Dimension-reduced Interior Point Method

Discussion 2

August 11, 2022

Substituting Conjugate Gradient

More directions are chosen for CG.

$$\mathbf{x}^k + \mathbf{d}^k = \mathbf{x}^k + \alpha^g(\mathbf{A}\mathbf{x}^k - \mathbf{b}) + \alpha^m \mathbf{m}^k$$

If $\mathbf{m}^k \approx \mathbf{x}^* - \mathbf{x}^k$, then $\alpha^m = 1$ solves the system.

Intuitively

- We choose $\hat{\mathbf{x}}$ heuristically to be close to $\mathbf{x}$
- Then take $\mathbf{m}^k = \hat{\mathbf{x}} - \mathbf{x}^k$

  e.g. $\hat{\mathbf{x}} = \mathrm{diag}(\mathbf{A})^{-1}\mathbf{b}$, $\hat{\mathbf{x}} = \mathrm{tridiag}(\mathbf{A})^{-1}\mathbf{b}$
- Return to CG after some steps

A little better than CG wehen $\mathbf{A}$ has special structure (e.g., dominant diagonal).

Potential Reduction Method

Still solving

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\|\mathbf{A}\mathbf{x}\|^2 \quad =: f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{e}^\top \mathbf{x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

with

$$\begin{aligned} \mathbf{d}^k & \leftarrow \alpha^g \mathbf{P}_\Delta[\nabla \varphi(\mathbf{x}^k)] + \alpha^m(\mathbf{x}^k - \mathbf{x}^{k-1}) \\ \mathbf{x}^k & \leftarrow \mathbf{x}^k + \mathbf{d}^k \end{aligned}$$

where $\mathbf{P}_\Delta[\cdot]$ is the orthogonal projection onto $\mathbf{e}^\top \mathbf{x} = 0$. α^g, α^d come from the following model

$$\begin{aligned} \min_{\mathbf{d}, \alpha^g, \alpha^m} \quad & \frac{1}{2}\mathbf{d}^\top \mathbf{H} \mathbf{d} + \mathbf{h}^\top \mathbf{d} \\ \text{subject to} \quad & \|\mathbf{X}^{-1}\mathbf{d}\| \leq \Delta \\ & \mathbf{d} = \alpha^g \mathbf{g}^k + \alpha^m \mathbf{m}^k \end{aligned}$$

Implementation

- Hessian vector product is relatively cheap

$$\nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k) = -\frac{\rho \nabla f(\mathbf{x}^k) \nabla f(\mathbf{x}^k)^\top}{f(\mathbf{x}^k)^2} + \rho \frac{\mathbf{A}^\top \mathbf{A}}{f(\mathbf{x}^k)} + (\mathbf{X}^k)^{-2}$$

- Trust radius β is adjusted by

$$\frac{m^\varphi(\alpha) - m^\varphi(0)}{\varphi(\mathbf{x}^k + \mathbf{d}^k) - \varphi(\mathbf{x}^k)}$$

  and $\beta \le 1$ to ensure feasibility

- Scaling is imposed to enhance stability

$$\begin{aligned} \min_{\mathbf{d},\alpha^g,\alpha^m} \quad & \frac{1}{2}\mathbf{d}^\top \mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}\mathbf{d} + (\mathbf{X}^{-1}\mathbf{h})^\top \mathbf{d} \\ \text{subject to} \quad & \|\mathbf{d}\| \le \Delta \\ & \mathbf{d} = \alpha^g \mathbf{X}\mathbf{g}^k + \alpha^m \mathbf{X}\mathbf{m}^k \end{aligned}$$

- Potential function reduces much faster and more stably

Performance

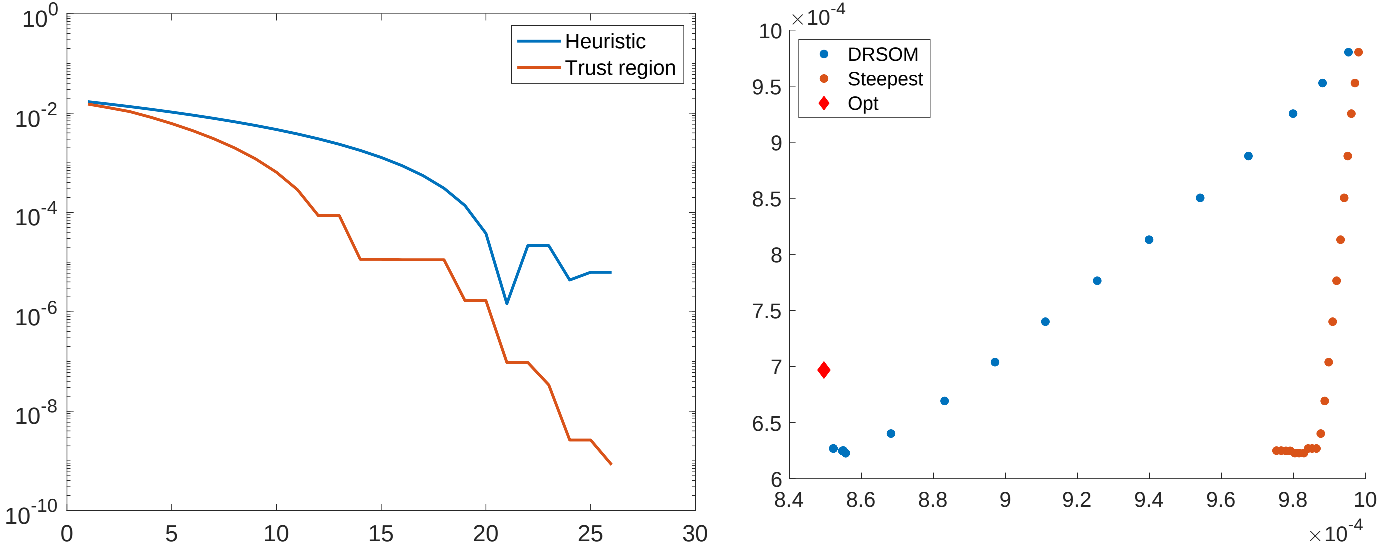

Figure 1. Left: $f(\mathbf{x}^k)$ Right: Momentum might accelerate convergence

- Reaching 1e-06 accuracy is easy on synthetic data

  Around 20 iterations are required

- Momentum plays an important role as $\alpha^m > \alpha^g$.