Sampling-Based Methods for Multi-Block Optimization Problems over Transport Polytopes

Yukuan Hu

State Key Laboratory of Scientific and Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, China

Joint work with **Mengyu Li** (RUC), **Xin Liu** (AMSS), and **Cheng Meng** (RUC)

The 14th International Conference on Numerical Optimization
and Numerical Linear Algebra (ICNONLA23)
Taiyuan, Shanxi
August 15, 2023

Multi-Block Optimization Problems over Transport Polytopes

$$\min_{X} \; f(X_1, \ldots, X_N), \;\; \text{s. t. } X_i \in \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i), \; i = 1, \ldots, N. \tag{1}$$

- $N \in \mathbb{N}$: number of variable blocks ($N \geq 1$).
- $f : \bigotimes_{i=1}^{N} \mathbb{R}^{m_i \times n_i} \to \mathbb{R}$: objective function.
- $X := (X_1, \ldots, X_N) \in \bigotimes_{i=1}^{N} \mathbb{R}^{m_i \times n_i}$: matrix variables.
- $\mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$: transport polytope ($i = 1, \ldots, N$).

$$\mathcal{U}(\mathbf{a}_i, \mathbf{b}_i) := \{T \in \mathbb{R}_{+}^{m_i \times n_i} \mid T\mathbf{1}_{n_i} = \mathbf{a}_i, \; T^{\top}\mathbf{1}_{m_i} = \mathbf{b}_i\},$$

where $\mathbf{a}_i \in \mathbb{R}_{+}^{m_i}$, $\mathbf{b}_i \in \mathbb{R}_{+}^{n_i}$ and "$\mathbf{1}_n$" is the all-ones vector in $\mathbb{R}^n$.

Applications

- Strongly correlated materials (e.g., transitional metal oxides). [Dagotto 2005]
- Wasserstein barycenters. [Carlier et al. 2015]
- Label distribution learning. [Zhao-Zhou 2018]
- ...

Related Works–General Multi-Block Optimization Problems

Block coordinate descent-type methods

- Representatives:
  - block coordinate descent (BCD) methods;
    [Bertsekas 2016; Wright 2015; Fercoq-Richtárik 2016; Hong et al. 2017; ...]
  - block conditional gradient (BCG) methods;
    [Lacoste-Julien et al. 2013; Beck et al. 2015; Braun et al. 2022; ...]
  - proximal alternating linearized minimization (PALM) methods;
    [Razaviyayn et al. 2013; Xu-Yin 2013; Bolte et al. 2014; Ochs 2019; H.-Liu 2023; ...]
  - stochastic versions (randomness in gradient calculations or update order);
    [Lacoste-Julien et al. 2013; Beck et al. 2015; Chen et al. 2019; Xiao et al. 2019]
    [Sun et al. 2020; Driggs et al. 2021; Hertrich-Steidl 2022; ...]
  - ...
- Formidable memory and computation burdens for large-scale problem (1).
  - In the aforementioned application, $K = 10^4 \sim 10^5$ for a **crude** discretization.

No one wants to store and operate on large-size full matrices directly!

Related Works–Optimal Transport Problems

Kantorovich formulation of general optimal transport (OT) problems

$$\min_{T} \quad \langle T, W \rangle, \quad \text{s. t. } T \in \mathcal{U}(\mathbf{a}, \mathbf{b}).$$

- First proposal, Kantorovich relaxation, theories and applications.
  [Monge 1781; Kantorovich 1942; Villani 2021; ...]
- $W \in \mathbb{R}^{m \times n}$: cost matrix.
- $T \in \mathbb{R}^{m \times n}$: transport plan.
- $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$: discrete marginal probability distributions.

Numerical methods for the OT problems

- Solving differential equations: special cost function.
  [Brenier 1997; Benamou et al. 2002; ...]
- Linear programming methods: cubic complexity.
  [Rubner et al. 1997; Pele-Werman 2009; ...]
- **Entropy regularization-based methods**.
  - Sinkhorn algorithm. [Sinkhorn-Knopp 1967; Cuturi 2013; ...]
  - Variants with partial updates: Greenkhorn, Randkhorn, Screenkhorn, ...
    [Altschuler et al. 2017; Lin et al. 2019; Alaya et al. 2019; Lin et al. 2022; ...]
  - Variants without quadratic complexities: Nys-Sink, Spar-Sink.
    [Kumar et al. 2012; Altschuler et al. 2019; Li et al. 2023; ...]

Related Works–Optimal Transport Problems (Cont.)

Spar-Sink [Li et al. 2023]

- Idea: solving the OT problem **restricted** on the **sampled support**.

$$\min_{T} \quad \left\langle T, \tilde{W} \right\rangle, \quad \text{s. t. } T \in \mathcal{U}(\mathbf{a}, \mathbf{b}),\ T_{\mathcal{I}^c} = 0.$$

  - $\tilde{W} \in \mathbb{R}^{m \times n}$: *effective* cost matrix.
  - $\mathcal{I} \subseteq \{(j,k) \mid j = 1, \ldots, m,\ k = 1, \ldots, n\}$: indices sampled from the beginning according to $\mathbf{a}$ and $\mathbf{b}$; $\mathcal{I}^c$ is its complementary set.
  - $T_{\mathcal{I}^c}$: the entries in T indexed by the set $\mathcal{I}^c$.

  Completely different from the well-known stochastic optimization methods.

- Computational cost in one iteration: $\mathcal{O}(|\mathcal{I}|) = o(mn)$.

It is nontrivial to apply the sampling technique to the multi-block, possibly nonconvex problem (1).

Contributions

- **Novel BCD-type methods combined with the sampling technique.**
  - Ingredients: importance sampling according to previous iterates, subproblems over restricted supports.
  - Complexity per iteration: $o\left(\sum_{i=1}^{N} m_i n_i\right)$.
- **Convergence and asymptotic properties of the proposed methods.**
  - Explicit characterization for sampling-induced errors.
  - Upper bounds for stationarity violations.
  - Best achieved stationarity violation $\to 0$ as $\sum_{i=1}^{N}(m_i + n_i) \to +\infty$.
- **Numerical experiments on strongly correlated electrons systems.**
  - High-quality solutions over finer meshes and higher dimensions.

Subproblems in the BCG Methods

$$\min_{X_i} \left\langle C_i^{(t)}, X_i - X_i^{(t)} \right\rangle, \quad \text{s. t. } X_i \in \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i).$$

- $C_i^{(t)} := \nabla_i f\left(X_{<i}^{(t+1)}, X_{\geq i}^{(t)}\right) \in \mathbb{R}^{m_i \times n_i}$.
- Essentially an OT problem, with $C_i^{(t)}$ as the cost matrix.
- Cubic computational complexity using linear programming methods.
- Quadratic spatial complexity for storing full matrices.

How to reduce the computational complexity?

Entropy regularization-based methods.

How to reduce the spatial complexity?

Sampling techniques.

Entropy Regularized Subproblems

$$\min_{X_i} \left\langle C_i^{(t)}, X_i - X_i^{(t)} \right\rangle + \lambda_i^{(t)} h(X_i), \quad \text{s. t. } X_i \mathbf{1}_{n_i} = \mathbf{a}_i, \; X_i^\top \mathbf{1}_{m_i} = \mathbf{b}_i.$$

- $\lambda_i^{(t)} > 0$: regularization parameter.
- $h(T) := \sum_{jk} t_{jk}(\log t_{jk} - 1)$: negative entropy of any $T = (t_{jk}) \in \mathbb{R}_+^{m_i \times n_i}$.
  (statistical) thermodynamics, quantum statistical mechanics, information theory.
- Early usage in the transportation field. [Wilson 1969]
- Wide adoption in the machine learning community. [Flamary et al. 2021]

Good things about entropy regularization

- Strongly convex problem $\Rightarrow$ differentiability of the optimal value w.r.t. $\mathbf{a}_i$, $\mathbf{b}_i$.
- Upper bound on the optimal value difference (related to $\mathbf{a}_i$, $\mathbf{b}_i$, and $\lambda_i^{(t)}$).
  [Blondel-Seguy-Rolet 2018; Genevay et al. 2019; Kerdoncuff-Emonet-Sebban 2021]
- Convergence of the optimal value and solution as $\lambda_i^{(t)} \to 0$.
  [Cominetti-San Martín 1994]
- Highly scalable iterative schemes for numerical resolution. [Cuturi 2013]
- Multiplicative formula for the optimal solution (given an optimal dual one).

Dual Form

$$\min_{\tilde{\mathbf{u}}_i,\tilde{\mathbf{v}}_i} \; q_i(\tilde{\mathbf{u}}_i,\tilde{\mathbf{v}}_i;\lambda_i^{(t)},\Psi_i^{(t)}) := \lambda_i^{(t)} \exp\left(\frac{\tilde{\mathbf{u}}_i}{\lambda_i^{(t)}}\right)^\top \Psi_i^{(t)} \exp\left(\frac{\tilde{\mathbf{v}}_i}{\lambda_i^{(t)}}\right) - \tilde{\mathbf{u}}_i^\top \mathbf{a}_i - \tilde{\mathbf{v}}_i^\top \mathbf{b}_i,$$

where $\tilde{\mathbf{u}}_i \in \mathbb{R}^{m_i}$, $\tilde{\mathbf{v}}_i \in \mathbb{R}^{n_i}$ are the dual variables,

$$\Psi_i^{(t)} := \exp\left(-C_i^{(t)}/\lambda_i^{(t)}\right) \in \mathbb{R}^{m_i \times n_i}$$

is called the (Gibbs) kernel matrix, "$\exp(\cdot)$" is entrywise exponential.

Relation between the optimal primal and dual solutions

$$X_i^{(t+1,\star)} := \mathrm{Diag}\left(\exp\left(\frac{\tilde{\mathbf{u}}_i^{(t,\star)}}{\lambda_i^{(t)}}\right)\right) \Psi_i^{(t)} \mathrm{Diag}\left(\exp\left(\frac{\tilde{\mathbf{v}}_i^{(t,\star)}}{\lambda_i^{(t)}}\right)\right). \qquad (2)$$

- $X_i^{(t+1,\star)} \in \mathbb{R}^{m_i \times n_i}$: optimal primal solution.
- $(\tilde{\mathbf{u}}_i^{(t,\star)}, \tilde{\mathbf{v}}_i^{(t,\star)}) \in \mathbb{R}^{m_i} \times \mathbb{R}^{n_i}$: optimal dual solution.

Dual BCD Method for the Entropy Regularized Subproblems

$$\tilde{\mathbf{u}}_i^{(t,s+1)} := \lambda_i^{(t)} \log\left(\mathbf{a}_i \oslash \left(\Psi_i^{(t)} \exp\left(\tilde{\mathbf{v}}_i^{(t,s)}/\lambda_i^{(t)}\right)\right)\right),$$
$$\tilde{\mathbf{v}}_i^{(t,s+1)} := \lambda_i^{(t)} \log\left(\mathbf{b}_i \oslash \left(\Psi_i^{(t)\top} \exp\left(\tilde{\mathbf{u}}_i^{(t,s+1)}/\lambda_i^{(t)}\right)\right)\right).$$

- "$\log(\cdot)$": entrywise logarithm; "$\oslash$": entrywise division.

Letting $\check{\mathbf{u}}_i^{(t,s)} := \exp\left(\tilde{\mathbf{u}}_i^{(t,s)}/\lambda_i^{(t)}\right) \in \mathbb{R}^{m_i}$, $\check{\mathbf{v}}_i^{(t,s)} := \exp\left(\mathbf{v}_i^{(t,s)}/\lambda_i^{(t)}\right) \in \mathbb{R}^{n_i}$.

$$\check{\mathbf{u}}_i^{(t,s+1)} := \mathbf{a}_i \oslash \left(\Psi_i^{(t)} \check{\mathbf{v}}_i^{(t,s)}\right), \quad \check{\mathbf{v}}_i^{(t,s+1)} := \mathbf{b}_i \oslash \left(\Psi_i^{(t)\top} \check{\mathbf{u}}_i^{(t,s+1)}\right).$$

- Matrix-vector multiplications: $\mathcal{O}(m_i n_i)$ complexity and high scalability.
- R-linear convergence rate. [Luo-Tseng 1992/1993]
- Warm start $\Rightarrow$ further acceleration.
- Other names: iterative proportional fitting procedure, RAS methods, **Sinkhorn algorithm**, ... [Deming-Stephan 1940; Sinkhorn 1964; Bacharach 1965]

Entropy Regularized Alternating Linearized Minimization Method

Algorithm 1: Entropy regularized alternating linearized minimization (**ERALM**) method.

Input: $X_i^{(0)} \in \mathbb{R}^{m_i \times n_i}$, $\mathbf{a}_i \in \mathbb{R}^{m_i}$, $\mathbf{b}_i \in \mathbb{R}^{n_i}$, $\tilde{\mathbf{v}}_i^{(0)} \in \mathbb{R}^{n_i}$ $(i = 1, \ldots, N)$, $t_{\max} \in \mathbb{N}$.

Set $t := 0$;

while *certain conditions not satisfied* **and** $t < t_{\max}$ **do**

  for $i = 1, \ldots, N$ **do**

    Select a regularization parameter $\lambda_i^{(t)} > 0$ and a step size $\alpha_i^{(t)} \in (0, 1]$;

    Compute $C_i^{(t)}$ and construct $\Psi_i^{(t)} \in \mathbb{R}^{m_i \times n_i}$;

    Starting from $\tilde{\mathbf{v}}_i^{(t)}$, solve the following subproblem

$$\min_{\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i} \; q_i(\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i; \lambda_i^{(t)}, \Psi_i^{(t)})$$

    using Sinkhorn algorithm to obtain $\tilde{\mathbf{u}}_i^{(t+1)} \in \mathbb{R}^{m_i}$ and $\tilde{\mathbf{v}}_i^{(t+1)} \in \mathbb{R}^{n_i}$;

    Update $\tilde{X}_i^{(t+1)} \in \mathbb{R}^{m_i \times n_i}$ as in the equation (2) with $\tilde{\mathbf{u}}_i^{(t+1)}$ and $\tilde{\mathbf{v}}_i^{(t+1)}$;

    Update $X_i^{(t+1)} := (1 - \alpha_i^{(t)}) X_i^{(t)} + \alpha_i^{(t)} \tilde{X}_i^{(t+1)} \in \mathbb{R}^{m_i \times n_i}$;

  end

  Set $t := t + 1$;

end

Output: Approximate solution $X^{(t)} \in \bigotimes_{i=1}^{N} \mathbb{R}^{m_i \times n_i}$.

Waiving the Full Matrices in the ERALM Method

Optimal solution of the subproblem

$$\tilde{X}_i^{(t+1,\star)} := \mathrm{Diag}\left(\exp\left(\frac{\tilde{\mathbf{u}}_i^{(t,\star)}}{\lambda_i^{(t)}}\right)\right)\Psi_i^{(t)}\mathrm{Diag}\left(\exp\left(\frac{\tilde{\mathbf{v}}_i^{(t,\star)}}{\lambda_i^{(t)}}\right)\right) \in \mathbb{R}^{m_i \times n_i},$$

where $(\tilde{\mathbf{u}}_i^{(t,\star)}, \tilde{\mathbf{v}}_i^{(t,\star)}) \in \mathbb{R}^{m_i} \times \mathbb{R}^{n_i}$ is an optimal dual solution.

Observation: $\Psi_{i,jk}^{(t)} = 0 \Rightarrow \tilde{x}_{i,jk}^{(t+1,\star)} = 0.$

Idea: use sparse matrices to approximate $\Psi_i^{(t)}$.

Matrix sparsification by **importance sampling**.

[Liu 1996/2004; Owen 2013]

Importance Sampling

- Optimal sampling probabilities: $p_{i,jk}^{(t,\star)} \propto \tilde{x}_{i,jk}^{(t+1,\star)}$.
- Alternative: $p_{i,jk}^{(t)\prime} \propto x_{i,jk}^{(t)}$.
  reasonable when $X^{(t)}$ is close to an optimal solution.
- Shrinkage strategy: interpolate between $p_{i,jk}^{(t)\prime}$ and $p_{i,jk}^{\prime\prime}$.

$$p_{i,jk}^{(t)} := \gamma p_{i,jk}^{(t)\prime} + (1-\gamma) p_{i,jk}^{\prime\prime} = \gamma \frac{x_{i,jk}^{(t)}}{\sum_{j',k'} x_{i,j'k'}^{(t)}} + (1-\gamma) \frac{\sqrt{a_{i,j} b_{i,k}}}{\sum_{j',k'} \sqrt{a_{i,j'} b_{i,k'}}}, \tag{3}$$

where $\gamma \in [0,1]$ is the interpolation factor. [Ma-Mahoney-Yu 2014; Yu et al. 2022]

Construction of Sparse Kernel Matrices

Let $\mathcal{I}_i^{(t)}$ be the sampled set of indices. The sparse approximation for $\Psi_i^{(t)}$:

$$\hat{\Psi}_{i,jk}^{(t)} := \begin{cases} \Psi_{i,jk}^{(t)} / \left(\left| \mathcal{I}_i^{(t)} \right| \cdot p_{i,jk}^{(t)} \right), & \text{if } (j,k) \in \mathcal{I}_i^{(t)}; \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where the adjustment factors $\left| \mathcal{I}_i^{(t)} \right| \cdot p_{i,jk}^{(t)}$ $(j = 1, \ldots, m_i,\ k = 1, \ldots, n_i)$ ensure the unbiasedness of the random approximation.

Sampling-Based ERALM Method

Algorithm 2: Sampling-based ERALM (**S-ERALM**) method.

Input: $X_i^{(0)} \in \mathbb{R}^{m_i \times n_i}$, $\mathbf{a}_i \in \mathbb{R}^{m_i}$, $\mathbf{b}_i \in \mathbb{R}^{n_i}$, $\tilde{\mathbf{v}}_i^{(0)} \in \mathbb{R}^{n_i}$ $(i = 1, \ldots, N)$, $\gamma \in (0, 1)$, $\{n_{s,i}\}_{i=1}^N \subseteq \mathbb{N}$, $t_{\max} \in \mathbb{N}$.

Set $t := 0$;
while *certain conditions not satisfied* **and** $t < t_{\max}$ **do**
  for $i = 1, \ldots, N$ **do**
    Select a regularization parameter $\lambda_i^{(t)} > 0$ and a step size $\alpha_i^{(t)} \in (0, 1]$;
    Randomly pick a subset $\mathcal{I}_i^{(t)} \subseteq \{(j, k) : j = 1, \ldots, m_i,\ k = 1, \ldots, n_i\}$ according to the probability distribution $P_i^{(t)} = (p_{i,jk}^{(t)}) \in \mathbb{R}^{m_i \times n_i}$ in the equation (3) such that $\left|\mathcal{I}_i^{(t)}\right| = n_{s,i}$;
    Construct the sparse approximate kernel matrix $\hat{\Psi}_i^{(t)} \in \mathbb{R}^{m_i \times n_i}$ in the equation (4);
    Starting from $\tilde{\mathbf{v}}_i^{(t)}$, solve the following subproblem

$$\min_{\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i} \ q_i(\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i; \lambda_i^{(t)}, \hat{\Psi}_i^{(t)})$$

    using Sinkhorn algorithm to obtain $\tilde{\mathbf{u}}_i^{(t+1)} \in \mathbb{R}^{m_i}$ and $\tilde{\mathbf{v}}_i^{(t+1)} \in \mathbb{R}^{n_i}$;
    Update $\tilde{X}_i^{(t+1)} \in \mathbb{R}^{m_i \times n_i}$ as in the equation (2) with $\tilde{\mathbf{u}}_i^{(t+1)}$ and $\tilde{\mathbf{v}}_i^{(t+1)}$;
    Update $X_i^{(t+1)} := (1 - \alpha_i^{(t)}) X_i^{(t)} + \alpha_i^{(t)} \tilde{X}_i^{(t+1)} \in \mathbb{R}^{m_i \times n_i}$;
  end
  Set $t := t + 1$;
end

Output: Approximate solution $X^{(t)} \in \bigotimes_{i=1}^N \mathbb{R}^{m_i \times n_i}$.

Primal Form of the Subproblem in the S-ERALM Method

$$\begin{aligned} \min_{X_i} \quad & \left\langle \hat{C}_i^{(t)}, X_i - X_i^{(t)} \right\rangle + \lambda_i^{(t)} h(X_i), \\ \text{s. t.} \quad & X_i \mathbf{1}_{n_i} = \mathbf{a}_i,\ X_i^\top \mathbf{1}_{m_i} = \mathbf{b}_i,\ (X_i)_{(\mathcal{I}_i^{(t)})^c} = 0, \end{aligned}$$

where $\hat{C}_i^{(t)} = (\hat{c}_{i,jk}^{(t)}) \in \mathbb{R}^{m_i \times n_i}$ is the **effective cost matrix**, defined as

$$\hat{c}_{i,jk}^{(t)} := \begin{cases} c_{i,jk}^{(t)} + \lambda_i^{(t)} \log\left(\left| \mathcal{I}_i^{(t)} \right| \cdot p_{i,jk}^{(t)} \right), & \text{if } (j,k) \in \mathcal{I}_i^{(t)}; \\ c_{i,jk}^{(t)}, & \text{otherwise.} \end{cases}$$

Only the entries indexed by $\mathcal{I}_i^{(t)}$ require to be computed.

Subproblems in the PALM Methods

$$\min_{X_i} \left\langle C_i^{(t)}, X_i - X_i^{(t)} \right\rangle + \frac{\mu_i^{(t)}}{2} \|X_i - X_i^{(t)}\|_{\mathrm{F}}^2, \quad \text{s. t. } X_i \in \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i).$$

- Equivalent to projecting $X_i^{(t)} - C_i^{(t)}/\mu_i^{(t)}$ onto $\mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$;
- Unknown computational complexity for the projection.
- Quadratic spatial complexity for storing full matrices.

How to reduce the complexities?

Kullback-Leibler (KL) divergence + sampling techniques.

KL Divergence-Based Subproblems

$$\begin{aligned} \min_{X_i} \quad & \left\langle C_i^{(t)}, X_i - X_i^{(t)} \right\rangle + \mu_i^{(t)} \mathrm{KL}(X_i; X_i^{(t)}), \\ \text{s. t.} \quad & X_i \mathbf{1}_{n_i} = \mathbf{a}_i,\ X_i^\top \mathbf{1}_{m_i} = \mathbf{b}_i. \end{aligned}$$

- $\mu_i^{(t)} > 0$: proximal parameter.
- $\mathrm{KL}(X_i; X_i^{(t)})$: KL divergence between X_i and $X_i^{(t)}$.

  $$\mathrm{KL}(T; T') := \sum_{j,k} [t_{jk}(\log t_{jk} - \log t'_{jk}) - (t_{jk} - t'_{jk})],\ \ \forall\, T, T' \in \mathbb{R}_+^{m_i \times n_i}.$$

  Bregman distance associated with the negative entropy. [Kullback-Leibler 1951; Bregman 1967]
- The highly scalable Sinkhorn algorithm & multiplicative formula still apply.

  $\Rightarrow$ KL divergence-based alternating linearized minimization (**KLALM**).
- **Difference**: kernel matrix $\Phi_i^{(t)} := \exp\left(-C_i^{(t)}/\mu_i^{(t)}\right) \odot X_i^{(t)} \in \mathbb{R}^{m_i \times n_i}$.

  $\Rightarrow$ importance sampling **only in a critical iteration** $\Rightarrow \hat{\Phi}_i^{(t)}$.

Sampling-Based KLALM Method

Algorithm 3: Sampling-based KLALM (**S-KLALM**) method.

Input: $X_i^{(0)} \in \mathbb{R}^{m_i \times n_i}$, $\mathbf{a}_i \in \mathbb{R}^{m_i}$, $\mathbf{b}_i \in \mathbb{R}^{n_i}$, $\mathbf{v}_i^{(0)} \in \mathbb{R}^{n_i}$ $(i = 1, \ldots, N)$, $\gamma \in (0, 1)$, $\{n_{s,i}\}_{i=1}^N \subseteq \mathbb{N}$, $\hat{t}$, $t_{\max} \in \mathbb{N}$.

Set $t := 0$ and $\mathcal{I}_i := \{(j, k) : j = 1, \ldots, m_i,\ k = 1, \ldots, n_i\}$ $(i = 1, \ldots, N)$;

while *certain conditions not satisfied* **and** $t < t_{\max}$ **do**

  for $i = 1, \ldots, N$ **do**

    Choose proximal parameter $\mu_i^{(t)} > 0$;

    if $t = \hat{t}$ **then**

      Randomly pick a subset $\mathcal{I}_i^{(t)} \subseteq \{(j, k) : j = 1, \ldots, m_i,\ k = 1, \ldots, n_i\}$ according to the probability distribution $P_i^{(t)} = (p_{i,jk}^{(t)}) \in \mathbb{R}^{m_i \times n_i}$ in the equation (3) such that $\left|\mathcal{I}_i^{(t)}\right| = n_{s,i}$;

    end

    if $t < \hat{t}$ **then**

      Let $\hat{\Phi}_i^{(t)} := \Phi_i^{(t)} \in \mathbb{R}^{m_i \times n_i}$;

    else

      Construct the sparse approximate kernel matrix $\hat{\Phi}_i^{(t)} \in \mathbb{R}^{m_i \times n_i}$ in the equation (4) with $\mathcal{I}_i^{(\hat{t})}$ and $P_n^{(\hat{t})}$;

    end

    Starting from $\mathbf{v}_i^{(t)}$, solve the following subproblem

$$\min_{\mathbf{u}_i, \mathbf{v}_i} \ q_i(\mathbf{u}_i, \mathbf{v}_i; \mu_i^{(t)}, \hat{\Phi}_i^{(t)})$$

    using Sinkhorn algorithm to obtain $\mathbf{u}_i^{(t+1)} \in \mathbb{R}^{m_i}$ and $\mathbf{v}_i^{(t+1)} \in \mathbb{R}^{n_i}$;

    Update $X_i^{(t+1)} \in \mathbb{R}^{m_i \times n_i}$ as in the equation (2) with $\mathbf{u}_i^{(t+1)}$ and $\mathbf{v}_i^{(t+1)}$;

  end

  Set $t := t + 1$;

end

Output: Approximate solution $X^{(t)} \in \bigotimes_{i=1}^N \mathbb{R}^{m_i \times n_i}$.

Assumptions

Assumption 1

The function f is block Lipschitz smooth over $\bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$, i.e., there exists an $L \geq 0$ such that, for $i = 1, \ldots, N$,

$$\|\nabla_i f(X) - \nabla_i f(X')\| \leq L\|X - X'\| \text{ for all } X, X' \in \bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i).$$

Assumption 2

(i) *For $i = 1, \ldots, N$, there exist $a_i > 0$, $b_i > 0$, and $q_i > 0$ such that, for any $m_i, n_i \in \mathbb{N}$, $\mathbf{a}_i^\top \mathbf{1}_{m_i} = a_i$, $\mathbf{b}_i^\top \mathbf{1}_{n_i} = b_i$, $\max_j a_{i,j} \leq q_i \cdot \min_j a_{i,j}$, and $\max_k b_{i,k} \leq q_i \cdot \min_k b_{i,k}$.*

(ii) *There exists a $\theta \geq 0$ such that, for any $m_i, n_i \in \mathbb{N}$, the block Lipschitz constant $L \leq \mathcal{O}(\sum_{i=1}^{N} (m_i + n_i)^\theta)$.*

$\Rightarrow \|\mathbf{a}_i\|_\infty = \mathcal{O}(1/m_i)$, $\|\mathbf{b}_i\|_\infty = \mathcal{O}(1/n_i)$, $-h(\mathbf{a}_i \mathbf{b}_i^\top) = \mathcal{O}(\log m_i n_i)$.

Residual Functions

Define the residual functions $R_i : \bigotimes_{i=1}^{N} \mathbb{R}^{m_i \times n_i} \to \mathbb{R}$ $(i = 1, \ldots, N)$ as

$$R_i(X) := \max_{T \in \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)} \langle \nabla_i f(X), X_i - T \rangle .$$

Let $R := \sum_{i=1}^{N} R_i$.

Lemma 1

For any $X \in \bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$, $R(X) \geq 0$ and $X \in \bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$ is a KKT point of the problem (1) if and only if $R(X) = 0$.

$R(X)$ can characterize the stationarity violation at X.

Theorem 1

Let $\{X^{(t)}\}$ *be the sequence generated by the* ERALM *method where the subproblems are exactly solved. Suppose that Assumption 1 holds. Then*

$$0 \leq R(X^{(t+1)}) \leq \frac{f(X^{(t)}) - f(X^{(t+1)})}{\underline{\alpha}^{(t)}} + N\bar{\lambda}^{(t)}\bar{h}\frac{\bar{\alpha}^{(t)}}{\underline{\alpha}^{(t)}} + 2LN(2N+1)\bar{d}^2\frac{\bar{\alpha}^{(t)2}}{\underline{\alpha}^{(t)}}$$

holds for any $t \geq 0$*, where*

$$\bar{\lambda}^{(t)} := \max_{i=1}^{N} \lambda_i^{(t)}, \quad d_i := \max\{\sqrt{m_i}\|\mathbf{a}_i\|_\infty, \sqrt{n_i}\|\mathbf{b}_i\|_\infty\}\ (i = 1, \ldots, N),$$

$$\underline{\alpha}^{(t)} := \min_{i=1}^{N} \alpha_i^{(t)}, \quad \bar{d} := \max_{i=1}^{N} d_i,$$

$$\bar{\alpha}^{(t)} := \max_{i=1}^{N} \alpha_i^{(t)}, \quad \bar{h} := -\min_{i=1}^{N} h(\mathbf{a}_i\mathbf{b}_i^\top).$$

Corollary 1

Let $\{X^{(t)}\}$ *be the sequence generated by the* ERALM *method where the subproblems are exactly solved. Suppose that Assumptions 1 and 2 hold and* $X^{(0)} \in \bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$ *is chosen such that* $f(X^{(0)}) \leq M$*, where* M *is irrelevant to* $\{(m_i, n_i)\}_{i=1}^{N}$*. Also assume that* $t_{\max} = \mathcal{O}(\sum_{i=1}^{N}(m_i + n_i)^{\eta})$ *with* $\eta > \theta$*,*

$$\begin{aligned} \lambda_i^{(t)} \equiv \lambda &:= o\left(\frac{1}{\sum_{i=1}^{N} \log m_i n_i}\right), \\ \alpha_i^{(t)} \equiv \alpha &:= \frac{1}{\bar{d}}\sqrt{\frac{f(X^{(0)}) - \underline{f}}{2LN(2N+1)t_{\max}}} \leq 1, \end{aligned} \tag{5}$$

for any $t \geq 0$ *and* $i = 1, \ldots, N$*, where* $\underline{f} \in \mathbb{R}$ *is not larger than the optimal value of the problem* (1)*. Then* $\min_{t=1}^{t_{\max}} R(X^{(t)}) \to 0$ *as* $\sum_{i=1}^{N}(m_i + n_i) \to +\infty$*.*

Assumption 3

(i) *There exist constants* $\nu \in (1/2, 1]$, c_1, c_2, $\hat{c}_2 > 0$ *such that, for any* $t \geq 0$ *and* $i = 1, \ldots, N$,

$$\|\Psi_i^{(t)}\|_2 \geq \frac{(m_i + n_i)^\nu}{c_1}, \quad \kappa(\Psi_i^{(t)}) \leq c_2, \quad \kappa(\hat{\Psi}_i^{(t)}) \leq \hat{c}_2.$$

(ii) *The interpolation factor* $\gamma < 1$ *and there exists an* $\varepsilon > 0$ *such that, for* $i = 1, \ldots, N$,

$$n_{s,i} \geq \frac{8(m_i + n_i)^{1-2\nu} \log^4(m_i + n_i)}{(1-\gamma) w_i \log^4(1+\varepsilon)},$$

where $w_i := \min_{j,k} \sqrt{a_{i,j} b_{i,k}} / \sum_{j',k'} \sqrt{a_{i,j'} b_{i,k'}}$.

Convergence of the S-ERALM Method (Cont.)

Theorem 2

Let $\{X^{(t)}\}$ *be the sequence generated by the* S-ERALM *method where the subproblems are feasible and exactly solved. Suppose that Assumptions 1 and 3 hold. Then, for any* $\zeta > 0$ *and* $m_i + n_i > 152$ $(i = 1, \ldots, N)$*, with probability no less than* $\prod_{i=1}^{N} \left(1 - 2(m_i + n_i)^{-16\zeta^2/\varepsilon^4}\right)$*,*

$$0 \leq R(X^{(t+1)}) \leq \frac{f(X^{(t)}) - f(X^{(t+1)})}{\underline{\alpha}^{(t)}} + 2N\bar{\lambda}^{(t)}\bar{h}\frac{\bar{\alpha}^{(t)}}{\underline{\alpha}^{(t)}} + 2LN(2N+1)\bar{d}^2\frac{\bar{\alpha}^{(t)2}}{\underline{\alpha}^{(t)}}$$

$$+\bar{\lambda}^{(t)}\bar{d}\frac{\bar{\alpha}^{(t)}}{\underline{\alpha}^{(t)}}\sum_{i=1}^{N}\sqrt{n_{s,i}}\log\frac{1}{(1-\gamma)w_i n_{s,i}} + \bar{\lambda}^{(t)}\frac{\bar{\alpha}^{(t)}}{\underline{\alpha}^{(t)}}\sum_{i=1}^{N}\frac{\hat{c}_2 c_3}{\log^2(m_i+n_i) - c_3}$$

holds for any $t \geq 0$*, where* $c_3 := c_1(1 + \varepsilon + \zeta)\log^2(1 + \varepsilon)$*.*

Convergence of the S-ERALM Method (Cont.)

Corollary 2

Let $\{X^{(t)}\}$ *be the sequence generated by the* S-ERALM *method where the subproblems are feasible and exactly solved. Suppose that Assumptions 1, 2, and 3 hold,* $X^{(0)} \in \bigotimes_{i=1}^{N} \mathcal{U}(\mathbf{a}_i, \mathbf{b}_i)$ *is chosen such that* $f(X^{(0)}) \leq M$*, and* $\max_{i=1}^{N}(m_i + n_i)/\min_{i=1}^{N}(m_i + n_i) \leq r$ *with* $r > 0$*, where* ν*,* c_1*,* c_2*,* c_2'*,* ε*,* M*,* r *are irrelevant to* $\{(m_i, n_i)\}_{i=1}^{N}$*. Also assume that* $t_{\max} = \mathcal{O}(\sum_{i=1}^{N}(m_i + n_i)^{\eta})$ *with* $\eta > \theta$*,*

$$n_{s,i} = \mathcal{O}\left((m_i + n_i)^{3-2\nu} \log^4(m_i + n_i)\right),$$

$$\lambda_i^{(t)} \equiv \hat{\lambda} := o\left(\frac{1}{\sum_{i=1}^{N}(m_i + n_i)^{1.5-\nu} \log^3(m_i n_i)}\right),$$

and $\alpha_i^{(t)} \equiv \alpha$ *as in the formula (5) for any* $t \geq 0$ *and* $i = 1, \ldots, N$*. Then* $\min_{t=1}^{t_{\max}} R(X^{(t)}) \to 0$ *as* $\sum_{i=1}^{N}(m_i + n_i) \to +\infty$ *with probability going to* 1*.*

Application for Strongly Correlated Materials

$$\begin{aligned} \min_{Y} \quad & \sum_{i=2}^{N_e} \langle Y_i, C + \beta I \rangle + \sum_{i<j} \left\langle Y_i, \Lambda^{-1} Y_j C + \beta Y_j \right\rangle, \\ \text{s. t.} \quad & Y_i \in \mathcal{U}(\boldsymbol{\varrho}, \boldsymbol{\varrho}),\ i = 2, \ldots, N_e. \end{aligned} \tag{6}$$

- $N_e \in \mathbb{N}$: number of electrons; $\beta > 0$: penalty parameter.
- $K \in \mathbb{N}$: number of non-overlapping finite elements $\{e_k\}_{k=1}^{K}$ discretizing a bounded domain $\Omega \subseteq \mathbb{R}^d$ ($d \in \{1, 2, 3\}$).
- $\boldsymbol{\varrho} := [\varrho_1, \ldots, \varrho_K]^\top \in \mathbb{R}^K$: density vector, where $\varrho_k := \int_{e_k} \rho(\mathbf{r})\, \mathrm{d}\mathbf{r}$ ($k = 1, \ldots, K$) and $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is the single-particle density; $\Lambda := \mathrm{Diag}(\boldsymbol{\varrho}) \in \mathbb{R}^{K \times K}$.
- $C = (c_{ij}) \in \mathbb{R}^{K \times K}$: Coulomb cost matrix.

$$c_{ij} := \begin{cases} \|\mathbf{d}_i - \mathbf{d}_j\|^{-1}, & \text{if } i \neq j; \\ 0, & \text{otherwise,} \end{cases}$$

  where $\mathbf{d}_k$ is the barycenter of element e_k ($k = 1, \ldots, K$).
- $Y := (Y_2, \ldots, Y_N) \in (\mathbb{R}^{K \times K})^{N_e - 1}$: couplings between the positions of the first and other electrons.

$y_{i,jk} = \text{Prob}\{\text{the first electron lies in } e_j \textbf{ and } \text{the } i\text{th electron lies in } e_k\}$.

Implementation Details

Optimization model

- Equi-mass or equi-size discretization (fixed later).
- Penalty parameter $\beta = 1$. [H. et al. 2023]
- Truncation on $\boldsymbol{\varrho}$: discard the entries smaller than 0.1% of $\|\boldsymbol{\varrho}\|_\infty$. [Dvurechensky-Gasnikov-Kroshnin 2018]

Outer loop

- Step sizes $\alpha_i^{(t)} = 1/(t+1)$ (ERALM and S-ERALM).
- Interpolation factor $\gamma = 0.99$ (S-ERALM and S-KLALM).
- Sample size $n_{s,i} = \lfloor K^{1.5} \rfloor$ (S-ERALM and S-KLALM).
- Critical iteration number $\hat{t} = 0$ (S-KLALM).
- Regularization/proximal parameters

$$\lambda_i^{(t)} = \|\tilde{\mathbf{v}}_i^{(t)}\|_\infty/(20\log(K)), \quad \mu_i^{(t)} = \|\mathbf{v}_i^{(t)}\|_\infty/(20\log(K)).$$

Implementation Details (Cont.)

Outer loop (cont.)

- Stopping rules: one of the following three holds.
  - $\Delta^{(t)} := \sum_{i=2}^{N_e} \|\Lambda^{-1}(Y_i^{(t)} - Y_i^{(k-1)})\|/(N_e - 1) \leq tol.$
  - $t \geq t_{\max}.$
  - CPU time in seconds $\leq T_{\max}.$

Inner loop

- Warm start with the previous dual iterates.
- Stopping rules: $\|Y_i^{(t+1)}\mathbf{1}_K - \boldsymbol{\varrho}\|_\infty \leq 10^{-6}$ or $s \geq s_{\max} = 20.$

Running environment

- CPU: Intel Xeon Gold 6242R CPU @ 3.10GHz.
- RAM: 510GB.
- Operating system: Ubuntu 20.04.
- Software: MATLAB R2019b.

Metrics of Interest

(i) Converged objective value (obj).

(ii) Approximate SCE potential $\mathbf{v}$. [Chen et al. 2014; H. et al. 2023]

Taking the ERALM *method for example,* $\mathbf{v} := \tilde{\mathbf{v}} - \min_{k=1}^{K}\{\tilde{v}_k\} \cdot \mathbf{1}_K \in \mathbb{R}^K$, *where*

$$\tilde{\mathbf{v}} := \frac{1}{N_e - 1} \sum_{i=2}^{N_e} \tilde{\mathbf{v}}_i \in \mathbb{R}^K$$

and $\{\tilde{\mathbf{v}}_i\}_{i=2}^{N_e}$ *are the dual solutions given by the Sinkhorn algorithm.*

(iii) Errors of objective value (err_obj) and SCE potential (err_sce).[1)]

$$\mathrm{err\_obj} := \left| \frac{f - f^\star}{f^\star} \right|, \quad \mathrm{err\_sce} := \frac{\|\mathbf{v} - \mathbf{v}^\star\|_1}{\|\mathbf{v}^\star\|_1},$$

where f *and* $f^\star \in \mathbb{R}$ *denote respectively the converged and optimal objective values of the problem* (6), $\mathbf{v}^\star \in \mathbb{R}^K$ *refers to the vector made up by the values of the SCE potential at barycenters.*

(iv) CPU time in seconds (T).

[1)] Available only if there admit explicit constructions of the optimal solutions and the SCE potentials, e.g., in 1D settings. [Colombo-De Pascale-Di Marino 2015; H.-Liu 2022]

Model Systems

Table 1: Information on the model systems.

System No.	Unnormalized single-particle densities ρ	Domains Ω	#electrons N_e
	1D systems		
1	$\cos(\pi x) + 1$	$[-1, 1]$	3
2	$2e^{-6(x+0.5)^2} + 1.5e^{-4(x-0.5)^2}$	$[-1.5, 1.5]$	3
3	$e^{-x^2/\sqrt{\pi}}$	$[-2, 2]$	7
4	$e^{-4(x+2)^2} + e^{-4(x+1.5)^2} + e^{-4(x+1)^2} + e^{-4(x+0.5)^2}$ $+e^{-4(x-2/3)^2} + e^{-4(x-4/3)^2} + e^{-4(x-2)^2}$	$[-3, 3]$	7
	2D systems		
5	$e^{-3(x^2+(y-0.96)^2)} + e^{-3((x-1.032)^2+(y+0.84)^2)}$ $+e^{-3((x+1.032)^2+(y+0.84)^2)}$	$[-2.5, 2.5]^2$	3
6	$2e^{-3(x^2+(y-1.2)^2)} + e^{-3((x-1.29)^2+(y+1.05)^2)}$ $+e^{-3((x+1.29)^2+(y+1.05)^2)}$	$[-3, 3]^2$	4
	3D systems		
7	$e^{-3((x+1)^2+(y+1)^2+(z+1)^2)} + e^{-3((x-1)^2+(y-1)^2+(z+1)^2)}$ $+e^{-3((x+1)^2+(y-1)^2+(z-1)^2)}$	$[-2, 2]^3$	3
8	$3e^{-4((x+1)^2+y^2+z^2)} + e^{-4((x-1)^2+y^2+z^2)}$	$[-2, 2] \times [-1, 1]^2$	4

Comparison Among Algorithms

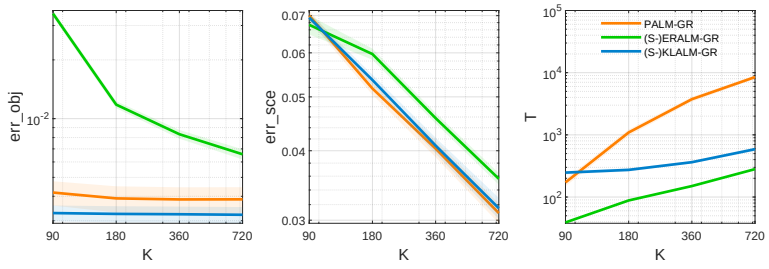
Algorithms

- **PALM-GR** = PALM + GR-based framework [H. et al. 2023].
- **S-ERALM-GR** = ERALM + S-ERALM + GR-based framework.
- **S-KLALM-GR** = KLALM + S-KLALM + GR-based framework.

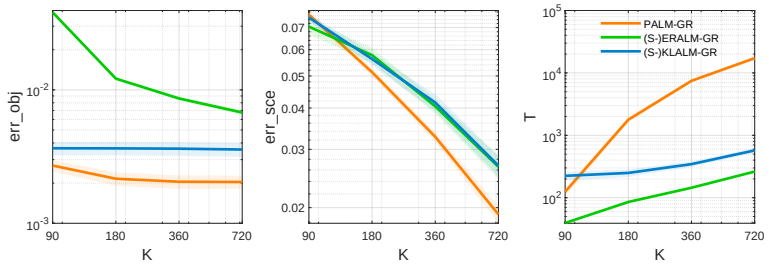
Settings

- Systems 1-4: equi-mass discretization ($\ell = 0$) / refinements ($\ell > 0$).
  - Systems 1 and 2: $K^{(0)} = 90 \xrightarrow[\text{grid refinements}]{\text{3 times}} K^{(3)} = 720.$
  - Systems 3 and 4: $K^{(0)} = 140 \xrightarrow[\text{grid refinements}]{\text{3 times}} K^{(3)} = 1120.$
- For each system, invoke the three algorithms with 10 trials.
- Stopping rules: $tol = 10^{-3} \cdot \sqrt{2}^{\log_2(K/K^{(0)})}$, $t_{\max} = 10^4$, $T_{\max} = 10^5$.

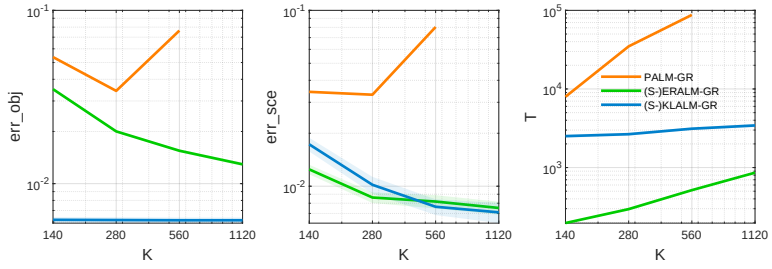
Comparison Among Algorithms (Cont.)



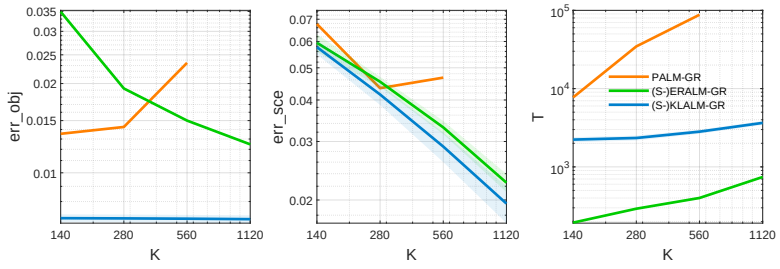
(a) System 1 ($N = 3$)



(b) System 2 ($N = 3$)



(c) System 3 ($N = 7$)



(d) System 4 ($N = 7$)

Fig. 1. $\mathrm{err\_obj}$, $\mathrm{err\_sce}$, and T of the PALM-GR, S-ERALM-GR, and S-KLALM-GR methods on the 1D systems with different K.

The S-KLALM-GR method stands out with

- comparable (or better) $\mathrm{err\_obj}$ and $\mathrm{err\_sce}$ to (than) the PALM-GR method;
- comparable $\mathrm{err\_sce}$ and T to the S-ERALM-GR method.

Simulations on 2D/3D Systems

Settings

- Systems 5-8: equi-size discretization ($\ell = 0$) / uniform refinements ($\ell > 0$).
  - Systems 5 and 6 (2D): $K^{(0)} = 900 \xrightarrow[\text{grid refinements}]{\text{3 times}} K^{(3)} = 57600.$
  - System 7 (3D): $K^{(0)} = 1728 \xrightarrow[\text{grid refinements}]{\text{2 times}} K^{(2)} = 110592.$
  - System 8 (3D): $K^{(0)} = 1000 \xrightarrow[\text{grid refinements}]{\text{2 times}} K^{(2)} = 64000.$
- Stopping rules: $t_{\max} = 10^4$, $T_{\max} = +\infty$,

$$tol = \begin{cases} 5 \times 10^{-3}, & \ell = 0; \\ 10^{-2} \times (\sqrt{2^d})^{\log_2(K/K^{(0)})}, & \ell > 0. \end{cases}$$

- Mappings between electron positions $\{\mathcal{T}_i\}_{i=2}^{N_e}$.

$$\mathcal{T}_i(\mathbf{d}_i) := \sum_{1 \le k \le K} y_{i,jk} \mathbf{d}_k / \varrho_j, \ j = 1, \ldots, K, \ i = 2, \ldots, N_e.$$

Simulations on 2D/3D Systems (Cont.)

Table 2: obj given by the S-KLALM-GR method when simulating the 2D/3D systems. K_{trunc} refers to #entries in $\boldsymbol{\varrho}$ that are larger than $0.999\|\boldsymbol{\varrho}\|_\infty$.

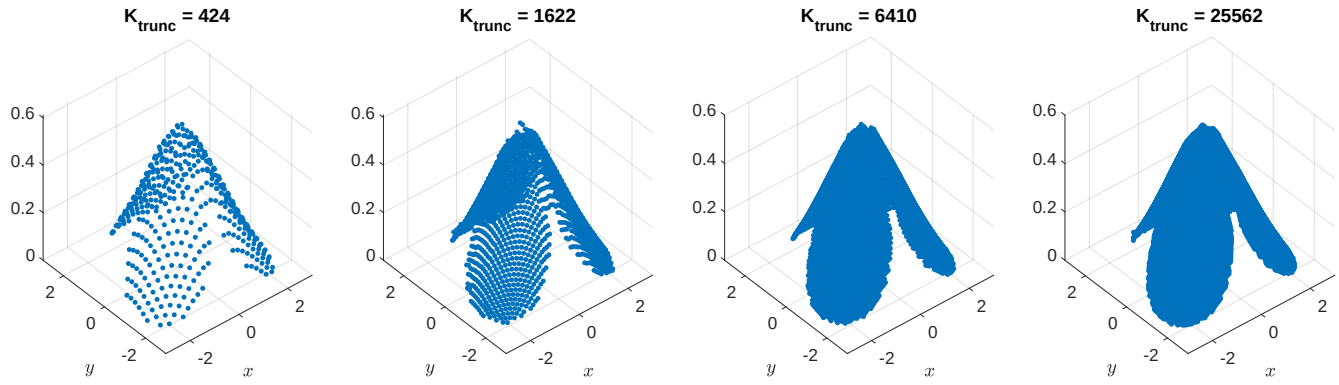
Step	System 5			System 6		
	K	K_{trunc}	obj	K	K_{trunc}	obj
0	900	424	1.1339	900	408	3.0690
1	3600	1622	1.1337	3600	1534	3.0690
2	14400	6410	1.1335	14400	6068	3.0677
3	57600	25562	1.1334	57600	24176	3.0667

(a) 2D systems

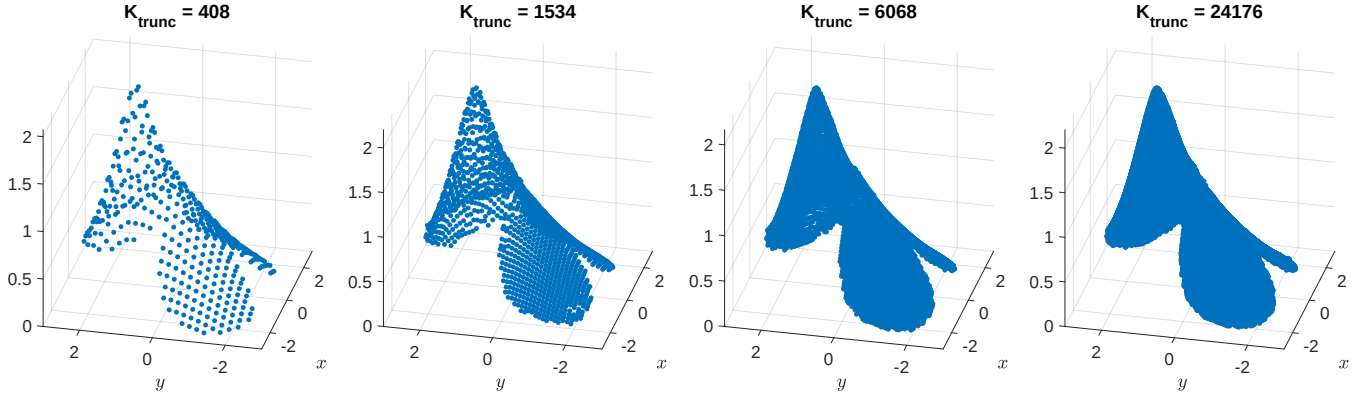
Step	System 7			System 8		
	K	K_{trunc}	obj	K	K_{trunc}	obj
0	1728	780	1.0202	1000	720	4.6193
1	13824	5628	1.0209	8000	5272	4.6716
2	110592	42936	1.0209	64000	40764	4.6833

(b) 3D systems

Simulations on 2D/3D Systems (Cont.)



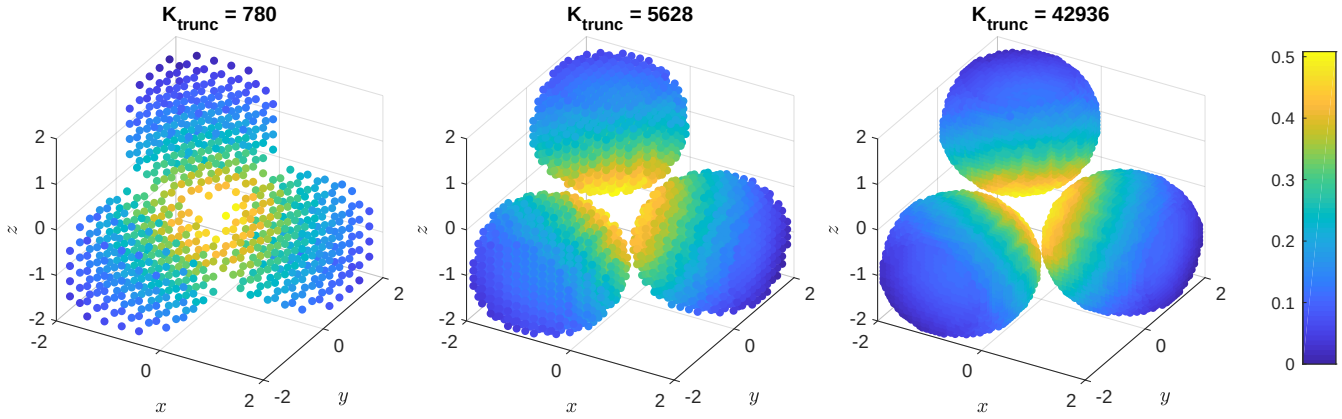
(a) System 5



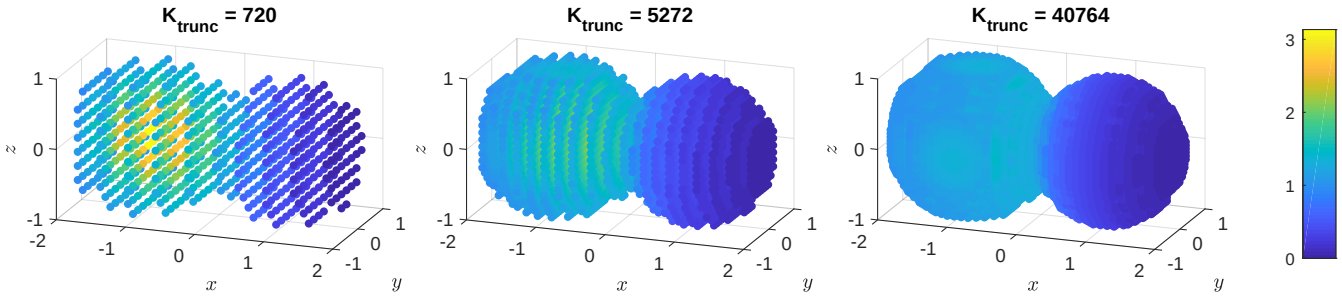
(b) System 6

Fig. 2. Approximate SCE potentials given by the S-KLALM-GR method on the 2D systems.

Simulations on 2D/3D Systems (Cont.)



(a) System 7



(b) System 8

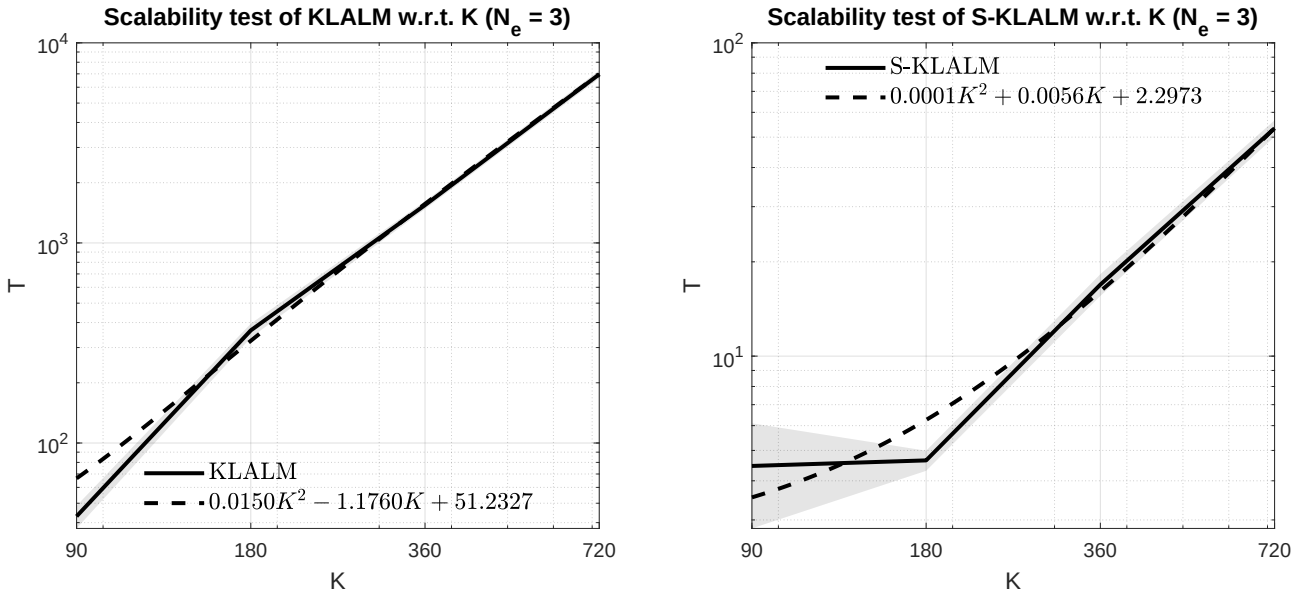
Fig. 3. Approximate SCE potentials given by the S-KLALM-GR method on the 3D systems.

Scalability Tests

Settings

- $\rho(x) \propto \cos(\pi x) + 1$, equi-mass discretization.
- Algorithms: the KLALM and S-KLALM methods.
- **Scalability w.r.t. K:**
  - $K = 90, 180, 360, 720, 1440, 2880$, fix $N_e = 3$.
  - Stopping rules: $tol = 10^{-3} \times \sqrt{2}^{\log_2(K/90)}$, $t_{\max} = 10^4$, $T_{\max} = +\infty$.
- **Scalability w.r.t. N_e:**
  - $N_e = 3, 6, 12, 24, 48$, fix $K = 144$.
  - Stopping rules: $tol = 10^{-3}$, $t_{\max} = 10^4$, $T_{\max} = +\infty$.
- Each K or N_e corresponds to 10 random trials.

Scalability Tests w.r.t. K

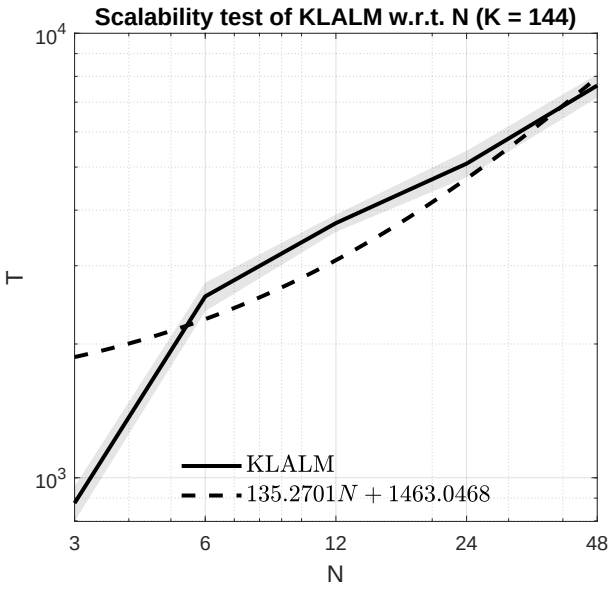


(a) Scalability of KLALM w.r.t. K (b) Scalability of S-KLALM w.r.t. K

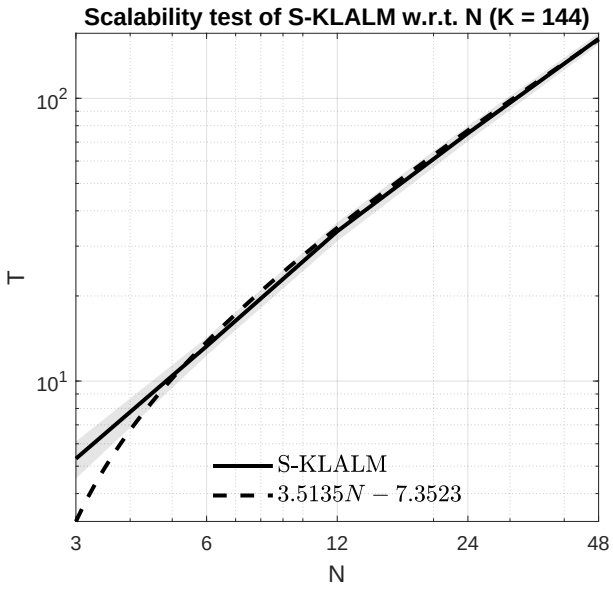
Fig. 4. Scalability tests of the KLALM and S-KLALM methods w.r.t. K.

Scalability w.r.t. $K \sim \mathcal{O}(K^2)$.

Scalability Tests w.r.t. N_e



(a) Scalability of KLALM w.r.t. N



(b) Scalability of S-KLALM w.r.t. N

Fig. 5. Scalability tests of the KLALM and S-KLALM methods w.r.t. N_e.

Scalability w.r.t. $N_e \sim \mathcal{O}(N_e)$.

Conclusions and Future Work

Conclusions

(i) Sampling-based BCD-type methods for the problem (1).
  - Highly scalable schemes for subproblems.
  - Considerable saving using random matrix sparsification.

(ii) Theoretical properties for the ERALM and S-ERALM methods.

(iii) Numerical simulations on several model strongly correlated systems.
  - Better scalabilities of the S-ERALM, S-KLALM methods than the PALM method.

Future work

- Theoretical properties for the KLALM and S-KLALM methods.
- Support identification for further acceleration.

Thanks for your attentions!

Email: ykhu@lsec.cc.ac.cn