SPLR: The Sparse-Plus-Low-Rank Quasi-Newton Method for Entropic-Regularized Optimal Transport

Chenrui Wang[1], Yixuan Qiu[1]

[1] Shanghai University of Finance and Economics

Problem Definition

Our goal is to solve the entropic-regularized optimal transport problem, and our approach is to study its dual problem. After removing one redundant degree of freedom, we obtain the *smooth* and *unconstrained convex* optimization problem:

$$\begin{aligned} f(x) =& -\mathcal{L}(\alpha, \beta) \\ =& \eta \sum \exp\{\eta^{-1}(\alpha_i + \beta_j - M_{ij})\} \\ & - \alpha^T a - \beta^T b, \end{aligned}$$

where $x = (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_{m-1})^T$.

Motivation

The main idea of SPLR is to provide an *accurate approximation* of the Hessian matrix while maintaining its *sparsity* as much as possible:

- **Sparse:** Some second-order methods have been proposed based on the idea of Hessian sparsification. SPLR also obtain an approximation of the true Hessian matrix by sparsification.
- **Low-Rank:** Similar to the BFGS update rule, we incorporate a low-rank correction term to enhance the approximation quality.

Contribution

Our main contributions are summarized as follows:

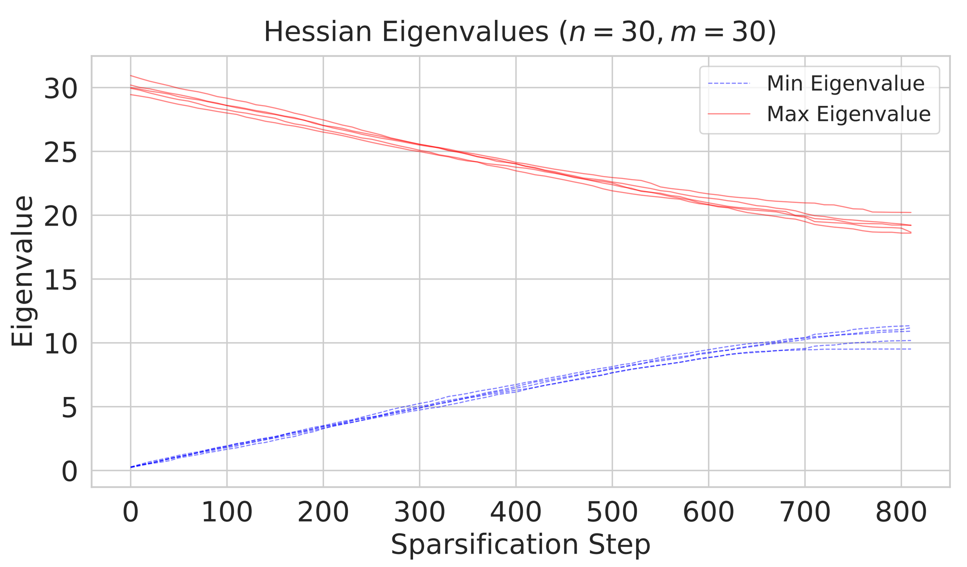
- New theoretical results are developed to understand Hessian sparsification.
- A new quasi-Newton method is proposed to solve entropic-regularized OT with *super-linear-like* convergence speed in practice.
- We provide convergence guarantees for the proposed method: SPLR enjoys a *global convergence* property and the convergence speed is at least *linear*.
- We conduct extensive numerical experiments to demonstrate the performance of SPLR on various entropic-regularized OT problems.

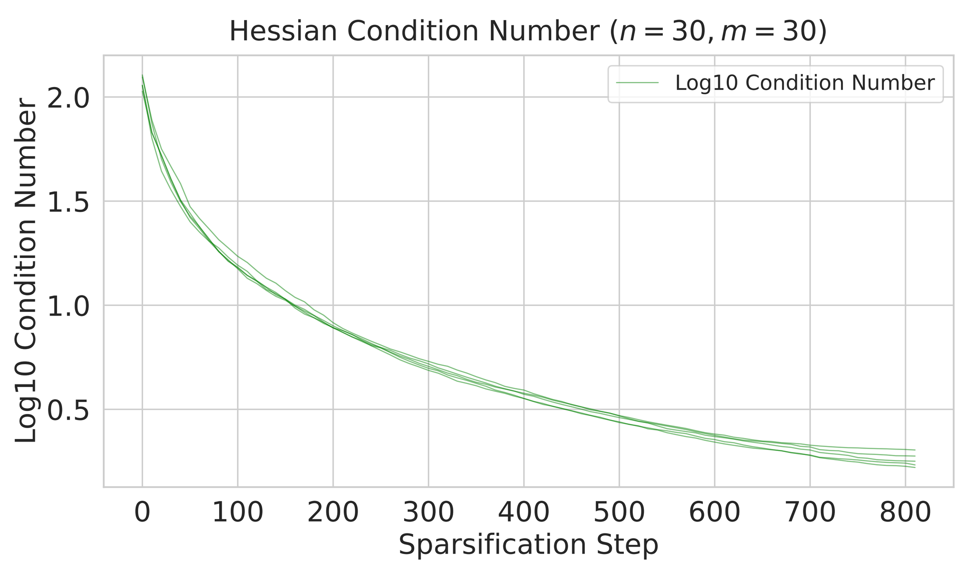
Sparsification Effect

Our sparsification scheme is designed based on the structure of the Hessian matrix, which preserves its diagonal elements and symmetry. The impact of this sparsification on the eigenvalues of the Hessian matrix is then analyzed under a *very weak* assumption, leading to the following results:

- *Any* valid sparsification scheme maintains positive definiteness.
- The sparsified Hessian matrix is guaranteed to have a smaller condition number. In particular, the biggest eigenvalue is reduced and the smallest eigenvalue is increased.
- The assumption is valid for almost *any* sparsification scheme, allowing for highly flexible algorithmic designs and *extremely sparse* matrices.

The assumption merely requires the existence of a power of the sparsified Hessian matrix in which all entries are strictly positive. A numerical verification is shown below:





Update Rule

Suppose that we are at the $(k+1)$-th iteration of the optimization procedure. We approximate H_{k+1} by a matrix B_{k+1} of the form:

$$H_{k+1} \approx B_{k+1} := (H_\Omega^{k+1}) + (auu^T + bvv^T) + (\tau_{k+1} I),$$

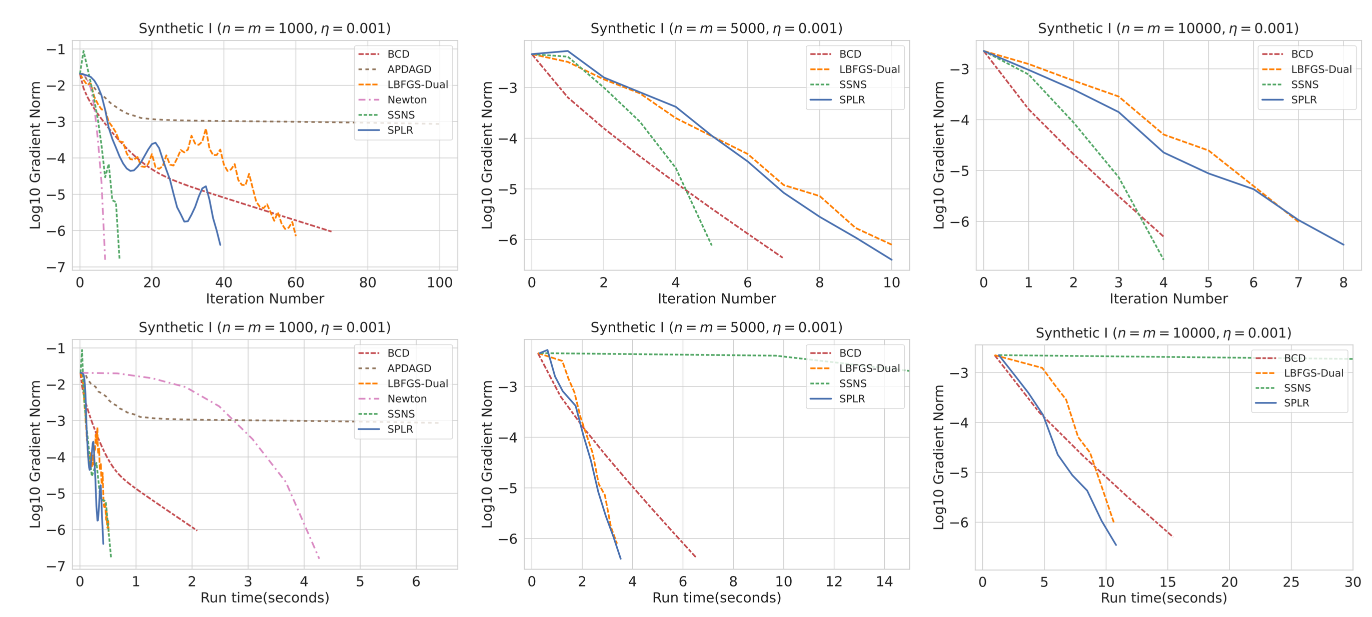
where H_Ω is the sparsified Hessian matrix according to some sparsification scheme Ω, $auu^T + bvv^T$ is a rank-two approximation term, and $\tau > 0$ is a shift parameter. Then x is updated according to the following rule:

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k),$$

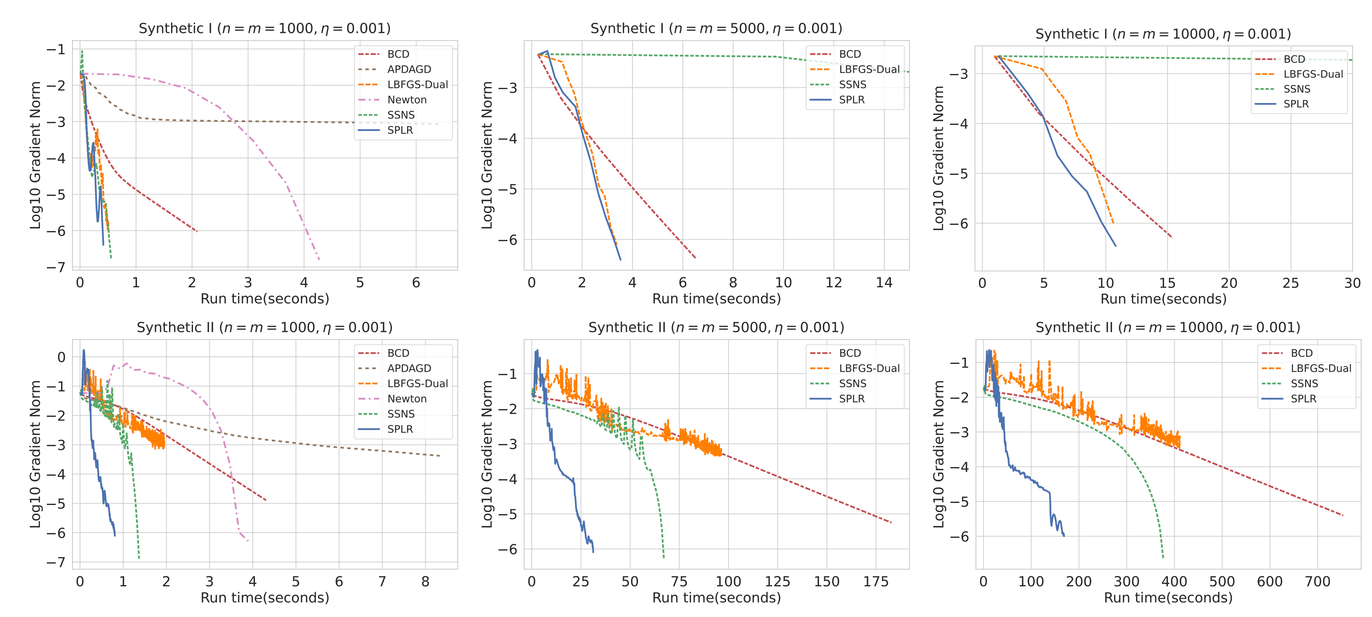
Experiments

We evaluate the performance of the proposed SPLR algorithm via a series of numerical experiments, and compare SPLR with a number of widely-used algorithms for solving entropic-regularized OT.

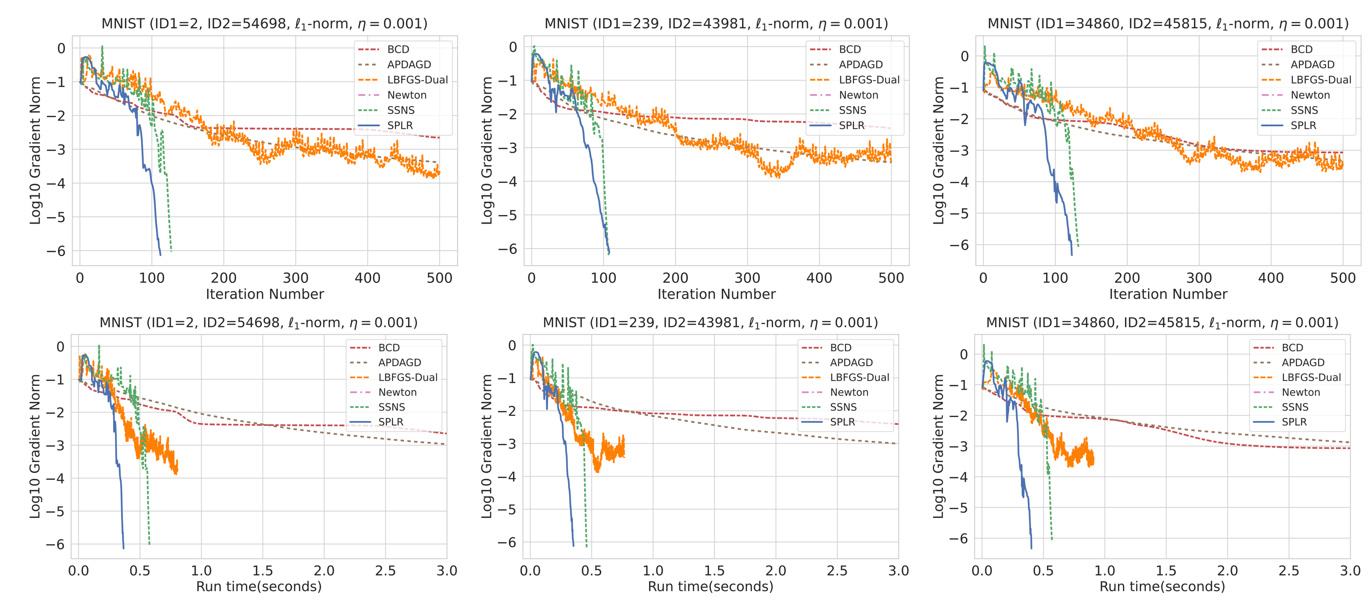
Synthetic Data I



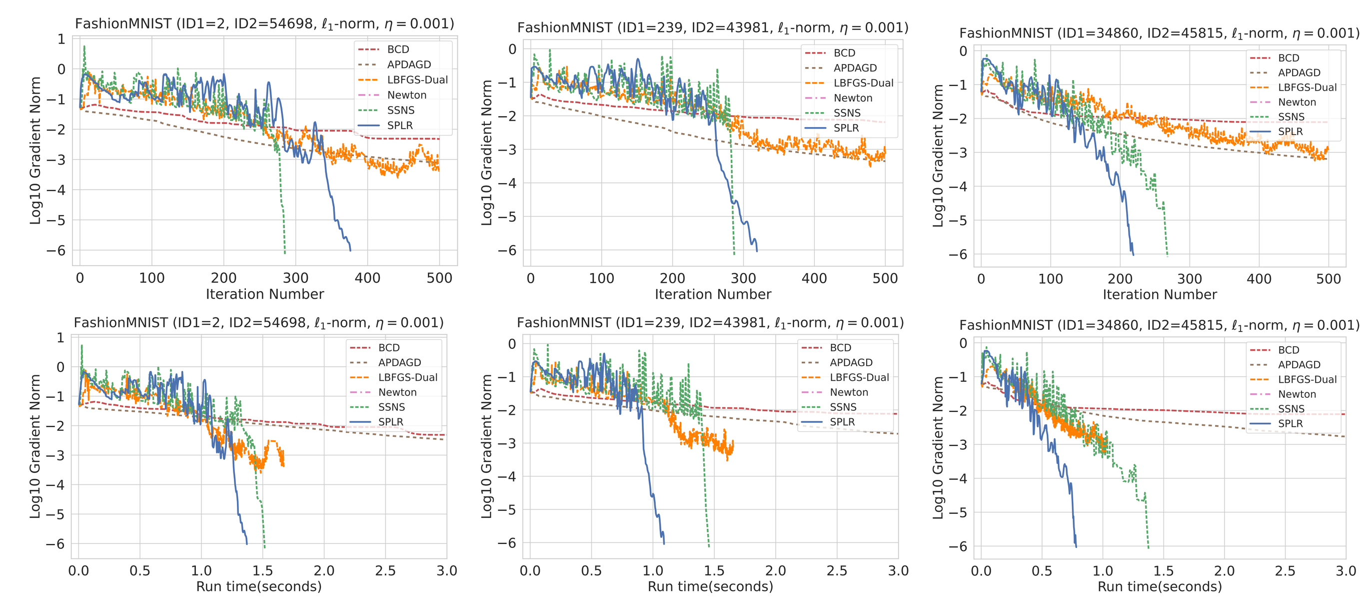
Synthetic Data II



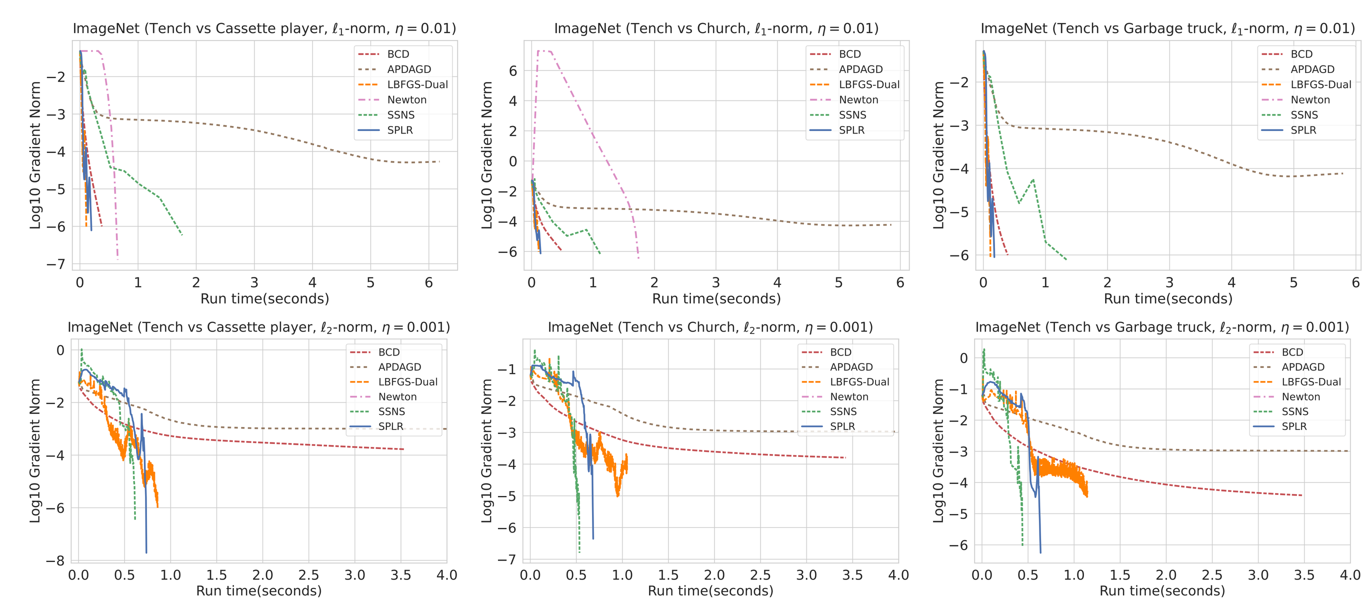
MNIST



Fashion-MNIST



Imagenet



Effect of low-rank terms