Retrospective Approximation Sequential Quadratic Programming for Stochastic Optimization with General Deterministic Nonlinear Constraints

Shagun Gupta,
University of Texas at Austin

Collaborators

Prof. Raghu Bollapragada,
UT Austin

Prof. Albert S. Berahas,
University of Michigan

Constrained Stochastic Optimization

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \; & f(x) = \mathbb{E}\left[F(x,\xi)\right] \\ s.t. \; & c_E(x) = 0 \\ & c_I(x) \leq 0\end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function
- $c_E : \mathbb{R}^n \to \mathbb{R}^{m_E}$ are general nonlinear equality constraints
- $c_I : \mathbb{R}^n \to \mathbb{R}^{m_I}$ are general nonlinear inequality constraints

Constrained Optimization

- **Implicit Methods**: Penalty Methods, Projection Methods, ...

$$\min_{x \in \mathbb{R}^n} f(x) \quad s.t. \; c_E(x) = 0, \; c_I(x) \leq 0$$

→ Unconstrained Surrogate → Local Model → Search Direction

Constrained Optimization

- **Implicit Methods**: Penalty Methods, Projection Methods, ...

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ s.t. \ & c_E(x) = 0 \\ & c_I(x) \leq 0 \end{aligned}$$

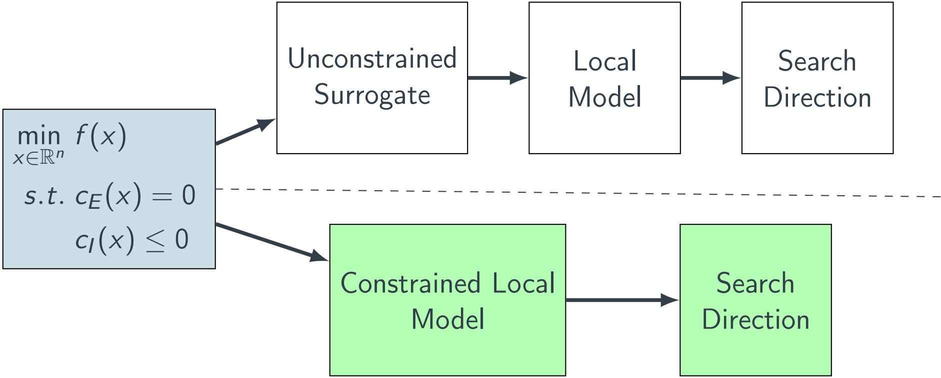
→ Unconstrained Surrogate → Local Model → Search Direction

Penalty Methods: $$\min_{x\in\mathbb{R}^n} \ f(x) + \rho \left\| c_E(x) \right\|^2 + \rho \left\| [c_I(x)]_+ \right\|^2$$

Proximal Methods: $$\min_{x\in\mathbb{R}^n} \ f(x) + \mathcal{I}_{\begin{bmatrix} c_E(x)=0 \\ c_I(x)\leq 0 \end{bmatrix}}(x)$$

Constrained Optimization

- Implicit Methods: Penalty Methods, Projection Methods, ...
- **Explicit Methods**: SQP Methods, Interior Point Methods, ...



Sequential Quadratic Programming

At iterate x_k:

1. Build local constrained model at x_k:

$$\min_{d_k \in \mathbb{R}^n} \frac{1}{2} d_k^T H_k d_k + d_k^T \nabla f(x_k)$$
$$s.t. \; c_E(x_k) + \nabla c_E(x_k)^T d_k = 0$$
$$c_I(x_k) + \nabla c_I(x_k)^T d_k \leq 0$$

Sequential Quadratic Programming

At iterate x_k:

1. Build local constrained model at x_k:

$$\begin{aligned} \min_{d_k \in \mathbb{R}^n} \; & \frac{1}{2} d_k^T H_k d_k + d_k^T \nabla f(x_k) \\ s.t. \; & c_E(x_k) + \nabla c_E(x_k)^T d_k = 0 \\ & c_I(x_k) + \nabla c_I(x_k)^T d_k \leq 0 \end{aligned}$$

2. Update the merit parameter τ_k in merit function $\phi(x, \tau)$ based on d_k

$$\phi(x, \tau) = \tau f(x) + \text{norm} \begin{pmatrix} c_E(x) \\ [c_I(x)]_+ \end{pmatrix}$$

3. Find step size α_k to update iterate: $x_{k+1} = x_k + \alpha_k d_k$

Stochastic Sequential Quadratic Programming

At iterate x_k:

1. Build local constrained model at x_k:

$$\begin{aligned} \min_{d_k \in \mathbb{R}^n} \ & \frac{1}{2} d_k^T H_k d_k + d_k^T g_k \\ s.t. \ & c_E(x_k) + \nabla c_E(x_k)^T d_k = 0 \\ & c_I(x_k) + \nabla c_I(x_k)^T d_k \leq 0 \end{aligned}$$

2. Update the merit parameter τ_k in merit function $\phi(x, \tau)$ based on d_k
3. Find step size α_k to update iterate: $x_{k+1} = x_k + \alpha_k d_k$

Stochastic Sequential Quadratic Programming

At iterate x_k:

1. Build local constrained model at x_k:

$$\begin{aligned} \min_{d_k \in \mathbb{R}^n} \ & \frac{1}{2} d_k^T H_k d_k + d_k^T g_k \\ s.t. \ & c_E(x_k) + \nabla c_E(x_k)^T d_k = 0 \\ & c_I(x_k) + \nabla c_I(x_k)^T d_k \leq 0 \end{aligned}$$

2. **Update the merit parameter τ_k in merit function $\phi(x, \tau)$ based on d_k**
3. **Find step size α_k to update iterate:** $x_{k+1} = x_k + \alpha_k d_k$

- Major redesign of algorithm components
  Berahas, Curtis, et al. 2021; Berahas, Bollapragada, et al. 2022; Na et al. 2023; O'Neill 2024, ...
- Conservative updates to safeguard against noise
- Slower convergence

Retrospective Approximation

Idea: Decouple the uncertainty from the optimization

Chen et al. 2001; Deng et al. 2009; Royset 2013; Jalilzadeh et al. 2016; Newton et al. 2024, ...

Retrospective Approximation

Idea: Decouple the uncertainty from the optimization

Chen et al. 2001; Deng et al. 2009; Royset 2013; Jalilzadeh et al. 2016; Newton et al. 2024, . . .

In iteration k:

1. Build a subsampled problem using sample set S_k:

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) = \frac{1}{|S_k|} \sum_{\xi \in S_k} F(x, \xi)$$

$$s.t.\ c_E(x) = 0$$

$$c_I(x) \leq 0$$

Retrospective Approximation

Idea: Decouple the uncertainty from the optimization

Chen et al. 2001; Deng et al. 2009; Royset 2013; Jalilzadeh et al. 2016; Newton et al. 2024, ...

In iteration k:

1. Build a subsampled problem using sample set S_k:

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) = \frac{1}{|S_k|} \sum_{\xi \in S_k} F(x, \xi)$$
$$s.t.\ c_E(x) = 0$$
$$c_I(x) \leq 0$$

2. Solve the subsampled problem up to a certain accuracy

Retrospective Approximation

Idea: Decouple the uncertainty from the optimization

Chen et al. 2001; Deng et al. 2009; Royset 2013; Jalilzadeh et al. 2016; Newton et al. 2024, ...

In iteration k:

1. Build a subsampled problem using sample set S_k:

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) = \frac{1}{|S_k|} \sum_{\xi \in S_k} F(x, \xi)$$
$$s.t. \; c_E(x) = 0$$
$$c_I(x) \leq 0$$

2. Solve the subsampled problem up to a certain accuracy
   - Can be achieved using **deterministic methods**
   - Requires control over the samples

Contributions

$$\boxed{\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x) = \mathbb{E}\left[F(x, \xi)\right] \\ s.t. \ & c_E(x) = 0 \\ & c_I(x) \leq 0 \end{aligned}}$$

1. Proposed a framework for equality constrained stochastic optimization problems that can use any deterministic solver
   - Proposed an instance with deterministic SQP solver
   - Established optimal theoretical complexity results
2. Proposed an algorithm for stochastic problems with general nonlinear constraints that uses the deterministic Robust SQP method
3. Illustrated benefits of the approach with numerical experiments

Retrospective Approximation

Algorithm Retrospective Approximation Constrained Optimization

Inputs: Initial iterate $x_{0,0}$, batch sizes $\{|S_k|\}$, termination tests $\{\mathcal{T}_k\}$.

```
for k = 0, 1, 2, ... do
    Construct the subsampled problem: min_{x∈ℝ^n} F_{S_k}(x) s.t. c_E(x) = 0, c_I(x) ≤ 0
    for j = 0, 1, 2, ..., do
        if T_k is satisfied; N_k = j, break
        Update x_{k,j+1}
    end for
    Set x_{k+1,0} = x_{k,N_k}
end for
```

Retrospective Approximation

Algorithm Retrospective Approximation Constrained Optimization

Inputs: Initial iterate $x_{0,0}$, batch sizes $\{|S_k|\}$, termination tests $\{\mathcal{T}_k\}$.

```
for k = 0, 1, 2, ... do
    Construct the subsampled problem: min_{x∈ℝ^n} F_{S_k}(x) s.t. c_E(x) = 0, c_I(x) ≤ 0
    for j = 0, 1, 2, ..., do
        if T_k is satisfied; N_k = j, break
        Update x_{k,j+1}
    end for
    Set x_{k+1,0} = x_{k,N_k}
end for
```

Questions to be addressed:

- Batch Size sequence $\{|S_k|\}$
- Termination Criterion sequence $\{\mathcal{T}_k\}$
- Deterministic Solver

Equality Constrained Optimization

$$\boxed{\begin{aligned}&\min_{x \in \mathbb{R}^n} \ f(x) \\ &\ s.t.\ c(x) = 0\end{aligned}}$$

- $c : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable, with $m \leq n$
- $\mathcal{L}(x, \lambda) = f(x) + \lambda^T c(x)$ where $\lambda \in \mathbb{R}^m$
- x^* is a stationary point if there exist λ^* such that,

$$\begin{aligned}\text{Lagrangian Gradient:} \quad & \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) + \nabla c(x^*)\lambda^* = 0 \\ \text{Feasible:} \quad & c(x^*) = 0\end{aligned}$$

With sample set S:

$$\begin{aligned}\text{Subsampled Problem:} \quad & \min_{x \in \mathbb{R}^n} \ F_S(x) \quad s.t. \quad c(x) = 0 \\ \text{Subsampled Lagrangian:} \quad & \mathcal{L}_S(x, \lambda) = F_S(x) + \lambda^T c(x)\end{aligned}$$

Analysis Equality Constrained Problems

Assumptions

Let χ be an open convex set containing all iterates

1. Linear Independence Constraint Qualification (LICQ)

$$rank(\nabla c(x)^T) = m \qquad \forall\, x \in \chi$$

2. $\|\nabla f(x)\| \leq \kappa_g \ \forall\, x \in \chi$
3. For any sample set S,

$$G_S = \max_{x \in \chi} \frac{\|\nabla f(x) - \nabla F_S(x)\|}{\epsilon_G + \|\nabla f(x)\|},$$

where $\epsilon_G > 0$. As $|S| \to \infty$, $E\left[G_S^2\right] \to 0$.

Pasupathy 2010; Newton et al. 2024

Termination Criterion Equality Constrained

In outer iteration $k \geq 0$, the inner loop is terminated when,

$$\mathcal{T}_k : \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k,$$

where $\gamma_k \in [0, 1)$ and $\epsilon_k \geq 0$.

Termination Criterion Equality Constrained

In outer iteration $k \geq 0$, the inner loop is terminated when,

$$\mathcal{T}_k : \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k,$$

where $\gamma_k \in [0, 1)$ and $\epsilon_k \geq 0$.

- Measures progress on the determinstic subsampled problem using KKT error
- Invariant of the chosen deterministic solver
- Requires dual variable estimate for the subsampled problem

$$\lambda_{k,j} = -(\nabla c(x_{k,j})^T \nabla c(x_{k,j}))^{-1} \nabla c(x_{k,j})^T \nabla F_{S_k}(x_{k,j})$$

Termination Criterion Equality Constrained

In outer iteration $k \geq 0$, the inner loop is terminated when,

$$\mathcal{T}_k : \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k,$$

where $\gamma_k \in [0, 1)$ and $\epsilon_k \geq 0$.

- Measures progress on the determinstic subsampled problem using KKT error
- Invariant of the chosen deterministic solver
- Requires dual variable estimate for the subsampled problem

$$\lambda_{k,j} = -(\nabla c(x_{k,j})^T \nabla c(x_{k,j}))^{-1} \nabla c(x_{k,j})^T \nabla F_{S_k}(x_{k,j})$$

Theorem (Informal)

Under the stated assumptions, if $E[\epsilon_k] \to 0$, $\{\gamma_k\} \leq \gamma < 1$, $\{|S_k|\} \to \infty$, and the inner loop terminates finitely, then,

$$\mathbb{E}\left[\left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,N_k}, \lambda_{k,N_k}) \\ c(x_{k,N_k}) \end{matrix} \right\|\right] \to 0.$$

Equality Constrained Framework Sampling

Adaptive Sampling Condition

In outer iteration $k \geq 0$, the sample set S_k is chosen such that,

$$\mathbb{E}\left[\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\|^2 \,|x_{k,0}\right] \leq \theta^2 \left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,0}, \lambda_{k-1,N_{k-1}}) \\ c(x_{k,0}) \end{matrix} \right\|^2 + a^2 \beta^{2k},$$

where $\theta \in [0, 1)$, $\beta \in [0, 1)$ and $a > 0$.

Equality Constrained Framework Sampling

Adaptive Sampling Condition

In outer iteration $k \geq 0$, the sample set S_k is chosen such that,

$$\mathbb{E}\left[\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\|^2 \,|x_{k,0}\right] \leq \theta^2 \left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,0}, \lambda_{k-1,N_{k-1}}) \\ c(x_{k,0}) \end{matrix} \right\|^2 + a^2\beta^{2k},$$

where $\theta \in [0,1)$, $\beta \in [0,1)$ and $a > 0$.

- Only tested at the start of outer iterations
- Invariant of the deterministic solver employed
- Requires true problem estimates (approximated in practice)
- When $\theta = 0$, can be satisfied using a geometric sequence, i.e.,

$$|S_{k+1}| = \left\lceil \frac{|S_k|}{\beta^2} \right\rceil$$

Sequential Quadratic Programming - Equality Constrained

$$\boxed{\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c(x) = 0}$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$

$$s.t.\ c(x_{k,j}) + \nabla c(x_{k,j})^T d_{k,j} = 0$$

Sequential Quadratic Programming - Equality Constrained

$$\boxed{\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c(x) = 0}$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$
$$s.t.\ c(x_{k,j}) + \nabla c(x_{k,j})^T d_{k,j} = 0$$

Step computation "Newton-SQP system":

- $\nabla c(x_{k,j})^T$ has linearly independent rows (LICQ)
- $H_{k,j}$ is positive definite over $Null\left(\nabla c(x_{k,j})^T\right)$

$$\begin{bmatrix} H_{k,j} & J_{k,j}^T \\ J_{k,j} & 0 \end{bmatrix} \begin{bmatrix} d_{k,j} \\ \delta_{k,j} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{bmatrix}$$

Sequential Quadratic Programming - Equality Constrained

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c(x) = 0$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\begin{bmatrix} H_{k,j} & J_{k,j}^T \\ J_{k,j} & 0 \end{bmatrix} \begin{bmatrix} d_{k,j} \\ \delta_{k,j} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{bmatrix}$$

2. Update the merit parameter $\tau_{k,j}$ in merit function $\phi_{S_k}(x, \tau)$ such that $\Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) >> 0$ to ensure $\phi'_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) << 0$

$$\phi_{S_k}(x, \tau) = \tau F_{S_k}(x) + \|c(x)\|_1$$

$$\Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) = -\tau_{k,j} \nabla F_{S_k}(x_{k,j})^T d_{k,j} + \|c_{k,j}\|_1 \leq -\phi'_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j})$$

Sequential Quadratic Programming - Equality Constrained

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t. \; c(x) = 0$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\begin{bmatrix} H_{k,j} & J_{k,j}^T \\ J_{k,j} & 0 \end{bmatrix} \begin{bmatrix} d_{k,j} \\ \delta_{k,j} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{bmatrix}$$

2. Update the merit parameter $\tau_{k,j}$ in merit function $\phi_{S_k}(x, \tau)$
3. Find step size $\alpha_{k,j}$ that satisfies the Armijo condition

$$\phi_{S_k}(x_{k,j} + \alpha_{k,j} d_{k,j}, \tau_{k,j}) \leq \phi_{S_k}(x_{k,j}, \tau_{k,j}) - \eta \alpha_{k,j} \Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j})$$

4. Update iterates: $(x_{k,j+1}, \lambda_{k,j+1}) = (x_{k,j}, \lambda_{k,j}) + \alpha_{k,j}(d_{k,j}, \delta_{k,j})$

Sequential Quadratic Programming - Equality Constrained

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c(x) = 0$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\left.\begin{bmatrix} H_{k,j} & J_{k,j}^T \\ J_{k,j} & 0 \end{bmatrix} \begin{bmatrix} d_{k,j} \\ \delta_{k,j} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{bmatrix} \right\} \text{Quasi Newton Approximation}$$

2. Update the merit parameter $\tau_{k,j}$ in merit function $\phi_{S_k}(x, \tau)$
3. Find step size $\alpha_{k,j}$ that satisfies the Armijo condition
4. Update iterates: $(x_{k,j+1}, \lambda_{k,j+1}) = (x_{k,j}, \lambda_{k,j}) + \alpha_{k,j}(d_{k,j}, \delta_{k,j})$

Sequential Quadratic Programming - Equality Constrained

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c(x) = 0$$

At $x_{k,j}$ and $\lambda_{k,j}$:

1. Build local constrained model at $x_{k,j}$:

$$\left.\begin{bmatrix} H_{k,j} & J_{k,j}^T \\ J_{k,j} & 0 \end{bmatrix} \begin{bmatrix} d_{k,j} \\ \delta_{k,j} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{bmatrix} + \begin{bmatrix} \rho_{k,j} \\ r_{k,j} \end{bmatrix} \right\} \text{Inexact Solutions}$$

2. Update the merit parameter $\tau_{k,j}$ in merit function $\phi_{S_k}(x, \tau)$
3. Find step size $\alpha_{k,j}$ that satisfies the Armijo condition
4. Update iterates: $(x_{k,j+1}, \lambda_{k,j+1}) = (x_{k,j}, \lambda_{k,j}) + \alpha_{k,j}(d_{k,j}, \delta_{k,j})$

Retrospective Approximation - SQP

Algorithm Retrospective Approximation - SQP Equality Constrained

Inputs: Initial iterate $x_{0,0}$ and dual variable $\lambda_{0,0}$, termination test parameters $\{\gamma_k\}$ and $\{\epsilon_k\}$, and sampling parameters a, β and θ.

for $k = 0, 1, 2, \dots$ **do**

  Choose S_k such that:

$$\mathbb{E}\left[\left\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\right\|^2 \middle| x_{k,0}\right] \leq \theta^2 \left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,0}, \lambda_{k-1,N_{k-1}}) \\ c(x_{k,0}) \end{matrix} \right\|^2 + a^2 \beta^{2k}$$

  Construct the subsampled problem: $\min_{x \in \mathbb{R}^n} F_{S_k}(x)$ $s.t.$ $c(x) = 0$

  for j = 0, 1, 2, ..., **do**

    if $\left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k$ **then**

      Set $N_k = j$, **break**

    end if

    Solve local constrained model for $d_{k,j}$, $\delta_{k,j}$

    Update merit parameter $\tau_{k,j}$

    Find step size $\alpha_{k,j}$

    Set $(x_{k,j+1}, \lambda_{k,j+1}) = (x_{k,j}, \lambda_{k,j}) + \alpha_{k,j}(d_{k,j}, \delta_{k,j})$

  end for

  Set $(x_{k+1,0}, \lambda_{k+1,0}) = (x_{k,N_k}, \lambda_{k,N_k})$

end for

Complexity of Stochastic SQP methods

Assumption Newton et al. 2024

CLT Scaling: $\mathbb{E}\left[G_{S_k}^2\right] \leq \frac{\kappa_G^2}{|S_k|}$ $\forall k \geq 0$, where $\kappa_G > 0$

Theorem (Informal)

Under the stated assumptions and approprate parameter selection, one achieves a solution satisfying,

$$\mathbb{E}\left[\left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,N_k}, \lambda_{k,N_k}) \\ c(x_{k,N_k}) \end{matrix} \right\|\right] \leq \epsilon,$$

with $\epsilon > 0$ in $K_\epsilon = \mathcal{O}\left(\log\left(\frac{1}{\epsilon}\right)\right)$ outer iterations, and

- $\mathcal{O}(\epsilon^{-2})$ inner iterations (SQP linear system solves, $\sum_{k=0}^{K_\epsilon} N_k$), and
- $\mathcal{O}(\epsilon^{-4})$ gradient evaluations ($\sum_{k=0}^{K_\epsilon} N_k |S_k|$).

Complexity of Stochastic SQP methods

SQP Methods	Linear Solves	Sample Gradients
Deterministic (Curtis, et al. 2024)	$\mathcal{O}(\epsilon^{-2})$	-
Stochastic (Curtis, et al. 2024)	$\mathcal{O}(\epsilon^{-4})$	$\mathcal{O}(\epsilon^{-4})$
Adaptive Sampling (Berahas, et al. 2022)	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2(\nu+1)}), \nu > 1$
Retrospective Approximation	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-4})$

Alternate Termination Criterion (SQP)

- Solver invariant termination criterion:

$$\mathcal{T}_k : \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k,$$

- SQP search direction termination criterion:

$$\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$$

- SQP merit function model:

$$\mathcal{T}_k : \Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) \leq \gamma_k \min\{\Delta l_{S_k}(x_{k,0}, \tau_{k,0}, d_{k,0}), \kappa_d \|d_{k,0}\|^2\} + \epsilon_k$$

Alternate Termination Criterion (SQP)

- Solver invariant termination criterion:

$$\mathcal{T}_k : \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \leq \gamma_k \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,0}, \lambda_{k,0}) \\ c(x_{k,0}) \end{matrix} \right\| + \epsilon_k,$$

- SQP search direction termination criterion:

$$\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$$

- SQP merit function model:

$$\mathcal{T}_k : \Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) \leq \gamma_k \min\{\Delta l_{S_k}(x_{k,0}, \tau_{k,0}, d_{k,0}), \kappa_d \|d_{k,0}\|^2\} + \epsilon_k$$

- Do not mix multiple quantities of different scales
- Do not require dual variable estimates
- Require the solution to the SQP subproblem for evaluation

Retrospective Approximation - SQP

Algorithm Retrospective Approximation - SQP Equality Constrained

Inputs: Initial iterate $x_{0,0}$ and dual variable $\lambda_{0,0}$, termination test parameters $\{\gamma_k\}$ and $\{\epsilon_k\}$, and sampling parameters a, β and θ.

for $k = 0, 1, 2, \ldots$ **do**

    Choose S_k such that:

$$\mathbb{E}\left[\left\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\right\|^2 |x_{k,0}\right] \leq \theta_k^2 \left\|d_{k,0}^{true}\right\|^2 + a^2\beta^{2k}$$

    Construct the subsampled problem: $\min_{x\in\mathbb{R}^n} F_{S_k}(x)\ s.t.\ c(x) = 0$

    for j = 0, 1, 2, ..., **do**

        Solve local constrained model for $d_{k,j}$, $\delta_{k,j}$

        Update merit parameter $\tau_{k,j}$

        if $\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$ **then**

            Set $N_k = j$, **break**

        end if

        Find step size $\alpha_{k,j}$

        Set $(x_{k,j+1}, \lambda_{k,j+1}) = (x_{k,j}, \lambda_{k,j}) + \alpha_{k,j}(d_{k,j}, \delta_{k,j})$

    end for

    Set $(x_{k+1,0}, \lambda_{k+1,0}) = (x_{k,N_k}, \lambda_{k,N_k})$

end for

Numerical Experiments

Termination Test:

$$\mathcal{T}_k : Z_{k,j} \leq \gamma_k Z_{k,0} + \epsilon_k$$

Adaptive Sampling: $\tilde{S}_{k-1}$ i.i.d. samples, independent of S_{k-1}, $|S_{k-1}| = |\tilde{S}_{k-1}|$

$$|S_k| = \min\left\{5|S_{k-1}|, \max\left\{|S_{k-1}|, \left\lceil \frac{Var_{\xi \in \tilde{S}_{k-1}}(\nabla F(x_{k,0}, \xi)|x_{k,0})}{\theta^2 \tilde{Z}_k^2} \right\rceil \right\}\right\}$$

- Algorithm to solve linear systems: MINRES
- Parameters: $\theta = 0.5$, $\epsilon_k = 10^{-6}$ and $\gamma_k = \gamma$
  - Merit function model condition, $\gamma = 0.1$
  - Step size norm condition, $\gamma = 0.5$

Multi Class Logistic Regression - Equality Regularization

$$\min_{x \in \mathbb{R}^n} \frac{1}{|\mathcal{S}|} \sum_{(y,t)\in\mathcal{S}} \sum_{i\in\mathcal{K}} t^i \log\left(\frac{1}{1+\exp(-y^T x^i)}\right)$$

$$s.t.\ \|x^i\|^2 = 1 \quad \forall i \in \mathcal{K}$$

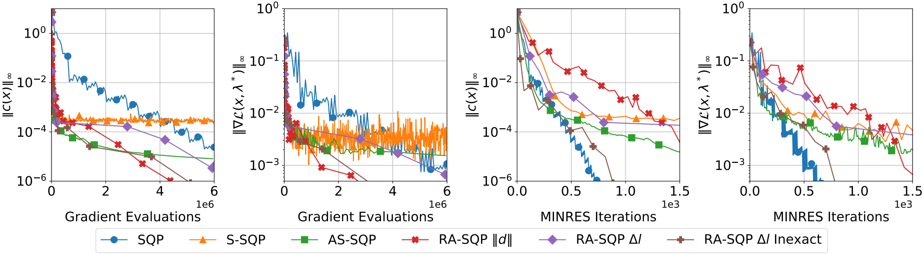


Figure: mnist $(n_f = 781,\ |\mathcal{K}| = 10,\ |\mathcal{S}| = 60,000,\ n = n_f|\mathcal{K}|)$

CUTEst problem set

- S2MPJ CUTEst problem set (Gratton et al. 2024)
- $F(x, \xi) = f(x) + \xi \|x - x_{init} - e_n\|^2$
- ξ is a uniformly distributed random variable in $[-0.1, 0.1]$

$$\min_{x \in \mathbb{R}^n} \mathbb{E}\left[F(x, \xi)\right] \quad s.t.\ c(x) = 0$$

- Total 88 problems selected that:
  1. Have only equality constraints and non constant objective function
  2. $m + n < 1000$
  3. Satisfy constraint qualification
- 10 seed runs per problem, each with a budget of 10^6 gradient evaluations

CUTEst problem set

- S2MPJ CUTEst problem set (Gratton et al. 2024)
- $F(x, \xi) = f(x) + \xi \|x - x_{init} - e_n\|^2$
- ξ is a uniformly distributed random variable in $[-0.1, 0.1]$

$$\min_{x \in \mathbb{R}^n} \mathbb{E}\left[F(x, \xi)\right] \quad s.t. \; c(x) = 0$$

- Total 88 problems selected that:
  1. Have only equality constraints and non constant objective function
  2. $m + n < 1000$
  3. Satisfy constraint qualification
- 10 seed runs per problem, each with a budget of 10^6 gradient evaluations

Solutions of ϵ_{tol} accuracy

Feasible Solution: $\|c(x_{out})\|_\infty \leq \epsilon_{tol} \max\{1, \|c(x_{init})\|_\infty\}$

Stationary Solution:
$$\|\nabla_x \mathcal{L}(x_{out}, \lambda_{out}^*)\|_\infty \leq \epsilon_{tol} \max\{1, \|\nabla_x \mathcal{L}(x_{init}, \lambda_{init}^*)\|_\infty\}$$
$$\|c(x_{out})\|_\infty \leq \epsilon_{tol} \max\{1, \|c(x_{init})\|_\infty\}$$

CUTEst problem set

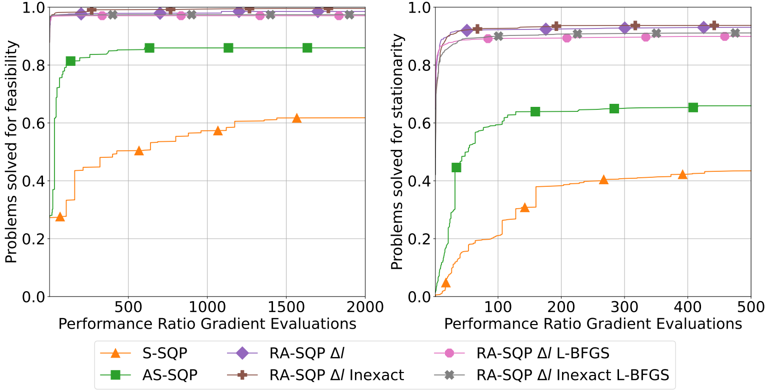


Figure: CUTEst performance profile equality constraints ($\epsilon_{tol} = 10^{-1}$)

CUTEst problem set

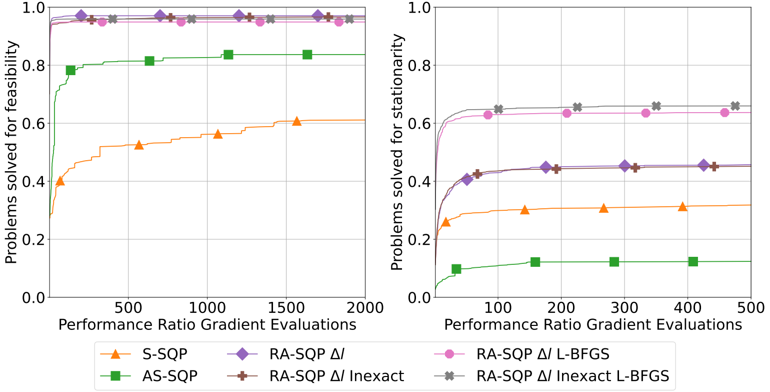


Figure: CUTEst performance profile equality constraints ($\epsilon_{tol} = 10^{-3}$)

Constrained Optimization

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x) \\ s.t. \ & c_E(x) = 0 \\ & c_I(x) \leq 0 \end{aligned}$$

- $\mathcal{L}(x, \lambda_E, \lambda_I) = f(x) + \lambda_E^T c_E(x) + \lambda_I^T c_I(x)$ where $\lambda_E \in \mathbb{R}^{m_E}$ and $\lambda_I \in \mathbb{R}^{m_I}$
- x^* is a stationary point if there exist λ_E^* and $\lambda_I^* \geq 0$ such that,

$$\begin{aligned} \text{Lagrangian Gradient:} \quad & \nabla_x \mathcal{L}(x^*, \lambda_E^*, \lambda_I^*) = 0 \\ \text{Feasible:} \quad & c_E(x^*) = 0 \quad \text{and} \quad c_I(x^*) \leq 0 \\ \text{Complimentary Slackness:} \quad & \lambda_I^* \odot c_I(x^*) = 0 \end{aligned}$$

Constrained Optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$s.t. \ c_E(x) = 0$$
$$c_I(x) \leq 0$$

- $\mathcal{L}(x, \lambda_E, \lambda_I) = f(x) + \lambda_E^T c_E(x) + \lambda_I^T c_I(x)$ where $\lambda_E \in \mathbb{R}^{m_E}$ and $\lambda_I \in \mathbb{R}^{m_I}$
- x^* is a stationary point if there exist λ_E^* and $\lambda_I^* \geq 0$ such that,

Lagrangian Gradient: $\nabla_x \mathcal{L}(x^*, \lambda_E^*, \lambda_I^*) = 0$

Feasible: $c_E(x^*) = 0$ and $c_I(x^*) \leq 0$

Complimentary Slackness: $\lambda_I^* \odot c_I(x^*) = 0$

- Combinatorial nature of constraints

Retrospective Approximation

Algorithm Retrospective Approximation

Inputs: Initial iterate $x_{0,0}$, batch sizes $\{|S_k|\}$, termination tests $\{\mathcal{T}_k\}$.

for $k = 0, 1, 2, \ldots$ **do**
  Construct the subsampled problem: $\min_{x \in \mathbb{R}^n} F_{S_k}(x)$ $s.t.$ $c_E(x) = 0, c_I(x) \leq 0$
  for j = 0, 1, 2, ..., **do**
    if $\mathcal{T}_k$ is satisfied; $N_k = j$, **break**
    Update $x_{k,j+1}$
  end for
  Set $x_{k+1,0} = x_{k,N_k}$
end for

Questions to be addressed:

- Batch Size sequence $\{|S_k|\}$
- Termination Criterion sequence $\{\mathcal{T}_k\}$
- Deterministic Solver

Extending Equality Constrained Framework

$$\left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\| \Rightarrow \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_E, \lambda_I) \\ c_E(x_{k,j}) \\ [c_I(x_{k,j})]_+ \\ \lambda_I \odot c_I(x_{k,j}) \\ [-\lambda_I]_+ \end{matrix} \right\|$$

- Scaling issues
  - Multiple errors combined due to the presence of inequality constraints
- Dual variable availability for the subsampled problem
  - Requires solving a Linear Program if not updated by deterministic solver

Extending Equality Constrained Framework

$$\left\| \begin{array}{c} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{array} \right\| \Rightarrow \left\| \begin{array}{c} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_E, \lambda_I) \\ c_E(x_{k,j}) \\ [c_I(x_{k,j})]_+ \\ \lambda_I \odot c_I(x_{k,j}) \\ [-\lambda_I]_+ \end{array} \right\|$$

- Scaling issues
  - Multiple errors combined due to the presence of inequality constraints
- Dual variable availability for the subsampled problem
  - Requires solving a Linear Program if not updated by deterministic solver

- We focus on a specific solver for inequality constraints - SQP

Inequality Constrained SQP

$$\boxed{\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0,\ c_I(x) \leq 0}$$

Local constrained model at $x_{k,j}$:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$

$$\begin{aligned} s.t.\ c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) d_{k,j} &= 0 \\ c_I(x_{k,j}) + \nabla c_I^T(x_{k,j}) d_{k,j} &\leq 0 \end{aligned}$$

Inequality Constrained SQP

$$\boxed{\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0,\ c_I(x) \leq 0}$$

Local constrained model at $x_{k,j}$:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$

$$s.t.\ c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) d_{k,j} = 0$$

$$c_I(x_{k,j}) + \nabla c_I^T(x_{k,j}) d_{k,j} \leq 0$$

1. **Feasibility of the local model**:
   - Not guaranteed under any constraint qualification (Nocedal et al. 1999)

Inequality Constrained SQP

$$\boxed{\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0,\ c_I(x) \leq 0}$$

Local constrained model at $x_{k,j}$:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$

$$s.t.\ c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) d_{k,j} = 0$$

$$c_I(x_{k,j}) + \nabla c_I^T(x_{k,j}) d_{k,j} \leq 0$$

1. **Feasibility of the local model**:
   - Not guaranteed under any constraint qualification (Nocedal et al. 1999)
2. **Feasibility of the true problem**:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|c_E(x)\|^2 + \frac{1}{2} \left\| [c_I(x)]_+ \right\|^2$$

   - A stationary point to constraint violation minimization may not be feasible
   - **Infeasible Stationary Points**

Robust SQP Burke et al. 1989

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0,\ c_I(x) \leq 0$$

At iterate $x_{k,j}$

$$\text{Feasibility:} \quad r_{k,j} = \min_{p_{k,j} \in \mathbb{R}^n} \text{norm} \begin{pmatrix} c_E(x_{k,j}) + \nabla c_E(x_{k,j})^T p_{k,j} \\ \left[c_I(x_{k,j}) + \nabla c_I(x_{k,j})^T p_{k,j}\right]_+ \end{pmatrix}$$

$$s.t.\ \text{norm}(p_{k,j}) \leq \sigma_p$$

Robust SQP Burke et al. 1989

$$\min_{x \in \mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0,\ c_I(x) \leq 0$$

At iterate $x_{k,j}$

Feasibility:

$$r_{k,j} = \min_{p_{k,j} \in \mathbb{R}^n} \text{norm} \begin{pmatrix} c_E(x_{k,j}) + \nabla c_E(x_{k,j})^T p_{k,j} \\ \left[c_I(x_{k,j}) + \nabla c_I(x_{k,j})^T p_{k,j}\right]_+ \end{pmatrix}$$

$$s.t.\ \text{norm}(p_{k,j}) \leq \sigma_p$$

Optimality:

$$\min_{d_{k,j} \in \mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$

$$s.t.\ \text{norm} \begin{pmatrix} c_E(x_{k,j}) + \nabla c_E(x_{k,j})^T d_{k,j} \\ \left[c_I(x_{k,j}) + \nabla c_I(x_{k,j})^T d_{k,j}\right]_+ \end{pmatrix} \leq r_{k,j}$$

$$\text{norm}(d_{k,j}) \leq \sigma_d$$

Robust SQP Burke et al. 1989

$$\min_{x\in\mathbb{R}^n} F_{S_k}(x) \quad s.t. \; c_E(x) = 0, \; c_I(x) \leq 0$$

At iterate $x_{k,j}$

Feasibility:
$$r_{k,j} = \min_{p_{k,j}\in\mathbb{R}^n} \text{norm}\begin{pmatrix} c_E(x_{k,j}) + \nabla c_E(x_{k,j})^T p_{k,j} \\ \left[c_I(x_{k,j}) + \nabla c_I(x_{k,j})^T p_{k,j}\right]_+ \end{pmatrix}$$
$$s.t. \; \text{norm}(p_{k,j}) \leq \sigma_p$$

Optimality:
$$\min_{d_{k,j}\in\mathbb{R}^n} \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j})$$
$$s.t. \; \text{norm}\begin{pmatrix} c_E(x_{k,j}) + \nabla c_E(x_{k,j})^T d_{k,j} \\ \left[c_I(x_{k,j}) + \nabla c_I(x_{k,j})^T d_{k,j}\right]_+ \end{pmatrix} \leq r_{k,j}$$
$$\text{norm}(d_{k,j}) \leq \sigma_d$$

- If $p_{k,j} = 0$ and $x_{k,j}$ is infeasible, $x_{k,j}$ is an infeasible stationary point
- If $d_{k,j} = 0$, $x_{k,j}$ is a stationary point

Termination Criterion and Sampling Condition

Termination Criterion

In outer iteration $k \geq 0$, the inner loop is terminated when,

$$\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k,$$

where $\gamma_k \in [0, 1)$ and $\epsilon_k \geq 0$.

Adaptive Sampling Condition

In outer iteration $k \geq 0$, the sample set S_k is chosen large enough such that,

$$\mathbb{E}\left[\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\|^2\right] \leq \theta^2 \left\|d_{k,0}^{true}\right\|^2 + a^2\beta^{2k},$$

where $\theta \in [0, 1)$, $\beta \in [0, 1)$ and $a > 0$.

Retrospective Approximation - Robust SQP

Algorithm Retrospective Approximation - Robust SQP

Inputs: Initial iterate $x_{0,0}$ and dual variable $\lambda_{0,0}$, termination test parameters $\{\gamma_k\}$ and $\{\epsilon_k\}$, and sampling parameters a, β and θ.

for $k = 0, 1, 2, \ldots$ **do**
  Choose S_k such that:

$$\mathbb{E}\left[\left\|\nabla F_{S_k}(x_{k,0}) - \nabla f(x_{k,0})\right\|^2\right] \leq \theta_k^2 \left\|d_{k,0}^{true}\right\|^2 + a^2\beta^{2k}$$

  Construct the subsampled problem: $\min_{x\in\mathbb{R}^n} F_{S_k}(x) \quad s.t.\ c_E(x) = 0, c_I(x) \leq 0$
  for j = 0, 1, 2, ..., **do**
    Solve feasability problem for $p_{k,j}$
    if $p_{k,j} = 0$ and $x_{k,j}$ is infeasible **then**
      return $x_{k,j}$ as **infeasible stationary point**
    end if
    Solve optimality problem for $d_{k,j}$
    if $\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$ **then**
      Set $N_k = j$, **break**
    end if
    Update merit parameter $\tau_{k,j}$ and find step size $\alpha_{k,j}$
    Set $x_{k,j+1} = x_{k,j} + \alpha_{k,j} d_{k,j}$
  end for
  Set $x_{k+1,0} = x_{k,N_k}$
end for

Analysis Inequality Constrained Problems

Assumptions

Let χ be an closed bounded convex set containing all the iterates.

1. The extended MFCQ (Mangasarian-Fromovitz constraint qualification) hold $\forall x \in \chi$, i.e.,
   - $rank(\nabla c_E(x)) = m_E$,
   - there exists $u \in \mathbb{R}^n$ such that:

$$\nabla c_E(x)^T u = 0 \quad \text{and} \quad [\nabla c_I(x)^T]_i u > 0 \quad \forall i \in \{i : [c_I(x)]_i \geq 0\}$$

2. $\mu_H I_n \preceq \{H_{k,j}\} \preceq \kappa_H I_n$ is a sequence of positive definite matrices

Analysis Inequality Constrained Problems

Assumptions

Let χ be an closed bounded convex set containing all the iterates.

1. The extended MFCQ (Mangasarian-Fromovitz constraint qualification) hold $\forall x \in \chi$, i.e.,
   - $rank(\nabla c_E(x)) = m_E$,
   - there exists $u \in \mathbb{R}^n$ such that:

   $$\nabla c_E(x)^T u = 0 \quad \text{and} \quad [\nabla c_I(x)^T]_i u > 0 \quad \forall i \in \{i : [c_I(x)]_i \geq 0\}$$
2. $\mu_H I_n \preceq \{H_{k,j}\} \preceq \kappa_H I_n$ is a sequence of positive definite matrices

Theorem (Informal)

Under the stated assumptions and appropriate parameter selection, the $\mathbb{E}\left[\|d_{k,N_k}^{true}\|\right] \to 0$ at a **linear rate** across outer iterations.

Analysis Inequality Constrained Problems

Assumptions

Let χ be an closed bounded convex set containing all the iterates.

1. The extended MFCQ (Mangasarian-Fromovitz constraint qualification) hold $\forall x \in \chi$, i.e.,
   - $rank(\nabla c_E(x)) = m_E$,
   - there exists $u \in \mathbb{R}^n$ such that:
   $$\nabla c_E(x)^T u = 0 \quad \text{and} \quad [\nabla c_I(x)^T]_i u > 0 \quad \forall i \in \{i : [c_I(x)]_i \geq 0\}$$
2. $\mu_H I_n \preceq \{H_{k,j}\} \preceq \kappa_H I_n$ is a sequence of positive definite matrices

Theorem (Informal)

Under the stated assumptions and appropriate parameter selection, the $\mathbb{E}\left[\|d^{true}_{k,N_k}\|\right] \to 0$ at a **linear rate** across outer iterations.

- $\mathbb{E}\left[\|d^{true}_{k,N_k}\|\right] \to 0$ does not necessarily imply convergence to a stationary point
- Only $d^{true}_{k,N_k} = 0$ confirms that x_{k,N_k} is a stationary point

Curtis et al. 2023; Qiu et al. 2023, ...

Numerical Experiments

Termination Test:

$$\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$$

Adaptive Sampling: $\tilde{S}_{k-1}$ i.i.d. samples, independent of S_{k-1}, $|S_{k-1}| = |\tilde{S}_{k-1}|$

$$|S_k| = \min\left\{|\mathcal{S}|, 5|S_{k-1}|, \max\left\{|S_{k-1}|, \left\lceil \frac{Var_{\xi \in \tilde{S}_{k-1}}(\nabla F(x_{k,0}, \xi)|x_{k,0})}{\theta^2 \tilde{d}_{k,0}^2} \right\rceil \right\}\right\}$$

Numerical Experiments

Termination Test:

$$\mathcal{T}_k : \|d_{k,j}\| \leq \gamma_k \|d_{k,0}\| + \epsilon_k$$

Adaptive Sampling: $\tilde{S}_{k-1}$ i.i.d. samples, independent of S_{k-1}, $|S_{k-1}| = |\tilde{S}_{k-1}|$

$$|S_k| = \min\left\{|\mathcal{S}|, 5|S_{k-1}|, \max\left\{|S_{k-1}|, \left\lceil \frac{Var_{\xi \in \tilde{S}_{k-1}}(\nabla F(x_{k,0},\xi)|x_{k,0})}{\theta^2 \tilde{d}_{k,0}^2} \right\rceil \right\}\right\}$$

- l_∞ and l_1 norm based robust-SQP subproblems
- Solver for Quadratic and Linear Programs: GUROBI (Barrier Method)
- Parameters: $\theta = 0.5$, $\epsilon_k = 10^{-6}$ and $\gamma_k = \gamma = 0.5$
- Stationarity Metric:

$$KKT(x) = \min_{\lambda_E \in \mathbb{R}^{m_E}, \lambda_I \in \mathbb{R}^{m_I}, \lambda_I \geq 0} \max\{\|\nabla_x \mathcal{L}(x, \lambda_E, \lambda_I)\|_\infty, \|\lambda_I \odot c_I(x)\|_\infty\}$$

Multi Class Logistic Regression - Inequality Regularization

$$\min_{x\in\mathbb{R}^n} \frac{1}{|\mathcal{S}|} \sum_{(y,t)\in\mathcal{S}} \sum_{i\in\mathcal{K}} t^i \log\left(\frac{1}{1+\exp(-y^T x^i)}\right)$$

$$s.t.\ \|x^i\|^2 \leq 1 \quad \forall i \in \mathcal{K}$$

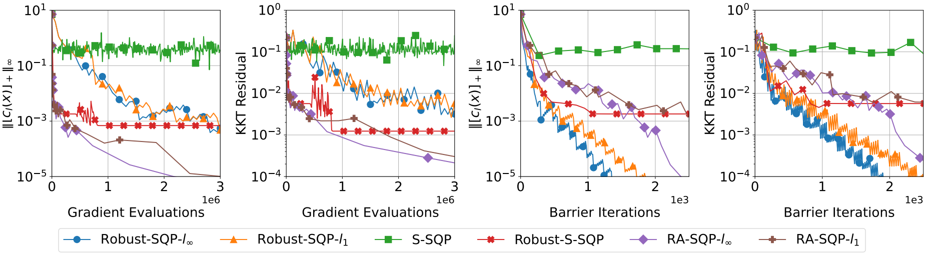


Figure: mnist $(n_f = 781,\ |\mathcal{K}| = 10,\ |\mathcal{S}| = 60,000,\ n = n_f|\mathcal{K}|)$

CUTEst problem set

- S2MPJ CUTEst problem set (Gratton et al. 2024)
- $F(x,\xi) = f(x) + \xi\|x - x_{init} - e_n\|^2$
- ξ is a uniformly distributed random variable in $[-0.1, 0.1]$

$$\min_{x\in\mathbb{R}^n} \mathbb{E}\left[F(x,\xi)\right] \quad s.t.\ c_E(x) = 0,\ c_I(x) \le 0$$

- Total 248 problems selected that:
  1. Have altеast one inequality (not bound constraints) constraint and non constant objective function
  2. $m_E + m_I + n < 2000$
  3. Satisfy constraint qualification
- 10 seed runs per problem, each with a budget of 10^6 gradient evaluations

CUTEst problem set

- S2MPJ CUTEst problem set (Gratton et al. 2024)
- $F(x, \xi) = f(x) + \xi \|x - x_{init} - e_n\|^2$
- ξ is a uniformly distributed random variable in $[-0.1, 0.1]$

$$\min_{x \in \mathbb{R}^n} \mathbb{E}[F(x, \xi)] \quad s.t. \; c_E(x) = 0, \; c_I(x) \leq 0$$

- Total 248 problems selected that:
  1. Have alteast one inequality (not bound constraints) constraint and non constant objective function
  2. $m_E + m_I + n < 2000$
  3. Satisfy constraint qualification
- 10 seed runs per problem, each with a budget of 10^6 gradient evaluations

Solutions of ϵ_{tol} accuracy

Feasible Solution: $\left\| \begin{matrix} c_E(x_{out}) \\ [c_I(x_{out})]_+ \end{matrix} \right\|_\infty \leq \epsilon_{tol} \max \left\{ 1, \left\| \begin{matrix} c_E(x_{init}) \\ [c_I(x_{init})]_+ \end{matrix} \right\|_\infty \right\}$

Stationary Solution: $KKT(x_{out}) \leq \epsilon_{tol} \max \{1, KKT(x_{init})\}$

A stationary solution must also be a feasible solution

CUTEst problem set

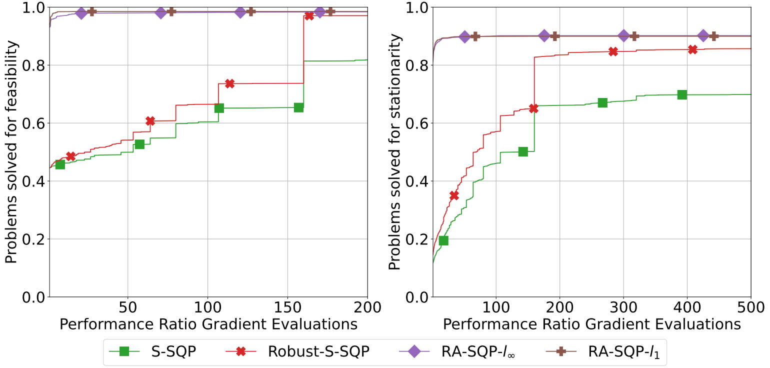


Figure: CUTEst performance profile inequality constraints ($\epsilon_{tol} = 10^{-1}$)

CUTEst problem set

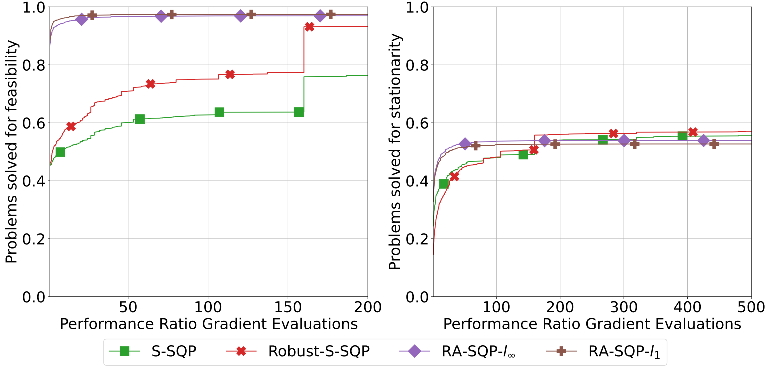


Figure: CUTEst performance profile inequality constraints ($\epsilon_{tol} = 10^{-3}$)

Final Remarks

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & f(x) = \mathbb{E}\left[F(x,\xi)\right] \\ s.t. \; & c_E(x) = 0 \\ & c_I(x) \leq 0 \end{aligned}$$

1. Proposed a framework for equality constrained stochastic optimization problems that can employ any deterministic solve
2. Proposed a variant of the framework for equality constraints that uses the deterministic SQP method
3. Proposed an algorithm for stochastic problems with general nonlinear constraints using the Robust SQP method
4. Illustrated the benefits of the proposed methods with numerical experiments

Manuscript available at: `https://arxiv.org/pdf/2505.19382`

Thank You!

Infeasiable SQP problems

Infeasible stationary points (Burke et al. 1989):

$$c_I(x) = \begin{bmatrix} x^2 + 1 \\ x \end{bmatrix} \leq 0$$

Infeasible Subproblem:

$$c_I(x) = \begin{bmatrix} x_1^2 + x_2^2 \\ -x_1 \end{bmatrix} \leq \begin{bmatrix} 5 \\ -2.05 \end{bmatrix}$$

At $(x_1, x_2) = (1.9, 0)$:

$$\begin{bmatrix} x_1^2 + x_2^2 - 5 + 2x_1 d_1 + 2x_2 d_2 \\ -x_1 + 2.05 - d_1 \end{bmatrix} = \begin{bmatrix} -0.0134 + d_1 \\ 0.15 - d_1 \end{bmatrix} \leq 0 \Rightarrow \begin{matrix} d_1 \leq 0.0134 \\ d_1 \geq 0.15 \end{matrix}$$

SQP Equality subproblem inexactness conditions

Inexactness condition I:

$$
\begin{aligned}
\Delta l_{S_k}(x_{k,j}, \tau_{k,j-1}, d_{k,j}) &\geq \epsilon_\sigma (1 - \epsilon_{feas}) \max\{\|c_{k,j}\|_1, \|r_{k,j}\| - \|c_{k,j}\|_1\} \\
&\quad + \epsilon_\sigma (1 - \epsilon_{feas}) \tau_{k,j-1} \max\{d_{k,j}^T H_{k,j} d_{k,j}, \epsilon_d \|d_{k,j}\|^2\}, \\
\left\| \begin{bmatrix} \rho_{k,j} \\ r_{k,j} \end{bmatrix} \right\| &\leq \kappa_T \min \left\{ \left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,j}, \lambda_{k,j}) \\ c(x_{k,j}) \end{matrix} \right\|, \|d_{k,j}\| \right\}, \\
\text{and} \qquad \|\rho_{k,j}\| &\leq \kappa' \max\{\|J_{k,j}\|, \|g_{S_k}(x_{k,j})\|\},
\end{aligned}
$$

Inexactness condition II:

$$
\|r_{k,j}\| \leq \epsilon_{feas} \|c_{k,j}\| \quad \text{and} \quad \|\rho_{k,j}\| \leq \epsilon_{opt} \|c_{k,j}\|,
$$

Merit Parameter Update

$$\begin{aligned}\Delta l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) &= l_{S_k}(x_{k,j}, \tau_{k,j}, 0) - l_{S_k}(x_{k,j}, \tau_{k,j}, d_{k,j}) \\ &= -\tau_{k,j}\nabla F_{S_k}(x_{k,j})^T d_{k,j} + \|c_{k,j}\|_1 - \|c_{k,j} + J_{k,j} d_{k,j}\|_1 \\ &= -\tau_{k,j}\nabla F_{S_k}(x_{k,j})^T d_{k,j} + \|c_{k,j}\|_1 - \|r_{k,j}\|_1\end{aligned}$$

$$\tau_{k,j}^{trial} = \begin{cases} \infty & \text{if } \nabla F_{S_k}(x_{k,j})^T d_{k,j} + \max\{d_{k,j}^T H_{k,j} d_{k,j}, 0\} \leq 0, \\ \frac{(1-\epsilon_\sigma)(\|c_{k,j}\|_1 - \|r_{k,j}\|_1)}{\nabla F_{S_k}(x_{k,j})^T d_{k,j} + \max\{d_{k,j}^T H_{k,j} d_{k,j}\},0} & \text{otherwise,} \end{cases}$$

$$\tau_{k,j} = \begin{cases} \tau_{k,j-1} & \text{if } \tau_{k,j-1} \leq \tau_{k,j}^{trial}, \\ (1-\epsilon_\tau)\tau_{k,j}^{trial} & \text{otherwise,} \end{cases}$$

Adaptive Sampling Derivation:

$$\frac{Var(\nabla F(x_{k,0}, \xi)|x_{k,0})}{|S_k|} \leq \tilde{\theta}^2 Z_k^2 + \tilde{a}^2 \tilde{\beta}^{2k} \quad \Rightarrow \quad |S_k| \geq \frac{Var(\nabla F(x_{k,0}, \xi)|x_{k,0})}{\tilde{\theta}^2 Z_k^2 + \tilde{a}^2 \tilde{\beta}^{2k}},$$

CUTEst problem set

Total number of problems : 1098

Number of unconstrained problems : 248

Number of with only equality problems : 208

Number of with atleast one inequality constraint (not bound constraints) : 278

CUTEst problem set equality

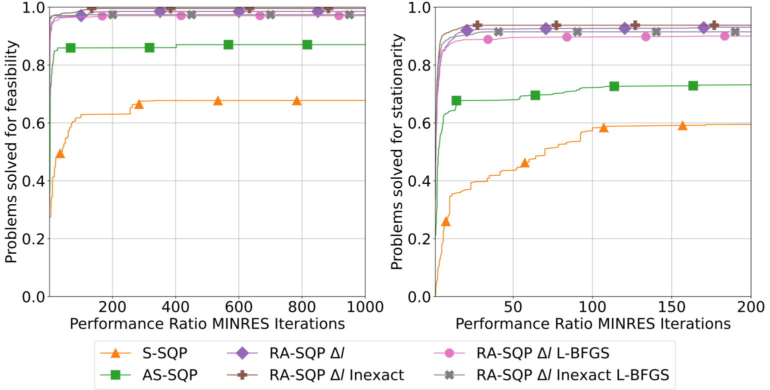


Figure: CUTEst performance profile equality constraints ($\epsilon_{tol} = 10^{-1}$)

CUTEst problem set equality

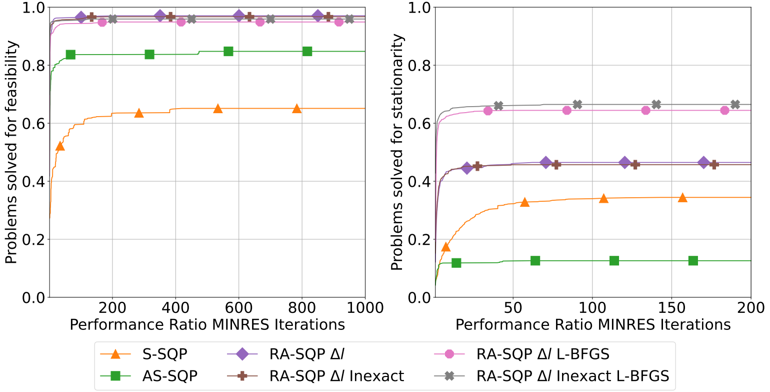


Figure: CUTEst performance profile equality constraints ($\epsilon_{tol} = 10^{-3}$)

CUTEst problem set inequality

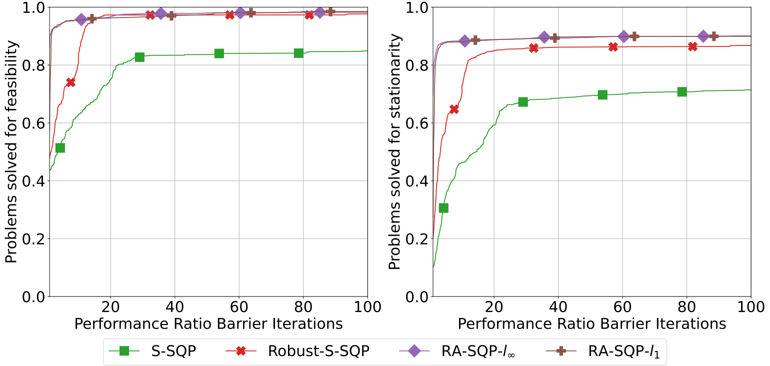


Figure: CUTEst performance profile inequality constraints ($\epsilon_{tol} = 10^{-1}$)

CUTEst problem set inequality

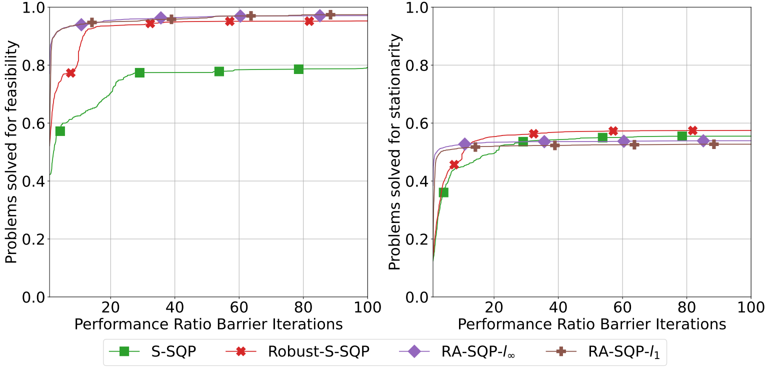


Figure: CUTEst performance profile inequality constraints ($\epsilon_{tol} = 10^{-3}$)

Robust SQP Burke et al. 1989

Using l_∞ norm at iterate $x_{k,j}$:

Feasibility:

$$\begin{aligned}
\min_{\substack{y\in\mathbb{R},\\ p_{k,j}\in\mathbb{R}^n}} \quad & y \\
s.t. \quad & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j})p_{k,j} \geq -ye_{m_E} \\
& c_E(x_{k,j}) + \nabla c_E^T(x_{k,j})p_{k,j} \leq ye_{m_E} \\
& c_I(x_{k,j}) + \nabla c_I^T(x_{k,j})p_{k,j} \leq ye_{m_I} \\
& y \geq 0, \|p_{k,j}\|_\infty \leq \sigma_p
\end{aligned}$$

Optimality:

$$\begin{aligned}
\min_{d_{k,j}\in\mathbb{R}^n} \quad & \frac{1}{2}d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j}) \\
s.t. \quad & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j})d_{k,j} \geq -ye_{m_E} \\
& c_E(x_{k,j}) + \nabla c_E^T(x_{k,j})d_{k,j} \leq ye_{m_E} \\
& c_I(x_{k,j}) + \nabla c_I^T(x_{k,j})d_{k,j} \leq ye_{m_I} \\
& \|d_{k,j}\|_\infty \leq \sigma_d
\end{aligned}$$

Robust SQP Burke et al. 1989

Using l_1 norm at iterate $x_{k,j}$:

Feasibility:

$$\begin{aligned} \min_{\substack{y_E \in \mathbb{R}^{m_E}, \\ y_I \in \mathbb{R}^{m_I}, \\ p_{k,j} \in \mathbb{R}^n}} \quad & e_{m_E}^T y_E + e_{m_I}^T y_I \\ s.t. \quad & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) p_{k,j} \geq -y_E \\ & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) p_{k,j} \leq y_E \\ & c_I(x_{k,j}) + \nabla c_I^T(x_{k,j}) p_{k,j} \leq y_I \\ & y_E \geq 0, y_I \geq 0, \|p_{k,j}\|_1 \leq \sigma_p, \end{aligned}$$

Optimality:

$$\begin{aligned} \min_{d_{k,j} \in \mathbb{R}^n} \quad & \frac{1}{2} d_{k,j}^T H_{k,j} d_{k,j} + d_{k,j}^T \nabla F_{S_k}(x_{k,j}) \\ s.t. \quad & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) d_{k,j} \geq -y_E \\ & c_E(x_{k,j}) + \nabla c_E^T(x_{k,j}) d_{k,j} \leq y_E \\ & c_I(x_{k,j}) + \nabla c_I^T(x_{k,j}) d_{k,j} \leq y_I \\ & \|d_{k,j}\|_1 \leq \sigma_d, \end{aligned}$$

Multi Class Logistic Regression - Equality Regularization

$$\min_{x \in \mathbb{R}^n} \frac{1}{|\mathcal{S}|} \sum_{(y,t)\in\mathcal{S}} \sum_{i\in\mathcal{K}} t^i \log\left(\frac{1}{1+\exp(-y^T x^i)}\right)$$

$$s.t.\ \|x^i\|^2 = 1 \quad \forall i \in \mathcal{K}$$

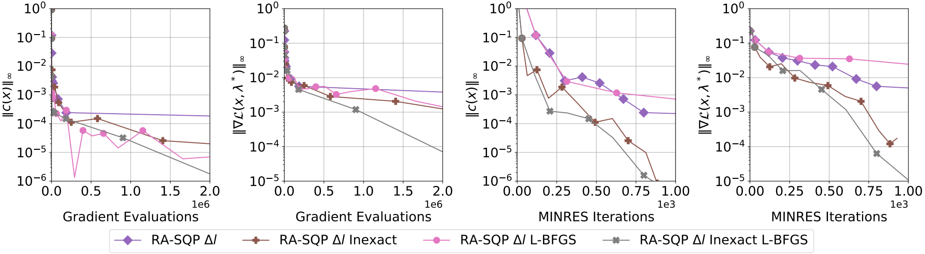


Figure: mnist $(n_f = 781,\ |\mathcal{K}| = 10,\ |\mathcal{S}| = 60,000,\ n = n_f|\mathcal{K}|)$

Multi Class Logistic Regression - Equality Regularization

$$\min_{x\in\mathbb{R}^n} \frac{1}{|\mathcal{S}|} \sum_{(y,t)\in\mathcal{S}} \sum_{i\in\mathcal{K}} t^i \log\left(\frac{1}{1+\exp(-y^T x^i)}\right)$$
$$s.t. \ \|x^i\|^2 = 1 \quad \forall i \in \mathcal{K}$$

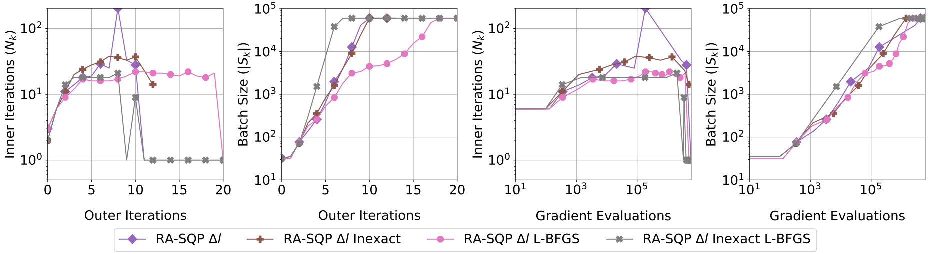


Figure: mnist $(n_f = 781,\ |\mathcal{K}| = 10,\ |\mathcal{S}| = 60,000,\ n = n_f|\mathcal{K}|)$

Equality Constrained Framework Complexity

Assumptions

Let χ be an open convex set containing all the iterates.

1. CLT Scaling: $\mathbb{E}\left[G_{S_k}^2\right] \leq \frac{\kappa_G^2}{|S_k|}$ $\forall k \geq 0$, where $\kappa_G > 0$.
2. Variance Lower Bound: $\mathrm{Var}\left(\nabla F(x_{k,0}, \xi) | x_{k,0}\right) \geq \kappa_G^2 \kappa_\sigma^2$ $\forall k \geq 0$, where $\kappa_\sigma > 0$.

Theorem (Informal)

Under the stated assumptions, if $0 \leq \gamma_k \leq \gamma < 1$ and $\epsilon_k = \omega\sqrt{\frac{\mathrm{Var}(\nabla F(x_{k,0},\xi)|x_{k,0})}{|S_k|}}$ where $\omega \geq 0$, and θ is chosen such that $\left[\gamma + \theta\left(\omega + \frac{(\epsilon_G + \kappa_g)}{\kappa_\sigma} + \gamma\right)\right] < 1$, then $\mathbb{E}\left[\left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,N_k}, \lambda_{k,N_k}) \\ c(x_{k,N_k}) \end{matrix} \right\|\right] \to 0$ at a **linear rate** across outer iterations.

Equality Constrained Framework Complexity

Assumption

The inner loop converges at a sublinear rate, i.e., a solution satisfying $\left\| \begin{matrix} \nabla_x \mathcal{L}_{S_k}(x_{k,N_k}, \lambda_{k,N_k}) \\ c(x_{k,N_k}) \end{matrix} \right\| \leq \epsilon$ with $\epsilon > 0$ is achieved in $\mathcal{O}(\epsilon^{-2})$ inner iterations.

Theorem (Informal)

Under the stated assumptions and parameter selection, one achieves a solution satisfying,

$$\mathbb{E}\left[\left\| \begin{matrix} \nabla_x \mathcal{L}(x_{k,N_k}, \lambda_{k,N_k}) \\ c(x_{k,N_k}) \end{matrix} \right\|\right] \leq \epsilon,$$

with $\epsilon > 0$ in $K_\epsilon = \mathcal{O}\left(\log\left(\frac{1}{\epsilon}\right)\right)$ outer iterations, and

- $\mathcal{O}(\epsilon^{-2})$ inner iterations (deterministic solver iterations, $\sum_{k=0}^{K_\epsilon} N_k$), and
- $\mathcal{O}(\epsilon^{-4})$ gradient evaluations ($\sum_{k=0}^{K_\epsilon} N_k |S_k|$).

Analysis Inequality Constrained Problems

Assumptions

Let χ be an closed bounded convex set containing all the iterates.

1. The extended MFCQ (Mangasarian-Fromovitz constraint qualification) hold $\forall x \in \chi$, i.e.,
   - $rank(\nabla c_E(x)) = m_E$,
   - $\exists u \in \mathbb{R}^n$ such that:

$$\nabla c_E(x)^T u = 0 \quad \text{and} \quad [\nabla c_I(x)^T]_i u > 0 \quad \forall i \in \{i : [c_I(x)]_i \geq 0\}$$

2. $\mu_H I_n \preceq \{H_{k,j}\} \preceq \kappa_H I_n$ is a sequence of positive definite matrices

Theorem (Informal)

Under the stated assumptions, if $0 \leq \{\gamma_k\} \leq \gamma < 1$, $\epsilon_k = \omega \sqrt{\frac{\mathrm{Var}(\nabla F(x_{k,0}) | \mathcal{F}_k)}{|S_k|}}$ where $\omega \geq 0$ and θ is chosen such that $\left[\gamma + \theta\left(\omega + \mu_H^{-1}\left(\frac{(\epsilon_G + \kappa_g)}{\kappa_\sigma} + \gamma\right)\right)\right] < 1$, then $\mathbb{E}\left[\|d_{k,N_k}^{true}\|\right] \to 0$ at a **linear rate** across outer iterations.