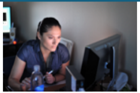
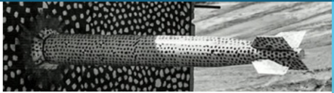
-8772C

National Laboratories

An Inexact Trust-Region Newton Method for Large-Scale Convex-Constrained Optimization





Drew P. Kouri

SIAM Conference on Optimization

Virtual

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell

dia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly ow subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2021-xxxx PE

Outline

1. Problem Formulation
2. Motivating Application
3. Trust-Regions for Convex Constraints
4. Spectral Projected Gradient Subproblem Solver
5. Numerical Results

RAPID OPTIMIZATION LIBRARY

Problem Formulation

We consider the optimization problem

$$\min_{x\in H}\ f(x) \quad \text{subject to} \quad x\in\mathcal{C},$$

where:

- H is a Hilbert space;
- $\mathcal{C}\subset H$ is a nonempty, closed, and convex set;
- $f:H\to\mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient.

Assumption: There exists $\gamma\in\mathbb{R}$ such that the level set L_γ is bounded, where

$$L_\gamma := \{x\in\mathcal{C}\,|\,f(x)\le\gamma\}.$$

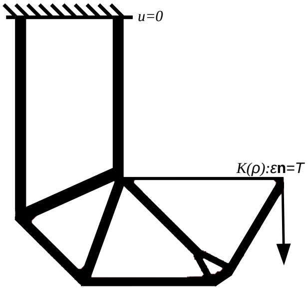
$\implies$ The problem has a solution and we can replace $\mathcal{C}$ by the closed convex hull of L_γ.

Notation: $\mathbf{P}_{\mathcal{C}}(x)$ is the (unique) projection of $x\in H$ onto $\mathcal{C}$ and is given by

$$\mathbf{P}_{\mathcal{C}}(x) := \operatorname*{arg\,min}_{y\in\mathcal{C}}\ \|x-y\|_H.$$

Motivating Application

Elastic Topology Optimization



Given a domain $\Omega \subset \mathbb{R}^d$ and a volume fraction $v \in (0,1)$,

$$\min_{\rho \in L^2(\Omega)} \int_{\Gamma_t} T(x) \cdot [S(\rho)](x)\, \mathrm{d}x$$

$$\text{subject to} \quad \int_\Omega \rho(x)\, \mathrm{d}x \le v|\Omega|, \quad 0 \le \rho \le 1 \text{ a.e.},$$

where $S(\rho)) = u \in (H^1(\Omega))^d$ solves

$$-\nabla \cdot (K(\rho) : \varepsilon) = 0 \qquad \text{in } \Omega$$

$$\varepsilon = \frac{1}{2}(\nabla u + \nabla u^\top) \qquad \text{in } \Omega$$

$$K(\rho) : \varepsilon \mathbf{n} = T \qquad \text{on } \Gamma_t$$

$$u = 0 \qquad \text{on } \Gamma_d$$

Goal: Determine a ***binary*** ρ that is maximally stiff and that satisfies the volume constraint.

Motivating Application

Challenges:

1. Binary solutions are difficult to compute (i.e., mixed integer PDE-constrained optimization);
2. Continuous (grey) solutions can be challenging to interpret (i.e., micro-structure, alloys, etc.);
3. PDE can be extremely expensive to solve—difficult to prove convergence to infinite-dimensional problem.

Common Solution: Solid **I**sotropic **M**aterial with **P**enalization

$$K(\rho) := K_0 + (K_1 - K_0)\rho^p \qquad \text{for} \qquad p > 1.$$

SIMP problem is **highly nonconvex** and has **infinitely many solutions**!

After discretization, SIMP can lead to **checkerboard (i.e., mesh-dependent) designs**!

Filtering: To enforce a length scale in the design, it is common to filter $\rho \leftarrow F(\rho)$, where $F : L^2(\Omega) \to L^2(\Omega)$ is a compact operator that perserves volume.

Motivating Application

Consider the optimization problem

$$\min_{\rho\in\mathbb{R}^2} \; T[S(\rho)](1) \quad \text{subject to} \quad \tfrac{1}{2}(\rho_1+\rho_2) \le v, \quad 0 \le \rho_1,\, \rho_2 \le 1,$$

where $S(\rho) = u$ solves

$$[K(\rho)u']' = 0 \quad \text{in } (0,1), \qquad u(0)=0, \qquad [K(\rho)u'](1) = T,$$

and

$$[K(\rho)](x) = \begin{cases} k_0 + (1-k_0)\rho_1^p & \text{if } x < 0.5 \\ k_0 + (1-k_0)\rho_2^p & \text{if } x > 0.5 \end{cases}.$$

For this problem, $S(\rho)$ can be computed analytically as

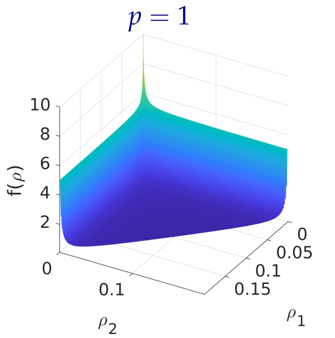
$$[S(\rho)](x) = \begin{cases} \frac{Tx}{k_0+(1-k_0)\rho_1^p} & \text{if } x < 0.5 \\ \frac{0.5T}{k_0+(1-k_0)\rho_1^p} - \frac{T(x-0.5)}{k_0+(1-k_0)\rho_2^p} & \text{if } x < 0.5 \end{cases}.$$

Motivating Application

An Illustrative Example

Substituting the PDE solution into the objective function yields the optimization problem

$$\min_{\rho\in\mathbb{R}^2} \frac{T^2}{2}\left(\frac{1}{k_0+(1-k_0)\rho_1^p}+\frac{1}{k_0+(1-k_0)\rho_2^p}\right) \quad \text{subject to} \quad \tfrac{1}{2}(\rho_1+\rho_2)\le v, \quad 0\le \rho_1,\ \rho_2\le 1.$$



$p = 1$ is convex — $p = 3$ produces multiple local minima!

Motivating Application

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe and Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

2. **Method of Moving Asymptotes:** A sequential convex optimization approach that uses rational approximations of the objective and constraints. The dual subproblem is commonly solved using nonlinear CG. This method is inherently **finite dimensional**.

   Svanberg, The method of moving asymptotes—A new method for structural optimization, IJNME, 1987.

3. **Augmented Lagrangian:** Robust, yet minimizing the penalty function at each iteration can be expensive.

4. **Interior Points:** Primal-dual line-search methods have been used succesfully. However, nonconvexity can lead to expensive inertia correction.

It can be extremely difficult to incorporate inexactness in these methods!

Trust-Regions for Convex Constraints

Trust-Region Subproblem: At each iteration, we approximately solve

$$\min_{x\in H}\ m_k(x) \quad \text{subject to} \quad x \in \mathcal{C},\ \ \|x - x_k\| \le \Delta_k,$$

where $\Delta_k > 0$ is the radius and $m_k : H \to \mathbb{R}$ is a model of the f near the iterate x_k.

Generalized Cauchy Point: A point along the projected gradient path

$$s_k^{\mathrm{GCP}} := d_k(t_k) \quad \text{where} \quad d_k(t) := \mathbf{P}_{\mathcal{C}}(x_k - tg_k) - x_k$$

and $g_k := \nabla m_k(x_k)$, that satisfies both

1. **Trust-Region Feasibility:** $\|s_k^{\mathrm{GCP}}\| \le \nu_1 \Delta_k$
2. **Sufficient Model Decrease:** $m_k(s_k^{\mathrm{GCP}}) - m_k(x_k) \le \mu_1(g_k, s_k^{\mathrm{GCP}})$

and at least one of the following conditions

1. **Sufficient Step Length:** $t_k \ge \nu_2 t_k'$ with $m_k(x_k + d_k(t_k')) - m_k(x_k) \ge \mu_2(g_k, d_k(t_k'))$
2. **Sufficient Step Length:** $t_k \ge \min\{\nu_3 \Delta_k / \|g_k\|, \nu_4\}$.

Trust-Regions for Convex Constraints

Require: An initial guess $x_0 \in \mathcal{C}$, initial trust-region radius $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$

for $k = 1, 2, \ldots$ **do**

Cauchy Point Computation: Compute a generalized Cauchy point $x_k^{\mathrm{GCP}} \in \mathcal{C}$

Step Computation: Compute a trial step $x_{k+1} \in \mathcal{C}$ that satisfies

$$m_k(x_k) - m_k(x_{k+1}) \ge \mu_3(m_k(x_k) - m_k(x_k^{\mathrm{GCP}}))$$

Step Acceptance: Compute ratio of actual and predicted reduction:

$$\rho_k := \frac{f(x_k) - f(x_{k+1})}{m_k(x_k) - m_k(x_{k+1})} < \eta_1 \qquad \Longrightarrow \qquad x_{k+1} \leftarrow x_k$$

Update Trust-Region Radius: $\Delta_{k+1} \in \begin{cases} [\gamma_1\Delta_k, \gamma_2\Delta_k] & \text{if } \rho_k < \eta_1 \\ [\gamma_2\Delta_k, \Delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2) \\ [\Delta_k, \infty) & \text{if } \rho_k \ge \eta_2 \end{cases}$

end for

Trust-Regions for Convex Constraints

Convergence of the form

$$\liminf_{k\to\infty} \|\mathbf{P}_{\mathcal{C}}(x_k - \nabla f(x_k)) - x_k\| = 0$$

was proved in Theorem 10 of

Toint, Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space, IMA Journal of Numerical Analysis, 1988.

This theory permits **inexact model gradients** that satisfy, e.g., $\exists\kappa > 0$ such that

$$\|\nabla f(x_k) - g_k\| \le \kappa \min\{\|\mathbf{P}_{\mathcal{C}}(x_k - g_k) - x_k\|, \Delta_k\}.$$

However, it does not account for **inexact objective function evaluations**!

Trust-Regions for Convex Constraints

In many applications, we can only hope to compute

$$f_k(x) \approx f(x) \qquad \text{and} \qquad \rho_k = \frac{f_k(x_k) - f_k(x_{k+1})}{m_k(x_k) - m_k(x_{k+1})}.$$

Fortunately, the **inexact objective function** criteria

$$|(f(x_k) - f(x_{k+1})) - (f_k(x_k) - f_k(x_{k+1}))| \leq K(\eta \min\{m_k(x_k) - m_k(x_{k+1}), r_k\})^{1/\omega},$$

where $K > 0$ and $\omega \in (0,1)$ are fixed, $\eta < \min\{\eta_1, 1-\eta_2\}$ and $r_k \searrow 0$, can be applied with **little change to the theory**!

This condition was used for unconstrained problems in

Kouri, et al., Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty, SISC, 2014.

Spectral Projected Gradient Subproblem Solver

Spectral Projected Gradient: A first-order method that approximates second-order information using a Barzilai-Borwein (spectral) step length.

Applied to our original optimization problem

$$\min_{x \in H} f(x) \qquad \text{subject to} \qquad x \in \mathcal{C},$$

the SPG method produces feasible iterations

$$x_{k+1} = x_k + \alpha_k s_k \qquad \text{where} \qquad s_k := (\mathbf{P}_{\mathcal{C}}(x_k - \lambda_k \nabla f(x_k)) - x_k).$$

The step length $\alpha_k > 0$ is computed using a nonmonotone line search and the spectral step length $\lambda_k > 0$ is computed as

$$\lambda_k := \max\left\{\lambda_{\min}, \min\left\{\lambda_{\max}, \frac{\alpha_k (s_{k-1}, s_{k-1})}{(\nabla f(x_k) - \nabla f(x_{k-1}), s_{k-1})}\right\}\right\}.$$

Birgin, et al., Nonmonotone spectral projected gradient methods on convex sets, SIOPT, 2000.

Spectral Projected Gradient Subproblem Solver

Model: For our trust-region method, we consider the quadratic model

$$m_k(x) := \frac{1}{2}(B_k(x - x_k), (x - x_k)) + (g_k, x - x_k) + f_k(x_k),$$

where $B_k : H \to H$ is linear operator that approximates the Hessian $\nabla^2 f(x_k)$.

Feasible Set: The trust-region subproblem feasible set is

$$\mathcal{C}_k := \{x \in \mathcal{C} \mid \|x - x_k\| \leq \Delta_k\}.$$

SPG Iteration: $x_{k,\ell+1} = x_{k,\ell} + \alpha_\ell s_\ell$ where $s_\ell = \mathbf{P}_{\mathcal{C}_k}(x_{k,\ell} - \lambda_\ell \nabla m_k(x_{k,\ell})) - x_{k,\ell}$

1. Start with $x_{k,0} = x_k + s_k^{\mathrm{GCP}}$ to ensure fraction of Caucy decrease;
2. Use exact line search to determine α_ℓ step length;
3. Spectral step length simplifies to

$$\lambda_\ell := \max\left\{\lambda_{\min}, \min\left\{\lambda_{\max}, \frac{(s_{\ell-1}, s_{\ell-1})}{(B_k s_{\ell-1}, s_{\ell-1})}\right\}\right\}.$$

Spectral Projected Gradient Subproblem Solver

Projection onto $\mathcal{C}_k$: The projecton of $x \in H$ onto $\mathcal{C}_k$ is given by

$$\mathbf{P}_{\mathcal{C}_k}(x) = \begin{cases} \mathbf{P}_{\mathcal{C}}(x) & \text{if } \|\mathbf{P}_{\mathcal{C}}(x) - x_k\| \le \Delta_k \\ \mathbf{P}_{\mathcal{C}}(x_k + t^\star(x - x_k)) & \text{if } \|\mathbf{P}_{\mathcal{C}}(x) - x_k\| > \Delta_k \end{cases},$$

where $t^\star \in [0,1]$ is any $t \in [0,1]$ that satisfies

$$\phi(t) := \|\mathbf{P}_{\mathcal{C}}(x_k + t(x - x_k)) - x_k\| - \Delta_k = 0.$$

Here, ϕ is nondecreasing and continuous on $[0,1]$ with $\phi(0) = -\Delta_k$ and $\phi(1) > 0$.

Can compute $\mathbf{P}_{\mathcal{C}_k}(x)$ by applying, e.g., Brent's method to $\phi(t)$.

Numerical Results

Algorithmic Comparison

- `TRSPG`: Convex-constrained trust-region method that uses the SPG subproblem solver.

  Kouri, A matrix-free trust-region Newton algorithm for convex-constrained optimization, Optimization Online, 2021.

- `LMTR`: Linearly-constrained trust-region method that uses truncated CG to approximately solve the subproblem. `TRON` is a popular implementation for bound-constrained problems.

  Lin and Moré, Newton's method for large bound-constrained optimization problems, SIOPT, 1999.

- `PQN`: Line-search BFGS method that uses SPG to compute the projected quasi-Newton step.

  Schmidt et al., Optimizing costly functions with simple constraints: A limited-memory projected quasi-Newton algorithms, Proceedings of the 12th International Conference on AISTATS, 2009.

- `SPG`: The nonmonotone spectral projected gradient method as previously described.

  Birgin, et al., Nonmonotone spectral projected gradient methods on convex sets, SIOPT, 2000.

- `AL-TRSPG`/`AL-LMTR`: Augmented Lagrangian using `TRSPG` and `LMTR`, respectively.

All methods are implemented in the Rapid Optimization Library!

`https://trilinos.github.io/rol.html`

Constraint Set: The set $\mathcal{C}$ has the same form for all examples:

$$\mathcal{C} = \{x \in L^2(\Omega) \mid a \le x \le b \;\text{ a.e.},\;\; (c,x) = d\},$$

where $\Omega \subset \mathbb{R}^n$ for $n = 2$ or 3, $a, b, c \in L^2(\Omega)$, $a \le b$ a.e., and $d \in \mathbb{R}$.

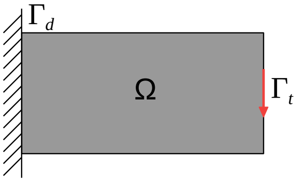
Projection Algorithm: Apply a *secant* method to the dual optimality conditions:

$$\text{Find} \qquad \lambda \in \mathbb{R} \qquad \text{such that} \qquad (c, \mathbf{P}_{[a,b]}(x - \lambda c)) = d.$$

Dai and Fletcher, New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds, Math Programming, 2006.

Numerical Results

Elastic Topology Optimization



Let $\Omega = (0,2) \times (0,1)$ and $v = 0.4$, and consider

$$\min_{\rho \in L^2(\Omega)} \int_{\Gamma_t} T(x) \cdot [S(\rho)](x)\,\mathrm{d}x$$

$$\text{subject to} \quad \int_\Omega \rho(x)\,\mathrm{d}x = v|\Omega|, \quad 0 \le \rho \le 1 \;\text{ a.e.},$$

where $S(\rho)) = u \in (H^1(\Omega))^2$ solves

$$\begin{aligned} -\nabla \cdot (K(\rho) : \varepsilon) &= 0 && \text{in } \Omega \\ \varepsilon &= \frac{1}{2}(\nabla u + \nabla u^\top) && \text{in } \Omega \\ K(\rho) : \varepsilon \mathbf{n} &= T && \text{on } \Gamma_t \\ u &= 0 && \text{on } \Gamma_d \end{aligned}$$

Numerical Results

Formulation: SIMP power $p = 3$ with Helmholtz filtering (radius$= 0.1$).

Discretization: Q1 FEM for displacement variables and piecewise constant for density.

Problem Size: 26,880 density degrees of freedom.

method	iter	fval	grad	hess	proj	time(s)
TRSPG	9	10	10	236	1200	16.49
LMTR	33	34	31	418	391	32.42
PQN	126	235	127	---	4972	164.49
SPG	84	90	85	---	170	52.36
AL-TRSPG	9	52	51	1153	---	61.98
AL-LMTR	11	276	263	4368	---	280.77

Numerical Results

We now consider the 3D domain $\Omega = (0,2) \times (0,1) \times (0,1)$.

Problem Size: 221,184 density degrees of freedom.

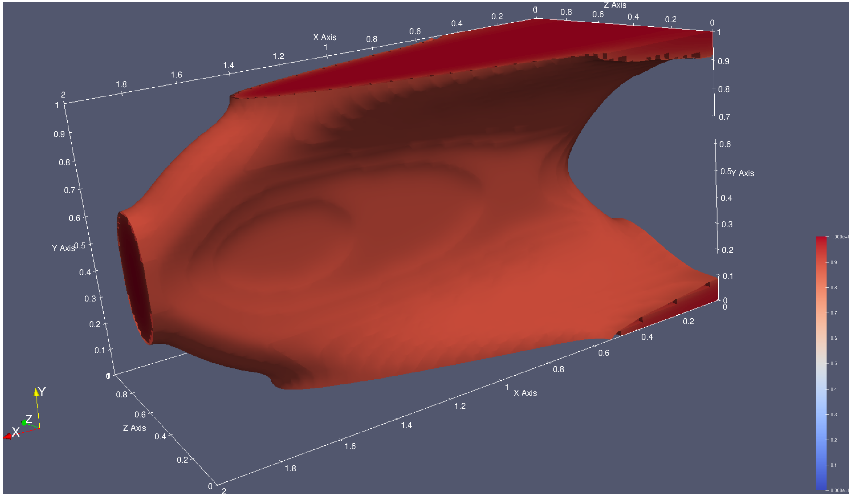
Inexact Solves: Solve using CG with AMG preconditioning.

- **Helmholtz Filter:** Requires ∼8 iterations to achieve the relative error of $\sim 10^{-12}$ — Considered to be **exact**.
- **Elasticity Equations:** Trust-region algorithm controls accuracy of linear solver.

k	$f(x_k)$	$\|d_k(1)\|$	$\|x_k - x_{k-1}\|$	Δ_k	fval	grad	hess	proj	obj tol	grad tol
0	1.0000	4.017e-2	---	20	1	1	0	4	1.000e-2	1.000e-2
1	0.7156	1.771e-2	2.000e1	50	2	2	28	96	1.000e-2	1.000e-2
2	0.4393	6.788e-3	5.000e1	50	3	3	55	204	1.000e-2	1.000e-2
3	0.3168	2.853e-3	5.000e1	125	4	4	82	405	1.000e-2	1.000e-2
4	0.1654	8.805e-4	1.250e2	125	5	5	109	639	1.000e-2	8.802e-3
5	0.1255	2.066e-5	1.250e2	125	6	6	143	707	1.000e-2	2.066e-4
6	0.1247	2.713e-6	6.272e1	312.5	7	7	171	765	1.461e-4	2.713e-5

Recall: $d_k(t) := \mathbf{P}_{\mathcal{C}}(x_k - t g_k) - x_k$

Filtered Density: 0.9 Countour

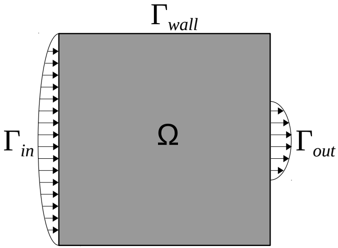


Numerical Results

Let $\Omega = (0,1)^2$ and $v = 0.4$, and consider

$$\min_{\rho\in L^2(\Omega)} \int_\Omega \{\nabla S_u(\rho)\cdot\nabla S_u(\rho) + \alpha(\rho)S_u(\rho)\cdot S_u(\rho)\}\,\mathrm{d}x$$

$$\text{subject to} \quad \int_\Omega \rho(x)\,\mathrm{d}x = v|\Omega|, \quad 0\le\rho\le 1 \ \text{ a.e.},$$



where $(S_u(\rho), S_p(\rho)) = (u,p) \in (H^1(\Omega))^2 \times L^2(\Omega)$ solves

$$\begin{aligned} -\nu\Delta u + u\cdot\nabla u + \nabla p &= -\alpha(\rho)u && \text{in } \Omega \\ \nabla\cdot u &= 0 && \text{in } \Omega \\ u &= u_{\text{in}} && \text{on } \Gamma_{\text{in}} \\ u &= 0 && \text{on } \Gamma_{\text{wall}} \\ \nu\nabla u\cdot\mathbf{n} - p\mathbf{n} &= 0 && \text{on } \Gamma_{\text{out}} \end{aligned}$$

Numerical Results

Formulation: RAMP material model with $\overline{\alpha} = 2.5 \times 10^4$, $\underline{\alpha} = 2.5 \times 10^{-4}$ and $q = 0.1$, i.e.,

$$\alpha(\rho) = \overline{\alpha} + (\underline{\alpha} - \overline{\alpha}) \frac{\rho(1+q)}{q+\rho}.$$

Discretization: Q2–Q1 FEM for state variables and piecewise constant for density.

Problem Size: 30,720 density degrees of freedom.

method	iter	fval	grad	hess	proj	time(s)
TRSPG	11	12	12	306	1600	724.76
LMTR	22	23	20	709	296	1487.11
PQN	84	85	85	---	3076	1884.58
SPG	139	353	140	---	280	4799.63
AL-TRSPG	5	22	22	437	---	1021.85
AL-LMTR	4	33	33	1151	---	2302.32

Conclusions:

- **Numerical solution** of infinite-dimensional problems requires **expensive approximations**
- Often, the objective function and its gradient can only be computed **inexactly**
- Convex-constrained trust region is **provably convergent** even with **inexact computations**
- **We can efficiently compute a trial step using the spectral projected gradient method**
- SPG trust-region subproblem solver is **matrix free**, but may **require** many projections onto $\mathcal{C}$
  Can we incorporate inexact projections (à la Garreis, Ulbrich, Birgin, et al.)?
- SPG trust-region method **outperforms** existing derivative-based methods:
  - Outperforms `SPG` and `PQN` because it uses second-order information
  - Outperforms `LMTR` because CG can terminate early, producing gradient-like steps
  - Outperforms `AL-TRSPG/AL-LMTR` methods by avoiding penalty function iterations

References:

- D. P. Kouri, A matrix-free trust-region Newton algorithm for convex-constrained optimization, Optimization Online, 2021.
- D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty, SISC, 2014.