A Matrix-free Algorithm for Reduced-space PDE-governed Optimization

Pengfei Meng

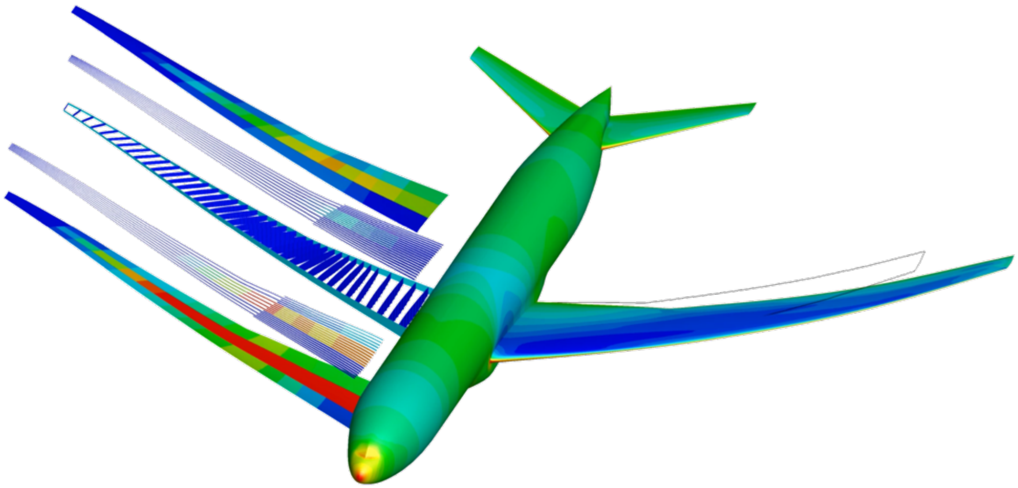
Outline

1. Introduction
2. Homotopy-Based Globalization
3. Iterative Solver & Preconditioner
4. Tests and Applications
5. Summary and Recommendations

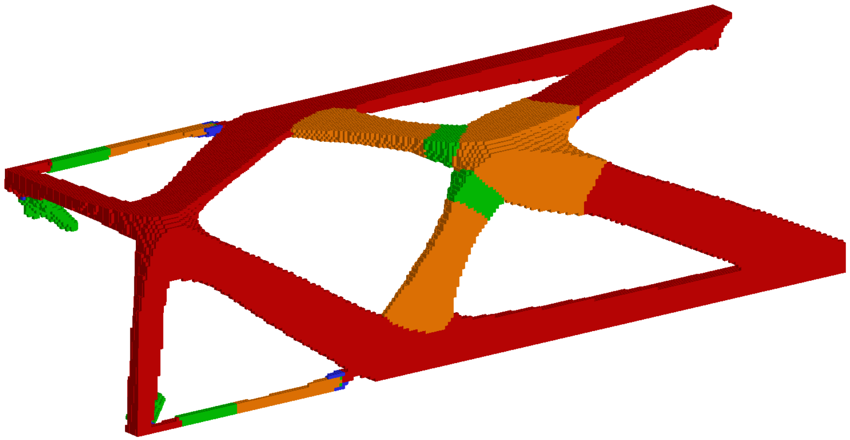
Improved engineering design is important to address climate change

Improved engineering design is important to address climate change

Optimization problems with state-based constraints are common, yet challenging to solve

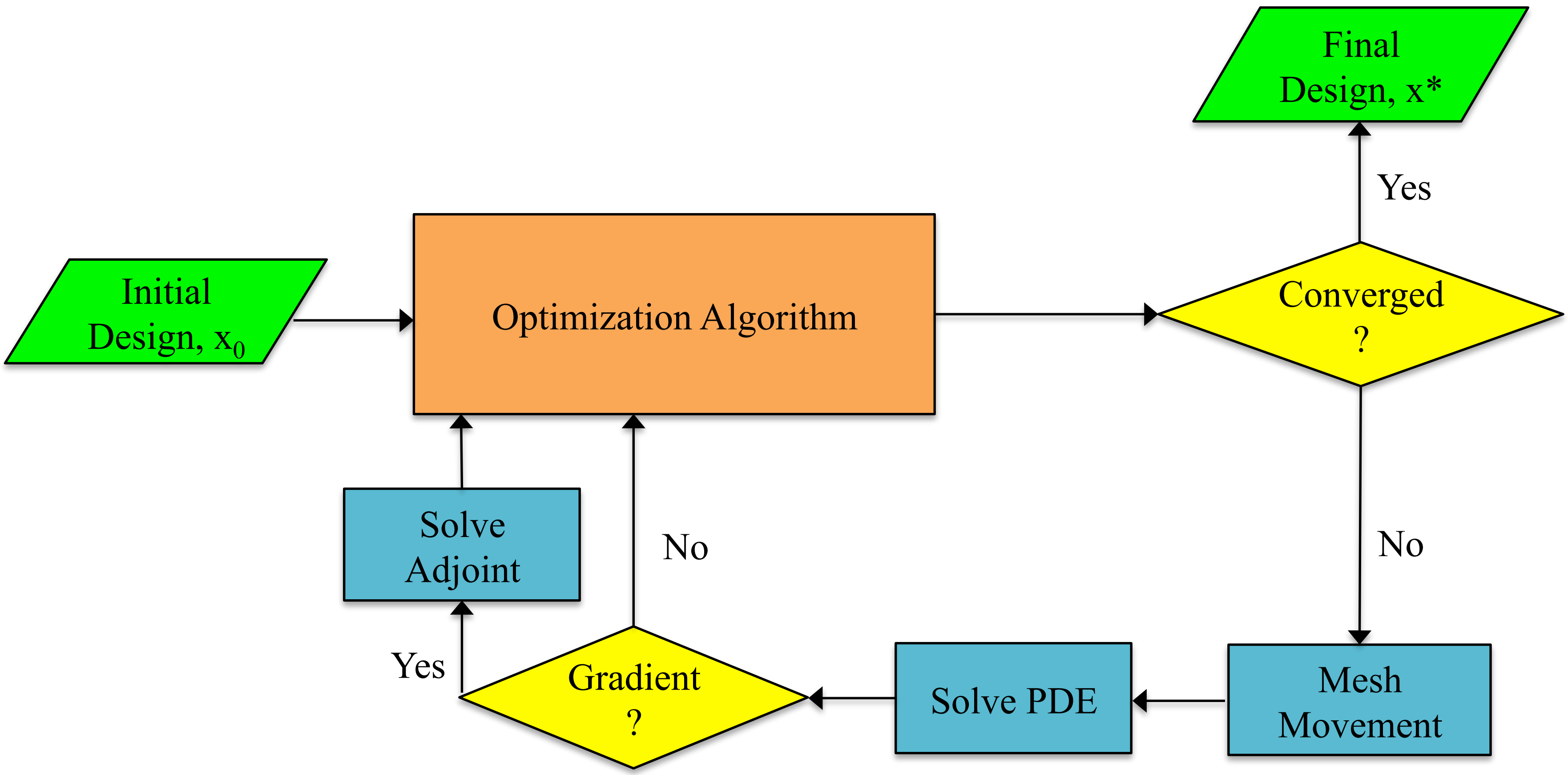


Aero-Structural Optimization
Kenway and Martins [1]



Topology Optimization
Kennedy [2]

Schematic Process of PDE-constrained Optimization



Conventional Optimization Algorithms and Their Limitations

Conventional optimization algorithms have been used on PDE-governed optimization problems

- When there are relatively few state-based constraints
- SNOPT, IPOPT, Knitro

Conventional optimization algorithms are unsuitable as they require the explicit total constraint Jacobian

- The gradient of each constraint requires the solution of a linear PDE (the adjoint)
- Cost is intractable when there are hundreds of constraints
- Storing the (dense) constraint Jacobian can also be expensive

Full-space and Reduced-space approaches are commonly used to solve PDE-constrained optimization problems

The generic formulation:

$$
\begin{array}{rll}
\min\limits_{x,u} & f(x,u) & \\
\text{subject to} & h(x,u) & = 0 \\
 & g(x,u) & \geq 0 \\
\text{governed by} & \mathcal{R}(x,u) & = 0
\end{array}
$$

The Lagrangian:

$$\mathcal{L}(x,u,\psi,s,\lambda_h,\lambda_g) = f(x,u) + \lambda_h^T h(x,u) + \lambda_g^T (g(x,u) - s) + \psi^T \mathcal{R}(x,u)$$

Optimal designs are defined by zeros of the first-order optimality conditions (KKT condition)

Adjoint Equation:

$$\partial_x \mathcal{L} = \partial_x f + \lambda_h^T \partial_x h + \lambda_g^T \partial_x g + \psi^T \partial_x \mathcal{R} = 0,$$

$$\partial_u \mathcal{L} = \partial_u f + \lambda_h^T \partial_u h + \lambda_g^T \partial_u g + \psi^T \partial_u \mathcal{R} = 0,$$

State Equation:

$$\partial_\psi \mathcal{L} = \mathcal{R} = 0,$$

$$\partial_{\lambda_h} \mathcal{L} = h = 0,$$

$$\partial_{\lambda_g} \mathcal{L} = g - s = 0,$$

Complementarity:

$$\begin{cases} -\mathsf{S}\Lambda_g e = 0, \\ s \geq 0, \quad \lambda_g \leq 0. \end{cases}$$

A coupled, nonlinear system of equations;
ill-conditioned;
huge size;

Full-space approach solve all the variables simultaneously, however,

- The KKT-system is large (at least 2 times the number of states), indefinite, ill-conditioned
- Globalization algorithms for nonlinear PDEs are difficult to incorporate
- If the optimization does not converge, the PDE is not even satisfied and cannot be used to inform decision making

Reduced-Space Inexact Newton Methods:

- Treat the state and adjoint vectors as implicit functions of the design variables
- Makes use of existing simulation solvers and adjoint solvers
- Much smaller optimization system than the full-space method
- Successfully implemented in unconstrained, equality-constrained problems

The reduced KKT conditions are as follows:

$$F(x, s, \lambda_h, \lambda_g) \equiv \begin{bmatrix} \nabla_x f + \lambda_h^T \nabla_x h + \lambda_g^T \nabla_x g \\ -\mathsf{S}\Lambda_g e \\ h \\ g - s \end{bmatrix} = 0,$$

$$\text{subject to} \quad s_i \geq 0, \quad \text{and} \quad \lambda_{gi} \leq 0 \qquad \forall i = 1, 2, \ldots, m,$$

Inexact-Newton methods solve the following linear system using Krylov methods at each iteration step:

$$(\nabla_q F) \Delta q^{(k)} = -F(q^{(k)})$$

$$\left\| (\nabla_q F) \Delta q^{(k)} + F(q^{(k)}) \right\| \leq \eta_k \left\| F(q^{(k)}) \right\|$$

Extending Reduced-Space Inexact-Newton methods to general nonlinear constraints encounters the following challenges

- Globalization: Avoid stationary points that are not local minimizers, i.e. handle nonconvexity.
- Preconditioning: Accelerate the Convergence of Krylov Iterative Methods

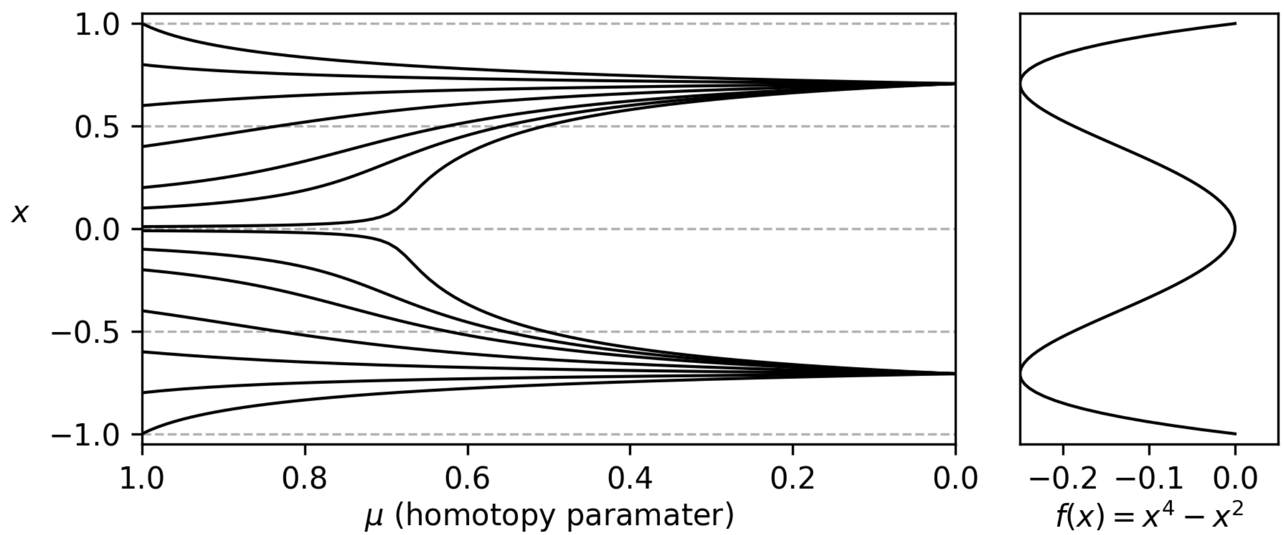
Outline

1. Introduction
2. Homotopy-Based Globalization
3. Iterative Solver & Preconditioner
4. Tests and Applications
5. Summary and Recommendations

A homotopy map defines a solution path from an easy solution to the desired solution

$$\min_x \quad f(x) = x^4 - x^2$$

$$H(x, \mu) = (1-\mu)\nabla f(x) + \mu(x - x_0) = 0, \quad x_0 \in [-1, 1.0]$$



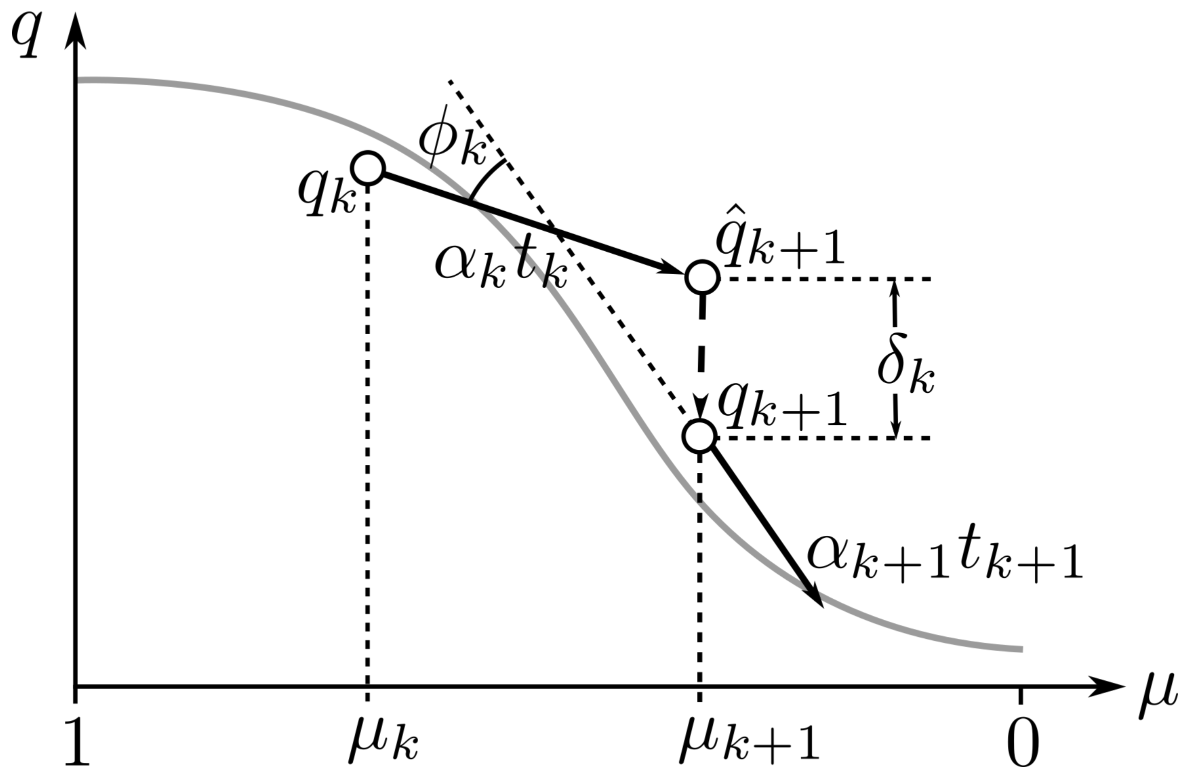
The Homotopy map for constrained optimization

$$
\begin{aligned}
H(q, q_0, \mu) &= (1-\mu)F(q) + \mu G(q) \\
&= (1-\mu)\begin{bmatrix} \nabla_x \mathcal{L} \\ -\mathsf{S}\Lambda_g e \\ h(x) \\ g(x) - s \end{bmatrix} + \mu \begin{bmatrix} x - x_0 \\ s - s_0 \\ -\lambda_h \\ -\lambda_g \end{bmatrix}
\end{aligned}
$$

The Jacobian of the Homotopy function

$$
\begin{aligned}
\nabla_q H &= (1-\mu)\nabla_q F(q) + \mu \nabla_q G(q, q_0) \\
&= (1-\mu)\begin{bmatrix} \nabla_{xx}\mathcal{L} & \mathbf{0} & \nabla_x h^T & \nabla_x g^T \\ \mathbf{0} & -\Lambda_g & \mathbf{0} & -\mathsf{S} \\ \nabla_x h & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \nabla_x g & -\mathsf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \mu \begin{bmatrix} \mathsf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathsf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathsf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathsf{I} \end{bmatrix}
\end{aligned}
$$

A predictor-corrector algorithm has been implemented to trace the homotopy zero-curve

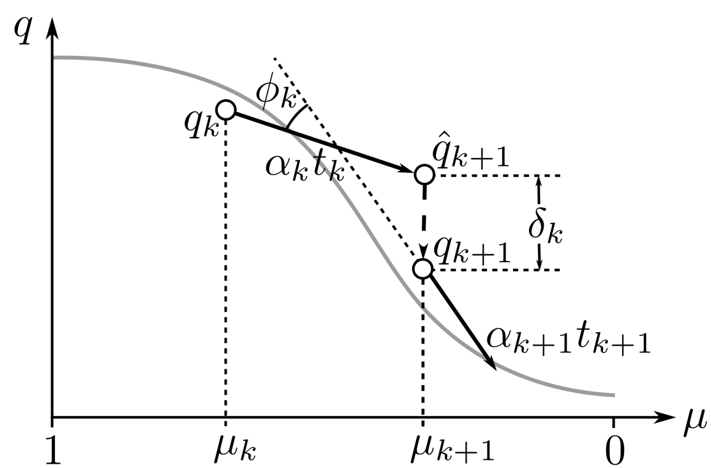


Adaptive step-size control on the predictor step:

Initial parameters: $\alpha_0, \delta_{\text{targ}}, \phi_{\text{targ}}$

Step-size factor: $\zeta_k \equiv \max\left(\sqrt{\delta_k/\delta_{\text{targ}}}, \phi_k/\phi_{\text{targ}}\right)$

Step-size: $\alpha_k = \min\left(\sqrt{\|q_k'\|^2 + 1}\Delta\mu_{\max}, \alpha_{\max}, \alpha_{k-1}/\zeta_k\right)$

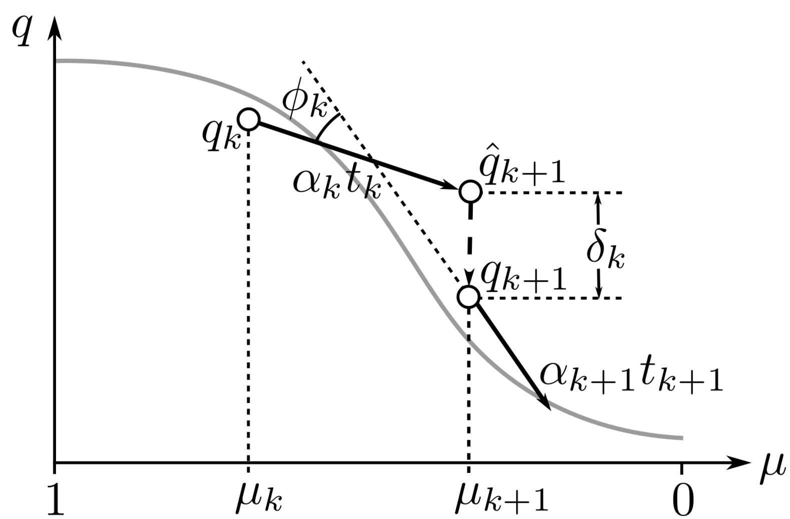


Safeguards on the slacks and inequality multipliers

Slack Variable $\alpha_{\max} = \max\{\alpha \in (0,1] \mid s + \alpha s' \geq \tau_s\}$

$$s \leftarrow \max(s, \tau_s)$$

Inequality Multipliers $\lambda_g \leftarrow \min(\lambda_g, 0)$



Numerical Experiments 1: Sphere Problem

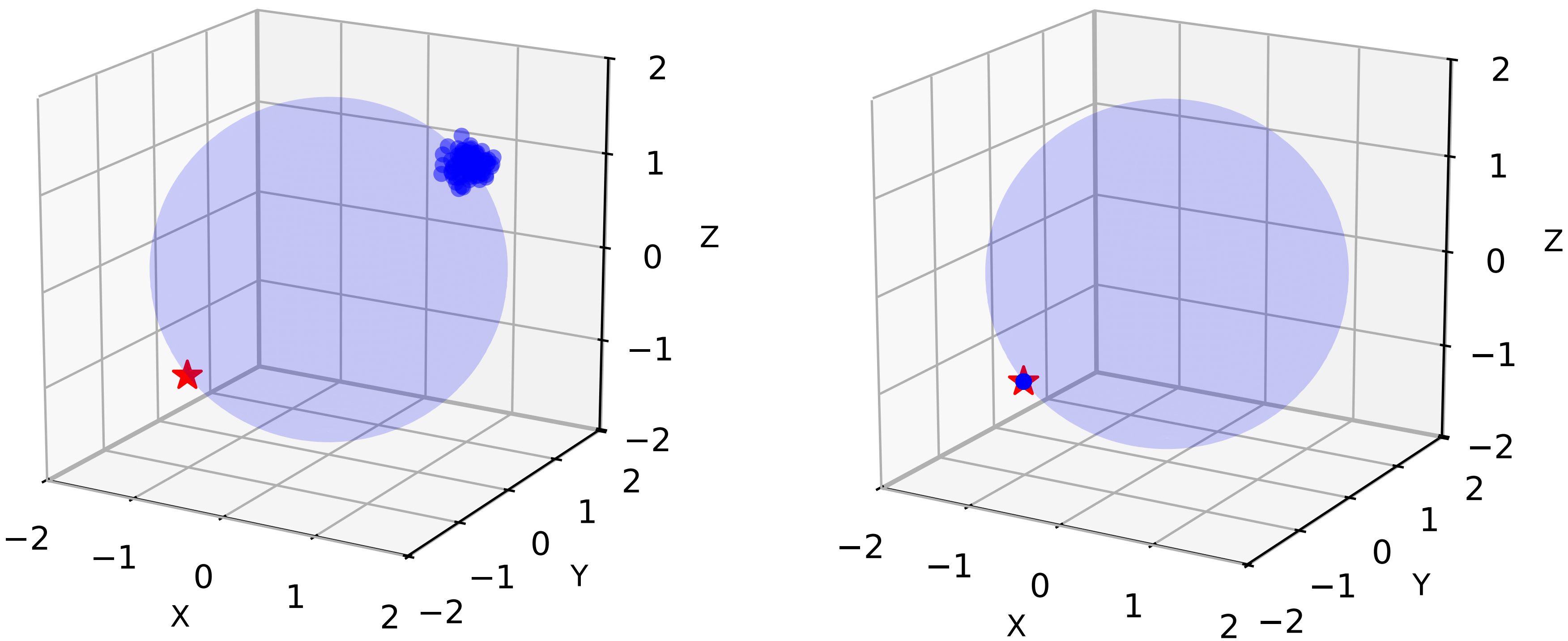
Problem Formulation:

$$\begin{aligned} \min_{x,y,z} \quad & x + y + z \\ \text{subject to} \quad & x^2 + y^2 + z^2 \leq 3 \end{aligned}$$

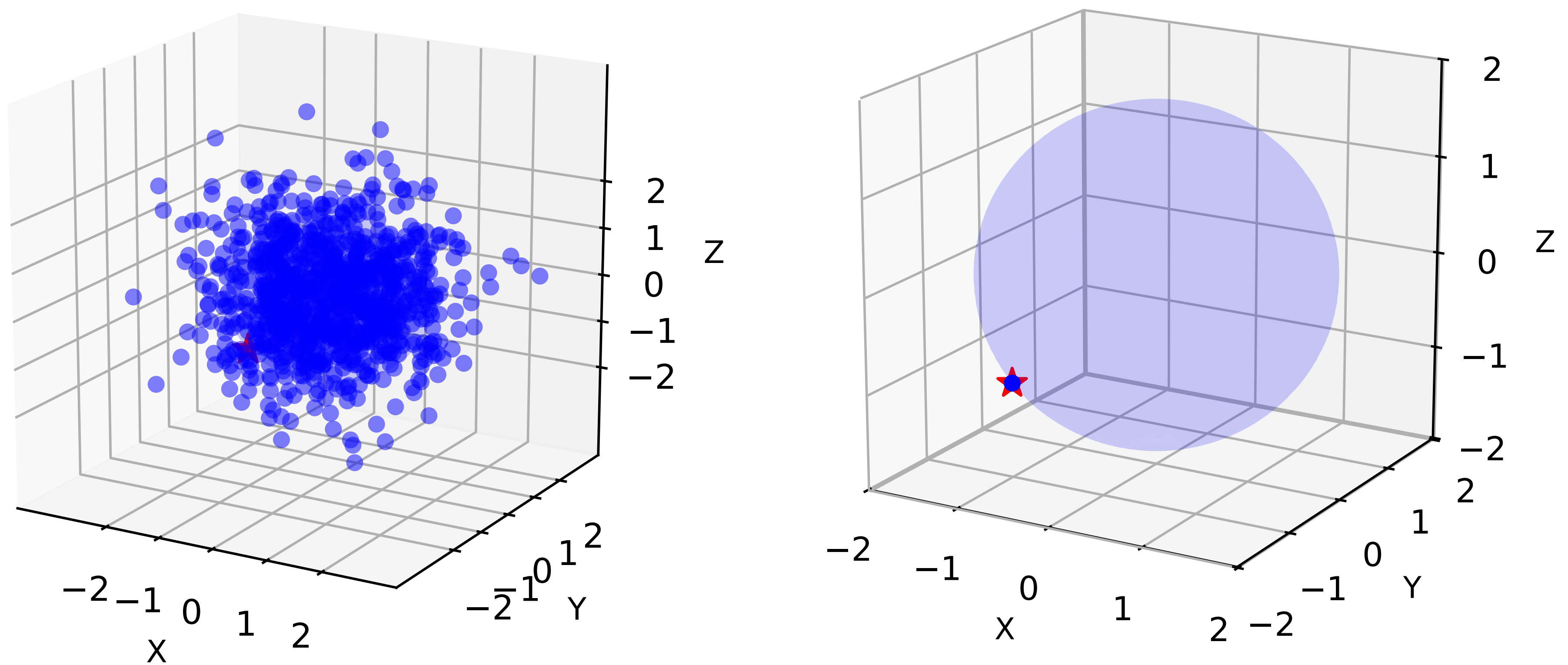
True Solution: $(-1, -1, -1)^T$

Local Maximizer: $(1, 1, 1)^T$

Numerical Experiments 1: Sphere Problem



Numerical Experiments 1: Sphere Problem



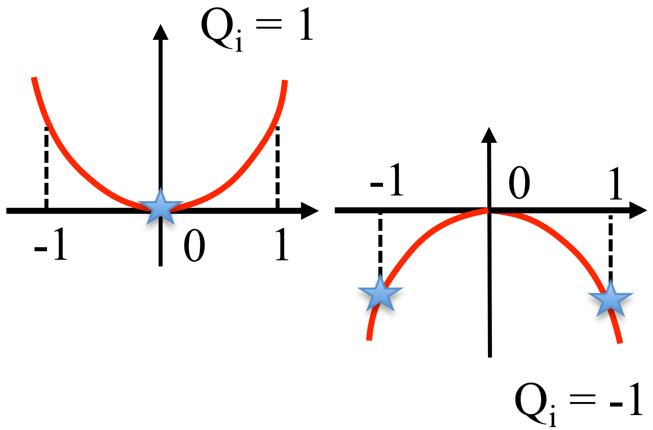
Numerical Experiments 2: Non-convex Problem

Problem Formulation:

$$\min_{x \in R^{100}} \quad \frac{1}{2} x^T \mathsf{Q} x$$
$$\text{subject to} \quad -1 \leq x_i \leq 1 \quad i = 1, 2, \ldots, 100$$

Randomly Generated

$$\text{diag}(\mathsf{Q}) = [1, -1, ...1, -1, -1] \in \mathbb{R}^n$$
$$x_0 \in [-2, 2]$$



Numerical Experiments 2: Non-convex Problem

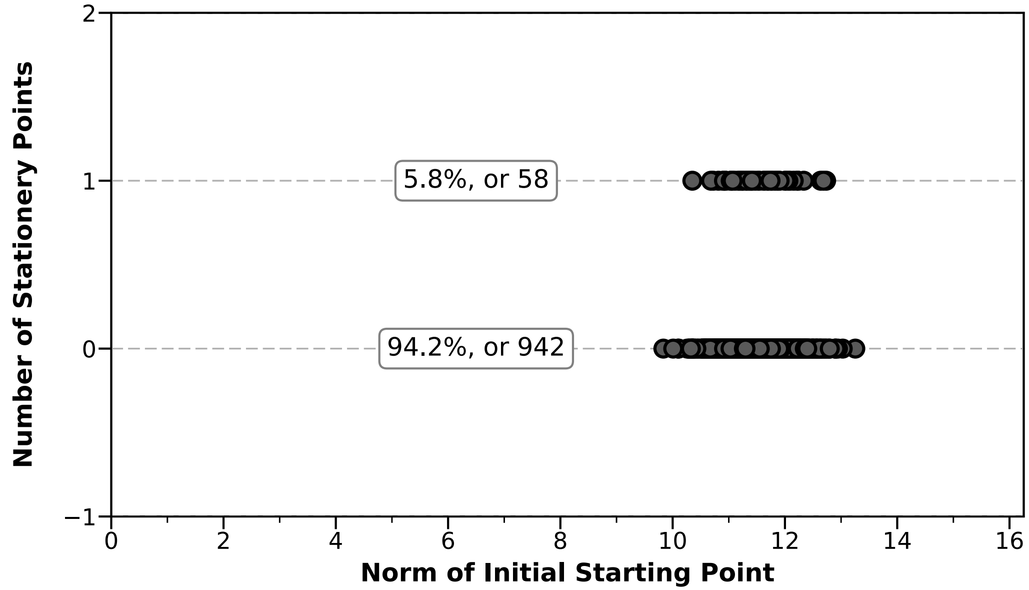
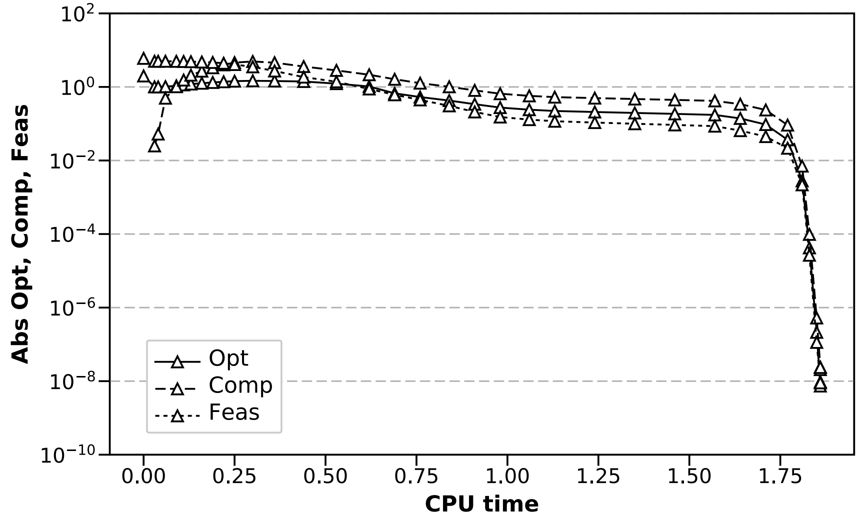
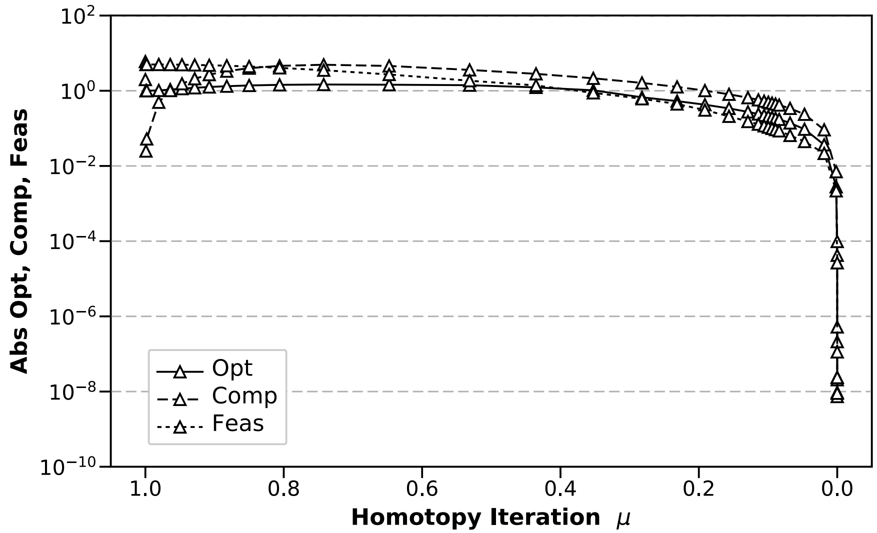


Table 2.1: Success Rate with Different Parameters

		ϵ_{krylov}					
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
τ and ϵ_H	10^{-1}	51%	90.0%	94.2%	94.6%	93.9%	93.8%
	10^{-2}	47.2%	93.1%	94.4%	93.9%	94.2%	94.5%

Numerical Experiments 2: Non-convex Problem





Outline

1. Introduction
2. Homotopy-Based Globalization
3. Iterative Solver & Preconditioner
4. Tests and Applications
5. Summary and Recommendations

The most expensive part of the algorithm is solving the linear systems

Predictor Step

$$(\nabla_q H)_k \, q_k' = -\nabla_\mu H_k = F(q_k) - G(q_k, q_0)$$

Newton Step in Corrector

$$(\nabla_q H)_k \, \Delta q_k = -H_k$$

The linear systems that arise are solved inexactly using FGMRES

In FGMRES, the solution basis Z_i satisfies

$$(\nabla_q H)\mathsf{Z}_i = \mathsf{V}_{i+1}\bar{\mathsf{H}}_i$$

Solve the least square problem to get the solution

$$\begin{aligned} y_i = \underset{y\in\mathbb{R}^i}{\operatorname{argmin}}\|b-(\nabla_q H)\mathsf{Z}_i y\| &= \underset{y\in\mathbb{R}^i}{\operatorname{argmin}}\|\mathsf{V}_{i+1}(\|b\|e_1-\bar{\mathsf{H}}_i y)\| \\ &= \underset{y\in\mathbb{R}^i}{\operatorname{argmin}}\|\,\|b\|e_1-\bar{\mathsf{H}}_i y\| \\ &\quad (\text{since } \mathsf{V}_{i+1}^T\mathsf{V}_{i+1}=\mathsf{I}) \end{aligned}$$

Preconditioner $\quad z_j = P_j(v_j), \quad \forall j = 1,2,\ldots,i$

Final solution $\quad x_i = \mathsf{Z}_i y_i$

Review on Preconditioners for Optimization

- A preconditioner transform the linear system to a better conditioned one with the same solutions:

$$P_j(u) \approx (\nabla_q H)^{-1} u$$

- The true KKT system in optimization is ill-conditioned
- General preconditioners based on explicit matrix not applicable
- Full-space preconditioners based on explicit constraint Jacobian not applicable

The preconditioner is designed to solve an approximation of the KKT system

Consider only $\mu = 0$

$$\begin{bmatrix} \nabla_{xx}\tilde{\mathcal{L}} & \mathbf{0} & \nabla_x\tilde{h}^T & \nabla_x\tilde{g}^T \\ \mathbf{0} & -\mathsf{\Lambda}_g & \mathbf{0} & -\mathsf{S} \\ \nabla_x\tilde{h} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \nabla_x\tilde{g} & -\mathsf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} v_x \\ v_s \\ v_h \\ v_g \end{bmatrix} = \begin{bmatrix} u_x \\ u_s \\ u_h \\ u_g \end{bmatrix}$$

For inequality-only constrained problems: the matrix-free preconditioner involves several steps

Step 1: reducing the preconditioning system into $[v_s, v_g]^T$ and its Schur complement

$$\left[\mathsf{W}_\mu + \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \Lambda_\mu \mathsf{A}_{g,\mu}\right] v_x = u_x - \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \left[(1-\mu) u_s - \Lambda_\mu u_g\right]$$

$$\begin{bmatrix} v_s \\ v_g \end{bmatrix} = \begin{bmatrix} \mathsf{C}_\mu^{-1} & 0 \\ 0 & \mathsf{C}_\mu^{-1} \end{bmatrix} \begin{bmatrix} -\mu \mathsf{I} & \mathsf{S}_\mu \\ (1-\mu)\mathsf{I} & -\Lambda_\mu \end{bmatrix} \begin{bmatrix} u_s \\ u_g - \mathsf{A}_{g,\mu} v_x \end{bmatrix}$$

$$\mathsf{C}_\mu \equiv \mu \Lambda_\mu - (1-\mu)\mathsf{S}_\mu = \mu(1-\mu)\Lambda_g - \mu^2 \mathsf{I} - (1-\mu)^2 \mathsf{S}$$

The matrix-free preconditioner involves several steps

Step 2: Lanczos SVD approximation

$$\tilde{\mathsf{A}}_{g,\mu}^{T}\mathsf{C}_{\mu}^{-1}\Lambda_{\mu}\tilde{\mathsf{A}}_{g,\mu} = M_{n\times k}\Gamma_{k\times k}N_{k\times n}^{\star}$$

Step 3: Sherman-Morrison formula to obtain the desired design update

$$B = A + UV$$

$$B^{-1} = A^{-1} - A^{-1}U(I_k + VA^{-1}U)^{-1}VA^{-1}$$

For general constrained problems, the Schur complement system is different

$$\begin{bmatrix} \mathsf{W}_\mu + \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \Lambda_\mu \mathsf{A}_{g,\mu} & \mathsf{A}_{h,\mu}^T \\ \mathsf{A}_{h,\mu} & -\mu \mathsf{I} \end{bmatrix} \begin{bmatrix} v_x \\ v_h \end{bmatrix} = \begin{bmatrix} u_x - \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \left[(1-\mu) u_s - \Lambda_\mu u_g\right] \\ u_h \end{bmatrix}$$

To solve it, the inverse of the Schur complement is used:

$$\begin{bmatrix} v_x \\ v_h \end{bmatrix} = \begin{bmatrix} I & 0 \\ \frac{1}{\bar{\mu}} \mathsf{A}_{h,\mu} & I \end{bmatrix} \begin{bmatrix} \left(\mathsf{W}_\mu + \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \Lambda_\mu \mathsf{A}_{g,\mu} + \frac{1}{\bar{\mu}} \mathsf{A}_{h,\mu}^T \mathsf{A}_{h,\mu} \right)^{-1} & 0 \\ 0 & -\frac{1}{\bar{\mu}} \mathsf{I} \end{bmatrix} \begin{bmatrix} I & \frac{1}{\bar{\mu}} \mathsf{A}_{h,\mu}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{u}_x \\ u_h \end{bmatrix}$$

The Lanczos SVD approximation is applied to the combined equality and inequality Jacobians:

$$\mathsf{W}_\mu + \mathsf{A}_\mu^T \Sigma_\mu \mathsf{A}_\mu = \mathsf{W}_\mu + \mathsf{A}_{g,\mu}^T \mathsf{C}_\mu^{-1} \Lambda_\mu \mathsf{A}_{g,\mu} + \frac{1}{\bar{\mu}} \mathsf{A}_{h,\mu}^T \mathsf{A}_{h,\mu}$$

$$\mathsf{A}_\mu = \begin{bmatrix} \mathsf{A}_{h,\mu} \\ \mathsf{A}_{g,\mu} \end{bmatrix} \quad \text{and} \quad \Sigma_\mu = \begin{bmatrix} \frac{1}{\bar{\mu}} \mathsf{I} & 0 \\ 0 & \mathsf{C}_\mu^{-1} \Lambda_\mu \end{bmatrix}$$

Numerical Experiments: Scalable Quadratic Problem

Problem Formulation: $$\min_{x \in R^n} \quad \frac{1}{2} x^T \mathsf{Q} x + g^T x$$
$$\text{subject to} \quad \mathsf{A}x \geq b$$

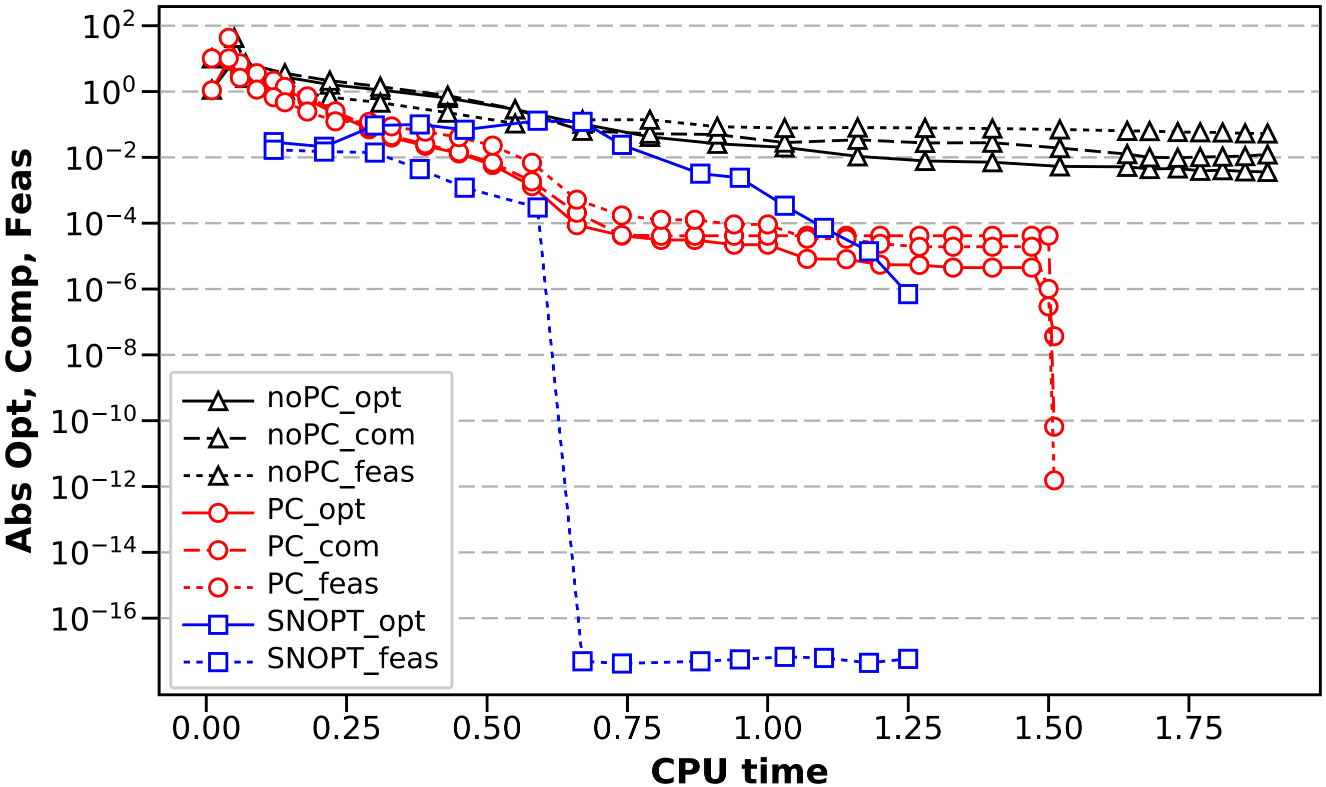
The singular values of Q and A are predefined:

$$\mathsf{Q}_{ii} = \begin{cases} \frac{1}{i}, & i = 1, 2, ..., \kappa, \\ \frac{1}{\kappa}, & i = \kappa + 1, ..., n, \end{cases}$$

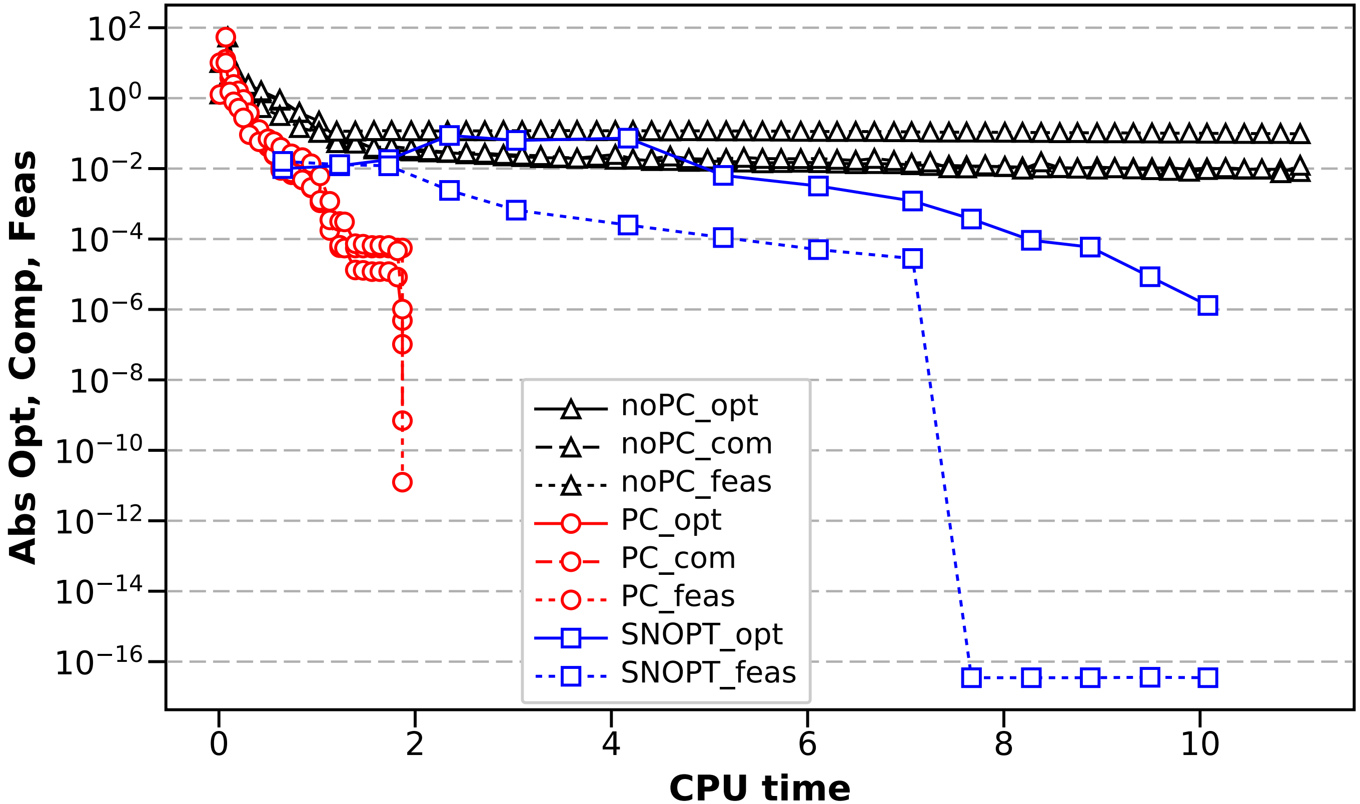
$$\mathsf{A} = \mathsf{Q}_L \mathsf{D} \mathsf{Q}_R \begin{cases} \mathsf{D}_{ii} = \begin{cases} \frac{1}{i^2}, & i = 1, 2, ..., \nu, \\ \frac{1}{\nu^2}, & i = \nu + 1, ..., n, \end{cases} \\ \mathsf{Q}_L \; \mathsf{Q}_R \end{cases}$$

Randomly generated orthonormal matrix

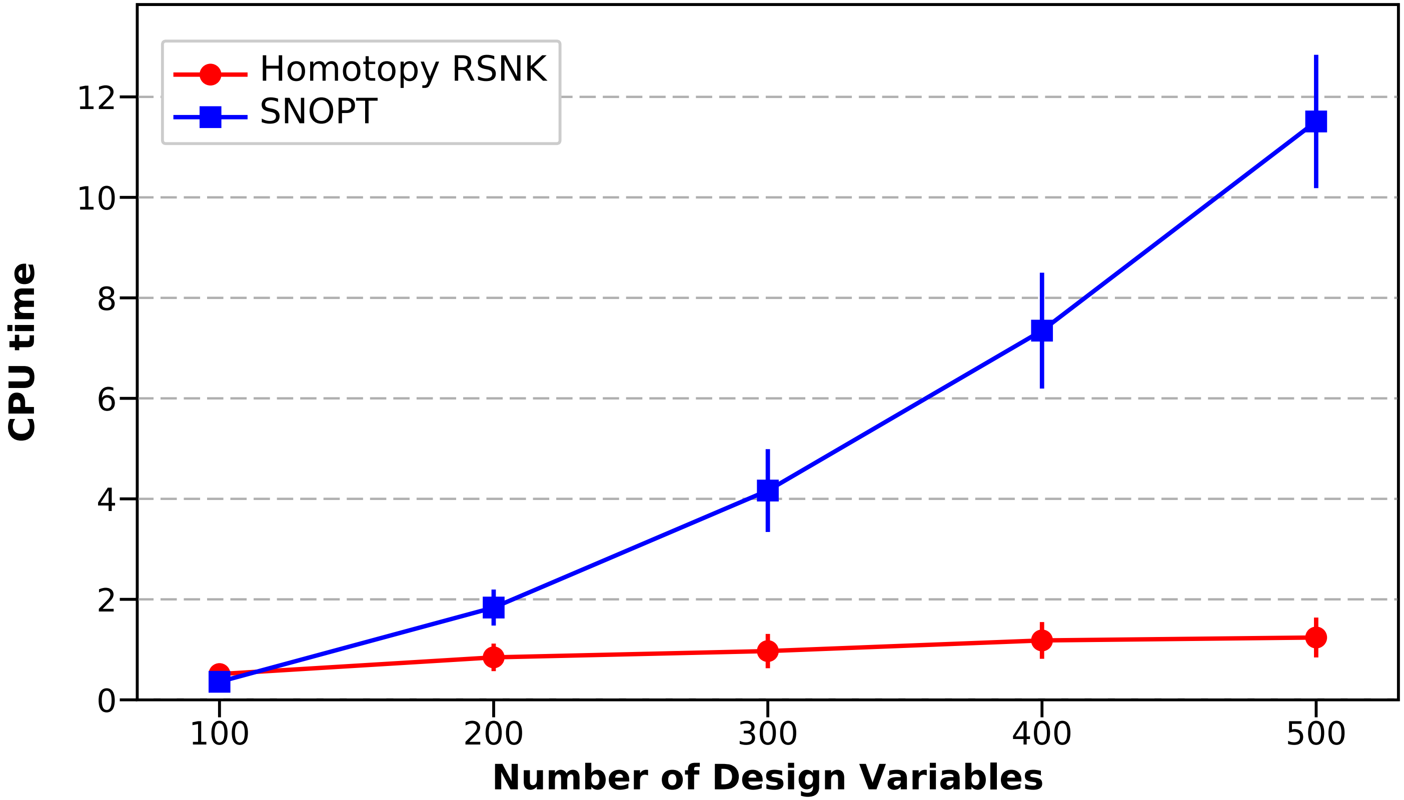
Convergence plots with number of design: 200



Convergence plots with number of design: 500



Comparison of scalability performance



Outline

1. Introduction
2. Homotopy-Based Globalization
3. Iterative Solver & Preconditioner
4. Tests and Applications
5. Summary and Recommendations

Software Implementation and Benchmarking

1. Kona : matrix-free optimization package for PDE-constrained problems

2. Convergence Criteria

$$\text{Optimality} = \max_j |(\nabla_x \mathcal{L})_j|,$$

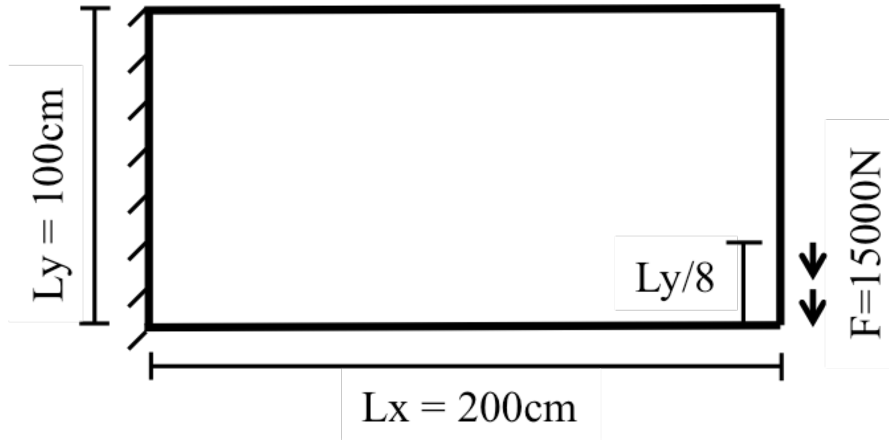
$$\text{Complementarity} = \max_j |s_j \lambda_j|,$$

$$\text{Feasibility} = \max_j |c_j|, \quad \text{where} \quad c = \left[h(x)^T, (g(x) - s)^T\right]$$

3. For benchmarking, we compare against SNOPT, which is a state-of-the-art active-set algorithm frequently used in the engineering optimization community

Test 1: A subset of CUTEr test problems

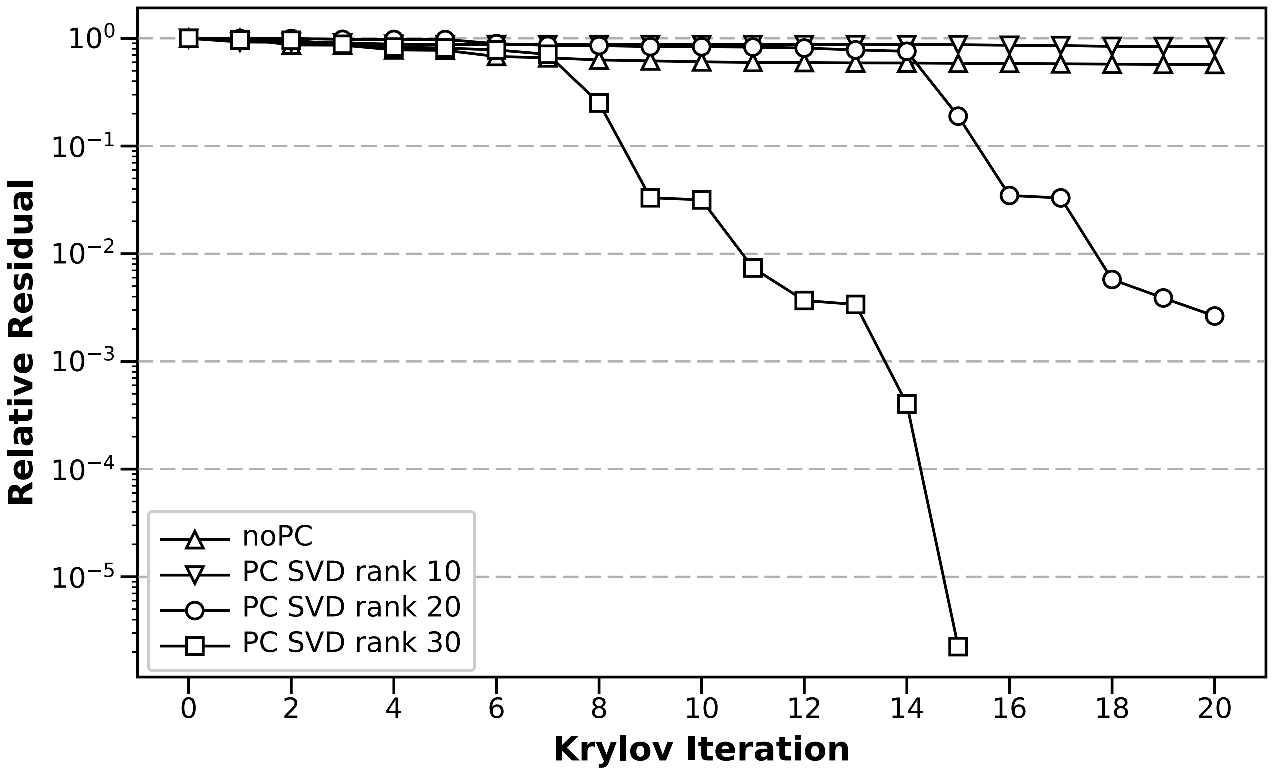
Test 2: Stress-Constrained Mass Minimization



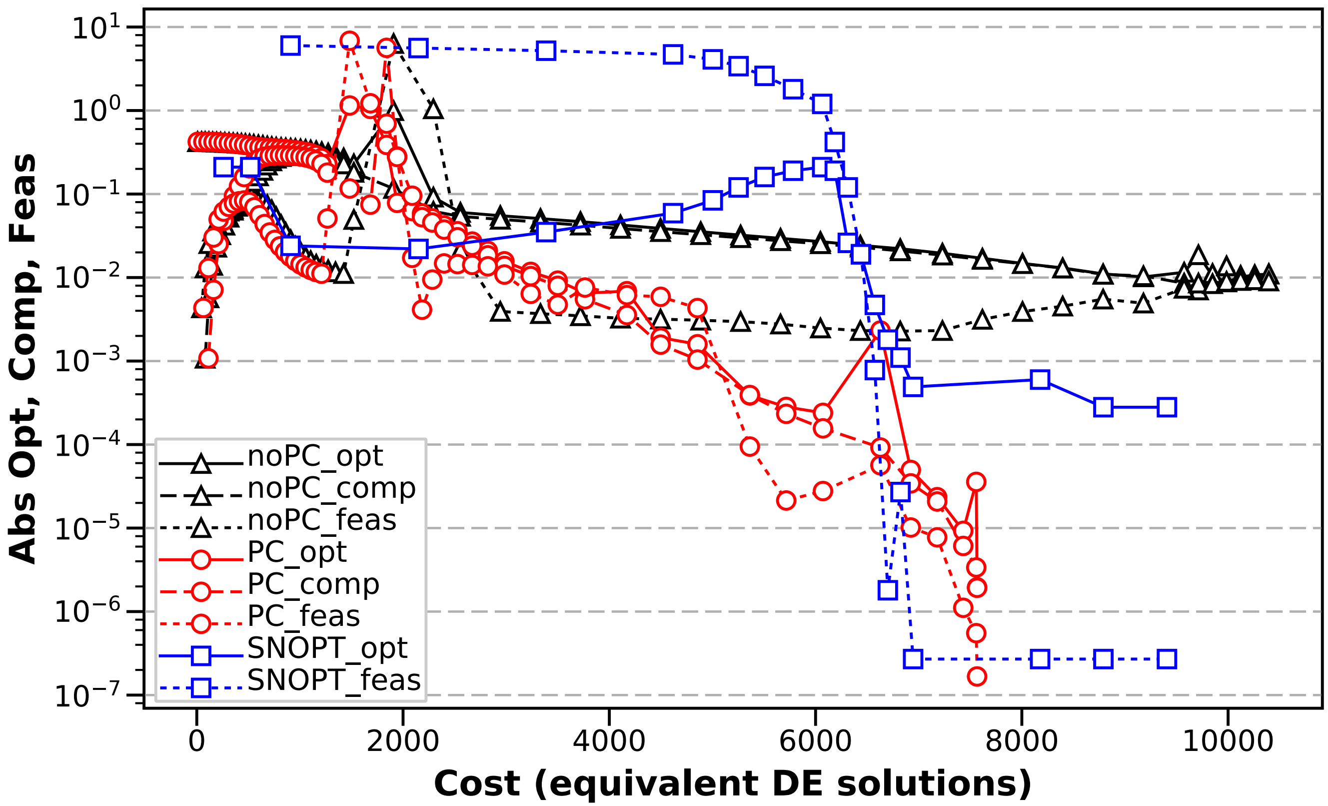
Case	nx	ny	Number of design
Small	16	8	128
Medium	32	16	512
Large	64	32	2048

$$
\begin{aligned}
\min_{x} \quad & \mathsf{mass}(x) \\
\text{subject to} \quad & \mathsf{stress}_i(x) \leq \sigma_{\max}, \forall\, i = 1, 2, \ldots, n, \\
& x_l \leq x_i \leq x_u, \qquad \forall\, i = 1, 2, \ldots, n,
\end{aligned}
$$

Effectiveness of Preconditioner

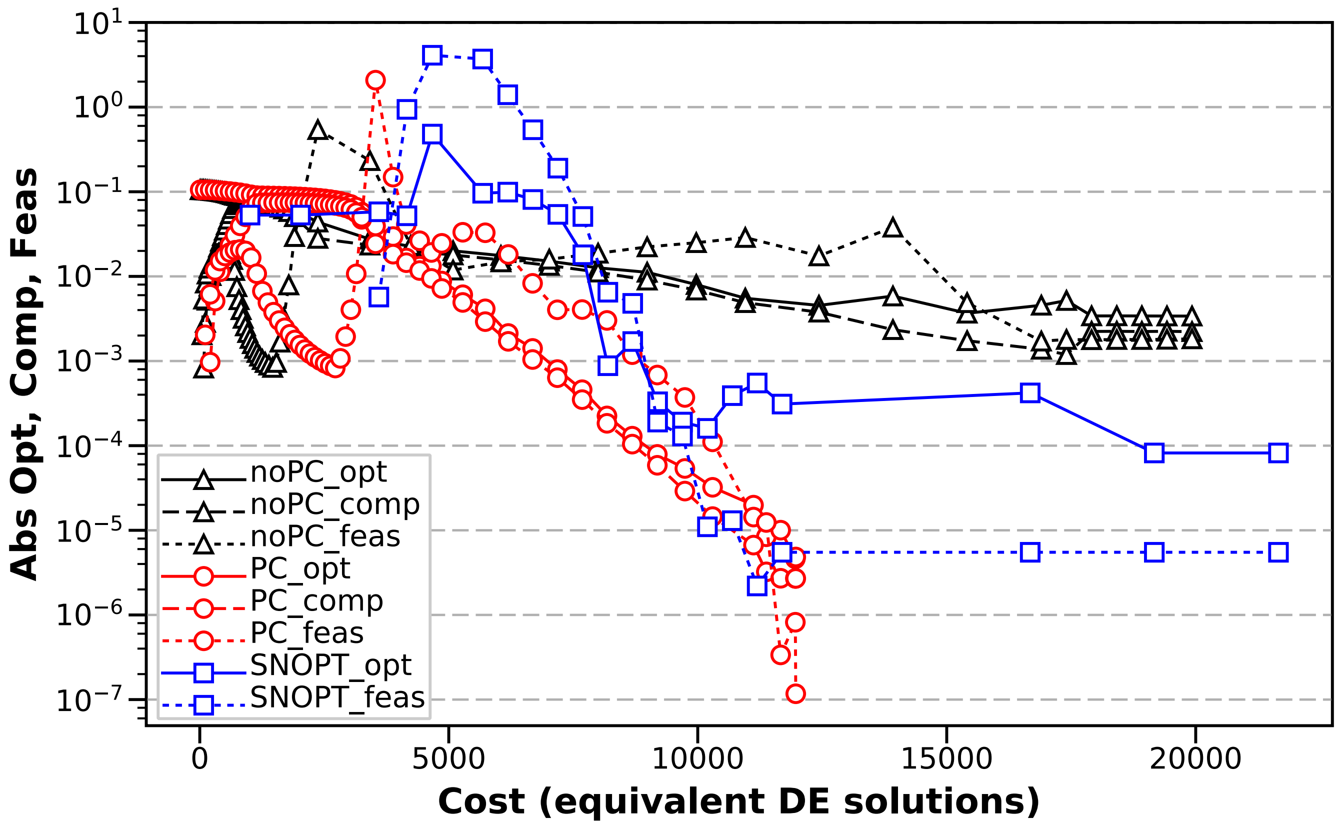


Convergence Plots: Kona vs. SNOPT



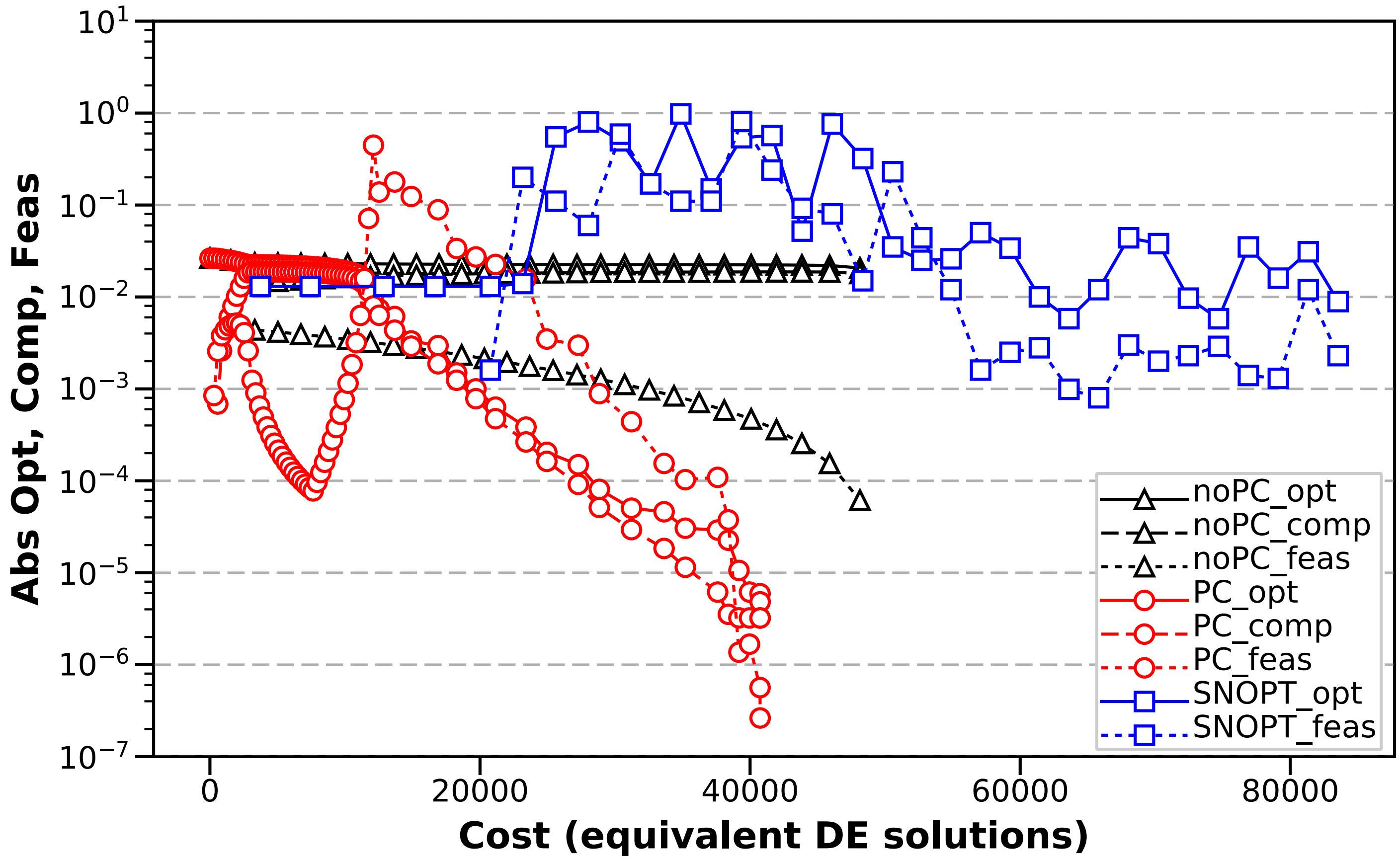
Small

Convergence Plots: Kona vs. SNOPT



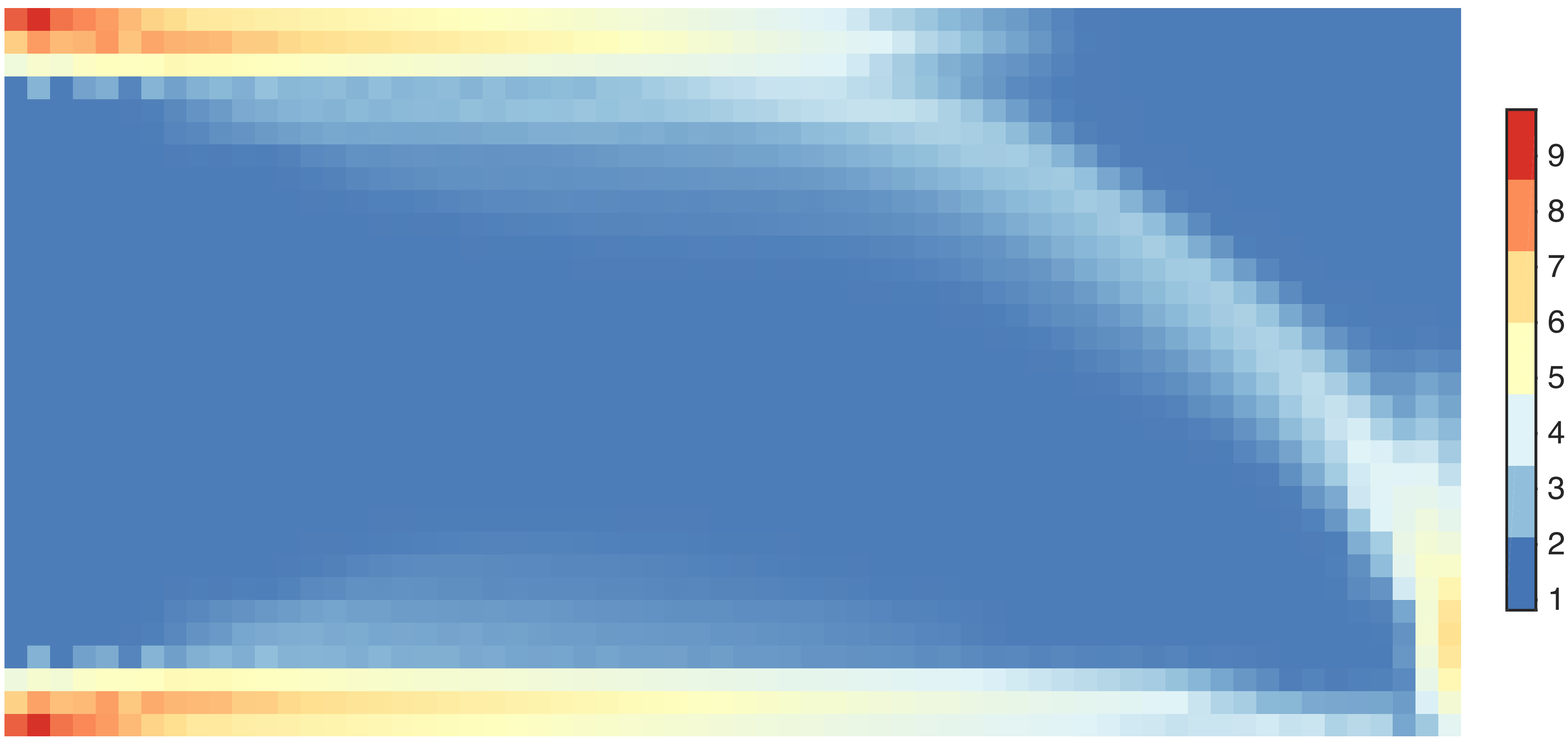
Medium

Convergence Plots: Kona vs. SNOPT

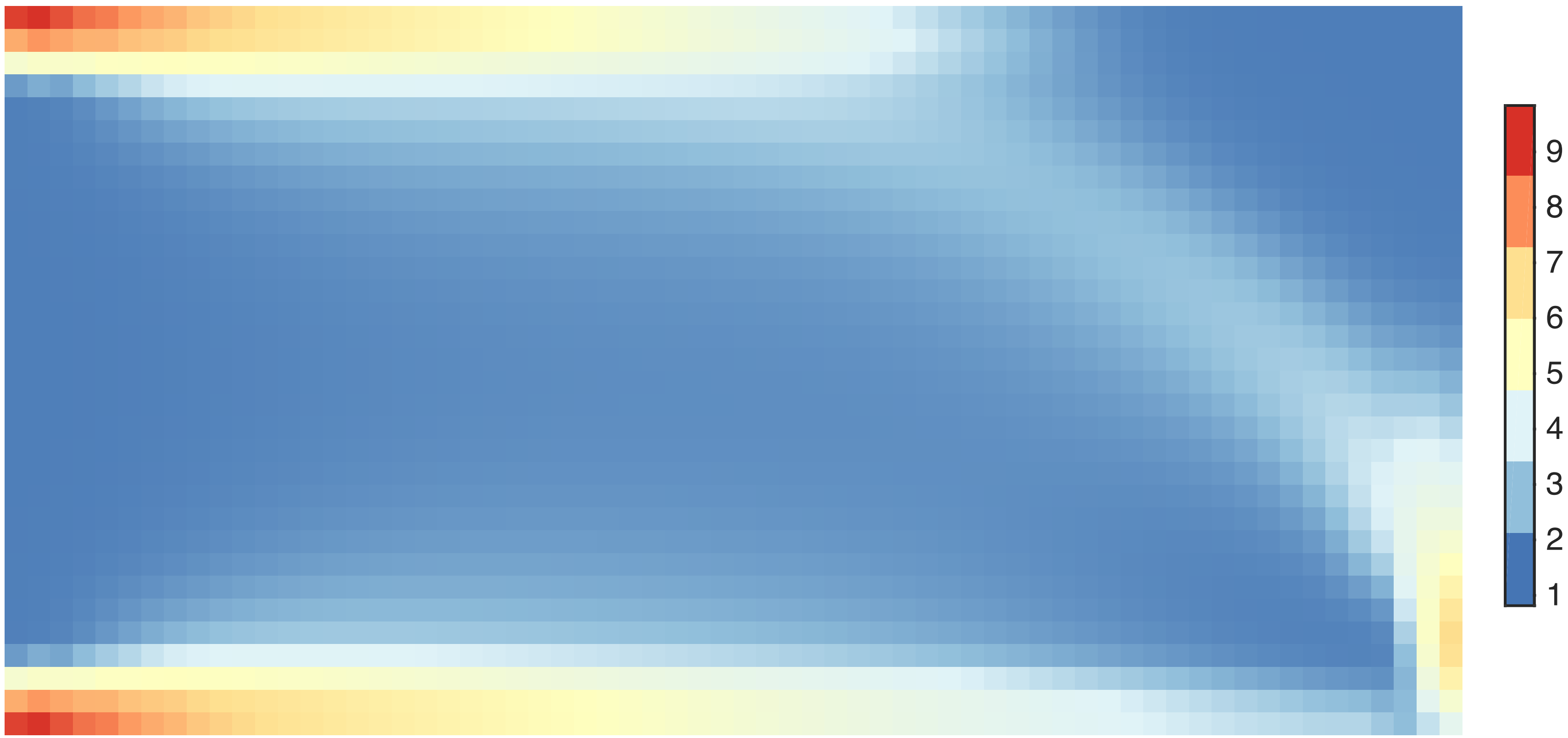


Large

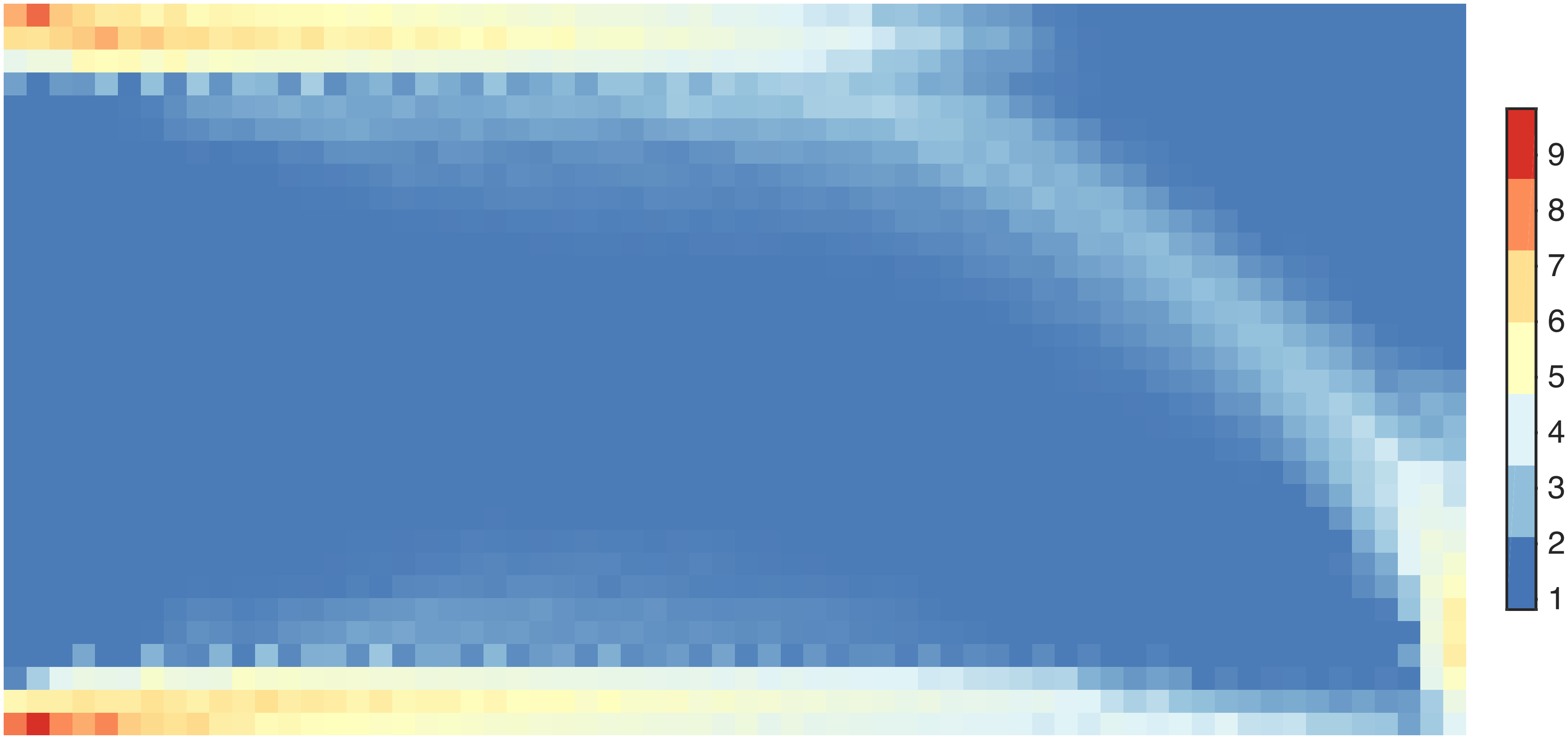
Thickness distribution of the large case with the preconditioner



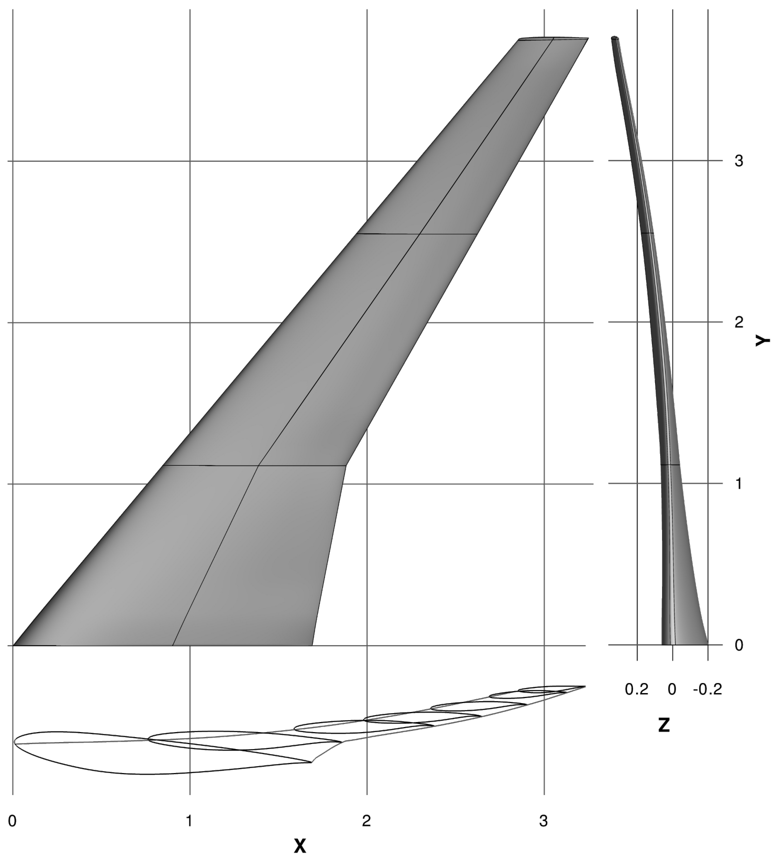
Thickness distribution of the large case without the preconditioner



Thickness distribution of the large case with SNOPT



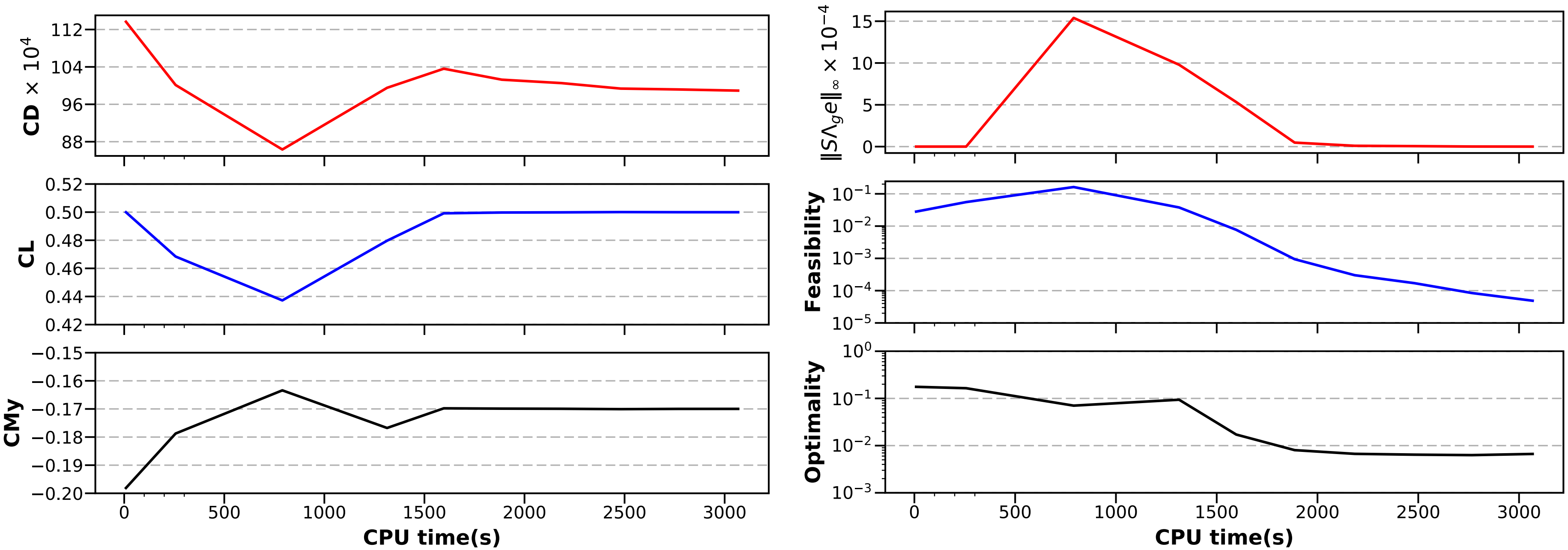
Test 3: Aerodynamic Shape Optimization Problem



$$\begin{aligned} \min_{x} \quad & C_D \\ \text{subject to} \quad & C_L \geq 0.5 \\ & C_{M_y} \geq -0.17 \\ & t \geq 0.25 t_{\text{base}} \\ & V \geq V_{\text{base}} \\ & \Delta z_{TE,\text{upper}} = -\Delta z_{TE,\text{lower}} \\ & \Delta z_{LE,\text{upper, root}} = -\Delta z_{LE,\text{lower, root}} \end{aligned}$$

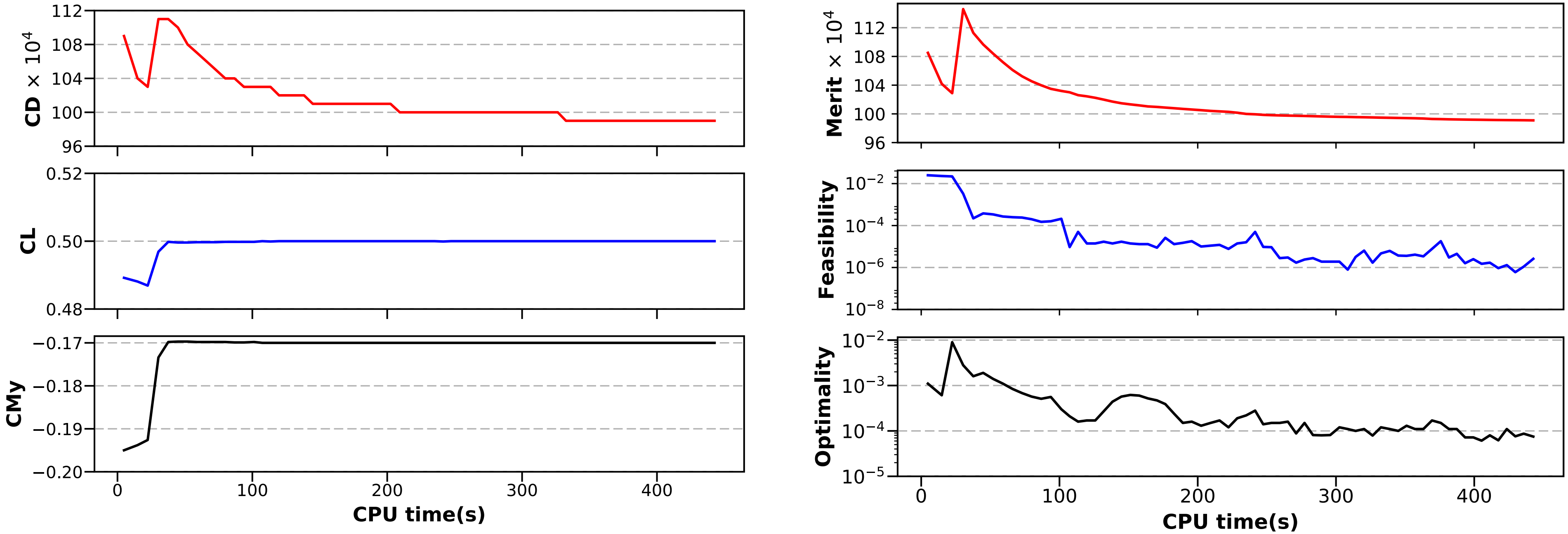
	192	**480**	**768**
Chordwise	12	20	24
Spanwise	8	12	16
Vertical	2	2	2

Aerodynamic coefficient history and optimization plots using Kona



Number of Design: 768

Aerodynamic coefficient history and optimization plots using SNOPT



Number of Design: 768

Outline

1. Introduction
2. Homotopy-Based Globalization
3. Iterative Solver & Preconditioner
4. Tests and Applications
5. Summary and Recommendations

Conclusions

- Developed a matrix-free optimization algorithm for reduced-space PDE-constrained optimization problems that have state-based constraints
- Globalization and nonconvexity are handled using a homotopy map and predictor-corrector method
- A novel and effective matrix-free preconditioner has been created for inequality-only and general constrained problems

Recommendations

- Provide preconditioners for different types of problems in Kona
- Separating nonlinear and linear constraints in Kona API
- Improve the robustness of the method in the context of nonconvex problems
- Improve the efficiency of the preconditioner for ASO problem and run the RANS equation
- Automatically determine the parameters

Thank you!
Questions?