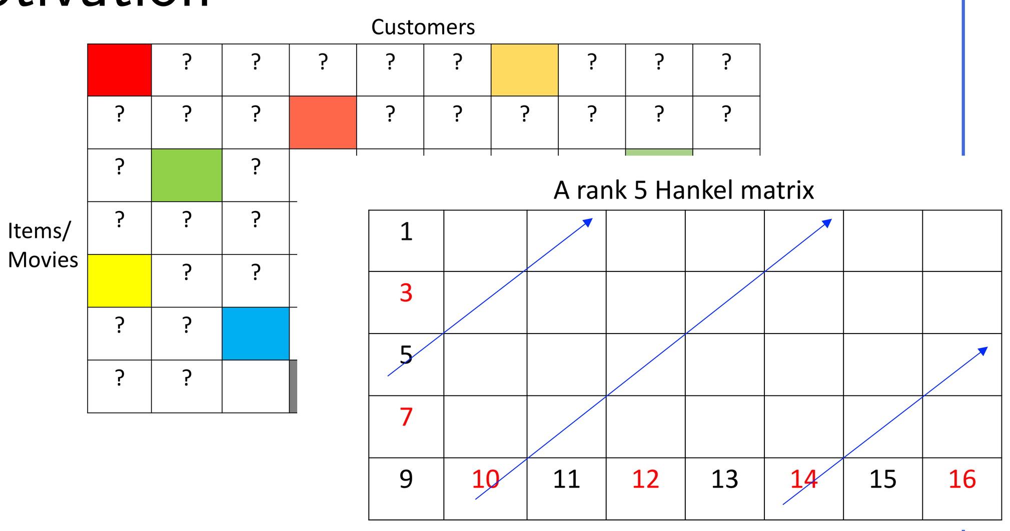
A unified framework for structured low-rank matrix learning

Pratik Jawanpuria and Bamdev Mishra

India Machine Learning Team, Amazon.com, India, {jawanpur, bamdevm}@amazon.com

Motivation



- Low-rank matrix learning with simple constraints.
- Convex loss: Robust / 0-1 matrix completion
- Element-wise: bounded, ordinal, Hankel,...
- $\mathbf{W} \succeq \mathbf{0}$.

Formulation

$$\min_{\mathbf{W}\in\mathbb{R}^{d\times T}} \underbrace{\|\mathbf{W}-\mathbf{Y}\|_F^2}_{\text{Any convex loss function}} + \lambda \underbrace{\|\mathbf{W}\|_*^2}_{\text{Square of trace norm}}$$

$$\text{subject to } \overbrace{\mathbf{A}_t w_t \leq \mathbf{b}}^{\text{Simple convex constraint}}, \quad \mathbf{W} = [w_1, \ldots, w_T], \quad \mathbf{Y} = [y_1, \ldots, y_T]$$

- We exploit a variational characterization of above to separate out low-rank and structural constraints onto different variables.

Contributions

- First generic framework for distinct large scale applications.
- Propose a saddle point formulation, where inner problem: black-box convex solvers outer problem: spherical constraint.
- Parallelizable across columns.
- Computationally efficient for $d \ll T$ or $T \ll d$ instances.
- Code: www.pratikjawanpuria.com

Proposed reformulations

$$\min_{\mathbf{W}\in\mathbb{R}^{d\times T}} \|\mathbf{W}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{W}\|_*^2 \quad \text{subject to } \mathbf{A}_t w_t \leq \mathbf{b}, \ \mathbf{W} = [w_1, \ldots, w_T], \ \mathbf{Y} = [y_1, \ldots, y_T]$$

$$\Longleftrightarrow$$

$$\min_{\Theta\succeq 0, \|\Theta\|_*=1} \min_{\mathbf{W}\in\mathbb{R}^{d\times T}} \|\mathbf{W}-\mathbf{Y}\|_F^2 + trace(\mathbf{W}^\top\Theta^\dagger\mathbf{W}) \quad \text{subject to } \mathbf{A}_t w_t \leq \mathbf{b}, \ \mathbf{W} = [w_1, \ldots, w_T], \ \mathbf{Y} = [y_1, \ldots, y_T]$$

Convex:

$$\overbrace{\min_{\Theta\succeq 0, \|\Theta\|_*=1} \underbrace{\max_{s_t, z_t \forall t} \sum_t y_t^\top z_t - \frac{\lambda}{2} z_t^\top z_t - \frac{1}{2}(z_t + \mathbf{A}_t^\top s_t)^\top \Theta (z_t + \mathbf{A}_t^\top s_t)}_{\text{Application specific algorithm}}}^{\text{Optimization over } \Theta}$$

$$\Theta = \mathbf{U}\mathbf{U}^\top, \mathbf{U} \in \mathbb{R}^{d\times r}, \|\Theta\|_* = 1 \Leftrightarrow \|\mathbf{U}\|_F = 1 \Leftrightarrow \text{Spectrahedron Manifold}$$

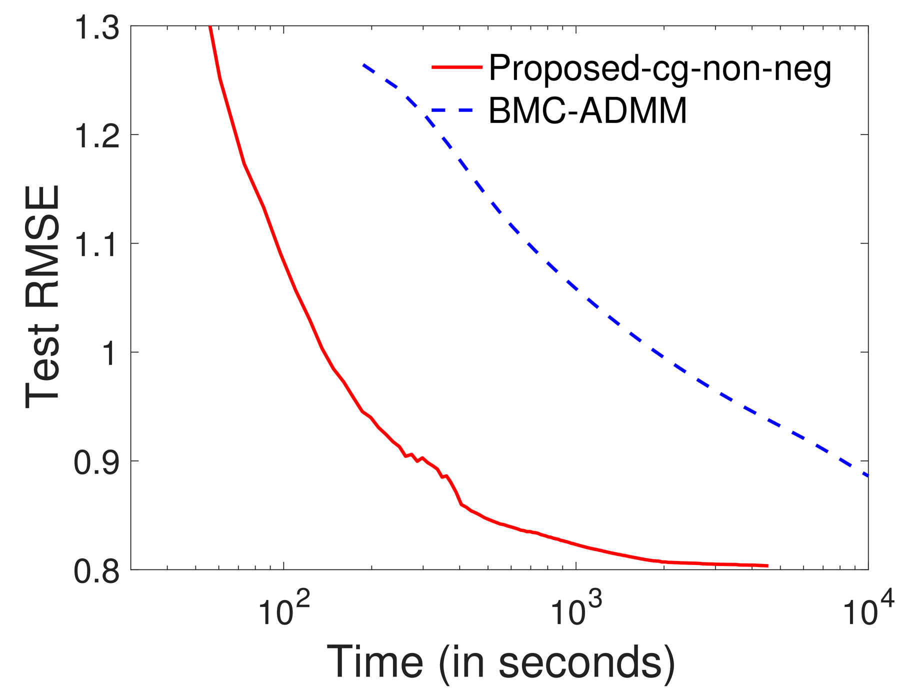
Nonconvex:

$$\overbrace{\min_{\mathbf{U}:\|\mathbf{U}\|_F=1} \underbrace{\max_{s_t, z_t \forall t} \sum_t y_t^\top z_t - \frac{\lambda}{2} z_t^\top z_t - \frac{1}{2}(z_t + \mathbf{A}_t^\top s_t)^\top \mathbf{U}\mathbf{U}^\top (z_t + \mathbf{A}_t^\top s_t)}_{\text{Application specific algorithm}}}^{\text{Optimization over } \mathbf{U}}$$

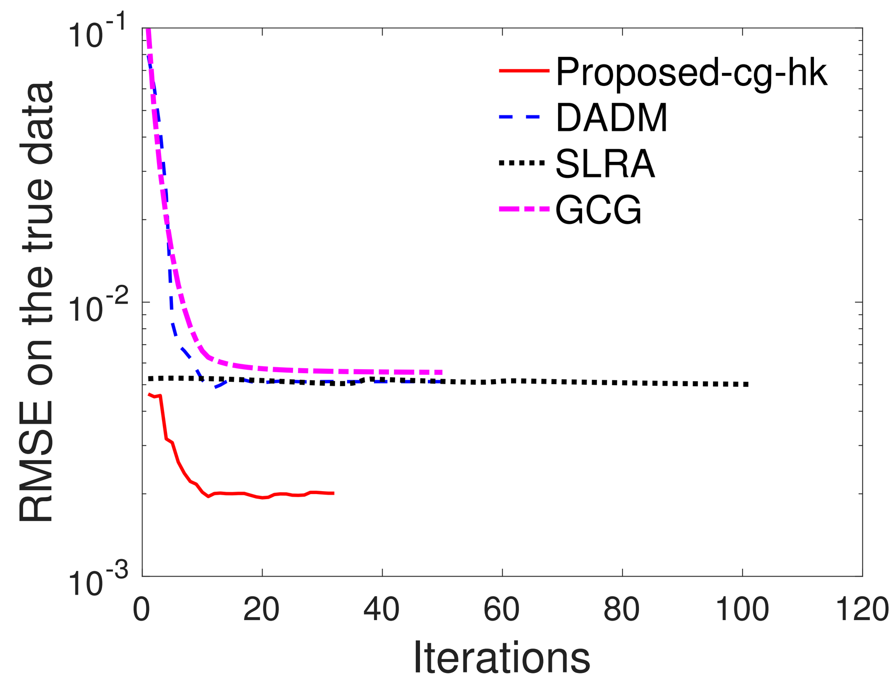
Rank r is chosen depending on computational requirements.
We provide explicit duality gap criterion to check global optimality for $r \leq d$.

Numerical comparisons

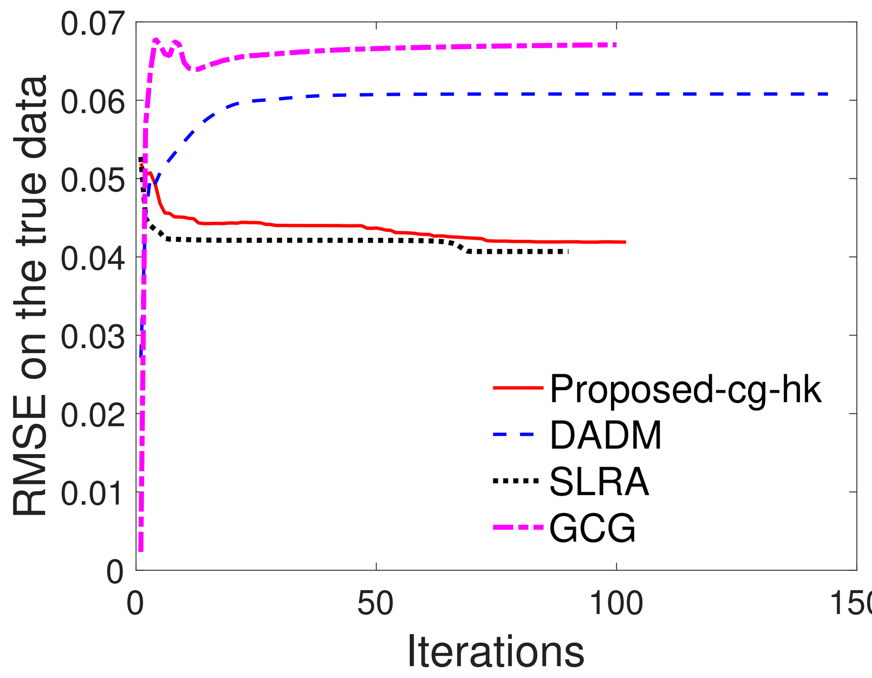
Nonnegative MC MovieLens20M data



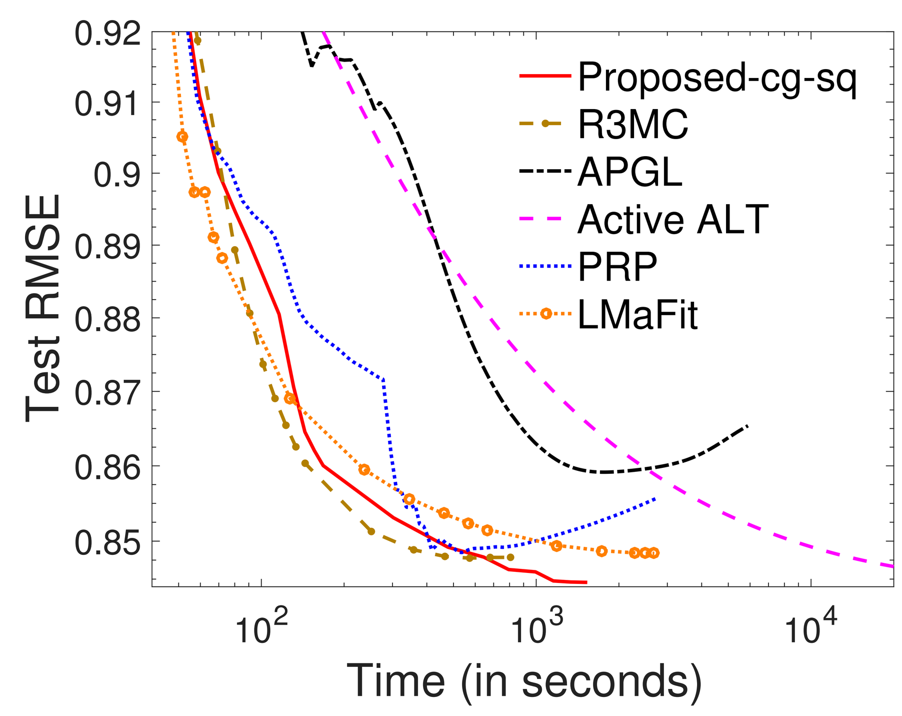
Hankel matrix learning on small-scale data set



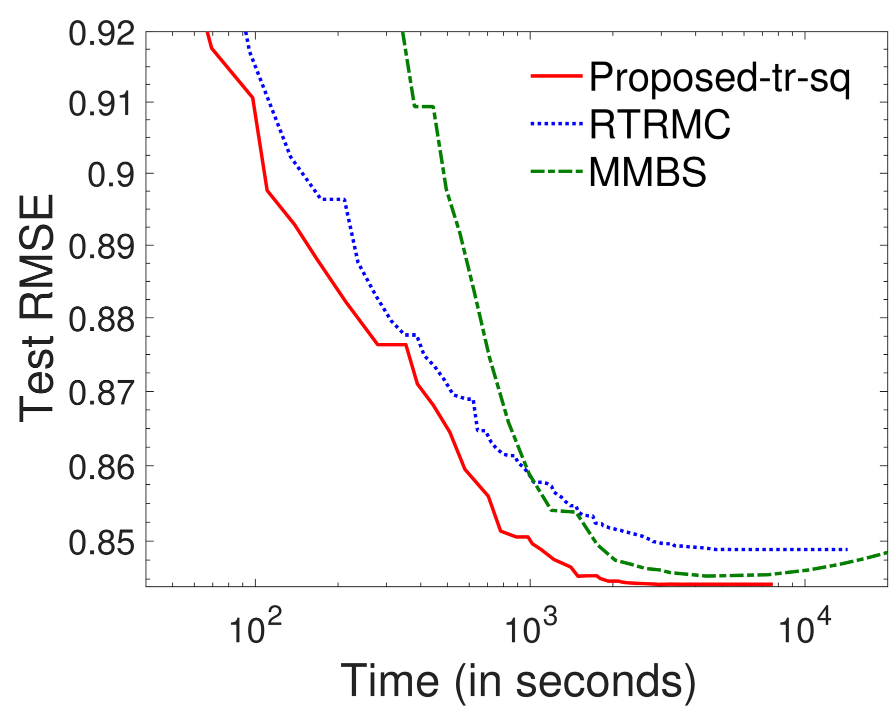
Hankel matrix learning on large-scale data set



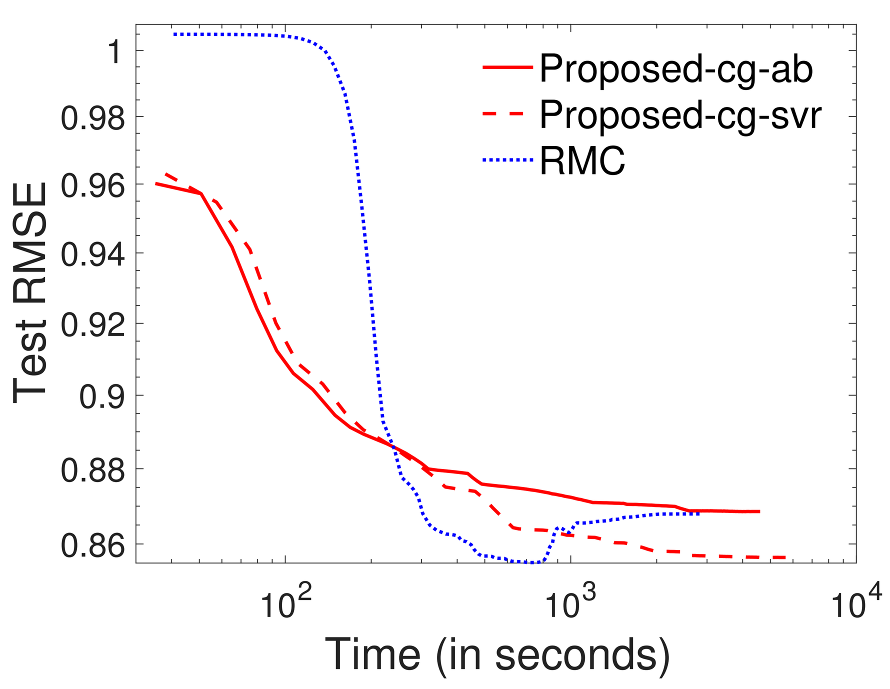
First-order ℓ2-loss algorithms on Netlfix data



Second-order ℓ2-loss algorithms on Netlfix data



ℓ1-loss minimization on Netlfix data



References

[1] P. Jawanpuria and B. Mishra. A saddle point approach to structured low-rank matrix learning in large-scale applications. arXiv:1704.07352, 2017.