Dimension-reduced Interior Point Method

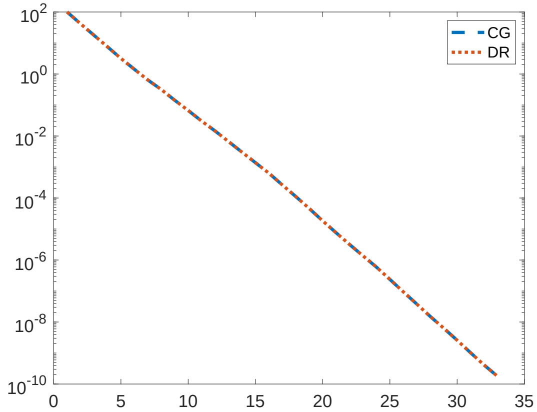
Discussion 1

August 4, 2022

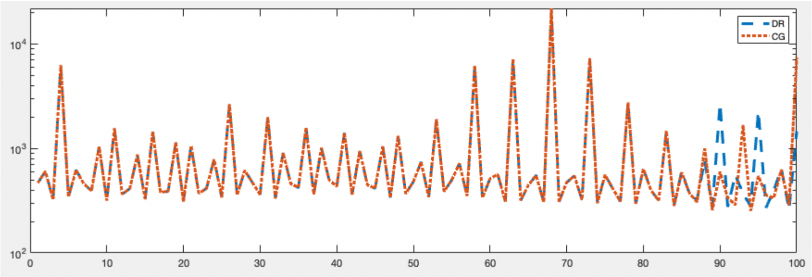
Substituting Conjugate Gradient

One direction applies dimension-reduced method to replace CG to solve $\mathbf{A}\mathbf{x} = \mathbf{b}$.

- The iterations are almost identical when $\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\| \geq 10^{-15}$.



- Sometimes better than CG when the problem is ill-conditioned



Potential Reduction Method

We start from the simple case of simplex-constrained QP

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

and use the potential function

$$\begin{aligned} \varphi(\mathbf{x}) \quad &:= \quad \rho \log\left(f(\mathbf{x})\right) - \sum_{i=1}^{n} \log x_i \\ \nabla \varphi(\mathbf{x}) \quad &= \quad \frac{\rho \nabla f(\mathbf{x})}{f(\mathbf{x})} - \mathbf{X}^{-1}\mathbf{e}. \end{aligned}$$

Potential reduction solves for $\Delta := \mathbf{x}^{k+1} - \mathbf{x}^k$ at each iteration.

$$\begin{aligned} \min_{\Delta} \quad & \langle \nabla \varphi(\mathbf{x}^k), \Delta \rangle \\ \text{subject to} \quad & \mathbf{e}^\top \Delta = 0 \\ & \|(\mathbf{X}^k)^{-1} \Delta\| \leq \beta, \end{aligned}$$

We follow the gradient projection framework and define $\mathbf{P}_{\Delta}[\mathbf{x}] := \left(\mathbf{I} - \frac{\mathbf{e}\mathbf{e}^\top}{\|\mathbf{e}\|^2}\right)\mathbf{x}$. Then we consider

$$\begin{aligned} \mathbf{x}^{k+\frac{1}{2}} &\leftarrow \mathbf{x}^k + \alpha^g \nabla \varphi(\mathbf{x}^k) + \alpha^m(\mathbf{x}^k - \mathbf{x}^{k-1}) \\ \mathbf{x}^{k+1} &\leftarrow \mathbf{P}_{\Delta}\left[\mathbf{x}^{k+\frac{1}{2}}\right], \end{aligned}$$

where α^g, α^d come through

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

$$\mathbf{H} := \begin{pmatrix} \langle \mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{g}^k \rangle & \langle \mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{m}^k \rangle \\ \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{g}^k \rangle & \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{m}^k \rangle \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} \|\mathbf{g}^k\|^2 \\ \langle \mathbf{g}^k, \mathbf{m}^k \rangle \end{pmatrix}$$

and

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

Preliminary observations

- The adaptive trust radius replaces original $\beta^k = \frac{1}{2 + \frac{\rho\gamma}{f(\mathbf{x}^k)}}$
- May not be strictly decreasing for some steps due to large β^k
- Some issues are observed

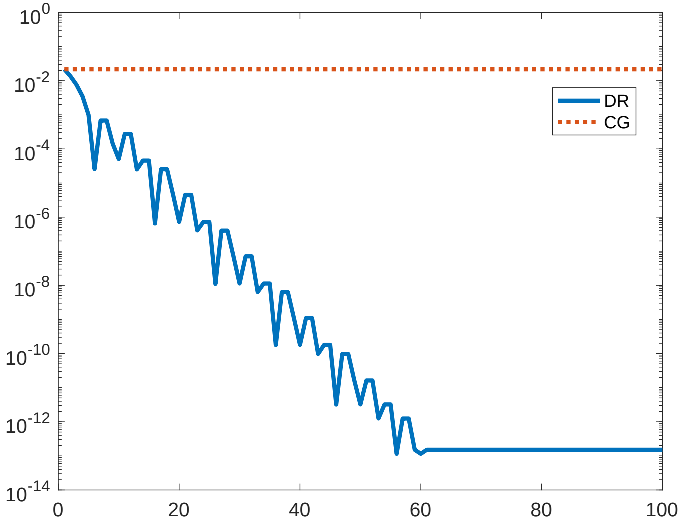


Figure 1. $\|\mathbf{A}^k\mathbf{x}\|^2 \sim k$

Issues with potential reduction

Recall that

$$\mathbf{H} = \begin{pmatrix} \langle \mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}} \varphi(\mathbf{x}^k) \mathbf{g}^k \rangle & \langle \mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}} \varphi(\mathbf{x}^k) \mathbf{m}^k \rangle \\ \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}} \varphi(\mathbf{x}^k) \mathbf{g}^k \rangle & \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}} \varphi(\mathbf{x}^k) \mathbf{m}^k \rangle \end{pmatrix}$$

and

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

$$\mathbf{g}^k = \frac{\rho \nabla f(\mathbf{x}^k)}{f(\mathbf{x}^k)} - (\mathbf{X}^k)^{-1} \mathbf{e}.$$

- The matrix $\mathbf{H}$ is almost always ill-conditioned since $|\langle \mathbf{g}^k, \nabla^2_{\mathbf{x},\mathbf{x}} \varphi(\mathbf{x}^k) \mathbf{g}^k \rangle| \gg 0$

  this drives α^g and α^d imbalanced and subproblem hard to solve
- Often only α^g works and this looks like steepest descent with adaptive β^k.

One possible direction is to consider the projected gradient in the model problem

$$\mathbf{H}^{\mathbf{P}} = \begin{pmatrix} \langle \mathbf{P}_{\Delta}[\mathbf{g}^k], \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{P}_{\Delta}[\mathbf{g}^k]\rangle & \langle \mathbf{P}_{\Delta}[\mathbf{g}^k], \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{m}^k\rangle \\ \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{P}_{\Delta}[\mathbf{g}^k]\rangle & \langle \mathbf{m}^k, \nabla^2_{\mathbf{x},\mathbf{x}}\varphi(\mathbf{x}^k)\mathbf{m}^k\rangle \end{pmatrix}.$$

Or look for other proper ways to replace steepest descent in potential reduction.

The other direction is to consider a more general formulation. e.g., the dual potential function

$$\rho\log(z - \mathbf{b}^\top \mathbf{y}) - \sum_{i=1}^{n} \log(\mathbf{c}_i - \mathbf{a}_i^\top \mathbf{y}).$$