston


J. Herring
Math UHouston


N. Himthani
Oden UTAustin


J. Kim
Math UHouston


M. Mehl
CS UStuttgart



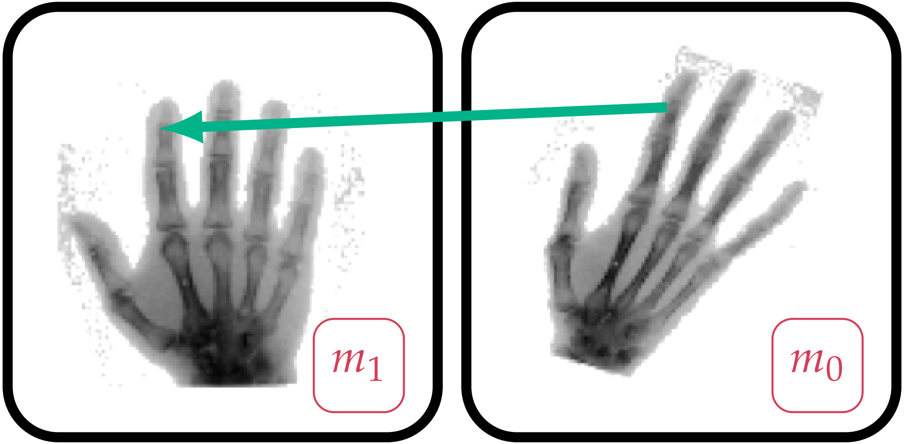
Inverse Problem

find a *plausible* map $\boldsymbol{y} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

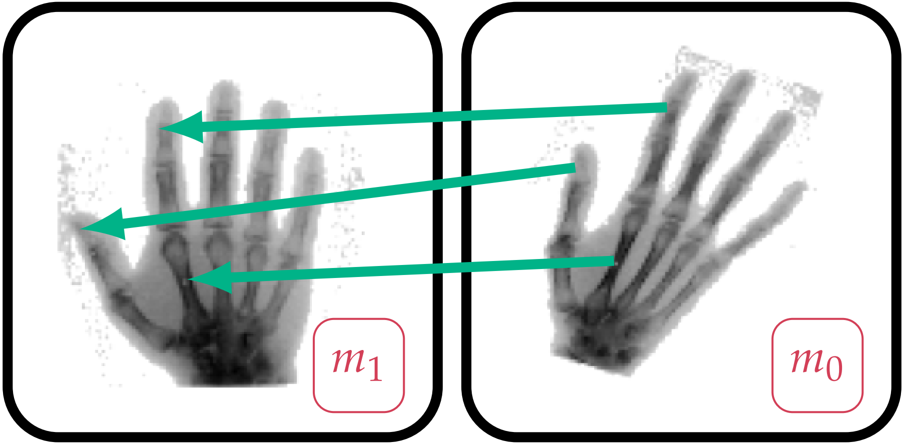
$$(m_0 \circ \boldsymbol{y})(\boldsymbol{x}) = m_1(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in \mathbb{R}^d$$

m_0 m_1 $m_0 \circ \boldsymbol{y}$

[Amit, 1994, Modersitzki, 2009, Modersitzki, 2004, Fischer and Modersitzki, 2008]

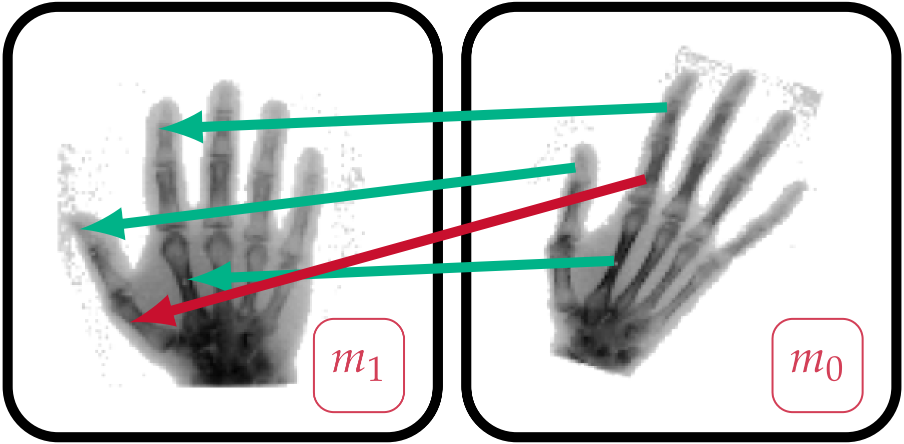


[Amit, 1994, Modersitzki, 2009, Modersitzki, 2004, Fischer and Modersitzki, 2008]



$$\boldsymbol{y} \in \mathrm{diff}(\Omega)$$

[Amit, 1994, Modersitzki, 2009, Modersitzki, 2004, Fischer and Modersitzki, 2008]



$$\boldsymbol{y} \notin \mathrm{diff}(\Omega)$$

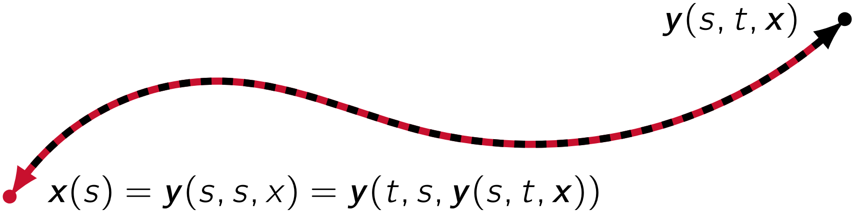
[Amit, 1994, Modersitzki, 2009, Modersitzki, 2004, Fischer and Modersitzki, 2008]

Building Blocks

Flows of Diffeomorphisms

introduce pseudo-time variable $t \in [0, 1]$ and parameterize $\boldsymbol{y}$ by $\boldsymbol{v}$

$$\partial_t \boldsymbol{y} = \boldsymbol{v}(\boldsymbol{y}), \quad \boldsymbol{y}(0) = \mathrm{id}_{\mathbb{R}^d}$$



[Younes, 2010]

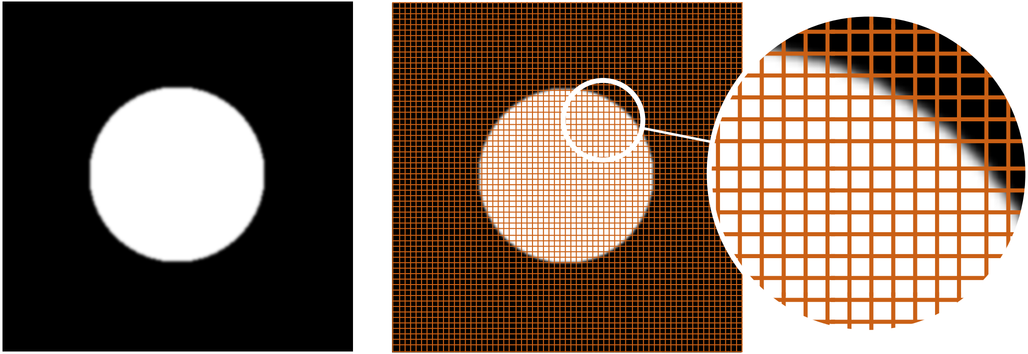
Optimal Control Problem (Prototype)

$$\underset{\boldsymbol{v},\,\boldsymbol{y}}{\text{minimize}} \;\; \operatorname{dist}(\boldsymbol{y}(1)\cdot m_0, m_1) + \operatorname{reg}(\boldsymbol{v})$$

$$\text{subject to} \;\; \partial_t \boldsymbol{y} = \boldsymbol{v}(\boldsymbol{y}), \quad \boldsymbol{y}(0) = \mathrm{id}_{\mathbb{R}^d}$$

Large Deformation Diffeomorphic Metric Mapping

[Younes, 2010, Beg et al., 2005]



Regularity

$$\partial_t \boldsymbol{y} = \boldsymbol{v}(\boldsymbol{y}), \quad \boldsymbol{y}(0) = \mathrm{id}$$

$$\boldsymbol{v} \in L^2([0,1], \mathcal{V}), \mathcal{V} \hookrightarrow W^{s,2}(\mathbb{R}^3)^3, s > 5/2$$

$$\implies \boldsymbol{y} \in G_{\mathcal{V}} \subseteq \mathrm{diff}(\mathbb{R}^3)$$

$$(\text{smoothness class } \; 1 \leq r \leq s - 3/2)$$

[Beg et al., 2005, Trouve, 1998, Dupuis et al., 1998]

Regularity

$$\int_0^1 \|\boldsymbol{v}(t)\|_{\mathcal{V}}^2 \, \mathrm{d}t = \int_0^1 \langle \mathcal{L}\boldsymbol{v}(t), \boldsymbol{v}(t)\rangle_{L^2(\Omega)^d} \, \mathrm{d}t$$

$$\mathcal{L} : \mathcal{V} \to \mathcal{V}^*, \quad \mathcal{L} := (1 - \gamma^2 \Delta)^\kappa \mathrm{id}, \quad \gamma, \kappa > 0$$

$$\mathrm{dist}_G(\mathrm{id}_{\mathbb{R}^d}, \boldsymbol{\phi})^2 = \inf_{\boldsymbol{v}} \left\{ \int_0^1 \|\boldsymbol{v}\|_{\mathcal{V}}^2 \, \mathrm{d}t : \boldsymbol{\phi} = \boldsymbol{y}(1) \right\}$$

$$\partial_t \boldsymbol{y} = \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{y}(0) = \mathrm{id}_{\mathbb{R}^d}$$

[Beg et al., 2005]

Regularity (RKHS)

$$\mathcal{V} \equiv \mathcal{V}_\kappa \quad \text{(RKHS with associated kernel } \kappa\text{)}$$

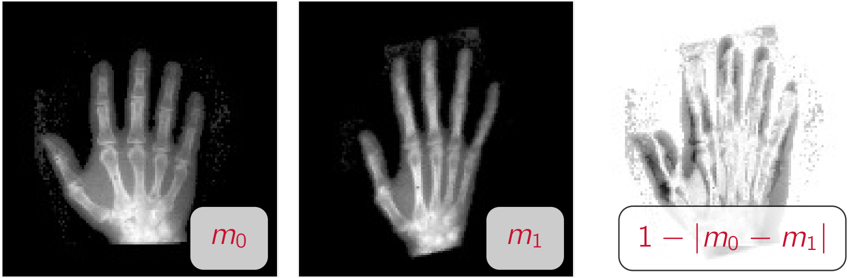
$$\boldsymbol{v}(t, \boldsymbol{x}) := \textstyle\sum_{j=1}^{q} \kappa(\boldsymbol{x}_j(t), \boldsymbol{x})\boldsymbol{\alpha}_j(t)$$

$$\|\boldsymbol{v}(t)\|_{\mathcal{V}}^2 = \sum_{j=1}^{q}\sum_{k=1}^{q} \kappa(\boldsymbol{x}_j(t), \boldsymbol{x}_k(t))\boldsymbol{\alpha}_j^\intercal(t)\boldsymbol{\alpha}_k(t)$$

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) \propto \exp(-0.5\|\boldsymbol{x} - \boldsymbol{y}\|_{\Sigma^{-1}}^2)$$

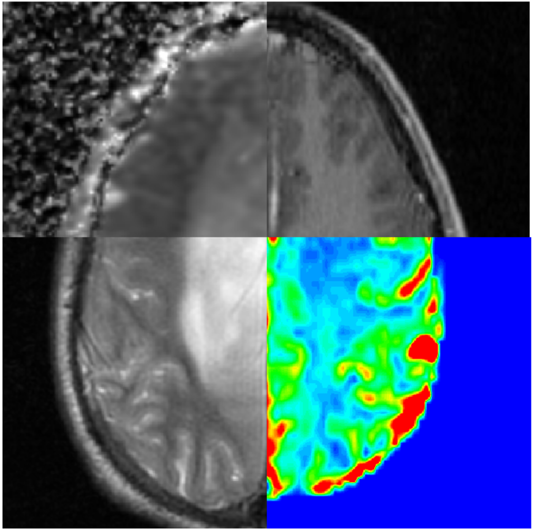
Distance Functional

$$\mathrm{dist}_{\mathrm{SSD}}(m_0, m_1) = \|m_0 - m_1\|^2_{L^2(\Omega)}$$



[Sotiras et al., 2013, Modersitzki, 2009]

Distance Functional



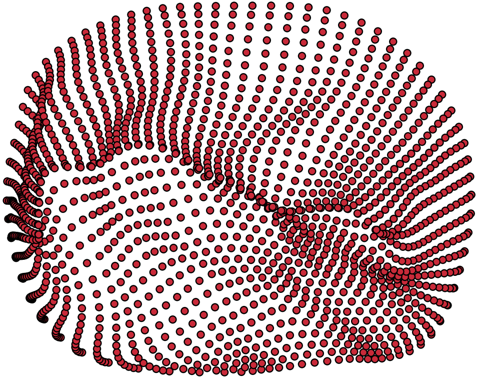
Distance Functional

$$\mathrm{dist}_{\mathrm{CC}}(m_0, m_1) = \frac{\langle m_1, m_0 \rangle_{L^2(\Omega)}}{\langle m_1, m_1 \rangle_{L^2(\Omega)} \langle m_0, m_0 \rangle_{L^2(\Omega)}}$$

$$\mathrm{dist}_{\mathrm{NGF}}(m_0, m_1) = \int_\Omega 1 - ((\tilde{\nabla} m_0)^\intercal \tilde{\nabla} m_1)^2 \,\mathrm{d}\boldsymbol{x}$$

[Sotiras et al., 2013, Modersitzki, 2009, Haber and Modersitzki, 2006]

Distance Functional (RKHS)



$$\boldsymbol{s}_j := \{\boldsymbol{x}_1^j, \ldots, \boldsymbol{x}_k^j\}, \quad j = 1, 2, \ldots$$

[Azencott et al., 2010]

Distance Functional (RKHS)

$$\begin{aligned}\mathrm{dist}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \tfrac{1}{k^2}\Big(&\textstyle\sum_{i=1}^{k}\sum_{j=1}^{k}\kappa(\boldsymbol{x}_i^1, \boldsymbol{x}_j^1)\\ &- \textstyle\sum_{i=1}^{k}\sum_{j=1}^{m} 2\kappa(\boldsymbol{x}_i^1, \boldsymbol{x}_j^2)\\ &+ \textstyle\sum_{i=1}^{k}\sum_{j=1}^{k}\kappa(\boldsymbol{x}_i^2, \boldsymbol{x}_j^2)\Big)\end{aligned}$$

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) \propto \exp(-0.5\|\boldsymbol{x} - \boldsymbol{y}\|^2_{\Sigma^{-1}})$$

[Azencott et al., 2010]

Formulations

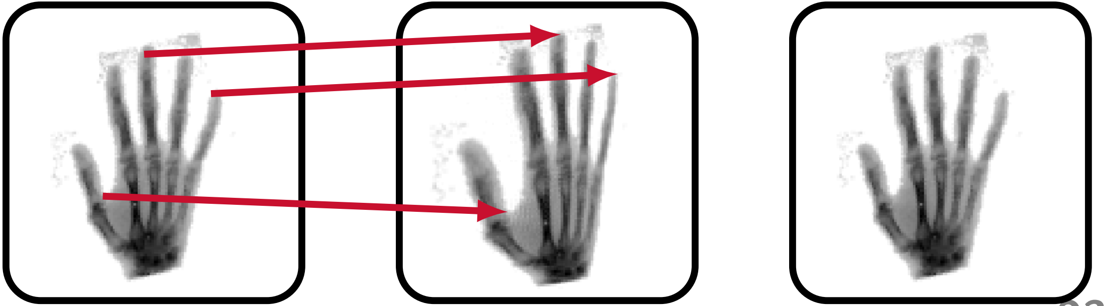
Optimal Control Problem

$$\begin{aligned} \underset{\mathbf{v},\, \mathbf{y}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{y}(1)\cdot m_0 - m_1\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|\mathbf{v}\|^2_{L^2([0,1],\mathcal{V})} \\ \text{subject to} \quad & \partial_t \mathbf{y} = \mathbf{v}(\mathbf{y}), \quad \mathbf{y}(0) = \mathrm{id}_{\mathbb{R}^d} \end{aligned}$$

[Younes, 2010, Beg et al., 2005]

Deformation Model

$$\begin{aligned} \partial_t m + \langle \mathbf{v}, \nabla m \rangle &= 0 \quad \text{in } \Omega \times (0,1] \\ m &= m_0 \quad \text{in } \Omega \times \{0\} \end{aligned}$$



Optimization Problem

$$\begin{aligned} \underset{\boldsymbol{v},\, m}{\text{minimize}} \quad & \frac{1}{2}\|m(1) - m_1\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|\boldsymbol{v}\|^2_{L^2([0,1],\mathcal{V})} \\ \text{subject to} \quad & \partial_t m + \langle \boldsymbol{v}, \nabla m \rangle = 0 \\ & m = m_0 \\ & (\operatorname{div} \boldsymbol{v} = 0) \end{aligned}$$

[Arguilière et al., 2016, Chen and Lorenz, 2012, Barbu and Marinoschi, 2016, Borzi et al., 2002, Hart et al., 2009, Herzog et al., 2019, Jarde and Ulbrich, 2019, Vialard et al., 2012]

Solver

Numerical Optimization

Lagrangian

$$\underset{\boldsymbol{v}, m}{\text{minimize}} \quad \mathcal{J}(\boldsymbol{v}) \quad \text{subject to} \;\; \mathcal{C}(\boldsymbol{v}, m) = 0$$

$$\mathcal{L}(\boldsymbol{v}, m, \lambda) := \mathcal{J}(\boldsymbol{v}) + \langle \lambda, \mathcal{C}(\boldsymbol{v}, m) \rangle_{L^2(\Omega)^k}$$

[Biegler et al., 2003, Borzi and Schulz, 2012, Hinze et al., 2009, Lions, 1971]

Optimality Conditions

$$\mathbf{g}(\mathbf{w}^\star) = \begin{bmatrix} \mathbf{g}^m \\ \mathbf{g}^v \\ \mathbf{g}^\lambda \end{bmatrix} (\mathbf{w}^\star) = \mathbf{0}, \quad \mathbf{w}^\star = \begin{bmatrix} \mathbf{m}^\star \\ \mathbf{v}^\star \\ \boldsymbol{\lambda}^\star \end{bmatrix} \in \mathbb{R}^n, \quad n \gg 1\text{e}6$$

$$\boldsymbol{g}(\boldsymbol{w}) = \partial_\varepsilon \mathcal{L}(\boldsymbol{w} + \varepsilon \tilde{\boldsymbol{w}})\big|_{\varepsilon=0}$$

(optimize-then-discretize)

[Biros and Ghattas, 2005a, Biros and Ghattas, 2005b, Haber and Ascher, 2001]

Full Space Method

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k \tilde{\mathbf{w}}_k$$

$$\underbrace{\begin{bmatrix} \mathbf{H}_{mm} & \mathbf{H}_{mv} & \mathbf{A}^\top \\ \mathbf{H}_{vm} & \mathbf{H}_{\text{reg}} & \mathbf{C}^\top \\ \mathbf{A} & \mathbf{C} & \mathbf{0} \end{bmatrix}_k}_{\mathbf{H}_k} \underbrace{\begin{bmatrix} \tilde{\mathbf{m}} \\ \tilde{\mathbf{v}} \\ \tilde{\boldsymbol{\lambda}} \end{bmatrix}_k}_{\tilde{\mathbf{w}}_k} = - \underbrace{\begin{bmatrix} \mathbf{g}^m \\ \mathbf{g}^v \\ \mathbf{g}^\lambda \end{bmatrix}_k}_{\mathbf{g}_k}$$

[Biros and Ghattas, 2005a, Biros and Ghattas, 2005b, Haber and Ascher, 2001]

Reduced Space Method

$$
\begin{aligned}
& \mathbf{g}^m = \mathbf{0} \quad \text{and} \quad \mathbf{g}^\lambda = \mathbf{0} \\
\Longrightarrow \quad & \tilde{\mathbf{m}} = -\mathbf{A}^{-1}\mathbf{C}\tilde{\mathbf{v}} \\
& \tilde{\boldsymbol{\lambda}} = -\mathbf{A}^{-\mathsf{T}}(\mathbf{H}_{mm}\tilde{\mathbf{m}} + \mathbf{H}_{mv}\tilde{\mathbf{v}})
\end{aligned}
$$

[Biros and Ghattas, 2005a, Biros and Ghattas, 2005b, Haber and Ascher, 2001]

Reduced Space Method

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \alpha_k \tilde{\mathbf{v}}_k$$

$$\tilde{\mathbf{v}}_k = -((\mathbf{H}_{\text{reg}} + \mathbf{H}_{\text{mis}})_k)^{-1} \mathbf{g}_k^v$$

$$\mathbf{H}_{\text{mis}} := \mathbf{C}^\mathsf{T} \mathbf{A}^{-\mathsf{T}} (\mathbf{H}_{mm} \mathbf{A}^{-1} \mathbf{C} - \mathbf{H}_{mv}) - \mathbf{H}_{vm} \mathbf{A}^{-1} \mathbf{C}$$

[Biros and Ghattas, 2005a, Biros and Ghattas, 2005b, Haber and Ascher, 2001]

Problem Formulation (Reminder)

$$
\begin{aligned}
\underset{\boldsymbol{v},m}{\text{minimize}} \quad & \frac{1}{2}\|m(1) - m_1\|^2_{L^2(\Omega)} + \frac{\beta}{2}\langle \mathcal{L}\boldsymbol{v}, \boldsymbol{v}\rangle_{L^2(\Omega)^d} \\
\text{subject to} \quad & \partial_t m + \langle \boldsymbol{v}, \nabla m\rangle = 0 \\
& m = m_0
\end{aligned}
$$

Reduced Gradient

$$\boldsymbol{g}^{v}(\boldsymbol{v}) := \beta\mathcal{L}\boldsymbol{v} + \mathcal{Q}\int_0^1 \lambda\,\nabla m\,\mathrm{d}t$$

$$\begin{aligned}
\partial_t m + \langle \boldsymbol{v}, \nabla m\rangle &= 0 && \text{in } \Omega \times (0,1] \\
m &= m_0 && \text{in } \Omega \times \{0\} \\
-\partial_t \lambda - \operatorname{div} \lambda \boldsymbol{v} &= 0 && \text{in } \Omega \times [0,1) \\
\lambda &= m_1 - m && \text{in } \Omega \times \{1\}
\end{aligned}$$

Newton–Krylov Method

$$\mathbf{H}_k^v \tilde{\mathbf{v}}_k = -\mathbf{g}_k^v, \quad \mathbf{v}_{k+1} = \mathbf{v}_k + \alpha_k \tilde{\mathbf{v}}_k$$

- globalized via Armijo line search
- (preconditioned) CG method
- matrix-free (only matvec required)
- inexactness (Eisenstat & Walker)

(Reduced) Hessian Matvec

$$\mathcal{H}^{v}[\tilde{\boldsymbol{v}}](\boldsymbol{v}) := \beta\mathcal{L}\tilde{\boldsymbol{v}} + \mathcal{Q}\int_0^1 \lambda\nabla\tilde{m} + \tilde{\lambda}\nabla m\,\mathrm{d}t$$

$$\begin{aligned}
\partial_t\tilde{m} + \langle\boldsymbol{v},\nabla\tilde{m}\rangle + \langle\tilde{\boldsymbol{v}},\nabla m\rangle &= 0 \quad && \text{in } \Omega\times(0,1] \\
\tilde{m} &= 0 \quad && \text{in } \Omega\times\{0\} \\
-\partial_t\tilde{\lambda} - \operatorname{div}(\tilde{\lambda}\boldsymbol{v} + \lambda\tilde{\boldsymbol{v}}) &= 0 \quad && \text{in } \Omega\times[0,1) \\
\tilde{\lambda} &= -\tilde{m} && \text{in } \Omega\times\{1\}
\end{aligned}$$

Computational Bottlenecks

- evaluating objective: 1 PDE solve
- evaluating gradient: 2 PDE solves
- Hessian matvec: 2 PDE solves

Computational Bottlenecks

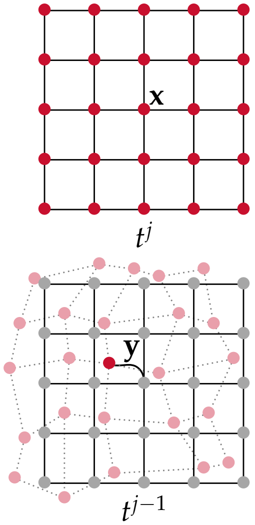
- efficient time integrator (fast PDE solves)
- effective preconditioner (few PDE solves)

PDE Solver

Time Integration

$$\partial_t u + \boldsymbol{v} \cdot \nabla u = f(u, \boldsymbol{v})$$

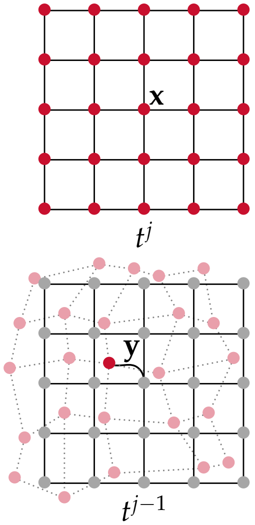
$$\begin{aligned} \mathrm{d}_t \mathbf{y} &= \mathbf{v}(\mathbf{y}) \quad \text{in} \quad [t^{j-1}, t^j) \\ \mathbf{y} &= \mathbf{x} \qquad \text{for} \quad t = t^j \end{aligned}$$



Time Integration

$$\partial_t u + \boldsymbol{v} \cdot \nabla u = f(u, \boldsymbol{v})$$

$$\mathrm{d}_t \mathbf{u}(\mathbf{y}) = \mathbf{f} \quad \text{in} \ \ (t^{j-1}, t^j]$$
$$\mathbf{u} = \mathbf{u}_0 \ \ \text{for} \ \ t = t^{j-1}$$



Preconditioner

Spectral Preconditioner

$$(\mathbf{H}_{\text{reg}} + \mathbf{H}_{\text{mis}})\tilde{\mathbf{v}} = -\mathbf{g}^{v}$$

$$(\mathbf{I} + \mathbf{H}_{\text{reg}}^{-1}\mathbf{H}_{\text{mis}})\tilde{\mathbf{v}} = -\mathbf{H}_{\text{reg}}^{-1}\mathbf{g}^{v}$$

2L Preconditioner

$$\mathbf{F}_H + \mathbf{F}_L = \mathbf{I}$$

$$\mathbf{H}\mathbf{e}_k = \mathbf{F}_H\mathbf{H}\mathbf{F}_H\mathbf{e}_k + \mathbf{F}_L\mathbf{H}\mathbf{F}_L\mathbf{e}_k$$

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}_L + \tilde{\mathbf{v}}_H$$

$$\mathbf{H}_L\tilde{\mathbf{v}}_L = (\mathbf{F}_L\mathbf{H}\mathbf{F}_L)\tilde{\mathbf{v}}_L = -\mathbf{F}_L\mathbf{g}$$

$$\mathbf{H}_H\tilde{\mathbf{v}}_H = (\mathbf{F}_H\mathbf{H}\mathbf{F}_H)\tilde{\mathbf{v}}_H = -\mathbf{F}_H\mathbf{g}$$

[Adavani and Biros, 2008, Biros and Doğan, 2008, Giraud et al., 2006, Kaltenbacher, 2003, Kaltenbacher, 2001, King, 1990]

2L Preconditioner

$$\tilde{\mathbf{H}}\mathbf{w} = -\mathbf{H}_{\mathrm{reg}}^{-1/2}\mathbf{g}$$

$$\mathbf{w} := \mathbf{H}_{\mathrm{reg}}^{1/2}\tilde{\mathbf{v}}, \qquad \tilde{\mathbf{H}} := \left(\mathbf{I} + \mathbf{H}_{\mathrm{reg}}^{-1/2}\mathbf{H}_{\mathrm{mis}}\mathbf{H}_{\mathrm{reg}}^{-1/2}\right)$$

2L Preconditioner

$$\mathbf{H}\mathbf{u} = \mathbf{s}, \quad \mathbf{u} = \mathbf{u}_L + \mathbf{u}_H \approx \mathbf{F}_L \mathbf{Q}_P \bar{\mathbf{u}}_L + \mathbf{F}_H \mathbf{s}$$

$$\bar{\mathbf{u}}_L \approx \tilde{\mathbf{H}}_C^{-1} \mathbf{Q}_R \mathbf{F}_L \mathbf{s}$$

2L Preconditioner

$$\tilde{\mathbf{H}}_c^G = \mathbf{Q}_R \tilde{\mathbf{H}} \mathbf{Q}_P$$

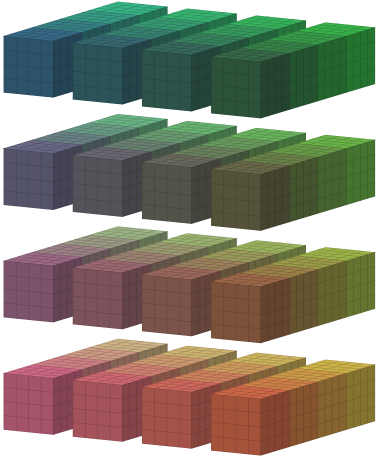
$$\tilde{\mathbf{H}}_c = \mathbf{I}_c + \mathbf{H}_{\mathrm{reg},c}^{-1/2} \mathbf{H}_{\mathrm{mis},c} \mathbf{H}_{\mathrm{reg},c}^{-1/2}$$

$$\mathbf{H}_{\mathrm{mis},c} = \mathbf{C}_c^\top \mathbf{A}_c^{-\top} (\mathbf{H}_{mm,c} \mathbf{A}_c^{-1} \mathbf{C}_c - \mathbf{H}_{mv,c}) - \mathbf{H}_{vm,c} \mathbf{A}_c^{-1} \mathbf{C}_c$$

Parallel Implementation

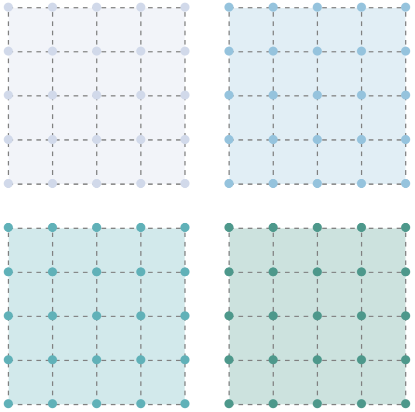
MPI Parallelism

- AccFFT
  http://accfft.org
- PETSc + TAO
  https://www.mcs.anl.gov/petsc/

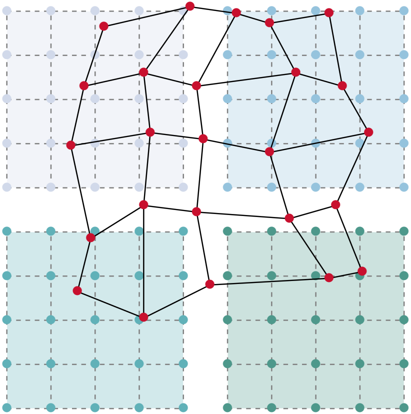


[Gholami et al., 2016, Munson et al., 2015, Balay et al., 2014]

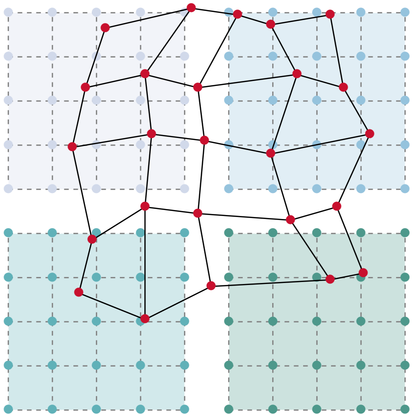
Parallel Semi-Lagrangian

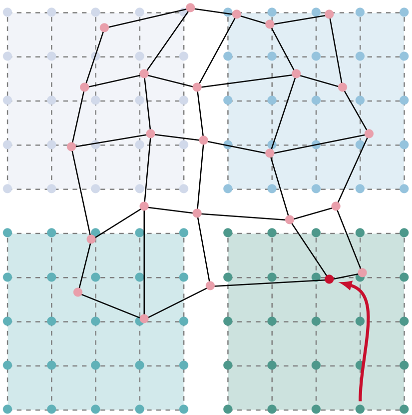


Parallel Semi-Lagrangian

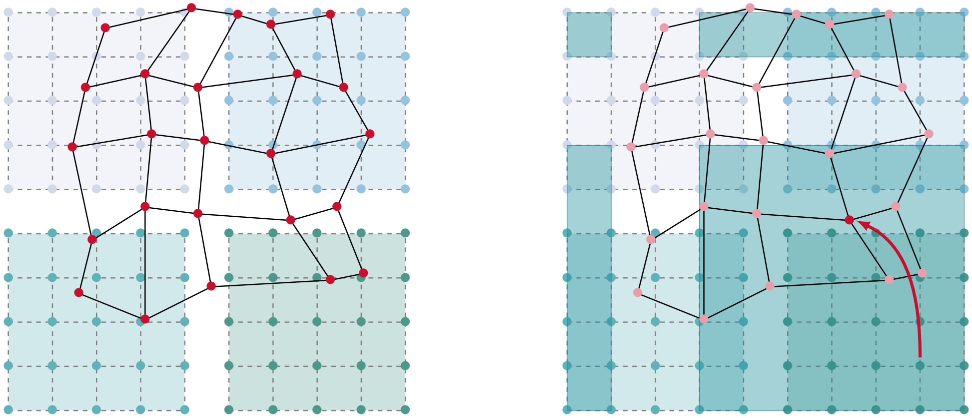


Parallel Semi-Lagrangian

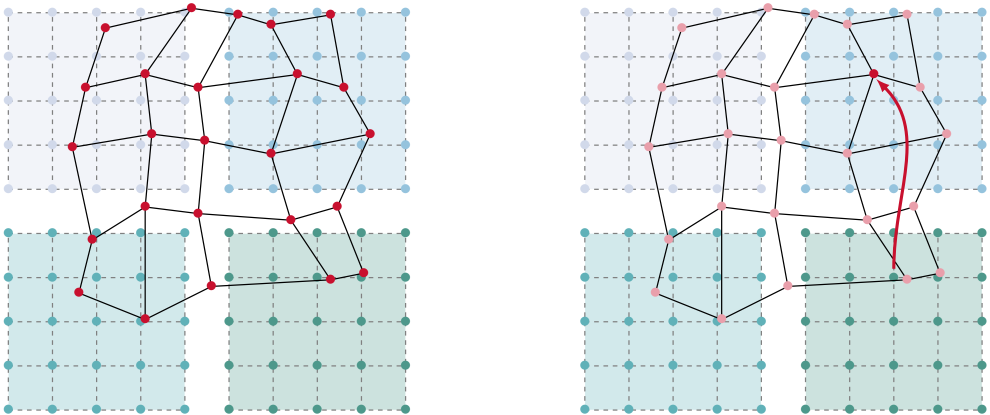




Parallel Semi-Lagrangian



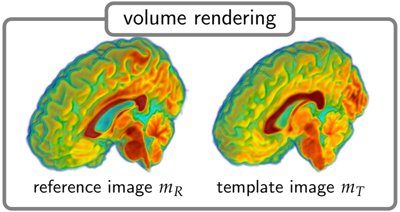
Parallel Semi-Lagrangian

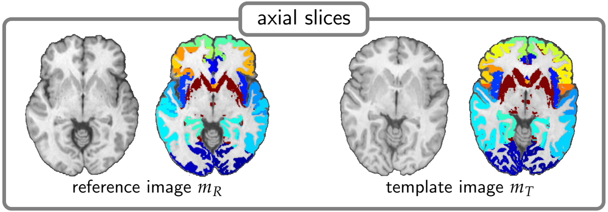


GPU Implementation

tag	variant
cpu-fft-cubic	FP32, CPU, FFT, cubic IP
gpu-fft-cubic	FP32, GPU, FFT, cubic IP
gpu-fd8-cubic	FP32, GPU, FD8, cubic IP
gpu-fd8-linear	FP32, GPU, FD8, trilinear IP

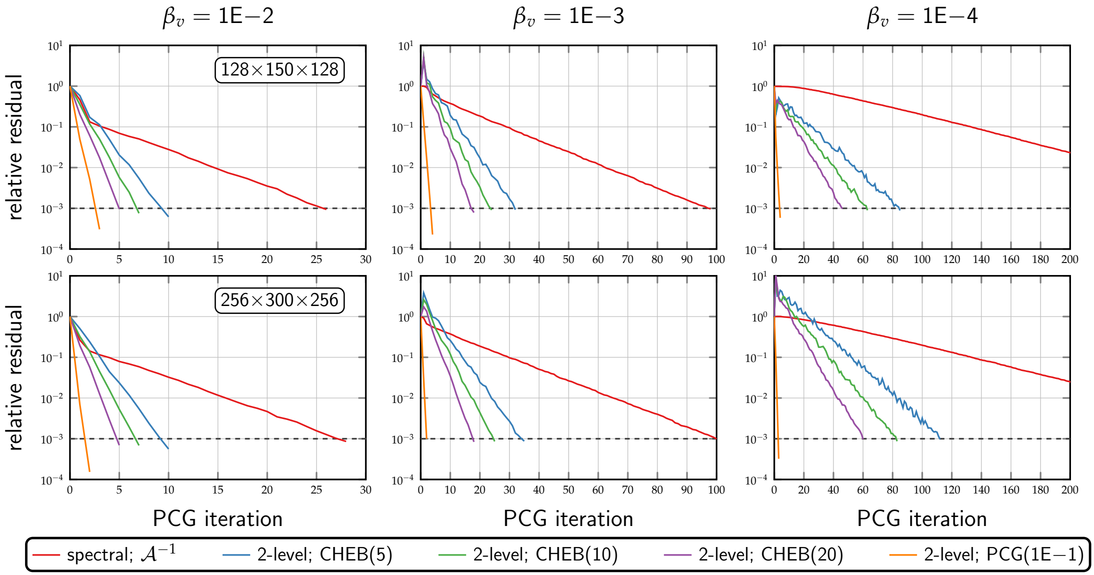
Results



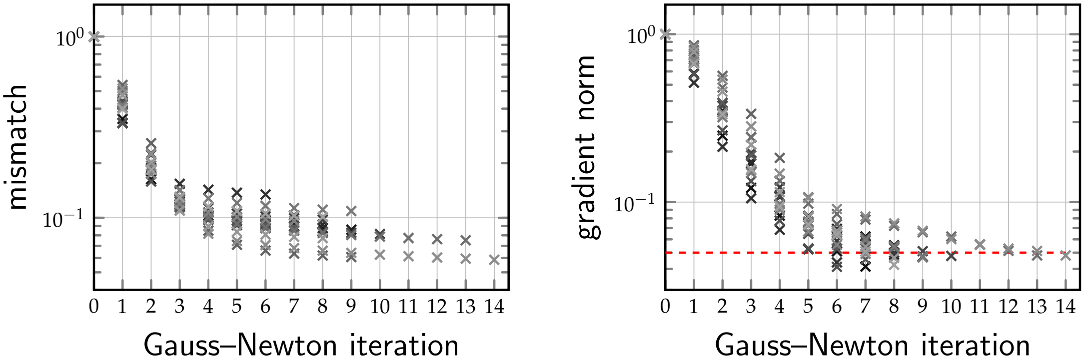


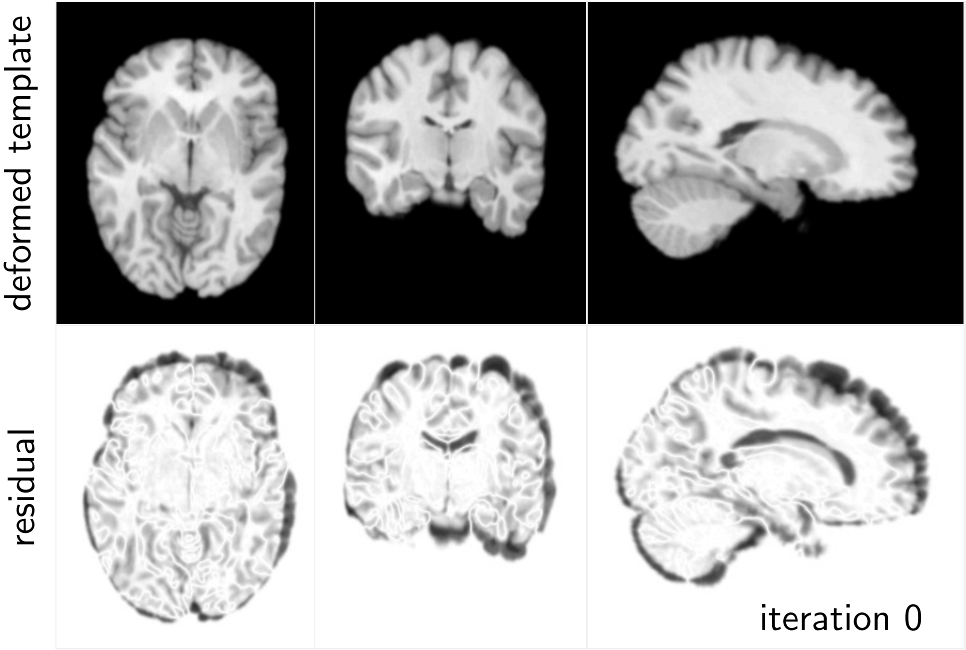
mean	max	min
5.2e−1	5.6e−1 (`na08`)	4.4e−1 (`na14`)

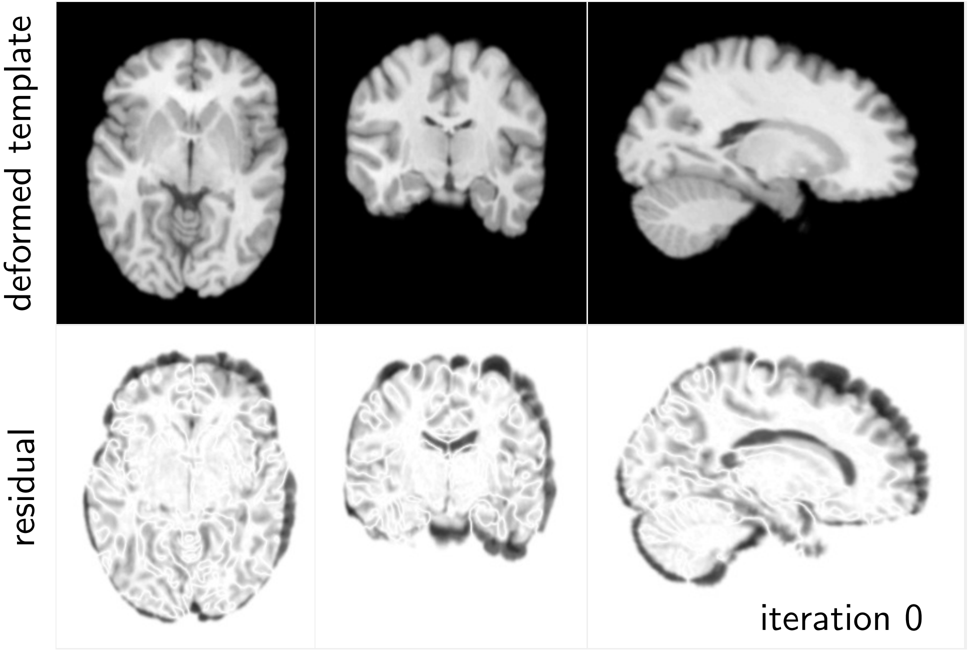
RCDC's Opuntia system (Intel ten-core Xeon E5-2680v2 at 2.8 GHz with 64 GB memory (2 sockets for a total of 20 cores))

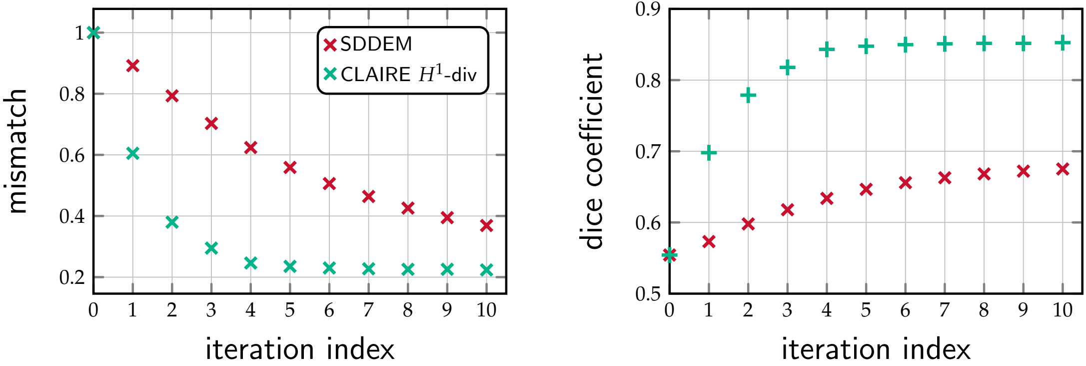


βv = 1E−2
βv = 1E−3
βv = 1E−4
128×150×128
256×300×256
relative residual
relative residual
PCG iteration
PCG iteration
PCG iteration
spectral; A−1
2-level; CHEB(5)
2-level; CHEB(10)
2-level; CHEB(20)
2-level; PCG(1E−1)

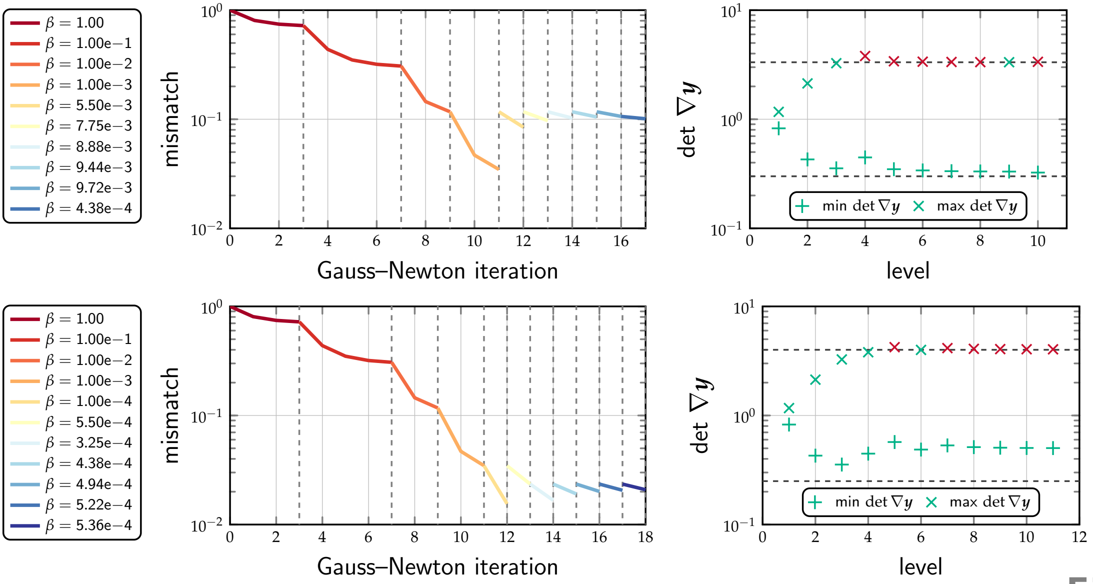

mismatch
10^0
10^{-1}
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Gauss–Newton iteration
gradient norm
10^0
10^{-1}
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Gauss–Newton iteration


deformed template
residual
iteration 0


deformed template
residual
iteration 0

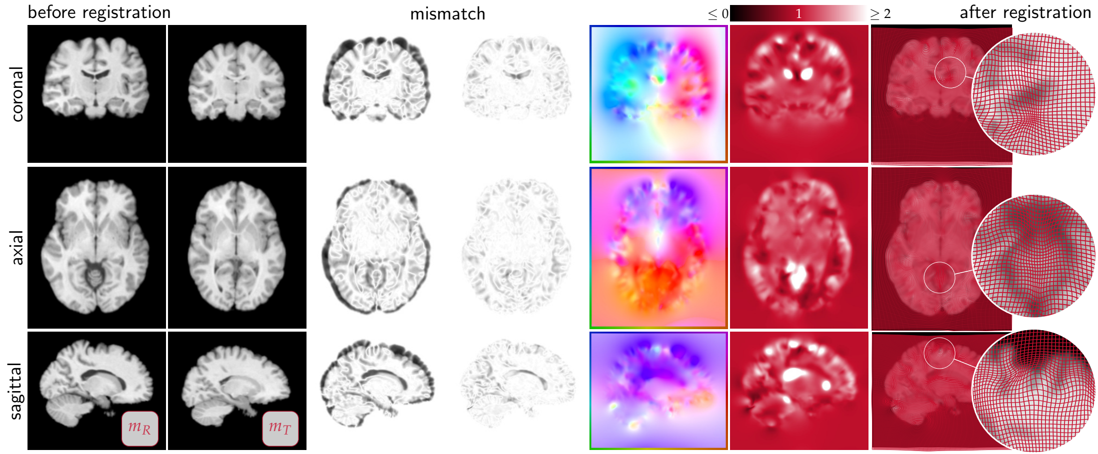


SDDEM
CLAIRE H^1-div
mismatch
iteration index
dice coefficient
iteration index



β = 1.00
β = 1.00e−1
β = 1.00e−2
β = 1.00e−3
β = 5.50e−3
β = 7.75e−3
β = 8.88e−3
β = 9.44e−3
β = 9.72e−3
β = 4.38e−4
mismatch
Gauss–Newton iteration
det ∇y
level
min det ∇y
max det ∇y
β = 1.00
β = 1.00e−1
β = 1.00e−2
β = 1.00e−3
β = 1.00e−4
β = 5.50e−4
β = 3.25e−4
β = 4.38e−4
β = 4.94e−4
β = 5.22e−4
β = 5.36e−4
mismatch
Gauss–Newton iteration
det ∇y
level
min det ∇y
max det ∇y

	dice		det $\nabla \boldsymbol{y}$		runtime
na02	5.5e−1	8.6e−1	4.7e−1	3.9	2.1e2
na03	5.0e−1	8.3e−1	4.8e−1	7.2	2.2e2
na04	5.2e−1	8.3e−1	3.4e−1	2.4e1	2.1e2
na05	5.6e−1	8.5e−1	4.2e−1	5.2	2.0e2
na06	5.6e−1	8.4e−1	5.2e−1	7.6	3.0e2
na07	5.3e−1	8.5e−1	2.9e−1	3.7	2.2e2
na08	5.6e−1	8.5e−1	3.3e−1	3.9	3.2e2
na09	5.1e−1	8.4e−1	5.3e−1	1.0e1	2.2e2
na10	4.8e−1	8.2e−1	6.0e−1	7.7	2.3e2
na11	4.6e−1	8.3e−1	3.4e−1	2.2e1	2.3e2
na12	5.2e−1	8.4e−1	5.1e−1	3.3e1	4.3e2
na13	5.3e−1	8.1e−1	3.3e−1	8.1	2.1e2
na14	4.4e−1	8.3e−1	3.3e−1	4.3	2.4e2
na15	5.0e−1	8.3e−1	3.3e−1	4.3	2.0e2
na16	5.5e−1	8.4e−1	3.7e−1	2.0e1	2.1e2
mean	5.2e−1	8.4e−1	4.1e−1	1.1e1	2.4e2



before registration
mismatch
≤ 0
1
≥ 2
after registration
coronal
axial
sagittal
m_R
m_T

β_v		#PDE	mismatch	runtime	speedup
1e−2	—	187	8.5e−2	6.0e2	
	PC	46	9.8e−2	9.3e1	6.5
	SC	67	8.8e−2	1.2e2	5.2
	GC	15,11,11	8.7e−2	3.5e1	17.1
1e−3	—	273	2.9e−2	9.0e2	
	PC	56	3.4e−2	1.6e2	5.6
	SC	83	2.8e−2	3.2e2	2.8
	GC	35,19,17	2.7e−2	1.4e2	6.3

Strong Scaling (Lonestar)

tasks	FFT	IP	sec	eff
2	48.0	43.4	2.4e2	100.0
8	48.0	44.5	6.7e1	87.6
32	51.8	41.3	1.8e1	81.4
128	58.6	36.5	4.6	79.5
512	53.1	42.2	1.5	60.5

Weak Scaling (Hazel Hen)

size	tasks	FFT	IP	sec	eff
1024^3	128	60.9	35.0	196.9	100.0
2048^3	1024	65.0	34.3	210.4	100.0
4096^3	8192	72.9	26.3	237.5	93.1

GPU Implementation (64^3)

dice		$\|g\|_{rel}$	#iter	#mv	sec
0.56	0.62	7.7e−3	12	58	1.82
	0.63	1.1e−2	12	54	0.23 (8)
0.50	0.61	8.0e−3	13	64	1.97
	0.61	1.6e−2	12	42	0.18 (11)
0.48	0.68	1.2e−2	12	48	1.61
	0.68	1.3e−2	12	44	0.18 (8)

CPU: dual socket Intel Skylake (Xeon Gold 5120); GPU: 32GB NVIDIA Tesla V100

GPU Implementation (128^3)

dice		$\|\|\mathbf{g}\|\|_{rel}$	#iter	#mv	sec
0.55	0.79	1.8e−2	14	70	13.36
	0.80	1.7e−2	12	63	0.75 (18)
0.51	0.79	1.8e−2	15	77	14.62
	0.79	1.7e−2	13	68	0.81 (18)
0.48	0.78	1.7e−2	15	84	15.93
	0.78	1.7e−2	15	82	0.96 (17)

CPU: dual socket Intel Skylake (Xeon Gold 5120); GPU: 32GB NVIDIA Tesla V100

GPU Implementation (256^3)

dice		$\|\mathbf{g}\|_{rel}$	#iter	#mv	sec
0.55	0.86	3.7e−2	14	81	146.69
	0.86	3.1e−2	14	75	5.87 (25)
0.50	0.83	3.6e−2	17	95	169.46
	0.83	3.1e−2	17	93	7.22 (24)
0.48	0.82	3.5e−2	18	103	184.78
	0.82	2.9e−2	17	94	7.29 (25)

CPU: dual socket Intel Skylake (Xeon Gold 5120); GPU: 32GB NVIDIA Tesla V100

Publications

Brunn, Himthania, Biros, Mehl & M (2019). *Fast GPU 3D diffeomorphic image registration*. Preprint (25 pages).

M, Gholami, Davatzikos, & Biros (2019). *CLAIRE: A parallel Newton–Krylov solver for constrained large deformation diffeomorphic image registration*, SIAM J Sci Comput (in press).

M, Gholami, Davatzikos & Biros (2018). *PDE-constrained optimization in medical image analysis*. Opt Eng, 19(3):765–812.

M & Biros (2017). *A semi-Lagrangian two-level preconditioned Newton–Krylov solver for constrained diffeomorphic image registration*. SIAM J Sci Comput, 39(6):B1064–B1101.

M & Ruthotto (2017). *A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration*. SIAM J Sci Comput, 39(5):B860–B885.

Gholami, M, Scheufele, Davatzikos, Mehl & Biros (2017). *A framework for scalable biophysics-based image analysis*. Proc ACM/IEEE Conf on Supercomputing.

M, Gholami & Biros (2016). *Distributed-memory large-deformation diffeomorphic 3D image registration*. Proc ACM/IEEE Conf on Supercomputing.

M & Biros (2016). *Constrained H^1-regularization schemes for diffeomorphic image registration*. SIAM J Imag Sci, 9(3):1154–1194.

M & Biros (2015). *An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration*. SIAM J Imag Sci, 8(2):1030–1069.

NVIDIA GPU Grant Program; Simons Foundation Award #586055; AFOSR grants FA9550-12-10484 and FA9550-11-10339; NSF grants DMS-1854853 and CCF-1337393; U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under DE-SC0010518 and DE-SC0009286; NIH grant 10042242; DARPA grant W911NF-115-2-0121; and TUM—Institute for Advanced Study, funded by the German Excellence Initiative (and the European Union Seventh Framework Programme under grant agreement 291763). Computing time on TACC systems was provided by an allocation from TACC and the NSF. Computing time on HLRS's Hazel Hen system was provided by an allocation of the federal project application ACID-44104.

‣ restart

References

Adavani, S. S. and Biros, G. (2008).
Multigrid algorithms for inverse problems with linear parabolic PDE constraints.
SIAM Journal on Scientific Computing, 31(1):369–397.

Amit, Y. (1994).
A nonlinear variational problem for image matching.
SIAM Journal on Scientific Computing, 15(1):207–224.

Arguilière, S., Trélat, E., Trouvé, A., and Younes, L. (2016).
Multiple shape registration using constrained optimal control.
SIAM J Imaging Sci.

Azencott, R., Glowinski, R., He, J., Jajoo, A., Lie, Y. P., Martynenko, A., Hoppe, R. H. W., Benzekry, S., and Little, S. H. (2010).
Diffeomorphic matching and dynamic deformable surfaces in 3D medical imaging.
Computational Methods in Applied Mathematics, 10(3):235–274.

Balay, S., Abhyankar, S., Adams, M. F., Brown, J., Brune, P., Buschelman, K., Eijkhout, V., Gropp, W. D., Kaushik, D., Knepley, M. G., McInnes, L. C., Rupp, K., Smith, B. F., and Zhang, H. (2014).
PETSc users manual.
Technical Report ANL-95/11 - Revision 3.5, Argonne National Laboratory.

Barbu, V. and Marinoschi, G. (2016).
An optimal control approach to the optical flow problem.
Systems & Control Letters, 87:1–9.

Beg, M. F., Miller, M. I., Trouve, A., and Younes, L. (2005).
Computing large deformation metric mappings via geodesic flows of diffeomorphisms.
International Journal of Computer Vision, 61(2):139–157.

Biegler, L. T., Ghattas, O., Heinkenschloss, M., and van Bloemen Waanders, B. (2003).
Large-scale PDE-constrained optimization.
Springer.

Biros, G. and Doğan, G. (2008).
A multilevel algorithm for inverse problems with elliptic PDE constraints.
Inverse Problems, 24(1–18).

Biros, G. and Ghattas, O. (2005a).
Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization—Part I: The Krylov-Schur solver.
SIAM Journal on Scientific Computing, 27(2):687–713.

Biros, G. and Ghattas, O. (2005b).
Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization—Part II: The Lagrange-Newton solver and its application to optimal control of steady viscous flows.
SIAM Journal on Scientific Computing, 27(2):714–739.

Borzi, A., Ito, K., and Kunisch, K. (2002).
An optimal control approach to optical flow computation.
International Journal for Numerical Methods in Fluids, 40(1–2):231–240.

Borzi, A. and Schulz, V. (2012).
Computational optimization of systems governed by partial differential equations.
SIAM, Philadelphia, Pennsylvania, US.

Chen, K. and Lorenz, D. A. (2012).
Image sequence interpolation based on optical flow, segmentation and optimal control.
Image Processing, IEEE Transactions on, 21(3):1020–1030.

Dupuis, P., Gernander, U., and Miller, M. I. (1998).
Variational problems on flows of diffeomorphisms for image matching.
Quarterly of Applied Mathematics, 56(3):587–600.

Fischer, B. and Modersitzki, J. (2008).
Ill-posed medicine – an introduction to image registration.
Inverse Problems, 24(3):1–16.

Gholami, A., Hill, J., Malhotra, D., and Biros, G. (2016).
AccFFT: A library for distributed-memory FFT on CPU and GPU architectures.
arXiv e-prints.
`https://arxiv.org/abs/1506.07933`.

Gholami, A., Mang, A., Scheufele, K., Davatzikos, C., Mehl, M., and Biros, G. (2017).
A framework for scalable biophysics-based image analysis.
In *Proc ACM/IEEE Conference on Supercomputing*, number 19, pages 19:1–19:13.
`https://doi.org/10.1145/3126908.3126930`.

Giraud, L., Ruiz, D., and Touhami, A. (2006).
A comparitive study of iterative solvers exploiting spectral information for SPD systems.
SIAM Journal on Scientific Computing, 27(5):1760–1786.

Haber, E. and Ascher, U. M. (2001).
Preconditioned all-at-once methods for large, sparse parameter estimation problems.
Inverse Problems, 17(6):1847–1864.

Haber, E. and Modersitzki, J. (2006).
Intensity gradient based registration and fusion of multi-modal images.
In *Proc Medical Image Computing and Computer-Assisted Intervention*, volume 4191, pages 726–733.

Hart, G. L., Zach, C., and Niethammer, M. (2009).
An optimal control approach for deformable registration.
In *Proc IEEE Conference on Computer Vision and Pattern Recognition*, pages 9–16.

Herzog, R., Pearson, J. W., and Stoll, M. (2019).
Fast iterative solvers for an optimal transport problem.
Advances in Computational Mathematics, 45:495–517.
`https://arxiv.org/abs/1801.04172`.

Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. (2009).
Optimization with PDE constraints.
Springer, Berlin, DE.

Jarde, P. P. and Ulbrich, M. (2019).
Existence of minimizers for optical flow based optimal control problems under mild regularity assumptions.
Preprint.

Kaltenbacher, B. (2001).
On the regularizing properties of a full multigrid method for ill-posed problems.
Inverse Problems, 17(4):767–788.

Kaltenbacher, B. (2003).
V-cycle convergence of some multigrid methods for ill-posed problems.
Mathematics of Computation, 72(244):1711–1730.

King, J. T. (1990).
On the construction of preconditioners by subspace decomposition.
Journal of Computational and Applied Mathematics, 29:195–205.

Lions, J. L. (1971).
Optimal control of systems governed by partial differential equations.
Springer.

Mang, A., Gholami, A., and Biros, G. (2016).
Distributed-memory large-deformation diffeomorphic 3D image registration.
In *Proc ACM/IEEE Conference on Supercomputing*, number 72.
https://doi.org/10.1109/SC.2016.71.

Mang, A., Gholami, A., Davatzikos, C., and Biros, G. (2019).
CLAIRE: A distributed-memory solver for constrained large deformation diffeomorphic image registration.
arXiv e-prints.
https://arxiv.org/abs/1808.04487.

Modersitzki, J. (2004).
Numerical methods for image registration.
Oxford University Press, New York.

Modersitzki, J. (2009).
FAIR: Flexible algorithms for image registration.
SIAM, Philadelphia, Pennsylvania, US.

Munson, T., Sarich, J., Wild, S., Benson, S., and McInnes, L. C. (2015).
TAO 3.6 users manual.
Argonne National Laboratory, Mathematics and Computer Science Division.

Sotiras, A., Davatzikos, C., and Paragios, N. (2013).
Deformable medical image registration: A survey.
Medical Imaging, IEEE Transactions on, 32(7):1153–1190.

Trouve, A. (1998).
Diffeomorphism groups and pattern matching in image analysis.
International Journal of Computer Vision, 28(3):213–221.

Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. (2012).
Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation.
International Journal of Computer Vision, 97:229–241.

Younes, L. (2010).
Shapes and diffeomorphisms.
Springer.