Large Scale Partial Correlation Screening with Uncertainty Quantification

Emily Neo, Peter Radchenko, Bala Rajaratnam
University of Sydney

Abstract

Identifying multivariate dependencies in high-dimensional data is an important problem in large-scale inference. This problem has motivated recent advances in mining (partial) correlations, which focus on the challenging ultra-high dimensional setting where the sample size, n, is fixed, while the number of features, p, grows without bound. The state-of-the-art method for partial correlation screening can lead to undesirable results. This paper introduces a novel principled framework for partial correlation screening with error control (PARSEC), which leverages the connection between partial correlations and regression coefficients. We establish the inferential properties of PARSEC when n is fixed and p grows super-exponentially. First, we provide "fixed-n-large-p" asymptotic expressions for the familywise error rate (FWER) and k-FWER. Equally importantly, our analysis leads to a novel discovery which permits the calculation of exact marginal p-values for controlling the false discovery rate (FDR), and also the positive FDR (pFDR). To our knowledge, no other competing approach in the "fixed-n-large-p" setting allows for error control across the spectrum of multiple hypothesis testing metrics. We establish the computational complexity of PARSEC and rigorously demonstrate its scalability to the large p setting. The theory and methods are successfully validated on simulated and real data, and PARSEC is shown to outperform the current state-of-the-art.

Keywords: ultra-high dimensional, multiple hypothesis testing, sample-starved, covariance

1 Introduction

Modern massive datasets have emerged out of our improved ability to collect, store and analyze data. Extracting and identifying complex associations and dependencies in such data, while addressing the unique challenges inherent to the increasing number of features, is a critically important task in statistical inference and machine learning. In the modern

ultra-high dimensional setting where the sample size, n, is fixed and the number of features, p, tends to infinity, spurious associations will be wrongly identified due to random chance alone. Indeed, it is well understood that an alarming proportion of scientific results reported in leading journals are not reproducible (Ioannidis 2005). Hence, there is a compelling need for a principled framework to identify the strongest and most statistically significant associations in the ultra-high dimensional setting, while providing theoretical safeguards against identifying spurious associations. The resulting significant associations can subsequently be represented as a correlation or partial correlation graph.

For a given set of features, sample correlation coefficients are a basic measure of bivariate or marginal (linear) dependency. Alternatively, partial correlation coefficients provide a multivariate measure of dependency, and have found widespread use. Notably, in the Gaussian setting, zero partial correlation coefficients imply conditional independence. In this setting, the partial correlation matrix is typically estimated using the inverse of the sample covariance matrix (Anderson 2003; Dempster 1972). However, this approach is problematic in high dimensions, as the inverse is not well-defined when $p > n$.

A number of methods for estimating sparse partial correlation matrices have been proposed for the high-dimensional asymptotic regime where both p and n tend to infinity. The corresponding approaches are often based on the ℓ_1-penalized likelihood framework, which adds an ℓ_1 penalty to either the Gaussian likelihood or some pseudo-likelihood - see, for example, Banerjee et al. (2008), Khare et al. (2015), and the references therein. These ℓ_1-based approaches are appropriate when both n and p are large and provide a framework for model selection by shrinking small partial correlations to zero. However, the focus of ℓ_1 approaches is on variable selection rather than hypothesis testing. Moreover, they also require the determination of the tuning/penalty parameter. Such ℓ_1-based approaches are also computationally expensive in large p data regimes due to the predominant use of iterative optimization algorithms. Furthermore, theoretical safeguards for the ℓ_1-penalized methods are established in the setting when both n and p go to infinity. For these reasons, the ℓ_1-penalized framework is not always amenable to the modern ultra-high dimensional setting where n is fixed and p goes to infinity - from both a statistical and a computational perspective.

Limited order partial correlation methods provide an alternative approach for estimating partial correlation matrices. These methods first employ preliminary statistical testing on bivariate marginal correlations or "q-order" partial correlations (where $q < p-2$) to reduce dimensionality of the conditioning set – see, for example, the useful works of Magwene and Kim (2004), Castelo and Roverato (2009) and references therein. Liang et al. (2015) notes that these methods do not evaluate full-order partial correlation coefficients, and as such they may result in estimates that are closer to marginal correlations. We note that the pre-processing step of other methods seem to also have the disadvantage that it

does not readily facilitate *statistical inference*. Hero and Rajaratnam (2012) developed an ultra-high dimensional approach with uncertainty quantification for the purpose of screening variables which are highly partially correlated with many others. Their approach is based on quantifying the distribution of the number of exceedances of the generalized inverse of the sample correlation matrix in the ultra-high dimensional setting where n is fixed and $p \to \infty$. While the emphasis of their approach is on screening highly connected variables (and not parameters), a natural adaptation of their approach, which we refer to as PCS-Hub, can be used for screening edges in the corresponding partial correlation graph. However, our empirical analysis demonstrates that PCS-Hub can be deficient in ultra-high dimensional settings. To our knowledge, this important problem has remained largely unsolved for more than 10 years.

In this paper, we introduce a principled framework for PARtial correlation Screening with Error Control (PARSEC), which moves beyond the model selection framework at the center of most existing high-dimensional partial correlation learning methods. PARSEC provides stable estimates of partial correlation coefficients by breaking down the partial correlation estimation problem into a series of simpler regression problems. PARSEC is inferentially sound, easy to use, computationally efficient and thus can be immediately deployed.

In Section 2 we introduce the specifics of the PARSEC method, and in Section 3 we establish theoretical properties of PARSEC and use these for inference in the ultra-high dimensional setting, thus going beyond model selection. In Section 4 we analyze the computational complexity of our approach, demonstrating its scalability. Section 5 provides extensive simulations to illustrate PARSEC's computational and inferential performance. Lastly, the efficacy of PARSEC is demonstrated in Section 6 on real applications, including breast cancer gene screening and a portfolio selection problem.

2 Methodology

In this section, we introduce PARSEC, a novel scalable method for partial correlation screening with uncertainty quantification in ultra-high dimensional data regimes. Section 2.1 provides preliminary information and notation, and in Section 2.2 we motivate and describe the PARSEC method in detail.

2.1 Preliminaries

We let $\mathbf{X} = [X_1, ..., X_p]^\top \in \mathbb{R}^p$ be a vector of random variables with mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}_{p\times p}$, and let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. We define the corresponding (marginal) correlations as $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, and partial correlations as $\rho^{ij} = -\omega_{ij}/\sqrt{\omega_{ii}\omega_{jj}}$. Assuming a sample size of n, let $\mathbb{X}$ denote the corresponding $n \times p$ data matrix, where column vector $\mathbf{X}_j$ and

row vector $\mathbf{X}_{(i)}$ represent the j-th column and the i-row, respectively: $\mathbb{X} = [\mathbf{X}_1, ..., \mathbf{X}_p] = [\mathbf{X}_{(1)}^\top, ..., \mathbf{X}_{(p)}^\top]^\top$. We define the sample mean of the j-th column as $\bar{X}_j = n^{-1}\sum_{i=1}^n X_{ij}$, write $\bar{\mathbf{X}} = [\bar{X}_1, ..., \bar{X}_p]$ for the vector of sample means, and define the sample covariance matrix as $\mathbf{S} = \sum_{i=1}^n (\mathbf{X}_{(i)} - \bar{\mathbf{X}})^\top (\mathbf{X}_{(i)} - \bar{\mathbf{X}})/(n-1)$. Now let $\mathbf{1}$ denote a column vector consisting of 1s. We calculate multivariate Z-scores by standardizing the columns of the data matrix: $\mathbf{Z}_j = (\mathbf{X}_j - \bar{\mathbf{X}}_j\mathbf{1})/\sqrt{\mathbf{S}_{jj}(n-1)}$, which permits the following representation of the sample correlation matrix: $\mathbf{R} = \mathbb{Z}^\top\mathbb{Z}$. We say that a random matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ has a *vector-elliptical* distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite covariance parameter $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ if its density can be expressed as follows (Anderson 2003):

$$f_{\mathbb{X}}(\mathbb{X}) = \det(\boldsymbol{\Sigma})^{-n/2} g(\mathrm{tr}((\mathbb{X} - \boldsymbol{\mu}\mathbf{1}^\top)\boldsymbol{\Sigma}^{-1}(\mathbb{X} - \boldsymbol{\mu}\mathbf{1}^\top)^\top)), \tag{1}$$

where the shape function $g : \mathbb{R} \to [0, \infty)$ is such that $\int f_{\mathbb{X}}(\mathbb{X})\, d\mathbb{X} = 1$.

In our theoretical analysis of partial correlation screening in ultra-high dimensions, we will exploit the so-called U-score representation: $\mathbf{R} = \mathbb{Z}^\top\mathbb{Z} = \mathbb{U}^\top\mathbb{U}$. Here, $\mathbb{U}$ is an $n-1$ by p matrix whose j-th column corresponds to the j-th feature. In particular, each feature is represented by a U-score, and all U-scores lie on the unit sphere, S_{n-2}, providing a lower-dimensional geometric representation of high-dimensional data and the dependencies therein. To achieve mean centering, U-scores project away components of $\mathbf{X}_j$ that are orthogonal to the $n-1$ dimensional hyperplane $\{\mathbf{u} \in \mathbb{R}^n : \mathbf{1}^\top\mathbf{u} = 0\}$. This mean centering changes the effective sample size from n to $n-1$. Furthermore, U-scores that are close to each other correspond to highly positively correlated features. Importantly, when the distribution of $\mathbb{X}$ is vector-elliptical and $\boldsymbol{\Sigma}$ is diagonal, the U-scores are uniformly distributed on the unit sphere. A deviation from uniformity in the distribution of the U-scores reveals the existence of dependency. U-scores have the desirable property of preserving correlations between features (as they are constructed through normalization and centering using an orthogonal matrix as described below). More formally, let $\mathbb{T}_{n\times n}$ be an orthogonal matrix of the form $[n^{-1/2}\mathbf{1}, \mathbb{T}_{2:n}]$. Here, $\mathbb{T}_{2:n}$ can be any orthogonal matrix whose columns are orthogonal to $\mathbf{1}$; in practice, we construct $\mathbb{T}_{2:n}$ using Gramm-Schmidt orthogonalization. The U-score matrix, $\mathbb{U}_{(n-1)\times p} = [\mathbf{U}_1, ..., \mathbf{U}_p]$, can then be obtained from the relation $\mathbb{U} = \mathbb{T}_{2:n}^\top\mathbb{Z}$. The sample correlation between $\mathbf{X}_j$ and $\mathbf{X}_k$, r_{jk}, thus has a simple relationship with the Euclidean distance between the corresponding U-scores:

$$r_{jk} = \mathbf{U}_j^\top\mathbf{U}_k = 1 - ||\mathbf{U}_j - \mathbf{U}_k||_2^2/2 \tag{2}$$

Figure 1 illustrates the distribution of U-scores for different covariance structures when $n = 4$ and $p = 500$. In the diagonal covariance setting, U-scores are uniformly distributed on the unit sphere. In contrast, dependent U-scores corresponding to block-diagonal covariance settings notably coalesce into clusters.

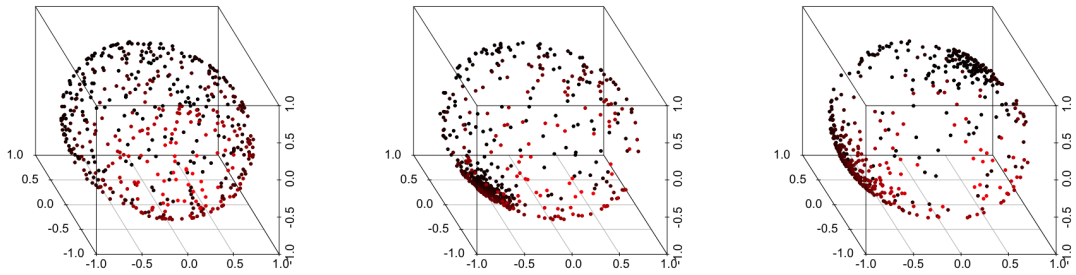


Figure 1: In the Gaussian setting when $p = 500$ and $n = 4$: U-scores associated with a diagonal covariance matrix (left); block-diagonal covariance with one block of size 200 (center); block-diagonal covariance with two blocks of size 200 and 100 (right). In the diagonal covariance setting, the U-scores are uniformly distributed on the unit sphere. However, in the block-diagonal settings, clusters of U-scores are evident.

Hero and Rajaratnam (2011) leveraged the U-score representation above to understand the asymptotic behavior of sample correlation coefficients in sample-deficient settings and screen for significant correlations. They established that in the ultra-high dimensional asymptotic regime where n is fixed and p goes to infinity, the number of false discoveries is approximated by a Poisson distribution. This approximation holds for a wide class of vector-elliptical distributions and permits the derivation of the critical screening level ρ that can control the family-wise error rate (FWER) at a pre-specified significance level. Hero and Rajaratnam (2012) extended the correlation screening framework for discovering highly connected vertices in partial correlation networks (which we term as "PCS-Hub"). They estimated the partial correlation matrix using the generalized inverse of the sample correlation matrix, $\mathbf{R}^{\dagger}$, by leveraging the following representation:

$$\mathbf{R}^{\dagger} = \mathbb{U}^{\top}[\mathbb{U}\mathbb{U}^{\top}]^{-2}\mathbb{U}. \tag{3}$$

Their resulting estimate of the partial correlation matrix is the standardized quantity $\mathbf{P} = \mathbf{D}_{\mathbf{R}^{\dagger}}^{-1/2}\mathbf{R}^{\dagger}\mathbf{D}_{\mathbf{R}^{\dagger}}^{-1/2}$, where $\mathbf{D}_{\mathbf{A}}$ denotes the diagonal matrix corresponding to the matrix $\mathbf{A}$. Equivalently,

$$\mathbf{P} = \mathbb{Y}^{\top}\mathbb{Y}, \qquad \text{where} \quad \mathbb{Y}_{(n-1)\times p} = [\mathbb{U}\mathbb{U}^{\top}]^{-1}\mathbb{U}\mathbf{D}_{\mathbb{U}^{\top}[\mathbb{U}\mathbb{U}^{\top}]^{-2}\mathbb{U}}^{-1/2}, \tag{4}$$

where the Y-scores provide an analog of the U-score representation for the partial correlation setting. The columns of $\mathbb{Y}$ are the so-called partial correlation Z-scores, which also lie on the unit sphere $\mathbf{S}_{n-2}$. This representation has been used to perform partial correlation screening with statistical error control - see Hero and Rajaratnam (2012). Despite these advances, the resulting PCS-Hub approach has some shortcomings: i) it can produce par-

tial correlation estimates that are similar to marginal correlation estimates (see Figure 2), and ii) it can fail to pick up a collection of strong signals (see Figure 3). Moreover, it does not also readily specify the distribution of partial correlations in the fixed n, fixed p setting. Hence, it is not clear how to use PCS-Hub to obtain exact marginal p-values for partial correlations, thus preventing a meaningful False Discovery Rate (FDR) and/or positive FDR (pFDR) analysis. These fundamental shortcomings have remained unresolved for the better part of the last 10 to 15 years, and to address these, we introduce the PARSEC method below.

2.2 The PARSEC method

The primary goal of the PARSEC method is to extract and "screen" high partial correlations in the noisy ultra-high dimensional setting. More specifically, PARSEC constructs stable estimates of scaled partial correlation coefficients, and quantifies the distribution of the resulting partial correlation estimates and the distribution of the number of exceedances in order to achieve a desired level of error control for a broad spectrum of error metrics (FWER, k-FWER, FDR, pFDR) - even in the challenging modern ultra-high dimensional setting when the sample size n is fixed and the dimension p tends to infinity.

As mentioned in the Introduction, the distinct advantage of the PCS-Hub method (Hero and Rajaratnam 2012) is that it quantifies the behavior of the distribution of partial correlations in the ultra-high dimensional regime (unlike ℓ_1-based methods), thus enabling uncertainty quantification even in this challenging setting. We show below that despite this important advantage, PCS-Hub can however be highly deficient in the modern fixed-n-large-p setting. Despite this large gap in the literature, to our knowledge, methods which provide more stable partial correlation estimates in the fixed-n-large-p setting, while achieving rigorous statistical error control, have not been proposed. This begs the question of whether there are alternative methods to stably estimate partial correlation coefficients in the ultra-high dimensional setting, which are still sufficiently tractable so that the distribution of partial correlation estimates can be precisely quantified.

In this paper, we note that one strategic and deliberate approach to circumventing the direct estimation of the inverse correlation matrix is to recognize and exploit the important relationship between partial correlation coefficients and regression coefficients. Recall that regression coefficients can be recast as scaled partial correlation coefficients. This connection at the population level provides an alternative approach to partial correlation estimation (at the sample level). We show in this paper that decoupling the partial correlation estimation problem into a series of regressions adequately addresses the deficiencies of directly estimating the inverse correlation matrix. Simultaneously, it uniquely allows for the precise quantification of the distribution of the partial correlation estimates in the highly challenging fixed-n-large-p setting - resulting in a novel and principled framework

for partial correlation screening with uncertainty quantification.

We now describe the proposed approach: the PARSEC method estimates matrix $\mathbf{H}$ of scaled partial correlations row by row, regressing $\mathbf{X}_j$ on the rest of the features for $j = 1, ..., p$. In fact, PARSEC performs this regression at the level of the U-scores instead of the original features, $\mathbf{X}_j$. More specifically, we treat the j-th U-score, $\mathbf{U}_j$, as the response variable, and the rest of the U-scores as predictors for $j = 1, ..., p$. Leveraging the U-score representation in (3) and denoting by $\mathbb{U}^{-j}$ the U-score matrix with the j-th column excluded, we derive the following expression for the vector of regression coefficient estimates: $((\mathbb{U}^{-j})^\top \mathbb{U}^{-j})^\dagger (\mathbb{U}^{-j})^\top \mathbf{U}_j$, which is the analogue of the classical regression formula $(\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbf{Y}$ at the level of U-scores. Our partial correlation estimates are then obtained by rescaling the estimated regression coefficients. More formally, we (a) compute the matrix of partial correlation Z-scores (cf. equation 4) that excludes the j-th feature:

$$\tilde{\mathbb{U}}^{-j} := (\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top)^{-1} \mathbb{U}^{-j} \mathbf{D}^{-\frac{1}{2}}_{\mathbb{U}^{-j\top}[\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top]^{-2}\mathbb{U}^{-j}};$$

and (b) define the j-th row of the PARSEC $\mathbb{H}$ matrix as follows:

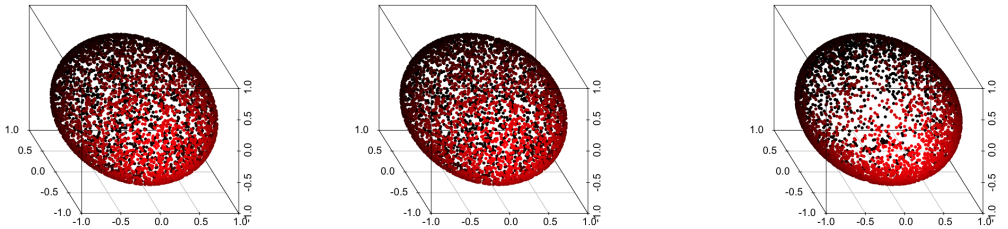
$$\big(H_{j1}, \ldots, H_{j(j-1)}, H_{j(j+1)}, \ldots, H_{jp}\big)^\top := (\tilde{\mathbb{U}}^{-j})^\top \mathbf{U}_j. \tag{5}$$

Observe that equation (5) is a vector containing inner products and parallels the correlation formula in (2), $r_{jk} = \mathbf{U}_j^\top \mathbf{U}_k$, and the PCS-Hub-based partial correlation formula in (4), $p_{jk} = \mathbf{Y}_j^\top \mathbf{Y}_k$. To construct PARSEC's scaled partial correlation matrix $\mathbf{H}$, we implement (5) for $j = 1, ..., p$. We note that matrix $\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top$, which is inverted in the definition of $\tilde{\mathbb{U}}^{-j}$, is $(n-1) \times (n-1)$ and hence low-dimensional in our setting of interest. Thus, the PARSEC approach is readily scalable and is ideally suited to the sample-starved regime when $p \gg n$.

The second component of the PARSEC framework provides a screening method which provides rigorous error control for detecting high partial correlations. Given a screening level ρ, PARSEC deems an estimate H_{jk} of the scaled partial correlation between features X_j and X_k a (screening) discovery if $|H_{jk}| \geq \rho$. This screening level, ρ, is obtained in a principled manner, using PARSEC's inferential properties. Specifically, in Section 3, we derive a Poisson approximation for the number of discoveries and also expressions for the mean number of discoveries in the ultra-high dimensional regime when n is fixed and the dimension p goes to infinity. These expressions allow for the specification of various multiple hypothesis testing measures for statistical error control, such as FWER, k-FWER (Romano and Wolf 2005), FDR and pFDR as a function of the screening level, ρ. These can then be used to identify the specific screening level, ρ, which leads to the desired level of error control.

Through empirical analysis, it can be easily seen that one of the major shortcom-

(a) U-scores, p=5000 (b) Y-scores, p=5000 (c) PARSEC scores, p=5000



(d) U-scores, p=10,000 (e) Y-scores, p=10,000 (f) PARSEC scores, p=10,000

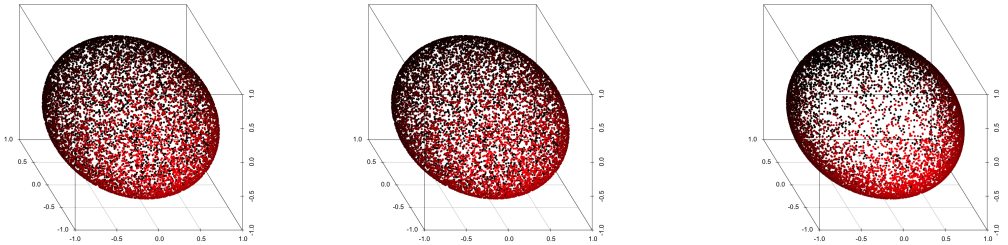


Figure 2: Geometric comparison of Z-type scores used in ultra-high dimensional screening methods for an AR(10) block structure with Gaussian features, block size = 50, $n = 4$ and varying p. U-scores are used in large-scale marginal correlation screening and Y-scores are used in partial correlation screening (PCS-Hub).

ings of the PCS-Hub approach is that it can yield partial correlation estimates that are very close to marginal correlation estimates in the high-dimensional sample-starved setting. Consequently, we find that PCS-Hub does not always produce reliable estimates when partial and marginal correlations are different. More specifically, the PCS-Hub approach of Hero and Rajaratnam (2012) uses the Moore-Penrose approach to estimate partial correlations, whereas the PARSEC regression approach in addition also uses the sample marginal correlations, which tend to be more stable (as they do not derive from inverses in high dimensions and can simply be calculated as inner products between features). One of the motivations for the newly proposed PARSEC method is to leverage more stable sample quantities to estimate partial correlations in the challenging ultra-high dimensional setting.

We now provide preliminary evidence that PARSEC can do better at identifying partial correlations than the PCS-Hub approach. Similar to the U-scores and Y-scores discussed in Section 2.1 for marginal correlation screening and PCS-Hub-based partial correlation screening, the analogous PARSEC scores can be constructed via the corresponding decomposition of the symmetric matrix $\tilde{\mathbf{H}}$, where $\tilde{\mathbf{H}}$ is obtained by transposing the upper triangle of PARSEC's $\mathbf{H}$ matrix. The closeness of the marginal correlation U-scores and partial correlation Y-scores, as seen in Figure 2, highlights the possible challenges faced

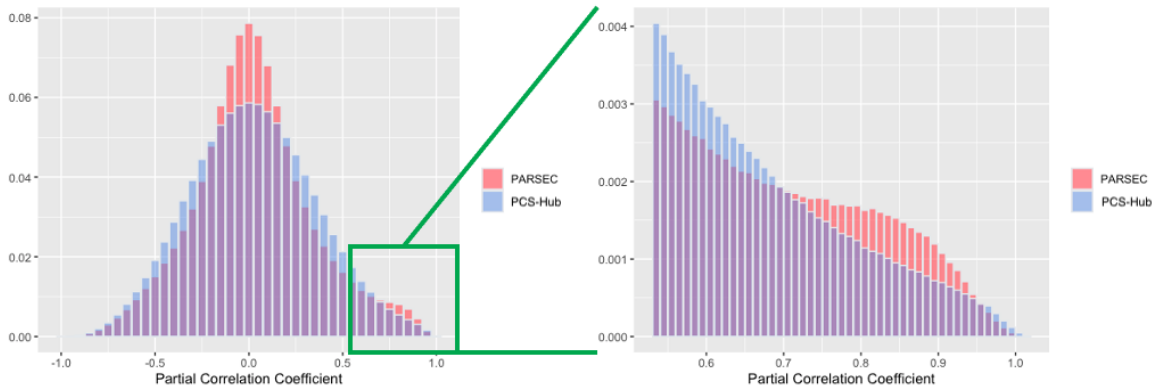


Figure 3: A comparison of the distribution of estimated partial correlation coefficients from PARSEC to the PCS-Hub method (Hero and Rajaratnam 2012) in an AR(10) setting. We simulate 1000 replications with Gaussian features where $n = 10$, $p = 100$, block size=20 and $\phi_1 = 0.8$ (where ϕ_1 is the coefficient on the first lag). Notably, the distribution of PARSEC's partial correlation coefficients is bi-modal. The mode located closer to 1 corresponds to the true AR(10) model coefficients. We magnify this area of the plot in the right figure. In addition to illustrating PARSEC's statistical power in identifying true signal, the left plot also illustrates PARSEC's effectiveness in controlling Type I error, as PARSEC's estimates of the null coefficients are shrunk closer to zero. In contrast, the distribution of PCS-Hub partial correlation coefficients is nearly symmetric around zero.

by the PCS-Hub-based Y-scores in instances when partial correlations are relatively high but the corresponding marginal correlations are low. In contrast, PARSEC scores are more distinguishable from marginal correlation U-scores. To understand whether these differences lead to better inference, we compare the distribution of partial correlation estimates stemming from PCS-Hub and PARSEC, in a setting where the true covariance model has high partial correlations - see Figure 3. In contrast to the PCS-Hub-based approach, a bi-modal distribution for PARSEC estimates is immediately identifiable, with a peak closer to 1 corresponding to the large nonzero partial correlations in the underlying model - illustrating PARSEC's improved Type II error control. The second peak closer to zero demonstrates that PARSEC also better shrinks the null coefficients to zero, illustrating PARSEC's improved Type I error control.

3 Theoretical properties of PARSEC

In this section we provide the theoretical properties of the PARSEC method in the ultra-high dimensional regime where the sample size n is fixed while the number of features p tends to infinity. In particular, we establish the limiting behavior of the number of PARSEC discoveries, which in turn allows us to derive fixed-n-large-p expressions of the corresponding family-wise error rate (FWER) and k-family-wise error rate (k-FWER). Equally

importantly, we also derive the exact marginal p-values for the PARSEC estimates of scaled partial correlation coefficients, which consequently allows us to achieve FDR and pFDR control.

3.1 Preliminaries

As our method competes with that of Hero and Rajaratnam (2012), for comparison purposes we adopt similar notation. Let N_{ρ_p} denote the number of PARSEC discoveries corresponding to the screening level $\rho_p \in (0,1)$, i.e., N_{ρ_p} is the number of entries located above the main diagonal in the PARSEC estimate of the scaled partial correlation matrix that exceed ρ_p in magnitude. We define the spherical cap probability,

$$P_0 = P_0(\rho_p, n) = a_n \int_{\rho_p}^{1} (1-u^2)^{(n-4)/2} du, \qquad \text{where} \qquad \mathrm{a}_n = \frac{2\Gamma([n-1]/2)}{\sqrt{\pi}\Gamma([n-2]/2)}, \tag{6}$$

and let $\|\Delta_{p,q}\|_1$ be the average weak dependency coefficient, defined in Hero and Rajaratnam (2011, equation (A.13)). We write $J\big(\overline{f_{\mathbf{U}_\bullet, \tilde{\mathbf{U}}^\bullet_{*-\bullet}}}\big)$ for the normalized integral of the average pairwise density involving a partial and a regular U-score, as formally defined in Appendix A.1 of this paper. For the remainder of this section, bounds $o(\cdot)$ and $O(\cdot)$, as well as their stochastic counterparts, are understood to hold in the setting where n is fixed and p tends to infinity. In the results that follow, we invoke either or both of the following commonly used assumptions from the literature.

A1. The random matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ has a vector-elliptical distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, covariance parameter $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and a differentiable density function that is uniformly bounded in p.

A2. The elements of $\mathbb{U}\mathbb{U}^\top - E\big(\mathbb{U}\mathbb{U}^\top\big)$ are of order $O_p(s_p)$, where $s_p = o(p)$.

As a concrete example, assumption A2 holds with $s_p = p^{1/2} + q_p$ - see Firouzi et al. (2017) - if the correlation matrix $\boldsymbol{\Omega}$ corresponding to the covariance matrix $\boldsymbol{\Sigma}$ is of the form $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2$, where $\boldsymbol{\Omega}_1$ is a block-sparse matrix of degree $q_p = o(p)$, and $\boldsymbol{\Omega}_2 = (\omega_{j,k})$ is such that $\omega_{ij} = O\big(f(|j-k|)\big)$ for some function f that satisfies $\lim_{t\to\infty} f(t) = 0$.

3.2 Theoretical Results

In this section, we establish a sequence of results on the limiting distribution of the number of PARSEC discoveries, N_{ρ_p}, as p tends to infinity but the sample size n remains fixed. These results allow us to undertake statistical inference in the ultra-high dimensional setting. Our first three theorems provide Poisson approximations for the distribution of the number of discoveries and asymptotic expressions for the corresponding expected value. The assumptions of these theorems get progressively stronger for reasons specified below. Finally, in the fourth result, we derive exact marginal p-values for PARSEC estimates of

scaled partial correlation coefficients. All the proofs are provided in Appendix A.

Result 1. The following theorem establishes a general result on the limiting behavior of the number of PARSEC discoveries. Employing assumption A1 leads to an approximation for the expected number of discoveries, and additionally assuming A2 leads to the Poisson approximation given below.

Theorem 1. *Let $\eta_p = p(p-1)P_0/2$ and $\delta_p = (s_p+1)/p$. If assumption A1 holds, then the expected number of discoveries is given as follows:*

$$E[N_{\rho_p}] = \eta_p J\big(\overline{f_{\mathbf{U}_\bullet, \tilde{\mathbf{U}}^\bullet_{*-\bullet}}}\big)] + O\Big(\eta_p \sqrt{1-\rho_p}\Big).$$

Now, let N_p^* denote a Poisson random variable with rate $E[N_p^*] = \eta_p J\big(\overline{f_{\mathbf{U}_\bullet, \tilde{\mathbf{U}}^\bullet_{*-\bullet}}}\big)$. If assumption A2 also holds, $(p-1)P_0 \leq 1$, $\delta_p = o\big(1-\rho_p\big)$ and

$$\eta_p \Big[\eta_p (l_p/p)^2 + \|\Delta_{p,l_p}\|_1 + \big[\sqrt{1-\rho_p} + \delta_p (1-\rho_p)^{-1}\big]\big(1 + E[N_p^*]\big)^{-1/2}\Big] = o(1) \tag{7}$$

for some arbitrary $l_p \in [p]$, which is allowed to depend on p, then

$$\sup_{k \in \mathbb{N}} \Big| P(N_{\rho_p} > k) - P(N_p^* > k) \Big| = o(1).$$

Remark 1. *We note that our approximation for the probability of more than k discoveries is uniform in k, while the corresponding approximation in Hero and Rajaratnam (2012) is stated only for $k = 0$. The uniformity of our approximation allows us to use it for k that is a function of p, for example, $k = p^{1/2}$.*

Remark 2. *Note that the term $\eta_p[\sqrt{1-\rho_p} + \delta_p(1-\rho_p)^{-1}]\big(1 + E[N_p^*]\big)^{-1/2}$ appears in equation (7), as compared to $p^2 P_0[p^{-1} + (1-\rho_p)^{1/2}]$ used in Hero and Rajaratnam (2012). If $\delta_p = O([1-\rho_p]^{3/2})$ and $E[N_p^*] \to \infty$, then our condition for obtaining the Poisson approximation is weaker.*

Remark 3. *Note that Theorem 1 does not necessarily require sparsity.*

Result 2. We now consider the case where $\mathbf{\Sigma}$ is a block-sparse matrix of degree q_p. Imposing block-sparsity yields a better approximation for the Poisson rate parameter and leads to a tractable and useful representation in the modern ultra-high dimensional setting.

Theorem 2. *Let $\eta_p = p(p-1)P_0/2$. If assumption A1 holds and $\mathbf{\Sigma}$ is block-sparse of degree q_p, then*

$$E[N_{\rho_p}] = \eta_p \left[1 + O\Big(\frac{q_p}{p}\Big)\right].$$

Now, let N_p^* denote a Poisson random variable with rate $E[N_p^*] = \eta_p$. If $(p-1)P_0 \leq 1$,

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2025.3606207

(a) $n = 10$, varying p (b) $p = 1000$, varying n

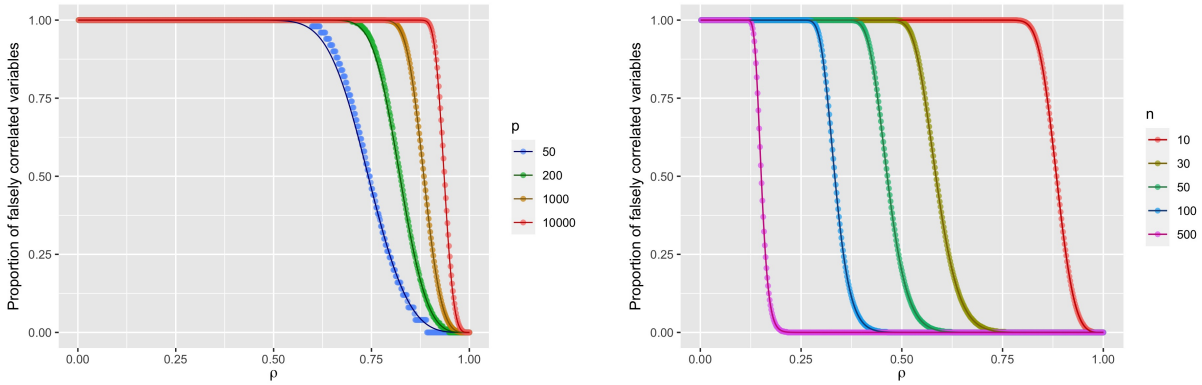


Figure 4: Illustration of the proportion of features in $\mathbb{X}$ with at least one partial correlation falsely identified by PARSEC under the null hypothesis of diagonal covariance at screening level ρ. In panel (a), $n = 10$ and p is varied; in panel (b), $p = 1000$ and n is varied. We compare the theoretical approximations (displayed as solid curves) with the corresponding median empirical proportions (displayed as circles) over 1000 numerical replications. It is clear that the large p theoretical expressions mirror the true empirical proportion of discovered features, even when the dimension p is small.

$n > 2$, $q_p = o(p)$ *and* $\eta_p^2\left(\frac{q_p}{p}\right)^2 + \sqrt{\eta_p}\left[\frac{1}{\sqrt{p}} + \frac{q_p}{p}\right](1-\rho_p)^{-1} = o(1)$, *then*

$$\sup_{k\in\mathbb{N}}\left|P(N_{\rho_p} > k) - P(N_p^* > k)\right| = o(1), \quad \textit{and} \tag{8}$$

$$\sup_{k\in\mathbb{N}}\left|P(N_{\rho_p} > k) - P(N_p^* > k)\right| = O\Big(\frac{1}{p} + \eta_p^2\Big(\frac{q_p}{p}\Big)^2 + \sqrt{\eta_p}\Big[\frac{\sqrt{\log(p)}}{\sqrt{p}} + \frac{q_p}{p}\Big](1-\rho_p)^{-1}\Big).$$

The above theorem subsumes the special case when the covariance matrix $\mathbf{\Sigma}$ is diagonal. This special case is important as it is required for setting FWER and k-FWER error control. In particular, the following corollary follows directly from Theorem 2 by setting $q_p = 0$.

Corollary 1. *If assumption A1 holds and* $\mathbf{\Sigma}$ *is diagonal, then* $E\big[N_{\rho_p}\big] = \eta_p$. *Now, let* N_p^* *denote a Poisson random variable with rate* $E[N_p^*] = \eta_p$. *If* $\sqrt{\eta_p/p}(1-\rho_p)^{-1} = o(1)$, *then*

$$\sup_{k\in\mathbb{N}}\left|P(N_{\rho_p} > k) - P(N_{\rho_p}^* > k)\right| = o(1).$$

Remark 4. *Note that Corollary 1 allows for the expected number of discoveries to go off to infinity. In particular, as* $P_0 = O\big([1-\rho_p]^{(n-2)/2}\big)$, *the condition imposed on* ρ_p *in the above result is satisfied when* $1-\rho_p = o\big(p^{-2/(n-6)}\big)$ *or, equivalently,* $E[N_p^*] = o\Big(p^{1-[4/(n-6)]}\Big)$. *Thus, this condition allows* $E[N_p^*] \to \infty$, *provided that* $n > 10$.

Figure 4 provides numerical validation of Corollary 1 in the null model setting, where $\Sigma = \mathbf{I}$. The theoretical approximations almost exactly mirror the empirical results, even when p is

as small as 50 (see Appendix A.7 for further details).

Result 3. In Appendix Section A.4, we complement Theorem 2 with a more specialized result, which specifies the limiting behavior ρ_p that allows the expected number of discoveries to converge to a finite limit as p goes to infinity.

Result 4. Our final theorem derives exact marginal p-values for the PARSEC estimates of scaled partial correlation coefficients. In the next subsection, we apply this result to control the FDR and pFDR.

Theorem 3. *Let P_0 denote the spherical cap probability as given in equation (6) and let H_{jk} denote the PARSEC partial correlation estimate between features X_j and X_k. Suppose that assumption A*1 *is satisfied and $\mathbf{\Sigma}$ is diagonal. Then for each non-negative ρ,*

$$P(|H_{jk}| > \rho) = P_0(\rho, n).$$

Remark 5. *Note that the above result is an "exact" result and holds for any fixed n and any fixed p. It is thus a "doubly finite" sample-dimension result. To our knowledge, no competing method for large-scale partial correlation screening (such as the gold standard PCS-Hub method) readily provides any type of distinct distribution for the estimated partial correlation in the large p setting, let alone an exact one. The above exact result for the PARSEC approach facilitates the calculation of exact marginal p-values.*

Remark 6. *We note that the assumptions invoked in order to prove our results are not restrictive for a number of compelling reasons. First, the main theorem (please see Theorem 1) does not assume sparsity, and even if sparsity is imposed, the block sparse matrix is allowed to grow in dimension as $p \to \infty$ and the result will still hold true (please see Theorem 2). Hence the assumptions are not strong in this regard, especially compared to work in high dimensional inference where sparsity is almost always assumed. In this sense, the results in the paper are quite general. Second, Assumptions A1 and A2 used in the paper are also invoked in the PCS-Hub paper and in Firouzi et al. (2017) respectively, both of which are screening papers that appear in IEEE Transactions on Information Theory. Thus, the assumptions invoked for the above results to hold are quite general and are not any more restrictive than those invoked in the literature. Third, neither assumption A1 nor assumption A2 are required to obtain the exact distribution of the distribution of the PARSEC partial correlation co-efficient (please see Theorem 3).*

3.3 Ultra high-dimensional statistical inference with PARSEC

We now demonstrate how the theoretical results stated above can be used for statistical inference (and not only for model selection) in the ultra high-dimensional settings where n

is fixed and p tends to infinity. In particular, we use these results to control the FWER, k-FWER, FDR and pFDR metrics for our proposed method.

k-FWER: Because FWER is a special case of k-FWER when $k = 0$, we will focus on the more general k-FWER metric. Let α be a pre-specified level of error control, k the allowable number of false discoveries, p the dimension, and n the sample size. By Corollary 1, if we choose ρ_p so that k is a $(1-\alpha)$-level quantile of the Poisson distribution with rate $p(p-1)P_0(\rho_p, n)/2$, using formula (6) for the spherical cap probability P_0, then $P(N_{\rho_p} > k) \approx \alpha$ for large p, thus allowing us to control the k-FWER at level α. The results of Corollary 1 can also be employed in other useful ways. First, Corollary 1 allows us to quantify the true level of error control achieved when a user pre-specifies a value for the screening level ρ_p (i.e., a type of generalization of the p-value in the multiple hypothesis testing context). Second, note that sometimes it may be difficult to specify the value of k. In such cases, given n, p, ρ_p and α, Corollary 1 allows us to quantify the number of false discoveries k that are being implicitly tolerated at that level of error control.

FDR: To control the FDR, we shall compute marginal p-values as follows. Recall that H_{jk} is the PARSEC estimate of the partial correlation between features X_j and X_k. Under the null hypothesis that $\mathbf{\Sigma}$ is diagonal, i.e. each partial correlation coefficient is zero, by Theorem 3, $P(|H_{jk}| > c) = P_0(c, n)$ for $c > 0$. Thus, the *exact* non-asymptotic marginal p-value corresponding to $\hat{\rho}^{jk} = H_{jk}$ is given by $P_0(|H_{jk}|, n)$. These marginal p-values for off-diagonal elements, together with the Benjamini-Hochberg (BH) and/or Benjamini-Yakuteli (BY) method, can be used to control the FDR. In addition to FDR, PARSEC is also readily amenable to controlling the pFDR (Storey 2002). We note that FDR control requires the calculation and ordering of $\binom{p}{2}$ p-values. To address this challenge, we implement a scalable FDR procedure described in Appendix B.2.

Finally, we note that our inferential approach is rich enough to cover Gaussian graphical modelling as a special case. In addition to learning the Gaussian graphical model structure, our approach can quantify the uncertainty associated with the estimated model - a property that is lacking in contemporary ultra-high dimensional methods. One of the other advantages of the PARSEC approach is that it can also be used to identify a conditional set for each feature, similar to useful methods proposed in the context of high dimensional regression (see Cho and Fryzlewicz (2012) and other similar methods).

4 Algorithms and computational properties

We now provide a suite of algorithms which facilitate the implementation of PARSEC, especially in ultra-high dimensions. In particular, Section 4.1 provides an algorithm for the

base implementation of PARSEC, followed by its highly scalable version. Lastly, we derive PARSEC's computational complexity in Section 4.2, illustrating PARSEC's scalability in large p settings.

4.1 The base and scalable implementation of PARSEC

The PARSEC approach described in Section 2 can also be expressed as a formal algorithm. This "base PARSEC algorithm" directly evaluates scaled partial correlation coefficients in matrix $\mathbf{H}$, through p parallelizable operations, and is provided in Algorithm A1 in Appendix B.1. As the base PARSEC algorithm requires p matrix inversions (Step 8 in Algorithm 1), it can thus be computationally expensive in ultra-high dimensional regimes. To address this challenge, we now propose an alternative and more scalable algorithm which exploits Sherman-Morrison-Woodbury rank one updates to speed up the evaluation of $\mathbf{H}$. Recall some of the previously used notation and formulas. Given a matrix $\mathbb{M}$, we write $\mathbf{M}_k$ for the k-th column of $\mathbb{M}$ and write M_{jk} for the element of $\mathbb{M}$ located in the j-th row and k-th column. The off-diagonal elements in the j-th row of matrix $\mathbf{H}$ are contained in the vector $\big(\tilde{\mathbb{U}}^{-j}\big)^\top \mathbf{U}_j$, where $\tilde{\mathbb{U}}^{-j} = \mathbb{C}_j \mathbb{D}_j^{-1/2}$ with $\mathbb{C}_j = \big(\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top\big)^{-1}\mathbb{U}^{-j}$, and $\mathbb{D}_j$ is a diagonal matrix whose diagonal equals to that of the matrix $(\mathbb{U}^{-j})^\top\big(\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top\big)^{-2}\mathbb{U}^{-j}$. Recall the matrix $\mathbb{U}$ which involves all the U-scores. We can compute $\big(\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top\big)^{-1}$ for each j using simple Sherman-Morrison-Woodbury rank one updates as follows:

$$\Big(\mathbb{U}^{-j}(\mathbb{U}^{-j})^\top\Big)^{-1} = \Big(\mathbb{U}\mathbb{U}^\top - \mathbf{U}_j{\mathbf{U}_j}^\top\Big)^{-1} = \mathbf{A} + \Big[\frac{1}{1-\mathbf{U}_j^\top\mathbf{A}\mathbf{U}_j}\Big]\mathbf{A}\mathbf{U}_j\mathbf{U}_j^\top\mathbf{A},$$

where $\mathbf{A} = \big(\mathbb{U}\mathbb{U}^\top\big)^{-1}$. Defining $\mathbf{B} = \mathbb{U}^\top\mathbf{A}\mathbb{U}$ and $\mathbf{F} = \mathbf{A}\mathbb{U}$, we can simplify the updates to

$$H_{jk} = \Big(\frac{B_{kj}}{1-B_{jj}}\Big)\Big\|\mathbf{F}_k + \frac{B_{kj}}{1-B_{jj}}\mathbf{F}_j\Big\|_2^{-1} \tag{9}$$

or, equivalently, $H_{jk} = \Big\|(1-B_{jj})B_{kj}^{-1}\mathbf{F}_k + \mathbf{F}_j\Big\|_2^{-1}$ - see Appendix C for more details. The procedure using equation (9), summarized in Algorithm 1, leads to significant computational savings.

4.2 Computational Complexity

The following theorem quantifies the computational complexity of the scalable implementation of PARSEC (see Appendix C for the proof).

Theorem 4. *The computational complexity of the unparallelized version of Algorithm 1 is $O(np^2 + n^2p + n^3)$, i.e., quadratic in p. The p-core computational complexity of the parallelized version of Algorithm 1 is $O(n^2p + n^3)$, i.e., linear in p.*

Algorithm 1: Scalable implementation of PARSEC

Input: $\mathbb{X}_{n\times p}$
Output: $\mathbf{H}_{p\times p}$
begin
  for $j = 1$ **to** p **do**
    $\mathbf{Z}_j = \dfrac{\mathbf{X}_j - \bar{X}_j\mathbf{1}}{\sqrt{\mathbf{S}_{jj}(n-1)}}$
  end
  Define $\mathbb{T} = [n^{-1/2}\mathbf{1}, \mathbf{T}_{2:n}]$, where $\mathbf{1}^\top\mathbf{T} = [\sqrt{n}, 0, ..., 0]$ and $\mathbf{T}_{2:n}^\top\mathbf{T}_{2:n} = \mathbf{I}_{n-1}$;
  Obtain $\mathbb{U}_{n-1\times p} = \mathbf{T}_{2:n}^\top\mathbb{Z}$;
  Obtain $\mathbf{A} = (\mathbb{U}\mathbb{U}^\top)^{-1}$;
  Obtain $\mathbf{B} = \mathbb{U}^\top\mathbf{A}\mathbb{U}$;
  Obtain $\mathbf{F} = \mathbf{A}\mathbb{U}$;
  for $k = 1$ **to** p **do**
    for $j = 1$ **to** p *except* $j = k$ **do**
      $H_{kj} = \Big(\dfrac{\mathbf{B}_{jk}}{1-\mathbf{B}_{kk}}\Big)\Big\|\mathbf{F}_j + \Big(\dfrac{\mathbf{B}_{kj}}{1-\mathbf{B}_{kk}}\Big)\mathbf{F}_k\Big\|_2^{-1}$;
    end
  end
end

	Method			
	Gaussian likelihood	**Pseudo-likelihood**	**PARSEC** (Algorithm 1)	
Property	(Banerjee et al. 2008)	(Khare et al. 2015)	**unparallelized**	**parallelized**
Computational complexity	$O(tp^4)$	$O(\min\{tnp^2, tp^3\})$	$O(np^2 + n^2p + n^3)$	$O(n^2p + n^3)$
Order in p	$O(p^4)$	$O(p^2)$	$O(p^2)$	$O(p)$
Order in t	$O(t)$	$O(t)$	$O(1)$	$O(1)$

Table 1: Computational complexity comparison with competing sparse partial correlation methods (t denotes the number of iterations required).

Computational Complexity: Table 1 compares PARSEC's computational complexity with that of state-of-the-art Gaussian likelihood and pseudo-likelihood methods. PARSEC's computational complexity is superior to both ℓ_1-based pseudo-likelihood and ℓ_1-based Gaussian likelihood approaches, irrespective of parallelization. In particular, unparallelized PARSEC's computational complexity is of order $O(p^2)$ and is therefore better than existing ℓ_1-penalized likelihood based methods because it is non-iterative. In addition and more importantly, PARSEC is amenable to parallelization: the computational complexity of the parallelized implementation of PARSEC is linear in p. This improvement stems from two sources: i) the parallelizable nature of PARSEC; and ii) the non-iterative nature of PARSEC as it yields a closed-form solution that does not require optimization. Parallelization also allows PARSEC's computational complexity to be competitive with the state-of-the-art PCS-Hub approach. As such, PARSEC is highly efficient.

Storage and Memory Complexity: In addition to improved computational complexity, PARSEC is also amenable to alternative data structures, which provide storage gains (improved "memory/storage complexity"). Concrete details are provided in Appendix D.

5 Numerical validation of statistical properties

We now proceed to validate the theoretical guarantees developed for PARSEC via numerical simulations and compare our approach to its competitor, PCS-Hub, in terms of both a) inferential and b) computational performance in the ultra-high dimensional setting.

5.1 Inferential properties and performance

Our assessment of PARSEC's inferential performance has two main components. We first validate the accuracy of PARSEC's theory by evaluating the error control achieved in the null model where $\Omega = \mathbf{I}$ (Type I error control). Thereafter, we assess PARSEC's screening performance in structured models, in terms of statistical power and ability to recover true underlying signals (Type II error control).

Assessing error control in the null model where $\Omega = \mathbf{I}$: First, we proceed to understand the level of error control afforded by PARSEC and PCS-Hub. To this end, we evaluate the accuracy of the asymptotic expressions developed for PARSEC and PCS-Hub in the null setting, i.e., when the true partial correlation matrix is the identity matrix. To reflect the ultra-high dimensional data regime, we simulate data with varying p and fixed n, where $n = 30$ and $p = 1,000$, $p = 10,000$ and even higher, all the way up to $p = 100,000$. To assess FWER and k-FWER error control, we specify partial correlation coefficient screening levels, ρ_p, for both PARSEC and PCS-Hub using the inferential procedures in Section 3.3. We also assess PARSEC's FDR control, using exact marginal p-values as outlined in Theorem 3. Recall from Section 2.1 that the PCS-Hub method does not provide a means to achieve FDR control in the partial correlation setting. As such, a direct comparison with PCS-Hub is not possible. The results in Table 2 show that PARSEC is competitive with PCS-Hub with respect to the highly conservative error control measure FWER. In contemporary applications where FWER may be too conservative, measures such as k-FWER and FDR can often be highly beneficial. Table 2 indicates that when k-FWER is used, PCS-Hub displays severe shortcomings whereas PARSEC maintains its performance: the proportion of PARSEC discoveries above k remains consistent with α when using k-FWER. Importantly, when considering FDR, the average number of PARSEC false discoveries is also consistently controlled by α, whereas PCS-Hub does not have a means to control FDR in the partial correlation setting. In addition to FDR, we note that PARSEC is able to control the pFDR. In Appendix Section E.1 we provide addi-

		$\mathbf{p}=10^3$		$\mathbf{p}=10^4$		$\mathbf{p}=10^5$	
		PARSEC	PCS-Hub	PARSEC	PCS-Hub	PARSEC	PCS-Hub
k-FWER (% of replications > k)							
FWER	$\alpha = 0.01$	0.009	0.004	0.009	0.01	0.013	0.014
	$\alpha = 0.05$	0.065	0.041	0.06	0.06	0.043	0.046
k-FWER	$\alpha = 0.01, k = 1\%$	0.018	0	0.013	0	0.012	0
	$\alpha = 0.01, k = 5\%$	0.019	0	0.012	0	0.009	0
	$\alpha = 0.05, k = 1\%$	0.078	0.001	0.056	0	0.047	0
	$\alpha = 0.05, k = 5\%$	0.072	0	0.063	0	0.053	0
FDR (averaged over all replications)							
FDR-BH	$\alpha = 0.01$	0.009	d.n.e	0.009	d.n.e	0.013	d.n.e
	$\alpha = 0.05$	0.066	d.n.e	0.051	d.n.e	0.043	d.n.e

Table 2: Error control with fixed n ($n = 30$) and increasing p in null models. Performance measures are evaluated over 1000 replications for each dimension. "d.n.e." denotes 'does not exist' as the PCS-Hub approach does not readily provide FDR control. Note, the value of k used in the k-FWER error control measure is proportional to the dimension p. We set k equal to the specified percentage of $p(p-1)/2$, the total number of all possible partial correlations.

tional assessment of PARSEC's pFDR error control. In summary, PARSEC's compelling performance across multiple error control measures highlights its versatility as a reliable inferential method in the modern ultra-high dimensional setting.

Screening performance in non-null models. We now assess PARSEC's screening performance in non-null models and investigate the setting when the p features are either Gaussian or heavy-tailed. We consider various structures for the covariance matrix $\mathbf{\Sigma}$, including auto-regressive (AR) block, block covariance, and star structures, which are summarized as follows:

- **AR block**. Features X_j with $j \leq a$ follow an AR model of order d; the rest of the features are independently generated. The first order coefficient in the AR model equals ϕ_1; the remaining AR coefficients equal $(1 - \phi_1)/(d - 1)$.
- **Block**. We set $\Sigma_{jk} = \rho \mathbf{1}\{j \neq k, j \leq a, k \leq a\} + 1\{j = k\}$.
- **Star**. The partial dependence among the features is represented by k "stars" or "hubs". Each star has a central feature connected to e other features, and there are no other connections within the star. All the nonzero elements of the inverse covariance are set equal to c. We consider two types of star structures: (i) the "Connected Star" case, where each pair of stars has a connection between them via at least one non-zero inverse covariance element; and (ii) the "Disconnected Star" case, where there are no connections between the stars.

In what follows, we undertake a comprehensive assessment of PARSEC's ability to identify

		PARSEC			PCS-Hub		
Structure	**Dimensions**	**AUC**	**FPR<.1**	**Timing**	**AUC**	**FPR<.1**	**Timing**
AR(1) Block, $\phi_1 = 0.7$							
	a=50, n=20	**0.991**	**0.925**	0.063	0.988	0.910	0.084
	a=100, n=50	**0.999**	**0.999**	0.101	**0.999**	**0.999**	0.117
	a=500, n=100	**0.999**	**0.999**	0.135	**0.999**	**0.999**	0.168
AR(2) Block, $\phi_1 = 0.7$							
	a=50, n=20	**0.981**	**0.895**	0.060	0.979	0.870	0.080
	a=100, n=50	**0.999**	**0.996**	0.108	0.998	0.982	0.118
	a=500, n=100	**0.999**	**0.999**	0.157	0.997	0.979	0.193
AR(5) Blocks, $\phi_1 = 0.7$							
	a=50, n=20	**0.999**	**0.999**	0.073	**0.999**	0.996	0.096
	a=100, n=50	**0.999**	**0.999**	0.142	**0.999**	0.998	0.158
	a=500, n=100	**0.999**	**0.992**	0.240	0.998	0.985	0.273
AR(10) Blocks, $\phi_1 = 0.7$							
	a=50, n=20	**0.849**	**0.542**	0.118	0.819	0.485	0.145
	a=100, n=50	**0.892**	**0.646**	0.170	0.806	0.489	0.184
	a=500, n=100	**0.884**	**0.638**	0.186	0.695	0.337	0.217
Block, $\sigma = 0.7$							
	a=5, n=20	**0.998**	**0.975**	0.051	0.996	0.961	0.069
	a=30, n=60	**0.998**	**0.983**	0.086	0.920	0.606	0.103
	a=50, n=100	**0.970**	**0.756**	0.122	0.589	0.124	0.152
Connected Star, $c = -0.35$**,** $n = 30$							
	k=5, e=2	**0.917**	**0.612**	0.049	0.904	0.576	0.059
	k=10, e=4	**0.974**	**0.839**	0.049	0.968	0.808	0.061
Disconnected Star, $c = -0.35$**,** $n = 30$							
	k=5, e=2	**0.868**	**0.498**	0.054	0.853	0.468	0.074
	k=20, e=2	**0.868**	**0.487**	0.054	0.851	0.470	0.074
	k=20, e=4	**0.917**	**0.630**	0.053	0.907	0.596	0.070
	k=50, e=4	**0.922**	**0.639**	0.052	0.911	0.604	0.071

Table 3: AUC values for the various covariance structures with $p = 1000$ and varying n. Each setting is replicated 1000 times. σ denotes the coefficients of non-zero elements simulated from a covariance matrix in block settings, ϕ_1 is the coefficient of the first order lag in AR block settings, and a provides the block size. c denotes the value of non-zero elements in the inverse covariance matrix in star structures. Performance is measured using median AUC, and median AUC where the FPR range is limited to less than 0.1. We also report median wall-times (in seconds). We highlight in bold the best method (highest AUC) in each setting. In all instances, PARSEC provides equal or superior performance.

true partial correlations while controlling the number of false discoveries using multiple commonly-used metrics which balance the false positive and false negative rates.

Table 3 compares the screening performance of PARSEC and PCS-Hub for the

above covariance structures using Area-Under-the-Curve (AUC), when $p = 1000$ and the sample size n is varied. We provide the median AUC over the entire ROC curve and also the median AUC for the part of the curve where the False Positive Rate (FPR) is less than 0.1. The latter measure assesses performance for more realistic values of FPR. In summary, PARSEC produces results which are always as good as and often much better than PCS-Hub across both measures and over all covariance structures considered, providing evidence of its superior screening performance. PARSEC also exhibits wall-times that are competitive with PCS-Hub. These timings are lower in all instances considered in Table 3. Moreover, we observe that when n is varied but p is fixed, PCS-Hub wall-times stay relatively constant. However, wall-times for PARSEC decrease for smaller n, that is the sample-starved settings PARSEC is designed for. Table 2 in Appendix Section E.2 also compares PARSEC and PCS-Hub in terms of another metric: Matthews Correlation Coefficient (MCC), a consolidated performance measure which integrates both the false positive and false negative rates. We observe that PARSEC uniformly obtains higher MCC values across all four error control measures, across all models, and across all signal strengths. Note that the AR block cases and the "connected star" model in Table 3 do not assume sparsity in the correlation matrix, highlighting the technical strengths of the PARSEC approach.

Though the composite measures AUC and MCC affirm PARSEC's strong consolidated performance, they do not readily assess statistical power when an error metric (such as FWER/k-FWER/FDR/pFDR) is fixed at a pre-specified level α (i.e., at a practically relevant point on the ROC curve). As an inferential method, PARSEC has the key advantage of being able to target a desired point on the ROC curve - it does not inadvertently quantify performance in regions of the curve which are not relevant to an inferential task. We investigate this distinct inferential question in Appendix Section E.2 where PARSEC is shown to achieve consistently higher sensitivity rates (i.e., statistical power) than PCS-Hub at pre-specified levels of error control. We note that PARSEC provides improved identification of true partial correlation coefficients and favorably shrinks null coefficients, as illustrated in Figure 3 (from Section 2.1) - see Appendix E.2 for further illustrations. Note also that the theoretical results for PARSEC hold for a wider class of distributions than the Gaussian (i.e., the class of vector-elliptical distributions). In an effort to understand PARSEC's performance in more general settings, we investigated its performance in the context of heavier tail distributions such as the multivariate T (see Appendix F). The results therein convincingly demonstrate PARSEC's superior performance in a broader class of distributions.

(a) Screening algorithm wall-times

	PARSEC		**PCS-Hub**
	Base	**Scalable**	
n=30, p=10^3	3.92	0.08	0.11
n=30, p=10^4	389.08	53.07	57.33
n=30, p=10^5	-	7376.76	11972.08

(b) FDR algorithm wall-times

	FDR	
	Original	**Iterative**
n=30, p=10^3	6.86	0.06
n=30, p=10^4	1442.10	20.63
n=30, p=10^5	-	2150.09

Table 4: Timing comparison of a) base versus scalable PARSEC algorithm and PCS-Hub (left), and b) the original versus the iterative FDR method (right), in the null setting where $\Omega = \mathbf{I}$. Timings provided are median wall-times over 100 simulated datasets for each structure, reported in seconds, limited to a maximum of 24 hours. Note, in the case when $p = 100,000$ the base PARSEC algorithm and FDR original algorithm do not yield a solution within 24 hours, affirming the benefits of the respective scalable versions.

5.2 Computational performance

Table 4 (a) compares wall-times for PARSEC's base implementation (Algorithm A1), PARSEC's scalable approach (Algorithm 1) and the PCS-Hub approach. We report the median wall-times (in seconds), calculated over 100 simulated datasets. It is clear that PARSEC's scalable implementation (Algorithm 1) exhibits competitive computational performance relative to the PCS-Hub approach, with significantly superior wall-times in higher dimensions such as $p = 100,000$. We further investigate the source of PARSEC's superior computational performance and demonstrate that it stems from all three aspects of computing: i) computational speed (i.e., processing speed), ii) storage needs, and iii) memory allocation requirements (see Appendix Section G).

6 Real Applications

We now proceed to demonstrate PARSEC's efficacy on modern high-dimensional applications, starting with Section 6.1 which illustrates how PARSEC's scalability to large p problems enables novel insights in breast cancer gene screening. Thereafter, Section 6.2 demonstrates how PARSEC can be used for down-stream applications of covariance estimation in finance.

6.1 Breast Cancer Gene Screening

Partial correlation graphs and network analysis are popular in cancer research as they can potentially be used to identify influential biomarkers for targeted treatment through the detection of highly-connected hubs (Grimes et al. 2019). Discovery of genetic alterations responsible for tumor growth or survival has revolutionized and driven much recent cancer

research, and hence the identification of hub genes can provide breakthrough knowledge on how breast cancer can be better treated (Giorgio et al. 2018). In particular, the reliable identification of novel genes that are pivotal in the development or prevention of the disease provides the basis for future research into potential gene therapeutic targets or treatment (Li et al. 2018). The sample-starved nature of gene expression data however presents significant challenges when constructing gene dependency networks.

In the context of gene expression data, one often has to deal with over 20,000 genes in a single analysis, but with sample sizes in the hundreds or even fewer. ℓ_1-based methods such as CONCORD are less adept at handling this many features. Hence, the standard approach in the literature is to first reduce the number of features using Cox regression models, and then build a partial correlation graph from the reduced set of genes (see Khare et al. 2015 and the references therein). This heuristic step via Cox regression serves to reduce the number of features to a more manageable one ($\sim$1,000). However, as this heuristic step is based on marginal associations between survival rates and individual genes, it may inadvertently eliminate an important gene, or a set of genes, which are jointly critical for understanding the targeted disease or outcome. In contrast, PARSEC's immediate scalability circumvents any such heuristic reduction step. This scalability permits us to jointly model all biomarkers simultaneously, and thus potentially discover gene expressions that have not yet been identified.

In addition to the considerable reduction in the number of genes, existing heuristic approaches determine the sparsity of the partial correlation graph by pre-specifying an ℓ_1-penalty/threshold level or the total number of edges of the resultant graph (Khare et al. 2015). This additional second layer of heuristics associated with ℓ_1-based methods is required as the "ground-truth" is often not available to determine the penalty parameter through cross-validation. Instead, PARSEC is rooted in a rigorous inferential framework which allows us to assess statistical significance - thus making the analysis more precise. In summary, PARSEC provides three significant advantages over existing approaches in such biomedical settings: i) scalability for handling full gene expression sets, ii) theoretical guarantees in sample-deficient settings, and iii) a principled approach for assessing statistical significance of hub genes.

We now proceed to illustrate the performance of PARSEC on a popular health application adapted from a breast cancer study, on which competing ℓ_1 methods have also been implemented - see Khare et al. (2015) and references therein. In particular, in previous applications, existing clinical information and univariate Cox regression analysis (p-value < 0.0003) were used to significantly reduce the original sample to approximately 1000 genes. PARSEC's scalability however allows us to readily consider the original full gene expression set comprised of all 24,481 gene expression levels. Note that the data contains missing values and thus the final analysis included 15,220 genes expression levels for 174

breast cancer patients (see Appendix H for additional details).

The hub genes with the 10 highest number of edges identified using a suite of error control metrics are reported in Table 7 in Appendix H. To identify hubs, we employ three approaches: i) k-FWER screening ($\alpha = 0.05$, k = 1% of p^2), ii) FWER screening ($\alpha = 0.05$) and iii) FDR control ($\alpha = 0.05$). As reported in Appendix H, PARSEC identifies a unique set of top hub genes not reported by other covariance estimation methods (see Khare et al. 2015). These top hub genes identified by PARSEC also feature prominently in recent (independent) biological studies as either being prognostic or as potential therapeutic targets. For example, there is growing evidence around the promise of HSPG2 as a target for treatment due to its negative association with survival for patients diagnosed with Triple Negative Breast Cancer (TNBC, Kalscheuer et al. 2019).

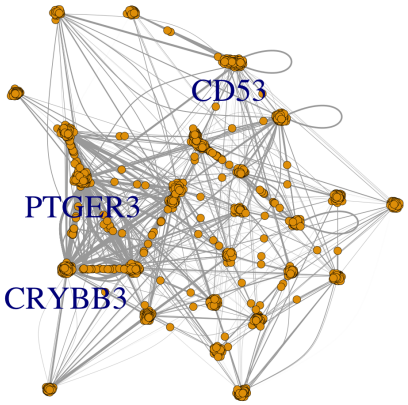


Figure 5: Network of genes screened, using k-FWER screening ($\alpha = 0.05$ and k = 1% of p^2). Edge weights (thickness) reflect the 1 minus p-values of partial correlation coefficients which exceed the screening level.

In Figure 5, we present the inferred network using the k-FWER screening approach with $\alpha = 0.05$ and k = 1% of p^2, where the edge weights represent the p-values of partial correlation coefficients which exceed the screening level. A number of centralized gene hubs are immediately visible, such as the cluster centralized around PTGER3 (EP3). The gene EP3 has been identified as a prognostic marker to breast cancer, and hence the nearest genes associated with this hub have been hypothesized to represent strong candidates for further research into potential therapeutic targets (Semmlinger et al. 2018).

In summary, the biomedical significance of the top hub genes identified by PARSEC provides empirical evidence of the method's efficacy. PARSEC's ability to simultaneously model the full gene expression set leads to the discovery of an entirely new set of hub genes previously not identified by ℓ_1 methods (recall that such methods employ a heuristic step to achieve variable reduction). In addition, PARSEC also successfully manages to identify the set of hub genes obtained via ℓ_1 methods. Appendix H provides a detailed analysis of these identified hub genes, including assessments of their statistical significance, which ℓ_1 methods are not able to quantify. As PARSEC decomposes the partial correlation learning problem into p separate regression problems, it also allows us to undertake uncertainty quantification at the level of each hub gene and not necessarily the entire graph. In particular, we can assess the statistical significance of the strongest partial correlations for each of these

biomarkers individually.

6.2 Minimum Variance Portfolio Selection

We now demonstrate how PARSEC can be leveraged in financial portfolio selection, where a stable down-stream estimate of the (inverse) covariance matrix is a critical input in determining optimal weights. For the sake of brevity a concise exposition is provided here - a detailed and comprehensive analysis can be found in Appendix I. We consider the minimum variance portfolio framework as implemented by Won et al. (2013). Let $\mathbf{\Sigma}_t$ denote the covariance matrix of the daily returns for period t. The minimum variance portfolio problem is defined as: min $\mathbf{w}_t^\top \mathbf{\Sigma}_t \mathbf{w}_t$ subject to $\mathbf{1}^\top \mathbf{w}_t = 1$, where $\mathbf{w}_t$ denotes portfolio weights, and has an analytic solution $\mathbf{w}_t^* = (\mathbf{1}^\top \mathbf{\Sigma}_t^{-1}\mathbf{1})^{-1}\mathbf{\Sigma}_t^{-1}\mathbf{1}$. Since the data is non-stationary, a rolling high-dimensional covariance estimate is required as the effective sample size is low. We re-estimate $\mathbf{\Sigma}_t$ repeatedly at the beginning of each investment period t, using a sample size of n daily (adjusted) returns preceding the period, addressing the challenge of non-stationarity in financial returns.

We consider securities in the S&P500 index and use a 20-year investment horizon starting from January 1, 1995 and ending on January 1, 2015. We re-estimate Σ_t using past data from the "estimation horizon" period, which is defined as data from the immediate past consisting of n days. These covariance estimates are then used to compute the portfolio weights, $\mathbf{w}_t^*$, at the beginning of each monthly investment (or "hold-out") period, with $\mathbf{w}_t^*$ held constant until the next investment period. We employ the PARSEC and PCS-Hub approaches to identify significant partial correlations by controlling either FWER, k-FWER or FDR. Once the graph structure is determined, we obtain estimates of the non-zero elements of the inverse covariance matrix by employing either a likelihood-based or a pseudo-likelihood-based estimation approach (see Appendix B). We also compare PARSEC to CONCORD, which is a leading ℓ_1-penalized pseudo-likelihood method.

We apply back-testing to compare the behavior of each portfolio using the performance metrics implemented in Won et al. (2013). We use the industry-standard measure, the Sharpe Ratio, to assess the effect of turnover and portfolio stability over the entire investment horizon. The results are reported in Table 5 under different estimation horizons. It is clear PARSEC provides uniformly competitive estimates across every estimation horizon. Superior performance is attained when using FWER screening ($\alpha = 0.05$) and FDR-BH screening ($\alpha = 0.05$). In contrast, the PCS-MPI method exhibits volatile performance.

		PARSEC					PCS-Hub			
$\mathbf{N}_{est}$	**CONCORD**, CV	**FWER,** $\alpha = 0.05$	**k-FWER,** $\alpha = 0.01$, $k = \approx 1000$	**k-FWER,** $\alpha = 0.05$, $k \approx 5000$	**FDR-BH,** $\alpha = 0.05$	**PARSEC** CV	**FWER,** $\alpha = 0.05$	**k-FWER,** $\alpha = 0.01$, $k \approx 1000$	**k-FWER,** $\alpha = 0.05$, $k \approx 5000$	**PCS-Hub** CV
1 month	-1.1979	**0.4244**	0.2212	-0.8092	**0.4239**	0.1066	0.4180	-0.3369	-1.1453	-0.2099
2 months	0.1939	**0.4228**	0.2965	-0.6791	**0.4240**	0.2321	**0.4203**	-0.9000	-0.3107	-0.3724
3 months	0.1677	**0.4203**	0.3158	-0.5948	**0.4199**	0.2404	0.4165	-0.8645	-1.0378	-0.5830
6 months	0.2737	**0.4442**	0.3808	0.1275	**0.4445**	0.3517	**0.4469**	-0.4426	-0.7094	0.1095
12 months	0.2593	**0.4621**	0.4343	0.3368	**0.4620**	0.4226	0.4462	-0.3906	-0.8427	-1.0158

Table 5: Comparison of Adjusted Sharpe Ratios across different methods. Similar to Khare et al. (2015), the highest Adjusted Sharpe Ratios for each estimation horizon, and values within 1% of this maximum, are highlighted in bold.

7 Conclusion

PARSEC yields a novel ultra-high dimensional method for partial correlation screening with statistical error control. It provides a highly tractable approach with superior screening and inferential performance. PARSEC's strength as a partial correlation screening approach also presents potential avenues for future work. First, PARSEC's scalability can be further improved by combining PARSEC with other large-scale screening methods such as the marginal correlation screening framework. Highly-scalable marginal correlation screening approaches could be used as a preliminary step for removing singleton features uncorrelated with the others. Thereafter, PARSEC can be exploited to identify significant partial correlations on a reduced subset of features within a shorter total run-time. Second, PARSEC is ideal for ultra-high dimensional, sample-starved applications, including the climate sciences, biomedical sciences and social sciences.

Acknowledgments: The authors gratefully acknowledge funding from the Australian Research Council (ARC)/Discovery Projects (DP).

References

Anderson, T. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, Wiley.

Banerjee, O., L. El Ghaoui, and A. d'Aspremont (2008), "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data", *Journal of Machine Learning Research* 9, pp. 485–516.

Castelo, R. and A. Roverato (2009), "Reverse engineering molecular regulatory networks from microarray data with qp-graphs", eng, *Journal of computational biology* 16 (2), pp. 213–227.

Cho, H. and P. Fryzlewicz (2012), "High dimensional variable selection via tilting", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74 (3), pp. 593–622.

Dempster, A. P. (1972), "Covariance Selection", *Biometrics* 28 (1), pp. 157–175.

Firouzi, H., A. Hero, and B. Rajaratnam (2017), "Two-Stage Sampling, Prediction and Adaptive Regression via Correlation Screening", *IEEE Transactions on Information Theory* 63 (1), pp. 698–714.

Giorgio, E. D., W. W. Hancock, and C. Brancolini (2018), "MEF2 and the tumorigenic process, hic sunt leones", *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1870 (2), pp. 261–273.

Grimes, T., S. S. Potter, and S. Datta (2019), "Integrating gene regulatory pathways into differential network analysis of gene expression data", *Scientific Reports* 9 (1), p. 5479.

Hero, A. and B. Rajaratnam (2011), "Large-Scale Correlation Screening", *Journal of the American Statistical Association* 106 (496), pp. 1540–1552.

— (2012), "Hub Discovery in Partial Correlation Graphs", *IEEE Transactions on Information Theory* 58 (9), pp. 6064–6078.

Ioannidis, J. P. A. (2005), "Why Most Published Research Findings Are False", *PLOS Medicine* 2 (8), null.

Kalscheuer, S. et al. (2019), "Discovery of HSPG2 (Perlecan) as a Therapeutic Target in Triple Negative Breast Cancer", *Scientific Reports* 9 (1), p. 12492.

Khare, K., S.-Y. Oh, and B. Rajaratnam (2015), "A convex pseudo-likelihood framework for high dimensional partial correlation estimation with convergence guarantees", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.

Li, J. et al. (2018), "Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design", *Scientific Reports* 8 (1), p. 622.

Liang, F., Q. Song, and P. Qiu (2015), "An Equivalent Measure of Partial Correlation Coefficients for High-Dimensional Gaussian Graphical Models", *Journal of the American Statistical Association* 110 (511), pp. 1248–1265.

Magwene, P. M. and J. Kim (2004), "Estimating genomic coexpression networks using first-order conditional independence", *Genome Biology* 5 (12), R100.

Romano, J. P. and M. Wolf (2005), "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing", *Journal of the American Statistical Association* 100 (469), pp. 94–108.

Semmlinger, A. et al. (2018), "EP3 (prostaglandin E2 receptor 3) expression is a prognostic factor for progression-free and overall survival in sporadic breast cancer", *BMC Cancer* 18 (1), p. 431.

Storey, J. D. (2002), "A direct approach to false discovery rates", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3), pp. 479–498.

Won, J.-H. et al. (2013), "Condition-number-regularized covariance estimation", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 75 (3), pp. 427–450.