Dimension-reduced Interior Point Method

Discussion 6

September 8, 2022

Negative Curvature of Hessian of Potential

Accelerate computation of negative curvature

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

Observations

- $\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top$ is dense and specially treated
- Customized Lanczos solver behaves 100x more efficient

  than direct eigen-decomposition
- Works fine for problems up to 10^{-3} accuracy
- For higher accuracy, Lanczos might be unstable

Numerical difficulty

The spectrum of ∇^2 is increasingly worse when optimization proceeds

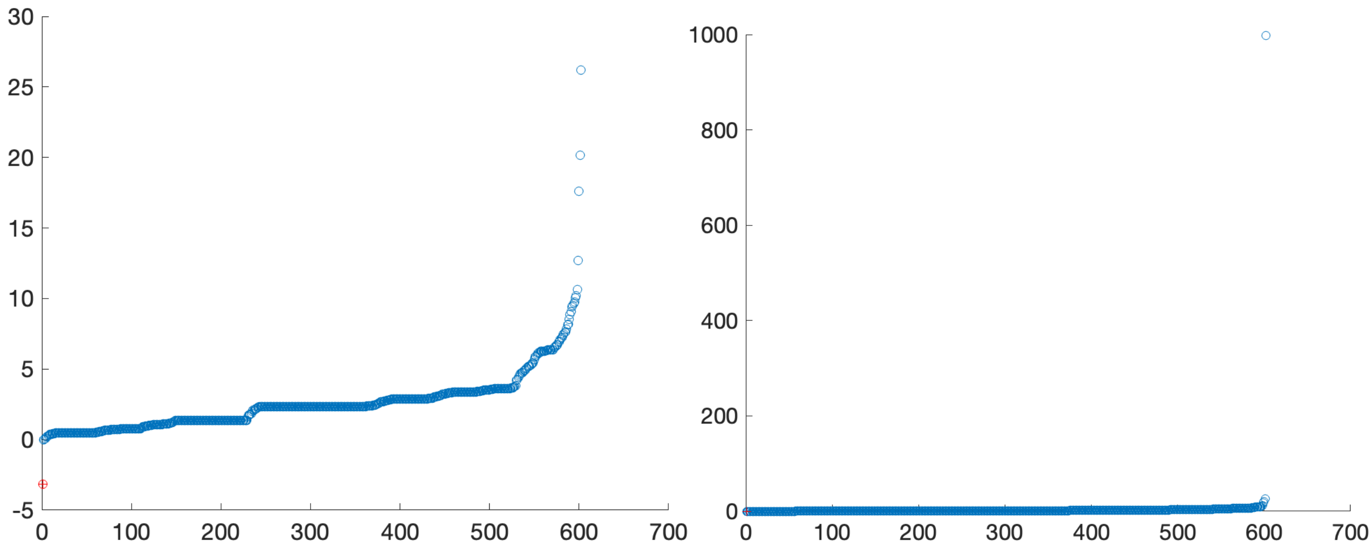

Figure 1. Spectrum at the beginning/at the end, ploted in relative scale

Left: min -9.36×10^5, max 2.24×10^{11}. Right: min -3.18×10^5, max 1.26×10^{14}

The positive eigen-value of the potential function increases and makes negative hard to find.

Improvement by scaling

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

$$\begin{aligned} \min_{\mathbf{v}} \quad & \langle \mathbf{v}, \nabla^2 \mathbf{v}\rangle \\ \text{subject to} \quad & \mathbf{e}^\top\mathbf{v} = 0 \\ & \|\mathbf{v}\| = 1 \end{aligned}$$

An alternative is to use scaled Hessian

$$\begin{aligned} \min_{\mathbf{v}} \quad & \langle \mathbf{v}, \mathbf{X}\nabla^2 \mathbf{X}\mathbf{v}\rangle \\ \text{subject to} \quad & \mathbf{e}^\top\mathbf{X}\mathbf{v} = 0 \\ & \|\mathbf{v}\| = 1 \end{aligned}$$

and the scaled Hessian is

$$-\frac{\rho\mathbf{X}\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top\mathbf{X}}{f(\mathbf{x})^2} + \frac{\rho\mathbf{X}\mathbf{A}^\top\mathbf{A}\mathbf{X}}{f(\mathbf{x})} + \mathbf{I}.$$

Further improvement

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

- The scaled direction makes spectrum more balanced
- Another way is to do diagonal scaling $\mathrm{diag}(\nabla^2_{\mathbf{x}\mathbf{x}}\varphi(\mathbf{x}))$
- Both solving LPs to 10^{-4} to 10^{-5} (but accurate eigen computation gives 10^{-7} to 10^{-8})

More attempts include

Method	Accuracy
Lanczos	10^{-4}
Lanczos + scaling (+ re-orthorgonalization)	$10^{-4} \sim 10^{-5}$
Shifted inverse power method	10^{-5} (solve linear system)
Full eigen-decomposition	$10^{-7} \sim 10^{-8}$

Table 1. Current attempts

One more cure: learning the support of negative curvature

The spectrum of ∇^2 is increasingly worse when optimization proceeds

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

Intuition: $\arg\min_i^k \{x_i\}$ is **not likely** to contribute to negative curvature.

- Use $\mathbf{X}^{-2}$ to predict the support of curvature $\mathbf{v}$
- Truncate α percent of the Hessian by $\mathbf{X}^{-2}$
- Reducing complexity and hardness to evaluate Hessian

Reduced support for Hessian

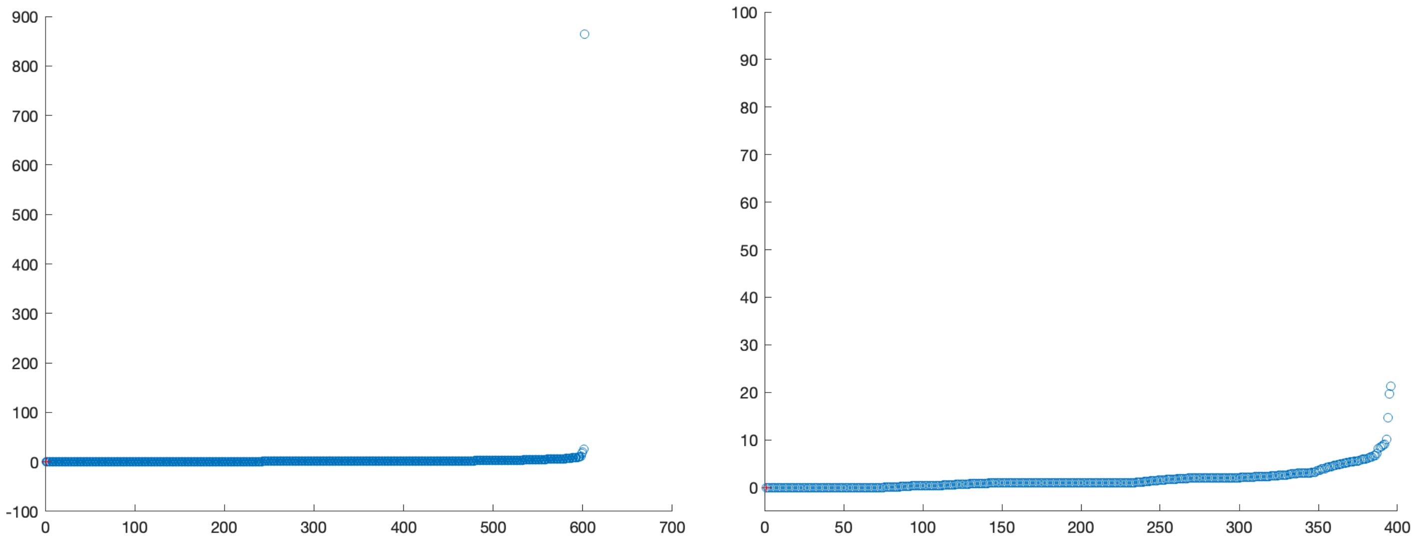

Figure 2. Spectrum of reduced support truncating 50% of dimension

- Retain negative curvature in general
- Need more tuning

Conclusions and future improvement

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

$$\begin{aligned} \min_{\mathbf{v}} \quad & \langle \mathbf{v}, \nabla^2 \mathbf{v} \rangle \\ \text{subject to} \quad & \mathbf{e}^\top \mathbf{v} = 0 \\ & \|\mathbf{v}\| = 1 \end{aligned}$$

- The customization greatly accelerate computation

  100x faster at the beginning
- The spectrum of Hessian is not well-distributed

  Lanczos stagnates since $\lambda_1(\nabla^2) \gg \lambda_{n-k}(\nabla^2)$, for k of interest
- Necessary to look for a method that uses the prior knowledge to accelerate

  We know that there exists exactly one negative eigen-value

  (Now experimenting on FEAST contour method)