

Original Research Article

Constrained Stochastic Space Search Method for Parameter Estimation in Biological Networks

ABSTRACT

Parameter estimation is an important part of computational systems biology – especially in studies on biological networks. Numerous stochastic search methods have been applied in parameter estimation in biological networks. In this paper, a constrained stochastic space search (CSSS) method for parameter estimation is proposed and evaluated for estimating the parameters of a genetic network described by differential equations. Both linear and nonlinear model formalisms were used for the data evaluation. The performance of the CSSS method was compared to the Integrated Controlled Random Search for Dynamic Systems (ICRS/DS) stochastic optimization algorithm. Compared to the ICRS/DS, the CSSS algorithm is faster with at least a 7-fold shorter convergence time. Independent replicates were run and identification performed. For the same initialization conditions prior to optimization, the CSSS had on average smaller relative mean errors than the ICRS/DS.

Keywords: Biological network; differential equations; optimization; parameter estimation.

1. INTRODUCTION

There is growing interest in parameter estimation and understanding the dynamics of biological networks. Numerous methods have been used so far to gain insight into various aspects of biological networks, e.g. (probabilistic) Bayesian networks, regression methods, Boolean methods, and differential equations. The ordinary differential equation (ODE) formalism is a popular method to model biological networks[1][2][3][4][5]. This formalism is used to model the regulatory status of genes, hence, rendering ODE suitable for describing gene regulation activity [6]. Generally, many parameters are required in ODEs to describe the dynamics and regulatory roles of various components of a biological network. In principle, the number of parameters increases with the level of details required to describe the regulatory mechanisms involved in a particular biological network. There is currently numerous global parameter estimation methods proposed in computational systems biology [7][8][9][10][11]. There are many space search methods that address parameter estimation problems in biological networks and the number keeps rising[12].

Although much progress has been made in developing powerful and accurate parameter estimation methods, such as that in [13][14], there is still need for improvement in the computational robustness and speed of existing methods. Many genetic network inverse modeling problems require robust parameter estimation routines to ensure that the parameters are obtained with high precision. The classical way to do this is by least squares estimation or maximum likelihood estimation is used [15]. Similarly, stochastic search based methods can be used – in which case a specified set of parameter constraints and heuristics

for initializations on some parameters for optimization are required [7][11][14]. The optimization is based on minimization of goal function to fit a given model to a dataset.

The challenge of accurately describing the dynamics and change in concentration of molecular species, e.g. messenger Ribose Nucleic Acid (mRNA) in the cell. In Computational Systems Biology, Bioinformatics and Biotechnology, parameter estimation remains a big challenge – especially in the study of biological networks. To help address this challenge, various methods have been proposed in literature; e.g., Englezos and Kalogerakis[16] provided insight into optimization methods used for parameter estimation. Currently, there are numerous approaches to estimating parameters from data. Fundamentally, the mathematical principles and problem formulation vary. Additionally, there exist differences between the most methods, for instance: (i) some methods are computationally more expensive, (ii) variation in levels of accuracy and precision in prediction of state variables and/or models parameters, and (iii) the scalability to larger dimensional networks (or compartmental systems). Even small functional modules in networks require a large number of parameters to describe the underlying kinetics and regulatory mechanisms between the molecular units (genes and/or proteins). This large number of parameters is required for accurate network reconstruction irrespective of the model formalism.

Here, the constrained stochastic space search (CSSS) method is proposed and its performance is compared to the Integrated Controlled Random Search for Dynamic Systems (ICRS/DS) which can be found in [17]. The ICRS/DS is a modification of the ICRS algorithm[18]. Much as the ICRS/DS is robust, it comes with weaknesses, e.g.: (i) it requires making heuristic guesses for the direction search tuning parameters, and (ii) it takes a long time to converge to an optimum solution, if a solution exists in the search space and the optimization tolerance level is sufficiently stringent. The CSSS method addresses these issues by using a technique of variance scaling on the parameters during optimization. The performances of these methods were validated on two test case networks through *in silico* experiments and optimization.

Other stochastic search algorithms such as the Genetic Algorithms and Particle (Swarm) Optimization have also been applied to solve parameter estimation in biochemical systems (see Bomezzi et al.[19] and Yang et al. [20]. Unlike the CSSS method, many of the stochastic algorithms in literature are quite complex and require curation of initial parameter values prior to optimization. Extensive insight into the performance of commonly used algorithms for parameter estimation in Systems Biology is found in the work of Ashyraliyev et al. [21]. The CSSS algorithm aims to get the best parameter estimates by optimizing the sum of squared errors (SSE) goal function. Just like the ICRS/DS, the CSSS algorithm requires pre-specification of constraints on the parameter for initialization. The parameter precision and accuracy levels were found to be influenced by network topology and the existence of a feedback and/or feed-forward loop. For instance, the existence of edges (which represents the regulatory strength between a target gene and its transcription factor) in a network being significantly stronger than others may cause a drastic shift in the system condition number - ultimately affecting how accurately parameter are estimated.

2. METHODS

2.1 Network representation

Consider a genetic network system that is represented by the differential equation

$$\dot{X} = f(X, u; \Omega); \quad Y = g(X, \varepsilon) \quad (1)$$

where $X = [x_1, \dots, x_n]^T \in \mathbb{R}^{N \times n}$ represents the state of the system which details the gene activities. Here $y_i = [y_{i1}, \dots, y_{iN}]^T \in \mathbb{R}^{N \times 1}$ is the measured activity of gene i in time, N - number of time points; the matrix of measured data is $Y = [y_1, \dots, y_n]^T \in \mathbb{R}^{N \times n}$, Ω - parameter space, ε - Gaussian measurement error term and u - external perturbation vector. The transition function $f(\cdot)$ captures the network dynamics and $g(\cdot)$ is the output or mRNA measurement function which can be obtained by transcriptomics experiments.

2.1.1 Linear representation

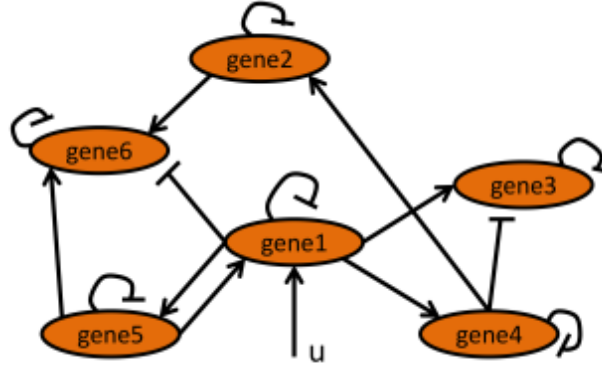
Let us consider the linear system representation:

$$\dot{X} = AX + bu \quad (2)$$

Here, A is the network connectivity matrix with typical attributes of real biological networks like activation, repression, auto-regulation (which is also often referred to as self-regulation) and feedback. The network is triggered using an external perturbation function:

$$u(t) = u(0)[1/(\beta + e^{Kt})] \quad (3)$$

β and K are nonnegative constants and $u(0)$ is the initial concentration of the triggering compound. The nature of $u(t)$ is based on many physical and biological systems in which the concentrations of an inducing compound is consumed[22][23].



98

99 **Fig. 1. A synthetic genetic network. The sharp headed ($\rightarrow$) and flat headed ($\dashv$) arrows**
 100 **respectively represent activation and repression of transcription.** The nodes
 101 **represent genes and u is the perturbation (input) signal, which can be nutrient uptake,**
 102 **glucose uptake etc depending the biological network under consideration.**

103 Consider a system of ODEs represented with the connectivity matrix constructed from the
 104 network in Figure 1. The fully connected matrix $A \in \mathbb{R}^{n \times n}$ that corresponds to this network
 105 has a total of $n^2 = 36$ parameters. However, not all the parameters need to be considered
 106 given that sparseness is a common attribute of biological networks [24][25]. Let $A(\Omega)$ denote
 107 the parameterized connectivity matrix which, in this case, is a sparse network with 16
 108 parameters – including the input vector coefficient b_1 . The diagonal entries θ_{ii} in (4) are the
 109 auto-regulation effects. The parameterized connectivity matrix is

$$A(\Omega) = \begin{pmatrix} -\theta_{11} & 0 & 0 & 0 & \theta_{15} & 0 \\ 0 & -\theta_{22} & 0 & \theta_{24} & 0 & 0 \\ \theta_{31} & 0 & -\theta_{33} & -\theta_{34} & 0 & 0 \\ \theta_{41} & 0 & 0 & -\theta_{44} & 0 & 0 \\ \theta_{51} & 0 & 0 & 0 & -\theta_{55} & 0 \\ -\theta_{61} & \theta_{62} & 0 & 0 & \theta_{65} & -\theta_{66} \end{pmatrix} \quad (4)$$

and the input vector

$$b = [b_1, 0, 0, 0, 0, 0]^T \quad (5)$$

The network is triggered by the signaling function in (3) as an external perturbation, therefore, enhancing the gene activities. For the ease of notations, let the parameter from (4) are represented by:

$$\omega = [\theta_{11}, \theta_{15}, \theta_{22}, \dots, \theta_{66}, b_1]^T \quad (6)$$

Several sets with different parameter values were assigned to the connectivity matrix to enable synthetic data generation. The initial parameters are depicted as:

$$\omega^{(0)} = [\theta_{11}^{(0)}, \theta_{15}^{(0)}, \theta_{22}^{(0)}, \dots, \theta_{66}^{(0)}, b_1^{(0)}]^T \quad (7)$$

The vector dimension $\dim(\omega^{(0)}) = m$ depends on the number of unknown parameters.

2.1.2 Boundary constraints on parameters

The restrictions imposed on the lower (L) and upper (U) parameter bounds are:

$$\Omega_i = \begin{cases} \theta_{ij}^L \leq \theta_{ij} \leq \theta_{ij}^U & \text{for all } i, j \\ b_1^L \leq b_1 \leq b_1^U \end{cases} \quad (8)$$

These constraints ensure that nonrealistic parameter values are avoided[26][27]. Once the model structure is known and the constraints on the parameters in (8) specified, then optimization of the goal function is initiated. The parameter values are updated during optimization until convergence at some optimal goal function value is obtained.

2.2 Nonlinear network representation

Consider the nonlinear model representation of a genetic network of three genes as in[28]. With a slight change of notation, the model is given by

$$\Gamma_{\text{nls}}(\Omega_i) = \begin{cases} \frac{dx_1}{dt} = k_{1s} \frac{1}{1+k_{13}x_3} - k_{1d}x_1 \\ \frac{dx_2}{dt} = k_{2s} \frac{k_{21}x_1}{1+k_{21}x_1} - k_{2d}x_2 \\ \frac{dx_3}{dt} = k_{3s} \frac{k_{31}x_1}{1+k_{31}x_1} \frac{k_{32}x_2}{1+k_{32}x_2} - k_{3d}x_3 \end{cases} \quad (9)$$

The term x_i represents the state for gene i . The equations in (9) describe a simple regulatory network in which gene 3 (x_3) represses transcription of gene 1 (x_1), gene 1 activates transcription of gene 2 (x_2), gene 1 (x_1) and gene 2 (x_2) collectively activate gene 3 (x_3). A

popular form of function used in modeling biological networks is the Hill function[29][30]. The parameter vector for this nonlinear system (nls) representation is given by

$$\omega_{\text{nls}} = [k_{1s}, k_{13}, k_{1d}, k_{2s}, k_{21}, k_{2d}, k_{3s}, k_{31}, k_{32}, k_{3d}]^T \quad (10)$$

In(9), the mRNA degradation is modeled as a first-order reaction and is represented by the constants k_{1d}, k_{2d} and k_{3d} ; The terms k_{1s}, k_{2s} and k_{3s} are the synthesis parameters and k_{ij} is the effective affinity constant for gene j activating i . The expression for the measured system outputs are as described in subsection 2.1. The constraints on the parameters in (9) were specified as:

$$\Omega_{i,\text{nls}} = k_{ij}^L \leq k_{ij} < k_{ij}^U \quad (11)$$

The dots in the parameter subscripts in (11) denote the corresponding indices in (10). This restriction and the initial guess $\omega_{\text{nls}}^{(0)}$ were used for the optimization.

2.3 Data generation and parameter estimation

The synthetic data was generated prior to parameter estimation. A comparison of the “true” and the estimated parameters was done. In the simulation, measurement noise was added to the data. The expression for gene i at time t is $y_i(t) = x_i(t) + \varepsilon_i(t)$, where $\varepsilon_i = \alpha \|x_i\|_1$ is the relative data measurement error, α is the measurement noise. A value of $\alpha = 0.05$ was considered at sampling moments in the *in silico* data simulations. The notation $\|\cdot\|_1$ represents the *City Block Distance* in the L_1 space. The SSE goal function is expressed as

$$J(\hat{\Omega}_i) = \arg \min \left(\xi_i^T(t|\hat{\Omega}_i) \xi_i(t|\hat{\Omega}_i) \right) \text{ for } i = 1, \dots, n \quad (12)$$

where $\xi_i(t|\hat{\Omega}_i) = y_i^{\text{obs}}(t) - y_i^{\text{est}}(t|\hat{\Omega}_i)$ is the residual vector, and $y_i^{\text{est}}(t|\hat{\Omega}_i)$ is the estimated expression for gene i conditioned on the parameter space.

2.4 Performance evaluation

A precision measure, relative errors were used on the individual parameters. The expression for computing the relative error on a parameter is given by

$$R_{\text{error},l} = |\omega_l^{\text{true}} - \hat{\omega}_l| / |\omega_l^{\text{true}}| \quad (13)$$

where ω_l^{true} and $\hat{\omega}_l$ represent the considerably “true” and estimated parameter values, respectively. Overall, two goodness-of-fit measures on parameter estimates were used, namely: the goal function and the relative error in parameter estimates. The mean and standard deviations of the relative errors in the vector $\hat{\omega}$ are given by (14) and (15), respectively.

$$\hat{\mu}_{R_{\text{error}}} = \frac{1}{p} \sum_{l=1}^p R_{\text{error},l}(\hat{\omega}) \quad (14)$$

$$\hat{\sigma}_{R_{\text{error}}} = \sqrt{\frac{1}{p-1} \sum_{l=1}^p (R_{\text{error},l}(\hat{\omega}) - \mu_{R_{\text{error}}})^2} \quad (15)$$

Here p is the number of independent identifications performed. Twenty different *in silico* experiments were performed (i.e. $p = 20$). Firstly, performing different experiments involved

169 making a nonnegative change $\delta\omega_l$ in the parameter vector (i.e. set $\omega_l := \omega_l \pm \delta\omega_l$) and then
 170 simulating the data – this was repeated p times. Secondly, independent identifications were
 171 performed using both the CSSS and ICRS/Ds methods. In each *in silico* experiment, a
 172 random number of parameters was kept constant while the others were varied around at
 173 most 20% of their starting values – this was repeated p times. The above different cases in
 174 data simulation and identification help to specify reliability of the optimization methods with
 175 respect to $\hat{\omega}$. Low values of $\bar{\mu}_{\text{Rerror}}$ are associated with a good performance, and low values
 176 of the mean relative standard errors $\hat{\sigma}_{\text{Rerror}}$ represent a high consistency in estimating ω . To
 177 assess the performance of the method over the parameter space, the average values,
 178 $\bar{\mu}_{\text{Rerror}}$ were computed using

$$179 \quad \hat{\mu}_{\text{Total}} = \frac{1}{m} \sum_{k=1}^m \hat{\mu}_{\text{Rerror},k} \quad (16)$$

180 This statistic is an unbiased estimator since $\bar{\mu}_{\text{Total}} \approx 0$. Similarly, the total variations in the
 181 estimate $\hat{\sigma}_{\text{Total}}$ were calculated using the expression

$$182 \quad \hat{\sigma}_{\text{Total}} = \sqrt{\frac{1}{m-1} \sum_{k=1}^m \left(\hat{\sigma}_{\text{Rerror},k} - E(\hat{\sigma}_{\text{Rerror}})_k \right)^2} \quad (17)$$

183 The mean standard deviation was calculated from

$$184 \quad \hat{\sigma}_{\text{mean}} = \frac{1}{m} \sum_{k=1}^m \hat{\sigma}_{\text{Rerror},k} \quad (18)$$

185 where $E(\cdot)$ is the expectation operator. Equation(16) and(18) provide additional information
 186 on the accuracy and consistency of the computed parameter estimates. A small confidence
 187 bound in(18) indicates replicable experiments with respect to the identification method used.
 188 The closer (16) gets to zero, the better the identification.

189 2.5 CSSS algorithm pseudo-code

190 The specifications of the subsequent steps in the CSSS algorithm are given in Table 1. The
 191 implementation and systems identification was done in MATLAB version 7.4.0. Setting an
 192 infinitesimal tolerance value $\delta_\varepsilon = 1e - 5$ (a relative deviation in goal functions) reduces the
 193 likelihood of the optimization getting stuck at a local minimum. This low tolerance value
 194 ensures that the error in the goal function is negligible, hence, increasing the accuracy in
 195 parameter estimation.

196 **Table 1. Pseudo-code for the CSSS algorithm. Here ℓ, l, s are the iteration indices, r :**
 197 **number of successful iterations satisfying the specified criterion, $\mathfrak{R}$: a definite integral**
 198 **domain, $\|\cdot\|_1$: the city block distance (or L_1 norm), $E(\cdot)$: is the expectation operator. $J^{(0)}$**
 199 **– initial value of goal function, $[y_1^{\text{est}}, \dots, y_n^{\text{est}}]^{(\ell)}$ and $[y_1^{\text{est}}, \dots, y_n^{\text{est}}]^{(\text{cg})}$ are the vectors of**
 200 **mRNA measurements (or model output for individual genes) during iteration and**
 201 **upon convergence, respectively. The superscripts “L” and “U” in ω^L and ω^U represent**
 202 **the lower and upper bounds, respectively. The other variables and parameters are as**
 203 **defined elsewhere in the manuscript.**

Require: $\{\omega^{(0)}, J^{(0)}, \Gamma(\Omega_i | Y_{N \times n})\}$
 Ensure: $\Gamma(\Omega_i)$ holds
 1: $\omega^{(0)}, J^{(0)}, \Gamma^{(0)}(\Omega_i); \ell \leftarrow 1$
 2: $J_\ell(\Omega_i), [y_1^{\text{est}}, \dots, y_n^{\text{est}}]^{(\ell)} \leftarrow \int_{\mathfrak{R}_\Omega \subset \mathfrak{R}} f(\cdot) d\mathfrak{R}_\Omega$
 3: $s \leftarrow 0$
 4: while $\{|J_{\ell+1}(\Omega_i) - J_\ell(\Omega_i)| / |J_\ell(\Omega_i)| > \delta_\varepsilon\}$ do
 5: $\omega^{(\ell+1)} \leftarrow \omega^{(\ell)}; s + 1 \leftarrow s$
 6: $\sigma^{2(\ell)} \leftarrow \min \left[E(\omega^U - \omega^{(\ell)})^2, E(\omega^{(\ell)} - \omega^L)^2 \right]$
 7: $l \leftarrow 0$
 8: while $s < 1$ do
 9: $r \leftarrow 0$
 10: while $(r < 1) = 1$ do
 11: $\omega^{(\ell+1)} \leftarrow \omega^{(\ell)} + \sigma^{2(\ell)} / \ell + 1; \sigma^{2(\ell)} \square \mathfrak{R}(0, I_m)$
 12: if $\|\omega^U - \omega^{(\ell)}\|_1 + \|\omega^{(\ell)} - \omega^L\|_1 \leq 2N$ then
 13: $r = 1; \omega^{(\ell+1)} \leftarrow \omega^{(\ell)}$
 14: end if
 15: end while
 16: $J_{\ell+1}(\Omega_i) \leftarrow \int_{\mathfrak{R}_\Omega \subset \mathfrak{R}} f(\cdot) d\mathfrak{R}_\Omega$
 17: if $J_{\ell+1}(\Omega_i) < J_\ell(\Omega_i)$ then
 18: $\ell + 1 \leftarrow \ell; \omega^{(\ell+1)} \leftarrow \omega^{(\ell)}$
 19: end if
 20: $\omega^{(cg)} \leftarrow \omega^{(\ell+1)}$
 21: end while
 22: end while
 23: return $\omega^{(cg)}, J^{(cg)}, [y_1^{\text{est}}, \dots, y_n^{\text{est}}]^{(cg)}, \ell + 1$

204

205 2.6 Pre-conditioner verification

206 The condition number, just like correlation coefficients, can be used to explain any
 207 anomalies in parameter estimates. If the condition number is large (i.e. deviates significantly
 208 from the value 1), then the gene expression values have similar time trajectories. This is an
 209 attribute that may lead to biased parameter estimates. Let the measured gene expression
 210 (mRNA levels) data be $Y_{N \times n}$, then the condition number for this matrix is defined as $\kappa(Y) :=$
 211 $\|Y\| \|Y^{-1}\|$ if $Y_{[N=n]}$ is a nonsingular matrix. If Y is a non-square matrix, then the analogous
 212 definition for the condition number is:

$$213 \quad \kappa(Y^T Y) := \kappa(Y^2) = \|(Y^T Y)\| \|(Y^T Y)^{-1}\| \quad (19)$$

214 3. RESULTS AND DISCUSSIONS

215 3.1 Model goodness-of-fit to data

On performing the identification, the results for the linear, the non-linear both for the CSSS and ICRD/DS methods all showed good model fits to the data. See, for instance Figure 2, which concerns results of the CSSS method for the non-linear case which involves three genes and a total of 10 parameters to be estimated. Individual data points such as those in Figure 2 represent averaged data values from duplicates, triplicates or more measurements – depending on the costs. The condition number for the linear and nonlinear network cases was found to be in the order of magnitude $\kappa(\cdot) = 1.7e + 4$ and $\kappa_{nls}(\cdot) = 1.2e + 4$, respectively. This order of magnitude in this number may compromise the accuracy of parameter estimates. It should be noted that by invoking a network by a single trigger most of the gene transcription profiles turn out to be correlated.

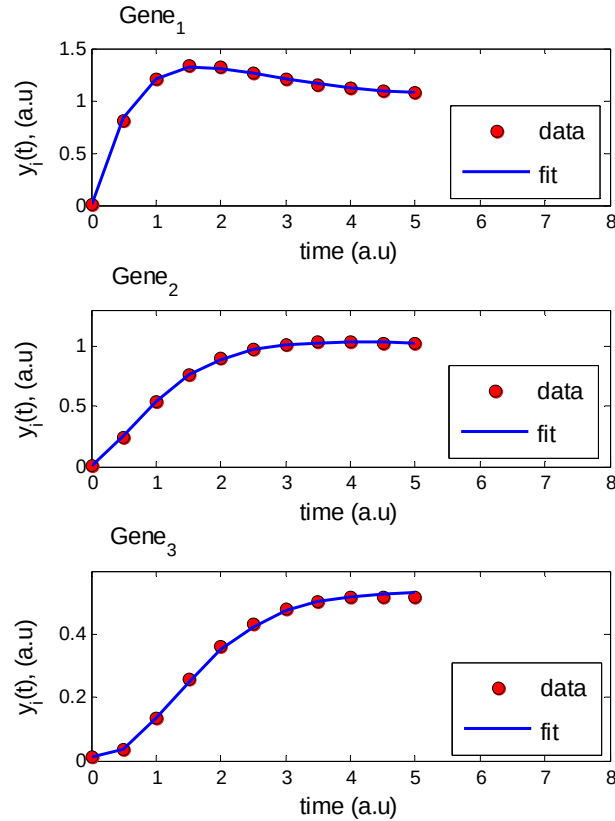


Fig. 2. Model fit to data using the CSSS method to the nonlinear example in section 2.2. The data points are indicated by points and the model fit by lines (measurement noise $\alpha = 0.05$, $N = 11$ data points and sampling step size of $\Delta t = 0.5$ units).

3.2 Identification from the linear model formalism

Parameter estimation with the CSSS and ICRS/DS algorithms is done for a 20 sets of parameters for the linear model formalism. The SSE (or goal function value) using the ICRS/DS algorithm was found to be $J_{ICRS/DS}^{(cg)} = 0.0191$, and $J_{CSSS}^{(cg)} = 0.0170$ for the CSSS method. These values indicate the lowest achievable goal function resulting from the optimization. It basically shows a marginally better performance for CSSS compared to ICR/DS (i.e. $J_{CSSS}^{(cg)} < J_{ICRS/DS}^{(cg)}$). Table 2 provides an example of the parameter estimates and the true parameters of one of the sets. For both methods there is a deviation from the true

values. The main reason is that a large number of parameters are estimated from a dataset with correlated response variables resulting in a high condition number. Identification with reduced system dimension (e.g. with gene 5 and 6 knocked out) shows that the CSSS method still yields results with lower variances on the relative errors. The overall relative error rates for the reduced system (with 9 parameters) were also lowered.

Table 2. Parameter estimates obtained by using the CSSS and ICRS/DS algorithm. Estimations were done using time course data from the linear network formalism (subsection 2.1) with measurement noise level of $\alpha = 0.05$ and $N = 20$ data points.

ω_{nls}	"true" value	ICRS/DS $\hat{\omega}_{\text{nls},i}^{(\text{cg})}$	CSSS $\hat{\omega}_{\text{nls},i}^{(\text{cg})}$
θ_{11}	2.0	0.6922	1.6589
θ_{15}	1.0	0.1398	0.6760
θ_{22}	2.1	2.0944	1.6817
θ_{24}	1.2	1.1965	0.9581
θ_{31}	2.5	2.7175	1.0710
θ_{33}	2.6	2.8466	1.1614
θ_{34}	1.6	1.7097	0.6087
θ_{41}	1.4	1.1515	0.6307
θ_{44}	2.9	2.3845	1.2947
θ_{51}	1.2	1.2235	0.9866
θ_{55}	1.1	1.1193	0.8993
θ_{61}	1.0	0.2683	0.3777
θ_{62}	2.6	2.4020	1.1511
θ_{65}	2.1	0.9770	1.6007
θ_{66}	2.1	1.5236	1.7599
b_1	1.0	0.5854	1.0011

246

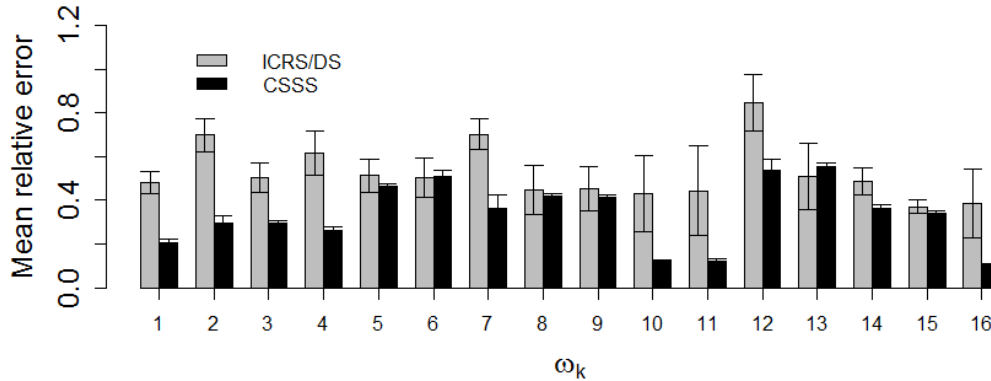
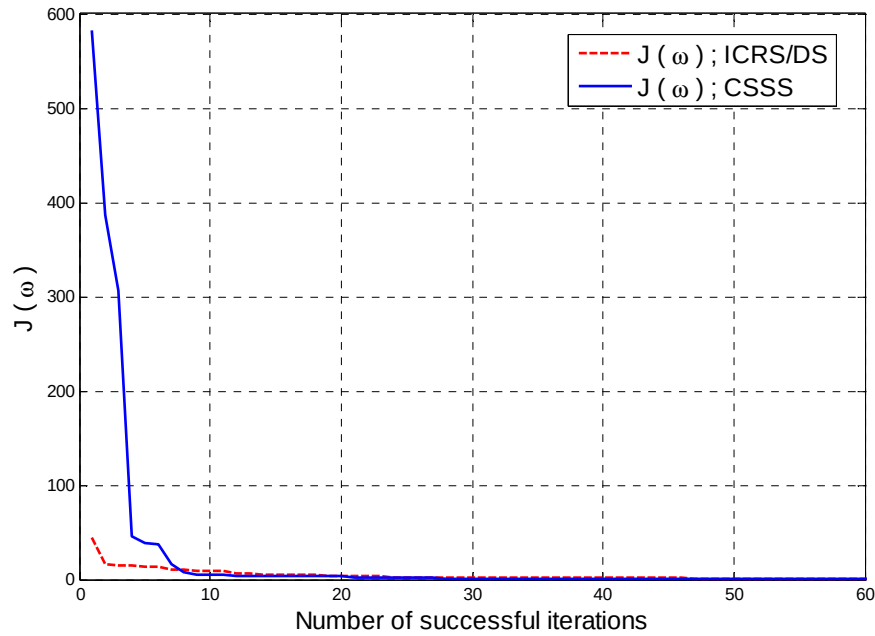


Fig. 3. Mean relative error (MRE) for the linear model formalism by the ICRS/DS and CSSS method, respectively. Analysis was done using dataset in which $\alpha = 0.05$. The error bars are the standard deviations ($\hat{\sigma}_{R_{\text{error},k}}$) to the MRE on parameter ω_k .

Figure 3 gives the mean relative error (MRE) for each parameter as estimated by both methods. Overall, the MRE for the CSSS method ($\bar{\mu}_{\text{Total}} = 0.4072$) is below that of the ICRS/DS ($\bar{\mu}_{\text{Total}} = 0.6001$). Moreover, an important property of the CSSS method is that the

254 MRE for an individual parameter is nearly constant while the MRE for the ICRS/DS shows a
 255 large variation (compare corresponding error bars). In other words the CSSS method is
 256 consistent in its results. The lower variation is a result of the variance scaling in step 11 of
 257 the algorithm. To estimate ω in the linear model set up, parameter estimation a choice of
 258 $\omega_l^{(0)} \in \{-1, +1\}$ was made. Hereby, -1 was chosen for the auto-regulation effects, -1 for the
 259 repressing and $+1$ for the activating effects of the transcription factors to a particular target
 260 gene. With this choice of initial parameter, the CSSS algorithm needs a significantly lower
 261 amount of iterations to achieve the convergence than ICRS/DS (Figure 3). If the initial
 262 parameters were not chosen according the above strategy, the convergence rate of both
 263 algorithms slightly worsens, but the performance of both algorithms remain invariantly
 264 comparable with CSSS marginally outperforming the ICRS/DS.



265 **Fig.4.** Plot of goal function convergence values using the CSSS and ICRS/DS
 266 algorithms. The figure axis narrowed for easy visibility – the final successful iteration
 267 number at which the goal function converges is about 150 and 900 for the ICRS/DS
 268 and CSSS, respectively.
 269

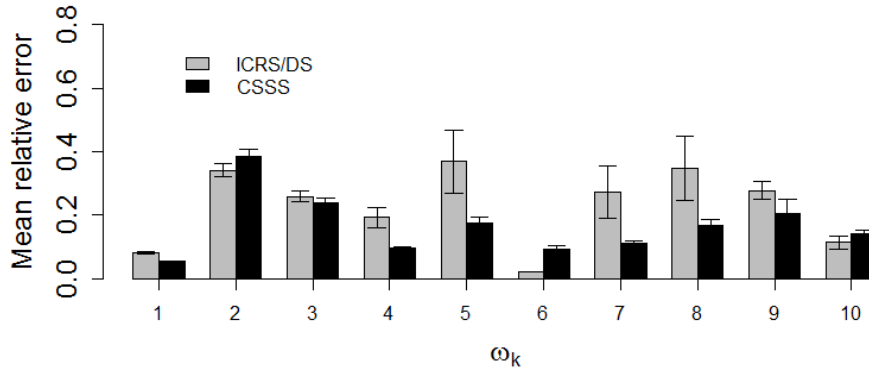
270 3.3 Identification from the nonlinear model formalism

271 The ODE nonlinear model formalism is known to be prone to having correlated parameters.
 272 This is problematic for network identification since it leads to reduced accuracy and precision
 273 in parameter estimates. Parameters were estimated for 20 different datasets. For the two
 274 methods being compared, the SSE values for the goal functions using the nonlinear
 275 identification problem were $J_{ICRS/DS}^{(cg)} = 0.0017$ and $J_{CSSS}^{(cg)} = 0.0014$. The parameters estimated
 276 for one of the datasets are shown in Table 3. Despite the correlation between parameters,
 277 for all these sets the parameters estimated are much closer to the true values compared to
 278 the linear model formalism. As expected, for the linear case with reduced number of
 279 parameters, better parameter estimates were observed.

280 **Table 3. Parameter estimates obtained by using the CSSS and ICRS/DS algorithm. The**
 281 **estimations were performed using with data having measurement noise $\alpha = 0.05$,**
 282 **$N = 11$ data points and sampling step size of $\Delta t = 0.5$ units.**

ω_{nls}	"true" value	ICRS/DS $\hat{\omega}_{\text{nls},i}^{(\text{cg})}$	CSSS $\hat{\omega}_{\text{nls},i}^{(\text{cg})}$
k_{1s}	2.0	2.0221	2.2348
k_{13}	2.0	1.4011	0.9060
k_{1d}	1.0	1.1755	1.5165
k_{2s}	2.0	1.4067	1.8152
k_{21}	1.0	1.8981	2.0154
k_{2d}	1.0	1.7361	1.8558
k_{3s}	2.0	3.1916	2.0154
k_{31}	1.0	0.2739	0.9287
k_{32}	1.0	4.0543	1.6019
k_{3d}	1.0	0.8022	0.7271

283 Figure 4 gives the MRE for the individual parameters. The average MRE for the CSSS is
 284 0.1670 with a deviation of $1.4e - 4$ and for the ICRS/DS method 0.2274 with a deviation of
 285 $1.5e - 3$. Again the CSSS is more consistent in finding the parameters. In contrast to the
 286 linear case no rules could be applied for the parameters initial guess, therefore initial values
 287 were randomly chosen from the interval $[0,5]$.

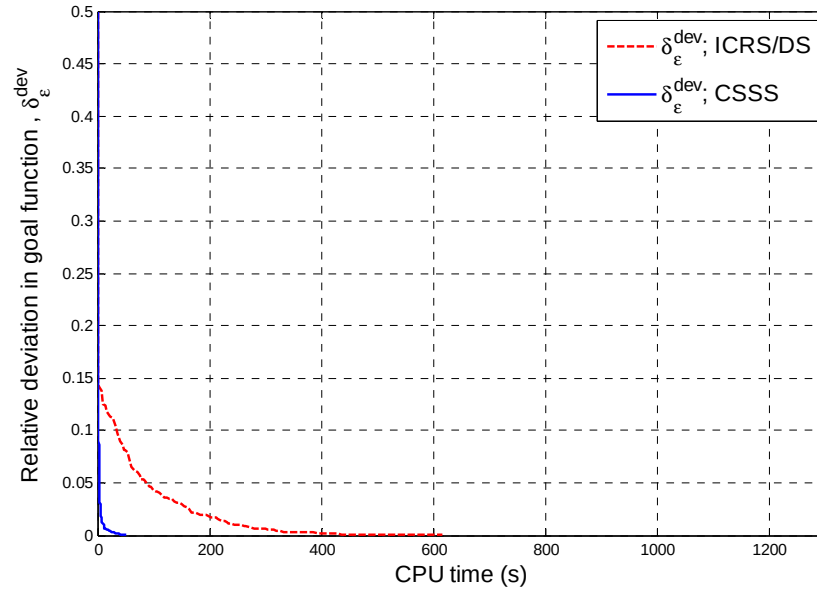


288
 289 **Fig. 5. Mean relative error (MRE) for the non-linear model formalism by the ICRS/DS**
 290 **and CSSS method, respectively. Analysis using data with a $\alpha = 0.05$ noise level. The**
 291 **error bars are the standard deviations ($\hat{\sigma}_{R_{\text{error},k}}$) to the MRE on parameter ω_k .**

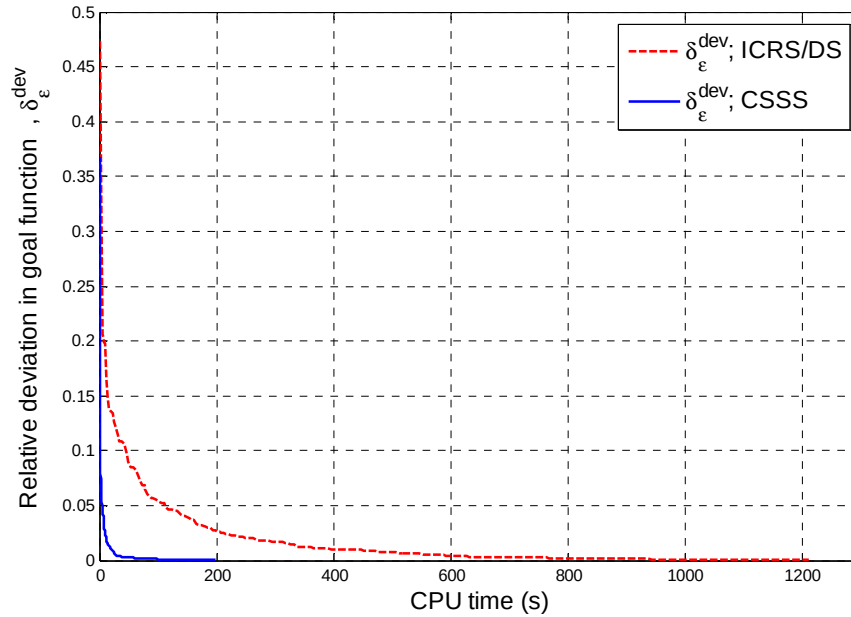
292 3.4 Comparison of algorithm performance: ICRS/DS versus CSSS

293 To assess the performance of the convergence times for the two identification methods
 294 under consideration, the optimization times to convergence were computed. For the same
 295 model and dataset, the CSSS algorithm converges faster than the ICRS/DS algorithm
 296 (Figure 7 and 8). This holds true even for a highly connected network (with $m = 16$
 297 parameters) used in the linear network as opposed to the nonlinear network with only three
 298 genes (with $m = 10$). During the optimization process, an iteration run was either rejected as
 299 unsuccessful or accepted as successful by checking if $\omega^{(\ell)}$ and $J^{(\ell)}$ satisfies the conditions in
 300 steps 12 and 17, respectively (Table 1). A distinguishing feature of stochastic optimization

301 methods is that much computational time is wasted on unsuccessful iterations. Compared to
 302 the ICRS/DS; steps 12 and 17 in the CSSS method ensure that the number of unsuccessful
 303 iterations is reduced.

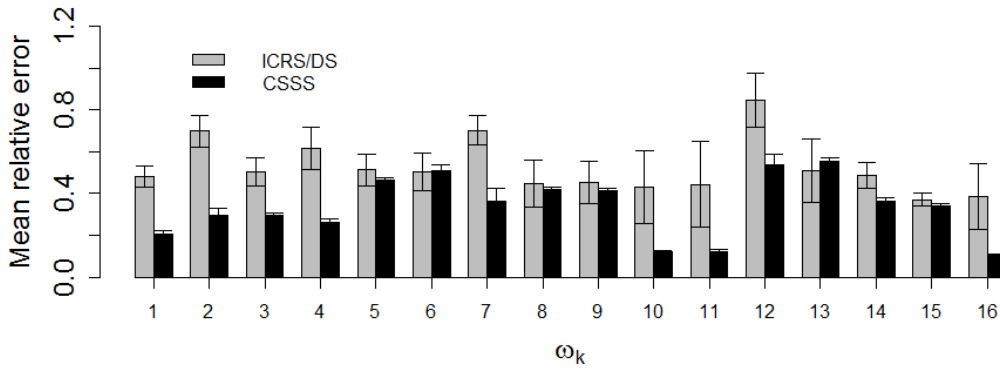


304 **Fig. 6. Comparison of relative deviations in goal functions $\delta_\epsilon^{\text{dev}}$ to convergence time**
 305 **for the CSSS algorithm versus with the ICRS/DS algorithm for the linear model**
 306 **formalism.**
 307



308 **Fig. 7. Comparison of relative deviations in goal function $\delta_\epsilon^{\text{dev}}$ to convergence time for**
 309 **the CSSS algorithm versus with the ICRS/DS algorithm for the non-linear model**
 310 **formalism.**
 311

312 An assessment on the parameter accuracy was performed with five times the number of
 313 time points for all the cases considered. A total of 100 data points were used for both the
 314 linear and nonlinear network representation problems. The results do not deviate much from
 315 those observed earlier-on in which the CSSS method was shown to be superior to the
 316 ICRS/DS method. Relative error analysis was performed with noise-free data and the results
 317 indicated that the standard deviations on the parameters are again lower for identification
 318 with the CSSS as compared to the ICRS/DS method. Therefore, irrespective of the
 319 measurement noise level in the dataset, the CSSS method is consistent and reliable in the
 320 parameter estimates – as seen from the low mean relative error and standard deviation
 321 values. This analysis indicates that the errors in the parameters are not caused by the noise
 322 level (see Figure 8) but are associated with the high condition number.



323 **Fig. 8. Mean relative error (MRE) for the linear model formalism by respectively the**
 324 **ICRS/DS and CSSS method, analysis with the noise-free dataset in which $\alpha = 0.0$. The**
 325 **error bars are the standard deviations ($\hat{\sigma}_{R_{error,k}}$) to the MRE on parameter ω_k .**
 326

327 Using the CSSS method, the averaged relative errors on the estimated parameters were
 328 less than those obtained using the ICRS/DS method (Figure 3 and 4). The CSSS method
 329 achieves the best accuracy results and is faster in convergence. The advantage of using the
 330 CSSS method over the ICRS/DS method is that the CSSS method does not require heuristic
 331 tuning of the parameters. In step 6: Table 1, the variance of the parameters is estimated and
 332 the minimum deviation from the two terms is used for updating the next parameter estimate.
 333 Scaling the variance term σ^2 by the iteration index $\ell + 1$ (step 11: Table 1) ensures that the
 334 variance decreases during optimization, leading to faster convergence (Figure 6 and 7). This
 335 increased speed of convergence is associated with less function rejection during the space
 336 search in the optimization process. In the presence of correlated parameters in a model, the
 337 likelihood of obtaining good parameter estimates is reduced because a change in one
 338 parameter value proportionally causes a shift in the other parameter value – this is a
 339 challenge in systems identification. In essence, applying a method to a parameter estimation
 340 problem for larger network dimensions may yield a model with good fits to the data, but with
 341 fading degrees of parameter accuracy. Quicker convergence and better parameter
 342 estimates can be achieved by good parameter initialization since this increases the
 343 likelihood of obtaining optimal solutions [31]. This statement seems relevant for the ICRS/DS
 344 method, but this work shows that with the fast converging CSSS algorithm the initial choice
 345 is not that important.

346 4. CONCLUSIONS

347 The results presented in this paper are based on two stochastic optimization methods
348 addressing the parameter identification problem from time course datasets. The findings
349 indicate that the performance of the CSSS is better than that of the ICRS/DS. The CSSS
350 method shows lower variation on relative error on the parameter estimates. The mean
351 relative errors on the parameters are also lower for the results obtained using the CSSS
352 method as opposed to those obtained from the ICRS/DS method. The CSSS method can be
353 used for parameter estimation in small sized networks. The CSSS method is computationally
354 efficient and out-performs the ICRS/DS in the convergence speed. The performance of the
355 proposed optimization method for both the linear and nonlinear model representation was
356 better than that of the ICRS/DS. Achieving accurate parameter identification remains one that
357 is crucial to address, especially with growing interests from understanding larger biological
358 networks, thereby, up-scaling the challenge. However, by decomposing networks into sub-
359 units, it is easier to estimate the parameters with much better precision.

360 COMPETING INTERESTS

361 The author declares no competing interests.

362 REFERENCES

- 363 1. Goodwin BC. Temporal Organization In Cells; A Dynamic Theory Of Cellular Control
364 Processes. Academic Press, New York. 1963
- 365 2. Hartemink AJ, Gifford DK, Jaakola TS, Young RA. Combining location and expression
366 data for principled discovery of genetic regulatory network models. Proc. Pac. Symp.
367 Biocomput. 2002;7:437–449
- 368 3. Zak D, Pearson RK, Vadigepalli R, Gonye G, Schwaber JS, Doyle III FJ. Continuous-time
369 identification of gene expression models. Omics. 2004;7:373–386
- 370 4. Savageau MA. Biochemical systems analysis: a study of function and design in molecular
371 biology. Addison Wesley Publishing Company. 1976
- 372 5. Kikuchi S, Tominaga D, Arita M, Takahashi K, Tomita M. Dynamic modeling of genetic
373 networks using genetic algorithm and S-system. Bioinformatics. 2003;19:643-650
- 374 6. Smolen P, Baxter DA, Byrne JH. Modeling transcriptional control in gene networks:
375 methods, recent results, and future directions. Bull. Math. Biol. 2000;62:247-292
- 376 7. Zwolak J, Tyson J, Watson L. Globally optimized parameters for a model of mitotic control
377 in frog egg extracts. IEE Proceedings Systems Biology. 2005;152:81-92
- 378 8. Chaitankar, V.P., Perkins EJ, Gong P, Deng Y, Zhang C. A novel gene network inference
379 algorithm using predictive minimum description length approach. BMC Syst. Biol. 2010;4:S7
- 380 9. Tsai KY, Wang FS. Evolutionary optimization with data collocation for reverse engineering
381 of biological networks. Bioinformatics. 2005;21:1180-1188; doi:10.1093/bioinformatics/bti099

- 382 10. Ashyraliyev M, Jaeger J, Blom JG. Parameter estimation and determinability analysis
383 applied to *Drosophila* gap gene circuits. BMC Syst. Biol. 2008;2:83
- 384 11. Polisetty PK, Voit EO, Gatzke EP. Identification of metabolic system parameters using
385 global optimization methods. Theoretical Biology and Medical Modelling. 2006;3:4
- 386 12. Torres NV, Voit EO. Pathway analysis and optimization in metabolic engineering. New
387 York, Cambridge University Press. 2002
- 388 13. Espocito WR, Floudas CA. Global optimization for the parameter estimation of
389 differential-algebraic systems. Ind. Eng. Chem. Res. 2000;39:1291-1310.
- 390 14. Papamichail I, Adjiman CS. A rigorous global optimization algorithm for problems with
391 ordinary differential equations. J. Global Optim. 2002;24:1-33
- 392 15. Li Z, Osborne MR, Prvan T. Parameter estimation of ordinary differential equations. IMA
393 J. Numer. Anal. 2005;25:264-285
- 394 16. P E, N K. Applied Parameter Estimation for Chemical Engineers. 1st Edition ed. CRC
395 Press: USA; 2000
- 396 17. Carrasco EF, Banga JR. Dynamic optimization of batch reactors using adaptive
397 stochastic algorithms. Ind. Eng. Chem. Res. 1997;36:2252-2261
- 398 18. Banga JR, Casares JJ. Integrated Controlled Random Search: application to a
399 wastewater treatment plant model. Inst. Chem. Eng. Symp. Ser. 1987;100:183
- 400 19. Besozzi D, Cazzaniga P, Mauri G, Pescini D, Vanneschi L. A Comparison of Genetic
401 Algorithms and Particle Swarm Optimization for Parameter Estimation in Stochastic
402 Biochemical Systems. Lecture Notes in Computer Science. 2009;5483:116-127
- 403 20. Yang F, Zhang C, Sun T. Comparison of Particle Swarm Optimization and Genetic
404 Algorithm for HMM training. ICPR 2008 19th International Conference on Pattern
405 Recognition IEEE Xplore Digital Library. 2008
- 406 21. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology:
407 parameter estimation for biochemical models. FEBS J. 2009;276:886-902;
408 doi:10.1111/j.1742-4658.2008.06844.x; 10.1111/j.1742-4658.2008.06844.x
- 409 22. Eungdamrong NJ, Iyengar R. Modeling cell signaling networks. Biol. Cell. 2004;96:355-
410 62
- 411 23. Omony J, de Graaff LH, van Straten G, van Boxtel AJB. Modeling and analysis of the
412 dynamic behavior of the XlnR regulon in *Aspergillus niger*. BMC Syst. Biol. 2011;5:S14
- 413 24. Hoguland M, Frigyesi A, Mitelman K. A gene fusion network in human neoplasia.
414 Oncogene. 2006;25:2674-8
- 415 25. Nacher JC, Ochiai T. Power-law distribution of gene expression fluctuations. Physics
416 letter A. 2008;372:6202-6206

- 417 26. Tucker W, V M. Parameter reconstruction for biochemical networks using interval
418 analysis. *Reliable Comput.* 2006;12:389-402
- 419 27. Tucker W, Kotalik Z, V M. Estimating parameters for generalized mass action models
420 using constraint propagation. *Math. Biosci.* 2007;208:607-620
- 421 28. Karleback G, Samir R. Modeling and analysis of gene regulatory networks. *Molecular*
422 *cell biology.* 2008;9:770-780
- 423 29. Hill AV. The possible effect of the aggregation of the molecules of haemoglobin on its
424 dissociation curves. *J. Physiol.* 1910;40(Suppl.):4-7
- 425 30. Polynikis A, Hogan SJ, di Bernardo M. Comparing different ODE modeling approaches
426 of gene regulatory networks. *J. Theor. Biol.* 2009;261:511-530
- 427 31. Averick BM, Carter RG, Moré JJ. The MINPACK-2 test problem collection. 1991
- 428