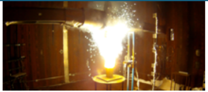
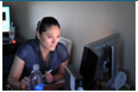
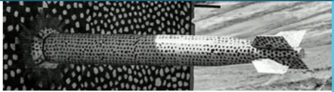
Sandia National Laboratories

ALESQP: An Augmented Lagrangian Equality-constrained SQP Method for Optimization with General Constraints







Harbir Antil, Drew Kouri and **Denis Ridzal**

Sayas Numerics Seminar

UMD, UMBC, GMU, UD

November 3, 2020

AFOSR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

LDRD
Laboratory Directed Research and Development

U.S. DEPARTMENT OF ENERGY

NNSA

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

SAND2020-11366 C

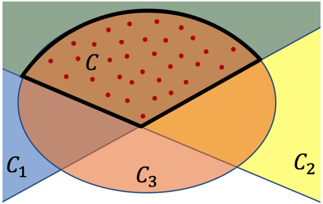
Problem Statement

Develop efficient algorithms to solve the **general optimization problem (OP)**,

$$\min_{x \in X} \; f(x) \qquad \text{subject to} \qquad g(x) = 0, \quad Tx \in C.$$

Notation, problem data and assumptions:

- $f : X \to \mathbb{R}$ and $g : X \to Y$ are continuously differentiable.
- f and $\|g(\cdot)\|_Y$ are weakly lower semicontinuous.
- We let $C = C_1 \cap \cdots \cap C_m \neq \emptyset$ with $C_i \subseteq Z$ nonempty, closed and convex sets.
- X, Y are real Banach spaces and Z is a real Hilbert space.
- $T \in \mathcal{L}(X, Z)$ and T^* is injective.



Set C as the intersection of two closed half-spaces, C_1 and C_2, and the closed ball C_3.

Additionally: The *projection* of $x \in X$ onto C is $\mathbf{P}_C(x)$, and the *distance* from x to C is $d_C(x)$, with

$$d_C(x) := \min_{y \in C} \|x - y\|_X = \|x - \mathbf{P}_C(x)\|_X.$$

The sets $\{C_1, \ldots, C_m\}$ are boundedly regular, i.e., for every bounded sequence $\{x_k\} \subset X$

$$\max_{i=1,\ldots,m} d_{C_i}(Tx_k) \to 0 \qquad \Longrightarrow \qquad d_C(Tx_k) \to 0.$$

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u,z} \; f(u,z) \qquad \text{subject to} \qquad g(u,z) = 0, \quad T_1 u \in C_1, \quad T_2 z \in C_2.$$

- Given a control z, the PDE $g(u,z) = 0$ is expensive to solve for the state $u = u(z)$.

  Keep PDE as constraint and solve it gradually using, e.g., trust-region SQP (no nonlinear solves).

- Linear (KKT-like) systems in SQP take advantage of iterative solvers and good preconditioners.

  Use matrix-free SQP that efficiently handles inexact linear system solves; also mesh adaptivity.[†,††]

- Catch: SQP with inexact linear system solves cannot directly handle general inequality constraints.

  Penalize $Tx \in C$ explicitly using augmented Lagrangian.

- Control and state constraint multipliers typically have different regularity properties, e.g., L^2 for control multipliers and measures for state multipliers, resulting in vastly different scales.

  Use separate penalties and multiplier estimates for control and state constraints.

- Many nonlinear programming methods exhibit mesh dependence, i.e., iterations grow with problem size.

  Prove convergence in infinite-dimensional Banach/Hilbert spaces.

[†] **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.
[††] **Ziems, Ulbrich (2011)**, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIOPT.

Example 1

State and control constrained semilinear PDE-constrained optimization

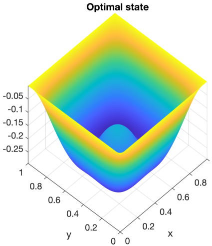
$$\min_{u \in H_0^1(\Omega),\, z \in L^2(\Omega)} \left\{ \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)} \right\}$$

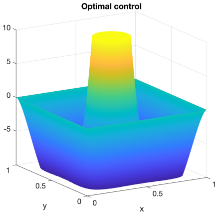
subject to

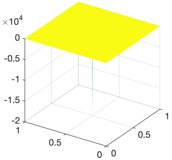
$$\begin{aligned} -\Delta u + u^3 &= z \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \\ u_a &\le u \quad \text{a.e. in } \Omega \\ z_a \le z &\le z_b \quad \text{a.e. in } \Omega, \end{aligned}$$

with

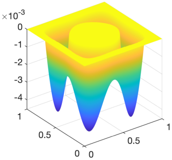
$$\begin{aligned} u_d &= -1, \quad \alpha = 10^{-3}, \quad z_a = -10, \quad z_b = 10, \\ u_a(x) &= -\tfrac{2}{3} + \tfrac{1}{2}\min\Big\{x_1 + x_2, \\ &\qquad \min\{1 + x_1 - x_2, \min\{1 - x_1 + x_2, 2 - x_1 - x_2\}\}\Big\}. \end{aligned}$$



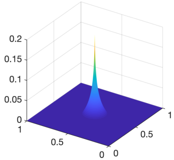




u_a multiplier



z_a multiplier



z_b multiplier

Börgens, Kanzow, Steck (2019), *Local and global analysis of multiplier methods for constrained optimization in Banach spaces*, SICON.

Example 2

Rényi entropy maximization with constraints

Applicable to other ∞-dim. problems, nonlinear T:

$$\max_{\rho \in L^p(\Omega)} \frac{1}{1-p} \log \left(\int_\Omega \rho(x)^p \, \mathrm{d}x \right)$$

subject to $\quad \rho \geq 0$ a.e.

$$\int_\Omega \rho(x) \, \mathrm{d}x = 1$$

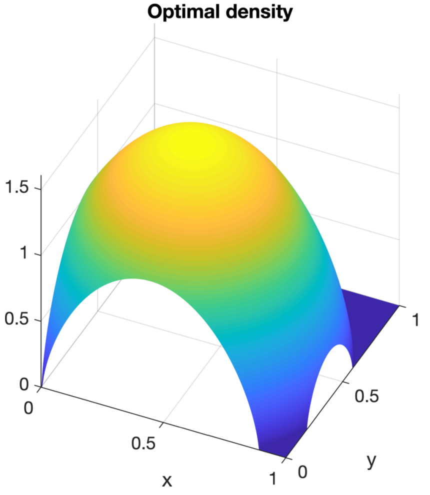
$$\int_\Omega \rho(x) x \, \mathrm{d}x = \mu$$

$$\det \left(\int_\Omega \rho(x)(x-\mu)(x-\mu)^\top \, \mathrm{d}x \right) \leq \sigma,$$

where $\mu = (0.45, 0.45)$ and

$$\sigma = \frac{1}{2} \det \left(\int_\Omega (x-\mu)(x-\mu)^\top \, \mathrm{d}x \right) \approx 0.00368.$$

Objective is p-order Rényi entropy, with $p = 2.5$.



Van Erven, Harremos (2014), *Rényi divergence and Kullback-Leibler divergence*, IEEE Trans. Inf. Theory.

Derivation of the Augmented Lagrangian

- **Exact penalty**: $I_C(y) = 0$ if $y \in C$ and $I_C(y) = \infty$ if $y \notin C$. **(OP)** is equivalent to:

$$\min_{x \in X} \{f(x) + I_C(Tx)\} \quad \text{subject to} \quad g(x) = 0.$$

- Specifically, $I_C(Tx) = I_{C_1}(Tx) + \ldots + I_{C_m}(Tx)$. It is **relaxed** **[Rockafellar, 1976]** using

$$\Psi_i(x, \lambda, r) := \sup_{\mu \in Z} \{(\mu, Tx)_Z - I^*_{C_i}(\mu) - \tfrac{1}{2r}\|\mu - \lambda\|^2_Z\},$$

  where $r > 0$, $i = 1, \ldots, m$, and $I^*_{C_i}(\mu) := \sup_{z \in C_i} (\mu, z)_Z$ is the Fenchel conjugate of I_{C_i}.

- Maximization problem is strongly concave and has the unique solution

$$\Lambda_i(x, \lambda, r) := r((r^{-1}\lambda + Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda + Tx)).$$

  This is the **usual augmented Lagrangian multiplier update**.

- Maintain **separate** multiplier estimates λ_i and penalty parameters r_i for each constraint $Tx \in C_i$ to account for **differing constraint scales**. Our **augmented Lagrangian** is

$$L(x, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m) := f(x) + \sum_{i=1}^{m} \tfrac{1}{2r_i}\|\Lambda_i(x, \lambda_i, r_i)\|^2_Z - \tfrac{1}{2r_i}\|\lambda_i\|^2_Z.$$

Examples of C and Multiplier Updates

- **Norm constraints:** Let $\rho > 0$ and $z_0 \in Z$. We consider the set

$$C = \{z \in Z \,|\, \|z - z_0\|_Z \leq \rho\} = z_0 + B_\rho^Z.$$

In this case,

$$\Lambda(x, \lambda, r) = \left(1 - \min\left\{1, \frac{r\rho}{\|\lambda + r(Tx - z_0)\|_Z}\right\}\right)(\lambda + r(Tx - z_0)).$$

- **Conical constraints:** Let $K \subset Z$ be pointed, closed, convex cone. Consider,

$$C = \{z \in Z \,|\, z - \ell \in K\} = \ell + K,$$

then

$$\Lambda(x, \lambda, r) = -(\lambda + r(Tx - \ell))_-.$$

- **Finitely many linear constraints:** Let $\{a_1, \ldots, a_n\} \subset X^* \setminus \{0\}$ be linearly independent, $\{b_1, \ldots, b_n\} \subset \mathbb{R}^n$ and $C = \{z \in \mathbb{R}^n \,|\, z_i \geq b_i\}$. Moreover, $(Tx)_i = \langle a_i, x\rangle_{X^*,X}$ for $i = 1, \ldots, n$. Then

$$\Lambda(x, \lambda, r)_i = \min\{0, \lambda_i + r(\langle a_i, x\rangle_{X^*,X} - b_i))\}, \quad i = 1, \ldots, m.$$

Properties of the Augmented Lagrangian

- For fixed $\lambda_i \in Z$ and $r_i > 0$, $i = 1, \ldots, m$, $L(\cdot, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m)$ is
  - **(i)** weakly lower semicontinuous if f is;
  - **(ii)** convex if f is; and
  - **(iii)** continuously Fréchet differentiable if f is.

  Moreover, in case (iii), the derivative of $L(\cdot, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m)$ is given by

$$L_x(x, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m) = f'(x) + \sum_{i=1}^{m} T^* \Lambda_i(x, \lambda_i, r_i),$$

  and if f' is **Lipschitz continuous,** $f \in C^{1,1}$, so is $L_x(\cdot, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m)$, $L \in C^{1,1}$.

- **Subproblem (SUB):** Find approximate stationary point of

$$\min_{x \in X} \; L(x, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m) \quad \text{subject to} \quad g(x) = 0.$$

  The solvers of **Heinkenschloss, Ridzal (2014)** and **Ziems, Ulbrich (2011)** *almost* apply in Hilbert space; they technically require $L \in C^2$, and we have $L \in C^{1,1}$.

ALESQP

Iteration k of **A**ugmented **L**agrangian with **E**quality-constrained **SQP**

Use ESQP to solve **(SUB)** and compute $(x^{(k)}, \zeta^{(k)}) \in X \times Y^*$ that satisfies

$$\|g(x^{(k)})\|_Y \le \delta^{(k)} \quad \text{and} \quad \|f'(x^{(k)}) + \textstyle\sum_i T^*\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}\|_{X^*} \le \varepsilon^{(k)}.$$

for $i = 1, \ldots, m$ **do**

  if $\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\|_Z > r_i^{(k)}\tau_i^{(k)}$ **then** // penalty parameter updates

    $r_i^{(k+1)} = \eta_i r_i^{(k)}$

    $\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)}, \theta_i\}$

    $\tau_i^{(k+1)} = \tau_i^{(0)}(\theta_i^{(k+1)})^{\alpha_i}$ // geometric decay

  else

    $r_i^{(k+1)} = r_i^{(k)}$

    $\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)}, \theta_i\}$

    $\tau_i^{(k+1)} = \tau_i^{(k)}(\theta_i^{(k+1)})^{\beta_i}$ // exponential decay

  end if

  if $\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)})\|_Z \le \nu_i(r_i^{(k+1)})^{\gamma_i}$ **then** // multiplier updates

    $\lambda_i^{(k+1)} = \Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)})$

  else

    $\lambda_i^{(k+1)} = \lambda_i^{(k)}$

  end if

end for // single update if $k > 1000$

Comparisons with Existing Approaches

- ALESQP incorporates **multiple penalty functions** to handle varying constraint scalings.
- ALESQP is related to **LANCELOT**† with two additional differences:
  - ALESQP can handle both finite and **infinite dimensional** spaces.
  - ALESQP keeps the **equality constraints explicit** and penalizes the constraints $Tx \in C$; this is related to the finite-dimensional **SECO**†† subproblem formulation.
- Existing infinite-dimensional augmented Lagrangians: **Börgens, Kanzow, Steck (2019)**. There are three fundamental differences:
  - **Different subproblems**: ALESQP treats the equality constraint (e.g., PDE) **explicitly** versus implicitly through an exact nonlinear solve. Subproblems are *only* **equality-constrained**.
  - **Inexact solves**: ALESQP enables **large-scale iterative linear solves and mesh adaptivity**.
  - **Convergence analysis**: ALESQP theory assumptions are weaker and the results are stronger.
- In a nutshell, ALESQP is an infinite-dimensional extension of the SECO formulation with the inner workings of LANCELOT + multiple penalty functions + inexact linear solves.

†**Conn, Gould, Toint (1991)**, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SINUM.
††**Birgin, Bueno, Martínez (2016)**, *Sequential equality-constrained optimization for NLP*, Comp. Optim. Appl.

Convergence Theory

- Let $\{\lambda_i^{(k)}\}$ be an infinite sequence of multipliers for the i^{th} constraint generated by ALESQP, and let $\gamma_i < 1/2$. If $r_i^{(k)} \to \infty$, then

$$\lim_{k\to\infty} \frac{1}{(r_i^{(k)})^\alpha} \|\lambda_i^{(k)}\|_Z = 0 \quad \forall\, \alpha > \gamma_i.$$

- Moreover, for the corresponding sequence of iterates $\{x^{(k)}\}$, the following are equivalent:

  (i) $\liminf\limits_{k\to\infty} d_{C_i}(Tx^{(k)}) = 0$

  (ii) $\liminf\limits_{k\to\infty} \frac{1}{r_i^{(k)}} \|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\|_Z = 0$

  (iii) $\liminf\limits_{k\to\infty} \frac{1}{r_i^{(k)}} \|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)})\|_Z = 0.$

- Moreover,
  - if the Lagrange multipliers λ_i are updated finitely many times, **or**
  - if the penalty parameters r_j are updated finitely many times for all $j = 1, \ldots, m$, **then**
  - the sequence of multipliers $\{\lambda_i^{(k)}\}$ converges strongly to some $\bar{\lambda}_i \in Z$.

Stopping Conditions: Let $\delta_* > 0$, $\varepsilon_* > 0$ and $\tau_* > 0$. ALESQP exits if

$$\begin{aligned} \|g(x^{(k)})\|_Y &\le \delta_* \\ \|L'(x^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}\|_{X^*} &\le \varepsilon_* \\ \max_{i=1,\dots,m} d_{C_i}(Tx^{(k)}) &\le \tau_* . \end{aligned}$$

- ALESQP either **exits after a finite number of iterations**; or produces a sequence that satisfies the **asymptotic stationarity conditions**:

$$\|g(x^{(k)})\|_Y \to 0$$

and

$$\limsup_{k\to\infty} \langle -(f'(x^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}), y - x^{(k)} \rangle_{X^*,X} \le 0 \quad \forall\, y \in T^{-1}(C).$$

Convergence Theory

- **Strong accumulation points of $\{x^{(k)}\}$ are stationary under mild assumptions.**
- **Finite Dimensions: ALESQP converges to a stationary point.**
- **Infinite Dimensions:** In general, **weak accumulation points** $\bar{x}$ of $\{x^{(k)}\}$ satisfy: $\exists\, \bar{t}_i > 0$ with $\bar{t}_1 + \ldots + \bar{t}_m = 1$ such that

$$T\bar{x} \in \bar{t}_1 C_1 + \ldots + \bar{t}_m C_m,$$

  but need not be **feasible**!
- **Weak accumulation points of $\{x^{(k)}\}$ are <u>feasible for many practical situations</u>:**
  - $m = 1$, **or**
  - T completely continuous, **or**
  - $x^{(k_j)}$ converges strongly to $\bar{x}$, **or**
  - $\exists X_0$ that is compactly embedded in X with $\{x^{(k_j)}\} \subset X_0$ and $x^{(k_j)} \rightharpoonup \bar{x}$ in X_0.
- Holds for finite-dimensional X and many PDECO problems with control/state constraints.
- Require strong assumptions on g to show that weak accumulation points are stationary.
- **Can extend theory to nonlinear T to handle, e.g., complementarity constraints.**

Subproblem Solver: Trust-region SQP

Start with the equality-constrained optimization problem:

$$\min_{x \in X} \; L(x) \quad \text{subject to} \quad g(x) = 0$$

where $L : X \to \mathbb{R}$ and $g : X \to Y$, for some **Hilbert spaces** X and Y, and where L is **Lipschitz continuously** and g is **twice continuously** Fréchet differentiable.

Define SQP Lagrangian functional $\mathscr{L} : X \times Y^* \to \mathbb{R}$:

$$\mathscr{L}(x, \zeta) = L(x) + \langle \zeta, g(x) \rangle_{Y^*, Y}$$

At j^{th} SQP iteration solve *nonconvex* quadratic trust-region subproblem:

$$\begin{aligned} \min_{s \in X} \quad & \frac{1}{2} \langle \nabla_{xx} \mathscr{L}(x_j, \zeta_j) s, s \rangle_{X^*, X} + \langle \nabla_x \mathscr{L}(x_j, \zeta_j), s \rangle_{X^*, X} + \mathscr{L}(x_j, \zeta_j) \\ \text{s.t.} \quad & g'(x_j) s + g(x_j) = 0 \,, \quad \|s\|_X \le \Delta_j \end{aligned}$$

Extended the Byrd-Omojokun composite-step approach to support inexact linear solvers.

Composite-step Method for Trust-region Subproblem

- **Trust-region step**: $s_j = n_j + t_j$

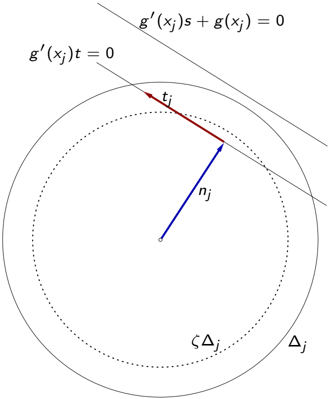
- **Quasi-normal step** n_j:

  reduces linear infeasibility

$$\begin{array}{ll} \min\limits_{n\in\mathcal{X}} & \|g'(x_j)n + g(x_j)\|_Y^2 \\ \text{s.t.} & \|n\|_X \le \zeta\Delta_j \end{array}$$

- **Tangential step** t_j:

  improves optimality while staying in the null space of the linearized constraints



$$\begin{array}{ll} \min\limits_{t\in X} & \frac{1}{2}\langle \nabla_{xx}\mathscr{L}(x_j,\zeta_j)(t+n_j), t+n_j\rangle_{X^*,X} + \langle \nabla_x\mathscr{L}(x_j,\zeta_j), t+n_j\rangle_{X^*,X} + \mathscr{L}(x_j,\zeta_j) \\ \text{s.t.} & g'(x_j)t = 0\,, \quad \|t+n_j\|_X \le \Delta_j \end{array}$$

Omojokun (1989), Byrd, Hribar, Nocedal (1997), Dennis, El-Alem, Maciel (1997)

Matrix-free Trust-region SQP Algorithm

1. Compute **quasi-normal step** n_j using **Powell dogleg**, where for the Newton step we solve an augmented system inexactly.

2. Solve **tangential subproblem** for $\widetilde{t}_j$ via **projected Steihaug-Toint CG**, where the projections are computed by solving augmented systems inexactly.

3. Restore linearized feasibility, for **tangential step** t_j, via another inexact projection.

4. Update **Lagrange multipliers** ζ_{j+1} by solving an augmented system inexactly.

5. Evaluate progress.

Heinkenschloss, Ridzal (2014)

Augmented System

$$\begin{pmatrix} I_{X,X^*} & g'(x_j)^* \\ g'(x_j) & 0 \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} + \begin{pmatrix} e^1 \\ e^2 \end{pmatrix}$$

- The size of $(e^1 \;\; e^2)$ is governed by the progress of the optimization algorithm:

$$\|e^1\|_{X^*} + \|e^2\|_Y \leq \mathscr{T}\left(\|b^1\|_{X^*}, \|b^2\|_Y, \|y^1\|_X, \Delta_j\right)$$

- In optimal control, augmented systems are:

$$\begin{pmatrix} I_{X,X^*} & g'^* \\ g' & 0 \end{pmatrix} \rightarrow \begin{pmatrix} I_{X_u,X_u^*} & & g_u^* \\ & I_{X_z,X_z^*} & g_z^* \\ g_u & g_z & 0 \end{pmatrix}$$

where u are the states and z the controls.

Projected CG Preconditioner

- Motivated by a comment on the "perfect preconditioner" for the projected CG method for optimization, from **Gould, Hribar, Nocedal (2002)**.[†]
- Idea: Replace I_{X,X^*} with **Augmented Lagrangian** derivatives:

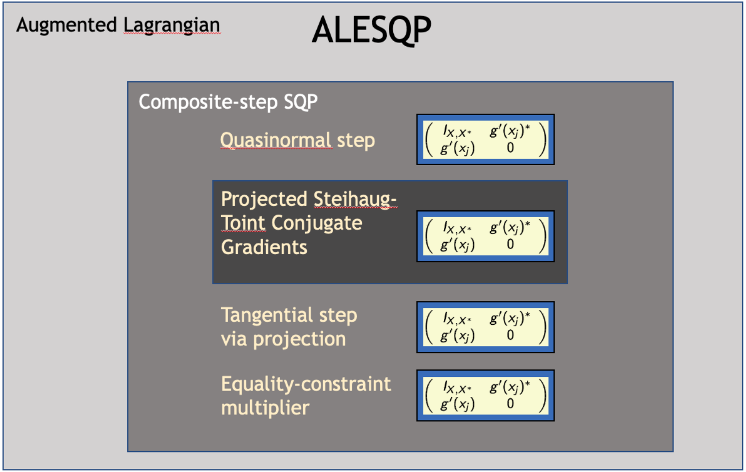
$$\begin{pmatrix} B(x_j) + T^*\left(\sum_{i=1}^m r_i^{(k)}\left(I_{Z,Z^*} - D_{ij}\right)\right)T & g'(x_j)^* \\ g'(x_j) & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ 0 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix},$$

  where $B(x_j) \in \mathcal{L}(X, X^*)$ is a nonnegative operator approximating $f''(x_j)$ and D_{ij} denotes the **Newton derivative**, see **Chen, Nashed, Qi (2000)**[††], of $\mathbf{P}_{C_i}((r_i^{(k)})^{-1}\lambda_i^{(k)} + Tx_j)$.
- This choice can accelerate projected CG significantly.
- At the same time, this system is more difficult to solve than a corresponding augmented system, and challenging to integrate into projected CG when using iterative linear solvers.

[†] **Gould, Hribar, Nocedal (2002)**, *On the solution of equality constrained quadratic programs*, SISC.

[††] **Chen, Nashed, Qi (2000)**, *Smoothing/semismooth methods for nondifferentiable operator equations*, SINUM.

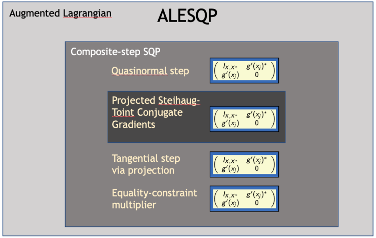
ALESQP Components



Note: Each box is an iterative process!

Numerical experiments

- Consider max entropy and semilinear PDECO problem.
- Each constraint has its own starting penalty parameter.
- Discretize infinite-dimensional objects using triangular finite elements on regular grids.
- Use diagonal **Riesz maps** in every ALESQP component.
- Consider direct (exact) and iterative (inexact) solves.



Iterative augmented system solves:

- We use MINRES with constant SPD preconditioners.
- Unpreconditioned for the max entropy problem.
- Schur complement preconditioner† for semilinear problem. Linearized forward and adjoint systems are "solved" using a few V-cycles of algebraic multigrid.

Schur Complement

$$\begin{bmatrix} I_{X_u^*,X_u} & & 0 \\ & I_{X_z^*,X_z} & 0 \\ 0 & 0 & g_u^{-*}\, I_{X_u,X_u^*}\, g_u^{-1} \end{bmatrix}$$

†**Rees, Dollar, Wathen (2010)**, *Optimal solvers for PDE-constrained optimization*, SISC.

Rényi Entropy Maximization with Constraints

$$\max_{\rho \in L^p(\Omega)} \; -\frac{2}{3} \log \left(\int_\Omega \rho(x)^{2.5} \, \mathrm{d}x \right)$$

subject to $\rho \geq 0$ a.e.

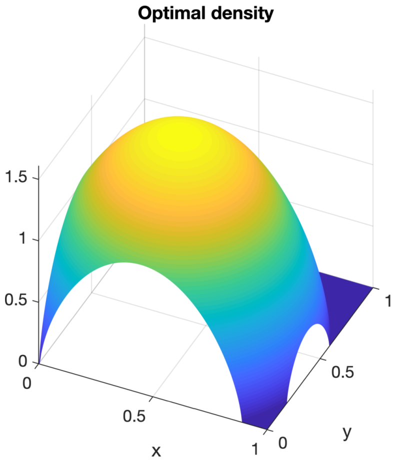
$$\int_\Omega \rho(x) \, \mathrm{d}x = 1$$

$$\int_\Omega \rho(x) x \, \mathrm{d}x = \mu$$

$$\det \left(\int_\Omega \rho(x)(x-\mu)(x-\mu)^\top \, \mathrm{d}x \right) \leq \sigma,$$

where $\mu = (0.45, 0.45)$ and

$$\sigma = \frac{1}{2} \det \left(\int_\Omega (x-\mu)(x-\mu)^\top \, \mathrm{d}x \right) \approx 0.00368.$$



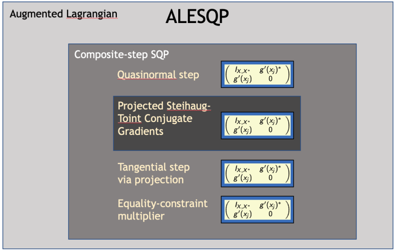
Up to 263,169 optimization variables, 3 equality constraints and 263,170 inequality constraints.

Maximum Rényi Entropy

Direct, i.e., exact linear solves

Mesh	AL	SQP	CG	normg	grad-lag	feas
64x64	11	56	249	1.11e-16	1.77e-08	1.42e-08
128x128	11	54	242	1.36e-16	3.58e-08	2.09e-07
256x256	12	67	342	3.55e-16	2.11e-08	2.27e-08
512x512	12	65	312	8.85e-16	1.15e-08	7.46e-09

- ALESQP performance for varying discretization (`Mesh`).
- `AL` is the total number of augmented Lagrangian iterations.
- `SQP` is the total number of SQP iterations.
- `CG` is the total number of conjugate gradient iterations.
- `normg` is the equality constraint violation $\|g(x^{(k)})\|_Y$.
- `grad-lag` is the gradient norm for subproblem Lagrangian, $\|L_k'(x^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}\|_{X^*}$.
- `feas` is the constraint violation $\max_i d_{C_i}(Tx^{(k)})$.



We observe that the AL, SQP, and CG iterations are nearly mesh independent.

Maximum Rényi Entropy

Mesh	AL	SQP	CG	normg	grad-lag	feas	avg.aug
64x64	11	57	261	2.85e-14	7.71e-08	1.49e-07	7.3
128x128	13	52	249	3.72e-16	2.77e-08	2.92e-09	7.7
256x256	11	45	177	3.60e-16	2.88e-08	9.72e-09	7.9
512x512	13	55	248	1.77e-15	2.94e-08	2.75e-08	8.2

- avg.aug are the average numbers of MINRES iterations per call.
- Again, the AL, SQP, and CG iterations are nearly mesh independent.
- avg.aug are nearly mesh independent.
- avg.aug are very small without preconditioning.
- feas is as small as 2.92e-09.
- Nominal linear system solver tolerances ranged from 1e-4 to 1e-1.

Maximum Rényi Entropy

Constraint feasibility with inexact solves

tol	AL	SQP	CG	normg	grad-lag	feas	avg.aug
1e-6	13	52	249	3.72e-16	2.77e-08	2.92e-09	7.7
1e-8	16	55	274	2.99e-16	3.15e-10	5.15e-11	8.0
1e-10	20	59	305	2.22e-16	3.90e-12	3.47e-13	8.4
1e-12	26	64	345	1.67e-16	2.38e-14	3.47e-15	9.1

- Fix the mesh resolution to 128x128.
- Vary augmented Lagrangian nominal tolerance tol.
- SQP iterations increase by 20% and CG iterations by 40% **to add 6 digits of accuracy**!
- avg.aug are tolerance independent.
- Smallest feas is 3.47e-15.
- Nominal linear system solver tolerances ranged from 1e-4 to 1e-1.

Optimization of Semilinear PDE with Constraints

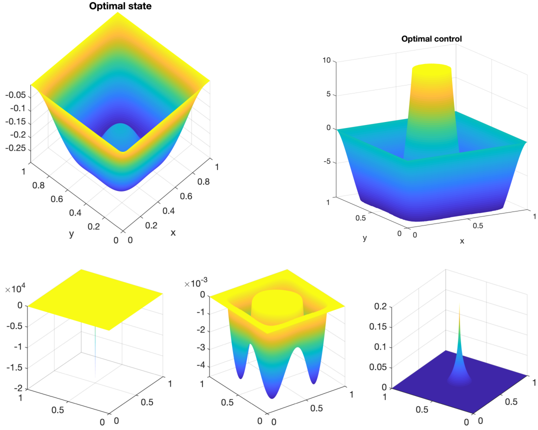
$$\min_{u\in H_0^1(\Omega),\, z\in L^2(\Omega)} \left\{ \frac{1}{2}\|u-u_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)} \right\}$$

subject to

$$\begin{aligned} -\Delta u + u^3 &= z \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \\ u_a &\le u \quad \text{a.e. in } \Omega \\ z_a \le z &\le z_b \quad \text{a.e. in } \Omega, \end{aligned}$$

with

$$\begin{aligned} u_d &= -1, \quad \alpha = 10^{-3}, \quad z_a = -10, \quad z_b = 10, \\ u_a(x) &= -\tfrac{2}{3} + \tfrac{1}{2}\min\Big\{x_1 + x_2, \\ &\quad \min\{1 + x_1 - x_2, \min\{1 - x_1 + x_2, 2 - x_1 - x_2\}\}\Big\}. \end{aligned}$$



u_a multiplier $\quad$ z_a multiplier $\quad$ z_b multiplier

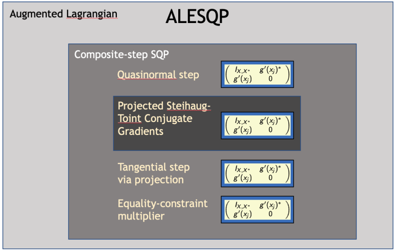
Up to 2,000,600 optimization variables, 970,299 equality constraints and 3,030,901 inequality constraints.

Optimization of Semilinear PDE

Direct, i.e., exact linear solves

Mesh	Control			State			Control + State		
	AL	SQP	CG	AL	SQP	CG	AL	SQP	CG
64x64	11	21	55	15	29	73	18	40	95
128x128	11	21	57	17	37	97	18	45	110
256x256	15	24	65	18	39	105	20	47	119
512x512	15	26	68	20	44	122	22	50	131

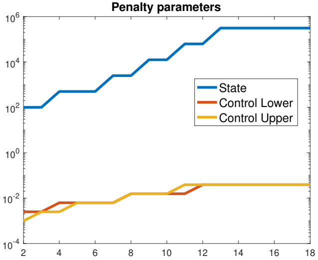
- ALESQP performance for varying **2D mesh** (`Mesh`).
- `Control`: only control constraints (use $z_b = -1$);
  `State`: only state constraints;
  `Control + State`: both control and state constraints.
- `AL` is the total number of augmented Lagrangian iterations.
- `SQP` is the total number of SQP iterations.
- `CG` is the total number of conjugate gradient iterations.



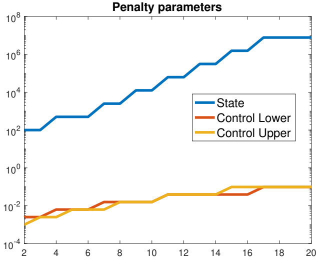
We observe that the AL, SQP, and CG iterations are nearly mesh independent.

Optimization of Semilinear PDE

64 × 64 mesh



256 × 256 mesh



- State-constraint penalties increase significantly.
- Control-constraint penalties do not increase much.
- **Underlines the need for multiple penalties!**
- Very mild mesh dependence in the state-constraint penalty parameter sequence.

Optimization of Semilinear PDE

Mesh	AL	SQP	CG	normg	grad-lag	feas	avg.aug	avg.augCG
24x24x24	32	38	74	3.21e-13	1.39e-10	5.70e-07	6.5	16.1
40x40x40	32	48	98	1.89e-10	2.97e-08	4.71e-07	7.3	30.8
64x64x64	29	50	128	1.04e-10	5.11e-08	6.05e-08	9.3	44.4
100x100x100	31	55	125	1.43e-10	8.58e-08	3.47e-07	11.7	54.4

- **3D version of the** `Control + State` **problem** with hexahedral mesh and $u_a = -\frac{1}{2}$.
- `avg.augCG` are the average numbers of MINRES iterations per call in CG.
  Recall: In CG we change the augmented system to account for AL penalty terms.
- `avg.aug` are the average numbers of MINRES iterations per call elsewhere.
- Again, the AL, SQP, and CG iterations are nearly mesh independent.
- `avg.aug` and `avg.augCG` are somewhat mesh dependent.
- `feas` is as small as `6.05e-08`.
- Nominal linear system solver tolerances ranged from `1e-6` to `1e-2`.

Optimization of Semilinear PDE

tol	AL	SQP	CG	normg	grad-lag	feas	avg.aug	avg.augCG
1e-6	32	48	98	1.89e-10	2.97e-08	4.71e-07	7.3	30.8
1e-8	49	55	112	3.21e-13	8.15e-11	1.20e-09	7.6	46.4
1e-10	63	61	124	1.82e-15	3.26e-13	5.30e-11	7.9	87.4
1e-12	80	66	134	1.82e-15	8.40e-14	2.76e-13	8.3	103.8

- Fix the mesh resolution to `40x40x40`.
- Vary augmented Lagrangian nominal tolerance `tol`.
- SQP and CG iterations increase by 35% **to add 6 digits of accuracy**!
- Very mild increase in `avg.aug` and moderate increase in `avg.augCG`.
- Smallest `feas` is `2.76e-13`.
- Nominal linear system solver tolerances ranged from `1e-6` to `1e-2`.

Summary

- ALESQP is well suited for infinite-dimensional optimization with general constraints.
- Augmented Lagrangian that penalizes inequalities, with equality-constrained subproblems.
- Subproblem solver: Matrix-free trust-region SQP method with iterative linear solvers.
- ALESQP uses a constraint decomposition with multiple penalties.
- In optimal control, ALESQP provides a unified framework to efficiently handle general constraints on both the state variables and the control variables.
- Exhibits fast convergence and remarkable accuracy, even with inexact linear solves.
- **ALESQP uses the inner SQP loop economically!**
- Two papers:
  - **ALESQP**: Submitted.
  - **Banach-space SQP, specialized for use with ALESQP**: In progress.