Joint Learning of Drug-Drug Combination and Drug-Drug Interaction via Coupled Tensor-Tensor Factorization with Side Information

Accepted at RECOMB 2026

Xiaoge Zhang[1], Zhengyu Fang[1], Kaiyu Tang[1], Huiyuan Chen[1], Jing Li[1]

[1]Case Western Reserve University, Cleveland, Ohio, USA
Emails: {xxz705, zxf177, kxt439, hxc501, jxl175}@case.edu

Abstract

Targeted drug therapies offer a promising approach for treating complex diseases, with combinational drug therapies often employed to enhance therapeutic efficacy. However, unintended drug-drug interactions may undermine treatment outcomes or cause adverse side effects. In this work, we propose a novel joint learning framework for the simultaneous prediction of effective drug combinations and drug-drug interactions, based on coupled tensor-tensor factorization. Specifically, we model drug combination therapies and DDI by representing drug-drug-disease associations and drug-drug interaction profiles as coupled three-way tensors. To address the challenges of data incompleteness and sparsity, the proposed model integrates auxiliary drug similarity information, such as chemical structure similarities, drug-specific side effects, drug target profiles, and drug inhibition data on cancer cell lines, within a multi-view learning framework. For optimization, we adopt a modified Alternating Direction Method of Multipliers (ADMM) algorithm that ensures convergence while enforcing non-negativity constraints. In addition to standard tensor completion tasks, we further evaluate the proposed method under a more realistic *new-drug prediction* setting, where all interactions involving a previously unseen drug are withheld. This scenario closely aligns with real-world applications, in which reliable predictions for emerging or under-studied compounds are essential. We evaluate the proposed method on a comprehensive dataset compiled from multiple sources, including DrugBank, CDCDB, SIDER, and PubChem. Our experiments show that SI-ADMM maintains robust performance and achieves the best results comparing to other tensor factorization approaches, with or without auxiliary information, particularly in the new-drug prediction setting. The implementation of our method is publicly available at: `https://github.com/Xiaoge-Zhang/SI-ADMM`.

Keywords: Machine Learning, Tensor Decomposition, Drug-Drug Interaction Prediction, Drug Combination Therapy

1 Introduction

In the treatment of complex diseases such as cancer and neurological disorders, combinational drug therapies have shown promise in improving treatment efficacy by targeting multiple biological pathways simultaneously [37]. However, as the number of available drugs continues to grow, the space of possible drug combinations expands rapidly, increasing the risk of undetected drug-drug interactions (DDIs) [46]. Such interactions can reduce therapeutic effectiveness, induce adverse side effects [17], and in severe cases lead to increased hospitalization [10], posing significant challenges to the screening and optimization of combinational drug therapies. Unexpected DDIs can therefore substantially undermine the benefits of combination treatments, either by diminishing efficacy or by introducing harmful effects. Consequently, there is a critical need for computational methods that can accurately predict both beneficial drug-drug synergies and adverse interactions to support safer and more effective therapeutic strategies. Strictly speaking, drug combination therapies represent one category of drug-drug interactions that rely on synergistic or additive effects to increase efficacy, reduce toxicity, and/or prevent drug resistance. Throughout this paper, we use the terms *drug combination* and *drug-drug synergy* interchangeably to denote positive interactions, while the term *drug-drug interaction* is primarily used to refer to adverse interactions that may cause unexpected side effects.

In recent years, a variety of machine learning and computational approaches have been proposed to predict DDIs and drug-drug synergies [52, 11, 12], with several methods leveraging tensor decomposition or tensor completion techniques [54]. Tensor-based models are particularly attractive because they can naturally integrate multi-relational information involving drugs, targets, and diseases within a unified representation. Despite these advantages, most existing studies treat DDI prediction and drug-drug synergy prediction as separate problems [38, 43], and relatively few efforts have attempted to jointly model these two closely related interaction types within a single computational framework.

The motivation for our study is rooted in the observation that drug-drug synergies and adverse DDIs are essentially two sides of the same coin: while one emphasizes positive therapeutic effects, the other focuses on identifying harmful interactions that may lead to side effects. Although the outcomes differ, these interaction types may share underlying biological and pharmacological mechanisms. Jointly modeling both forms of interactions has the potential to improve predictive accuracy and provide richer insight into interaction effects. However, this problem cannot be reduced to a simple binary classification task that assigns positive or negative labels to drug pairs, as interaction outcomes are highly context-dependent, influenced by factors such as disease specificity and interaction mechanisms.

To address these challenges, we propose a novel joint learning framework based on coupled tensor-tensor decomposition that simultaneously models drug-drug-disease associations and adverse drug-drug interaction profiles. To mitigate the severe sparsity inherent in these data sources, we incorporate multiple forms of auxiliary drug side information—including chemical structures, individual drug side effects, and drug target profiles—within a multi-view learning architecture. Building on this formulation, we develop an efficient optimization algorithm based on the Alternating Direction Method of Multipliers (ADMM) [4, 35, 20, 41, 5], enhanced with structured side information, which we refer to as **SI-ADMM**. We evaluate the proposed method on a comprehensive dataset compiled from DrugBank [21], PubChem [19], DrugCombDB [30], SIDER [25], and BindingDB [13]. Experimental results demonstrate that SI-ADMM consistently outperforms existing tensor decomposition methods in predicting both drug-drug synergies and adverse interactions, while ablation studies further confirm its robustness across a wide range of parameter settings.

The main contributions of this paper are summarized as follows:

1. We propose a unified joint learning framework that integrates drug synergy prediction and adverse interaction prediction via coupled tensor-tensor decomposition.

2. We develop SI-ADMM, an enhanced ADMM-based optimization algorithm that incorporates auxiliary information to mitigate data sparsity.

3. We demonstrate the effectiveness of SI-ADMM on large-scale datasets compiled from multiple biomedical databases, achieving competitive performance over state-of-the-art tensor decomposition methods.

4. We conduct comprehensive ablation studies showing the robustness of SI-ADMM, further validating its practical utility.

2 Related Work

Tensor decomposition has become an effective and widely used framework for modeling and predicting drug-drug interactions (DDIs) and drug-drug synergies, particularly due to its ability to capture higher-order relationships among multiple biological entities. Classical tensor decomposition techniques, including Tucker decomposition and CANDECOMP/PARAFAC (CP) decomposition [22], have been extensively applied to biomedical data analysis. In the context of DDI prediction, these methods are typically used to recover missing entries in partially observed tensors by learning low-rank latent factors that encode interaction patterns among drugs, diseases, or experimental conditions.

Beyond purely latent factor models, a substantial body of work has emphasized the importance of incorporating auxiliary or side information to improve predictive performance and biological interpretability. Side information may include chemical structure similarities, molecular fingerprints, gene expression profiles, protein-protein interaction networks, pathway annotations, or known drug-target associations. Prior studies have consistently shown that integrating such heterogeneous information sources helps mitigate data sparsity and enhances generalization, particularly in cold-start or new-drug scenarios [9, 6]. These approaches align with the guilt-by-association principle [7, 24], where drugs sharing similar biological or chemical characteristics are more likely to exhibit similar interaction behaviors.

Several tensor-based models have explicitly incorporated side information through regularization or constrained factorization. For example, Zhang et al. [52] proposed a constrained tensor factorization (CTF-DDI) model that embeds biological constraints and similarity information into the tensor decomposition process, enabling improved robustness under extreme sparsity. Ding et al. [11] introduced a deep tensor factorization (DTF) framework for anticancer drug synergy prediction, leveraging deep neural architectures to fuse multiple data modalities. While these methods demonstrate strong empirical performance, they typically focus on a single tensor and do not explicitly model interactions across multiple related relational tensors.

In parallel, coupled and joint factorization methods have been explored in broader machine learning and signal processing communities. Joint tensor decomposition and coupled tensor-matrix factorization have shown success in applications such as multi-view learning, recommender systems, and image processing [43]. In the biomedical domain, coupled tensor-matrix decomposition has been applied to tasks such as drug-target interaction prediction and multi-relational biological data integration [15]. These methods exploit shared latent factors across different data sources to capture consistent underlying structures while preserving modality-specific information.

Despite these advances, the application of joint tensor decomposition to drug discovery remains relatively limited. In particular, existing DDI and drug synergy models often treat different relational datasets independently, even when they involve the same set of drugs. In this work, we address this gap by jointly decomposing two related tensors: one encoding drug-drug-disease relationships and the other representing drug-drug interaction outcomes. By sharing latent drug factors across both tensors, our approach enables information transfer between complementary tasks, facilitating simultaneous prediction of drug combination efficacy and potential adverse interactions. Furthermore, we incorporate multiple auxiliary similarity matrices into the optimization framework, allowing the model to leverage diverse biological prior knowledge while maintaining a unified and interpretable latent representation.

3 Problem Formulation

3.1 Prediction as Tensor Recovery

To simultaneously predict drug-drug combinations and adverse drug-drug interaction within their context, we assume that drug-drug-disease relationships are encoded in a three-way $drug \times drug \times disease$ tensor $\mathcal{X}$ of size $n \times n \times m$, where n and m represent the number of drugs and the number of diseases, respectively. An entry $\mathcal{X}_{ijk} = 1$ indicates that drug i and drug j have a synergistic effect on disease k, while entries are set to 0 otherwise.

We assume that drug-drug interactions are encoded in a three-way $drug \times drug \times DDI$ tensor $\mathcal{Y}$ of size $n \times n \times t$, where n and t denote the number of drugs and the number of DDI types, respectively. An entry $\mathcal{Y}_{ijk} = 1$ indicates that drug i and drug j are reported to DDI k when used together, with entries set to 0 otherwise. We only have incomplete observations for both tensors $\mathcal{X}$ and $\mathcal{Y}$, which are denoted as $\mathcal{T}$ and $\mathcal{O}$, respectively. The goal of joint drug-drug combination and drug-drug interaction prediction is to recover $\mathcal{X}$ and $\mathcal{Y}$ from $\mathcal{T}$ and $\mathcal{O}$. We formulate this problem as a coupled tensor-tensor completion problem because the two tensors share the same set of drugs. Throughout the paper, we use calligraphic letters ($\mathcal{X}, \mathcal{Y}$) to denote tensors, and bold uppercase letters to denote matrices.

3.2 Side Information for Data Sparsity

In real-world applications, tensors $\mathcal{T}$ and $\mathcal{O}$ are often sparse, with many unknown entries. Recovering these unobserved entries based solely on the tensors is challenging. To address this, we incorporate additional auxiliary information and formulate the problem under the multi-view learning framework. In this work, we primarily focus on side information related to drugs, because the same set of drugs is present in both tensors.

We further assume that all the side information can be organized into similarity/kernel matrices, which are then integrated into the learning process and jointly decomposed with the two tensors. We denote the similarity matrices constructed from n_a views of drugs as $\mathscr{S} = \{\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(n_a)}\}$. In practice, these similarity matrices are constructed based on different properties of drugs such as their chemical structures and side effect information when drugs are used alone. We will discuss more side information in Section 5. The overall architecture of our proposed approach is illustrated in Fig 1.

4 Methods

4.1 Tensor Decomposition

To recover the drug-drug-disease tensor $\mathcal{X}$ and the DDI tensor $\mathcal{Y}$, we adopt the INDSCAL decomposition method, which is a special case of CP decomposition [22]. Given a three-dimensional tensor $\mathcal{M}$, the goal of CP decomposition is to represent $\mathcal{M}$ as a summation of a set of rank one tensors:

$$\mathcal{M} \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r$$

where R is a hyperparameter, representing the rank after the decomposition. Vectors a_r, b_r, c_r are latent factors, and $\circ$ denote the vector outer product.

In our problem setup, both tensor $\mathcal{X}$ and tensor $\mathcal{Y}$ have drugs as their first and second mode, which imposes an additional constraint where the latent drug factor in the decomposition must be identical for both tensors, essentially creating a symmetric structure across the drug-drug slices of the tensors. INDSCAL decomposition is precisely for such a structure. More specifically, we want to decompose the two tensors $\mathcal{X}$ and $\mathcal{Y}$ simultaneously as follows:

$$\begin{aligned} \mathcal{X} &\approx \textstyle\sum_{r=1}^{R} u_r \circ u_r \circ v_r \\ \mathcal{Y} &\approx \textstyle\sum_{r=1}^{R} u_r \circ u_r \circ w_r \end{aligned}$$

where u_r, v_r, and w_r represent the drug, disease, and DDI latent factors for the rth rank component, respectively.

4.2 Incorporating Side Information

To overcome the difficulties of decomposing highly sparse tensors, we propose a novel addition to the loss function to incorporate side information of drugs, as mentioned in the problem formulation section. One challenge to directly incorporating similarity matrices from side information is that many of them have significant missing values. To address this challenge, we further decompose each of the similarity matrices into its own latent space and require all lower rank matrices in the latent spaces share some commonality. Finally, the tensor decomposition loss and the side information decomposition loss are combined into one loss function, which is defined as follows:

$$\begin{aligned} \min f = {} & a \left\| \mathcal{X} - [\![U, U, V]\!] \right\|_F^2 + b \left\| \mathcal{Y} - [\![U, U, W]\!] \right\|_F^2 \\ & + \sum_{i=1}^{n_a} \alpha_i \left(\left\| A^{(i)} \circ (S_A^{(i)} - U^{(i)} U^{(i)^T}) \right\|_F^2 + \left\| U^{(i)} Q^{(i)} - U \right\|_F^2 \right) \\ & \text{s.t.} \quad \mathcal{P}_{\Omega_1}(\mathcal{X}) = \mathcal{P}_{\Omega_1}(\mathcal{T}), \mathcal{P}_{\Omega_2}(\mathcal{Y}) = \mathcal{P}_{\Omega_2}(\mathcal{O}), \\ & \mathbf{U} \geq 0,\ \mathbf{V} \geq 0,\ \mathbf{W} \geq 0,\ \mathbf{U^{(i)}} \geq 0 \end{aligned} \tag{1}$$

where $[\mathbf{U}, \mathbf{U}, \mathbf{V}] = \sum_r u_r \circ u_r \circ v_r$ is a shorthand for the sum of rank-one tensors; $\mathbf{U}$ is the latent factor matrix of the drug mode for both tensor $\mathcal{X}$ and tensor $\mathcal{Y}$; $\mathbf{U^{(i)}}$ is the latent drug factor matrix from the i-th view of side information;

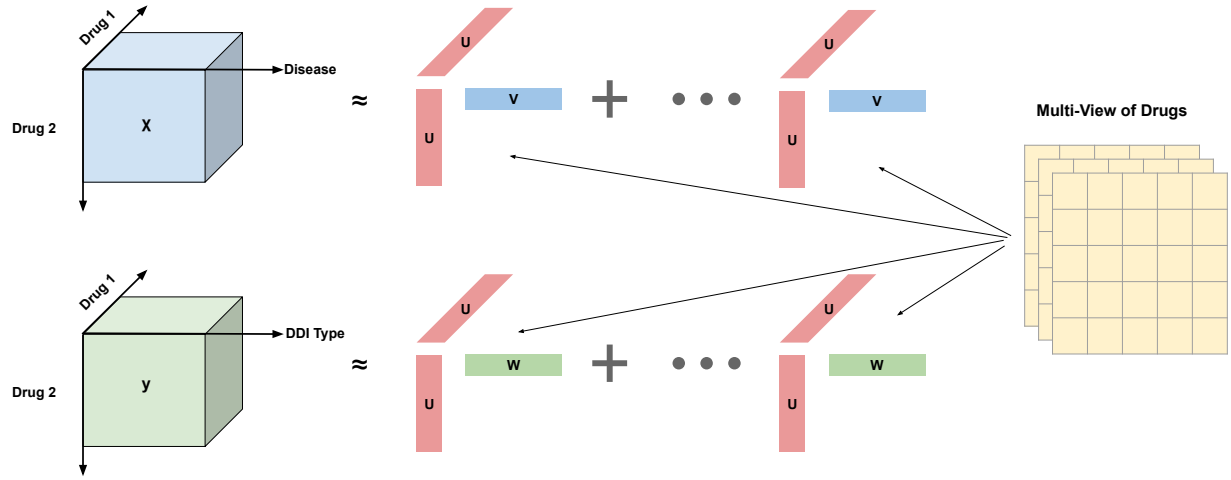


Figure 1: The architecture of SI-ADMM showing the joint tensor-tensor decomposition is aided by auxiliary side information from different views of drugs.

$\mathbf{V}$ is the latent disease factor matrix for tensor $\mathcal{X}$, and $\mathbf{W}$ is the latent factor matrix for drug-drug interaction tensor $\mathcal{Y}$. a and b are weight factors which control the weight of the reconstruction losses of $\mathcal{X}$ and $\mathcal{Y}$ in the case of two tensors having disproportionate sizes. Ω_i represents the set of observed entries, and $\mathcal{P}_{\Omega_i}(\cdot)$ retains the entries in Ω_i while zeroing out the others [18, 47]. The non-negativity constraints on the latent factors $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{U^{(i)}}$ help to yield more interpretable and intuitive decomposition results [35, 49].

The last term simultaneously decomposes all drug similarity matrices $S_A^{(i)}$ into latent factor matrices $\mathbf{U^{(i)}}$, and further requires that $\mathbf{U^{(i)}}$ should be close to the drug latent factor matrix $\mathbf{U}$ from the tensor decomposition, up to a scaling factor $\mathbf{Q^{(i)}}$. $A^{(i)}$ is a binary masking matrix for each side of information to mitigate the impact of missing values in the side information. The entry has a value of one for the existing value in the side information matrix and zero otherwise. In this work, we use $\circ$ to represent Hadamard Product. This last term can also be viewed as an application of guilt-by-association principle [7, 24]. The idea is to use matrix factorization [7, 8, 32] to learn multiple incomplete kernel matrices from multiple data sources representing different views of drugs. Because all the views represent the same entities, they should share some common latent structures. By factorizing all views of drugs into similar latent factors and requiring them to be consistent with the latent factor from the tensors, we can effectively link all the side information together, and link all the side information with the two main tensors representing drug combinations and adverse drug-drug interactions. In addition, by factorizing multiple views simultaneously, this term also mitigates the practical problem of missing data contained in each data source, as suggested by existing work [31, 9]. The purpose of the scaling matrix $\mathbf{Q^{(i)}}$ is to make different views comparable to each other after the decomposition, which is a diagonal matrix that sums over the columns of $U^{(i)}$: $\mathbf{Q^{(i)}} = \text{Diag}\left(\sum_\tau \mathbf{U}^{(\mathbf{i})}_{\tau,1}, \sum_\tau \mathbf{U}^{(\mathbf{i})}_{\tau,2}, \ldots, \sum_\tau \mathbf{U}^{(\mathbf{i})}_{\tau,R}\right)$. The weight parameter α_i represents the relative importance of the i-th view of drugs, *i.e.*, how influential $\mathbf{S^{(i)}}$ is in shaping the decomposition, and can be determined based on prior knowledge about the side information or through empirical studies.

4.3 Learning Algorithm

Many algorithms have been proposed to solve the non-negative tensor decomposition problem [49, 22, 35, 1, 2]. The majority of the optimization algorithms use multiplicative update rules [49] or the gradient descent method [1]. However, the non-negativity constraints may bring a significant computational burden with slow convergence and complicate the development of parallel algorithms to compute large datasets [28]. Also, in such cases, the INDSCAL model is not well understood when the factorization is symmetric [2, 22]. In this work, we adapt the ADMM algorithm to optimize the latent factor matrices based on the partially augmented Lagrangian method. Notice that the objective function in Eq.(1) is non-convex w.r.t. $\mathbf{U}, \mathbf{V}, \mathbf{W}$, and $\mathbf{U^{(i)}}$ all together and it involves fourth-order terms w.r.t. $\mathbf{U}$ and $\mathbf{U^{(i)}}$, which is hard to optimize directly. To mitigate such problems, SI-ADMM consists of the following steps. First,

Algorithm 1 Pseudo-code of SI-ADMM

Require: Drug-Drug-Disease tensor $\mathcal{X}$, DDI tensor $\mathcal{Y}$, side information matrices $\boldsymbol{S_A^i}$, number of side information n_a, penalty parameter ρ, η, maximum iterations $iter$, gradient descent iterations $inner_iter$, tolerance tol, no improvement limit $limit$.

Randomly initialize $\mathbf{U}, \mathbf{D}, \mathbf{V}, \mathbf{W}, \mathbf{C^{(i)}}, \mathbf{U^{(i)}}$
Set $\mathcal{X}^{(0)} = \mathcal{T}, \mathcal{Y}^{(0)} = \mathcal{O}$, $F^{(0)} = {Y^{(0)}}^{(i)} = 0$, $\rho^{(0)} = \eta^{(0)} = 10^{-7}$
repeat
    Update U, D, V, W by Eq.(4) - (7)
    Update F by Eq.(10)
    for $i = 1$ to n_a **do**
        for $j_{U^{(i)}} = 1$ to $inner_iter$ **do**
            Update $U^{(i)}$ by Eq.(8)
        end for
        for $j_{C^{(i)}} = 1$ to $inner_iter$ **do**
            Update $C^{(i)}$ by Eq.(9)
        end for
        Update $Y^{(i)}$ by Eq.(10)
    end for
    Update $\mathcal{X}^{(t+1)}$ and $\mathcal{Y}^{(t+1)}$ by Eq.(12)
    Update ρ and η by Eq.(11)
until $\frac{\|\mathcal{X}^{(t+1)}-\mathcal{X}^{(t)}\|_F^2}{\|\mathcal{X}^{(t)}\|_F^2} + \frac{\|\mathcal{Y}^{(t+1)}-\mathcal{Y}^{(t)}\|_F^2}{\|\mathcal{Y}^{(t)}\|_F^2} \leq tol$
or No loss improvement iterations $\geq limit$
return $\mathcal{X}$ and $\mathcal{Y}$

it converts the objective function to a function of a lower order by using the variable splitting technique. Second, to solve the optimization problem with constraints, we rely on the augmented Lagrangian method, and the new objective function is derived. The third step consists of the core of ADMM, which updates one variable while fixing the others. The details of these steps are discussed in the sequel.

4.3.1 Lower Order Function Using Variable Splitting

To use the variable splitting technique, we define two sets of new variables $\mathbf{D}$ and $\mathbf{C}^{(i)}$ to replace some occurrences of $\mathbf{U}$ and $\mathbf{U}^{(i)}$ in Eq.(1), and simultaneously add the constraints $\mathbf{D} = \mathbf{U}$ and $\mathbf{C}^{(i)} = \mathbf{U}^{(i)}$ to the system, which enables the reduction of the order of the objective function from 4 to 2 and makes the variables more separable for the ADMM algorithm. The new objective function becomes:

$$\begin{aligned}
\min f = {} & a \left\|\mathcal{X} - [\![U, D, V]\!]\right\|_F^2 + b \left\|\mathcal{Y} - [\![U, D, W]\!]\right\|_F^2 \\
& + \sum_{i=1}^{n_a} \alpha_i \left(\left\| A^{(i)} \circ (S_A^{(i)} - C^{(i)} {U^{(i)}}^T) \right\|_F^2 + \left\| U^{(i)} Q^{(i)} - U \right\|_F^2 \right) \\
& \text{s.t.} \quad \mathbf{D} = \mathbf{U}; \mathbf{C^{(i)}} = \mathbf{U^{(i)}}; \\
& P_{\Omega 1}(\mathcal{X}) = P_{\Omega 1}(\mathcal{T}); P_{\Omega 2}(\mathcal{Y}) = P_{\Omega 2}(\mathcal{O}); \\
& \mathbf{U} \geq 0; \mathbf{D} \geq 0; \mathbf{V} \geq 0; \mathbf{W} \geq 0; \mathbf{U^{(i)}} \geq 0; \mathbf{C^{(i)}} \geq 0
\end{aligned} \tag{2}$$

4.3.2 Partially Augmented Lagrangian Function

To utilize the ADMM algorithm to solve the optimization problem, we form the following partially augmented Lagrangian function $\mathcal{L}$:

$$
\begin{aligned}
\mathcal{L} =& f + \langle F, \mathbf{D} - \mathbf{U} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{U}\|_F^2 \\
&+ \sum_{i=1}^{n_a} \left(\left\langle Y^{(i)}, \mathbf{C^{(i)}} - \mathbf{U^{(i)}} \right\rangle + \frac{\eta}{2} \left\| \mathbf{C^{(i)}} - \mathbf{U^{(i)}} \right\|_F^2 \right) \\
&\text{s.t.} \quad P_{\Omega 1}(\mathcal{X}) = P_{\Omega 1}(\mathcal{T}); P_{\Omega 2}(\mathcal{Y}) = P_{\Omega 2}(\mathcal{O}); \\
&\mathbf{U} \geq 0; \mathbf{D} \geq 0; \mathbf{V} \geq 0; \mathbf{W} \geq 0; \mathbf{U^{(i)}} \geq 0; \mathbf{C^{(i)}} \geq 0
\end{aligned} \tag{3}
$$

where f is the objective function in Eq.(2). The notation $\langle A, B \rangle$ denotes the inner product of two matrices. The additional linear order terms are added to handle the constraints $\mathbf{U} = \mathbf{D}$ and $\mathbf{C}^{(i)} = \mathbf{U}^{(i)}$, where F and $Y^{(i)}$ in these terms are Lagrange multiplier. The additional second-order terms are for regularization. Penalty parameters ρ and η are applied to the constraint of $U = D$ and side information constraints of $C^{(i)} = U^{(i)}$, respectively.

4.3.3 The SI-ADMM Algorithm

Following the original ADMM framework, we adopt an alternating optimization strategy, where each latent matrix is updated sequentially while keeping the others fixed until convergence is achieved. The update process for each matrix consists of three key steps:

1. **Term Identification:** Extract all terms from the partial augmented Lagrangian that depend on the matrix to be updated.
2. **Derivative Computation:** Calculate the partial derivative with respect to the target matrix, and derive the closed-form solution by setting the derivative to zero.
3. **Matrix Update:** Apply the derived update rule and use the updated matrix in the subsequent optimization steps.

We present full derivations for the updates of U, the Lagrangian multipliers, penalty terms, and associated tensor components in the following section. Derivations for the remaining latent matrices are provided in the supplementary appendix.

1. Update of U The terms in the Lagrangian $\mathcal{L}$ involving U are expressed as:

$$
\begin{aligned}
\mathcal{L}_U = a \|\mathcal{X} - [\![U, D, V]\!]\|_F^2 + b \|\mathcal{Y} - [\![U, D, W]\!]\|_F^2 \\
+ \sum_{i=1}^{n_a} \alpha_i \left\| U^{(i)} Q^{(i)} - U \right\|_F^2 \\
+ \langle F, D - U \rangle + \frac{\rho}{2} \|D - U\|_F^2
\end{aligned}
$$

Since the direct computation of the tensor reconstruction norm is non-trivial, we leverage the matricization and Khatri-Rao product approximation introduced in [22], where the tensor decomposition can be written as:

$$
\begin{aligned}
\mathcal{M}_{(1)} \approx& A(C \odot B)^T, \\
\mathcal{M}_{(2)} \approx& B(C \odot A)^T, \\
\mathcal{M}_{(3)} \approx& C(B \odot A)^T,
\end{aligned}
$$

Here, $\mathcal{M}$ is a 3-D tensor and A, B, C are the latent factor matrices. $\mathcal{M}_{(1)}, \mathcal{M}_{(2)}, \mathcal{M}_{(3)}$ are the mode-1, mode-2, and mode-3 matricization of the tensor [22]. $\odot$ is Khatri-Rao Product. Using this approximation rule, we can rewrite the

terms involving U into:

$$\begin{aligned}\mathcal{L}_U = & a \left\| U(V \odot D) - \mathcal{X}(1) \right\|_F^2 + b \left\| U(W \odot D) - \mathcal{Y}(1) \right\|_F^2 \\ & + \sum_{i=1}^{n_a} \alpha_i \left\| U^{(i)} Q^{(i)} - U \right\|_F^2 \\ & + \frac{\rho}{2} \left\| U - D - \frac{F}{\rho} \right\|_F^2\end{aligned}$$

Defining $\Pi = V^{(t)} \odot D^{(t)}$ and $\Theta = W^{(t)} \odot D^{(t)}$, and following the matrix derivative identities provided in [36], the partial derivative with respect to U becomes:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial U} = & 2aU\Pi^T\Pi - 2a\mathcal{X}(1)\Pi + 2bU\Theta^T\Theta - 2b\mathcal{Y}(1)\Theta \\ & - \sum_{i=1}^{n_a} 2\alpha_i U^{(i)^{(t)}} Q^{(i)} + \sum_{i=1}^{n_a} 2\alpha_i U + \rho U - \rho D^{(t)} - F\end{aligned}$$

Here t is the tth iteration of optimization. By equating the derivative to zero, the update rule for U is derived as:

$$\begin{aligned}U^{(t+1)} = & (2a\mathcal{X}(1)\Pi + 2b\mathcal{Y}(1)\Theta + \sum_{i=1}^{n_a} 2\alpha_i U^{(i)^{(t)}} Q^{(i)} + \rho^{(t)} D^{(t)} + F^{(t)}) \\ & (2a\Pi^T\Pi + 2b\Theta^T\Theta + \sum_{i=1}^{n_a} 2\alpha_i I + \rho I)^{-1}\end{aligned} \tag{4}$$

2. Update of Lagrangian Multipliers

Unlike standard ADMM, which updates the Lagrangian multipliers for all imposed constraints, our framework updates the multipliers specifically for the variable splitting constraints, namely $U = D$ and $U^{(i)} = C^{(i)}$. The updates are performed as follows:

$$\begin{aligned}F^{(t+1)} &= F^{(t)} + \rho^{(t)} \left(D^{(t+1)} - U^{(t+1)} \right) \\ Y^{(i)^{(t+1)}} &= Y^{(i)^{(t)}} + \eta^{(t)} \left(C^{(i)^{(t+1)}} - U^{(i)^{(t+1)}} \right)\end{aligned} \tag{5}$$

3. Update of Penalty Parameters

To enhance convergence speed, we follow the adaptive penalty update strategy proposed in [29], updating the penalty parameters as:

$$\begin{aligned}\rho^{(t+1)} &= \min(d\rho^{(t)}, \rho_{\max}) \\ \eta^{(t+1)} &= \min(d\eta^{(t)}, \eta_{\max})\end{aligned} \tag{6}$$

where d is a scalar larger than 1, and $\rho_{\max}$ and $\eta_{\max}$ are upper bounds to prevent the penalties from becoming excessively large.

4. Update of Predicted Tensors

At each iteration, the unobserved entries of the tensors are estimated by combining the observed data with the reconstructed values from the current latent factors, according to the following update rules:

$$\begin{aligned}\mathcal{X}^{(t+1)} &= \mathcal{T} + \mathcal{C}_X \circ [\![U^{(t+1)}, D^{(t+1)}, V^{(t+1)}]\!] \\ \mathcal{Y}^{(t+1)} &= \mathcal{O} + \mathcal{C}_Y \circ [\![U^{(t+1)}, D^{(t+1)}, W^{(t+1)}]\!]\end{aligned} \tag{7}$$

Here, $\mathcal{X}^{(t+1)}$ and $\mathcal{Y}^{(t+1)}$ represent the predicted tensors at iteration $t+1$, while $\mathcal{T}$ and $\mathcal{O}$ denote the observed parts of the tensors, as defined in Section 3.1. The complement masks $\mathcal{C}_X$ and $\mathcal{C}_Y$ have the same dimensions as $\mathcal{T}$ and $\mathcal{O}$, respectively, containing zeros at observed entries and ones at unobserved entries. $\circ$ means element-wise product.

5. Summary of the SI-ADMM Algorithm

Combining all the update procedures described above, as well as the initialization and termination steps, the complete SI-ADMM algorithm is summarized in Algorithm 1. For the termination step, we check the weighted reconstruction loss of both tensors $\mathcal{X}$ and $\mathcal{Y}$ against a tolerance threshold at the end of each iteration (line 12 in Algorithm 1), and the algorithm stops when the difference between two consecutive runs is smaller than the threshold or the number of iterations with no improvement in reconstruction loss reaches the limit.

4.4 Complexity Analysis For SI-ADMM

The main computational cost arises from updating the matrices $U, D, V, W, U^{(i)}$, and $C^{(i)}$. The most expensive terms in these updates are as follows. First, the term $\mathcal{X}_{(1)}V \odot D + \mathcal{Y}_{(1)}W \odot D$, where $\odot$ denotes the Khatri-Rao product, requires $\mathcal{O}(N^2 M_{\text{disease}} R + N^2 M_{\text{DDI}} R)$ operations, with N the number of drugs, M_{disease} the number of diseases, and M_{DDI} the number of DDI types. Second, the term $(V \odot D)^\top (V \odot D) + (W \odot D)^\top (W \odot D)$ can be equivalently written as $V^\top V D^\top D + W^\top W D^\top D$, with complexity $\mathcal{O}((N + M_{\text{disease}})R^2 + (N + M_{\text{DDI}})R^2)$. Third, computing the matrix inverse required when separating U, D, V, and W involves a Cholesky decomposition, which has complexity $\mathcal{O}(R^3)$. Finally, solving the update rule as a whole requires $\mathcal{O}(NR^2)$. Therefore, the overall complexity of SI-ADMM is

$$\mathcal{O}\big(N^2 R(M_{\text{disease}} + M_{\text{DDI}}) + (N + M_{\text{disease}} + M_{\text{DDI}})R^2 + R^3\big)\,.$$

Although the rank R appears in higher-order terms, in practice it is much smaller than the number of drugs N or $(M_{\text{disease}} + M_{\text{DDI}})$.

5 EXPERIMENTS

5.1 Dataset Preparation

5.1.1 Data Preparation for $\mathcal{X}$ and $\mathcal{Y}$ tensor

To construct the drug-drug combination tensor and DDI tensor, we curated a task-specific dataset based on two primary sources: CDCDB[1] [39] and DrugBank[2] [21]. CDCDB is a comprehensive repository that integrates drug combination information from multiple resources, including ClinicalTrials.gov[3] [51], international patents, and the FDA Orange Book[4] [45]. In this work, we focused on drug combinations involving exactly two distinct drugs, extracted from both patents and ClinicalTrials.gov. We did not include combinations from the FDA Orange Book because they did not specify medical conditions associated with drug combinations.

In total, we collect 12033 unique drug-drug-disease triplets that involve 2652 unique drugs and 10156 unique medical conditions. The fine-grained classification of medical conditions results in an extreme sparse tensor and makes it practically infeasible to predict any meaningful drug combinations for any diseases. To address this, we map the disease conditions to the second level disease classifications based on the International Classification of Diseases (ICD) hierarchy[5] [50]. This reduces the number of diseases down to 238.

We construct the DDI tensor from the DDI dataset of the DrugBank database[6], which contains manually curated drug-drug interactions from drug labels as well as primary literature. DrugBank provides 86 types of DDI among drug pairs that cover 1706 unique drugs. Using both datasets and cross-referencing various identifiers, we obtain our own dataset consisting of 1070 unique drugs, 238 different disease types, and 81 types of DDI. Therefore the dimensions of $\mathcal{X}$ and $\mathcal{Y}$ tensors are (1070, 1070, 238) and (1070, 1070, 81), respectively. Both tensors are very sparse. The percentages of entries that equal 1 are 0.0025% and 0.106% for $\mathcal{X}$ and $\mathcal{Y}$ tensors, respectively.

[1] https://icc.ise.bgu.ac.il/medical_ai/CDCDB/
[2] https://tdcommons.ai/multi_pred_tasks/ddi
[3] https://clinicaltrials.gov/
[4] https://www.accessdata.fda.gov/scripts/cder/ob/
[5] https://icd.who.int/icdapi
[6] https://tdcommons.ai/multi_pred_tasks/ddi

5.1.2 Formulation of Side Information matrices

We obtained side information similarity matrices from four different sources: the Tainimoto similarity [44] based on SMILE[48] strings, the cosine similarities among the side effect vectors, the global alignment scores of drugs' target sequences, and cosine similarities among the drugs' inhibition scores (IC50) on NCI-60 human cancer cell lines [34]. A detailed description of how we formed such side information matrices is provided below.

Tanimoto scores are derived from the SMILES representations of all drugs. We employ the RDKit [26] Python library to compute the Tanimoto similarity between all pairs of drugs, resulting in a complete 1070×1070 drug-drug similarity matrix without missing.

Side Effect Information is obtained from the SIDER database [25]. Only 776 drugs out of 1070 drugs have side effect information. For each drug, we construct an n-dimensional binary vector, where n represents the total number of side effects recorded in SIDER. An entry in the vector is set to 1 if the drug is associated with the corresponding side effect, and 0 otherwise. We then compute the pairwise cosine similarity between these vectors to generate the drug-drug similarity matrix.

Target Sequence Alignment scores are derived from the target protein sequences associated with each drug in the DrugBank dataset. Out of 1070 drugs, only 890 drugs have their target sequences. For every pair of drugs with target sequences, we apply the Smith-Waterman alignment algorithm [40] to compute the scores between their target sequences. The self-alignment scores of target sequences are then normalized to produce a drug-drug similarity matrix.

IC_{50} scores are obtained from experimental data provided by the NCI-60 project [34]. Out of 1070 drugs, only 235 drugs have IC50 data. For each such drug, we construct a 60-dimensional vector representing its IC_{50} values (a measure of inhibitory potency) across the 60 cancer cell lines. Pairwise cosine similarity is then computed between these vectors to generate the drug-drug similarity matrix.

5.2 Experiment Settings

5.2.1 Random Prediction Task

To ensure a fair evaluation, we adopt the following data splitting procedure. Specifically, to simulate partially observed tensors $\mathcal{O}$ and $\mathcal{T}$ from the fully observed tensors $\mathcal{X}$ and $\mathcal{Y}$, we randomly mask out 10% of the entries with value 1. These masked entries serve as the positive samples in the testing set. An equal number of randomly selected zero-valued entries are included as negative samples in the testing set. Once the testing set is defined, the resulting partially observed tensors are used as input to the learning algorithm.

We evaluate the performance using two main metrics - the Area Under the ROC Curve (AUC) and Area Under the Precision-Recall Curve (AUPR) - but also include other metrics such as F1-score, accuracy, recall, specificity, and precision. For all baseline methods and SI-ADMM, we perform five independent runs with different random seeds and report the average performance along with the standard deviation.

5.2.2 New-Drug Prediction Task

To better reflect real-world usage, we also evaluate SI-ADMM and the baselines in a new–drug prediction setting. In this scenario, the model must infer potential drug-drug combinations or DDIs for a previously unseen drug, where all of its interaction entries in the tested tensor are masked. Thus, the model can rely only on side information matrices and cross-tensor signals (e.g., drug-drug combination information when predicting DDIs, and vice versa).

Formally, given a target drug and a tensor (either the DDI tensor or the drug-drug combination tensor), all entries whose row or column correspond to that drug are set to zero and treated as test entries. Under this setting, the CP baseline–which incorporates no side information–cannot produce meaningful predictions. Its ALS updates are described in [22], and when all entries for a drug are masked, the updates return zero vectors. Therefore, we compare SI-ADMM only with TDRC, TFAI, and CTF for this task.

We select the top 10 most interacted drugs from each tensor (based on the number of known interactions) for evaluation, yielding 20 drugs in total. Each drug is tested individually under the new–drug protocol for SI-ADMM and the baselines.

Performance is assessed using hit rate, which is most appropriate for this ranking-based retrieval problem. Since all true interactions involving the target drug are masked, metrics such as RMSE, MAE, accuracy, and AUROC become uninformative due to extreme class imbalance and the dominance of zeros. Hit rate instead measures whether the model ranks true interacting partners within the top-K predictions, aligning with practical drug discovery workflows.

5.3 Baseline Approaches

To the best of our knowledge, no existing approach adopts the same formulation as the one proposed in SI-ADMM. Consequently, a direct comparison with methods performing the exact same analysis is not feasible. Instead, we compare our method against baseline tensor decomposition approaches that independently decompose individual 3-D tensors, either with or without incorporating additional side information. These baseline methods are described in the following sections.

- **CP**: CANDECOMP/PARAFAC (CP) [22] is a standard tensor decomposition method (see Section 4.1). We use the implementation provided by the TensorLy Python library [23], which optimizes the latent factor matrices via Alternating Least Squares (ALS). CP does not incorporate side information and serves as a fundamental baseline.
- **Tensor Factorization using Auxiliary Information (TF-AI)**: TF-AI [33] incorporates symmetric side information matrices for each mode of the target tensor to guide the decomposition process. It introduces two regularization strategies, tailored for dense and sparse tensors respectively. We include TF-AI as an early method that integrates side information in the form of similarity matrices.
- **TDRC**: TDRC [16] is a tensor decomposition method originally proposed for miRNA-disease-type prediction. It incorporates side information as relational constraints in the optimization objective. We include TDRC as a baseline that leverages side information differently from our SI-ADMM framework, focusing on relational regularization.
- **CTF-DDI**: CTF-DDI [14] is a recent tensor decomposition method designed for DDI prediction, incorporating side information directly into its optimization framework. CTF-DDI integrates two types of side information—structural and biological similarities—by averaging their values. We consider CTF-DDI as a strong and competitive baseline.

5.4 Hyper-parameters Settings

5.4.1 Hyper-parameters Tuning for SI-ADMM

Our model consists of several hyper-parameters, including the rank, initial values of penalties (ρ and η), and weights for the main tensors (a and b) and side information (α_i). Rank is set to 60 by running and evaluating the algorithm using the following rank numbers: $\{10, 20, 30, 40, 50, 60\}$, the penalty is set to 10^-7 with a 15% increase every iteration and is capped at 10^5. The penalties are used solely to accelerate the convergence of the algorithm, so adjustment of this doesn't affect the outcome of the program. For weighting the loss contributions of the tensors and side information matrices, we apply a linear scaling based on the inverse of their total magnitude. Specifically, the weights for each tensor or matrix are:

$$w_T = \frac{\frac{1}{sum(T)}}{\sum_T \frac{1}{sum(T)}} \quad where \quad T \in \{X, Y, S_A^{(i)}\} \tag{8}$$

where $sum(T)$ denotes the sum of all elements, since tensors and side information matrices are all non-negative in T. The resulting w_T values serve as the weights (i.e., a, b, and α_i) in the SI-ADMM optimization framework, balancing the contributions of each tensor and side information matrix. For the dataset used in the SI-ADMM, the weights a, b, and α_i are [0.5764 0.04189 0.08469 0.09330 0.02414 0.1795]. For the inner update iterations that updates $U^{(i)}$ and $C^{(i)}$, We set the inner iteration number to 10 after testing the program with five different random seeds across the following inner update iteration counts: 10, 20, 30, 40, 50, 60, 70, 80.

5.4.2 Hyper-parameter Tuning for Baselines

For CTF-DDI [14], we followed the hyperparameter settings reported in their original paper and utilized Tanimoto similarity and target sequence alignments similarity as structural side information, IC50 similarity and side effect similarity as biological side information. For other baseline methods, the primary parameter was the rank number, and we evaluated different rank numbers in $\{10, 20, 30, 40, 50, 60\}$ and reported their best performance, averaging the metrics over both tensor-recovery tasks.

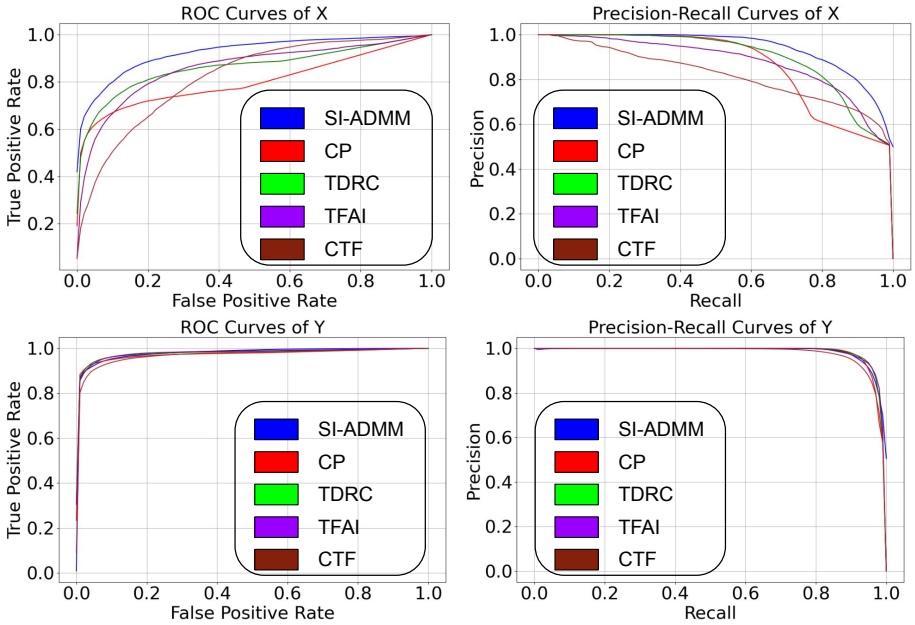


Figure 2: ROC and Precision-Recall curves for different models on tensors $\mathcal{X}$ (top) and $\mathcal{Y}$ (bottom).

Table 1: Results of SI-ADMM and baselines on tensor $\mathcal{X}$ prediction (drug-drug combination)

Methods \ Metrics (%)	**AUPR**	**AUC**	**F1**	**Acc**	**Rec**	**Spec**	**Prec**
CP (50)	88.12±0.6	82.59±0.82	77.97±0.97	79.99±0.73	70.87±1.85	89.12±1.39	86.72±1.3
TF-AI (60)	91.49±0.52	89.24±0.59	**82.8±0.9**	82.66±0.98	**83.48±1.57***	81.83±2.23	82.17±1.65
TDRC (60)	90.86±0.37	87.03±0.5	82.31±0.46	82.62±0.62	80.82±1.28	84.43±2.19	83.9±1.71
CTF-DDI (60)	90.73±0.29	86.94±0.51	82.33±0.46	82.94±0.44	79.46±1.2	86.42±1.39	85.43±1.16
SI-ADMM (60)	**93.97±0.61***	**92.58±0.94***	82.40±3.50	**83.95±1.83**	76.76±2.06	**91.13±6.41**	**90.72±5.6***

Note: * indicates statistically significant improvement compared with the second-best result using the paired t-test ($p < 0.05$). "(60)" indicates the rank number used for the model. Best results are bolded.

6 RESULTS

6.1 Results on Random Prediction Task

Tables 1 and 2 present the prediction results for the drug-drug combination prediction task and the DDI prediction task, respectively, for baseline models and SI-ADMM. The results demonstrate that SI-ADMM consistently achieves the highest performance among all the tested methods in terms of AUC and AUPR (Fig 2), as well as specificity and precision for both the drug-drug combination and DDI prediction tasks. Notably, for the drug-drug-disease association prediction task, SI-ADMM achieves significant improvements in AUPR, AUC, and Precision (all p-values are smaller than 0.05 comparing with the second best method based on the paired t-test). It also achieves the highest score in accuracy and the second highest in F1 score that is nearly indistinguishable from the highest score (from CTF-DDI). For DDI prediction, SI-ADMM is the best approach and achieves the best score in four measures including AUPR and AUC (Fig 2). TDRF is the close second and achieves the best score in the other three measures. These findings highlight the competitive discriminative capability of SI-ADMM in both drug-drug combination and DDI prediction tasks.

It is also worth noting that all baseline methods perform independent decompositions for each tensor, resulting in separate sets of three latent factor matrices for each task. In contrast, SI-ADMM jointly decomposes both tensors using a shared set of four latent factor matrices, namely U, D, V, and W. The superior performance of SI-ADMM not only demonstrates that the shared drug latent factors can effectively capture meaningful representations for both tasks but also highlights the model's strong learning capability despite operating with fewer degrees of freedom in terms of the number of latent factor matrices.

Table 2: Results of SI-ADMM and baselines on tensor $\mathcal{Y}$ prediction (DDI)

Methods \ Metrics (%)	**AUPR**	**AUC**	**F1**	**Acc**	**Rec**	**Spec**	**Prec**
CP (50)	98.30±0.08	97.50±0.10	91.99±0.16	92.21±0.17	89.41±0.61	95.02±0.70	94.73±0.67
TF-AI (60)	97.51±0.03	97.66±0.04	94.16±0.14	94.19±0.13	**93.63±0.56**	94.75±0.51	94.70±0.47
TDRC (60)	98.48±0.06	97.85±0.09	**94.34±0.18**	**94.41±0.18**	93.21±0.29	95.61±0.40	95.50±0.38
CTF-DDI (60)	98.48±0.04	97.85±0.07	94.29±0.10	94.40±0.11	92.53±0.36	96.27±0.52	96.13±0.51
SI-ADMM (60)	**98.53±0.06**	**98.21±0.09***	93.12±0.55	93.42±0.46	89.10±1.56	**97.74±0.84***	**97.55±0.86***

Note: * indicates statistically significant improvement compared with the second-best result using the paired t-test ($p < 0.05$). "(60)" indicates the rank number used for the model. Best results are bolded.

Table 3: Top-10 predictions from tensor $\mathcal{X}$

Rank	**Drug Pair**	**Disease**
1-2	(Docetaxel/DB01248, Cisplatin/DB00515)	Malignant neoplasms
3-4	(Cisplatin/DB00515, Gemcitabine/DB00441)	Malignant neoplasms
5	(Docetaxel/DB01248, Doxorubicin/DB00997)	Malignant neoplasms
6-7	(Irinotecan/DB00762, Cisplatin/DB00515)	Malignant neoplasms
8	(Ibuprofen/DB01050, Acetaminophen/DB00316)	Arthropathies
9-10	(Lenalidomide/DB00480, Bortezomib/DB00188)	Neoplasms

Note: Drug pairs are unordered; inverse pairs may appear at lower ranks in the full prediction list.

6.2 Further Analysis of SI-ADMM

To further analyze the structural properties of the proposed SI-ADMM model, we examine the correspondence between the original tensors $\mathcal{X}$ and $\mathcal{Y}$ and their predicted counterparts at a global interaction level. For each tensor, interactions are aggregated by summing over the third mode (i.e., the drug-drug-disease or DDI dimension), yielding a symmetric drug-drug matrix in which each entry represents the overall interaction strength between a pair of drugs. We then compute row (equivalently, column) sums of the aggregated matrix and rank drugs according to their interaction intensities. To ensure a consistent basis for comparison, the predicted tensors are aggregated and visualized using the same drug ranking derived from the original tensors. The resulting ranked heatmaps, shown in Figure 3, facilitate a direct comparison of large-scale interaction structures between the original and reconstructed tensors. As observed in the figure, the predicted heatmaps closely match their original counterparts in terms of overall intensity distributions and prominent structural patterns, including dense interaction blocks and banded regions, indicating that SI-ADMM preserves the global organization of drug-drug interactions rather than overfitting individual tensor entries.

To complement this structural visualization, we quantitatively assess the agreement between the original tensors and their predictions using global similarity metrics. Direct entry-wise comparison is unreliable due to the extreme sparsity of the original tensors; therefore, all quantitative evaluations are conducted on the aggregated two-dimensional interaction matrices described above. This aggregation retains meaningful global interaction structure while reducing sparsity-induced noise, enabling a more stable and informative assessment. Specifically, we compute Pearson correlation [3], Spearman rank correlation [42], and relative Frobenius error between the aggregated original and predicted matrices. These measures respectively capture linear correspondence, rank-level consistency, and overall reconstruction fidelity, and are commonly used to evaluate structural similarity in high-dimensional matrix and tensor analysis.

The quantitative results demonstrate a strong structural correspondence for tensor $\mathcal{Y}$, with high Pearson (0.79) and Spearman (0.59) correlations and a low relative Frobenius error (0.65), suggesting that SI-ADMM effectively recovers dominant drug-drug interaction patterns. For tensor $\mathcal{X}$, a moderate Pearson correlation (approximately 0.46) indicates that the model captures coarse interaction intensity structure, while weaker rank-level agreement reflects heterogeneity across disease-specific tensor slices. In contrast, entry-wise correlations computed directly on the full three-dimensional tensors are close to zero, which is expected given the high dimensionality and sparsity of biomedical interaction data and does not contradict the model's ability to recover meaningful global structure. Taken together, these results provide evidence that SI-ADMM captures latent interaction patterns in a structurally consistent manner at an analytically relevant scale.

Table 4: Top-10 predictions from tensor $\mathcal{Y}$

Rank	Drug Pair	DDI Type
1-2	(Acenocoumarol/DB01418, Warfarin/DB00682)	Increasing anticoagulant activities
3-4	(Acenocoumarol/DB01418, Phenprocoumon/DB00946)	Increasing anticoagulant activities
5-8	(Avanafil/DB06237, DB00484)	Increasing antihypertensive activities
6-7	(Warfarin/DB00682, Bivalirudin/DB00006)	Increasing anticoagulant activities
9-10	(Rivaroxaban/DB06228, Bivalirudin/DB00006)	Increasing anticoagulant activities

Note: Drug pairs are unordered; inverse pairs may appear at lower ranks in the full prediction list.

Table 5: Results of SI-ADMM and baselines on New-Drug prediction (top-10 most interacted drugs).

Method	HR@100 $\mathcal{X}$	HR@100 $\mathcal{Y}$
TF-AI	10.00±3.79	24.40±7.20
TDRC	17.60±8.80	29.00±10.04
CTF-DDI	22.00±11.73	33.60±10.50
SI-ADMM	**29.80±17.18**	**55.20±16.50***

Note: * indicates statistically significant improvement compared with the second-best result using the paired t-test ($p < 0.05$). Best results are bolded.

6.3 Results on New-Drug Prediction Task

Table 5 summarizes the results of New-Drug Prediction for SI-ADMM and the applicable baselines measured using hit rate. SI-ADMM consistently achieves strong hit-rate performance across the most interacted drugs for both tensors, with particularly notable gains on the $\mathcal{Y}$ tensor. These results demonstrate that the proposed joint learning framework can effectively exploit side information and cross-tensor coupling to infer meaningful interaction patterns even in the absence of historical interaction data. Compared to competing baselines, the consistently higher hit rates indicate that SI-ADMM generalizes more reliably under cold-start conditions, which closely reflect real-world deployment scenarios where newly developed or insufficiently studied drugs are common. Consequently, the strong performance advantage observed in this new-drug setting suggests that SI-ADMM has a higher likelihood of delivering improved predictive accuracy in practical applications, where robust interaction prediction based on limited information is essential.

6.4 Case Analysis

After running the algorithm, we examined the entries with the top 10 highest predicted values for both tensors. The top 10 results for tensor $\mathcal{X}$ are shown in Table 3, where the drug pairs are represented using their DrugBank IDs. We observe that although our model does not directly impose symmetric constraints on the drug pairs in the recovered tensors, the prediction results show many drug pairs (8 out of 10) of the top-ranked predictions are symmetric entries. For example, the prediction values of entry (Ibuprofen/DB01050, Acetaminophen/DB00316) and entry (Acetaminophen/DB00316, Ibuprofen/DB01050) ranked top 2 and they were for the same disease. This is also true for tensor $\mathcal{Y}$, in which case the top 10 predictions comprise five unique drug pairs.

Furthermore, the suggested drug combinations are clinically plausible based on their known functions and descriptions. For example, Ibuprofen/DB01050 and Acetaminophen/DB00316 are two commonly used over the counter medicines for pain and inflammation management. Their combination aligns well with their predicted disease association (arthropathies). All other drugs (Docetaxel/DB01248, Cisplatin/DB00515, Gemcitabine/DB00441, Doxorubicin/DB00997, Irinotecan/DB00762, Lenalidomide/DB00480, and Bortezomib/DB00188) are commonly used in cancer treatment, supporting their predicted links to various neoplasms. In terms of the DDI prediction, the top 10 predictions are composed by five symmetric pairs and is shown in Table 4. 4 out of 5 pairs involves in the same DDI type: *one drug may increase the anticoagulant activities of the other drug*. The drugs in these four pairs (Acenocoumarol/DB01418, Warfarin/DB00682, Phenprocoumon/DB00946, Bivalirudin/DB00006, Rivaroxaban/DB06228) are all blood thinner, which supports the predicted DDI type.

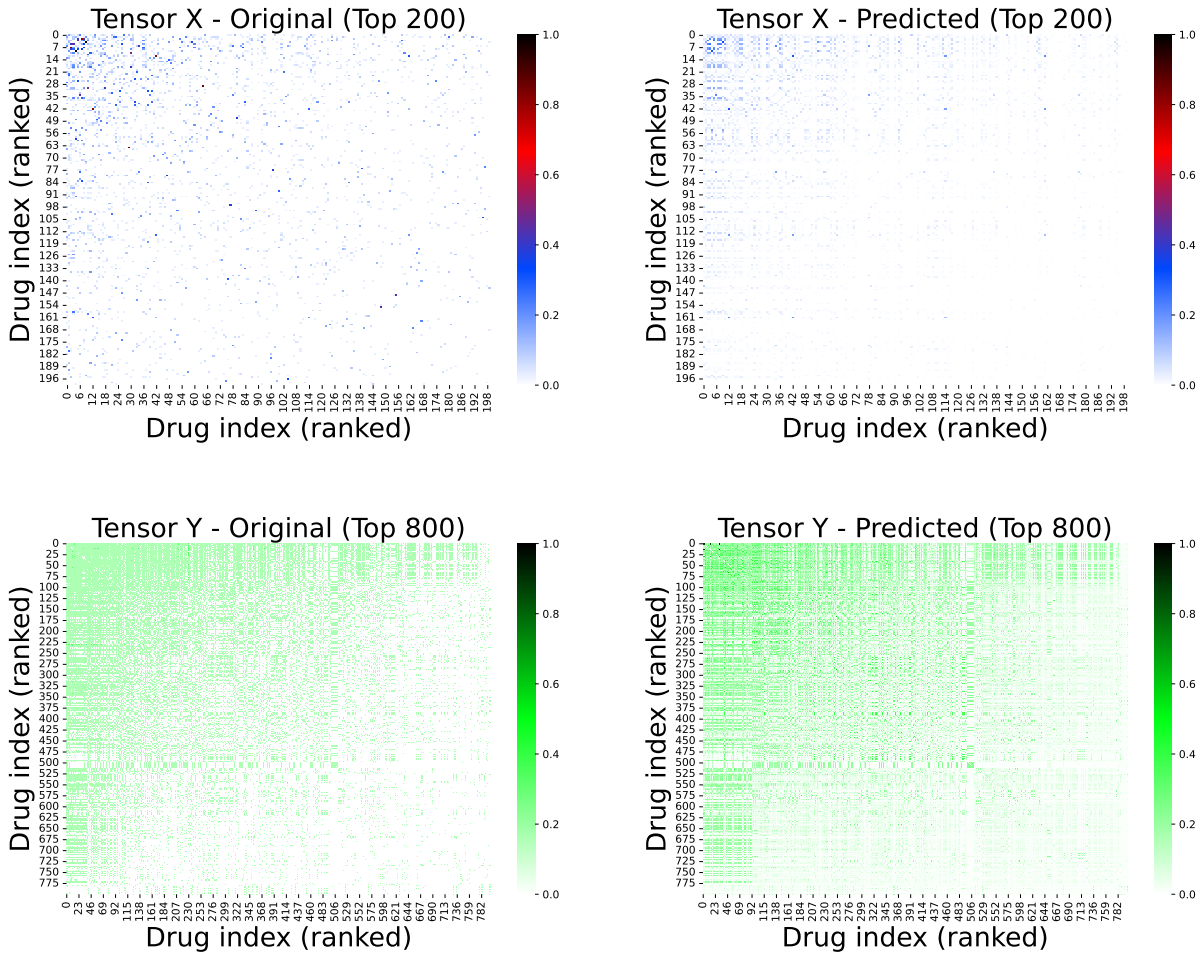


Figure 3: Ranked heatmaps of aggregated tensors. Tensors are summed over the third mode and reordered by row-sum ranking from the original data, which is consistently applied to predictions. The predicted heatmaps recover key block and band structures, demonstrating preservation of global interaction patterns. Owing to sparsity, the top 200 drugs for $\mathcal{X}$ and top 800 drugs for $\mathcal{Y}$ are shown.

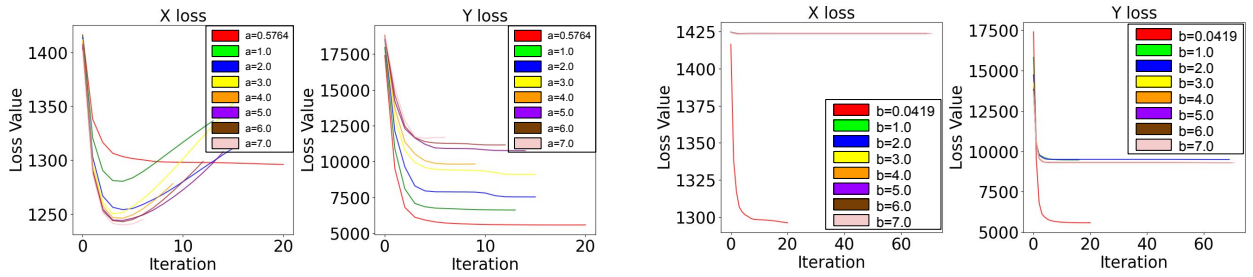
all the 5 unique drug pairs in the top 10 predictions involve a common drug, Hydroxyzine/DB00557 and all drug pairs are associated with the same DDI type: *one drug may increase the central nervous system (CNS) depressant activities of the other drug*. Hydroxyzine is a prescription medication used to treat anxiety and tension, while its pairing drugs—Buprenorphine/DB00921, Tapentadol/ DB06204, Zolpidem/DB00425, Droperidol/DB00450, and Thalidomide/DB01041 — are known to act as painkillers, neural suppressants, or anti-inflammatory drug. This supports the plausibility of the predicted interactions, as combining these drugs could indeed potentiate CNS depressant effects.

6.5 Ablation Studies

To show the robustness of the model, we carried out the following ablation studies: the impact of tensor weights (a and b) on the tensor reconstruction losses and the impact of the rank parameter on performance of all the approaches.

6.5.1 Impact of Tensor Weights on Reconstruction Losses

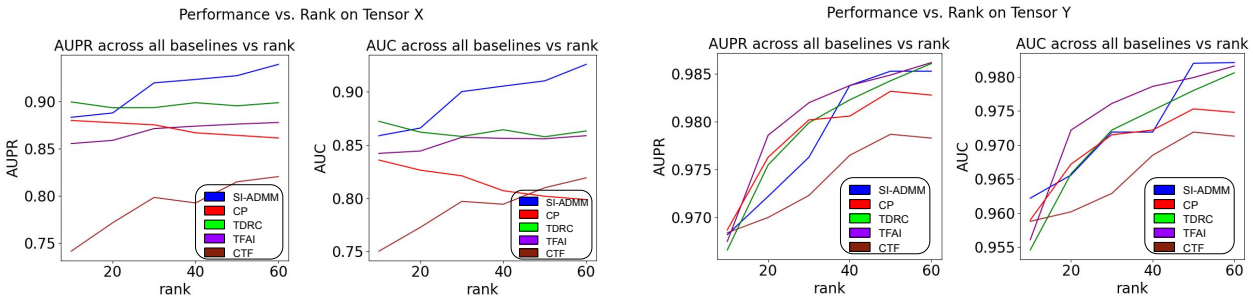
To evaluate the effectiveness of the weight selection in Eq. (13), we varied the tensor weights a and b for $\mathcal{X}$ and $\mathcal{Y}$ individually while fixing all other weights, and compared the resulting reconstruction losses on the test indices



(a) Loss curves for increasing a

(b) Loss curves for increasing b

Figure 4: Model loss curves with increasing a (left) and b (right) weights.



(a) Tensor $\mathcal{X}$

(b) Tensor $\mathcal{Y}$

Figure 5: Model performance across ranks.

between the original and predicted tensors. As shown in Figure 4, increasing a or b reduces the reconstruction loss for the corresponding tensor ($\mathcal{X}$ or $\mathcal{Y}$), but does so at the expense of a significant increase in the loss of the other tensor whose weight remains unchanged. The weights specified in Eq. (13) therefore achieve the best trade-off, minimizing the overall reconstruction loss across both tensors. We also observe that the reconstruction loss of an individual tensor (e.g., $\mathcal{X}$) does not necessarily decrease monotonically during optimization, which can be attributed to the fact that the overall objective function consists of multiple coupled terms, and the loss associated with $\mathcal{X}$ represents only one component of the joint optimization objective.

6.5.2 Impact of Rank on Performance

To evaluate the effect of rank on model performance, we conducted experiments using all methods with rank settings of $\{10, 20, 30, 40, 50, 60\}$. The corresponding results, measured by AUPR and AUC of both tensors $\mathcal{X}$ and $\mathcal{Y}$, are presented in Figure 5. As shown in the two sub-figures on the bottom row, all models exhibit improved performance on the DDI prediction task (tensor $\mathcal{Y}$) as the rank increases, which is expected as higher rank in the latent space can capture more information. However, for the drug-drug combination prediction task (tensor $\mathcal{X}$), only SI-ADMM and TF-AI show improved performance as the rank increases. The performance of TDRC and CTF-DDI fluctuate when the value of rank changes. The performance of CP even declines with higher ranks (the two sub-figures on the top row in Figure 5). This indicates that these models face challenges in handling tensor data with varying structures and sparsity levels, two primary features that differentiate tensors $\mathcal{X}$ and $\mathcal{Y}$. Notably, SI-ADMM demonstrates consistent performance improvement across both tensors as the rank increases.

6.6 Runtime Analysis for SI-ADMM and Baselines

To evaluate the runtime of the SI-ADMM model, we conducted five independent runs for each method. The results are shown in figure 6. Since all models employ an early termination mechanism, we report the per-iteration runtime as a function of rank. For SI-ADMM in particular, the per-iteration runtime exhibits no significant increment as the rank grows. Accounting for early termination, the average total runtime of SI-ADMM still demonstrates no increase with respect to increasing rank. This relationship is consistent with the complexity analysis since the number of

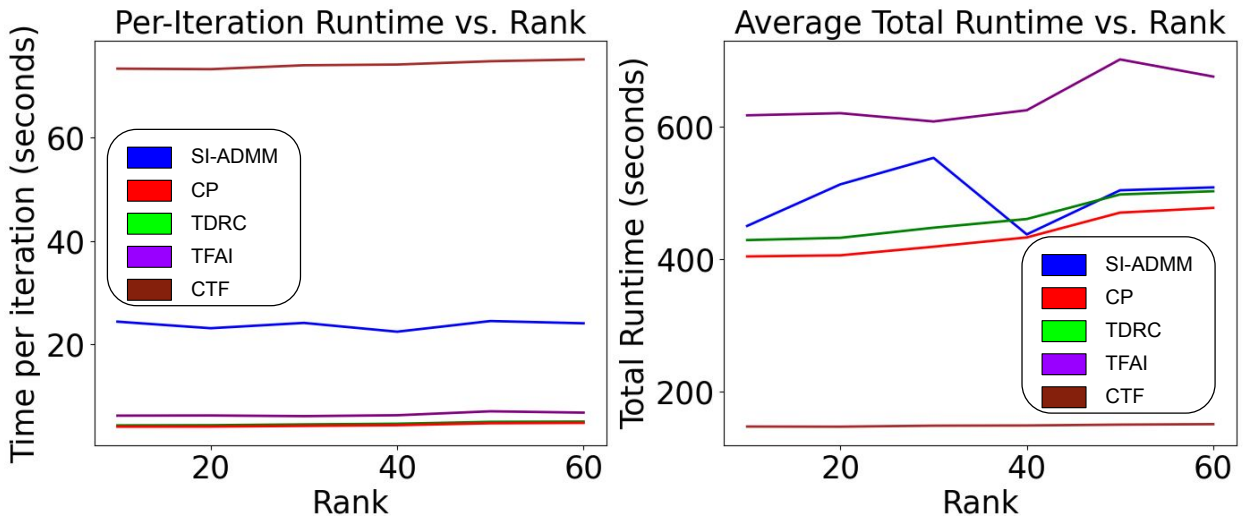


Figure 6: Runtime analysis of SI-ADMM and baselines.

Left: per-iteration runtime (seconds) vs. rank. Right: average total runtime (five runs) vs. rank.

ranks (≤ 60) is much smaller than the number of drugs (1069). So the term that has the highest quantity becomes $N^2R(M_{\text{disease}} + M_{\text{DDI}})$, which makes the computation burden non-significant.

7 CONCLUSIONS

Drug-drug combinational therapies and drug-drug adverse effects are two closely related aspects of the same underlying phenomenon. In this paper, we proposed a joint tensor decomposition model, SI-ADMM, that simultaneously predicts disease-specific drug-drug combinations as well as DDIs, while mitigating sparsity through structured integration of auxiliary drug side information. We further developed an efficient optimization algorithm to solve the resulting joint tensor decomposition and reconstruction problem, and extensive experiments on real-world datasets demonstrated the effectiveness and robustness of the proposed SI-ADMM method. Looking ahead, our future work will focus on incorporating richer sources of side information and enhancing the model's capacity for nonlinear transformations. In particular, we plan to integrate deep learning modules, such as graph neural networks (GNNs), to extract latent representations from molecular structures and heterogeneous biomedical networks [12, 27, 53]. In addition, to further improve expressive power, we intend to introduce neural network components after optimizing the latent factor tensors, allowing nonlinear transformations to be seamlessly incorporated while retaining the core optimization framework established in this work.

References

[1] Evrim Acar, Tamara G. Kolda, and Daniel M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations, 2011.

[2] Miguel Ramos de Araujo, Pedro Manuel Pinto Ribeiro, and Christos Faloutsos. Tensorcast: Forecasting with context using coupled tensors (best paper award). In *2017 IEEE International Conference on Data Mining (ICDM)*, page 71–80. IEEE, November 2017.

[3] Stephen Blyth. Karl pearson and the correlation curve. *International Statistical Review / Revue Internationale de Statistique*, 62(3):393, December 1994.

[4] Stephen Boyd. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2010.

[5] Bokai Cao, Lifang He, Xiaokai Wei, Mengqi Xing, Philip S. Yu, Heide Klumpp, and Alex D. Leow. *t-BNE: Tensor-based Brain Network Embedding*, page 189–197. Society for Industrial and Applied Mathematics, June 2017.

[6] Huiyuan Chen, Sudha K Iyengar, and Jing Li. Large-scale analysis of drug combinations by integrating multiple heterogeneous information networks. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, New York, NY, USA, September 2019. ACM.

[7] Huiyuan Chen and Jing Li. A flexible and robust multi-source learning algorithm for drug repositioning. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '17, page 510–515. ACM, August 2017.

[8] Huiyuan Chen and Jing Li. Learning multiple similarities of users and items in recommender systems. In *2017 IEEE International Conference on Data Mining (ICDM)*, volume 1, page 811–816. IEEE, November 2017.

[9] Huiyuan Chen and Jing Li. Drugcom: Synergistic discovery of drug combinations using tensor decomposition. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 899–904, 2018.

[10] S. Dechanont, S. Maphanta, B. Butthum, and C. Kongkaew. Hospital admissions/visits associated with drug-drug interactions: a systematic review and meta-analysis. *Pharmacoepidemiology and Drug Safety*, 23(5):489–497, May 2014. Epub 2014 Mar 10.

[11] Yiheng Ding, Hande Küçük-McGinty, et al. Dtf: Deep tensor factorization for predicting anticancer drug synergy. *Briefings in Bioinformatics*, 22(2):232–244, 2020.

[12] Zhengyu Fang, Xiaoge Zhang, Anyin Zhao, Xiao Li, Huiyuan Chen, and Jing Li. Recent developments in gnns for drug discovery. *arXiv preprint arXiv:2506.01302*, 2025.

[13] Michael K. Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong. Bindingdb in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, 44(D1):D1045–D1053, October 2015.

[14] Guosheng Han, Lingzhi Peng, Aocheng Ding, Yan Zhang, and Xuan Lin. Ctf-ddi: Constrained tensor factorization for drug–drug interactions prediction. *Future Generation Computer Systems*, 161:26–34, December 2024.

[15] Peilin Hu, Yang Yang, et al. Coupled matrix–matrix and coupled tensor–matrix completion methods for predicting drug–target interactions. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1057–1064, 2016.

[16] Feng Huang, Xiang Yue, Zhankun Xiong, Zhouxin Yu, Shichao Liu, and Wen Zhang. Tensor decomposition with relational constraints for predicting multiple types of microrna-disease associations. *Briefings in Bioinformatics*, 22(3), July 2020.

[17] Huaqiao Jiang, Yanhua Lin, Weifang Ren, Zhonghong Fang, Yujuan Liu, Xiaofang Tan, Xiaoqun Lv, and Ning Zhang. Adverse drug reactions and correlations with drug–drug interactions: A retrospective study of reports from 2011 to 2020. *Frontiers in Pharmacology*, 13, 2022.

[18] Hiroyuki Kasai and Bamdev Mishra. Low-rank tensor completion: a riemannian manifold preconditioning approach. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1012–1021, New York, New York, USA, 20–22 Jun 2016. PMLR.

[19] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. Pubchem 2023 update. *Nucleic Acids Research*, 51(D1):D1373–D1380, October 2022.

[20] Yejin Kim, Jimeng Sun, Hwanjo Yu, and Xiaoqian Jiang. Federated tensor factorization for computational phenotyping. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17. ACM, August 2017.

[21] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun (Lucy) Chin, Seth A Strawbridge, Marysol Garcia-Patino, Ray Kruger, Aadhavya Sivakumaran, Selena Sanford, Rahil Doshi, Nitya Khetarpal, Omolola Fatokun, Daphnee Doucet, Ashley Zubkowski, Dorsa Yahya Rayat, Hayley Jackson, Karxena Harford, Afia Anjum, Mahi Zakir, Fei Wang, Siyang Tian, Brian Lee, Jaanus Liigand, Harrison Peters, Ruo Qi (Rachel) Wang, Tue Nguyen, Denise So, Matthew Sharp, Rodolfo da Silva, Cyrella Gabriel, Joshua Scantlebury, Marissa Jasinski, David Ackerman, Timothy Jewison, Tanvir Sajed, Vasuk Gautam, and David S Wishart. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1):D1265–D1275, November 2023.

[22] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[23] Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. Tensorly: Tensor learning in python. *Journal of Machine Learning Research*, 20(26):1–6, 2019.

[24] Danai Koutra, Tai-You Ke, U. Kang, Duen Horng Chau, Hsing-Kuo Kenneth Pao, and Christos Faloutsos. *Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms*, page 245–260. Springer Berlin Heidelberg, 2011.

[25] Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic Acids Research*, 44(D1):D1075–D1079, October 2015.

[26] Greg Landrum and contributors. Rdkit: Open-source cheminformatics software, 2025. Accessed: 2025-05-19.

[27] Tianyi Li, Zhengyu Fang, Xiaoge Zhang, Kaiyu Tang, Huiyuan Chen, Zhimeng Jiang, Tianxiang Zhao, Rong Xu, Feixiong Cheng, Xiao Li, and Jing Li. Druglm: A unified framework to enhance drug-target interaction predictions by incorporating textual embeddings via language models. July 2025.

[28] Athanasios P. Liavas and Nicholas D. Sidiropoulos. Parallel algorithms for constrained tensor factorization via alternating direction method of multipliers. *IEEE Transactions on Signal Processing*, 63(20):5450–5463, October 2015.

[29] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

[30] Hui Liu, Wenhao Zhang, Bo Zou, Jinxian Wang, Yuanyuan Deng, and Lei Deng. Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Research*, October 2019.

[31] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. *Multi-View Clustering via Joint Nonnegative Matrix Factorization*, pages 252–260.

[32] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, May 2013.

[33] Atsuhiro Narita, Kohei Hayashi, Ryota Tomioka, and Hisashi Kashima. Tensor factorization using auxiliary information. *Data Mining and Knowledge Discovery*, 25(2):298–324, July 2012.

[34] National Cancer Institute. Nci-60 human tumor cell lines screen, 2025. Accessed: 2025-05-19.

[35] Evangelos E. Papalexakis, Christos Faloutsos, and Nicholas D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology*, 8(2):1–44, October 2016.

[36] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook, November 2012. Version 20121115.

[37] K Preuer, R Lewis, S Hochreiter, et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature Communications*, 11(1):5057, 2020.

[38] Mihee Ryu, Jiyoung Yang, et al. Drugcom: Synergistic discovery of drug combinations using tensor decomposition. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2252–2260, 2019.

[39] Guy Shtar, Louise Azulay, Omer Nizri, Lior Rokach, and Bracha Shapira. Cdcdb: A large and continuously updated drug combination database. *Scientific Data*, 9(1), June 2022.

[40] T.F. Smith and M.S. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197, March 1981.

[41] Qingquan Song, Xiao Huang, Hancheng Ge, James Caverlee, and Xia Hu. Multi-aspect streaming tensor completion. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17. ACM, August 2017.

[42] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72, January 1904.

[43] Cheng Sun et al. Toward super-resolution image construction based on joint tensor decomposition. *Remote Sensing*, 12(16):2535, 2020.

[44] T. T. Tanimoto. An elementary mathematical theory of classification and prediction. Technical report, IBM Corporation, 1958. IBM Internal Report.

[45] U.S. Food and Drug Administration. Approved drug products with therapeutic equivalence evaluations (orange book), 2025. Accessed: 2025-05-19.

[46] Santiago Vilar, Eugenio Uriarte, Lourdes Santana, Tal Lorberbaum, George Hripcsak, Carol Friedman, and Nicholas P Tatonetti. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols*, pages 2147–2163, 2014.

[47] Yichen Wang, Robert Chen, Joydeep Ghosh, Joshua C. Denny, Abel Kho, You Chen, Bradley A. Malin, and Jimeng Sun. Rubik: Knowledge guided tensor factorization and completion for health data analytics. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 1265–1274. ACM, August 2015.

[48] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.

[49] Max Welling and Markus Weber. Positive tensor factorization. *Pattern Recognition Letters*, 22(12):1255–1261, October 2001.

[50] World Health Organization. International classification of diseases, 11th revision (icd-11), 2019. Accessed: 2025-05-19.

[51] Deborah A. Zarin, Tony Tse, Rebecca J. Williams, Robert M. Califf, and Nicholas C. Ide. The clinicaltrials.gov results database — update and key issues. *New England Journal of Medicine*, 364(9):852–860, March 2011.

[52] Xiao Zhang, Yu Cui, et al. Ctf-ddi: Constrained tensor factorization for drug-drug interactions prediction. *Future Generation Computer Systems*, 156:276–288, 2024.

[53] Anyin Zhao, Zuquan Chen, Zhengyu Fang, Xiaoge Zhang, and Jing Li. Dual-modality representation learning for molecular property prediction. In *International Symposium on Bioinformatics Research and Applications*, pages 34–47. Springer, 2025.

[54] Marinka Zitnik and Blaz Zupan. Modeling drug combination effects via latent tensor reconstruction. *Bioinformatics*, 37(Suppl_1):i93–i101, 2021.

Supplementary Appendix

This appendix provides the detailed derivations of the latent matrices $D, V, W, C^{(i)}, U^{(i)}$

Update of D

To calculate the partial derivative respect to D from equation (2), we can first separate out terms containing D.The terms in the Lagrangian $\mathscr{L}$ that involve D are:

$$\begin{aligned}\mathscr{L}_D = a\left\|X - [\![U, D, V]\!]\right\|_F^2 + b\left\|Y - [\![U, D, W]\!]\right\|_F^2 \\ + \langle F, D - U\rangle + \frac{\rho}{2}\left\|D - U\right\|_F^2\end{aligned}$$

Using the approximation rule on the p.463 on the tensor review, the terms can be rewritten as:

$$\begin{aligned}\mathscr{L}_D = a\left\|D(V \odot U) - X(2)\right\|_F^2 + b\left\|D(W \odot U) - Y(2)\right\|_F^2 \\ + \frac{\rho}{2}\left\|D + \frac{F}{\rho} - U\right\|_F^2\end{aligned}$$

$X(2), Y(2)$ are the second order matricizations of tensor X and Y. here set $\Phi = V \odot U$ and $\Psi = W \odot U$, using equation 116 and 102 in Matrix Cookbook, the partial derivative of partial lagrangian function respects to D is:

$$\begin{aligned}\frac{\partial \mathscr{L}}{\partial D} = 2aD\Phi^T\Phi - 2aX(2)\Phi + 2bD\Psi^T\Psi - 2bY(2)\Psi \\ + \rho D + F - \rho U\end{aligned}$$

solve for D by setting the partial derivative to zero:

$$\begin{aligned}0 = 2aD\Phi^T\Phi - 2aX(2)\Phi + 2bD\Psi^T\Psi - 2bY(2)\Psi \\ + \rho D + F - \rho U\end{aligned}$$

$$\begin{aligned}D = (2aX(2)\Phi + 2bY(2)\Psi - F + \rho U) \\ (2a\Phi^T\Phi + 2b\Psi^T\Psi + \rho I)^{-1}\end{aligned} \tag{9}$$

Update of V

To calculate the partial derivative with respect to V from equation (2), we can first separate out terms containing V.The terms in the Lagrangian $\mathscr{L}$ that involve V are:

$$\mathscr{L}_V = a\left\|X - [\![U, D, V]\!]\right\|_F^2$$

Using the approximation rule on the p.463 on the tensor review, the terms can be rewritten as:

$$\mathscr{L}_V = a\left\|V(D \odot U) - X(3)\right\|_F^2$$

$X(3)$ is the third order matricization of tensor X. Here set $\Xi = D \odot U$, using equations 116 and 102 in Matrix Cookbook, the partial derivative of partial lagrangian function respects to V is:

$$\frac{\partial \mathscr{L}}{\partial V} = 2aV\Xi^T\Xi - 2aX(3)\Xi$$

solve for V by setting the partial derivative to zero:

$$0 = 2aV\Xi^T\Xi - 2aX(3)\Xi$$

$$\begin{aligned} V = &(X(3)\Xi) \\ &(\Xi^T\Xi)^{-1} \end{aligned} \tag{10}$$

Update of W

To calculate the partial derivative with respect to W from equation (2), we can first separate out terms containing W.The terms in the Lagrangian $\mathscr{L}$ that involve W are:

$$\mathscr{L}_W = b \left\| Y - [\![U, D, W]\!] \right\|_F^2$$

Using the approximation rule on the p.463 on the tensor review, the terms can be rewritten as:

$$\mathscr{L}_W = a \left\| W(D \odot U) - Y(3) \right\|_F^2$$

$Y(3)$ is the third order matricization of tensor Y. Here set $\Omega = D \odot U$, using equations 116 and 102 in Matrix Cookbook, the partial derivative of partial Lagrangianll function respects to W is:

$$\frac{\partial \mathscr{L}}{\partial W} = 2bW\Omega^T\Omega - 2bY(3)\Omega$$

solve for W by setting the partial derivative to zero:

$$0 = 2bW\Omega^T\Omega - 2bY(3)\Omega$$

$$\begin{aligned} W = &(Y(3)\Omega) \\ &(\Omega^T\Omega)^{-1} \end{aligned} \tag{11}$$

Update of $U^{(i)}$ and $C^{(i)}$

After the addition of the masking matrix on each side of the term in the second degree norm, the related terms for $U^{(i)}$ become:

$$\mathscr{L}_{U^{(i)}} = \alpha_i \left\| A^{(i)} \cdot (S_A^{(i)} - C^{(i)}U^{(i)^T}) \right\|_F^2 + \alpha_i \left\| U^{(i)}Q^{(i)} - U \right\|_F^2 + \frac{\eta}{2} \left\| U^{(i)} - C^{(i)} - \frac{Y^{(i)}}{\eta} \right\|_F^2$$

Here $\cdot$ is the element-wise multiplication, and $A^{(i)}$ is a symmetric masking with values 1 for non-missing values and 0 for the missing value. The derivatives of the second and third terms are calculated in previous section, and the derivative of $\alpha_i \left\| A^{(i)} \cdot (S_A^{(i)} - C^{(i)} {U^{(i)}}^T) \right\|_F^2$ can be calculated as the following:

define matrix $Z = C^{(i)} {U^{(i)}}^T - S_A^{(i)}$, and Matrix $W = S_A^{(i)} \cdot Z$, Then the term's differential can be derived as:

$$\begin{aligned} f &= W : W \\ df &= 2W : dW \\ &= 2W : (A^{(i)} \cdot dZ) \\ &= 2W : (dC^{(i)} {U^{(i)}}^T + C^{(i)} d{U^{(i)}}^T) \\ &= 2WU^{(i)} : dC^{(i)} + 2W^T C^{(i)} : dU^{(i)} \end{aligned}$$

Here, : is the Frobenius norm product, and the partial derivative for $C^{(i)}$ and $U^{(i)}$can be calculated as:

$$\begin{aligned} \frac{\partial f}{\partial U^{(i)}} &= 2\alpha_i (A^{(i)} \cdot (C^{(i)} {U^{(i)}}^T - S_A^{(i)}))^T C^{(i)} \\ &= 2\alpha_i (A^{(i)^T} \cdot U^{(i)} C^{(i)^T}) C^{(i)} - 2\alpha_i (A^{(i)^T} \cdot S_A^{(i)^T}) C^{(i)} \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial C^{(i)}} &= 2\alpha_i (A^{(i)} \cdot (C^{(i)} {U^{(i)}}^T - S_A^{(i)})) U^{(i)} \\ &= 2\alpha_i (A^{(i)} \cdot C^{(i)} {U^{(i)}}^T) U^{(i)} - 2\alpha_i (A^{(i)} \cdot S_A^{(i)}) U^{(i)} \end{aligned}$$

Putting the calculation of partial derivatives for all terms involves $C^{(i)}$ and $U^{(i)}$ together, we got the over all partial derivative of the lagrangian function for the two matrices as:

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial U^{(i)}} = \; & 2\alpha_i (A^{(i)^T} \cdot U^{(i)} C^{(i)^T}) C^{(i)} - 2\alpha_i (A^{(i)^T} \cdot S_A^{(i)^T}) C^{(i)} \\ & + 2\alpha_i U^{(i)} Q^{(i)} Q^{(i)^T} - 2\alpha_i U Q^{(i)^T} \\ & + \eta U^{(i)} - \eta C^{(i)} - Y^{(i)} \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial C^{(i)}} = \; & 2\alpha_i (A^{(i)} \cdot C^{(i)} {U^{(i)}}^T) U^{(i)} - 2\alpha_i (A^{(i)} \cdot S_A^{(i)}) U^{(i)} \\ & + \eta C^{(i)} + Y^{(i)} - \eta U^{(i)} \end{aligned}$$

Since it is challenging to derive the analytic solve for $U^{(i)}$ and $U^{(i)}$ that separates them from the Hadamard product, we uses gradient descent to update $U^{(i)}$:

$$U^{(i)^{t+1}} = U^{(i)^t} - \gamma^{U^{(i)}} \frac{\partial \mathcal{L}}{\partial U^{(i)}}$$

and use the similar gradient descent to update $C^{(i)}$:

$$C^{(i)^{t+1}} = C^{(i)^t} - \gamma^{C^{(i)}} \frac{\partial \mathcal{L}}{\partial C^{(i)}}$$

Here $\gamma^{U^{(i)}}$ and $\gamma^{C^{(i)}}$ are the learning rates.