A Scalable Interior-Point Gauss-Newton Method for PDE-Constrained Optimization with Bound Constraints

MFEM Community Workshop 2025 – Portland, Oregon

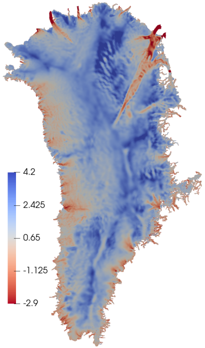
Tucker Hartland, Cosmin G. Petra, Noémi Petra, and Jingyi Wang

LLNL-PRES-863065

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Motivation – PDE-constrained optimization with bound constraints

- **Example:** inversion (from surface flow observations) for the *non-negative* ice sheet basal friction field.
- **Challenge:** inequality constraints lead to computational challenges via complementarity in the KKT conditions.
- **Goal:** develop a scalable computational framework for such PDE- and bound-constrained optimization problems.



Picture from: T. Hartland, G. Stadler, K. Liegeois, M. Perego, and N. Petra. "Hierarchical off-diagonal low-rank approximation of Hessians in inverse problems, with application to ice sheet model initialization", Inverse Problems, 2023.

PDE- and bound-constrained optimization problem structure

General problem statement

$$\min_{(u,\rho)} \mathcal{J}(u,\rho) := \mathcal{J}_{\text{misfit}}(u) + \mathcal{J}_{\text{reg}}(\rho)$$

such that

$$\rho(t,\boldsymbol{x}) \geq \rho_\ell(t,\boldsymbol{x}), \text{ on } [0,T]\times\overline{\Omega} \text{ and}$$
$$r(u,\rho) = 0, \text{ in } (0,T)\times\Omega$$

Notation:

1. $u = u(t,\boldsymbol{x}), \rho = \rho(t,\boldsymbol{x})$, the state and parameter (respectively).
2. Ω – spatial domain, $[0,T]$ – time domain.
3. ρ_ℓ – lower-bound constraint, r – PDE residual.

PDE- and bound-constrained optimization common approach

General problem **common "reduced-space" approach**

$$\min_{\beta} \mathcal{J}(\beta) := \mathcal{J}(u(\rho), \rho), \quad \rho = \exp(\beta) + \rho_\ell$$

where $u(\rho)$ defined implicitly by $r(u(\rho), \rho) = 0$

Pros and cons

PDE- and bound-constrained optimization common approach

General problem **common "reduced-space" approach**

$$\min_{\beta} \mathcal{J}(\beta) := \mathcal{J}(u(\rho), \rho), \quad \rho = \exp(\beta) + \rho_\ell$$

where $u(\rho)$ defined implicitly by $r(u(\rho), \rho) = 0$

Pros and cons

+ The optimization problem is a reduced unconstrained problem and so simple optimization methods (Newton, Gauss-Newton, etc) are applicable.

PDE- and bound-constrained optimization common approach

General problem **common "reduced-space" approach**

$$\min_{\beta} \mathcal{J}(\beta) := \mathcal{J}(u(\rho), \rho), \quad \rho = \exp(\beta) + \rho_\ell$$

where $u(\rho)$ defined implicitly by $r(u(\rho), \rho) = 0$

Pros and cons

\+ The optimization problem is a reduced unconstrained problem and so simple optimization methods (Newton, Gauss-Newton, etc) are applicable.

\- Each objective evaluation requires a PDE-solve.

PDE- and bound-constrained optimization common approach

General problem **common "reduced-space" approach**

$$\min_{\beta} \mathcal{J}(\beta) := \mathcal{J}(u(\rho), \rho), \quad \rho = \exp(\beta) + \rho_\ell$$

where $u(\rho)$ defined implicitly by $r(u(\rho), \rho) = 0$

Pros and cons

\+ The optimization problem is a reduced unconstrained problem and so simple optimization methods (Newton, Gauss-Newton, etc) are applicable.

\- Each objective evaluation requires a PDE-solve.

\- The reparametrization $\rho = \exp(\beta) + \rho_\ell$ can introduce higher-order nonlinearities and does not make use of constrained numerical optimization research.

PDE- and bound-constrained optimization example problem I

Nonlinear elliptic example problem

$$\min_{(u,\rho)} \mathcal{J}(u,\rho) := \underbrace{\frac{1}{2}\int_{\Omega_{\text{obs}}} (u(\boldsymbol{x}) - u_d(\boldsymbol{x}))^2 \,\mathrm{d}\boldsymbol{x}}_{\text{data-misfit}} + \underbrace{\frac{1}{2}\int_{\Omega} (\gamma_1 \rho(\boldsymbol{x})^2 + \gamma_2 \nabla\rho \cdot \nabla\rho)\mathrm{d}\boldsymbol{x}}_{\text{regularization}}$$

such that

$$\rho(\boldsymbol{x}) \geq \rho_\ell(\boldsymbol{x}), \text{ on } \overline{\Omega} = [0,1]^2 \text{ and}$$

$$\begin{cases} -\nabla \cdot (\rho \nabla u) + u + u^3/3 & = g \quad \text{in } \Omega \\ \rho \nabla u \cdot \boldsymbol{n} & = 0 \quad \text{on } \partial\Omega \end{cases}$$

PDE- and bound-constrained optimization example problem II

Linear parabolic example problem

$$\min_{(u,\rho)} \mathcal{J}(u,\rho) := \underbrace{\frac{1}{2}\int_{\Omega_{\text{obs}}} \left(u(T,\boldsymbol{x}) - u_d(\boldsymbol{x})\right)^2 \mathrm{d}\boldsymbol{x}}_{\text{data-misfit}} + \underbrace{\frac{1}{2}\int_{\Omega} (\gamma_1 \rho(\boldsymbol{x})^2 + \gamma_2 \nabla\rho \cdot \nabla\rho) \mathrm{d}\boldsymbol{x}}_{\text{regularization}}$$

such that

$\rho(\boldsymbol{x}) \geq \rho_\ell(\boldsymbol{x})$, on $\overline{\Omega} = [0,1]^2$ (periodic) and

$$\begin{cases} \partial u/\partial t - \nabla \cdot (\kappa \nabla u) & = g(\boldsymbol{x}) \quad \text{in } (0,T) \times \overline{\Omega} \\ u(t,\boldsymbol{x})|_{t=0} & = \rho(\boldsymbol{x}) \quad \text{in } \overline{\Omega} \end{cases}$$

Infinite-dimensional interior-point (IP) optimality system

1. Cast the PDE into weak form (assumed here time independent)

$$\text{Find } u \in H^1(\Omega) \text{ such that } \quad c(u, \rho, \lambda) = 0, \ \ \forall \lambda \in H^1(\Omega).$$

2. Define the log-barrier penalized Lagrangian function with PDE-constraint Lagrange multiplier λ and log-barrier penalty parameter $\mu > 0$

$$\mathcal{L}(u, \rho, \lambda) := \underbrace{\mathcal{J}(u, \rho)}_{\text{Objective}} \underbrace{- \mu \int_\Omega \log\left(\rho - \rho_\ell\right) \mathrm{d}\boldsymbol{x}}_{\text{log-barrier term}} + \underbrace{c(u, \rho, \lambda)}_{\text{PDE constraint}} \ .$$

3. Solve the interior-point regularized nonlinear optimality system for $\mu \to 0^+$

$$\mathcal{L}_u \tilde{u} = 0, \ \ \forall \tilde{u} \in H^1(\Omega), \qquad \{\text{stationarity}\}$$

$$\mathcal{L}_\rho \tilde{\rho} = 0, \ \ \forall \tilde{\rho} \in H^1(\Omega) \cap L^\infty(\Omega), \qquad \{\text{stationarity}\}$$

$$\mathcal{L}_\lambda \tilde{\lambda} = 0, \ \ \forall \tilde{\lambda} \in H^1(\Omega), \qquad \{\text{feasibility}\}$$

“Outer” IP-Gauss-Newton method

1. Construct the $\mu > 0$, nonlinear continuation system

$$\mathcal{L}_u \tilde{u} = 0, \;\; \forall \tilde{u} \in H^1(\Omega), \qquad \{\text{stationarity}\}$$

$$\mathcal{L}_\rho \tilde{\rho} = 0, \;\; \forall \tilde{\rho} \in H^1(\Omega) \cap L^\infty(\Omega), \qquad \{\text{stationarity}\}$$

$$\mathcal{L}_\lambda \tilde{\lambda} = 0, \;\; \forall \tilde{\lambda} \in H^1(\Omega), \qquad \{\text{feasibility}\}$$

2. Use the Gauss-Newton method, with stopping criteria defined by mass-weighted norms, to inexactly solve log-barrier subproblems ($\mu \searrow 0$), with a filter line-search IPM for robust convergence

More details can be found in:

A. Wächter and L.T. Biegler. *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Mathematical Programming, 2006.

T. Hartland, C.G. Petra, N. Petra, J. Wang. *A scalable interior-point Gauss-Newton method for PDE-constrained optimization with bound constraints*, arXiv, 2024 (in review).

IP-Gauss-Newton linear solve is a critical computational step

– The IP-Gauss-Newton linear system

$$\underbrace{\begin{pmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & \boldsymbol{0} & \boldsymbol{J_u}^\top \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & \boldsymbol{J_\rho}^\top \\ \boldsymbol{J_u} & \boldsymbol{J_\rho} & \boldsymbol{0} \end{pmatrix}}_{\boldsymbol{A}} \begin{pmatrix} \color{blue}{\hat{\boldsymbol{u}}} \\ \color{red}{\hat{\boldsymbol{\rho}}} \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b_u} \\ \boldsymbol{b_\rho} \\ \boldsymbol{b_\lambda} \end{pmatrix}$$

that must be solved for the search direction $(\color{blue}{\hat{\boldsymbol{u}}}, \color{red}{\hat{\boldsymbol{\rho}}}, \hat{\boldsymbol{\lambda}})$ at each "outer" optimization step.

IP-Gauss-Newton linear solve is a critical computational step

- The IP-Gauss-Newton linear system

$$\underbrace{\begin{pmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & \boldsymbol{0} & {\boldsymbol{J}_{\boldsymbol{u}}}^{\top} \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & {\boldsymbol{J}_{\boldsymbol{\rho}}}^{\top} \\ \boldsymbol{J}_{\boldsymbol{u}} & \boldsymbol{J}_{\boldsymbol{\rho}} & \boldsymbol{0} \end{pmatrix}}_{\boldsymbol{A}} \begin{pmatrix} \hat{\boldsymbol{u}} \\ \hat{\boldsymbol{\rho}} \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}_{\boldsymbol{u}} \\ \boldsymbol{b}_{\boldsymbol{\rho}} \\ \boldsymbol{b}_{\boldsymbol{\lambda}} \end{pmatrix}$$

  that must be solved for the search direction $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\lambda}})$ at each "outer" optimization step.

- $\boldsymbol{J}_{\boldsymbol{u}}$, discretized linearized forward PDE, $\boldsymbol{R} = \gamma_1 \boldsymbol{M} + \gamma_2 \boldsymbol{K}$, regularization

IP-Gauss-Newton linear solve is a critical computational step

- The IP-Gauss-Newton linear system

$$\underbrace{\begin{pmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & \boldsymbol{0} & \boldsymbol{J_u}^\top \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & \boldsymbol{J_\rho}^\top \\ \boldsymbol{J_u} & \boldsymbol{J_\rho} & \boldsymbol{0} \end{pmatrix}}_{\boldsymbol{A}} \begin{pmatrix} \hat{\boldsymbol{u}} \\ \hat{\boldsymbol{\rho}} \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b_u} \\ \boldsymbol{b_\rho} \\ \boldsymbol{b_\lambda} \end{pmatrix}$$

  that must be solved for the search direction $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\lambda}})$ at each "outer" optimization step.

- $\boldsymbol{J_u}$, discretized linearized forward PDE, $\boldsymbol{R} = \gamma_1 \boldsymbol{M} + \gamma_2 \boldsymbol{K}$, regularization
- $\boldsymbol{H}_{\text{log-barrier}} = \boldsymbol{M}_{\text{lumped}} \boldsymbol{D}$, the log-barrier Hessian, a **positive** definite **diagonal** matrix that generally is ***very* ill-conditioned** $(\mu \searrow 0)$.

IP-Gauss-Newton linear system preconditioner

- **Goal:** efficient iterative solution of the IP-Gauss-Newton linear system $\boldsymbol{Ax} = \boldsymbol{b}$
- **Strategy 1:** GMRES with the block Gauss-Seidel preconditioner,

$$\tilde{\boldsymbol{A}} := \begin{pmatrix} \boldsymbol{H_{u,u}} & \boldsymbol{0} & \boldsymbol{J_u}^\top \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & \boldsymbol{J_\rho}^\top \\ \boldsymbol{J_u} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

J. Pestana, T. Rees. *Null-space preconditioners for saddle point systems*, SIAM Journal on Matrix Analysis and Applications, 2016

G. Biros, O. Ghattas. *Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. Part I: The Krylov-Schur solver*, SIAM Journal on Scientific Computing, 2005

- **Strategy 2:** Log-barrier and regularization, $\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}}$ preconditioned CG for the equivalent "reduced-space" Schur complement system

The Gauss-Seidel preconditioned IP-Gauss-Newton matrix

The preconditioned matrix $\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}$, is similar to $\begin{bmatrix} \boldsymbol{I} & \begin{bmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & {\boldsymbol{J}_{\boldsymbol{u}}}^\top \\ \boldsymbol{J}_{\boldsymbol{u}} & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} & {\boldsymbol{J}_{\rho}}^\top \end{bmatrix} \\ \boldsymbol{0} & \boldsymbol{I} + (\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}})^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}} \end{bmatrix}$.

$\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ is the positive semi-definite "reduced-space" data-misfit Gauss-Newton Hessian

$$\hat{\boldsymbol{H}}_{\boldsymbol{d}} = ({\boldsymbol{J}_{\boldsymbol{u}}}^{-1}\boldsymbol{J}_{\rho})^\top \boldsymbol{H}_{u,u}({\boldsymbol{J}_{\boldsymbol{u}}}^{-1}\boldsymbol{J}_{\rho})$$

Particular to typical ill-posed PDE-constrained optimization problems:

- eigenvalues of $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ decay rapidly to zero and in a mesh independent fashion

More details can be found in:

O. Ghattas, K. Wilcox. *Learning physics-based models from data: perspectives from inverse problems and model reduction*, Acta Numerica (2021).

Block Gauss-Seidel preconditioner effectively clusters spectrum

Bounds on the eigenvalues of preconditioned matrix

$$1 \leq \lambda_j(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}) \leq \begin{cases} 1 + \lambda_j(\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}), & 1 \leq j \leq \dim(\boldsymbol{\rho}), \\ 1, & \dim(\boldsymbol{\rho}) + 1 \leq j \leq \dim(\boldsymbol{\rho}) + 2 \cdot \dim(\boldsymbol{u}), \end{cases}$$

- eigenvalues of $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ decay rapidly to zero and in a mesh independent fashion
- $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ does not contain components specific to IPM

Block Gauss-Seidel preconditioner effectively clusters spectrum

Bounds on the eigenvalues of preconditioned matrix

$$1 \leq \lambda_j(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}) \leq \begin{cases} 1 + \lambda_j(\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}), & 1 \leq j \leq \dim(\boldsymbol{\rho}), \\ 1, & \dim(\boldsymbol{\rho}) + 1 \leq j \leq \dim(\boldsymbol{\rho}) + 2 \cdot \dim(\boldsymbol{u}), \end{cases}$$

- eigenvalues of $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ decay rapidly to zero and in a mesh independent fashion
- $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ does not contain components specific to IPM

Conclusions to be drawn from the eigenvalue bounds

- eigenvalues of $\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}$ decay rapidly to one and in a mesh independent fashion

Block Gauss-Seidel preconditioner effectively clusters spectrum

Bounds on the eigenvalues of preconditioned matrix

$$1 \leq \lambda_j(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}) \leq \begin{cases} 1 + \lambda_j(\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}), & 1 \leq j \leq \dim(\boldsymbol{\rho}), \\ 1, & \dim(\boldsymbol{\rho}) + 1 \leq j \leq \dim(\boldsymbol{\rho}) + 2 \cdot \dim(\boldsymbol{u}), \end{cases}$$

- eigenvalues of $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ decay rapidly to zero and in a mesh independent fashion
- $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ does not contain components specific to IPM

Conclusions to be drawn from the eigenvalue bounds

- eigenvalues of $\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}$ decay rapidly to one and in a mesh independent fashion
- spectrum of $\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}$ is largely independent of IPM ill-conditioning

Block Gauss-Seidel preconditioner effectively clusters spectrum

Bounds on the eigenvalues of preconditioned matrix

$$1 \leq \lambda_j(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}) \leq \begin{cases} 1 + \lambda_j(\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}), & 1 \leq j \leq \dim(\boldsymbol{\rho}), \\ 1, & \dim(\boldsymbol{\rho}) + 1 \leq j \leq \dim(\boldsymbol{\rho}) + 2 \cdot \dim(\boldsymbol{u}), \end{cases}$$

- eigenvalues of $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ decay rapidly to zero and in a mesh independent fashion
- $\boldsymbol{R}^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$ does not contain components specific to IPM

Ill-conditioning due to IPM is factored out by the preconditioner.

Details in: T. Hartland, C.G. Petra, N. Petra, J. Wang. *A scalable interior-point Gauss-Newton method for PDE-constrained optimization with bound constraints*, arXiv, 2024 (in review).

The cost to apply the block Gauss-Seidel preconditioner

To compute

$$\boldsymbol{x} = \tilde{\boldsymbol{A}}^{-1}\boldsymbol{b} = \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & \boldsymbol{0} & {\boldsymbol{J}_{\boldsymbol{u}}}^{\top} \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & {\boldsymbol{J}_{\rho}}^{\top} \\ \boldsymbol{J}_{\boldsymbol{u}} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_{\boldsymbol{u}} \\ \boldsymbol{b}_{\rho} \\ \boldsymbol{b}_{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{\boldsymbol{u}} \\ \boldsymbol{x}_{\rho} \\ \boldsymbol{x}_{\lambda} \end{pmatrix}$$

1. $\boldsymbol{x}_{\boldsymbol{u}} = {\boldsymbol{J}_{\boldsymbol{u}}}^{-1}\boldsymbol{b}_{\lambda}$ {linearized forward PDE solve}
2. $\boldsymbol{x}_{\lambda} = {\boldsymbol{J}_{\boldsymbol{u}}}^{-\top}(\boldsymbol{b}_{\boldsymbol{u}} - \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}}\boldsymbol{x}_{\boldsymbol{u}})$ {adjoint PDE solve}
3. $\boldsymbol{x}_{\rho} = (\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}})^{-1}\left(\boldsymbol{b}_{\rho} - {\boldsymbol{J}_{\rho}}^{\top}\boldsymbol{x}_{\lambda}\right)$ {AMG-CG solve}

The cost to apply the block Gauss-Seidel preconditioner

To compute

$$\boldsymbol{x} = \tilde{\boldsymbol{A}}^{-1}\boldsymbol{b} = \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}} & \boldsymbol{0} & \boldsymbol{J}_{\boldsymbol{u}}^{\top} \\ \boldsymbol{0} & \boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} & \boldsymbol{J}_{\boldsymbol{\rho}}^{\top} \\ \boldsymbol{J}_{\boldsymbol{u}} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_{\boldsymbol{u}} \\ \boldsymbol{b}_{\boldsymbol{\rho}} \\ \boldsymbol{b}_{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{\boldsymbol{u}} \\ \boldsymbol{x}_{\boldsymbol{\rho}} \\ \boldsymbol{x}_{\boldsymbol{\lambda}} \end{pmatrix}$$

1. $\boldsymbol{x}_{\boldsymbol{u}} = \boldsymbol{J}_{\boldsymbol{u}}^{-1}\boldsymbol{b}_{\boldsymbol{\lambda}}$ {linearized forward PDE solve}
2. $\boldsymbol{x}_{\boldsymbol{\lambda}} = \boldsymbol{J}_{\boldsymbol{u}}^{-\top}\left(\boldsymbol{b}_{\boldsymbol{u}} - \boldsymbol{H}_{\boldsymbol{u},\boldsymbol{u}}\boldsymbol{x}_{\boldsymbol{u}}\right)$ {adjoint PDE solve}
3. $\boldsymbol{x}_{\boldsymbol{\rho}} = \left(\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}}\right)^{-1}\left(\boldsymbol{b}_{\boldsymbol{\rho}} - \boldsymbol{J}_{\boldsymbol{\rho}}^{\top}\boldsymbol{x}_{\boldsymbol{\lambda}}\right)$ {AMG-CG solve}

   The log-barrier Hessian improves the **diagonal dominance** of $\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}}$

$$\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} = \gamma_1 \boldsymbol{M} + \gamma_2 \boldsymbol{K} + \underbrace{\boldsymbol{M}_{\text{lumped}}\boldsymbol{D}}_{\text{log-barrier Hessian}}$$

Sketch of Preconditioned CG approach

Idea: form Schur complement system by eliminating $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{\lambda}}$

$$(\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} + \hat{\boldsymbol{H}}_{\boldsymbol{d}})\hat{\rho} = \boldsymbol{b}$$

Use preconditioner: $\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}}$. The eigenvalues

$$(\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}})^{-1}\left(\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}} + \hat{\boldsymbol{H}}_{\boldsymbol{d}}\right) = \boldsymbol{I} + (\boldsymbol{R} + \boldsymbol{H}_{\text{log-barrier}})^{-1}\hat{\boldsymbol{H}}_{\boldsymbol{d}}$$

of this preconditioned system are intimately related to the eigenvalues of $\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}$.

MFEM-based implementation of the IP-Gauss-Newton framework

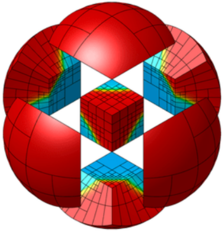
Features utilized

- Modular object-oriented design;
- Distributed memory parallelism (MPI);
- Finite element discretization, mesh refinement, Krylov subspace solvers (MFEM);
- Algebraic multigrid preconditioners (hypre);

More details can be found in:

J. Andrej, et al. *High performance finite elements with MFEM*, The International Journal of High Performance Computing Applications, 2024.

hypre. *High Performance Preconditioners*, https://llnl.gov/casc/hypre



MFEM is a *free*, *lightweight*, *scalable* C++ library for finite element methods.

Nonlinear elliptic PDE- and bound-constrained example problem

$$\min_{(u,\rho)} \mathcal{J}(u,\rho) := \underbrace{\frac{1}{2}\int_{\Omega_{\text{obs}}} (u(\boldsymbol{x}) - u_d(\boldsymbol{x}))^2 \,\mathrm{d}\boldsymbol{x}}_{\text{data-misfit}} + \underbrace{\frac{1}{2}\int_{\Omega} (\gamma_1 \rho(\boldsymbol{x})^2 + \gamma_2 \nabla\rho \cdot \nabla\rho)\mathrm{d}\boldsymbol{x}}_{\text{regularization}}$$

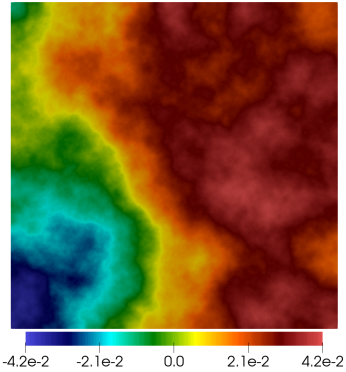
such that

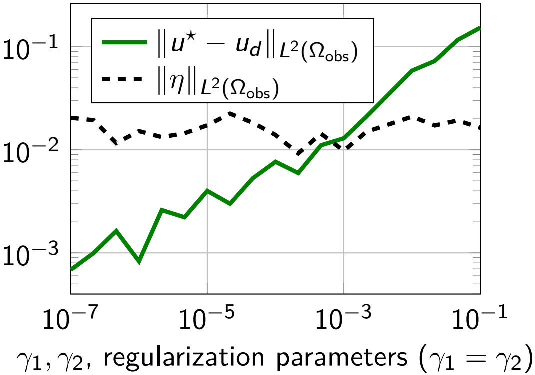
$$\rho(\boldsymbol{x}) \geq \rho_\ell(\boldsymbol{x}) = 1.0, \text{ on } \overline{\Omega} = [0,1]^2 \text{ and}$$

$$\begin{cases} -\nabla \cdot (\rho \nabla u) + u + u^3/3 & = g \quad \text{in } \Omega \\ \rho \nabla u \cdot \boldsymbol{n} & = 0 \quad \text{on } \partial\Omega \end{cases}$$

- $\Omega_{\text{obs}} = (0, 1/2) \times (0, 1)$, observation domain
- $u_d(\boldsymbol{x}) = \cos(\pi\, \boldsymbol{x}_1)\cos(\pi\, \boldsymbol{x}_2) + \eta(\boldsymbol{x})$, noisy ($\eta$) data

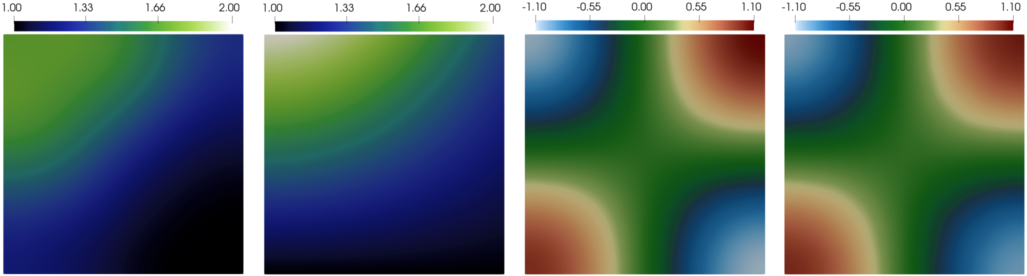
Regularization determined by the Morozov discrepancy principle





Left: spatial structure of a random noise sample η (5% relative noise level).
Right: Seminorms of the discrepancy $(u^\star - u_d)$ and noise η as functions of the regularization.

Nonlinear elliptic PDE- and bound-constrained problem solution



(a) $\rho^\star$ (b) ρ_{true} (c) $u^\star$ (d) u_d

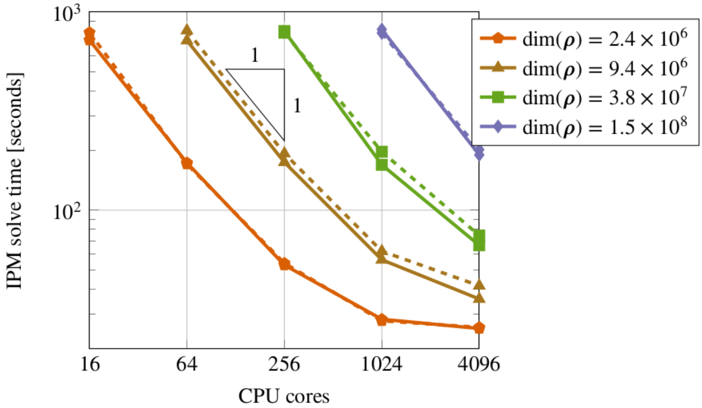
Computed optimum $\rho^\star$ (left), ρ_{true} (middle left), computed optimum state $u^\star$ (middle right) and noisy state data u_d (right).

Solution computed with a mesh independent number of iterations

$\dim(\rho)$	Average interior-point per optimizer solve	Average GMRES iterations per linear solve	Average CG iterations per linear solve
148 609	28.4	6.50	6.76
591 361	28.2	6.48	6.72
2 362 369	28.8	6.51	6.68
9 443 329	28.3	6.49	6.85
37 761 025	28.7	6.42	6.87
151 019 521	29.0	6.52	6.75

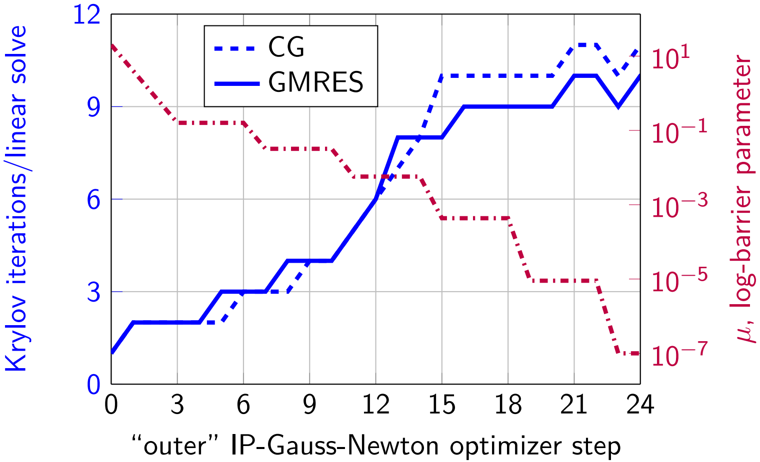
Figure: Outer interior-point optimizer and inner Krylov iteration counts. 10^{-6} interior-point optimizer absolute tolerance and 10^{-8} relative linear solve tolerance.

Strong scaling (quartz) of the IP-Gauss-Newton-Krylov method



GMRES: solid lines, CG: dashed lines. Timings obtained on Intel Xeon E5-2695 v4 chips.

Preconditioners are robust with respect to IPM ill-conditioning



Parabolic PDE- and bound-constrained example problem

Linear parabolic example problem

$$\min_{(u,\rho)} \mathcal{J}(u,\rho) := \underbrace{\frac{1}{2}\int_{\Omega_{\text{obs}}} \left(u(T,\boldsymbol{x}) - u_d(\boldsymbol{x})\right)^2 \mathrm{d}\boldsymbol{x}}_{\text{data-misfit}} + \underbrace{\frac{1}{2}\int_{\Omega} (\gamma_1 \rho(\boldsymbol{x})^2 + \gamma_2 \nabla\rho \cdot \nabla\rho) \mathrm{d}\boldsymbol{x}}_{\text{regularization}}$$

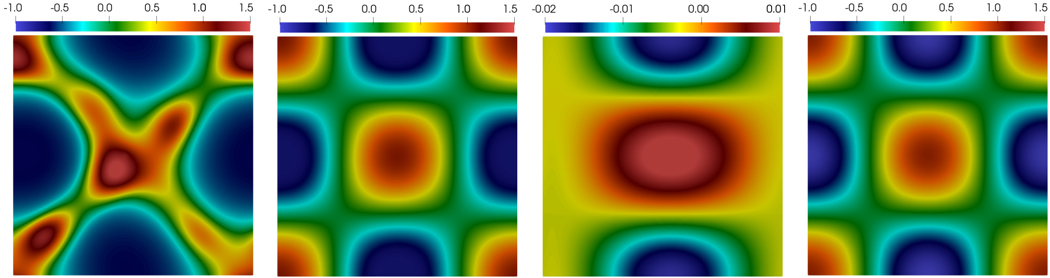
such that

$\rho(\boldsymbol{x}) \geq \rho_\ell(\boldsymbol{x}) = -3/4$, on $\Omega = (0,1)^2$ (periodic) and

$$\begin{cases} \partial u/\partial t - \nabla \cdot (\kappa \nabla u) & = g(\boldsymbol{x}) \quad \text{in } (0,T)\times\Omega \\ u(t,\boldsymbol{x})|_{t=0} & = \rho(\boldsymbol{x}) \quad \text{on } \Omega \end{cases}$$

- $\rho_{\text{true}}(\boldsymbol{x}) = \cos(\pi\, x_1)\cos(\pi\, x_2)$,

Parabolic PDE- and bound-constrained problem solution



(a) $\rho^\star$ ($\gamma_2 = 10^{-9}$) (b) $\rho^\star$ ($\gamma_2 = 10^{-5}$) (c) $\rho^\star$ ($\gamma_2 = 10^{-1}$) (d) ρ_{true}

Parameter reconstruction $\rho^\star$ with various regularization parameters (left, middle left and middle right), and true parameter ρ_{true} (right).

Solution computed with a mesh independent number of iterations for a wide range of regularization parameter values

$\dim(\rho)$	# iter GMRES, (# iter IP)				
	$\gamma_2 = 10^{-10}$	$\gamma_2 = 10^{-8}$	$\gamma_2 = 10^{-6}$	$\gamma_2 = 10^{-4}$	$\gamma_2 = 10^{-2}$
9.2×10^3	16.8 (14)	16.7 (14)	15.4 (13)	11.8 (11)	7.1 (7)
3.7×10^4	18.7 (15)	17.2 (14)	15.4 (13)	12.1 (11)	7.4 (7)
1.5×10^5	19.5 (15)	18.9 (14)	15.5 (13)	11.9 (11)	7.3 (7)
5.9×10^5	20.2 (15)	20.5 (15)	15.4 (13)	12.1 (11)	7.1 (7)
2.4×10^6	20.7 (14)	20.1 (14)	15.8 (13)	11.9 (11)	7.4 (7)

Table: Algorithmic scaling of the IP-Gauss-Newton method with block Gauss-Seidel preconditioned GMRES solves for the parabolic time-dependent PDE- and bound-constrained optimization with backward Euler time step $\Delta t = 0.01$. The absolute tolerance of the outer optimization loop is 10^{-6} and the relative tolerance of the block AMG-CG solves is 10^{-13}.

Conclusions and future work

- Algorithmically scalable IP-Gauss-Newton method for PDE- and bound-constrained optimization problems that respects the nature of the infinite-dimensional problem.

- Ill-conditioning of IP-Gauss-Newton linear systems handled by preconditioners that exploit PDE-constrained optimization problem structure.

- Interested in applying this framework to a broader set of problems and discussions/feedback from the MFEM community.

- The IPM solver (without PDE-constrained examples) is available at https://github.com/LLNL/ContinuationSolvers.

- Details in: T. Hartland, C.G. Petra, N. Petra, J. Wang. *A scalable interior-point Gauss-Newton method for PDE-constrained optimization with bound constraints*, arXiv, 2024 (in review).

Thank you for your attention.

Presenter (**Tucker Hartland**) contact info – hartland1@llnl.gov

This material is based upon work supported by the LDRD Project 23-ERD-017, the Exascale Computing Project (17-SC-20-SC) and by the NSF under Grant No. DMS-1840265 and CAREER-1654311.

Lawrence Livermore National Laboratory