HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2021 August 11.

Published in final edited form as:

J Am Stat Assoc. 2019 ; 114(526): 657–667. doi:10.1080/01621459.2017.1423074.

Parameter Estimation and Variable Selection for Big Systems of Linear Ordinary Differential Equations: A Matrix-Based Approach

Leqin Wu[#],
Department of Mathematics, Jinan University, Guangzhou, China

Xing Qiu[#],
Department of Biostatistics and Computational Biology University of Rochester, Rochester, New York, U.S.A.

Ya-xiang Yuan,
Academy of Mathematics and System Sciences Chinese Academy of Sciences, Beijing, China.

Hulin Wu
Department of Biostatistics, University of Texas Health Science Center at Houston, Houston, TX, U.S.A.

[#] These authors contributed equally to this work.

Abstract

Ordinary differential equations (ODEs) are widely used to model the dynamic behavior of a complex system. Parameter estimation and variable selection for a “Big System” with linear ODEs are very challenging due to the need of nonlinear optimization in an ultra-high dimensional parameter space. In this article, we develop a parameter estimation and variable selection method based on the ideas of similarity transformation and separable least squares (SLS). Simulation studies demonstrate that the proposed matrix-based SLS method could be used to estimate the coefficient matrix more accurately and perform variable selection for a linear ODE system with thousands of dimensions and millions of parameters much better than the direct least squares (LS) method and the vector-based two-stage method that are currently available. We applied this new method to two real data sets: a yeast cell cycle gene expression data set with 30 dimensions and 930 unknown parameters and the Standard & Poor 1500 index stock price data with 1250 dimensions and 1,563,750 unknown parameters, to illustrate the utility and numerical performance of the proposed parameter estimation and variable selection method for big systems in practice.

Keywords

Complex system; Ordinary differential equation; Matrix-based variable selection; High Dimension; Eigenvalue updating algorithm; Separable least squares

Correspondence should be sent to HW (hulin.wu@uth.tmc.edu).

1 INTRODUCTION

Ordinary Differential Equations(ODEs) are widely used to model the dynamic behavior of a complex system (Butcher, 2014; Commenges et al., 2011; De Jong, 2002; Hemker, 1972; Holter et al., 2001; Huang et al., 2006; Lavielle et al., 2011; Li et al., 2011; Lu et al., 2011; Ramsay et al., 2007). It is typical that the parameters which characterize the system must be estimated from the data in many real world applications. Parameter estimation of ODEs, also known as the inverse problem, has been studied by using the least squares (Li et al., 2005; Xue et al., 2010), the likelihood (Commenges et al., 2011; Lavielle et al., 2011), and Bayesian (Putter et al., 2002; Huang and Wu, 2006; Huang et al., 2006, 2010) approaches. Several other methods, such as the principal differential analysis and generalized profiling approaches (Ramsay et al., 2007; Poyton et al., 2006; Ramsay, 1996; Ramsay and Silverman, 1998) and the two-stage methods (Hemker, 1972; Varah, 1982; Chen and Wu, 2008a,b; Liang and Wu, 2008) are also proposed.

As an example, ODE is one of the popular models to quantify the dynamic gene regulatory networks (DGRNs) (Bonneau et al., 2006; Li et al., 2011; De Jong, 2002; Sakamoto and Iba, 2001; Yeung et al., 2002; Voit, 2000; Holter et al., 2001; Spieth et al., 2006). based on the high-dimensional time-course gene expression data from microarray (Schena et al., 1995; Lockhart et al., 1996) and RNA-seq (Wang et al., 2009; Garber et al., 2011). However, due to the high computational cost and model identifiability issues, most of the aforementioned parameter estimation methods are only good for small-scale systems containing at most a few dozens of variables (De Jong, 2002; Sakamoto and Iba, 2001; Yeung et al., 2002; Voit, 2000; Holter et al., 2001; Spieth et al., 2006). Recently, Lu et al. (2011) developed a procedure for reconstructing DGRNs based on linear homogeneous ODE systems. In this approach, differentially expressed genes (DEGs) are first clustered into co-expressed modules (Luan and Li, 2003; Ma et al., 2006) based on temporal patterns of their expressions in order to reduce the dimension and ease the identifiability problem. In general, for a d-dimensional linear ODE system, there are $p = d^2 + d$ parameters that need to be estimated. Here d^2 is the total number of unknown parameters in the ODE coefficient matrix and d is the number of initial conditions of the state variables that also need to be estimated. As an example, even after dimension-reduction, the resulted ODE-based DGRN for the yeast cell cycle application in Lu et al. (2011) still contains $d = 41$ dimensions (co-expression modules) and 1,722 unknown parameters that need to be estimated from the discrete, noisy time-course gene expression data. An important fact that we would like to point out is that the solutions to linear ODE systems are matrix exponential functions (Butcher, 2014) that are highly *nonlinear*. If we directly use the standard nonlinear least squares (NLS) approach (Xue et al., 2010) to estimate the parameters in the linear ODE system, we need to compute the matrix exponentials in order to evaluate the discrepancy between the observed variable and its corresponding prediction (or estimate) based on the model. However, matrix exponentials are known to be numerically unstable and cannot be computed efficiently (Moler and Van Loan, 2003). As an alternative, we may numerically solve the ODE system repeatedly, using methods such as the Runge-Kutta algorithm, to evaluate the NLS objective function. However, such a high-dimensional NLS problem is not

only hard to solve from a computational perspective, but is also prone to being trapped in local optima, which may be far away from the true global solution.

Based on the above considerations, Lu et al. (2011) applied the two-stage method (Chen and Wu, 2008a,b; Liang and Wu, 2008) to decouple the ODE coefficient matrix into d number of d-dimensional vectors; then they applied the SCAD method (Fan and Li, 2001) for parameter estimation and variable selection simultaneously for each of the row-vectors (equations) separately. This approach is straightforward to implement and computationally efficient. However, such a vector-based variable selection method ignores the wealth of structural information that is inherently possessed in the ODE coefficient matrix and it heavily depends on the good estimation of the derivatives of state variables that are sensitive to measurement errors. Consequently, it often leads to inaccurate parameter estimation and poor variable selection results (Ding and Wu, 2014).

In this paper, we propose a novel matrix-based approach to avoid the poor estimates of the vector-based two-stage method and the computational problem of the NLS method (Xue et al., 2010). At the heart of the proposed method is a special form of the separable least squares (SLS) method (Ruhe and Wedin, 1980) based on the Jordan Canonical Decomposition (JCD) of the coefficient matrix, which essentially transforms the original nonlinear optimization problem into an equivalent problem in which only d number of eigenvalues, instead of all $d^2 + d$ parameters, need to be estimated via a nonlinear optimization algorithm. The rest can be obtained by a closed-form formula with little computational cost. We further exploit the analytic form of the solution to the linear ODE system after the similarity transformation used in JCD to avoid numerically solving the original ODE system in evaluating the NLS objective function. Moreover, the derived analytic form of the objective function has analytic gradients which can be computed stably and efficiently. The estimates of the original unknown parameters are recovered from the closed-form functions of eigenvalue estimates of the coefficient matrix. In simulation studies, we show that the new approach is not only much faster, but also reaches the global optima much more frequently and produces more accurate and stable estimates than the alternative methods. Finally, we apply the proposed method to two real-world applications, one is a DGRN modeling with $d = 30$ dimensions and $p = 930$ unknown parameters, and another is stock market system modeling with $d = 1250$ dimensions and $p = 1{,}563{,}750$ unknown parameters, to demonstrate that large linear ODE systems can be recovered well using the proposed approach.

2 Models and Methods

2.1 Model Description

We consider the parameter estimation problem for the following high-dimensional homogeneous linear ODE system

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t), \quad \forall t \in [T_1, T_2], \\ \boldsymbol{x}(T_1) = \boldsymbol{x}_0, \end{cases} \tag{1}$$

where

$$\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_d(t))^T \tag{2}$$

is a d-dimensional state variable vector on a range satisfying

$$0 \le T_1 < T_2 < \infty .$$

The coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ and initial condition $\boldsymbol{x}_0 \in \mathbb{R}^d$ are the unknown parameters in the system which need to be estimated using the observed data.

In real world applications, we assume that $\boldsymbol{x}(t)$ are measured with independent errors at finite time points $(t_1, t_2, \ldots, t_n)$, and the measurement errors at each time point follow a Gaussian distribution with non-singular covariance matrix Σ_ϵ, i.e.,

$$y_i(t_j) = x_i(t_j) + \epsilon_{ij}, \quad \epsilon_{\cdot j} \in \mathbb{R}^d \sim N(0, \Sigma_\epsilon), \quad i = 1, \ldots, d; j = 1, \ldots, n . \tag{3}$$

For convenience, we denote the $d \times n$-dimensional data matrix $\{y_i(t_j)\}$ collectively as $\boldsymbol{y}$.

Based on the maximum likelihood principle, the inverse problem of estimating $\boldsymbol{A}$ and $\boldsymbol{x}_0$ can be formulated as the following nonlinear weighted least squares (NWLS) minimization problem

$$\min_{\boldsymbol{A}, \boldsymbol{x}_0} \| \boldsymbol{y}(t) - \boldsymbol{x}(\boldsymbol{A}, \boldsymbol{x}_0)(t) \|^2_{\Sigma_\epsilon} . \tag{4}$$

Depending on the context, $y_i(t)$ can either represent the observed curves or discrete data at time t; $\boldsymbol{x}(\boldsymbol{A}, \boldsymbol{x}_0)(t) := e^{t\boldsymbol{A}}\boldsymbol{x}_0$ is the solution curve of the ODE system (1) with parameters $\boldsymbol{A}$ and $\boldsymbol{x}_0$. The dimension of the above optimization space is $p = d^2 + d$. Two typical choices of norm $\| \cdot \|_{\Sigma_\epsilon}$ used in Equation (4), the weighted Euclidean-metric for discrete observations and the weighted L^2-metric for functions, are given in the Supplementary Text (Section S2).

2.2 Similarity Transformation and Separable Least-Squares

The optimization problem (4), in principle, can be numerically solved directly via any suitable nonlinear optimization algorithm designed for nonlinear least-squares problems, which is termed as the "direct LS method" in this study. In practice, when the parameter space is large (*e.g.*, $d \ge 100$), the dimension of the nonlinear optimization problem could be very large ($p = d^2 + d \geqslant 10,100$ if $d \geqslant 100$), which is difficult to solve numerically and likely to be trapped in local solutions. In this subsection, we propose a method based on the similarity transformation (ST) and separable least squares (SLS), aiming to reduce the nonlinear optimization dimension significantly, so that we could expand our capability to handle big ODE systems.

For ODE system (1), we further assume that the coefficient matrix $\boldsymbol{A}$ has no multiplicity in its spectrum (no two eigenvalues are exactly identical). This assumption does not lead to much loss of generality, because such matrices only form a zero-measure set (w.r.t. either $d\mu$, the standard Lebesgue measure on $\mathbb{R}^{d\times d}$, or any probability measure that is absolutely continuous w.r.t. $d\mu$, such as the probability measure associated with real random matrices such as Ginibre ensemble, Gaussian orthogonal ensemble, Wishart ensemble, etc.) in the space of all d-by-d matrices (Ginibre, 1965; Lehmann and Sommers, 1991; Tao, 2012).

Remark.—We would like to point out that if $\boldsymbol{A}$ does have multiplicity in its spectrum, then there exist other coefficient matrices which can generate exactly the same curves (or data points). In other words, the system is theoretically not identifiable. If no other structural information of $\boldsymbol{A}$ is given *a priori*, we will not be able to recover $\boldsymbol{A}$ even if we are given infinitely many observations without noise.

Under this assumption, the real Jordan canonical form of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}, \tag{5}$$

where $\boldsymbol{\Lambda}$ is a block diagonal matrix with only two types of blocks: a 1×1 block containing one real eigenvalue of $\boldsymbol{A}$; or a 2×2 block $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ which corresponds with a pair of conjugate complex eigenvalues $a \pm bi$. We can always choose an appropriate arrangement of $\boldsymbol{Q}$ such that the diagonal blocks $\boldsymbol{\Lambda}$ are organized as follows

$$\boldsymbol{\Lambda} = \begin{bmatrix} a_1 & b_1 & & & & & & \\ -b_1 & a_1 & & & & & & \\ & & \ddots & & & & & \\ & & & a_k & b_k & & & \\ & & & -b_k & a_k & & & \\ & & & & & c_1 & & \\ & & & & & & \ddots & \\ & & & & & & & c_{d-2k} \end{bmatrix}, \tag{6}$$

where $k \le \frac{n}{2}$ is a non-negative integer indicating the number of 2×2 blocks in $\boldsymbol{\Lambda}$.

Theorem 2.1.—*The optimization problem* (4) *is equivalent to*

$$\min_{\boldsymbol{\Lambda},\boldsymbol{Q}} \left\| \boldsymbol{y}(t) - \boldsymbol{Q}^T \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) \right\|_{\Sigma_\epsilon}^2, \tag{7}$$

where $\boldsymbol{e}$ *is a constant vector formed by k pairs of* $(0,1)^T$ *'s and* $d-2k$ *of 1's as follows*

$$\boldsymbol{e} = (\overbrace{0, 1, \ldots, 0, 1,}^{k \text{ of } (0,1)s} \quad \overbrace{1, \ldots, 1}^{d-2k \text{ of } 1's})^T. \tag{8}$$

Proof. Please see the Supplementary Text (Section S6.1).

Remark.—It is worth noting that Theorem 2.1 converts the original optimization problem (4) with $d^2 + d$ unknown parameters into the optimization problem (7), which also has a total number of $d^2 + d$ unknown parameters (i.e., d^2 unknown parameters in $\boldsymbol{Q}$ and d eigenvalues in $\boldsymbol{\Lambda}$). Our current choice of $\boldsymbol{e}$ in Eqn (8) ensures that the new transformed ODE system has a simple solution (see below), which enables us to derive a closed-form solution for $\boldsymbol{Q}$ (see Eqn (13)) when $\boldsymbol{\Lambda}$ is given. We can then apply the separable LS method to estimate $\boldsymbol{\Lambda}$, which is a nonlinear optimization of only d unknown parameters. After $\boldsymbol{\Lambda}$ is estimated, $\boldsymbol{Q}$ (with d^2 unknown parameters) can be computed by closed-form solution Eqn (13) with very little computational cost.

It is well known that for the given block-diagonalized matrix $\boldsymbol{\Lambda}$, the ODE system

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{\Lambda}\boldsymbol{x}(t), & \forall t \in [T_1, T_2], \\ \boldsymbol{x}(T_1) = \boldsymbol{e}, \end{cases} \tag{9}$$

has the following solution

$$\begin{cases} x_{2j-1}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) = \exp(a_j t)\sin b_j t, \\ x_{2j}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) = \exp(a_j t)\cos b_j t, \end{cases} \quad j = 1, 2, \ldots, k \tag{10}$$

$$x_j(\boldsymbol{\Lambda}, \boldsymbol{e})(t) = \exp(c_{j-2k} t), \quad j = 2k+1, 2k+2, \ldots, d\,. \tag{11}$$

Here $x_i(\boldsymbol{\Lambda}, \boldsymbol{e})(t)$ is the ith component of the solution vector $\boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t)$, and $\boldsymbol{e}$ is a constant vector given in (8).

Notice that for a fixed $\boldsymbol{\Lambda}$, optimizing the objective function in (7) with respect to $\boldsymbol{Q}$ can be reduced to a linear regression problem with a closed-form solution, i.e., for a given $\boldsymbol{\Lambda}$

$$\min_{Q} \left\| \boldsymbol{y}(t) - \boldsymbol{Q}^T \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) \right\|_{\Sigma_\epsilon}^2 \tag{12}$$

gives a closed-form solution (see Supplementary Text Section S6.2 for the deduction).

$$\boldsymbol{Q}(\boldsymbol{\Lambda}) = \langle \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t), \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) \rangle^{-1} \cdot \langle \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t), \boldsymbol{y}(t) \rangle\,. \tag{13}$$

Using the closed-form solutions (10), (11), and (13), the optimization problem (7) can be transformed into an equivalent problem that only involves $\boldsymbol{\Lambda}$ by the separable LS principle

$$\min_{\boldsymbol{\Lambda}} \left\| \boldsymbol{y}(t) - \boldsymbol{Q}^T(\boldsymbol{\Lambda}) \boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t) \right\|_{\Sigma_\epsilon}^2. \tag{14}$$

Despite the fact that $\boldsymbol{\Lambda}$ is a $d \times d$ matrix, it only contains d unknown parameters that need to be estimated, due to its special structure. Moreover, the objective function (14) is

continuously differentiable with respect to the parameters in $\mathbf{\Lambda}$. Therefore, we can derive the analytical formulas for the gradients of parameters, which will accelerate the nonlinear optimization procedure dramatically (see Supplementary Text, Section S6.4).

Once we obtain $\widehat{\mathbf{\Lambda}}$, an estimate of $\mathbf{\Lambda}$ that minimizes (14), we can immediately obtain $\widehat{\boldsymbol{Q}}$, the regression estimate of matrix $\boldsymbol{Q}$, from Equation (13). The estimates of the original ODE coefficient matrix and initial conditions can then be computed as

$$\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{Q}}\widehat{\mathbf{\Lambda}}\widehat{\boldsymbol{Q}}^{-1}, \quad \hat{\boldsymbol{x}}_0 = \widehat{\boldsymbol{Q}}\boldsymbol{e}, \tag{15}$$

where $\boldsymbol{e}$ is given in (8).

The following theorem shows that $\widehat{\boldsymbol{A}}$, $\hat{\boldsymbol{x}}_0$ are indeed the optimal solution of (4).

Theorem 2.2.—*$\widehat{\mathbf{\Lambda}}$ is a minimizer of* (14) *if and only if* $\left(\widehat{\boldsymbol{A}}, \hat{\boldsymbol{x}}_0\right)$ *generated by* Equation (15) *is a minimizer of* (4).

Proof. See the Supplementary Text (Section S6.3).

Based on the bijection between the local minimizers of original least-squares problem (4) and the reformulated problem (14), together with the fact that the two objective functions have identical values at corresponding local minimizers, we obtain the following corollary immediately.

Corollary 2.3.—*$\widehat{\mathbf{\Lambda}}$ is a global minimizer of* (14) *if and only if* $\left(\widehat{\boldsymbol{A}}, \hat{\boldsymbol{x}}_0\right)$ *generated by* Equation (15) *is a global minimizer of* (4).

The essence of Theorem 2.2 and Corollary 2.3 is that the original NLS optimization problem (4), which is of dimension d^2+d, is equivalent to an eigenvalue estimation problem, which is a nonlinear optimization problem of dimension d. This is a dramatic dimension reduction for the nonlinear optimization problem. We require the number of distinct data points $n > d$ in order to avoid the identifiability problem, although the total number of unknown parameters $p = d^2 + d$ can be *much greater* than n. We provide the pseudo-code of the parameter estimation algorithm, called the Similarity Transformation-based Separable Least Squares (ST-SLS), in the Supplementary Text (Section S1). While we choose the Levenberg-Marquardt Algorithm (LMA) to solve the reformulated optimization problem (14) in the ST-SLS Algorithm based on its flexibility, it can be replaced by any suitable optimization algorithm in principle.

We would like to point out that although the above method is developed for the homogeneous linear ODE model (1), it can be applied to heterogeneous linear ODE models with a simple mathematical technique that adds an additional constant term to the state variable $\boldsymbol{x}(t)$. Detailed discussion is given in the Supplementary Text (Section S4).

2.3 Asymptotic Variance and Inference

In this section, we provide the asymptotic variance-covariance matrix estimation for $\hat{\boldsymbol{A}}$, which represents the uncertainty in parameter estimation. The proofs of the two theorems in this section are provided in the Supplementary Text (Section S6.5).

In what follows we consider $\boldsymbol{A}$ as vec($\boldsymbol{A}$), which is a d^2-dimensional vector of parameters such that $\mathrm{vec}(\boldsymbol{A})_{(l-1)d+k} := \boldsymbol{A}_{lk}$. We define $\boldsymbol{D}(\boldsymbol{A}, t)$ as the following $d \times d^2$-dimensional matrix function

$$\boldsymbol{D}_{i,(l-1)d+k}(\boldsymbol{A}, t) = \frac{\partial\left(e^{t\boldsymbol{A}}\boldsymbol{x}_0\right)_i}{\partial A_{kl}}, \quad i = 1, \ldots, d; l = 1, \ldots, d; k = 1, \ldots, d\,. \tag{16}$$

Apparently, $\boldsymbol{D}_{i,(l-1)d+k}(\boldsymbol{A}, t)$ is the Jacobian matrix of the solution curves, $\boldsymbol{x}(t) := e^{t\boldsymbol{A}}\boldsymbol{x}_0$, with respect to $\boldsymbol{A}$ evaluated at time t. We can then express the total Fisher information matrix of $\boldsymbol{A}$ as a function of $\boldsymbol{D}(\boldsymbol{A}, t)$.

Theorem 2.4.— *Given* $\boldsymbol{x}_0$ *and* Σ_ϵ *(covariance matrix of the measurement error), the total Fisher information matrix pertain to the estimation of ODE system* (1) *is*

$$\boldsymbol{I}(\boldsymbol{A}) = \sum_{j=1}^{n} \boldsymbol{D}^T\left(\boldsymbol{A}, t_j\right)\Sigma_\epsilon^{-1}\boldsymbol{D}\left(\boldsymbol{A}, t_j\right). \tag{17}$$

Obviously, $\boldsymbol{D}^T\left(\boldsymbol{A}, t_j\right)\Sigma_\epsilon^{-1}\boldsymbol{D}\left(\boldsymbol{A}, t_j\right)$ is positive-semi-definite for all j. Because $\boldsymbol{I}(\boldsymbol{A})$ is a summation of positive-semi-definite matrices, it must be positive-semi-definite. Consequently, $\boldsymbol{I}(\boldsymbol{A})$ is positive-definite as long as it is of full rank.

Based on Theorem 2.4, we have the following asymptotic results for $\hat{\boldsymbol{A}}$.

Theorem 2.5.—*Assume that*

- **a.** $\hat{\boldsymbol{A}}$ *is a unique global minimizer of* (14).
- **b.** $\boldsymbol{I}(\boldsymbol{A})$, *the total Fisher information matrix, is of full-rank (hence positive-definite).*
- **c.** *The ODE system* (1), *given that* $\boldsymbol{A}$ *is the true system matrix and* $\boldsymbol{x}_0$ *is the true initial condition, is identifiable in the following sense. If there is a matrix* $\boldsymbol{B}$ *such that* $e^{t_j\boldsymbol{B}}\boldsymbol{x}_0 = e^{t_j\boldsymbol{A}}\boldsymbol{x}_0$ *for all* $j = 1, 2, \ldots, n$, *then* $\boldsymbol{B} = \boldsymbol{A}$.

With these assumptions, when $m \to \infty$, $\hat{\boldsymbol{A}}$ *converges in distribution to a normal distribution with the correct mean* ($\boldsymbol{A}$) *and covariance matrix* $\boldsymbol{I}(\boldsymbol{A})^{-1}$

$$\hat{A} \sim AN\left(A, I(A)^{-1}\right). \tag{18}$$

By definition, the asymptotic variance of $\widehat{A}_{ij}$ is the diagonal element of $\boldsymbol{I}(\boldsymbol{A})^{-1}$. More specifically, $\mathrm{var}\left(\widehat{A}_{ij}\right) = \boldsymbol{I}(\boldsymbol{A})^{-1}_{(j-1)d+i,(j-1)d+i}$ asymptotically. With these variance estimates, we can test the null hypothesis $H_{0,ij}: A_{ij} = 0$ against the corresponding alternative hypothesis $H_{1,ij}: A_{ij} \neq 0$ by using the standardized network strength, $z_{ij} := \frac{A_{ij}}{\sqrt{\mathrm{var}\left(\widehat{A}_{ij}\right)}}$, as the test statistic. Such a statistic follows an asymptotic standard normal distribution under $H_{0,ij}$. Because we need to test a large number (d^2) of hypotheses, a suitable multiple testing procedure, such as the Holm-Bonferroni procedure Holm (1979), Šidák procedure (Šidák, 1967), Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) needs to be applied to control for the overall type I error. The confidence interval for parameter estimates could also be constructed based on the asymptotic results.

2.4 Variable Selection

For problems with *a priori* information that the coefficient matrix $\boldsymbol{A}$ is a sparse matrix, it is advantageous to add a regularized term imposing sparsity on the coefficient matrix estimate.

$$\min_{\boldsymbol{A},\, x_0} \|\boldsymbol{y}(t) - \boldsymbol{x}(\boldsymbol{A}, \boldsymbol{x}_0)(t)\|^2_{\Sigma_\epsilon} + \rho(\boldsymbol{A})\,. \tag{19}$$

Possible choices of the penalty term $\rho(\boldsymbol{A})$ include LASSO (Tibshirani, 2011), SCAD (Fan and Li, 2001), MCP (Zhang, 2010), etc.

Taking the similarity transformation as we did in previous subsections leads to

$$\min_{\boldsymbol{\Lambda}} \left\|\mathbf{y}(t) - \boldsymbol{Q}^T(\boldsymbol{\Lambda})\boldsymbol{x}(\boldsymbol{\Lambda}, \boldsymbol{e})(t)\right\|^2_{\Sigma_\epsilon} + \rho\left(\boldsymbol{Q}(\boldsymbol{\Lambda})\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}(\boldsymbol{\Lambda})\right), \tag{20}$$

where $\boldsymbol{Q}(\boldsymbol{\Lambda})$ is defined by (13), respectively. We can apply the same optimization algorithm to solve the above problem.

Notice that minimizing the objective function in (20) should result in a sparse estimate for matrix $\boldsymbol{A}$ theoretically. However, we still need to use the separable LS estimate (13) for $\boldsymbol{Q}$, which is the optimal solution of (12), instead of (20). This approximation may not shrink the estimates of true zero-elements of $\boldsymbol{A}$ to exactly zero. Thus, we need to determine a numerical threshold c such that if $\left|\widehat{A}_{ij}\right| < c$, we replace $\widehat{A}_{ij}$ by zero. Similar idea has been adopted for removing estimates with small nonzero values due to numerical errors for L_1 regularized regression algorithms such as LASSO (Yukawa et al., 2012; Combettes and Wajs, 2005). One simple method for determining the threshold is to use the variance estimate in (17) and (18) in Section 2.3 to formulate an asymptotic z-test to check whether $A_{ij} \neq 0$. However, this method is not applicable to large networks or systems, because it requires the estimation of a $d^2 \times d^2$-dimensional covariance matrix which is computationally infeasible when d is large. An alternative way is to select a threshold to classify the estimated coefficients into two groups: zero and non-zero groups, based on standard classification methods such as the K-nearest-neighbor (KNN) algorithm.

Note that the closed-form gradient of parameters for the objective function (20) is not available. Therefore, a derivative-free optimization (DFO) algorithm such as the NEWUOA (Powell, 2006; Zhang et al., 2010) needs to be used and the computational cost is higher compared to that of the ST-SLS Algorithm in this case (see more discussions in the Supplementary Text, Section S3). The pseudo-code of ST-SLS with variable selection (ST-SLS-VS) is provided in the Supplementary Text (Section S1).

3 Simulation Studies

3.1 Design of Simulation Experiments

In this section, we compare the proposed ST-SLS and ST-SLS-VS methods based on the eigenvalue estimation framework with the existing methods via simulation studies with different dimensions and different noise levels. Notice that it is not trivial to design high-dimensional ODE simulation experiments. To generate reasonable ODE simulation models, special cares must be taken in order to avoid the collinearity of the simulated system and ease the identifiability problem. In particular, the eigenvalues of the coefficient matrix $\boldsymbol{A}$ need to be bounded away from each other. Based on these considerations, we obtain $\boldsymbol{A}$ by first generating its eigenvalues with good properties, then randomly generating its eigenvectors. More specifically, we first generate the eigenvalues by their real parts and imaginary parts separately. The real part of each eigenvalue should be non-positive in order to make the system stable. But it cannot be too negative, otherwise the ODE solution as an exponential function of time will decay to zero very rapidly, which may produce an ODE system numerically unidentifiable (Miao et al., 2011). In our simulation experiments, the real parts of eigenvalues are generated from a uniform distribution on [−0.7,0]. The imaginary parts of eigenvalues are only required to be bounded away from each other. For example, a typical choice, employed in our simulation studies, is $\pm 2\boldsymbol{\pi}, \pm 4\boldsymbol{\pi}, \ldots, \pm d\boldsymbol{\pi}$ with a small Gaussian noise added. Once $\boldsymbol{\Lambda}$ is generated, we multiply it by a randomly generated non-singular matrix $\boldsymbol{Q}$ to create the coefficient matrix, i.e.,

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}. \tag{21}$$

Technically $\boldsymbol{Q}$ can be any invertible square matrix, however for the variable selection experiment, the coefficient matrix should be sparse. Hence we use matrix $\boldsymbol{Q}$ with a special block-diagonal structure, which guarantees the sparsity of both $\boldsymbol{Q}$ and its inverse, consequently $\boldsymbol{A}$ can be generated as a sparse matrix.

Once the ODE coefficient matrix is generated, observed data are generated from ODE model (1) using its analytical solution. The time-points of observations are distributed evenly on the interval [0,1]. Random noise is added to the simulated data from the ODE system, which is *i.i.d.* Gaussian noise with a distribution $N\left(0, (\alpha\sigma)^2\right)$, where σ is taken to be the sample standard deviation of the original data, α controls the noise level, which is taken as 0, 0.1, or 0.3 respectively; in which 0 stands for the noise-free case. The dimension of simulation models is set as d = 30,100,300,1000, respectively. All results are given based on 1000 random simulations, except for the 1000-dimensional case, which is based on 100

simulations due to high computational cost. All simulations were performed on a laptop running Xubuntu 14.04 operating system with 2.5GHz CPU and 8G of RAM.

3.2 Parameter Estimation Comparisons

In this subsection, we present the results for parameter estimation comparisons between the proposed ST-SLS method and the direct-LS method from simulation studies. For fair comparisons, the Levenberg-Marquardt algorithm is employed as the optimization solver for both the ST-SLS method (iteratively updating $\boldsymbol{\Lambda}$) and the direct-LS method (estimating $\boldsymbol{A}$ by directly minimizing the LS objective function). Please refer to the Supplementary Text (Section S3) for more details on the optimization algorithms.

The simulation results of parameter estimation comparisons are reported in Table 1. We compare the two methods in computational cost, goodness-of-fit, and parameter estimation accuracy for different ODE system dimensions (d) and different noise levels (α). The computational cost is quantified by the CPU time (in seconds) used to run the algorithms. The goodness-of-fit is evaluated by the Relative Residue Sum-of-Squares (RRSS) of model fitting, which is the objective function value at the final solution, divided by the squared Frobenius norm of the data matrix. The overall parameter estimates are evaluated by the Relative Estimation Error (REE), which is defined as

$$\mathrm{REE}(\boldsymbol{A}) = \frac{\| \widehat{\boldsymbol{A}} - \boldsymbol{A} \|_F}{\|\boldsymbol{A}_0\|_F} \times 100\%, \qquad \mathrm{REE}(\boldsymbol{x}_0) = \frac{\|\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2}{\|\boldsymbol{x}_0\|_2} \times 100\%, \tag{22}$$

where $(\boldsymbol{A}, \boldsymbol{x}_0)$ are the true parameters and $\left(\widehat{\boldsymbol{A}}, \hat{\boldsymbol{x}}_0\right)$ are their corresponding estimates.

From Table 1, we see that, for the noise-free case ($\alpha = 0$), both methods for all the dimensions produced good parameter estimates with perfect fit. The computational time increases with the system dimension (d) as expected. This demonstrates that both methods are good under the ideal case of no measurement error. When the measurement noise is added to the data, the direct-LS method produces poor results. The relative error of fitting (RRSS) of the direct-LS method could go up to 30–45%, and the estimation error (REE) could be 6 to 12 fold difference between the estimate and the true value of the coefficient matrix $\boldsymbol{A}$. This indicates that the direct-LS method likely converges to local solutions which can be far away from the true solution. For the cases of higher dimensions ($d = 300$ and 1000), the direct-LS method does not converge and fails to obtain the estimates. On the contrary, our new ST-SLS method produces reasonable results for all the simulation cases. The relative RSS of model fitting is very low (<1%) and much smaller than that of the direct-LS method for all simulation cases, suggesting that our new method fits the model very well. The REE of coefficient matrix $\boldsymbol{A}$ ranges from <1% to 20.6% and the REE for initial value estimates is even smaller. So the ST-SLS algorithm produces very good estimates for all unknown parameters. We also observe that for both methods, the estimation for initial state $\hat{\boldsymbol{x}}_0$ is much better than the estimate of coefficient matrix $\boldsymbol{A}$. This is because the estimate of $\boldsymbol{x}_0$ is the fitted solution evaluated at time $t = 0$ and the model fitting is always very good. In addition, the proposed ST-SLS is very fast and produces results in a few seconds for the cases of low or medium dimensions ($d = 10$ to 100), which require many

hours of CPU time for the direct-LS method. For the high-dimensional case ($d = 300$ and 1000) for which the direct-LS method fails to obtain the results, the ST-SLS algorithm is still able to obtain good results in a few minutes ($d = 300$) or a few hours ($d = 1000$) on a regular PC, which demonstrates the scalability of our new method for handling large systems.

3.3 Variable Selection Comparisons

For high-dimensional ODE variable selection, the only existing computationally feasible method is the two-stage method (Lu et al., 2011). In this subsection, we compare our new ST-SLS-VS algorithm (equipped with three different regularized terms) with the two-stage method in terms of variable selection performance for big ODE systems.

We performed 1000 simulation runs with different noise levels for $d = 30$, 100, and 300 respectively; and 100 simulation runs for $d = 1000$ due to the high computational cost. The results based on the average of these simulation runs are reported in Table 2. In this simulation study, we compared the Sensitivity (SEN) and Specificity (SPE), which measure the true and false positive rates of variable selection, respectively, between the two-stage method (Lu et al., 2011) and the proposed ST-SLS-VS methods.

As we pointed out in Section 2.4, no closed-form gradient formula of the objective function (20) is available to implement the ST-SLS-VS algorithm. We have to use a derivative-free optimization algorithm, such as NEWUOA (Powell, 2006; Zhang et al., 2010) for optimization (see the Supplementary Text, Section S3), which requires a higher computational cost. For example, the ST-SLS-VS algorithm produced the results for an average of about 5 minutes for dimension $d = 300$ cases and 5–6 hours for dimension $d = 1000$ cases with a high noise level, which is slower than that of the ST-SLS algorithm. The reason that the ST-SLS-VS algorithm is slower is two-fold: no closed-form gradient can be used and the objective function is more complicated to evaluate. We implemented the ST-SLS-VS algorithm using Fortran while the two-stage method is implemented in R, that is why we did not compare the computational cost between the ST-SLS-VS algorithm and the two-stage method. But in general, the two-stage method is much faster because it converts the linear ODE parameter estimation into linear regression model fitting.

From Table 2 and 3, we can see that the sensitivity of the two-stage method, ranging from 85% to 95%, is generally good for most simulation cases, but its specificity is very low (ranging from 44% to 67%) with noisy data. In comparison, the performance of the proposed ST-SLS-VS method was very stable, and the three choices of regularization terms produced similar results. Our methods not only identified the exactly correct results in all noise-free cases, but also had very good sensitivity (mostly higher than 80%) and specificity (97–100%) in other cases. Overall, our new method outperforms the existing two-stage method in variable selection.

4 Real Data Analysis

We applied the proposed ODE parameter estimation and variable selection methods to two application data sets to illustrate their utility and scalability for large scale systems. The first one is a set of time-course microarray data collected from yeast culture at stationary

phase (Aragon et al., 2006) with a medium-size system of 30 dimensions and 930 unknown parameters. The second one comprises of 10-year historic daily values of stocks that were indexed by the Standard & Poor with a large system of 1,250 dimensions and 1,563,750 unknown parameters.

4.1 Time-course Yeast Microarray Data Analysis

The first application example is a subseries of time-course gene expression data (Gene Expression Omnibus number GSE3688) collected from yeast cells in stationary-phase cultures with the oxidative stress exposure (Aragon et al., 2006). These data were collected every 1-minute for 35 minutes, with an additional final time point at 60 minutes (a total of 37 time points) using microarray. We applied the functional principal component analysis approach (Wu and Wu, 2013) and identified top 30 significant genes related to cycle regulations (Spellman et al., 1998). Our goal is to study the regulatory relationships among these 30 genes using a linear ODE model.

We applied our proposed methods and the developed ST-SLS/ST-SLS-VS algorithms described in Section 2 to the gene expression data and recovered a dynamic network for the top 30 significant yeast cell cycle-related genes. We obtained the estimated dynamic system coefficient matrix ($\widehat{\boldsymbol{A}}$) and the standard deviation for each edge using Equations (17) and (18) in Section 2.3. As discussed and suggested in Section 2.4, we used the two-sided z-test with the Holm-Bonferroni multiple testing procedure (Holm, 1979) and determined the network sparsity by controlling the familywise error rate at 0.05. The resulted network has a sparsity of 95% and is illustrated in Figure 1. Note that 14 isolated genes (FLC2, PET9, RDH54, BEM1, BUD3, NDC80, MMR1, CAR2, SPT21, GCV2, WHI3, ARG1, GNT, and YKR012C) are not included in this plot. The reconstructed gene regulatory network is provided in the Supplementary Table S1.

From Figure 1, we see that SST2, PUT1, ZSP1, DSN1, and SPC34 are central hub nodes with the largest number of adjacent edges (network degree). According to the Saccharomyces Genome Database (SGD) (Cherry et al., 1998), SST2 encodes GTPase-activating protein for GPA1P, which is required to prevent receptor-independent signaling of the mating pathway. The null mutation of this gene leads to increased cell size and decreased growth rate. PUT1 encodes proline oxidase and the mutation of this gene results in the inability of yeast to grow when proline is the sole nitrogen source. ZSP1 is a protein of unknown function but is known to interact with PHO88, which is a member gene of the phosphate metabolism pathway. DSN1 is an essential component of the MIND kinetochore complex and is known to play an important role in attachment of spindle microtubules to kinetochore involved in meiotic sister chromatid segregation. SPC34 is a spindle pole component, which is an essential subunit of the Dam1 complex (DASH complex). Both DSN1 and SPC34 are components of the kinetochore and their connection is well established (Tanaka et al., 2005; Pramila et al., 2006). The connection between SPC34 and SST2 has also been documented (Montpetit et al., 2005).

Other network connections identified by our methods are novel and may help generate hypotheses for further investigations. For example, ZSP1 is an under-studied gene which is only known to interact with PHO88. We discovered that it had a strong connection

with PHO89, which is another member gene in the phosphate metabolism pathway. This observation suggests that ZSP1 may play a more important role in phosphate metabolism than what we currently know. The strong connection between PUT1 and SST2 is somewhat surprising and interesting because PUT1 and SST2 seem to fulfill very different biological functions. PUT1 is critical for *S. cerevisiae* to digest proline, which is the most abundant source of nitrogen in grapes, the natural environment of wild yeast (Huang and Brandriss, 2000). SST2 is best known for its function in regulating mating response, which seems to be unrelated to proline digestion. However, SST2 is also known to be involved in cell proliferation (Lopez et al., 1997) and growth, especially in a nutrient-limited environment (Lopez et al., 2001; Boer et al., 2003). Our findings suggest that PUT1 and SST2 might have an intimate relationship in the interplay between nitrogen metabolism and cell growth. In addition, we found that three genes (YLR297W, YMR253C, YPR174C) in the network have no clear biological annotation in literature. Among them, YLR297W is a regulator of PUT1 and SST2, the two most connected hub nodes. Our results may provide useful insights for future experimental investigations of biological functions of these genes.

4.2 Standard & Poor Stock Market Data Analysis

Traditionally stochastic differential equation (SDE) models such as the Black-Scholes-Merton Model (Black and Scholes, 1973; Merton, 1973; Øksendal, 2013) is used for modeling stock market price data. It is known that the corresponding ODE model could be used to describe the mean behavior of the SDE (Ahmed, 1998) (Theorem 1 in Chapter 2). Here we apply the linear ODE model to stock price data from the S&P 1500 (also known as S&P Composite 1500 Index) to investigate the long-term dynamic interactions of stock price changes for the companies in the S&P 1500. The data used in this study cover 10-year span of daily closing price of these stocks from 2004 to 2014 (2,668 trading days). The original index contains 1,501 stocks, of which 251 were removed from the analysis due to missingness and other data issues. Based on the remaining 1,250 stocks, we reconstructed a linear ODE system of $d = 1{,}250$-dimensions, or $p = 1{,}563{,}750$ unknown parameters. Our variable selection algorithm produced a network of sparsity of 97.3%. This reconstructed network is provided in Supplementary Table S2.

Table 4 lists the top ten companies (nodes) that have the highest network degree in this graph. One interesting observation is that most of these highly connected companies are *not* the largest corporations by market capitalization, such as Apple Inc. or Exxon Mobile. Instead, four of them provide the basic IT infrastructure such as telephone service or network hardware; two of them are related to healthcare services; three provide financial services, which can also be considered as the “infrastructure” for modern economy. In summary, most connected companies are not the largest or most famous ones indexed by the Standard & Poor, but those that provide the fundamental infrastructure for the entire economy.

To better understand the interactions of these companies from a more focused perspective, we divided the stocks into sectors and reconstructed the sub-network for each sector. More specifically, we downloaded the list of stocks issued by 500 large-cap companies indexed by the Standard & Poor as of October 12, 2015, among which 421 nodes (companies) are

not isolated nodes. These companies were further divided into nine sectors according to the Global Industry Classification Standard (GICS)^{SM}. In sub-network construction for each sector, we retain edges with absolute strength greater than 95% of all edges in order to make the results comparable across sectors. We define the hubs as the top 10% most connected (measured by network degrees) companies within each sector, which are listed in Table 5. These sub-networks are illustrated in individual figures and are provided as one compressed file (Supplementary File S3).

We noticed some interesting results from Table 5. For example, Harman Internaltional Industries and Amazon are the two most connected companies in the Consumer Discretionary sector as expected, because both companies have wide varieties of products that may influence or be influenced by other industrial leaders. However, it is somewhat surprising to see that Wynn Resorts, which is a developer and operator of high-end hotels and casinos, ranked the third among all 63 companies in this category. Further investigation shows that all 13 connections related to the Wynn Resorts are *inward* connections, which means that the Wynn Resorts is highly dependent to the performance of many other companies in this sector, but its stock price does not have high impact to other companies. This observation may suggest that we may use hotel and casino performance as a "litmus test" of the overall fitness of consumer spending.

5 Discussion

In this paper, we present a new ODE parameter estimation and model selection framework which is based on estimating the eigenvalues of the linear ODE coefficient matrix instead of directly estimating its entries. This new approach dramatically reduces the dimension of the corresponding nonlinear optimization problem from $p = d^2 + d$ to d, and the rest of the d^2 parameters can be obtained from a closed-form formula that does not require extensive computation. As a result, our proposed algorithms are much faster and more stable than competing procedures and can be easily scaled up to handle large ODE systems. Moreover, our reformulation of the problem provides closed-form gradients of the objective function, that can be used to further accelerate and stabilize the computation.

In simulation studies, we demonstrate that the new ST-SLS method is much stabler and faster than the competing method to locate the global solution of the high-dimensional optimization problem, which leads to better performance for parameter estimation for big ODE systems. The superior performance of our new ST-SLS estimation method and the corresponding variable selection algorithm is not only due to the capability of significant dimension reduction and the availability of closed-form gradients of the objective function, but also the fact that the coupled ODE information is used efficiently.

We also applied our new algorithms to two real world appliactions to illustrate their usability in practice; one is the yeast cell cycle gene expression data with 30 dimensions and another is the Standard & Poor Index stock price data with 1,250 dimensions. Our analysis results show that the new methods could effectively recover high-dimensional dynamic networks based on observed time-course data.

Our proposed methods are applicable to the general high-dimensional linear ODE model that is identifiable in theory, but some attentions should be paid in practical implementations. In practice, the linear ODE is theoretically identifiable if the eigenvalues of the coefficient matrix are distinct; but the ODE model may have numerical or statistical identifiability problems (Miao et al., 2011) when several eigenvalues have zero or near-zero imaginary parts (e.g., more than 2 real eigenvalues are present), this is because more real eigenvalues indicate more exponential terms in the ODE solution and the power of exponential terms is difficult to distinguish and identify numerically, which is similar to the multi-collinearity problem in linear regression. Also notice that our proposed methods require the number of distinct data points to be greater than the dimension of the ODE system, i.e., $n > d$, although the total number of unknown parameters $p = d^2 + d$ can be greater than n. This requirement is needed to avoid the identifiability problem. In general, the identifiability problem has to be dealt before our method can be applied. Usually the model needs to be modified or some variables can be combined to reduce the identifiability problem, but this is beyond the scope of this paper. Motivated readers can find more information on this topic in Miao et al. (2011).

In this Big Data era, it is a common task to build dynamic relationships among many components or elements in a big system based on more and more affordable frequent time-course data, so that the complex networks can be reconstructed and analyzed (Liu et al., 2011; Barabasi et al., 2011). A linear ODE system is a simple yet powerful model that can be used to describe dynamic relationships among elements of a big system. Future extension of similar ideas in this article to high-dimensional nonlinear ODE systems (Wu et al., 2014) and/or systems with partially observed variables (Wu et al., 2015), although challenging, is warranted. We believe that the field of identification and analysis of high-dimensional, complex dynamic systems is still in its infancy despite its wide applications in practice. We hope that our work will motivate more research in this direction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. Hongqi Xue for his thoughtful discussions and helpful comments on the earlier version of this paper, and Dr. Michelle Carey for her suggestion of the stock market data. This work is supported in part by NIH RO1 AI087135, Respiratory Pathogens Research Center (NIAID contract number HHSN272201200005C), the University of Rochester CTSA award number UL1 TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health, the University of Rochester Center for AIDS Research (NIH 5 P30 AI078498-08), and National Natural Science Foundation of China Grants 11526096 and 11601185.

REFERENCES

Ahmed N. (1998), Linear and Nonlinear Filtering for Scientists and Engineers, World Scientific.

Aragon AD, Quiñones GA, Thomas EV, Roy S, and Werner-Washburne M. (2006), “Release of extraction-resistant mRNA in stationary phase Saccharomyces cerevisiae produces a massive increase in transcript abundance in response to stress,” Genome Biology, 7, R9. [PubMed: 16507144]

Barabasi A, Gulbahce N, and Loscalzo J. (2011), "Network medicine: a network-based approach to human disease," Nature Reviews Genetics, 12, 56–68.

Benjamini Y. and Hochberg Y. (1995), "Controlling the false discovery rate: a practical and powerful approach to multiple testing," Journal of the royal statistical society. Series B (Methodological), 57, 289–300.

Black F. and Scholes M. (1973), "The pricing of options and corporate liabilities," Journal of political economy, 81, 637–654.

Boer VM, de Winde JH, Pronk JT, and Piper MD (2003), "The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur," Journal of Biological Chemistry, 278, 3265–3274.

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, and Thorsson V. (2006), "The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo," Genome biology, 7, R36. [PubMed: 16686963]

Butcher J. (2014), "Ordinary differential equations," in Walter Gautschi, Springer, vol. 3, pp. 7–8.

Chen J. and Wu H. (2008a), "Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics," Journal of the American Statistical Association, 103, 369–384.

Chen J. (2008b), "Estimation of time-varying parameters in deterministic dynamic models," Statistica Sinica, 18, 987–1006.

Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, et al. (1998), "SGD: Saccharomyces genome database," Nucleic Acids Research, 26, 73–79. [PubMed: 9399804]

Combettes PL and Wajs VR (2005), "Signal recovery by proximal forward-backward splitting," SIAM Journal on Multiscale Modeling and Simulation, 4, 1168–1200.

Commenges D, Jolly D, Drylewicz J, Putter H, and Thiébaut R. (2011), "Inference in HIV dynamics models via hierarchical likelihood," Computational Statistics & Data Analysis, 55, 446–456.

De Jong H. (2002), "Modeling and simulation of genetic regulatory systems: a literature review," Journal of Computational Biology, 9, 67–103. [PubMed: 11911796]

Ding A. and Wu H. (2014), "Estimation of ordinary differential equation parameters using constrained local polynomial regression," Statistica Sinica, 24, 1613–1631. [PubMed: 26401093]

Fan J. and Li R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," Journal of the American Statistical Association, 96, 1348–1360.

Garber M, Grabherr MG, Guttman M, and Trapnell C. (2011), "Computational methods for transcriptome annotation and quantification using RNA-seq," Nature Methods, 8, 469–477. [PubMed: 21623353]

Ginibre J. (1965), "Statistical ensembles of complex, quaternion, and real matrices," Journal of Mathematical Physics, 6, 440.

Hemker PW (1972), "Numerical methods for differential equations in system simulation and in parameter estimation," Analysis and Simulation of Biochemical Systems, 28, 59–80.

Holm S. (1979), "A simple sequentially rejective multiple test procedure," Scandinavian journal of statistics, 6, 65–70.

Holter NS, Maritan A, Cieplak M, Fedoroff NV, and Banavar JR (2001), "Dynamic modeling of gene expression data," Proceedings of the National Academy of Sciences, 98, 1693–1698.

Huang HL and Brandriss MC (2000), "The regulator of the yeast proline utilization pathway is differentially phosphorylated in response to the quality of the nitrogen source," Molecular and cellular biology, 20, 892–899. [PubMed: 10629046]

Huang Y, Liu D, and Wu H. (2006), "Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system," Biometrics, 62, 413–423. [PubMed: 16918905]

Huang Y. and Wu H. (2006), "A Bayesian approach for estimating antiviral efficacy in HIV dynamic models," Journal of Applied Statistics, 33, 155–174.

Huang Y, Wu H, and Acosta EP (2010), "Hierarchical Bayesian inference for HIV dynamic differential equation models incorporating multiple treatment factors," Biometrical Journal, 52, 470–486. [PubMed: 20661953]

Lavielle M, Samson A, Karina Fermin A, and Mentré F. (2011), "Maximum Likelihood Estimation of Long-Term HIV Dynamic Models and Antiviral Response," Biometrics, 67, 250–259. [PubMed: 20486926]

Lehmann N. and Sommers H-J (1991), "Eigenvalue statistics of random real matrices," Physical Review Letters, 67, 941–944. [PubMed: 10045029]

Li Z, Li P, Krishnan A, and Liu J. (2011), "Large-scale dynamic gene regulatory network inference combining differential equation models with local dynamic Bayesian network analysis," Bioinformatics, 27, 2686–2691. [PubMed: 21816876]

Li Z, Osborne MR, and Prvan T. (2005), "Parameter estimation of ordinary differential equations," IMA Journal of Numerical Analysis, 25, 264–285.

Liang H. and Wu H. (2008), "Parameter estimation for differential equation models using a framework of measurement error in regression models," Journal of the American Statistical Association, 103, 1570–1583. [PubMed: 19956350]

Liu Y, Slotine JJ, and Barabasi AL (2011), "Controllability of complex networks," Nature, 473, 167–73. [PubMed: 21562557]

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Norton H, et al. (1996), "Expression monitoring by hybridization to high-density oligonucleotide arrays," Nature Biotechnology, 14, 1675–1680.

Lopez F, Estève J-P, Buscail L, Delesque N, Saint-Laurent N, Théveniau M, Nahmias C, Vaysse N, and Susini C. (1997), "The tyrosine phosphatase SHP-1 associates with the sst2 somatostatin receptor and is an essential component of sst2-mediated inhibitory growth signaling," Journal of Biological Chemistry, 272, 24448–24454.

Lopez F, Ferjoux G, Cordelier P, Saint-Laurent N, Estève J-P, Vaysse N, Buscail L, and Susini C. (2001), "Neuronal nitric oxide synthase: a substrate for SHP-1 involved in sst2 somatostatin receptor growth inhibitory signaling," The FASEB Journal, 15, 2300–2302. [PubMed: 11511520]

Lu T, Liang H, Li H, and Wu H. (2011), "High-Dimensional ODEs Coupled With Mixed-Effects Modeling Techniques for Dynamic Gene Regulatory Network Identification," Journal of the American Statistical Association, 106, 1242–1258. [PubMed: 23204614]

Luan Y. and Li H. (2003), "Clustering of time-course gene expression data using a mixed-effects model with B-splines," Bioinformatics, 19, 474–482. [PubMed: 12611802]

Ma P, Castillo-Davis CI, Zhong W, and Liu JS (2006), "A data-driven clustering method for time course gene expression data," Nucleic Acids Research, 34, 1261–1269. [PubMed: 16510852]

Merton RC (1973), "Theory of rational option pricing," The Bell Journal of economics and management science, 4, 141–183.

Miao H, Xia X, Perelson AS, and Wu H. (2011), "On identifiability of nonlinear ODE models and applications in viral dynamics," SIAM review, 53, 3–39. [PubMed: 21785515]

Moler C. and Van Loan C. (2003), "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," SIAM review, 45, 3–49.

Montpetit B, Thorne K, Barrett I, Andrews K, Jadusingh R, Hieter P, and Measday V. (2005), "Genome-wide synthetic lethal screens identify an interaction between the nuclear envelope protein, APQ12P, and the kinetochore in Saccharomyces cerevisiae," Genetics, 171, 489–501. [PubMed: 15998715]

Øksendal B. (2013), Stochastic differential equations: an introduction with applications, Springer Science & Business Media, 5th ed.

Powell M. (2006), "The NEWUOA software for unconstrained optimization without derivatives," Large-Scale Nonlinear Optimization, 83, 255–297.

Poyton A, Varziri MS, McAuley KB, McLellan P, and Ramsay J. (2006), "Parameter estimation in continuous-time dynamic models using principal differential analysis," Computers & Chemical Engineering, 30, 698–708.

Pramila T, Wu W, Miles S, Noble WS, and Breeden LL (2006), "The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle," Genes & development, 20, 2266–2278. [PubMed: 16912276]

Putter H, Heisterkamp S, Lange J, and De Wolf F. (2002), "A Bayesian approach to parameter estimation in HIV dynamical models," Statistics in Medicine, 21, 2199–2214. [PubMed: 12210633]

Ramsay J. (1996), "Principal differential analysis: Data reduction by differential operators," Journal of the Royal Statistical Society. Series B (Methodological), 58, 495–508.

Ramsay J. and Silverman BW (1998), "Functional data analysis," Statistics and Computing, 8, 401–403.

Ramsay JO, Hooker G, Campbell D, and Cao J. (2007), "Parameter estimation for differential equations: a generalized smoothing approach (with Discussion)," Journal of the Royal Statistical Society, 69, 741–796.

Ruhe A. and Wedin PA (1980), "Algorithms for separable nonlinear least squares problems," SIAM Review, 22, 318–337.

Sakamoto E. and Iba H. (2001), "Inferring a system of differential equations for a gene regulatory network by using genetic programming," in Proceedings of the 2001 Congress on Evolutionary Computation, vol. 1, pp. 720–726.

Schena M, Shalon D, Davis RW, and Brown PO (1995), "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," Science, 270, 467–470. [PubMed: 7569999]

Šidák Z. (1967), "Rectangular confidence regions for the means of multivariate normal distributions," Journal of the American Statistical Association, 62, 626–633.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, and Futcher B. (1998), "Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization," Molecular Biology of the Cell, 9, 3273–3297. [PubMed: 9843569]

Spieth C, Hassis N, and Streichert F. (2006), "Comparing mathematical models on the problem of network inference," in Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, ACM, pp. 279–286.

Tanaka K, Mukae N, Dewar H, van Breugel M, James EK, Prescott AR, Antony C, and Tanaka TU (2005), "Molecular mechanisms of kinetochore capture by spindle microtubules," Nature, 434, 987–994. [PubMed: 15846338]

Tao T. (2012), Topics in random matrix theory, vol. 132, American Mathematical Society Providence, RI.

Tibshirani R. (2011), "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society, 73, 267–288.

Varah JM (1982), "A spline least squares method for numerical parameter estimation in differential equations," SIAM Journal on Scientific and Statistical Computing, 3, 28–46.

Voit EO (2000), Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists, Cambridge University Press.

Wang Z, Gerstein M, and Snyder M. (2009), "RNA-Seq: a revolutionary tool for transcriptomics," Nature Reviews Genetics, 10, 57–63.

Wu H, Lu T, Xue H, and Liang H. (2014), "Sparse Additive ODEs for Dynamic Gene Regulatory Network Modeling," Journal of the American Statistical Association, 109, 700–716. [PubMed: 25061254]

Wu H, Miao H, Xue H, Topham D, and Zand M. (2015), "Quantifying Immune Response to Influenza Virus Infection via Multivariate Nonlinear ODE Models with Partially Observed State Variables and Time-Varying Parameters," Statistics in Biosciences, 7, 147–166. [PubMed: 26085850]

Wu S. and Wu H. (2013), "More powerful significant testing for time course gene expression data using functional principal component analysis approaches," BMC Bioinformatics, 14, 6. [PubMed: 23323795]

Xue H, Miao H, and Wu H. (2010), "Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error," Annals of Statistics, 38, 2351–2387. [PubMed: 21132064]

Yeung MS, Tegnér J, and Collins JJ (2002), "Reverse engineering gene networks using singular value decomposition and robust regression," Proceedings of the National Academy of Sciences, 99, 6163–6168.

Yukawa M, Tawara Y, Yamagishi M, and Yamada I. (2012), "Sparsity-aware adaptive filters based on lp-norm inspired soft-thresholding technique," in 2012 IEEE International Symposium on Circuits and Systems, IEEE, pp. 2749–2752.

Zhang CH (2010), "NEARLY UNBIASED VARIABLE SELECTION UNDER MINIMAX CONCAVE PENALTY," Annals of Statistics, 38, 894–942.

Zhang H, Conn AR, and Scheinberg K. (2010), "A derivative-free algorithm for least-squares minimization," SIAM Journal on Optimization, 20, 3555–3576.

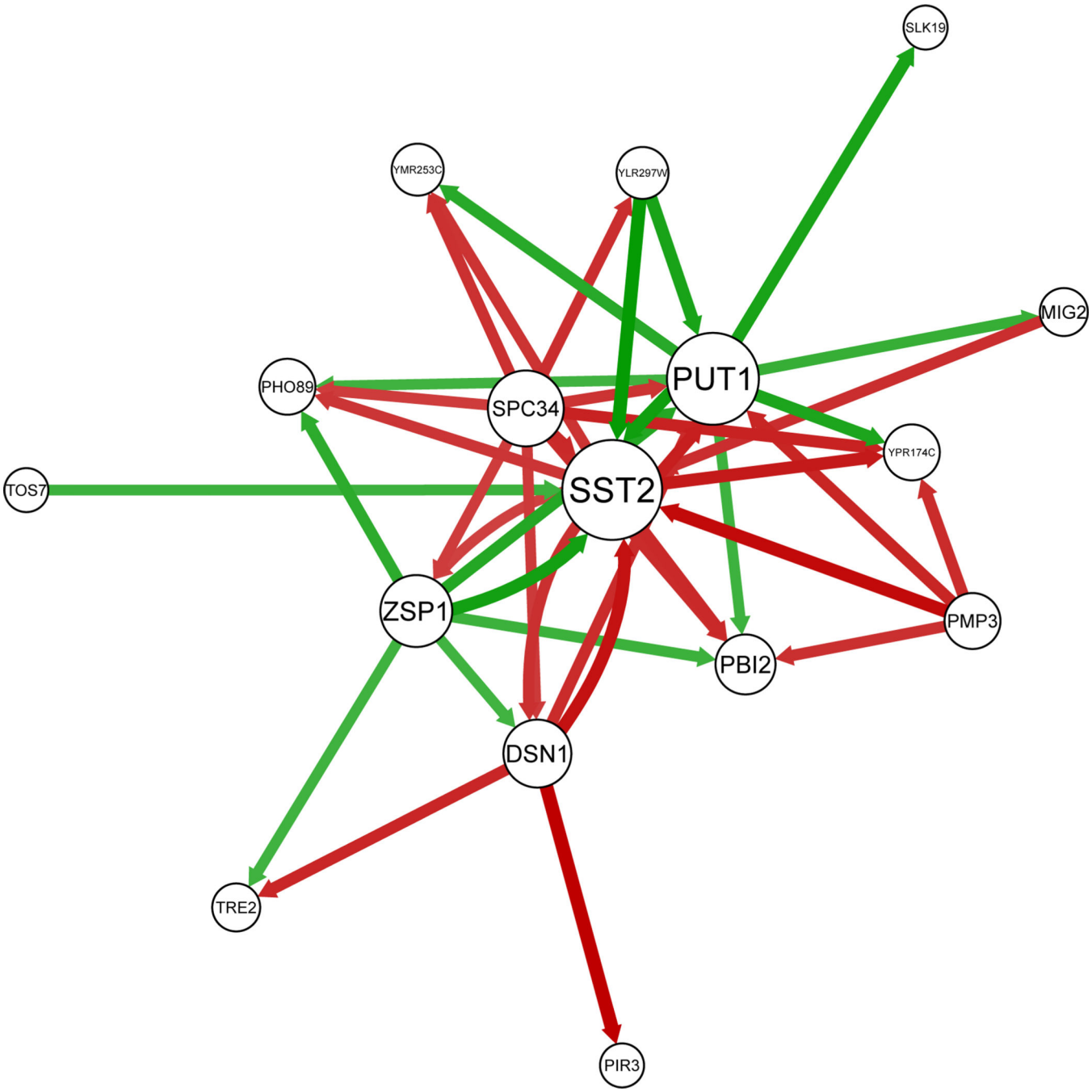

Figure 1:
Gene regulatory network reconstructed from the time-course yeast microarray data. Different sizes of nodes indicate network degree, which is defined as the number of adjacent edges of a node in the reconstructed network. Positive (negative) regulations are colored in green (red).

Table 1:

Parameter estimation comparisons between the direct-LS method and our new ST-SLS method. The computational cost (CPU time in second), relative RSS (RRSS) and REE are given as a percentage, based on the average of 1000 simulation runs (with the standard deviation in brackets) for most cases except for the Direct-LS method with $d = 100$ and the ST-SLS method with $d = 1000$, of which only 100 simulation runs were used due to the high computational cost.

		Direct-LS				ST-SLS			
Dimension(p)	**Noise(α)**	**Time in second**	**RRSS(%)**	**REE(A) (%)**	**REE(x_0) (%)**	**Time in second**	**RRSS(%)**	**REE(A) (%)**	**REE(x_0)(%)**
30	0	852	0	0	0	0.005	0	0	0
30	0.1	908(632)	30(12)	610(350)	43(15)	0.013(0.0005)	0.019(0.0018)	0.21(0.050)	0.020(0.0062)
30	0.3	866(349)	41(15)	1220(700)	58(31)	0.014(0.0008)	0.17(0.016)	2.1(0.64)	0.21(0.0056)
100	0	14982	0	0	0	1.0	0	0	0
100	0.1	16010(3972)	35(13)	760(300)	42(15)	1.2(0.49)	0.017(0.0054)	0.97(0.091)	0.023(0.0034)
100	0.3	55293(41769)	45(31)	1111(812)	73(51)	1.2(0.50)	0.17(0.0098)	1.7(0.99)	0.021(0.0030)
300	0	-	-	-	-	93	0	0	0
300	0.1	-	-	-	-	99(5.0)	0.017(0.0016)	2.3(0.41)	0.023(0.0019)
300	0.3	-	-	-	-	101(2.8)	0.015(0.0014)	4.0(0.74)	0.021(0.0017)
1000	0	-	-	-	-	7613	0	0	0
1000	0.1	-	-	-	-	8437(13)	0.018(0.0020)	8.8(2.9)	0.030(0.0094)
1000	0.3	-	-	-	-	8434(22)	0.015(0.0041)	20.6(16.7)	0.020(0.0069)

Table 2:

ODE variable selection comparisons between the 2-stage method and our new ST-SLS-VS algorithm. Both sensitivity (SEN) and specificity (SPE) are given as a percentage based on 1000 simulation runs for the cases d = 30,100,300 and 100 simulation runs for the case d = 1000 with the standard deviation in brackets.

		2-Stage Method		ST-SLS-VS(LASSO)	
Dimension(d)	**Noise(α)**	**SEN%**	**SPE%**	**SEN%**	**SPE%**
30	0	91.7	98.8	100	100
30	0.1	93.3(1.1)	63.6(1.3)	100(0)	100(0)
30	0.3	85.6(3.8)	43.5(1.5)	100(0)	100(0)
100	0	94	99.5	100	100
100	0.1	94.7(1.3)	67.3(0.3)	100(0)	100(0)
100	0.3	89.8(1.3)	65.9(0.2)	100(0)	99(0.1)
300	0	96.3	99.7	100	100
300	0.1	93.9(0.1)	62.1(0.2)	100(0)	99(0.1)
300	0.3	88.3(0.5)	62.5(0.2)	91(0.9)	98(0.1)
1000	0	98.0	99.9	100	100
1000	0.1	90.8(2.7)	60.2(1.5)	85(3.9)	99(0.1)
1000	0.3	85.1(4.3)	53.3(6.2)	80(5.5)	97(0.1)

Table 3:

ODE variable selection comparisons between the 2-stage method and our new ST-SLS-VS algorithm. Both sensitivity (SEN) and specificity (SPE) are given as a percentage based on 1000 simulation runs for the cases d = 30,100,300 and 100 simulation runs for the case d = 1000 with the standard deviation in brackets.

		ST-SLS-VS(SCAD)		ST-SLS-VS(MCP)	
Dimension(d)	**Noise(α)**	**SEN%**	**SPE%**	**SEN%**	**SPE%**
30	0	100	100	100	100
30	0.1	100(0)	100(0)	100(0)	100(0)
30	0.3	100(0)	100(0)	100(0)	100(0)
100	0	100	100	100	100
100	0.1	100(0)	100(0)	100(0)	100(0)
100	0.3	100(0)	100(0)	100(0)	100(0)
300	0	100	100	100	100
300	0.1	100(0)	99(0.1)	100(0)	99(0.1)
300	0.3	92(0.9)	98(0.1)	92(0.9)	98(0.1)
1000	0	100	100	100	100
1000	0.1	87(4.1)	98(0.1)	86(4.3)	98(0.1)
1000	0.3	81(5.4)	93(0.8)	82(5.4)	93(0.9)

Table 4:

Top ten most influential companies ranked by the network degree, which is defined as the number of adjacent edges of a node in the reconstructed network.

	Company	Category	Network Degree
TDS	Telephone & Data Systems Inc	IT Infrastructure	1162
UVE	Universal Insurance Holdings Inc	Financial	696
AKRX	Akorn Inc	Energy	648
NDAQ	Nasdaq OMX Group/The	Financial	613
GIS	General Mills Inc	Food	589
ABAX	Abaxis Inc	Healthcare	552
CMTL	Comtech Telecommunications	IT Infrastructure	548
NTGR	Netgear Inc	IT Infrastructure	545
NTCT	Netscout Systems Inc	IT Infrastructure	530
SBRA	Sabra Health Care REIT	Financial/Healthcare	515

Table 5:

Most influential companies in each sector defined by the network degree, which is the number of its adjacent edges in the reconstructed network.

	Company	Sector	Degree
HAR	Harman Int'l Industries	Consumer Discretionary	20
AMZN	Amazon.com Inc	Consumer Discretionary	16
WYNN	Wynn Resorts Ltd	Consumer Discretionary	13
GIS	General Mills	Consumer Staples	8
ADM	Archer-Daniels-Midland Co	Consumer Staples	7
CHK	Chesapeake Energy	Energy	9
DO	Diamond Offshore Drilling	Energy	7
RRC	Range Resources Corp.	Energy	7
NDAQ	NASDAQ OMX Group	Financials	43
ACE	ACE Limited	Financials	19
FITB	Fifth Third Bancorp	Financials	16
TMK	Torchmark Corp.	Financials	12
BXP	Boston Properties	Financials	10
PNC	PNC Financial Services	Financials	10
ENDP	Endo International	Health Care	13
JNJ	Johnson & Johnson	Health Care	10
SYK	Stryker Corp.	Health Care	9
FLS	Flowserve Corporation	Industrials	20
FLIR	FLIR Systems	Industrials	16
ITW	Illinois Tool Works	Industrials	14
NFLX	Netflix Inc.	Information Technology	11
PAYX	Paychex Inc.	Information Technology	11
BLL	Ball Corp	Materials	9
T	AT&T Inc	Telecommunications Services	3
AEE	Ameren Corp	Utilities	5
NEE	NextEra Energy	Utilities	5
SCG	SCANA Corp	Utilities	5