scientific reports

Check for updates

OPEN

Meta graphical lasso: uncovering hidden interactions among latent mechanisms

Koji Maruhashi[1]✉, Hisashi Kashima[2], Satoru Miyano[3] & Heewon Park[3,4]✉

In complex systems, it's crucial to uncover latent mechanisms and their context-dependent relationships. This is especially true in medical research, where identifying unknown cancer mechanisms and their impact on phenomena like drug resistance is vital. Directly observing these mechanisms is challenging due to measurement complexities, leading to an approach that infers latent mechanisms from observed variable distributions. Despite machine learning advancements enabling sophisticated generative models, their black-box nature complicates the interpretation of complex latent mechanisms. A promising method for understanding these mechanisms involves estimating latent factors through linear projection, though there's no assurance that inferences made under specific conditions will remain valid across contexts. We propose a novel solution, suggesting data, even from systems appearing complex, can often be explained by sparse dependencies among a few common latent factors, regardless of the situation. This simplification allows for modeling that yields significant insights across diverse fields. We demonstrate this with datasets from finance, where we capture societal trends from stock price movements, and medicine, where we uncover new insights into cancer drug resistance through gene expression analysis.

Keywords Graphical model, Graphical lasso, Latent factor, Stiefel manifolds

In guiding scientific discovery, it is crucial to estimate the latent mechanisms behind the generative process of observed data. For instance, in the financial domain, understanding the overall movement of a vast array of stock prices, not just individual stocks, through societal trends can lead to strategies that guide better futures across various domains. In the medical field, comprehending the behavior of tens of thousands of genes through underlying biological latent mechanisms, such as pathways, can help understand the essence of complex cancer metastasis, invasion, and drug resistance, contributing to the establishment of unprecedented treatment methods.

One approach to estimating these latent mechanisms is through learning the data's generative model. In the realm of machine learning research, numerous methods have been proposed for learning generative models from extensive observational data[1–3]. While these methods often provide models that can accurately reconstruct data, their black-box nature makes it difficult for humans to understand what the latent mechanisms are. To address this, some methods have been proposed to learn models associated with known mechanisms[4,5], but finding unknown mechanisms remains challenging, limiting the potential for new insights. Therefore, a method is needed that allows humans to intuitively understand and accurately explain the data's generative process through latent mechanisms.

Extracting latent factors through linear projection is one promising method for deriving latent mechanisms from data with tens of thousands of variables. Principal Component Analysis (PCA)[6] and Independent Component Analysis (ICA)[7] are typical methods in this regard. These methods can provide clues to the global structure behind the data, facilitating human comprehension. However, these methods assume that latent factors are orthogonal or independent, complicating the modeling of interactions between latent mechanisms. Additionally, there is no guarantee that latent mechanisms identified under one situation will apply to datasets observed in another situation. Methods have also been proposed that estimate latent factors assuming they have sparse interactions among themselves[8,9], allowing some degree of modeling between hidden mechanisms' interactions, but their applicability across different datasets is not guaranteed.

[1]Fujitsu Research, 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 2118588, Kanagawa, Japan. [2]Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan. [3]M &D Data Science Center, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan. [4]School of Mathematics, Statistics and Data Science, Sungshin Women's University, 2, Bomun-ro 34da-gil, Seongbuk-gu, Seoul 02844, Republic of Korea. ✉email: maruhashi.koji@fujitsu.com; heewonn.park@gmail.com

An alternative approach involves estimating the dependencies among all observed variables first, and then inferring common latent mechanisms across datasets based on the structure of these dependencies. Assuming a Gaussian distribution for the data, the precision matrix represents the dependencies between variables. That is, for two variables that are conditionally independent, the corresponding value in the precision matrix is zero, whereas for variables that are dependent on each other, the value is non-zero. Graphical Lasso[10,11] assumes sparsity in these dependencies and employs an L1 regularization (LASSO) on the precision matrix for estimation. However, as the number of variables increases, the dependencies become more complex, making it challenging for humans to understand the overall picture and infer hidden mechanisms. Assuming in advance that there are common or similar structures across datasets, methods have been proposed to estimate the dependencies among observed variables across multiple datasets simultaneously[12,13]. Some of these methods involve clustering the observed variables into common groups across datasets and then estimating the dependencies within each cluster, assuming sparsity in these dependencies[14]. The groups of variables identified using these methods can be considered as common hidden mechanisms across datasets. However, the dependencies among these hidden mechanisms still need to be understood in terms of individual variable dependencies, making it challenging for humans to comprehend. Furthermore, methods that rigorously estimate these individual variable dependencies require computational resources proportional to at least the number of variable combinations, making it difficult to compute for data with tens of thousands of variables.

To solve these problems, we propose a new method named Meta Graphical Lasso (MGLASSO), which enables the estimation of easily understandable latent mechanisms (Fig. 1). We assume that even data generated based on seemingly complex systems can be explained by a few latent factors. A key difference from traditional latent factor methods is that our latent factors are common across multiple datasets, and we assume that there are sparse dependencies between a few latent factors within each dataset, whereas the remaining latent factors are white noise. Based on this assumption, we propose a Gaussian distribution-based graphical modeling that assumes common latent factors and sparse dependencies between them across multiple datasets. This model is learned using an algorithm that combines the well-known Graphical Lasso[11] with gradient descent on Stiefel Manifolds[15]. It should be noted that calculating the dependencies among a small number of latent factors is significantly faster compared to the conventional methods that compute the dependencies among all variables. We demonstrate

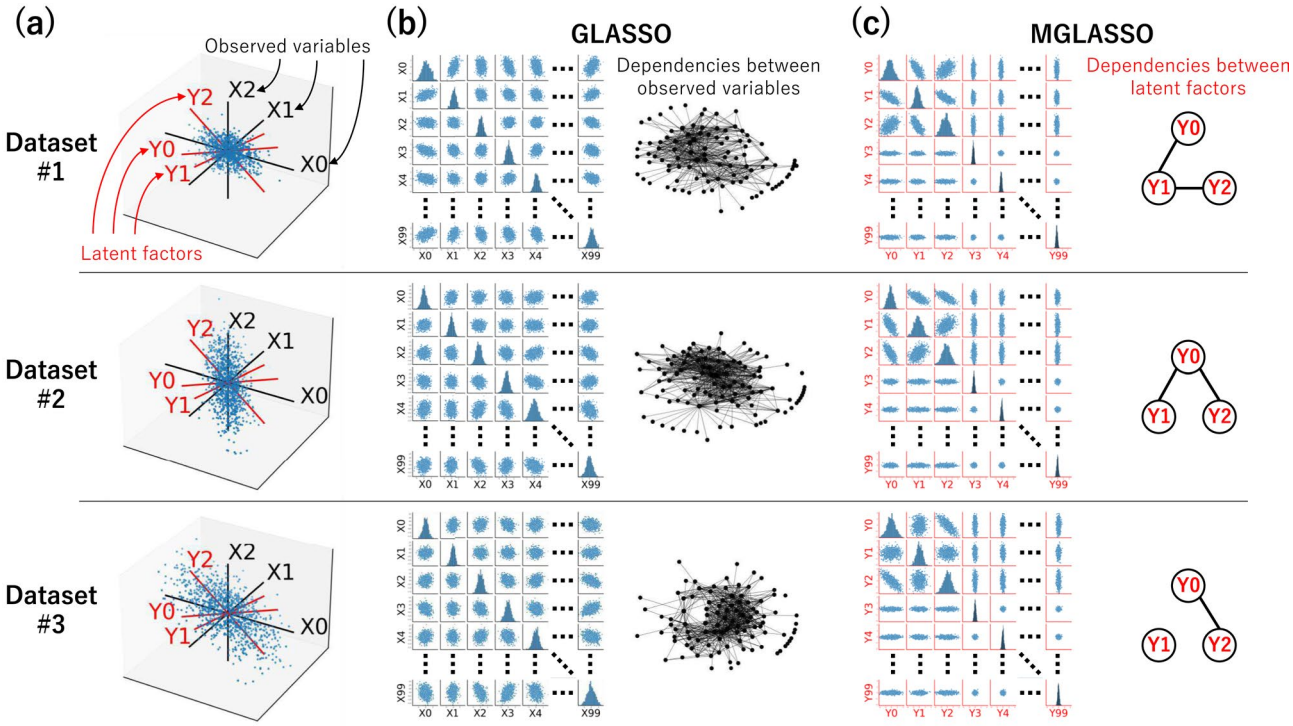

Figure 1. Overview of the proposed method. (**a**) There are multiple datasets observed under different contexts, each containing a large number of variables (here, 100). It is important to note that this plot visualizes only the dimensions corresponding to three observed variables ($X0$, $X1$, $X2$). The red line represents the projection in the three-dimensional space of latent factors ($Y0$, $Y1$, $Y2$) identified by the proposed method, Meta Graphical Lasso (MGLASSO), which discovers common latent factors across different contexts. (**b**) Based on the variance-covariance structure among the variables ($X0, X1, \ldots, X99$), it is possible to estimate sparse dependencies between variables using Graphical Lasso (GLASSO). However, even with 100 variables, the dependencies are too complex to identify latent mechanisms common across all datasets. (**c**) MGLASSO identifies a few latent factors ($Y0$, $Y1$, $Y2$) with sparse dependencies across any dataset, and treats the remaining latent factors ($Y3, Y4, \ldots, Y99$) as white noise with uniform variance. This facilitates the easy understanding of common latent mechanisms and their variations in dependencies across all datasets by humans.

through experiments with synthetic data that traditional latent factor estimation methods cannot accurately estimate sparse dependencies between latent factors, whereas our method can generally accurately estimate them.

While this modeling based on very simple assumptions does not precisely represent the reality of generative mechanisms, we empirically demonstrate that it can lead to surprisingly rich insights using real-world data from diverse domains. One example is in the financial domain, where we report the results applied to several years' worth of movements of thousands of stocks in the US market. We show that the overall movement of stock prices across the entire US market can be explained by a few key societal trends, going beyond traditional local analysis focused on individual stock movements to open new possibilities for comprehensive analysis from massive amounts of stocks. Another example is in the medical domain, where we report the application results to a dataset of gene expression levels related to drug resistance in lung cancer. We show that several stages of drug resistance can be understood through variations in interactions between a few common latent factors. This allows us to capture comprehensive interactions between biological processes that could not be discovered through individual gene interaction analysis, marking a significant advancement in understanding the essence of cancer drug resistance mechanisms and paving the way for developing more fundamental cancer treatments.

Our contributions can be summarized as follows:

- We propose a new graphical model, Meta Graphical Lasso (MGLASSO), that allows for understanding the generative process of datasets observed in various situations through a very simple assumption, namely, the sparse interactions between common latent mechanisms.
- We provide an algorithm that combines the well-known Graphical Lasso with gradient descent on Stiefel Manifolds to learn our proposed model. We demonstrate through synthetic data that our method can model more accurately than traditional methods.
- We empirically demonstrate the universal effectiveness of our method using datasets from two entirely different domains, namely finance and medicine. These results open the possibility of global analysis that can comprehensively understand all observed variables, moving beyond traditional local analysis based on a few specific observed variables.

Following, we define the problem we address in "Preliminaries", and propose our method in "Proposed method". We then evaluate performance using synthetic data in "Empirical results", and demonstrate how our method brings rich insights from large data in two different real-world domains: finance and medicine. Finally, we conclude in "Conclusion".

Preliminaries

Notations

Scalar quantities are represented by lowercase letters (e.g., a), vectors are indicated by boldface lowercase letters (e.g., $\mathbf{a}$), and matrices are signified by boldface uppercase letters (e.g., $\mathbf{A}$). The i^{th} element of a vector $\mathbf{a}$ is denoted as a_i. The transpose of a matrix $\mathbf{A}$ is represented as $\mathbf{A}^T$. $\text{tr}(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$, $|\boldsymbol{A}|$ represents the determinant of $\boldsymbol{A}$, and $|\boldsymbol{A}|_1$ is the L1 norm of $\boldsymbol{A}$, *i.e.*, the sum of the absolute values of the element of $\boldsymbol{A}$. $sign(\boldsymbol{A})$ represents a matrix that indicates the sign of each element in $\boldsymbol{A}$, *i.e.*, positive elements are marked with $+1$, negative elements with -1, and zero elements are assigned another value. The identity matrix is denoted by $\mathbf{I}$, while $\mathbf{1}$ and $\mathbf{0}$ represent vectors or matrices filled with ones and zeros, respectively. The notation $\langle \cdot, \cdot \rangle$ is used to denote the inner product. Element-wise operations, specifically multiplication and division, are represented by $*$ and $\oslash$, respectively. The derivatives of a function E with respect to the elements of $\mathbf{A}$ are expressed as $\frac{\partial E}{\partial \mathbf{A}}$.

Problem definition

In this work, we tackle the problem of estimating latent mechanisms through latent factors represented as linear projections of variables, based on data observed in multiple distinct contexts. We hypothesize that the dependencies among latent mechanisms are sparse in each context. Consider that we are given K datasets, each observed under different conditions, with V variables. The kth dataset is represented by a matrix $\mathbf{X}_k$ of size $V \times N_k$, containing observational data from N_k samples. Our problem is to represent the generative mechanism of the variables across these K datasets using an interpretable model that captures the sparse dependencies between a small number of latent factors, M, represented through the linear projection of variables, and to provide an efficient method for learning this model.

Sparse Gaussian graphical models

In the domain of high-dimensional data analysis, particularly when dealing with datasets that comprise N observations across V variables, represented as a matrix $\mathbf{X}$ of dimensions $V \times N$, the estimation of a sparse inverse covariance matrix, or precision matrix, $\mathbf{\Sigma}^{-1}$, becomes paramount. The Graphical Lasso algorithm, as delineated by Friedman et al.[11], offers a robust solution to this challenge. This algorithm operates by optimizing an objective function designed to estimate $\mathbf{\Sigma}^{-1}$ in a manner that encourages sparsity, which is particularly beneficial in understanding the conditional independence structure among the variables. The objective function E to be minimized is the Gaussian log-likelihood of data given by:

$$E = \frac{1}{N}\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{X}^T) - \log|\mathbf{\Sigma}^{-1}| + \rho|\mathbf{\Sigma}^{-1}|_1, \tag{1}$$

where $|\mathbf{\Sigma}^{-1}|_1$ enforces sparsity. The parameter ρ is a regularization coefficient that plays a crucial role in controlling the degree of sparsity in $\mathbf{\Sigma}^{-1}$, enabling the extraction of meaningful insights regarding the underlying

network of variable interactions. The optimal $\mathbf{\Sigma}^{-1}$ is obtained by solving $\frac{\partial E}{\partial \mathbf{\Sigma}^{-1}} = \mathbf{0}$, where the derivative of the objective function with respect to $\mathbf{\Sigma}^{-1}$ is given by

$$\frac{\partial E}{\partial \mathbf{\Sigma}^{-1}} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^T - \mathbf{\Sigma} + \rho \text{sign}(\mathbf{\Sigma}^{-1}). \tag{2}$$

Solutions to this problem are known through the use of the Coordinate Descent method[11] and the Alternating Direction Method of Multipliers (ADMM)[16].

Proposed method

In this section, we introduce a novel graphical model designed to estimate latent mechanisms common across datasets. Subsequently, we present a learning algorithm that combines the Graphical Lasso with gradient descent on Stiefel Manifolds, facilitating the efficient discovery of these latent factors.

Objective function

We propose a new graphical model, the Meta Graphical Lasso (MGLASSO). Our model projects observational data, which includes V variables, into a space formed by distinct basis vectors and applies a graphical model based on the Gaussian distribution within this projected space (Fig. 1). We assume the following:

1. Latent factors represented through projection into an M-dimensional subspace possess arbitrary variances and exhibit sparse dependencies among themselves.
2. The remaining $V - M$ dimensional subspace consists of white noise, meaning it contains only independent latent factors with a common variance.

Based on these assumptions, by searching for basis vectors that best explain the observed data, we expect to uncover M-dimensional subspaces with significant variance and sparse interdependencies across many datasets. As demonstrated in subsequent sections, modeling based on these remarkably simple assumptions can lead to astonishingly rich insights.

We define $\mathbf{D}$ as the $V \times M$ projection matrix for projecting into the M-dimensional subspace and $\widetilde{\mathbf{D}}$ as the $V \times (V - M)$ projection matrix for the remaining subspace. The column vectors of $\mathbf{D}$ and $\widetilde{\mathbf{D}}$, being basis vectors, fulfill the orthonormal condition; that is, $\mathbf{D}^T\mathbf{D} = \mathbf{I}$, $\widetilde{\mathbf{D}}^T\widetilde{\mathbf{D}} = \mathbf{I}$, and $\mathbf{D}^T\widetilde{\mathbf{D}} = \mathbf{0}$. Under our assumptions, the projection of the k^{th} dataset's observed data, $\mathbf{x}_k$, adheres to a Gaussian distribution as shown by:

$$\left(\mathbf{D}\widetilde{\mathbf{D}}\right)^T \mathbf{x}_k \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \varepsilon_k \mathbf{I} \end{pmatrix}\right), \tag{3}$$

where $\mathbf{\Sigma}_k$ is an $M \times M$ matrix, and ε_k represents the common variance for latent factors in the $V - M$ dimensional subspace. $\left(\mathbf{D}\widetilde{\mathbf{D}}\right)$ and $\begin{pmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \varepsilon_k \mathbf{I} \end{pmatrix}$ are block matrices of size $V \times V$. This equivalently means $\mathbf{x}_k$ is distributed according to:

$$\mathbf{x}_k \sim \mathcal{N}\left(\mathbf{0}, \left(\mathbf{D}\widetilde{\mathbf{D}}\right) \begin{pmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \varepsilon_k \mathbf{I} \end{pmatrix} \left(\mathbf{D}\widetilde{\mathbf{D}}\right)^T\right). \tag{4}$$

Given $\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}^T = \mathbf{I} - \mathbf{D}\mathbf{D}^T$, we can further express this as:

$$\mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Phi}_k), \text{ where } \mathbf{\Phi}_k = \boldsymbol{D}\mathbf{\Sigma}_k\boldsymbol{D}^T + \varepsilon_k(\mathbf{I} - \boldsymbol{D}\boldsymbol{D}^T). \tag{5}$$

Consequently, the precision matrix is $\mathbf{\Phi}_k^{-1} = \boldsymbol{D}\mathbf{\Sigma}_k^{-1}\boldsymbol{D}^T + \varepsilon_k^{-1}(\mathbf{I} - \boldsymbol{D}\boldsymbol{D}^T)$. To enforce sparsity in the precision matrix of the M-dimensional latent factors, $\mathbf{\Sigma}_k^{-1}$, we apply L1 regularization similar to the Graphical Lasso method. Hence, we define the Gaussian log-likelihood of data penalized by L1 regularization as:

$$L_k = \frac{1}{N_k}\text{tr}(\mathbf{\Phi}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T) - \log(\left|\mathbf{\Phi}_k^{-1}\right|) + \rho\left|\mathbf{\Sigma}_k^{-1}\right|_1, \tag{6}$$

and aim to minimize $E = \sum_k L_k$. This process estimates parameters ε_k, $\mathbf{\Sigma}_k^{-1}$, and $\boldsymbol{D}$, consistent with our model's assumptions.

Learning algorithm

Given the objective function is non-convex, we propose a methodology for seeking local optima. As will be demonstrated, ε_k and $\mathbf{\Sigma}_k^{-1}$ can compute the optimal solution that minimizes E when $\boldsymbol{D}$ is held fixed. Consequently, we propose an approach that iteratively updates ε_k and $\mathbf{\Sigma}_k^{-1}$ to their optimal solutions with $\boldsymbol{D}$ fixed, while optimizing $\boldsymbol{D}$ via stochastic gradient descent (SGD)[17]. The comprehensive algorithm is delineated in Algorithm 1.

Optimizing projection matrix on Stiefel manifolds

Firstly, we present a methodology for optimizing the projection matrix $\boldsymbol{D}$ via SGD. Since $\boldsymbol{D}$ satisfies the orthonormal condition $\mathbf{D}^T\mathbf{D} = \mathbf{I}$, this optimization occurs on Stiefel Manifolds, a type of Riemannian manifold[15]. Consequently, we adopt the simplified formulation introduced by Maruhashi et al.[18] Specifically, we introduce a latent variable $\boldsymbol{Z}$ on the tangent space and utilize its Singular Value Decomposition (SVD)[19],

$$\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T, \tag{7}$$

to calculate $\boldsymbol{D}$ to pull back from the tangent space to the manifold as

$$\boldsymbol{D} = \boldsymbol{P}\boldsymbol{Q}^T, \tag{8}$$

subsequently optimizing $\boldsymbol{Z}$ using SGD. Here, $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}$ and $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, with $\boldsymbol{S}$ being a diagonal matrix with non-diagonal elements set to zero.

Theorem 1 *The gradient of the objective function E with respect to* $\boldsymbol{Z}$ *on the tangent space can be calculated as follows*:

$$\frac{\partial E}{\partial \boldsymbol{Z}} = \frac{\partial E}{\partial \boldsymbol{D}} - \frac{1}{2}\boldsymbol{D}\left(\boldsymbol{D}^T\frac{\partial E}{\partial \boldsymbol{D}} + \frac{\partial E}{\partial \boldsymbol{D}^T}\boldsymbol{D}\right), \tag{9}$$

where

$$\frac{\partial E}{\partial \boldsymbol{D}} = \sum_k \frac{2}{N_k}\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D}\left(\boldsymbol{\Sigma}_k^{-1} - \varepsilon_k^{-1}\boldsymbol{I}\right). \tag{10}$$

Proof According to Maruhashi et al.[18], when Eqs. 7 and 8 are satisfied,

$$\frac{\partial E}{\partial \boldsymbol{Z}} = \boldsymbol{P}\left[\left(\boldsymbol{P}^T\frac{\partial E}{\partial \boldsymbol{D}}\boldsymbol{Q} - \boldsymbol{Q}^T\frac{\partial E}{\partial \boldsymbol{D}^T}\boldsymbol{P}\right) \oslash (\boldsymbol{S}\boldsymbol{1} + \boldsymbol{1}\boldsymbol{S})\right]\boldsymbol{Q}^T + (\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^T)\frac{\partial E}{\partial \boldsymbol{D}}\boldsymbol{Q}\boldsymbol{S}^{-1}\boldsymbol{Q}^T. \tag{11}$$

Assuming the previous $\boldsymbol{D}$ as $\boldsymbol{Z}$ implies that $\boldsymbol{P} = \boldsymbol{D}$ and $\boldsymbol{S} = \boldsymbol{Q} = \boldsymbol{I}$, thus,

$$\frac{\partial E}{\partial \boldsymbol{Z}} = \frac{1}{2}\boldsymbol{D}\left(\boldsymbol{D}^T\frac{\partial E}{\partial \boldsymbol{D}} - \frac{\partial E}{\partial \boldsymbol{D}^T}\boldsymbol{D}\right) + (\boldsymbol{I} - \boldsymbol{D}\boldsymbol{D}^T)\frac{\partial E}{\partial \boldsymbol{D}}. \tag{12}$$

By transforming this expression, we can derive Eq. (9). Furthermore, by differentiating Eq. (6) with respect to each element of $\boldsymbol{D}$, we obtain Eq. (10). □

We update $\boldsymbol{Z}$ on the tangent space using the gradient $\frac{\partial E}{\partial \boldsymbol{Z}}$, computed via Eq. (9). Subsequently, a new $\boldsymbol{D}$ is calculated to pull back from the tangent space to the manifold employing Eqs. (7) and (8). By iterating this process, we can optimize $\boldsymbol{D}$.

Precision matrix for latent factors
Here, we address the optimal solution for $\boldsymbol{\Sigma}_k^{-1}$ when $\boldsymbol{D}$ is held fixed. The gradient of the objective function E with respect to $\boldsymbol{\Sigma}_k^{-1}$ is

$$\frac{\partial E}{\partial \boldsymbol{\Sigma}_k^{-1}} = \frac{1}{N_k}\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D} - \boldsymbol{\Sigma}_k + \rho\,\mathrm{sign}(\boldsymbol{\Sigma}_k^{-1}). \tag{13}$$

The optimal solution for $\boldsymbol{\Sigma}_k^{-1}$ is obtained by solving $\frac{\partial L_k}{\partial \boldsymbol{\Sigma}_k^{-1}} = \boldsymbol{0}$, which aligns with the Graphical Lasso approach (Eq. 2) when the precision matrix is $\frac{1}{N_k}\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D}$.

Common variance for remaining subspace
Furthermore, we describe the optimal solution for ε_k when $\boldsymbol{D}$ is held fixed. Considering

$$\log(|\boldsymbol{\Phi}_k^{-1}|) = \log(|\boldsymbol{\Sigma}_k^{-1}|) - (V - M)\log(\varepsilon_k), \tag{14}$$

we derive

$$\frac{\partial E}{\partial \varepsilon_k} = -\frac{1}{N_k}\left(\mathrm{tr}(\boldsymbol{X}_k\boldsymbol{X}_k^T) - \mathrm{tr}(\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D})\right)\varepsilon_k^{-2} + (V - M)\varepsilon_k^{-1}. \tag{15}$$

By setting $\frac{\partial E}{\partial \varepsilon_k} = 0$, we compute

$$\varepsilon_k = (V - M)^{-1}\frac{1}{N_k}\left(\mathrm{tr}(\boldsymbol{X}_k\boldsymbol{X}_k^T) - \mathrm{tr}(\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D})\right). \tag{16}$$

This formula allows for the computation of ε_k.

Input: $\boldsymbol{X}_k, \rho, M$
Output: $\boldsymbol{\Sigma}_k^{-1}, \varepsilon_k, \boldsymbol{D}$
initialize $\boldsymbol{\Sigma}_k^{-1}, \varepsilon_k, \boldsymbol{D}$
repeat
  $\varepsilon_k \leftarrow (V-M)^{-1}\frac{1}{N_k}\left(tr(\boldsymbol{X}_k\boldsymbol{X}_k^T) - tr(\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D})\right)$
  $\boldsymbol{\Sigma}_k^{-1} \leftarrow GraphicalLasso\left(\frac{1}{N_k}\boldsymbol{D}^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D}, \rho\right)$
  $\frac{\partial E}{\partial \boldsymbol{D}} \leftarrow \sum_k \frac{2}{N_k}\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{D}\left(\boldsymbol{\Sigma}_k^{-1} - \varepsilon_k^{-1}\boldsymbol{I}\right)$
  $\boldsymbol{Z} \leftarrow \boldsymbol{D}$
  update $\boldsymbol{Z}$ via $\frac{\partial E}{\partial \boldsymbol{Z}} = \frac{\partial E}{\partial \boldsymbol{D}} - \frac{1}{2}\boldsymbol{D}\left(\boldsymbol{D}^T\frac{\partial E}{\partial \boldsymbol{D}} + \frac{\partial E}{\partial \boldsymbol{D}^T}\boldsymbol{D}\right)$
  $\boldsymbol{P}, \boldsymbol{S}, \boldsymbol{Q}^T \leftarrow svd(\boldsymbol{Z})$
  $\boldsymbol{D} \leftarrow \boldsymbol{P}\boldsymbol{Q}^T$
until convergence
return $\boldsymbol{\Sigma}_k^{-1}, \varepsilon_k, \boldsymbol{D}$

Algorithm 1. Meta graphical Lasso (MGLASSO)

Convergence and initialization
While $\boldsymbol{\Sigma}_k^{-1}$ and ε_k can be optimized if D is fixed, D does not necessarily converge to an optimal solution. However, setting appropriate initial values for D can somewhat mitigate solution instability. We propose using the top M principal components obtained by applying PCA to the combined data of all datasets as the initial values for D. Adopting this initialization allows for stable solutions unaffected by the randomness of SGD due to mini-batches, as confirmed across all datasets used in the next section. Additionally, we demonstrate that for at least the two real-world datasets we handled, namely the Stock Price dataset and the Gene Expression dataset, similar conclusions can be drawn even when D is randomly initialized (Supplementary Information).

Computational complexity
We discuss the time complexity of the proposed method. Here, noting that $M < V$, we describe the computational complexity per epoch. The computational complexity for ε_k, $\frac{\partial E}{\partial \boldsymbol{D}}$, and $\frac{\partial E}{\partial \boldsymbol{Z}}$ is $O(V(M\sum_k N_k + \sum_k {N_k}^2 + M^2))$. Additionally, the computational complexity for the graphical lasso, assuming the number of iterations is I, is at most $O(IM^3)$. Furthermore, performing SVD to obtain only the top M singular vectors of a matrix $\boldsymbol{Z}$ of size $V \times M$ can be efficiently done in $O(VM^2)$ using the well-known Lanczos method[20]. Summarizing the above, the overall time complexity of the proposed method per epoch is $O(V(M\sum_k N_k + \sum_k {N_k}^2 + M^2) + IM^3)$. This shows that the computational complexity is proportional to the number of variables even when V is very large, such as in gene expression data with tens of thousands of variables. For such data, methods that rigorously estimate dependencies at the level of individual variables[14], which require computational complexity proportional to at least the number of variable combinations, are difficult to apply.

Parameter selection
We propose using the Bayesian Information Criterion (BIC)[21] as a guideline for selecting the hyperparameters M and ρ in the proposed method. BIC is calculated as

$$BIC = -2\ln(L) + k\ln(n), \tag{17}$$

where L is the likelihood function, n is the number of observations, and k is the number of independent variables. In the proposed method, L is given by $E = \sum_k L_k$, and n is $\sum_k N_k$. The number of independent variables k is the sum of the number of non-zero elements in $\boldsymbol{\Sigma}_k^{-1}$ across all k (i.e., $\sum_k NNZ(\boldsymbol{\Sigma}_k^{-1})$), the number of ε_k parameters (i.e., K), and the size of $\mathbf{D}$, which is $V \times M$. Note that the degrees of freedom for $\mathbf{D}$ are reduced by $M(M+1)/2$ due to the orthogonality constraint. Therefore, the number of independent variables k is given by $\sum_k NNZ(\boldsymbol{\Sigma}_k^{-1}) + K + V \times M - M(M+1)/2$.

Empirical results

Synthesized dataset

Here, we utilize synthetic data generated based on latent factors with sparse dependencies to validate whether these sparse dependencies can be accurately estimated.

Data generation
We generate K datasets consisting of N samples, each containing V observed variables. These datasets share M common latent factors, among which we assume sparse dependencies. The procedure for generating these datasets is outlined below:

Creating latent factor To generate a matrix $\boldsymbol{D}$ that satisfies the orthonormal condition $\boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{I}$, we first create a matrix $\boldsymbol{Z}$ of the same size as $\boldsymbol{D}$ with random values sampled from a uniform distribution over [0, 1). Then, using $\boldsymbol{P}$ and $\boldsymbol{Q}$ obtained from SVD $\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T$, we generate $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{Q}^T$. Here, $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}$ and $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, where $\boldsymbol{S}$ is a square matrix with non-diagonal elements being zero.

Generating precision matrix For each of the K datasets, we generate the precision matrix of latent factors, $\mathbf{\Sigma}_k^{-1}$. Specifically, we first generate 20 M-dimensional vectors with random values sampled from a normal distribution with mean 0 and standard deviation 1. The precision matrix for these samples is then estimated using Graphical Lasso with an L1 regularization weight of 0.3, and this estimated matrix is designated as $\mathbf{\Sigma}_k^{-1}$.

Generating white noise The common variance ε_k for the remaining $V - M$ dimensions of latent factors is generated as the absolute value of random values sampled from a normal distribution with mean 0 and standard deviation 0.1. This assumes that the variance of the remaining latent factors is smaller than that of the M-dimensional latent factors.

Synthesizing samples The covariance matrix $\mathbf{\Phi}_k$ is defined as $\mathbf{\Phi}_k = \boldsymbol{D}\mathbf{\Sigma}_k\boldsymbol{D}^T + \varepsilon_k(\boldsymbol{I} - \boldsymbol{D}\boldsymbol{D}^T)$, and N samples are drawn from a Gaussian distribution represented by $\mathcal{N}(\mathbf{0}, \mathbf{\Phi}_k)$.

In this experiment, the number of latent factors M is set to 10, and the number of datasets K is 10. The number of observed variables V is tested in variations of 100, 1000, and 10, 000. Additionally, the number of samples N is tested for 200 and 1000. The MGLASSO model is trained using SGD for 5000 epochs, ensuring sufficient convergence. The L1 regularization weight ρ is selected for each V and N to achieve sparsity levels comparable to the ground truth. The resultant latent factors are reordered to closely match the ground truth latent factors. Specifically, we first select the latent factor closest in Euclidean distance to any of the ground truth latent factors as its corresponding latent factor and subsequently match the remaining factors based on the closest Euclidean distance. We report the average results obtained from repeating this verification process 10 times.

Comparison methods

Our goal is to propose a method using latent factors derived through linear projection, which is one of the effective ways to infer latent mechanisms in a form easily understandable by humans. Therefore, we compare our proposed method with the main traditional methods using latent factors through linear projection, such as PCA and ICA, to demonstrate that our approach can appropriately estimate the sparse dependencies among latent factors.

We apply traditional methods to the same synthetic data used in the evaluation of MGLASSO for our assessment. However, since PCA and ICA can only consider a single dataset, we apply PCA and ICA to a concatenated dataset comprising all individual datasets to estimate latent factors. It was observed that directly concatenating leads to an inability to estimate the dependencies among latent factors of datasets other than those with significant variance. Thus, we normalize the variance of the variables in each dataset to 1 before concatenation. For PCA, the major M principal components are regarded as latent factors. In the case of ICA, the M independent components extracted using the FastICA[7] algorithm are considered latent factors. Furthermore, similar to the procedure with MGLASSO, the estimated latent factors are reordered according to their correspondence with the ground truth. The estimated latent factors are then projected into an M-dimensional space using a projection matrix $\boldsymbol{D}'$, comprising column vectors of the latent factors, and a precision matrix is obtained by applying Graphical Lasso to each dataset.

Results on synthesized datasets

The results of the evaluation experiments using synthetic data are displayed in Table 1. Table 1 initially shows that a larger sample size N facilitates better alignment with the ground truth, however, increasing the number of observed variables V complicates this alignment. The proposed method consistently shows better results across all evaluation metrics compared to PCA or ICA. Notably, although the alignment of space projected by M latent factors (Space) is relatively good in PCA and ICA, the inner product of matched latent factors (Dist) and the Kullback-Leibler divergence of the Gaussian distributions (KLD) significantly outperform those in PCA or ICA. This indicates that, while PCA and ICA capture subspaces with a significant variance of latent factors, MGLASSO significantly excels in identifying latent factors from those subspaces with sparse dependencies across all datasets. It can also be confirmed that the KL divergence between the learned multivariate Gaussian distribution and the ground truth multivariate Gaussian distribution is smallest when $M = 10$ (Table 2).

An example of sparse dependencies in the estimated latent space is shown in Fig. 2, illustrating the case when $V = 10,000, N = 1000$. Here, the partial correlation coefficients, commonly used to understand dependencies between variables, are displayed. When the precision matrix is $\mathbf{\Lambda}$, the partial correlation coefficient between the ith and jth variables is given by $-\frac{\mathbf{\Lambda}ij}{\sqrt{\mathbf{\Lambda}ii}\sqrt{\mathbf{\Lambda}_{jj}}}$. It is observed that PCA and ICA are largely unsuccessful in accurately estimating the ground truth dependencies. In contrast, MGLASSO is almost entirely successful in accurately estimating the ground truth dependencies.

In MGLASSO, the weight ρ for L1 regularization is confirmed to generally provide optimal BIC for each V and N (Table 3). Additionally, the number of latent factors M generally shows optimal BIC when $M = 10$, which corresponds to the ground truth (Table 4). However, for $V = 10,000$, the smallest BIC is obtained with $M = 5$, with $M = 10$ resulting in the next smallest BIC. Overall, these findings suggest that BIC serves as a reasonably good criterion for selecting ρ and M when the dataset adheres to the assumption of being generated based on a multivariate Gaussian distribution.

Experiments with synthetic data strongly suggest that accurately estimating sparse dependencies among latent factors across numerous datasets is extremely challenging with traditional methods of extracting latent factors, yet the proposed method is a potent solution to this challenge.

Stock price dataset

We demonstrate that our proposed method can unveil significant societal trends behind the movements of numerous stocks by applying it to data on the movements of a large number of stocks. We apply our method to stock price data from the US market from 2009 to 2015 downloaded from the corresponding URL (https://github.

method	V	N	ρ	Space	Eps	Dist	KLD
PCA	100	200	0.09	9.919 ± 0.013	0.00823 ± 0.01018	0.630 ± 0.051	0.121 ± 0.027
ICA	100	200	0.09	9.919 ± 0.013	0.00823 ± 0.01019	0.674 ± 0.058	0.098 ± 0.017
MGLASSO	100	200	0.09	**9.963 ± 0.009**	**0.00681 ± 0.00618**	**0.958 ± 0.014**	**0.039 ± 0.008**
PCA	100	1000	0.09	9.956 ± 0.005	0.00450 ± 0.00548	0.695 ± 0.043	0.094 ± 0.023
ICA	100	1000	0.09	9.956 ± 0.005	0.00450 ± 0.00548	0.773 ± 0.089	0.070 ± 0.015
MGLASSO	100	1000	0.06	**9.993 ± 0.002**	**0.00438 ± 0.00602**	**0.992 ± 0.004**	**0.015 ± 0.003**
PCA	1000	200	0.09	9.543 ± 0.124	0.00731 ± 0.01563	0.643 ± 0.040	0.136 ± 0.024
ICA	1000	200	0.09	9.543 ± 0.125	0.00731 ± 0.01563	0.647 ± 0.052	0.121 ± 0.028
MGLASSO	1000	200	0.09	**9.576 ± 0.095**	**0.00216 ± 0.00174**	**0.912 ± 0.036**	**0.048 ± 0.019**
PCA	1000	1000	0.09	9.920 ± 0.012	**0.00107 ± 0.00109**	0.729 ± 0.070	0.085 ± 0.029
ICA	1000	1000	0.09	9.920 ± 0.012	**0.00107 ± 0.00109**	0.778 ± 0.096	0.068 ± 0.021
MGLASSO	1000	1000	0.06	**9.927 ± 0.012**	0.00126 ± 0.00139	**0.985 ± 0.005**	**0.016 ± 0.005**
PCA	10,000	200	0.09	7.279 ± 0.517	0.00127 ± 0.00088	0.459 ± 0.040	0.152 ± 0.020
ICA	10,000	200	0.09	7.278 ± 0.517	0.00127 ± 0.00089	0.480 ± 0.059	0.155 ± 0.026
MGLASSO	10,000	200	0.09	**7.363 ± 0.473**	**0.00080 ± 0.00032**	**0.649 ± 0.096**	**0.133 ± 0.081**
PCA	10,000	1000	0.09	9.312 ± 0.116	0.00025 ± 0.00018	0.679 ± 0.071	0.098 ± 0.021
ICA	10,000	1000	0.09	9.312 ± 0.116	0.00025 ± 0.00018	0.737 ± 0.066	0.079 ± 0.014
MGLASSO	10,000	1000	0.06	**9.331 ± 0.105**	**0.00025 ± 0.00015**	**0.925 ± 0.012**	**0.016 ± 0.006**

Table 1. Evaluation using synthetic data. Space: the degree of alignment with the ground truth of the space spanned by M latent factors. The mean and standard deviation of 10 trials are reported. The degree of space alignment is calculated as the sum of the singular values of the square matrix $\boldsymbol{D}^T\boldsymbol{D}'$, where $\boldsymbol{D}$ is the ground truth projection matrix and $\boldsymbol{D}'$ is the estimated projection matrix. This represents the sum of inner products of corresponding basis vectors when $\boldsymbol{D}$ and $\boldsymbol{D}'$ are rotated to maximize the sum of inner products of basis vectors, with a higher value indicating better alignment, and the maximum value of $M = 10$ taken when perfectly aligned. It's noted that in ICA, latent factors are not necessarily orthogonal, hence the left singular vectors of $\boldsymbol{D}'$ are computed and used for the basis vectors. Eps: the average difference of ε_k from the ground truth across all datasets. The mean and standard deviation of 10 trials are presented. Dist: the average inner product between M latent factors aligned according to the ground truth and the latent factors of the ground truth, averaged over M latent factors, and further averaged over K datasets. A value close to 1 signifies good correspondence between the ground truth and the latent factors. The mean and standard deviation of 10 trials are reported. KLD: the average Kullback–Leibler divergence of the Gaussian distributions $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ followed by M latent factors from the ground truth, averaged over K datasets. A value close to 0 indicates that the distribution of latent factors of the ground truth has been well estimated. The mean and standard deviation of 10 trials are reported. Significant values are in bold.

V	N	ρ	$M = 5$	$M = 10$	$M = 20$
100	200	0.09	3.3590×10^2	$\mathbf{3.2840 \times 10^2}$	3.3770×10^2
100	1000	0.06	3.8278×10^2	$\mathbf{3.7131 \times 10^2}$	3.8055×10^2
1000	200	0.09	3.6268×10^3	$\mathbf{3.6169 \times 10^3}$	3.6302×10^3
1000	1000	0.06	4.3129×10^3	$\mathbf{4.3112 \times 10^3}$	4.3205×10^3
10,000	200	0.09	4.4532×10^4	$\mathbf{4.4521 \times 10^4}$	4.4539×10^4
10,000	1000	0.06	4.1279×10^4	$\mathbf{4.1267 \times 10^4}$	4.1274×10^4

Table 2. KL divergence between the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_k)$ with different M obtained using MGLASSO and that of the ground truth for synthetic data. The table shows the average values over five trials and indicates that the KL divergence is smallest when $M = 10$, which corresponds to the ground truth (bold).

com/eliangcs/pystock-data), selecting $V = 4,234$ stocks with data available on the same days. We use this data as datasets divided by year. Specifically, we create one dataset for each year from 2009 to 2015, each containing $N_0 = 252, N_1 = 252, N_2 = 252, N_3 = 250, N_4 = 252, N_5 = 252$, and $N_6 = 54$ days, respectively. The stock prices are converted to logarithmic values and standardized so that the mean is 0 for each year. We use $M = 10$ as the number of latent factors, a number that is manageable for human analysis. For the L1 regularization weight, we adopt $\rho = 3$, indicating moderate sparsity. It can be observed that the BIC shows better values compared to when using different ρ (Table 5). On the other hand, although the number of latent factors M does not always yield the best BIC, it provides relatively reasonable BIC values compared to poorly fitting models such as $M = 2$ (see Supplementary Information). Considering that real-world data is not as simple as being represented by a

V	N	$\rho = 0.03$	$\rho = 0.06$	$\rho = 0.09$
100	200	2.9708×10^5	2.9643×10^5	$\mathbf{\textit{2.9614} \times 10^5}$
100	1000	1.1667×10^6	$\mathbf{\textit{1.1661} \times 10^6}$	1.1661×10^6
1000	200	2.2391×10^6	2.2385×10^6	$\mathbf{\textit{2.2381} \times 10^6}$
1000	1000	1.1539×10^7	$\mathbf{\textit{1.1536} \times 10^7}$	1.1537×10^7
10,000	200	1.8425×10^7	1.8426×10^7	$\mathbf{\textit{1.8424} \times 10^7}$
10,000	1000	1.0128×10^8	$\textit{1.0128} \times 10^8$	$\mathbf{1.0128 \times 10^8}$

Table 3. BIC values for MGLASSO applied to synthetic data using different ρ. The table shows the average values over five trials. The ρ values that yield sparsity levels similar to the ground truth are in italics. It can be observed that BIC is generally optimal when ρ is chosen in this manner. Significant values are in bold.

V	N	ρ	$M = 5$	$M = 10$	$M = 20$
100	200	0.09	3.1822×10^5	$\mathbf{\textit{2.9614} \times 10^5}$	2.9995×10^5
100	1000	0.06	1.3371×10^6	$\mathbf{\textit{1.1661} \times 10^6}$	1.1727×10^6
1000	200	0.09	2.2434×10^6	$\mathbf{\textit{2.2381} \times 10^6}$	2.2822×10^6
1000	1000	0.06	1.1607×10^7	$\mathbf{\textit{1.1536} \times 10^7}$	1.1541×10^7
10,000	200	0.09	$\mathbf{1.8145 \times 10^7}$	$\textit{1.8424} \times 10^7$	1.9042×10^7
10,000	1000	0.06	$\mathbf{1.0110 \times 10^8}$	$\textit{1.0128} \times 10^8$	1.0202×10^8

Table 4. BIC values when applying MGLASSO with different M on synthetic data. The table shows the average values over five trials. It is observed that BIC is generally optimal for the ground truth $M = 10$ (italics). However, for $V = 10,000$, $M = 5$ is optimal, with $M = 10$ showing the next smallest value. Significant values are in bold.

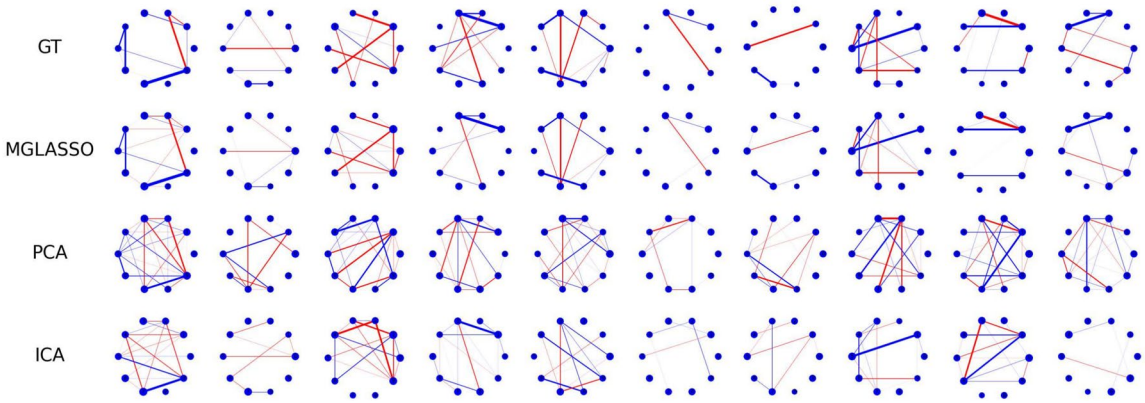

Figure 2. Graphical representation of the dependencies between latent factors estimated from synthetic data. Edges represent non-zero partial correlation coefficients between latent factors, with blue edges indicating positive partial correlations and red edges indicating negative ones. The thickness of an edge corresponds to the magnitude of the absolute value of the partial correlation coefficient. Additionally, the size of a node reflects the variance of the latent factor, that is, the magnitude of the diagonal elements in the covariance matrix. PCA and ICA fail to accurately capture the ground truth dependencies, whereas MGLASSO successfully approximates them with high fidelity. *GT* ground truth, *MGLASSO, PCA, ICA* dependencies between latent factors estimated using MGLASSO, PCA, and ICA, respectively.

Dataset	$\rho = 3$	$\rho = 30$	$\rho = 300$
Stock price	$\mathbf{-1.2629 \times 10^7}$	-1.2625×10^7	-1.2624×10^7
Gene expression	3.1412×10^7	$\mathbf{3.1410 \times 10^7}$	3.141×10^7

Table 5. BIC values for MGLASSO applied to U.S. stock price data and gene expression data using different ρ . The table shows that the BIC is optimal for the ρ values selected in this evaluation experiment, as indicated by the bold entries.

single multivariate Gaussian distribution, it can be said that striving for the best BIC and other criteria is not very meaningful. Therefore, from the perspective of interpretability, we provide an analysis using $M = 10$ in this case. The model is learned over 5,000 steps using stochastic gradient descent.

Figure 3 shows the daily magnitude of latent factors estimated using the projection matrix derived by MGLASSO. Latent factors with clear trends and those without can be observed each year, with factors showing clear trends likely having dependencies on that year. This suggests that our method proactively extracts latent factors that become sparsely dependent annually. Indeed, years with significant variance in latent factors can be seen from Fig. 4. Moreover, the sparsity of dependencies between latent factors can be confirmed from Fig. 5. Thus, our method successfully extracts latent factors that show significant movements each year and demonstrate sparse dependencies annually.

Table 6 presents the characteristics of stocks that significantly contribute to each latent factor. In this data, all stocks are classified by sector, and stocks within each sector are further categorized by industry. Analyzing the behavior of latent factors shown in Fig. 3, 4, and 5 based on these characteristics leads to the following interpretations:

1. Latent factor #1 is associated with the "Other Precious Metals & Mining Silver" industry, and latent factor #3 is linked to the "Electronic Components" industry. In 2010, these industries showed a dependency relationship, which reversed by 2012. This could reflect the impact of significant international contexts, such as China's export restrictions on rare metals in 2010 and the tensions over rare metals between the US and Europe in 2012[22].
2. Latent factor #8 is strongly related to the energy sector, particularly the "Oil & Gas Drilling" industry, showing significant movements in 2012 and 2014. This period marked a crucial phase for the U.S. oil and gas drilling

Latent factors	p/n	Sector	p-value	Industry	p-value
0	pos	Basic materials	0.006	–	
	neg	–		–	
1	pos	Healthcare	<0.001	Biotechnology	0.002
	neg	Basic materials	0.005	Other precious metals & Mining	<0.001
		Communication services	0.036	–	
2	pos	Basic materials	<0.001	Gold	<0.001
		Healthcare	<0.001	Biotechnology	<0.001
	neg	Healthcare	<0.001	Biotechnology	0.049
		Communication services	0.043	- -	
3	pos	Healthcare	<0.001	Biotechnology	0.012
	neg	Technology	<0.001	Electronic components	<0.001
		Healthcare	0.026	–	
4	pos	Basic materials	<0.001	Gold	0.011
	neg	Healthcare	<0.001	–	
5	pos	Consumer cyclical	<0.001	–	
		Healthcare	0.008	–	
	neg	–		–	
6	pos	Healthcare	<0.001	Biotechnology	0.049
		Technology	0.035	Communication Equipment	0.029
	neg	-		-	
7	pos	Healthcare	<0.001	Biotechnology	0.007
	neg	Communication services	<0.001	Entertainment	0.024
				Publishing	<0.001
8	pos	Energy	<0.001	Oil & gas drilling	0.020
		Basic Materials	0.003	Other Precious Metals & Mining	<0.001
	neg	Healthcare	0.004	Biotechnology	<0.001
9	pos	Healthcare	<0.001	Biotechnology	0.006
	neg	Healthcare	0.001	–	

Table 6. Sectors and industries significantly associated with the 10 latent factors identified by MGLASSO from US stock price data. Initially, stocks with linear projection weights deviating more than two standard deviations from their mean are classified as characteristic stock groups for each latent factor, based on the sign (positive or negative) of their linear projection weights. A log-likelihood ratio test is then conducted between these characteristic groups of stocks and all sectors, identifying sectors with a p-value less than 0.05. Additionally, within these sectors, a log-likelihood ratio test is performed between the characteristic groups of stocks and industries, marking industries with a p-value less than 0.05. Sectors or industries without any associated p-values less than 0.05 are denoted by '–'.

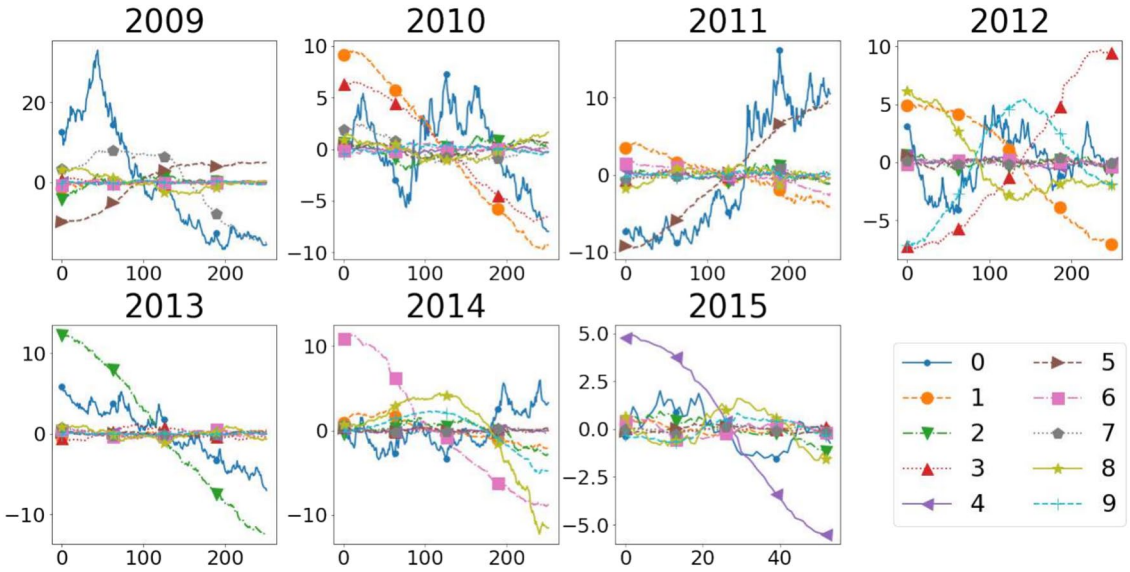

Figure 3. Daily plots of the values of latent factors extracted by MGLASSO. Horizontal axis: Days arranged in chronological order. Vertical axis: Values of the latent factors.

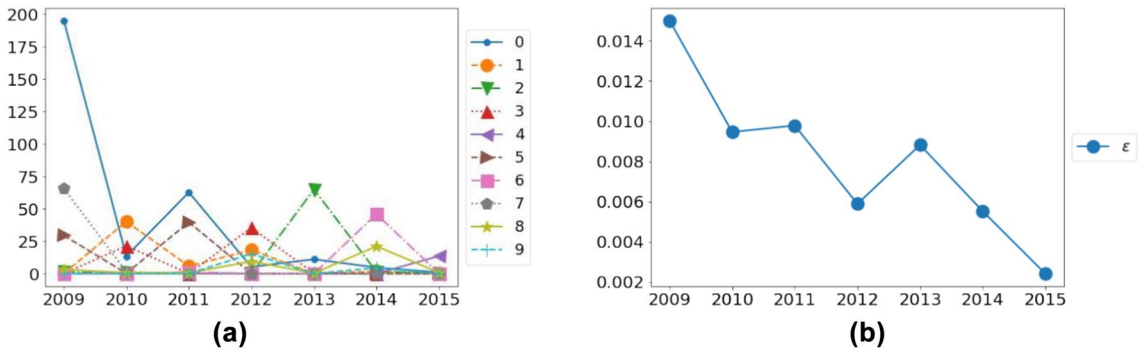

Figure 4. The variances of the latent factors estimated from US stock price data. (**A**) The variance of latent factors, *i.e.*, the diagonal components of the covariance matrix $\mathbf{\Sigma}_k$ of latent factors, plotted annually. (**B**) The common variance ε_k of the remaining latent factors plotted annually.

Figure 5. Graph representation of the dependency among latent factors estimated from the stock price data of the United States. Edges represent the nonzero partial correlation coefficients between latent factors. Blue edges indicate positive, and red edges indicate negative partial correlation coefficients, with the thickness of the edges representing the magnitude of the absolute value of the partial correlation coefficients. Additionally, the size of the nodes reflects the variance of the latent factors, i.e., the magnitude of the diagonal components of the covariance matrix.

industry, notably due to the "Shale Revolution," which garnered significant attention for its rapid production increases, economic benefits, and environmental impacts[23]. Additionally, latent variable #9, moving in strong correlation with #8 in the same years, is closely associated with healthcare, suggesting potential insights from companies related to both sectors.

3. Latent factor #2 is strongly associated with the "Gold" industry and exhibited significant variation in 2013. While gold prices are influenced by various factors and are difficult to attribute to any single cause, 2013 was marked by significant government interventions in gold prices, potentially reflecting these actions' strong impact.

4. Latent variables #0 and #5 experienced significant fluctuations with a strong dependency relationship between them in 2009 and 2011. Since no closely related industry could be identified for these latent variables, the analysis presents challenges. However, examining related companies from different perspectives may yield intriguing insights.

Such insights could never be obtained from traditional analyses focused solely on the movements of a few representative stocks. Additionally, while stock price movements are strongly influenced by major societal situations occurring each year, it is challenging to understand stock price movements over several years with a consistent societal situation by identifying major latent factors annually. Our proposed method provides a consistent explanation for all stock price movements over a certain period, including several years, and objectively estimates major societal situations affecting stock prices through characteristic interactions each year, without being hindered by experts' preconceptions.

Gene expression dataset

We demonstrate that applying our proposed method to gene expression data of cancer cells observed in actual drug resistance of cancer yields surprisingly profound insights.

Pathway network analysis for uncovering mechanisms of acquired 5-FU resistance

We apply our method to analyze pathway networks varying depending 5-Fluorouracil (5-FU) sensitivities of cell lines for uncovering the mechanisms of acquired 5-FU resistance of cancer cell lines. We use dataset from the "Sanger Genomics of Drug Sensitivity in Cancer (GDSC) dataset from the Cancer Genome Project (downloaded from "https://www.cancerrxgene.org/downloads/bulk download"), i.e., expression levels of $V = 17,737$ genes for 1018 cell lines and IC50 values of anti-cancer drugs. We focus on the 5-FU that is a chemotherapeutic drug commonly used for solid cancers[24]. The dataset is divided into four bins according to the value of IC50 of 5-FU, i.e., samples with IC50 of $(-\infty, -\sigma)$ (B1), $[-\sigma, 0)$ (B2), $[0, \sigma)$ (B3), $[\sigma, \infty)$ (B4), where σ is the standard deviation of IC50 values. The datasets B1, B2, B3, and B4 include $N_0 = 160$, $N_1 = 266$, $N_2 = 305$, and $N_3 = 160$ cell lines, respectively. We try $M = 20$ as the feasible number of latent factors for humans to analyze the dependencies between the factors, and extract the 20 latent factors by using the MGLASSO. We choose $\rho = 30$ so that the graph of dependencies between the latent factors is sparse enough but not completely independent of each other. As with the Stock Price dataset, it can be observed that the BIC shows better values compared to when using different ρ (Table 5). On the other hand, although the number of latent factors M does not always yield the best BIC (see Supplementary Information), from the perspective of interpretability, we provide an analysis using $M = 20$ in this case, similar to the approach taken with the Stock Price dataset. The model is trained by SGD with 20, 000 epochs. Note that we observe that the loss function seems almost completely converged at the 15, 000 epoch and few dependencies are added or removed after that. Additionally, the resulting $\boldsymbol{D}$, ε_k and $\boldsymbol{\Sigma}_k^{-1}$ are almost the same for three trials.

We perform the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to reveal biological pathways behind the latent factors. For the 100 genes having the largest scores of each latent factor, KEGG pathway analysis is performed. Table 7 shows the most enriched pathways of each latent factor (*i.e.*, each pathway shows the smallest p.value). That is, the latent factors can be interpreted as the corresponding biological pathways, and the 20 pathways in Table 7 may be crucial to understand molecular mechanisms of 5-FU resistance of cancer cell lines. In order to uncover molecular mechanism of the 5-FU resistance of cancer cell lines hidden behind the molecular interplays between thousands of genes, we construct the networks of the extracted 20 latent factors (i.e., pathway networks). Figure 7 shows the estimated pathway networks varying depending on 5-FU sensitivity. The variances of these factors in each bin are shown in Fig. 6.

We focus on the following interactions between the latent factors.

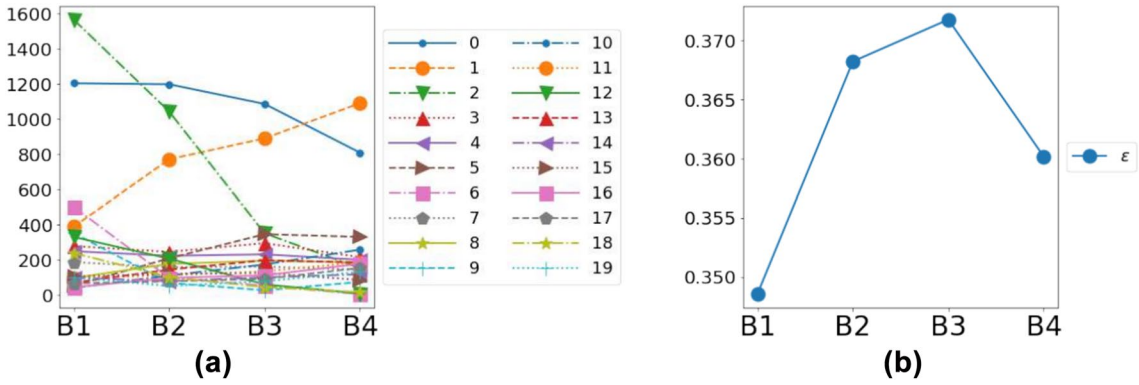

Figure 6. The variances of the latent factors estimated from gene expression data. (**A**) The variance of latent factors, i.e., the diagonal components of the covariance matrix $\boldsymbol{\Sigma}_k$ of latent factors, plotted annually. (**B**) The common variance ε_k of the remaining latent factors plotted annually.

Latent factor	The most enriched pathway
0	hsa04512:ECM-receptor interaction
1	hsa04510:Focal adhesion
2	hsa04145:Phagosome
3	hsa04915:Estrogen signaling pathway
4	hsa04110:Cell cycle
5	hsa04916:Melanogenesis
6	hsa05202:Transcriptional misregulation in cancer
7	hsa05202:Transcriptional misregulation in cancer
8	hsa05222:Small cell lung cancer
9	hsa05202:Transcriptional misregulation in cancer
10	hsa04610:Complement and coagulation cascades
11	hsa04974:Protein digestion and absorption
12	hsa04612:Antigen processing and presentation
13	hsa05146:Amoebiasis
14	hsa04110:Cell cycle
15	hsa04350:TGF-beta signaling pathway
16	hsa05204:Chemical carcinogenesis—DNA adducts
17	hsa03250:Viral life cycle—HIV-1
18	hsa04064:NF-kappa B signaling pathway
19	hsa05416:Viral myocarditis

Table 7. The most enriched pathways of 20 latent factors.

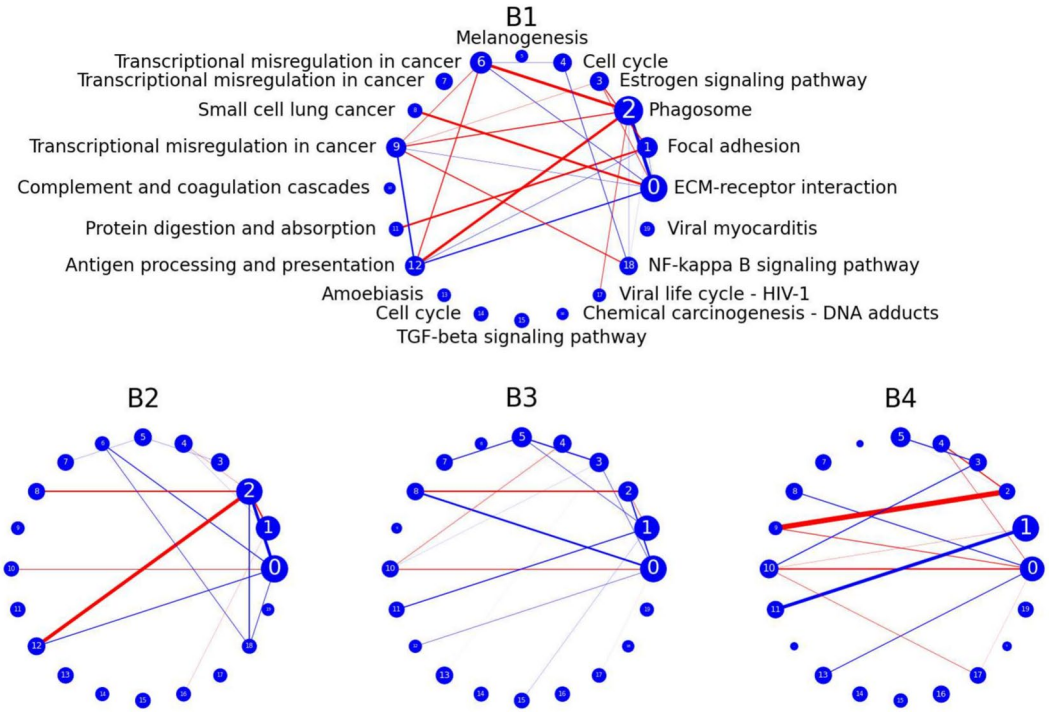

Figure 7. Networks of latent factors of expression levels of genes corresponding from 5-FU sensitive to resistance cell lines (i.e., B1 → B2 → B3 → B4), where each edge indicates the partial correlation, the element of the precision matrix divided by the square root of the corresponding diagonal element. Blue edges indicate the positive partial correlations and red edges indicate the negative. The size of the nodes indicates the variances of the latent factors, the diagonal elements of the covariance matrix.

- 0 (hsa04512: ECM-receptor interaction) and 2 (hsa04145: phagosome)
- 2 (hsa04145: phagosome) and 6 (hsa05202: transcriptional misregulation in cancer)
- 2 (hsa04145: phagosome) and 12 (hsa04612: antigen processing and presentation)
- 1 (hsa04510: focal adhesion) and 11 (hsa04974: protein digestion and absorption)

The positive interaction between the latent factors "0 (hsa04512: ECM-receptor interaction) and 2 (hsa04145: phagosome)" can be considered as 5-FU sensitive specific characteristic. The strong positive interaction becomes weaker from 5-FU-sensitive to non-sensitive cell lines, and the interaction disappeared in 5-FU resistant cell lines (i.e., B4). Furthermore, the interaction between the latent factors "2 (hsa04145: phagosome) and 6 (hsa05202: transcriptional misregulation in cancer)" and "2 (hsa04145: phagosome) and 12 (hsa04612: antigen processing and presentation)" are also considered as the 5-FU sensitive specific interplays. We also consider the interaction between the latent factors "1 (hsa04510: focal adhesion) and 11 (hsa04974: protein digestion and absorption)" as a crucial interplay for uncovering the mechanisms of 5-FU sensitivity in cancer cell lines, because the negative interaction in 5-FU sensitive cell lines (i.e., B1) disappeared in B2, and becomes the positive interaction in 5-FU resistance cell lines (B3 and B4).

In the gene regulatory network, the hub genes connected with many other genes are considered as crucial markers that play vital roles in biological process and gene regulation[25]. Thus, the hub-pathway can be also considered a crucial marker in the pathway networks. The hubness of the pathways "hsa04145: phagosome", "hsa05202: transcriptional misregulation in cancer" and "hsa04612: antigen processing and presentation" are considered as 5-FU sensitive specific characteristic, while the their hubnesses disappeared in 5-FU resistance cell lines. It implies that their activities are 5-FU sensitive specific characteristics.

Table 8 shows the enriched genes in the identified pathways, where the column "Evidences" indicates previous studies that recovered the pathways as key components to understand the mechanisms of gastric cancer and/or 5-FU.

- hsa04145: Phagosome
  Guo et al.[26] demonstrated that TPM4 expression is related to 5-FU drug sensitivity and TPM4 is a target to predict the 5-FU sensitivity in gastric cancer. The positively co-expressed genes with TPM4 are enriched in "hsa04145: Phagosome" and "hsa05202: Transcriptional misregulation in cancer". Differentially expression genes (DEGs) between fluorouracil-resistant and -sensitive gastric cell lines are enriched in the pathway "hsa04145: Phagosome"[28].
- hsa05202: Transcriptional misregulation in cancer
  This pathway was identified as one of biological process of DEGs between multi-drug resistance cell (*i.e.*, 5-FU, paclitaxel, mitomycin, vinorelbine tartrate, cisplatin and gemcitabine) and cisplatin resistance cell[29].
- hsa04612: Antigen processing and presentation
  The putative targets of DEGs between control and treatment with 5-FU involve in "hsa04612: Antigen processing and presentation"[27].

It can be seen through the Table 8 that the identified pathways have strong evidences as crucial pathways to understand the 5-FU sensitivity. It implies that the proposed MGLASSO provides biologically reliable results for identifying crucial pathways to understand the complex mechanisms of 5-FU resistance of cancer cell lines.

Figure 8 shows expression levels of the genes enriched in the 5-FU sensitive-specific pathways, where red and blue boxes indicate expression levels in 5-FU sensitive and resistance cell lines, respectively. The enriched genes in drug sensitive-specific pathways show relatively high expression levels in 5-FU sensitive cell lines compared with

KEGG pathway	Genes	Evidences
ECM-receptor interaction	SDC4, LAMB3, ITGA2,	26
	LAMA3, LAMC2, CD44	
Focal adhesion	COL1A2, CAV1, TNC,	26,27
	FN1, VEGFC, COL6A3, MYLK	
∗Phagosome	TUBB2B, NCF4, ITGB2,	26,28
	HLA-B, HLA-DRA, CYBA,	
	CORO1A, CTSS, HLA-DPA1	
∗Transcriptional misregulation in cancer	LYL1, CCNA1, HOXA9, HHEX,	29,30
	BCL11B, FLT3, LMO2, MPO, ELANE	
Protein digestion and absorption	COL17A1, COL16A1, COL1A2,	26
	COL4A2, COL5A1, COL7A1, COL6A3	
∗Antigen processing and presentation	CD74, HLA-DRA, IFI30,	27
	HSPA1A, HLA-DPA1	

Table 8. The enriched genes of the pathways and evidences related to 5-FU sensitivity, where ∗ indicates the 5-FU sensitive-specific pathways (i.e., their strong interactions and hubness are observed in the networks of 5-FU sensitive cell lines).

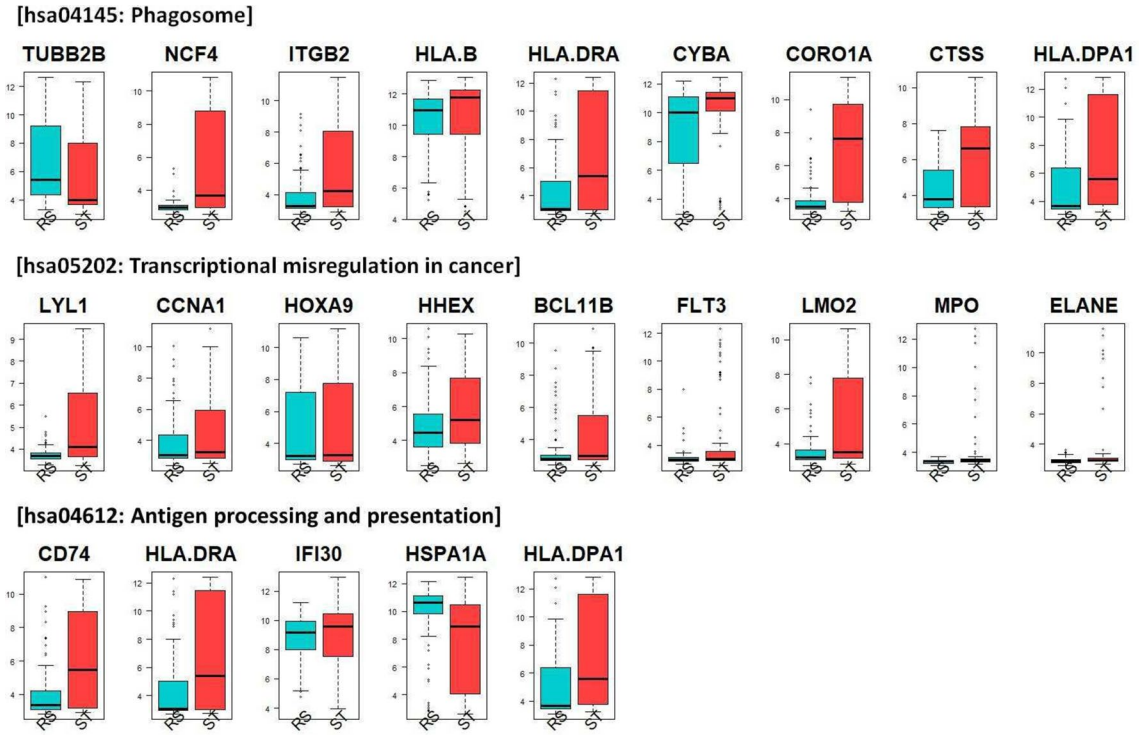

Figure 8. Expression levels of genes in 5-FU sensitive-specific pathways, where red and blue boxes indicate 5-FU sensitive (ST) and resistance (RS) cell lines, respectively.

resistance cell lines. Furthermore, high expression variances of the genes can be also observed in drug sensitive cell lines. In short, the enriched genes in 5-FU sensitive-specific pathways show high activities and diversity in drug sensitive cell lines. It implies that the identified pathways and enriched genes may have key information to uncover 5-FU sensitive/resistance mechanisms of cancer cell lines.

We suggest though our results and literature that catalyzing the identified 5-FU-specific sensitive pathways (hsa04145: Phagosome, hsa05202: Transcriptional misregulation in cancer, hsa04612: Antigen processing and presentation) and expressions of the enriched genes in the pathways may play crucial roles to prevent acquisition of 5-FU resistance of cancer cells.

To the best of our knowledge, this is the first computational study on biological pathways network for uncovering mechanisms of acquired drug resistance of cancer cell lines. We expect that our method will be a useful tool to reveal complex biological mechanisms hidden behind molecular interplays between giant number of genes.

Conclusion

Even when a multitude of variables appear to interact in complex and seemingly distinct ways across various situations, it is possible to comprehend the essential generative mechanisms of a complex system by identifying common mechanisms underlying these situations and interpreting each situation through the interactions among these common mechanisms. We proposed a novel method for extracting latent factors through linear projection, which is one of the primary methods for estimating generative mechanisms that are comprehensible to humans. Our approach assumes that the generative mechanisms of multiple datasets can be broadly explained by sparse dependencies among a few common latent factors. It models the data distribution as a Gaussian distribution, hypothesizing common latent factors across datasets, with dataset-specific sparse dependencies among these latent factors, learned through an algorithm that combines Graphical Lasso with gradient descent on Stiefel Manifolds. Experiments with synthetic data demonstrated that, whereas traditional latent factor extraction methods based on linear projection fail to estimate dependencies among latent factors, our proposed method can accurately do so. Although our assumption is simplistic and not capable of detailed explanations of complex systems, we showed with data from two vastly different domains-finance and healthcare-that this simple assumption could lead to surprisingly rich insights. In the financial domain, we discovered from the data of thousands of stock prices over several years that these prices were strongly influenced by notable societal trends during that period. In the healthcare domain, we demonstrated that gene expression data from cancer cells with varying degrees of drug resistance could provide comprehensive insights and new findings on the molecular mechanisms of drug resistance acquisition. However, the proposed method has several limitations. For instance, there is no guarantee of convergence to the optimal solution. Additionally, the assumption behind setting the same ρ for all datasets is that factors influencing sparsity estimation, such as the number of samples and the variance of each variable, are approximately equal across all datasets. Furthermore, like some previous works[12–14], the proposed method could potentially be improved by assuming similarity in the dependencies between variables across datasets. Addressing these issues remains a topic for future work. We are continuing efforts to enhance the

flexibility and scalability of our proposed method and expect that its application across a broad range of fields will open pathways from local analyses of relationships among a few variables to global analyses of interactions across all variables.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. The source code for the proposed algorithm is available at https://github.com/marucozy/mglasso.

Received: 7 April 2024; Accepted: 30 July 2024
Published online: 05 August 2024

References

1. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *ICLR* (2014).
2. Kipf, T. N. & Welling, M. Variational graph auto-encoders. *CoRR*. arXiv:1611.07308 (2016).
3. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. In *ICML*. 1278–1286 (2014).
4. Seninge, L., Anastopoulos, I., Ding, H. & Stuart, J. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nat. Commun.* **12**, 5684 (2021).
5. Qoku, A. & Buettner, F. Encoding domain knowledge in multi-view latent variable models: A bayesian approach with structured sparsity. In *AISTATS*. 11545–11562 (2023).
6. Greenacre, M. *et al.* Principal component analysis. *Nat. Rev. Methods Prim.* **2**, 100 (2022).
7. Hyvärinen, A. & Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **13**, 411–430 (2000).
8. He, Y., Qi, Y., Kavukcuoglu, K. & Park, H. Learning the dependency structure of latent factors. In *NIPS*. Vol. 25 (2012).
9. Celik, S. *et al.* Extracting a low-dimensional description of multiple gene expression datasets reveals a potential driver for tumor-associated stroma in ovarian cancer. *Genome Med.* **8**, 66 (2016).
10. Banerjee, O., Ghaoui, L. E., d'Aspremont, A. & Natsoulis, G. Convex optimization techniques for fitting sparse gaussian graphical models. In *ICML*. 89–96 (2006).
11. Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441 (2007).
12. Honorio, J. & Samaras, D. Multi-task learning of gaussian graphical models. In *ICML*. 447–454 (2010).
13. Hara, S. & Washio, T. Common substructure learning of multiple graphical gaussian models. In *ECML PKDD. Lecture Notes in Computer Science*. Vol. 6912. 1–16 (Springer, 2011).
14. Yang, S., Lu, Z., Shen, X., Wonka, P. & Ye, J. Fused multiple graphical lasso. *SIAM J. Optim.* **25**, 916–943 (2015).
15. Absil, P.-A., Mahony, R. & Sepulchre, R. *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, 2007).
16. Boyd, S. P., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2011).
17. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. In *ICLR* (2014).
18. Maruhashi, K. *et al.* Learning multi-way relations via tensor decomposition with neural networks. In *AAAI*. 3770–3777 (2018).
19. Golub, G. H. & Van Loan, C. F. *Matrix Computations* 4 edn. (Johns Hopkins University Press, 2013).
20. Lanczos, C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bureau Standard* **45**, 255–282 (1950).
21. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
22. Morrison, W. M. & Tang, R. China's rare earth industry and export regime: Economic and trade implications for the United States. In *CRS Report for Congress R42510, Congressional Research Service* (2012).
23. Gilje, E., Ready, R. & Roussanov, N. Fracking, drilling, and asset pricing: Estimating the economic benefits of the shale revolution. In*National Bureau of Economic Research Working Paper No. 22914* (2016).
24. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat. Commun.* **10**, 4571 (2019).
25. Yu, D., Lim, J., Wang, X., Liang, F. & Xiao, G. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinform.* **18**, 186 (2017).
26. Guo, Q. *et al.* Integrated pan-cancer analysis and experimental verification of the roles of tropomyosin 4 in gastric cancer. *Front. Immunol.* **14** (2023).
27. Shah, M. Y., Pan, X., Fix, L. N., Farwell, M. A. & Zhang, B. 5-fluorouracil drug alters the microRNA expression profiles in mcf-7 breast cancer cells. *J. Cell. Physiol.* **226**, 1868–1878 (2011).
28. Pan, J., Dai, Q., Xiang, Z., Liu, B. & Li, C. Three biomarkers predict gastric cancer patients' susceptibility to fluorouracil-based chemotherapy. *J. Cancer* **10**, 2953–2960 (2019).
29. Huang, H. *et al.* Identification and validation of NOLC1 as a potential target for enhancing sensitivity in multidrug resistant non-small cell lung cancer cells. *Cell. Mol. Biol. Lett.* **23**, 54 (2018).
30. Chen, W.-X. *et al.* Analysis of miRNA signature differentially expressed in exosomes from adriamycin-resistant and parental human breast cancer cells. *Biosci. Rep.* **38**, BSR20181090 (2018).

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 22H03692.

Author contributions

K.M. developed the methodology, conducted experiments, and wrote the overall manuscript. H.K. supervised regarding methodology. S.M. supervised the study. H.P. analyzed the biological results and wrote the related manuscript. All authors have read and approved the final version of the manuscript.

Competing interests

Dr. Maruhashi is an employee of Fujitsu Limited. Other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68959-7.

Correspondence and requests for materials should be addressed to K.M. or H.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.