2025 IEEE 64th Conference on Decision and Control (CDC)
December 10-12, 2025. Rio de Janeiro, Brazil

Tensor Completion via Integer Optimization

Xin Chen*, Sukanya Kudva*, Yongzheng Dai, Anil Aswani, and Chen Chen

2025 IEEE 64th Conference on Decision and Control (CDC) | 979-8-3315-2627-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CDC57313.2025.11312503

Abstract— **The tensor completion problem is to fill-in unobserved entries of a partially observed tensor. However, past approaches to tensor completion either achieved the information-theoretic rate but lacked practical algorithms, or proposed polynomial-time algorithms that require an exponentially-larger number of samples for low estimation error. In this paper, we develop a novel tensor completion algorithm to tackle this challenge by achieving both provable convergence (in numerical tolerance) in a linear number of oracle steps and the information-theoretic rate. We formulate tensor completion as a convex optimization problem constrained using a gauge-based tensor norm and provide proofs of properties of this norm including its computational complexity and tensor rank surrogacy. We formulate this norm such that linear separation problems over the gauge unit-ball can be solved using integer optimization. This enables the use of Frank-Wolfe variant to build our algorithm. We demonstrate effectiveness of our method using experiments with simulated data and with an application towards providing low-computation predictions of battery storage flows that may be beneficial in billion-device-scale integration of electromobility systems with the grid.**

I. INTRODUCTION

A tensor is a multilinear operator, and it can be represented as an array of numbers referenced by multiple indices. For a tensor $\psi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$ of order p, we use $\psi_x := \psi_{x_1,\ldots,x_p}$ to refer to an entry corresponding to the indices $x = (x_1, \ldots, x_p)$, where $x_i \in [r_i]$ and $[s] := \{1, \ldots, s\}$. Furthermore, we define parameters: $\rho = \sum_i r_i$, $\pi = \prod_i r_i$, and $\mathcal{R} = [r_1] \times \cdots \times [r_p]$. Though matrices are special cases of tensors, problems like computing rank, singular values, and nuclear norm for tensors with $p \geq 3$ are NP-hard [1].

In this paper, we address one such difficult problem called *tensor completion*. Here, a small subset of tensor entries are observed – possibly with noise. Under an assumption of low-rankness, the problem is to fill-in the remaining, unobserved entries – and remove noise, if any. Some applications of tensor completion include recommendation systems [2], [3], information diffusion [4], nonparametric regression [5], computer vision [6], [7], [8], bioinformatics [9], [10] and control systems [11], [12].

A. Related Work

The information-theoretic rate for estimation error in tensor completion is $\sqrt{k \cdot \sum_i r_i / n}$, where: k is tensor rank, r_i is the i-th dimension of the tensor, and n is the number of samples [13], [14], [15]. Past work has studied tradeoffs between computational complexity and this rate. [16].

Initial work on tensor completion used decomposition methods based on CP and Tucker decomposition [17]. Other approaches accounted for robustness to outliers and data corruptions [18], [15], or imposed fixed-rank constraints [19]. Alternative approaches used various tensor norms as a convex surrogate for tensor rank [14], [20], [21]. Our approach falls into this last category, where we give a new construction of the nuclear-∞ norm [22] using a gauge function to allow the use of integer optimization.

Past approaches to tensor completion are computationally efficient without achieving the information-theoretic rate and rely on heuristics to compute non-certifiably optimal solutions for NP-hard problems, where optimal solutions (if found) theoretically achieve the information-theoretic rate [19], [20], [15]. Some specialized tensor completion algorithms have successfully achieved the information-theoretic rate in practical computation time, such as rank-1 nonnegative tensor completion formulated as conic optimization with exponential constraints [5], symmetric tensor completion using a variant of the Frank-Wolfe algorithm [23] and two-stage non-convex optimization [24], and first-order optimization methods leveraging integer optimization for a weak separation oracle in nonnegative tensors [25], [26].

B. Contribution

Our algorithm does not assume tensor properties like nonnegativity or symmetricity and works for general tensors. It is guaranteed to reach a global optimum in a linear (in the required accuracy as represented in bits) number of oracle calls, where the oracle solves a weak separation problem using integer optimization and a heuristic. More importantly, it attains the information-theoretic rate, i.e., data efficiency, while finding optimal solutions on instances of general tensors of sizes up to $10^{\times 7}$ within minutes.

The main idea behind our algorithm is to define the tensor completion problem using a gauge norm constructed using a convex polytope with its vertices as rank-1 tensors. We relate the gauge norm to tensor rank and analyze the norm's computational and statistical complexities in relation to NP-hardness and low Rademacher complexity, respectively. Consequently, the tensor completion problem using gauge norm is also NP-hard to solve to arbitrary accuracy. Nevertheless, since the formulation is a convex optimization problem, we design an algorithm using a Franke-Wolfe-like first order method called Blended Conditional Gradients (BCG) [27]. We construct a weak separation oracle for BCG using an integer linear optimization formulation and deploy an additional heuristic

*Equal contribution (Alphabetical order).

This material is based upon work supported by the National Science Foundation under Grant DGE-2125913 and Grant CMMI-1847666.

XC is with Stanford jenxchen at stanford dot edu, SK and AA are with UC Berkeley {sukanya_kudva,aaswani} at berkeley dot edu, YD and CC are with Ohio State {dai.651, chen.8018} at osu dot edu.

to accelerate the computation of oracle calls.

C. Going from Nonnegative to General Tensors

This paper extends an approach for nonnegative tensors [25] to general tensors, which we explain below is a nontrivial generalization. The key idea in [25] was to define a gauge norm using a 0-1 polytope representing the convex hull of rank-1 nonnegative tensors with $\{0, 1\}$ entries. The polytope was such that linear separation problems over it could be written using integer linear constraints. For illustration consider the set $\{\zeta = \prod_{k=1}^{p} \nu_k : \nu_k \in \{0,1\}\}$. Its linearization $\{\zeta : 0 \le \zeta \le \nu_k, \sum_k \nu_k + (1-p) \le \zeta, \nu_k \in \{0,1\}\}$ is easily possible since $\zeta = 0$ if any single $\nu_k = 0$, and $\zeta = 1$ if all $\nu_k = 1$. The natural generalization of this idea, pursued in this paper, is a gauge norm using a polytope that is the convex hull of all (general) rank-1 tensors with entries ± 1. However, for the new set $\{\zeta = \prod_{k=1}^{p} \nu_k : \nu_k \in \{-1,+1\}\}$ satisfies $\zeta = +1$ if an even number of $\nu_k = -1$, and $\zeta = -1$ if an odd number of $\nu_k = -1$. Hence, the linearization is fundamentally different, more challenging, and requires new computational design and theoretical analysis.

General tensors also have well-posedness issues that do not occur for nonnegative tensors. In particular, the best low-rank approximation problem is not well-posed for tensors [28]. In contrast, the best low-rank approximation problem *is* well-posed for nonnegative tensors [29]. Given these phenomena for general tensors, there is a technical challenge in determining whether a specific gauge norm, as defined above, can act as a convex surrogate for the rank of tensors.

D. Applications to Electromobility and Battery Storage

Increased use of renewable energy and the rapid growth of electric vehicles (EVs) has introduced new complexities in energy grid management, namely integration of consumers who store surplus energy to share with the grid [30]. This has resulted in a bidirectional and dynamic energy flow, making grid management more intricate. EVs serve as a prime example, as they both consume significant energy and can store and reinject surplus energy into the grid [31].

Making predictions of electricity demand, generation, and storage is beneficial to optimal operation of the grid, but scaling model estimation and predictions to billion-device-scale systems is challenging because of the high computational burden of typical nonparametric models (e.g., random forests, neural networks). Here, we propose the use of tensor completion to construct nonparametric models because they have extremely low computation requirements for prediction. Indeed, these use of tensor completion for regression has been previously considered for bioengineered systems [5]. Here, we use household energy storage data from Ireland [32] to demonstrate a proof of concept in this direction.

II. Preliminaries

By definition, a rank-1 tensor can be written as the tensor product of vectors, that is $\psi = \bigotimes_{k=1}^{p} \theta^{(k)}$, where $\theta^{(k)} \in \mathbb{R}^{r_k}$. Equivalently, each tensor entry is $\psi_x := \psi_{x_1,\cdots,x_p} = \prod_{k=1}^{p} \theta^{(k)}_{x_k}$, where $\theta^{(k)}_{x_k}$ is the x_k-th element of vector $\theta^{(k)}$ for any index value $x_k \in [r_k]$. When obvious, we will use θ_{x_k} instead of $\theta^{(k)}_{x_k}$. Let $\mathcal{B}_\lambda$ be the set of rank-1 tensors such that each entry of the tensor has absolute values less than or equal to $\lambda \in \mathbb{R}_+$:

$$\mathcal{B}_\lambda = \{\psi : \psi_x = \lambda \cdot \textstyle\prod_{k=1}^{p} \theta_{x_k},\ \theta_{x_k} \in [-1,1] \text{ for } x \in \mathcal{R}\}. \quad (1)$$

Then, the rank of a tensor is defined as:

$$\text{rank}(\psi) = \min\{q \mid \psi = \textstyle\sum_{k=1}^{q} \psi^k, \psi^k \in \mathcal{B}_\infty \text{ for } k \in [q]\},$$

and its CP decomposition is given by $\psi = \sum_{k=1}^{\text{rank}(\psi)} \psi^k$.

The tensor completion problem we consider begins with n observations of the tensor, which are denoted by the pairs $(x\langle i\rangle, y\langle i\rangle) \in \mathcal{R} \times \mathbb{R}$ for $i \in [n]$. Here, $y\langle i\rangle$ is the (possibly noisy) observation of the tensor entry $\psi_{x\langle i\rangle}$. We note that the $x\langle i\rangle$ are assumed to be independent and identically distributed in our model, which means that any given entry of the tensor may be observed multiple times within the n observations. Our approach is to solve the tensor completion problem using a least squares formulation:

$$\begin{aligned} \widehat{\psi} \in \arg\min_{\psi}\ & \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(y\langle i\rangle - \psi_{x\langle i\rangle}\right)^2 \\ \text{s.t. } & \|\psi\|_{\pm} \le \lambda \end{aligned} \quad (2)$$

where the constraint uses a gauge norm $\|\psi\|_{\pm}$ defined below.

III. Gauge Norm for Tensors

We construct a norm for general tensors using a gauge function [33], [34], [25]. Let $\mathcal{S}_\lambda$ be the set of rank-1 tensors such that each entry of the tensor has an absolute value of some $\lambda \in \mathbb{R}_+$:

$$\mathcal{S}_\lambda = \{\psi : \psi_x = \lambda \cdot \textstyle\prod_{k=1}^{p} \theta_{x_k},\ \theta_{x_k} \in \{-1,1\} \text{ for } x \in \mathcal{R}\}. \quad (3)$$

Proposition 1: The convex hulls of the sets $\mathcal{B}_\lambda$ and C_λ are the same, i.e. $C_\lambda := \text{conv}(\mathcal{B}_\lambda) = \text{conv}(\mathcal{S}_\lambda)$.

Proof: The proof for this proposition is similar to that of Proposition 2.1 of [25], but where a multilinear optimization problem is formulated by restricting the entries to $[-1,1]$ instead of $[0,1]$ as in (4). We proceed by showing two set inclusions $\text{conv}(\mathcal{S}_\lambda) \subseteq \text{conv}(\mathcal{B}_\lambda)$ and $\text{conv}(\mathcal{B}_\lambda) \subseteq \text{conv}(\mathcal{S}_\lambda)$. The first inclusion is immediate by definition since we have $\mathcal{S}_\lambda \subset \mathcal{B}_\lambda$. The second inclusion can be proved by contradiction. Suppose instead that $\text{conv}(B_\lambda) \not\subset \text{conv}(\mathcal{S}_\lambda)$. Then there exists a tensor $\psi' \in B_\lambda$ with $\psi' \notin \text{conv}(\mathcal{S}_\lambda)$. By the hyperplane separation theorem, there exists $\varphi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$ and $\delta > 0$ such that $\langle \varphi, \psi' \rangle \ge \langle \varphi, \psi \rangle + \delta$ for all $\psi \in \text{conv}(\mathcal{S}_\lambda)$, where $\langle \cdot, \cdot \rangle$ is the usual inner product that is defined as the summation of elementwise multiplication. Now consider the multilinear optimization problem

$$\begin{aligned} \max\ & \langle \varphi, \psi \rangle \\ \text{s.t. } & \psi_x = \lambda \cdot \textstyle\prod_{k=1}^{p} \theta_{x_k}, && \text{for } x \in \mathcal{R} \\ & \theta_{x_k} \in [-1,1], && \text{for } x \in \mathcal{R} \end{aligned} \quad (4)$$

Proposition 2.1 of [35] shows there exists a global optimum ψ'' of (4) with $\psi'' \in \mathcal{S}_\lambda$. By construction, $\langle \varphi, \psi'' \rangle \ge \langle \varphi, \psi' \rangle$, which implies $\langle \varphi, \psi'' \rangle \ge \langle \varphi, \psi \rangle + \delta$ for $\psi \in \text{conv}(\mathcal{S}_\lambda)$. This statement is a contradiction since $\psi'' \in \mathcal{S}_\lambda \subseteq \text{conv}(\mathcal{S}_\lambda)$. ■

Remark 1: Note $\mathcal{B}_\lambda = \lambda\mathcal{B}_1$, $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$, and $C_\lambda = \lambda C_1$.

C_λ is useful because it is a convex polytope with its vertices as the points in $\mathcal{S}_\lambda$, and we use it to define a gauge norm.

Proposition 2: The function defined as

$$\|\psi\|_\pm := \inf\{\lambda \geq 0 \mid \psi \in \lambda C_1\} \tag{5}$$

is a norm for all tensors $\psi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$.

Proof: We use the result from Example 3.50 of [36] to conclude that $\|\cdot\|_\pm$ is a norm. To apply the result, we check that the required conditions on C_1 hold.

By definition C_1 is convex, closed, and bounded. C_1 is also symmetric since for every $a \in C_1$, $-a \in C_1$ is also true. To see this, let $a = \sum_i \lambda_i \psi_i$ where $\psi_i \in \mathcal{S}_1$, $\lambda_i \in [0,1]$ and $\sum_i \lambda_i = 1$ and notice $-\psi_i \in \mathcal{S}_1$. Symmetry and convexity also ensure $0 \in C_1$ since for any $a \in C_1$, $\frac{1}{2}(a + (-a)) = 0$.

The final, non-trivial condition required is that C_1 has a non-empty interior. To prove this, we use Theorem 2.4 of [37] that the dimension of C_1 is the maximum of dimensions of simplices included in it. We construct a simplex in C_1 that has dimension $\pi = \Pi_i r_i$, and hence C_1 has a dimension of at least π. But the flattened vector space of tensors also has dimension π; so, the dimension of C_1 is at most π. Hence, C_1 has a dimension π, which is the full dimension of the space, implying that the set C_1 must have a non-empty interior.

The rest of the proof constructs such a simplex of dimension π. Consider a polytope $D = \text{conv}(0 \cup \{d^x\}_{x \in \mathcal{R}})$. Here, for any $x = (x_1, \cdots, x_p) \in \mathcal{R}$, consider $d^x = \bigotimes_{k=1}^p \beta^{x_k}$ with each vector $\beta^{x_k} \in \mathbb{R}^{r_k}$ defined as $\beta^{x_k} = \mathbb{1}$ if $x_k = 1$ and $\beta^{x_k} = f_{x_k}$ if $x_k \neq 1$, where $\mathbb{1}$ is a vector of one's and f_j is a vector with -1 in position j and one's elsewhere. One can verify that $\{\beta^{x_k}\}_{x_k \in [r_k]}$ are linearly independent vectors and make a complete basis for $\mathbb{R}^{r_k}$. Since the tensor product of linearly independent vectors gives linearly independent tensors, the tensors $\{d^x\}_{x \in \mathcal{R}}$ are all linearly independent. Consequently, $\{d^x - 0\}_{x \in \mathcal{R}}$ are linearly independent and $\{d^x\}_{x \in \mathcal{R}} \cup 0$ are affinely independent. By definition, the polytope D, which is a convex hull of $|\mathcal{R}| + 1 = \pi + 1$ affinely independent points, is a simplex of dimension π. Note that all the points are in C_1, and hence so is the simplex.

With this, C_1 is shown to satisfy all the required conditions, which means that the proposition holds. ∎

To develop a practical algorithm using this norm, our next result provides an alternative characterization of the vertices of C_λ using linear inequality constraints of (binary) integer variables. This is unlike, say, the nuclear norm under the ℓ_2 norm that has a nonlinear (continuous) feasible region.

Proposition 3: We have $\widehat{\mathcal{S}}_\lambda = \mathcal{S}_\lambda$ for the set

$$\begin{aligned}
\widehat{\mathcal{S}}_\lambda = \big\{\psi :\ & \psi_x = \lambda \cdot y_{x,1} && x \in \mathcal{R} \\
& y_{x,k} \geq (-\theta_{x_k} - y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \geq (\theta_{x_k} + y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leq (\theta_{x_k} - y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leq (-\theta_{x_k} + y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,p} = \theta_{x_p} && x \in \mathcal{R} \\
& \theta_{x_k} \in \{-1, 1\} && x \in \mathcal{R} \\
& \theta^k \in \mathbb{R}^{r_k},\ y_{x,k} \in \mathbb{R} && k \in [p], x \in \mathcal{R}\big\}.
\end{aligned}$$

Proof: We show the constraints defining sets $\mathcal{S}_\lambda$ and $\widehat{\mathcal{S}}_\lambda$ are equivalent. Consider some $x \in \mathcal{R}$. From the definition of $\mathcal{S}_\lambda$ in (3), we have $\psi_x = \lambda \prod_{i=1}^p \theta_{x_k}$. Define $y_{x,k} = \prod_{i=k}^p \theta_{x_k}$, for $k \in [p]$ so that $\psi_x = \lambda y_{x,1}$. Or equivalently, in a recursive relationship, $y_{x,k} = \theta_{x_k} y_{x,k+1}$ for $k \in [p-1]$ and $y_{x,p} = \theta_{x_p}$. The recursive constraints can be thought of as a negated-XOR relation, and linearized by transformations for conjunctive and disjunctive statements (see Section 2.5 of [38]). These linearized constraints correspond to constraints 2-5 in the definition of $\widehat{\mathcal{S}}_\lambda$ as given above. So for each $x \in \mathcal{R}$, the tensor ψ_x is defined the same in both $\mathcal{S}_\lambda$ and $\widehat{\mathcal{S}}_\lambda$. ∎

Remark 2: The norm defined in (5) is equivalent to the nuclear-∞ norm by Proposition 4.3 of [22].

A. Relation between Gauge Norm and Tensor Rank

The underlying issue is related to the fact that the best low-rank approximation problem is not well-posed for tensors [28], which is in sharp contrast to the case of nonnegative tensors for which the best low-rank approximation problem is well-posed [29]. We impose a regularity condition to eliminate such pathological behavior of tensors.

Assumption 1 (Regularity Condition): Consider a class of tensors defined by the set

$$\Gamma = \big\{\psi : \exists \text{ CP decomposition of } \psi \text{ with terms } \psi^k \text{ s.t. } \|\psi^k\|_{\max} \leq \|\psi\|_{\max} \text{ for } k \in [\text{rank}(\psi)]\big\}. \tag{6}$$

This class is such that each tensor ψ has its largest entry at least as large as the largest entry of each CP term ψ^k.

Remark 3: A CP decomposition always exists for a finite-valued tensor, but it may not be unique. The class defined above asks that the regularity condition holds for at least one CP decomposition, but does not make any statement about holding for all the possible CP decompositions.

The following proposition suggests that the norm $\|\psi\|_\pm$, which is convex, can be a useful surrogate for tensor rank.

Proposition 4: For any $\psi \in \Gamma$ that satisfies Assumption 1, we have $\|\psi\|_{\max} \leq \|\psi\|_\pm \leq \text{rank}(\psi) \cdot \|\psi\|_{\max}$.

Proof: By the CP decomposition and the triangle inequality: $\|\psi\|_\pm \leq \sum_{k=1}^{\text{rank}(\psi)} \|\psi^k\|_\pm = \sum_{k=1}^{\text{rank}(\psi)} \|\psi^k\|_{\max}$, where $\psi^k \in \mathcal{B}_\infty$. The last equality follows from $\|\psi^k\|_\pm = \|\psi^k\|_{\max}$ when $\psi^k \in \mathcal{B}_\infty$. Using $\|\psi^k\|_{\max} \leq \|\psi\|_{\max}$ from Assumption 1 gives the right-side inequality.

The proof of the left-side inequality is similar to Proposition 2.4 of [25]. For any $\lambda \geq 0$, if $\psi \in \mathcal{S}_\lambda$, then by definition $\|\psi\|_{\max} = \lambda$. By the convexity of norms we have that: if $\psi \in C_\lambda$, then $\|\psi\|_{\max} \leq \lambda$. This means that $\forall\, \lambda \geq 0$ we have $C_\lambda \subseteq \mathcal{U}_\lambda := \{\psi : \|\psi\|_{\max} \leq \lambda\}$, and thus $\inf\{\lambda \mid \psi \in \lambda \mathcal{U}_1\} \leq \inf\{\lambda \mid \psi \in \lambda C_1\}$. But this is equivalent to $\|\psi\|_{\max} \leq \|\psi\|_\pm$, which proves the left-side inequality. ∎

Remark 4: Tensor rank surrogacy is not always possible. It does not hold for a gauge norm defined using the l_1-ball.

B. Complexity Analysis of Norm

We show that calculating the norm $\|\cdot\|_\pm$ is NP-complete.

*Proposition 5 (**Computational complexity**):* It is in fact NP-complete to determine if $\|\psi\|_\pm \leq K$ for $\psi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$.

Proof: Note that $\|\varphi\|_\circ = \sup\{|\langle\varphi,\psi\rangle| \mid \|\psi\|_\pm \le 1\} = \sup\{\langle\varphi,\psi\rangle \mid \psi \in C_1\}$ is the dual norm for $\|\cdot\|_\pm$. The approximation of $\|\cdot\|_\circ$ can be reduced to approximation of the norm $\|\cdot\|_\pm$ in polynomial time from theorems 3 and 10 of [39]. Our main idea is to give a polynomial-time reduction of an NP-Hard problem to an approximation of $\|\cdot\|_\circ$. In particular, we prove that calculating the $\infty,1$ subordinate matrix norm is polynomial-time reducible to $\sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$. Without loss of generality, assume $p = 2$ and $d := r_1 = r_2$. The decision version of $\sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$ is: *Question:* Does there exist $\theta_{x_k} \in \{-1,1\}$ for all $x_k \in [d]$ with $k = 1,2$ such that for a given L we have $\sum_{x_1=1}^d \sum_{x_2=1}^d \varphi_{x_1x_2}\theta_{x_1}\theta_{x_2} \ge L$? From Proposition 1 of [40], we have $\|W\|_{\infty,1} = \{\max \sum_{i,j} W_{ij}x_iy_j \mid x_i, y_j \in \{-1,1\}\}$ for a matrix W. The decision version of the $\infty,1$ subordinate matrix norm for the special case of M-matrices can be written as: *Question:* Does there exist $x_i, y_i \in \{-1,1\}$ for all $i \in [d]$ such that for a given $L' \ge 0$ and a *symmetric, positive definite* matrix $W \in \mathbb{R}^{d\times d}$ satisfying $w_{ij} \le 0$ for all $i \ne j$ and $\sum_{i=1}^d \sum_{j=1}^d w_{ij}x_iy_j \ge L'$? Clearly, setting $L = L'$ and $\varphi = W$ is a valid polynomial time reduction. Since $C_1 = \text{conv}(\mathcal{S}_1)$ and $\langle\varphi,\psi\rangle$ is linear, $\|\varphi\|_\circ = \sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$. The result now follows since approximately solving any ∞,p subordinate matrix norm, where $p \in [1,\infty)$, to arbitrary accuracy is NP-hard [41]. In particular, Theorem 5 of [40] shows that it is an NP-hard problem for M-matrices. When combined with Corollary 3.2 in [25], this establishes NP-completeness. ∎

Rademacher complexity, from computational learning theory, is used to characterize the richness of a class of functions [42], [43]. Roughly speaking, function classes with lower Rademacher complexity can be learned using less samples. From the following proposition, one can check that the norm $\|\cdot\|_\pm$ has an exponentially smaller Rademacher complexity than the max and Frobenius norms for tensors.

*Proposition 6 (**Stochastic Complexity**):* We have the relation $\mathsf{R}(C_\lambda) \le \mathsf{W}(C_\lambda) \le 2\lambda\sqrt{\rho/n}$, where $\mathsf{R}(\cdot)$ and $\mathsf{W}(\cdot)$ are the Rademacher and worst case Rademacher complexities.

Proof: The proof is similar to Proposition 3.3 of [25]. By definition, $\mathsf{R}(C_\lambda) = \mathbb{E}_\sigma(\sup_{\Psi\in C_\lambda} \frac{1}{n}|\sum_{i=1}^n \sigma_i \cdot \psi_{x\langle i\rangle}|)$ and $\mathsf{W}(C_\lambda) = \sup_X \mathsf{R}(C_\lambda)$, where σ_i are independent and identically distributed (i.i.d.) Rademacher random variables (i.e., $\sigma_i = \pm 1$ with probability $\frac{1}{2}$) [42], [43]. Hence, $\mathsf{R}(C_\lambda) \le \mathsf{W}(C_\lambda)$. Recall that $C_\lambda = \text{conv}(\mathcal{S}_\lambda)$ by Proposition 1. This means $\mathsf{W}(C_\lambda) = \mathsf{W}(\mathcal{S}_\lambda)$ [44], [42]. Next, observe that

$$\begin{aligned}\mathsf{W}(C_\lambda) &= \mathsf{W}(\mathcal{S}_\lambda)\\ &= \sup_X \mathbb{E}_\sigma\big(\sup_{\psi\in\mathcal{S}_\lambda} \tfrac{1}{n}\big|\textstyle\sum_{i=1}^n \sigma_i\cdot\psi_{x\langle i\rangle}\big|\big)\\ &= \sup_X \mathbb{E}_\sigma\big(\max_{\psi\in\mathcal{S}_\lambda} \tfrac{1}{n}\cdot\textstyle\sum_{i=1}^n \sigma_i\cdot\psi_{x\langle i\rangle}\big)\\ &\le \sup_X r\sqrt{2\log \#\mathcal{S}_\lambda}/n\end{aligned} \tag{7}$$

where in the last line since the set $\mathcal{S}_\lambda$ is finite, we used the Finite Class Lemma [45] with $r = \max_{\psi\in\mathcal{S}_\lambda}\sqrt{\sum_{i=1}^n(\psi_{x\langle i\rangle})^2} \le \lambda\sqrt{n}$, and replaced the supremum with a maximum. This inequality on r is because $\mathcal{S}_\lambda$ consists of tensors whose entries are from $\{-\lambda,\lambda\}$. Thus $\mathsf{W}(C_\lambda) \le \lambda\sqrt{(2\log 2)\cdot\rho/n} \le \lambda\cdot 2\sqrt{\rho/n}$. ∎

IV. Algorithm for Tensor Completion

A. Complexity Analysis of Tensor Completion

By interpreting the tensor completion problem (2) as a *convex aggregation* problem [46], [47], [48] for a finite set of functions, one can arrive at the generalization bound for the solution. Interestingly, these generalization bounds are the same in the special case of nonnegative tensors [25]. We believe this is because the proof for nonnegative tensors did not exploit the non-negativity, and hence could have led to non-tight generalization bounds. For completeness, we state these statistical guarantees in the following two results:

Proposition 7 ([48]): Suppose $|y| \le b$ almost surely. Given any $\delta > 0$, with probability at least $1 - 4\delta$ we have that $\mathbb{E}\big((y-\psi_x)^2\big) \le \min_{\varphi\in C_\lambda}\mathbb{E}\big((y-\varphi_x)^2\big) + c_0\cdot\max\big[b^2,\lambda^2\big]\cdot\max\big[\zeta_n, \frac{\log(1/\delta)}{n}\big]$, where c_0 is an absolute constant and

$$\zeta_n = \begin{cases}\frac{2^\rho}{n}, & \text{if } 2^\rho \le \sqrt{n}\\ \sqrt{\frac{1}{n}\log\left(\frac{e2^\rho}{\sqrt{n}}\right)}, & \text{if } 2^\rho > \sqrt{n}\end{cases} \tag{8}$$

Corollary 1 ([25]): Suppose $\psi \in \Gamma$ is a tensor (satisfying Assumption 1) with $\text{rank}(\psi) = k$ and $\|\psi\|_{\max} \le \mu$. Under an additive noise model, if $(x\langle i\rangle, y\langle i\rangle)$ are independent and identically distributed with $|y\langle i\rangle - \psi_{x\langle i\rangle}| \le e$ almost surely and $\mathbb{E}y\langle i\rangle = \psi_{x\langle i\rangle}$. Then given any $\delta > 0$, with probability at least $1 - 4\delta$ we have

$$\mathbb{E}\big((y-\widehat{\psi}_x)^2\big) \le e^2 + c_0\cdot\big(\mu k + e\big)^2\cdot\max\big[\zeta_n, \tfrac{\log(1/\delta)}{n}\big], \tag{9}$$

where ζ_n is as in (8) and c_0 is an absolute constant.

Remark 5: The above result achieves the information-theoretic rate when the rank $k = O(1)$.

Unsurprisingly, the tensor completion problem (2) is NP-hard, because approximating the norm $\|\cdot\|_\pm$ is NP-hard and there is a polynomial-time reduction of the problem (2) to the NP-hard weak membership problem [25].

Proposition 8: Tensor completion (2) is NP-hard to solve to arbitrary accuracy, and its decision version is NP-complete.

Proof: The proof for this proposition is similar to the proof of Proposition 4.4 of [25]. Define the ball of radius $\delta > 0$ centered at a tensor ψ to be $B(\psi,\delta) = \{\varphi : \|\varphi-\psi\|_F \le \delta\}$. Next define $W(C_1,\delta) = \bigcup_{\psi\in C_1} B(\psi,\delta)$ and $W(C_1,-\delta) = \{\psi \in C_1 : B(\psi,\delta) \subseteq C_1\}$. The weak membership problem for C_1 is that given a tensor ψ and a $\delta > 0$ decide whether $\psi \in W(C_1,\delta)$ or $\psi \notin W(C_1,-\delta)$. Theorem 10 of [39] shows that approximation of the norm $\|\cdot\|_\pm$ is polynomial-time reducible to the weak membership problem for C_1. Since Proposition 5 shows that approximation of the norm $\|\cdot\|_\pm$ is NP-hard, the result follows if we can reduce the weak membership problem to (2).

Suppose we are given inputs ψ and δ for the weak membership problem. Choose $x\langle i\rangle$ for $i = 1,\ldots,\pi$ such that each element in $\mathcal{R}$ is enumerated exactly once. Next choose $y\langle i\rangle = \psi_{x\langle i\rangle}$ and $\lambda = 1$. Finally, note if we solve (2) and the minimum objective value is less than or equal to δ, then we have $\psi \in W(C_1,\delta)$; otherwise, we have $\psi \notin W(C_1,-\delta)$. The result follows since this was the desired reduction. ∎

B. Numerical Algorithm

Despite being NP-hard, the optimization problem (2) is convex. This enables application of various first-order convex optimization algorithms. In particular, we use a variant of the Frank-Wolfe algorithm called *Blended Conditional Gradients* (BCG) [27]. The main alternative approaches for iterative optimization seem to run into issues inherent to the implicit description of the feasible region; for instance, we do not have an effective barrier function available for interior point methods, nor do we have an efficient projection oracle that can be leveraged in projected gradient descent.

BCG makes iterate updates based on the gradient of the objective function at the current iterate. It needs a weak separation oracle to find a new vertex that reduces a gradient-based linear objective. The weak separation oracle is given in Algorithm 1. The output of the oracle should either give a vertex that accomplishes separation, or a certificate that separation is not possible. We design our weak separation oracle using two algorithms, an integer optimization problem in (10) and an alternating maximization heuristic.

Since integer optimization is likely to be more computationally expensive than the alternating heuristic, we first try the latter several times with different randomized initializations. The heuristic adapts Algorithm 2, where solutions are explored by toggling between ± 1 [25]. The following (nonlinear) separation objective is minimized over binary θ, given a BCG-generated parameter c: $z_M(\theta) := \sum_{x\in\mathcal{R}}\langle c_x, \psi_x - \lambda \cdot \prod_{k=1}^{p} \theta_{x_k}\rangle$. The heuristic considers one dimension at a time, setting the corresponding entries of θ to optimize within the local neighbourhood defined by the given dimension (which can be done by simply considering the signs of c). It runs in polynomial time (linear in the size of θ) and has been seen to speed up the computation in our simulations by reducing calls to the integer optimization solver. However, the heuristic is merely that and cannot in general give a certificate for the non-existence of separation, which requires global optimization. Indeed, even for a matrix ($p = 2$), z_M represents a generic form of NP-hard Quadratic Unconstrained Binary Optimization (QUBO) [49].

If the heuristic cannot yield a separating cut, we implement the following integer optimization. Note $\langle\cdot,\cdot\rangle$ is the dot product of tensors obtained by flattening them into vectors.

$$
\begin{aligned}
&\max_{\varphi,\theta,y} \langle c, \psi - \varphi\rangle \\
&\text{s.t. } \varphi_x = \lambda y_{x,1} && x \in \mathcal{R} \\
&\quad y_{x,k} \geq (-\theta_{x_k} - y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
&\quad y_{x,k} \geq (\theta_{x_k} + y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
&\quad y_{x,k} \leq (\theta_{x_k} - y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
&\quad y_{x,k} \leq (-\theta_{x_k} + y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
&\quad y_{x,p} = \theta_{x_p} && x \in \mathcal{R} \\
&\quad \theta_{x_k} \in \{-1, 1\} && k \in [p], x \in \mathcal{R}
\end{aligned}
\tag{10}
$$

Note that, in principle, for some binary variable $q \in \{-1, 1\}$, one can directly branch on the disjunction: $q = -1 \vee q = 1$. However, in implementation we choose instead to reformulate via 0-1 auxiliary variables, $\sigma := (q + 1)/2$, as enforcing $\sigma \in \{0, 1\}$ obviates the need for imposing $q \in \{-1, 1\}$. Integer optimization solvers typically can only handle 0-1 binary variables by default instead of $\{-1, 1\}$ due to the more extensive set of techniques developed for 0-1 variables.

Algorithm 1 Weak Separation Oracle for C_λ

Input: linear objective $c \in \mathbb{R}^{r_1\times\cdots\times r_p}$, point $\psi \in C_\lambda$, accuracy $K \geq 1$, gap estimate $\Phi > 0$, norm bound λ
Output: Either (1) vertex $\varphi \in \mathcal{S}_\lambda$ with $\langle c, \psi - \varphi\rangle \geq \Phi/K$, or (2) **false:**
$\langle c, \psi - \varphi\rangle \leq \Phi$ for all $\varphi \in C_\lambda$

Algorithm 2 Alternating Maximization

Input: linear objective $c \in \mathbb{R}^{r_1\times\cdots\times r_p}$, point $\psi \in C_\lambda$, norm bound λ, incumbent solution $\hat{\theta} \in \mathcal{S}_\lambda$, objective function $z_M(\theta) := \sum_{x\in\mathcal{R}}\langle c_x, \psi_x - \lambda \cdot \prod_{k=1}^{p} \theta_{x_k}\rangle$.
Output: Best known solution θ

```
θ ← θ̂
z ← z_M(θ)
for i = 1 to p do
  for k = 1 to r_i do
    θ_k^(i) ← −1 · θ_k^(i)
    if z_M(θ) > z then
      z ← z_M(θ)
    else
      θ_k^(i) ← −1 · θ_k^(i)
    end if
  end for
end for
```

Since we are looking for weak separation, the integer optimization solver is made to terminate when a solution with an objective greater than Φ/K is found. In case of no such solution, the dual bound z from the solver serves as a no-separation certificate satisfying $\langle c, \psi - \varphi\rangle \leq z \leq \Phi$.

Proposition 9: Using BCG, the tensor completion problem in (2) can be solved in a linear (in the required accuracy as represented in bits) number of calls to the Weak Separation Oracle that is presented in Algorithm 1.

Proof: By Theorem 3.1 of [27], BCG converges linearly when the feasible set is a polytope and the objective function is smooth and strongly convex. Note the feasible set $\|\psi\|_\pm \leq \lambda$ in Problem (2) is a polytope. The objective is smooth and it can be made strongly convex by projecting the feasible space onto the set of unique, observed tensor entries U [25]. Specifically, we use the equivalent reformulation in which we change the feasible set from $\{\psi : \|\psi\|_\pm \leq \lambda\} = C_\lambda$ to $\mathrm{Proj}_U(C_\lambda)$ where the projection is done over the unique indices specified by the set U. Strong convexity of the objective results from strict convexity and quadraticity of the objective on the compact projected space. Since the conditions for linear convergence are satisfied by our problem, our algorithm terminates in a linear number of calls to Algorithm 1. ∎

In fact, we find the BCG algorithm to be practically

efficient for our problem since the weak linear separation oracle can accept (sufficiently good) suboptimal solutions that may be found at early termination of a global solver. When we design our weak separation oracle calls to an integer optimization problem, we explicitly define a tolerance for early termination. This works out well as integer optimization solvers usually discover near-optimal solutions fast and subsequently work hard to certify them.

V. Numerical Experiments

We perform numerical experiments to demonstrate scalability and efficacy of our tensor completion algorithm[1]. We compare its performance with benchmark algorithms, including the often-called 'workhorse' for numerical tensor problems, alternating least squares (ALS) [51], and two state-of-the-art algorithms implemented in the PyTen[2], known as the simple low-rank tensor completion (SiLRTC) algorithm [53] and the trace norm regularized CP decomposition (TNCP) algorithm [54].

A. Experiments on Simulated Tensor Data

We use simulated data to demonstrate the scalability of our algorithm with increasing tensor dimension and tensor order. The true tensor ψ is constructed by taking a random convex combination of a random set of ten points from $\mathcal{S}_1$ so that $\|\psi\|_\pm \leq 1$ and $\text{rank}(\psi) \leq 10$. To minimize the influence of hyper-parameter selection in the results, we use these ground truth values for the value of λ in our algorithm and the value of k in ALS and TNCP. Since ALS tends to perform better under L2 regularization [55], we chose its corresponding hyperparameter to make it favorable to its accuracy. Each experiment was performed with 100 repetitions, and the normalized mean squared error (NMSE), i.e., $\|\widehat{\psi} - \psi\|_F^2 / \|\psi\|_F^2$ was calculated. NMSE is a stricter measure than the error metric used in Corollary 1 because the statistical guarantee is not normalized.

In Figure 1, we use tensors of order $p = 3$ with dimensions increasing from $r = 10$ to $r = 100$. In each experiment, we observe $n = 1000$ samples, allowing for repetition in indices. Our approach yields greater accuracy in a lower size of r while all algorithms do not perform at a satisfactory level (NMSE below 1) as r increases close to 100.

In Figure 2, we consider tensors of increasing order p with dimension $r_i = 10$ for $i = 1, \ldots, p$. We observed $n = 10,000$ samples, again randomly sampled with replacement. we observe that our algorithm achieves consistently higher accuracy as compared to the other methods. In these plots, while our algorithm takes more computation time, it still converges within minutes even for tensors with 10^7 entries.

In Figure 3, we construct tensors of sizes $10^{\times 7}$. In each experiment, the sample size is increased by one order of magnitude as the percentage of total tensor entries, starting from 0.01% (using random sampling with replacement). The results show our algorithm achieves much higher accuracy while requiring greater computation time. Yet our algorithm converges within minutes for most cases, except for the case where the sample percent is 10^{-2}% (i.e., about three hours).

B. Experiments on Household Energy Storage Data

We run experiments on local weather and energy data from residential households in Ireland for the year 2020 [32]. Using tensor completion, we construct a nonparameteric model to predict how much energy (in kWh) is charged into storage batteries in an hour in one household. The inputs of this model are local temperature (in °C), percentage of battery charge, day of week (Monday-Friday) and hour of day (0-23). We map values of these inputs into 10, 5, 5, 24 uniform buckets, respectively, so that our nonparametric model is a $10 \times 5 \times 5 \times 24$ tensor. This model structure is beneficial for integration of model predictions for billion-device-scale systems because predictions have extremely low computational burden since model predictions simply become lookups into a table: the point at which a prediction is to be made is mapped to a set of indices based on the defined buckets, and the model prediction is simply the completed tensor value at that corresponding set of indices.

We randomly divide our dataset into non-overlapping train and test sets comprising 20% and 80% of total data, respectively. To run the experiments, we randomly sample (with replacement) $s = 1\%$ to $s = 10\%$ of total tensor entries in the nonparameteric model (i.e. $s \times 10 \times 5 \times 5 \times 24 = 6000 \times s$ entries). For each case of sample percentage s, we obtain the completed tensor $\widehat{\psi}$ and calculate the root mean square error (RMSE) on the test set X_{test}, i.e. $\sqrt{\sum_{x \in X_{\text{test}}} \|\widehat{\psi}_x - \psi_x^{\text{test}}\|^2 / \#X_{\text{test}}}$. To tune the hyperparameters of our and other algorithms, we do a grid search over the values in $[0.5 \cdot \|\psi^{\text{train}}\|_{\max}, 1.5 \cdot \|\psi^{\text{train}}\|_{\max}]$, where $\|\psi^{\text{train}}\|_{\max}$ is the maximum entry in the training set. The results plotted in Figure 4 show that our model outperforms other models at low sample percentages ranging from 1 to 10%, although with a slightly larger computation time for model training. However, we recall that model predictions have extremely low computational load, as described above.

VI. Conclusion

We developed a (general) tensor completion algorithm by defining a tensor norm using the gauge of a specific polytope. The method balances computation and information-theoretic limits: it provably converges in a linear number of integer optimization oracle calls while achieving the information-theoretic sample complexity rate. Numerical experiments confirm its scalability and accuracy, and an application to household energy data shows its potential for building nonparametric models of battery charging. Unlike other nonparametric models (e.g., random forests, neural networks), our approach is extremely efficient at prediction, although computationally intensive to train, making it promising for billion-devicescale energy grid management. Future work

[1]Experiments were performed on a computer server running a Linux OS, with 16GB of RAM and an Intel Xeon Processor E5-2650L v3 (30M Cache, 1.80 GHz) that has 12 cores and became available in the year 2014. The algorithm was implemented in Python 3, and Gurobi v9.1 [50] was used as an integer optimization solver for the separation oracle (10).

[2] PyTen is available at `https://github.com/datamllab/pyten` package [52] under a GPL 2 license.

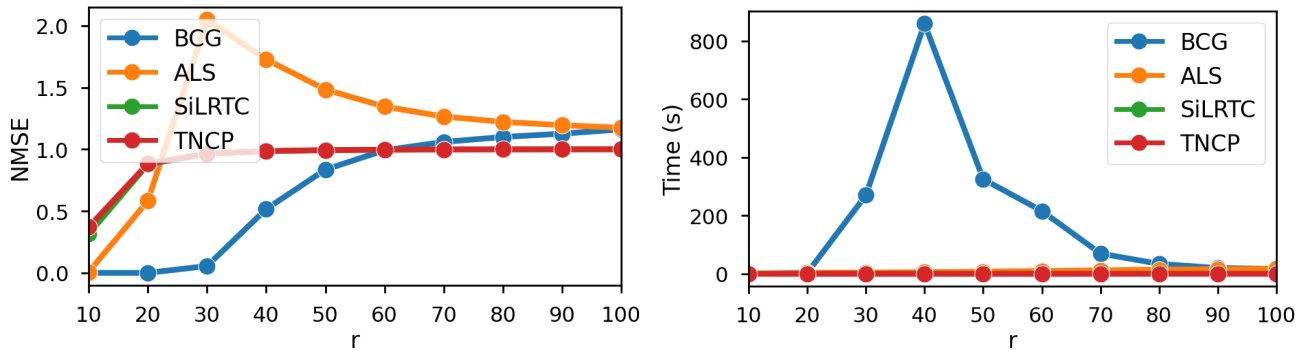

Fig. 1: NMSE and computation time (in s) for order-3 tensors with size $r \times r \times r$ and $n = 1000$ samples.

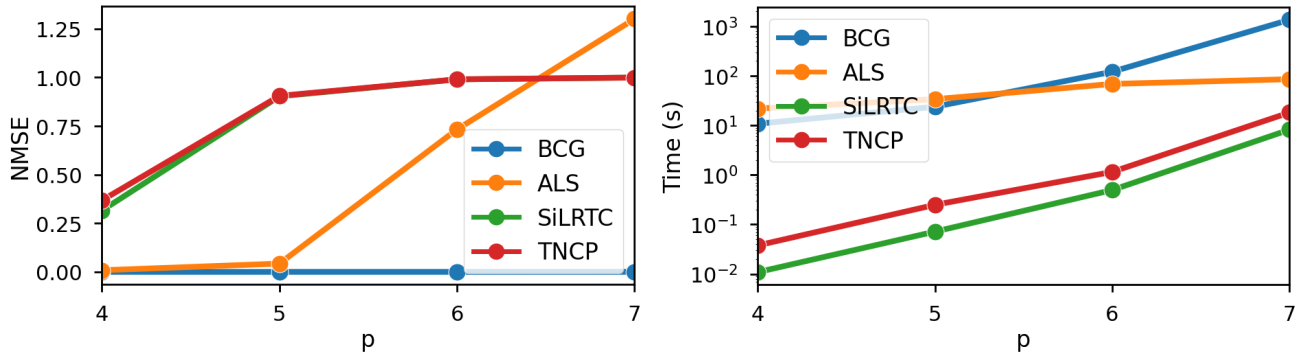

Fig. 2: NMSE and computation time (in s) for increasing order tensors with size $10^{\times p}$ and $n = 10,000$ samples.

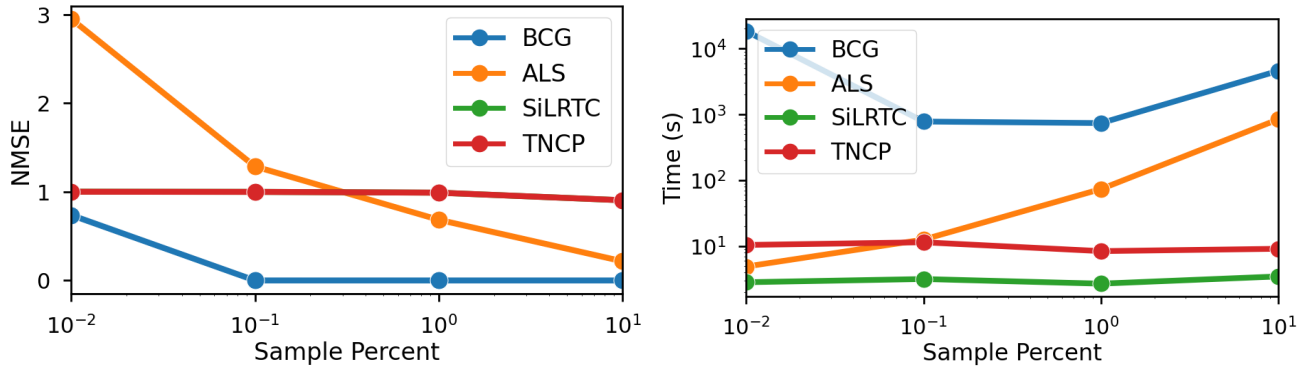

Fig. 3: NMSE and computation time (in s) for tensors with size $10^{\times 7}$ and increasing n samples.

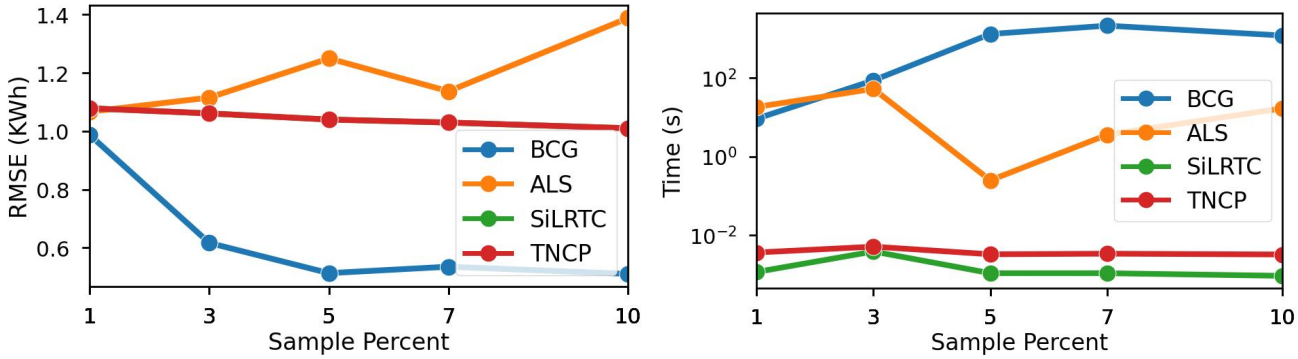

Fig. 4: RMSE (kWh) and computation time(s) for the tensor model predicting energy recharged into storage batteries.

will focus on accelerating training to further enhance performance.

References

[1] C. Hillar and L.-H. Lim, "Most tensor problems are np-hard," *J. ACM*, vol. 60, pp. 45:1–45:39, Nov. 2013.

[2] H. Ge, J. Caverlee, and H. Lu, "Taper: A contextual tensor-based approach for personalized expert recommendation," RecSys 16, p. 261268, Sept. 2016.

[3] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *RecSys 10*, p. 7986, Sept. 2010.

[4] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*. Cambridge University Press, 1 ed., Apr. 2014.

[5] A. Aswani, "Low-rank approximation and completion of positive tensors," *SIMAX*, vol. 37, no. 3, pp. 1337–1364, 2016.

[6] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, p. 494504, Feb. 2012.

[7] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. K. Suykens, "Tensor versus matrix completion: A comparison with application to spectral data," *IEEE Signal Proc. Lett.*, vol. 18, p. 403406, July 2011.

[8] H. Tan, B. Cheng, J. Feng, G. Feng, W. Wang, and Y.-J. Zhang, "Low-n-rank tensor recovery based on multi-linear augmented lagrange multiplier method," *Neurocomputing*, vol. 119, p. 144152, Nov. 2013.

[9] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Rank regularization and bayesian inference for tensor completion and extrapolation," *IEEE Transactions on Signal Processing*, vol. 61, p. 56895703, Nov. 2013.

[10] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, p. 4156, Mar. 2011.

[11] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, p. 21232133, Aug. 2016.

[12] Z. Zhang, S. Mou, K. Paynabar, and J. Shi, "Tensor-based temporal control for partially observed high-dimensional streaming data," *Technometrics*, vol. 66, p. 227239, Apr. 2024.

[13] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse problems*, vol. 27, no. 2, p. 025010, 2011.

[14] A. Montanari and N. Sun, "Spectral algorithms for tensor completion," *Comm. Pure Appl. Math.*, vol. 71, no. 11, pp. 2381–2425, 2018.

[15] S. Jain, A. Gutierrez, and J. Haupt, "Noisy tensor completion for tensors with a sparse canonical polyadic factor," in *IEEE ISIT*, p. 21532157, June 2017.

[16] B. Barak and A. Moitra, "Noisy tensor completion via the sum-of-squares hierarchy," in *COLT*, pp. 417–445, PMLR, 2016.

[17] G. Tomasi and R. Bro, "Parafac and missing values," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, p. 163180, Feb. 2005.

[18] S. Javed, T. Bouwmans, and S. K. Jung, "Stochastic decomposition into low rank and sparse tensor for robust background subtraction," in *ICDP-15*, p. 16, July 2015.

[19] D. Kressner, M. Steinlechner, and B. Vandereycken, "Low-rank tensor completion by riemannian optimization," *BIT Numerical Mathematics*, vol. 54, no. 2, pp. 447–468, 2014.

[20] M. Yuan and C.-H. Zhang, "On tensor completion via nuclear norm minimization," *Foundations of Computational Mathematics*, vol. 16, no. 4, pp. 1031–1068, 2016.

[21] Z. Hu, F. Nie, R. Wang, and X. Li, "Low rank regularization: A review," *Neural Networks*, vol. 136, p. 218232, 2021.

[22] S. Friedland and L.-H. Lim, "Nuclear norm of higher-order tensors," *Mathematics of Computation*, vol. 87, no. 311, 2017.

[23] N. Rao, P. Shah, and S. Wright, "Forwardbackward greedy algorithms for atomic norm regularization," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5798–5811, 2015.

[24] C. Cai, G. Li, H. V. Poor, and Y. Chen, "Nonconvex low-rank tensor completion from noisy data," *Neurips*, vol. 32, 2019.

[25] C. Bugg, C. Chen, and A. Aswani, "Nonnegative tensor completion via integer optimization," *Neurips*, vol. 35, pp. 10008–10020, 2022.

[26] W. Pan, A. Aswani, and C. Chen, "Accelerated non-negative tensor completion via integer programming," *Frontiers in Applied Mathematics and Statistics*, vol. 9, 2023.

[27] G. Braun, S. Pokutta, D. Tu, and S. Wright, "Blended conditonal gradients," in *ICML*, pp. 735–743, PMLR, 2019.

[28] V. de Silva and L. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[29] Y. Qi, P. Comon, and L.-H. Lim, "Uniqueness of nonnegative tensor approximations," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2170–2183, 2016.

[30] R. Zafar, A. Mahmood, S. Razzaq, W. Ali, U. Naeem, and K. Shehzad, "Prosumer based energy management and sharing in smart grid," *Renew. Sustain. Energy Rev.*, vol. 82, p. 16751684, 2018.

[31] A. Tavakoli, S. Saha, M. T. Arif, M. E. Haque, N. Mendis, and A. M. Oo, "Impacts of grid integration of solar pv and electric vehicle on grid stability, power quality and energy economics: a review," *IET Energy Systems Integration*, vol. 2, p. 243260, Sept. 2020.

[32] R. Trivedi, M. Bahloul, A. Saif, S. Patra, and S. Khadem, "Comprehensive dataset on electrical load profiles for energy community in ireland," *Scientific Data*, vol. 11, p. 621, June 2024.

[33] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[34] M. Jaggi, "Revisiting frank-wolfe: Projection-free sparse convex optimization," in *ICML*, p. 427435, 2013.

[35] R. Drenick, "Multilinear programming: Duality theories," *JOTA*, vol. 72, no. 3, pp. 459–486, 1992.

[36] R. Rockafellar and R. Wets, *Variational Analysis*. Springer, 2009.

[37] R. T. Rockafellar, *Convex Analysis: (PMS-28)*. Princeton University Press, Apr. 2015.

[38] M. Conforti, G. Cornuéjols, and G. Zambelli, *Integer Programming*, vol. 271. Springer International Publishing, 2014.

[39] S. Friedland and L.-H. Lim, "The computational complexity of duality," *SIOPT*, vol. 26, no. 4, pp. 2378–2393, 2016.

[40] J. Rohn, "Computing the norm $\|a\|_{\infty,1}$ is np-hard," *Linear and Multilinear Algebra*, vol. 47, no. 3, p. 195204, 2000.

[41] J. M. Hendrickx and A. Olshevsky, "Matrix p -norms are np-hard to approximate if $p \neq 1, 2, \infty$," *SIMAX*, vol. 31, no. 5, p. 28022812, 2010.

[42] P. Bartlett and S. Mendelson, "Rademacher and gaussian complexities: Risk bounds and structural results," *J. Mach. Learn. Res.*, 2002.

[43] N. Srebro, K. Sridharan, and A. Tewari, "Smoothness, low noise and fast rates," in *Neurips*, pp. 2199–2207, 2010.

[44] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[45] P. Massart, "Some applications of concentration inequalities to statistics," *Annales de la faculté des sciences de Toulouse Sér. 6*, vol. 9, no. 2, pp. 245–303, 2000.

[46] A. Nemirovski, "Topics in non-parametric statistics," *Ecole dEté de Probabilités de Saint-Flour*, vol. 28, p. 85, 2000.

[47] A. B. Tsybakov, "Optimal rates of aggregation," in *Learning theory and kernel machines*, pp. 303–313, Springer, 2003.

[48] G. Lecué, "Empirical risk minimization is optimal for the convex aggregation problem," *Bernoulli*, vol. 19, no. 5B, pp. 2153–2166, 2013.

[49] M. Lewis and F. Glover, "Quadratic unconstrained binary optimization problem preprocessing: Theory and empirical analysis," *Networks*, vol. 70, no. 2, pp. 79–97, 2017.

[50] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2021.

[51] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[52] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *ACM TKDD*, vol. 13, no. 1, pp. 1–48, 2019.

[53] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 208–220, 2012.

[54] Y. Liu, F. Shang, L. Jiao, J. Cheng, and H. Cheng, "Trace norm regularized candecomp/parafac decomposition with missing data," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2437–2448, 2014.

[55] C. Navasca, L. De Lathauwer, and S. Kindermann, "Swamp reducing technique for tensor decomposition," in *2008 16th European Signal Processing Conference*, pp. 1–5, IEEE, 2008.