Interior-Point Method for Nuclear Norm Approximation with Application to System Identification

Author: Zhang Liu and Lieven Vandenberghe
Presentation by: Martin Hellkvist

May 8, 2019

Summary

- Nuclear Norm: $||X||_* =$ sum of singular values
- $\min ||X||_*$ typically has low rank solutions
  - Can be cast as Semidefinite Program (SDP)
- $\min \operatorname{rank}(X)$ is NP-hard
  - System Identification
  - Machine Learning
  - Computer Vision
- Contribution:
  - Describe a more efficient interior-point method for high order problems with the structure that follows $||\cdot||_*$

Nuclear Norm Approximation Problem

$$
\begin{aligned}
&\text{minimize } ||\mathcal{A}(x) - B||_* \\
&B \in \mathbb{R}^{p \times q} \\
&\mathcal{A}(x) = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n
\end{aligned}
\tag{1}
$$

is convex:

$$
\begin{aligned}
&||\mathcal{A}(\alpha x + (1-\alpha) y) - B||_* = \\
&||\alpha(\mathcal{A}(x) - B) + (1-\alpha)(\mathcal{A}(y) - B)||_* \leq \\
&||\alpha(\mathcal{A}(x) - B)||_* + ||(1-\alpha)(\mathcal{A}(y) - B)||_* = \\
&\alpha||\mathcal{A}(x) - B||_* + (1-\alpha)||\mathcal{A}(y) - B||_*
\end{aligned}
\tag{2}
$$

Nuclear Norm Approximation Problem

Can be cast as

$$
\begin{array}{ll}
\text{minimize} & (\mathbf{tr} U + \mathbf{tr} V)/2 \\
\text{subject to} & \begin{bmatrix} U & (\mathcal{A}(x) - B)^T \\ \mathcal{A}(x) - B & V \end{bmatrix} \succeq 0
\end{array}
\tag{3}
$$

Where $U = U^T \in \mathbb{R}^{q \times q}$ and $V = V^T \in \mathbb{R}^{p \times p}$ are new variables, and the problem is an SDP which is very difficult to solve for $q \approx 100$ and $p \approx 100$ for general purpose interior-point methods where $\mathcal{O}(p^2 q^2 n)$

Custom Interior-Point Method

Exploiting the structure of

$$\begin{array}{ll} \text{minimize} & (\mathbf{tr}\, U + \mathbf{tr}\, V)/2 \\ \text{subject to} & \begin{bmatrix} U & (\mathcal{A}(x) - B)^T \\ \mathcal{A}(x) - B & V \end{bmatrix} \succeq 0 \end{array} \tag{4}$$

We need to solve the following linear equation system

$$\mathcal{A}_{\text{adj}}(\Delta Z) = r, \quad \begin{bmatrix} \Delta U & \mathcal{A}(\Delta x)^T \\ \mathcal{A}(\Delta x) & \Delta V \end{bmatrix} + T \begin{bmatrix} 0 & \Delta Z^T \\ \Delta Z & 0 \end{bmatrix} T = R \tag{5}$$

which can be solved by first finding Δx

$$\tilde{\mathcal{A}}_{\text{adj}}(\mathcal{S}^{-1}(\tilde{\mathcal{A}}(\Delta x))) = \tilde{\mathcal{A}}_{\text{adj}}(\mathcal{S}^{-1}(\tilde{R}_{21})) - r \tag{6}$$

which gives ΔZ and in turn ΔU and ΔV

Key step

Solving for Δx looks complicated

$$\tilde{\mathcal{A}}_{\text{adj}}(\mathcal{S}^{-1}(\tilde{\mathcal{A}}(\Delta x))) = \tilde{\mathcal{A}}_{\text{adj}}(\mathcal{S}^{-1}(\tilde{R}_{21})) - r \tag{7}$$

However ,

$$\mathcal{S}(X) = \mathcal{L}(\mathcal{L}_{\text{adj}}(X)) \tag{8}$$

where the inverses is "simply" defined as:

$$\mathcal{L}_{ij}^{-1} = \begin{cases} (X_{ij} - \sigma_i \sigma_j X_{ji})/\sqrt{1 - \sigma_i^2 \sigma_j^2}, & i < j \\ X_{ii}/\sqrt{1 + \sigma_i^2}, & i = j \\ X_{ij}, & i > j \end{cases} \tag{9}$$

and its adjoint $\mathcal{L}_{\text{adj}}^{-1}$ in a similar fashion

Complexity reduction

- From $\mathcal{O}(p^2q^2n)$ to $\mathcal{O}(pqn^2)$
- Works for large problems

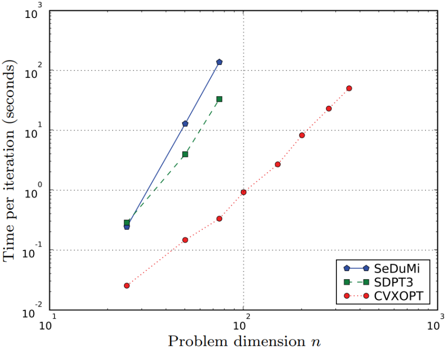

Figure: Solving randomly generated problems for $p = q = n$