Dimension-reduced Interior Point Method

Discussion 7

September 15, 2022

Current Progress

Theory

- $\mathcal{O}\big(\varepsilon^{-1}\log(\frac{1}{\varepsilon})\big)$ convergence without extra assumption
- $\mathcal{O}\big(\varepsilon^{-3/4}\log(\frac{1}{\varepsilon})\big)$ convergence with assumption on the Hessian

Practice

- The Lanczos solver has been tuned preliminarily
- Now a first order method with accuracy 10^{-5} $(10^{-8}$ if full eigen-decomposition is allowed)

Now transforming `MATLAB` into `C` implementation. (In one or two weeks)

Overview of Dimension-reduced Potential Reduction IPM

We consider the model of simplex-constrained QP

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

in the framework of potential reduction

$$\varphi(\mathbf{x}) := \rho \log\left(f(\mathbf{x})\right) - \sum_{i=1}^{n} \log x_i.$$

We update in a dimension-reduced fashion

$$\begin{aligned} \mathbf{d}^k &\leftarrow \alpha^g \mathbf{g}^k + \alpha^m \mathbf{m}^k \\ \mathbf{x}^k &\leftarrow \mathbf{x}^k + \mathbf{d}^k \end{aligned}$$

Subproblem and Directions

Trust-region controls the accuracy of approximation

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

where

$$\begin{aligned} \mathbf{g} \leftarrow \nabla\varphi(\mathbf{x}) &= \frac{\rho\nabla f(\mathbf{x})}{f(\mathbf{x})} - \mathbf{X}^{-1}\mathbf{e} \\ \mathbf{H} \leftarrow \nabla^2\varphi(\mathbf{x}) &= -\frac{\rho\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top}{f(\mathbf{x})^2} + \rho\frac{\mathbf{A}^\top\mathbf{A}}{f(\mathbf{x})} + \mathbf{X}^{-2} \\ \mathbf{m}^k &\in \{\mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{v}_{\min}^{\mathbf{e}_\perp}(\mathbf{H})\} \end{aligned}$$

The subproblem solvable using bisection/Newton's method

Heuristics

- **Radius adjustment**

  The trust radius $\Delta \leq 1$ is adjusted by

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

  by conventional trust region rules

- **Line-search**

  Line-search procedure is employed to compute

$$\alpha_{\max} = \max\{\gamma : \mathbf{x}^k + \gamma \mathbf{d}^\alpha \geq \mathbf{0}\}$$

  and search is performed over the potential function

Computational Aspects

- The algorithm is sensitive to numerical due to the potential function
- Simplex is magnified by $\mathbf{e}^\top \mathbf{x} = n$
- Hessian $\nabla^2 \varphi(\mathbf{x}^k)$ has special structures to exploit

$$
\begin{aligned}
\langle \mathbf{a}, \nabla^2 \varphi(\mathbf{x}^k)\mathbf{a} \rangle &= \left\langle \mathbf{a}, -\frac{\rho \nabla f(\mathbf{x}^k) \nabla f(\mathbf{x}^k)^\top \mathbf{a}}{f(\mathbf{x}^k)^2} \right\rangle + \frac{\|\mathbf{A}\mathbf{a}\|^2}{f(\mathbf{x}^k)} + \|(\mathbf{X}^k)^{-1}\mathbf{a}\|^2 \\
&= -\rho \left(\frac{\nabla f(\mathbf{x}^k)^\top \mathbf{a}}{f(\mathbf{x}^k)} \right)^2 + \frac{\|\mathbf{A}\mathbf{a}\|^2}{f(\mathbf{x}^k)} + \|(\mathbf{X}^k)^{-1}\mathbf{a}\|^2 \\
\langle \mathbf{a}, \nabla^2 \varphi(\mathbf{x}^k)\mathbf{b} \rangle &= \left\langle \mathbf{a}, -\frac{\rho \nabla f(\mathbf{x}^k) \nabla f(\mathbf{x}^k)^\top \mathbf{b}}{f(\mathbf{x}^k)^2} \right\rangle + \frac{\langle \mathbf{A}\mathbf{a}, \mathbf{A}\mathbf{b} \rangle}{f(\mathbf{x}^k)} + \langle \mathbf{a}, (\mathbf{X}^k)^{-2}\mathbf{b} \rangle \\
&= -\rho \left(\frac{\nabla f(\mathbf{x}^k)^\top \mathbf{a}}{f(\mathbf{x}^k)} \right) \left(\frac{\nabla f(\mathbf{x}^k)^\top \mathbf{b}}{f(\mathbf{x}^k)} \right) + \frac{\langle \mathbf{A}\mathbf{a}, \mathbf{A}\mathbf{b} \rangle}{f(\mathbf{x}^k)} + \langle \mathbf{a}, (\mathbf{X}^k)^{-2}\mathbf{b} \rangle.
\end{aligned}
$$

The most important and ill-conditioned step: negative curvature computation

Negative Curvature Computation

Negative curvature is quite efficient in accelerating convergence but is hard to compute.

- Direct method (cheaper but less stable)

$$\begin{array}{ll} \min\limits_{\|\mathbf{v}\|=1} & \mathbf{v}^\top \left\{ \frac{2\rho \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^2} - \frac{4\rho \mathbf{A}^\top \mathbf{A}\mathbf{u}\mathbf{u}^\top \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\} \mathbf{v} \\ \text{subject to} & \mathbf{e}^\top \mathbf{v}_{\mathbf{x}} = 0. \end{array}$$

- Scaled Hessian (expensive but more stable)

$$\begin{array}{ll} \min\limits_{\|\mathbf{v}\|=1} & \mathbf{v}^\top \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \left\{ \frac{2\rho \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^2} - \frac{4\rho \mathbf{A}^\top \mathbf{A}\mathbf{u}\mathbf{u}^\top \mathbf{A}^\top \mathbf{A}}{\|\mathbf{A}\mathbf{u}\|^4} + \begin{pmatrix} \mathbf{0}_m & \\ & \mathbf{X}^{-2} \end{pmatrix} \right\} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \\ \text{subject to} & \mathbf{x}^\top \mathbf{v}_{\mathbf{x}} = 0. \end{array}$$

- Reduced support (cheap and more stable)

$$\mathbf{v}_k = 0, k \leq \delta$$

Last, a customized Lanczos procedure is employed to solve for $\mathbf{v}$.

Customized Lanczos

A customized Lanczos procedure

$$
\begin{aligned}
\mathbf{x}' &\leftarrow \frac{\mathbf{x}}{\|\mathbf{x}\|} \\
\mathbf{v} &\leftarrow \begin{pmatrix} \mathbf{v_y} \\ \mathbf{v_x} - (\mathbf{x}'^\top \mathbf{v_x})\mathbf{x}' \end{pmatrix} \\
\mathbf{u}_1 &\leftarrow \begin{pmatrix} \mathbf{0} \\ \mathbf{v_x} - (\mathbf{x}'^\top \mathbf{v_x})\mathbf{x}' \end{pmatrix} \\
\mathbf{u}_2 &\leftarrow \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{x}'\mathbf{x}'^\top \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{A}^\top \mathbf{A} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \\
\mathbf{u}_3 &\leftarrow \mathbf{g}^\top \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{v} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{I}_n - \mathbf{x}'\mathbf{x}'^\top \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \\ & \mathbf{X} \end{pmatrix} \mathbf{g} \\
\text{Assemble} &\leftarrow f^2 \mathbf{u}_1 + \rho f \mathbf{u}_2 - \rho \mathbf{u}_3
\end{aligned}
$$

Re-orthogonalization is employed to improve convergence.

Early stop adapted from SDPT3 reduces number of iterations.

Numerical Experiments

Instance	Pinf.	Dinf.	Compl.	Instance	Pinf.	Dinf.	Compl.
ADLITTLE	3.326e-07	5.581e-07	7.617e-06	SC105	3.096e-06	2.351e-06	1.324e-05
AFIRO	3.597e-06	3.487e-06	1.467e-05	SC205	2.591e-05	2.077e-05	9.955e-05
AGG2	4.608e-04	9.277e-05	1.121e-03	SC50A	1.062e-05	5.956e-06	4.219e-05
BANDM	3.869e-06	2.935e-06	2.232e-05	SC50B	1.562e-06	1.259e-06	7.846e-06
BEACONFD	9.495e-07	1.642e-06	1.753e-05	SCAGR25	7.654e-06	4.435e-06	1.075e-04
BLEND	1.479e-06	2.673e-06	9.001e-06	SCAGR7	9.592e-07	3.253e-07	7.192e-06
BOEING2	1.841e-05	1.571e-06	3.407e-05	SCFXM1	3.558e-05	2.766e-05	7.141e-05
BORE3D	2.493e-05	7.667e-05	1.895e-04	SCORPION	1.174e-06	1.328e-06	1.249e-05
BRANDY	3.477e-05	1.398e-05	7.888e-05	SCTAP1	9.530e-07	1.702e-06	9.649e-06
FINNIS	3.468e-05	3.486e-05	3.622e-04	SEBA	2.459e-08	1.014e-07	5.075e-07
FORPLAN	3.323e-05	1.922e-05	1.717e-04	SHARE1B	2.614e-05	2.470e-05	1.034e-04
GFRD-PNC	4.032e-04	1.004e-04	3.425e-04	SHARE2B	8.259e-04	1.968e-04	5.570e-04
GROW7	3.069e-04	4.716e-05	6.817e-04	STAIR	4.136e-04	7.065e-06	2.526e-05
ISRAEL	2.326e-03	1.776e-04	9.293e-04	STANDATA	6.528e-06	9.310e-06	1.734e-04
KB2	4.759e-06	3.436e-05	4.135e-06	STANDGUB	1.175e-04	3.768e-05	6.312e-04
LOTFI	1.445e-06	1.496e-06	3.365e-05	STOCFOR1	2.798e-05	2.346e-05	1.125e-04
MODSZK1	5.552e-05	5.675e-04	2.214e-03	VTP-BASE	1.414e-05	2.472e-06	2.246e-05
RECIPELP	7.721e-06	9.249e-06	1.676e-05				

Table 1. Computation without full eigen-decomposition (1000 iterations)

The method solves LPs to

- $10^{-8} \sim 10^{-10}$ if full-eigen decomposition is allowed
- $10^{-4} \sim 10^{-6}$ if Lanczos is employed

Now optimizing the code and enhancing its implementation