
Convex Total Least Squares

Dmitry Malioutov

IBM Research, 1101 Kitchawan Road, Yorktown Heights, NY 10598 USA

DMALIOUTOV@US.IBM.COM

Nikolai Slavov

Departments of Physics and Biology, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

NSLAVOV@ALUM.MIT.EDU

Abstract

We study the total least squares (TLS) problem that generalizes least squares regression by allowing measurement errors in both dependent and independent variables. TLS is widely used in applied fields including computer vision, system identification and econometrics. The special case when all dependent and independent variables have the same level of uncorrelated Gaussian noise, known as ordinary TLS, can be solved by singular value decomposition (SVD). However, SVD cannot solve many important practical TLS problems with realistic noise structure, such as having varying measurement noise, known structure on the errors, or large outliers requiring robust error-norms. To solve such problems, we develop convex relaxation approaches for a general class of structured TLS (STLS). We show both theoretically and experimentally, that while the plain nuclear norm relaxation incurs large approximation errors for STLS, the re-weighted nuclear norm approach is very effective, and achieves better accuracy on challenging STLS problems than popular non-convex solvers. We describe a fast solution based on augmented Lagrangian formulation, and apply our approach to an important class of biological problems that use population average measurements to infer cell-type and physiological-state specific expression levels that are very hard to measure directly.

1. Introduction

Total least squares is a powerful generalization of ordinary least squares (LS) which allows errors in the measured explanatory variables (Golub & Van Loan, 1980). It has become an indispensable tool in a variety of disciplines including chemometrics, system identification, astronomy, computer vision, and econometrics (Markovsky & Van Huffel, 2007). Consider a least squares problem $\mathbf{y} \approx X\boldsymbol{\beta}$, where we would like to find coefficients $\boldsymbol{\beta}$ to best predict the target vector \mathbf{y} based on measured variables X . The usual assumption is that X is known exactly, and that the errors come from i.i.d. additive Gaussian noise \mathbf{n} : $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{n}$. The LS problem has a simple closed-form solution by minimizing $\|\mathbf{y} - X\boldsymbol{\beta}\|_2^2$ with respect to $\boldsymbol{\beta}$. In many applications not only \mathbf{y} but also X is known only approximately, $X = X_0 + E_x$, where X_0 are the uncorrupted values, and E_x are the unknown errors in observed variables. The total least squares (TLS) formulation, or errors in variables regression, tries to jointly minimize errors in \mathbf{y} and in X (ℓ_2 -norm of \mathbf{n} and Frobenius norm of E_x):

$$\min_{\mathbf{n}, E_x, \boldsymbol{\beta}} \|\mathbf{n}\|_2^2 + \|E_x\|_F^2 \text{ where } \mathbf{y} = (X - E_x)\boldsymbol{\beta} + \mathbf{n} \quad (1)$$

While the optimization problem in this form is not convex, it can in fact be reformulated as finding the closest rank-deficient matrix to a given matrix, and solved in closed form via the singular value decomposition (SVD) (Golub & Van Loan, 1980).

Many error-in-variables problems of practical interest have additional information: for example, a subset of the entries in X may be known exactly, we may know different entries with varying accuracy, and in general X may exhibit a certain structure, e.g. block-diagonal, Toeplitz, or Hankel in system identification literature (Markovsky et al., 2005). Furthermore, it is often important to use an error-norm robust to outliers, e.g. Huber loss or ℓ_1 -loss. Unfortunately, with rare exceptions¹, none of these problems allow an efficient solution, and the state of the art approach is to solve them

¹A closed form solution exists when subsets of columns are

by local optimization methods (Markovsky & Usevich, 2014; Zhu et al., 2011; Srebro & Jaakkola, 2003). The only available guarantee is typically the ability to reach a stationary point of the non-convex objective.

In this paper we propose a principled formulation for STLS based on convex relaxations of matrix rank. Our approach uses the re-weighted nuclear norm relaxation (Fazel et al., 2001) and is highly flexible: it can handle very general linear structure on errors, including arbitrary weights (changing noise for different entries), patterns of observed and unobserved errors, Toeplitz and Hankel structures, and even norms other than the Frobenius norm. The nuclear norm relaxation has been successfully used for a range of machine learning problems involving rank constraints, including low-rank matrix completion, low-order system approximation, and robust PCA (Cai et al., 2010; Chandrasekaran et al., 2011). The STLS problem is conceptually different in that we do not seek low-rank solutions, but on the contrary nearly full-rank solutions. We show both theoretically and experimentally that while the plain nuclear norm formulation incurs large approximation errors, these can be dramatically improved by using the re-weighted nuclear norm. We suggest fast first-order methods based on Augmented Lagrangian multipliers (Bertsekas, 1982) to compute the STLS solution. As part of ALM we derive new updates for the re-weighted nuclear-norm based on solving the Sylvester’s equation, which can also be used for many other machine learning tasks relying on matrix-rank, including matrix completion and robust PCA.

As a case study of our approach to STLS we consider an important application in biology, quantification of cellular heterogeneity (Slavov & Botstein, 2011). We develop a new representation for the problem as a large structured linear system, and extend it to handle noise by a structured TLS problem with block-diagonal error structure. Experiments demonstrate the effectiveness of STLS in recovering physiological-state specific expression levels from aggregate measurements.

1.1. Total Least Squares

We first review the solution of ordinary TLS problems. We simplify the notation from (1): combining our noisy data X and \mathbf{y} into one matrix, $\bar{A} \triangleq [X \ -\mathbf{y}]$, and the errors into $E \triangleq [E_x \ -\mathbf{n}]$ we have

$$\min \|E\|_F^2 \quad \text{where } (\bar{A} - E) \begin{bmatrix} \beta \\ 1 \end{bmatrix} = 0. \quad (2)$$

The matrix \bar{A} is in general full-rank, and a solution can be obtained by finding a rank-deficient matrix closest fully known; a Fourier transform based approach can handle block-circulant errors E_x (Beck & Ben-Tal, 2005).

to \bar{A} in terms of the Frobenius norm. This finds smallest errors E_x and \mathbf{n} such that $\mathbf{y} + \mathbf{n}$ is in the range space of $X - E_x$. The closest rank-deficient matrix is simply obtained by computing the SVD, $\bar{A} = USV^T$ and setting the smallest singular value to be zero.

Structured TLS problems (Markovsky & Van Huffel, 2007) allow more realistic errors E_x : with subsets of measurements that may be known exactly; weights reflecting different measurement noise for each entry; requiring linear structure of errors E_x such as Toeplitz that is crucial in deconvolution problems in signal processing. Unfortunately, the SVD does not apply to any of these more general versions of TLS (Srebro & Jaakkola, 2003; Markovsky & Van Huffel, 2007). Existing solutions to structured TLS problems formulate a non-convex optimization problem and attempt to solve it by local optimization (Markovsky & Usevich, 2014) that suffers from local optima and lack of guarantees on accuracy. We follow a different route and use a convex relaxation for the STLS problem.

2. STLS via a nuclear norm relaxation

The STLS problem in a general form can be described as follows (Markovsky & Van Huffel, 2007). Using the notation in Section 1.1, suppose our observed matrix \bar{A} is $M \times N$ with full column rank. We aim to find a nearby rank-deficient matrix A , $\text{rank}(A) \leq N - 1$, where the errors E have a certain linear structure:

$$\min \|W \odot E\|_F^2, \quad \text{where } \text{rank}(A) \leq N - 1 \\ A = \bar{A} - E, \quad \text{and } \mathcal{L}(E) = \mathbf{b} \quad (3)$$

The key components here are the linear equalities that E has to satisfy, $\mathcal{L}(E) = \mathbf{b}$. This notation represents a set of linear constraints $\text{tr}(L_i^T E) = b_i$, for $i = 1, \dots, J$. In our application to cell heterogeneity quantification these constraints correspond to knowing certain entries of A exactly, i.e. $E_{ij} = 0$ for some subset of entries, while other entries vary freely. One may require other linear structure such as Toeplitz or Hankel. We also allow an element-wise weighting $W \odot E$, with $W_{i,j} \geq 0$ on the errors, as some observations may be measured with higher accuracy than others. Finally, while we focus on the Frobenius norm of the error, any other convex error metric, for example, mean absolute error, or robust Huber loss, could be used instead. The main difficulty in the formulation is posed by the non-convex rank constraint. The STLS problem is a special case of the structured low-rank approximation problem, where rank is exactly $N - 1$ (Markovsky & Usevich, 2014). Next, we propose a tractable formulation for STLS based on convex relaxations of matrix rank.

We start by formulating the nuclear-norm relaxation for TLS and then improve upon it by using the re-

weighted nuclear norm. The nuclear norm $\|A\|_*$ is a popular relaxation used to convexify rank constraints (Cai et al., 2010), and it is defined as the sum of the singular values of the matrix A , i.e. $\|A\|_* = \sum_i \sigma_i(A)$. It can be viewed as the ℓ_1 -norm of the singular value spectrum² favoring few non-zero singular values, i.e., matrices with low-rank. Our initial nuclear norm relaxation for the STLS problem is:

$$\begin{aligned} \min \|A\|_* + \alpha \|W \odot E\|_F^2 \text{ such that} \\ A = \bar{A} - E, \text{ and } \mathcal{L}(E) = \mathbf{b} \end{aligned} \quad (4)$$

The parameter α balances error residuals vs. the nuclear norm (proxy for rank). We chose the largest α , i.e. smallest nuclear norm penalty, that still produces $\text{rank}(A) \leq N - 1$. This can be achieved by a simple binary search over α . In contrast to matrix completion and robust PCA, the STLS problem aims to find almost fully dense solutions with rank $N - 1$, so it requires different analysis tools. We present theoretical analysis specifically for the STLS problem in Section 4. Next, we describe the re-weighted nuclear norm, which, as we show in Section 4, is better suited for the STLS problem than the plain nuclear norm.

2.1. Reweighted nuclear norm and the log-determinant heuristic for rank

A very effective improvement of the nuclear norm comes from re-weighting it (Fazel et al., 2001; Mohan & Fazel, 2010) based on the log-determinant heuristic for rank. To motivate it, we first describe a closely related approach in the vector case (where instead of searching for low-rank matrices one would like to find sparse vectors). Suppose that we seek a sparse solution to a general convex optimization problem. A popular approach penalizes the ℓ_1 -norm of the solution \mathbf{x} $\|\mathbf{x}\|_1 = \sum_i |x_i|$ to encourage sparse solutions. A dramatic improvement in finding sparse signals can be obtained simply by using the weighted ℓ_1 -norm, i.e. $\sum_i w_i |x_i|$ with suitable positive weights w_i (Candes et al., 2008) instead of a plain ℓ_1 -norm. Ideally the weights would be based on the unknown signal, to provide a closer approximation to sparsity (ℓ_0 -norm) by penalizing large elements less than small ones. A practical solution first solves a problem involving the unweighted ℓ_1 -norm, and uses the solution $\hat{\mathbf{x}}$ to define the weights $w_i = \frac{1}{\delta + |\hat{x}_i|}$, with δ a small positive constant. This iterative approach can be seen as an iterative local linearization of the concave log-penalty for sparsity, $\sum_i \log(\delta + |x_i|)$ (Fazel et al., 2001; Candes et al., 2008). In both empirical and emerging theoretical studies (Needell, 2009; Khajehnejad et al., 2009)

²For diagonal matrices A the nuclear norm is exactly equivalent to the ℓ_1 -norm of the diagonal elements.

re-weighting the ℓ_1 -norm has been shown to provide a tighter relaxation of sparsity.

In a similar way, the re-weighted nuclear norm tries to penalize large singular values less than small ones by introducing positive weights. There is an analogous direct connection to the iterative linearization for the concave log-det relaxation of rank (Mohan & Fazel, 2010). Recall that the problem of minimizing the nuclear norm subject to convex set constraints \mathcal{C} ,

$$\min \|A\|_* \text{ such that } A \in \mathcal{C}, \quad (5)$$

has a semi-definite programming (SDP) representation (Fazel et al., 2001). Introducing auxiliary symmetric p.s.d. matrix variables $Y, Z \succeq 0$, we rewrite it as:

$$\min_{A,Y,Z} \text{tr}(Y) + \text{tr}(Z) \quad \text{s.t.} \begin{bmatrix} Y & A \\ A^T & Z \end{bmatrix} \succeq 0, \quad A \in \mathcal{C} \quad (6)$$

Instead of using the convex nuclear norm relaxation, it has been suggested to use the concave log-det approximation to rank:

$$\begin{aligned} \min_{A,Y,Z} \log \det(Y + \delta I) + \log \det(Z + \delta I) \\ \text{s.t.} \begin{bmatrix} Y & A \\ A^T & Z \end{bmatrix} \succeq 0, \quad A \in \mathcal{C} \end{aligned} \quad (7)$$

Here I is the identity matrix and δ is a small positive constant. The log-det relaxation provides a closer approximation to rank than the nuclear norm, but it is more challenging to optimize. By iteratively linearizing this objective one obtains a sequence of weighted nuclear-norm problems (Mohan & Fazel, 2010):

$$\begin{aligned} \min_{A,Y,Z} \text{tr}((Y^k + \delta I)^{-1}Y) + \text{tr}((Z^k + \delta I)^{-1}Z) \\ \text{s.t.} \begin{bmatrix} Y & A \\ A^T & Z \end{bmatrix} \succeq 0, \quad A \in \mathcal{C} \end{aligned} \quad (8)$$

where Y^k, Z^k are obtained from the previous iteration, and Y^0, Z^0 are initialized as I . Let $W_1^k = (Y^k + \delta I)^{-1/2}$ and $W_2^k = (Z^k + \delta I)^{-1/2}$ then the problem is equivalent to a weighted nuclear norm optimization in each iteration k :

$$\min_{A,Y,Z} \|W_1^k A W_2^k\|_* \quad \text{s.t. } A \in \mathcal{C} \quad (9)$$

The re-weighted nuclear norm approach iteratively solves convex weighted nuclear norm problems in (9):

Re-weighted nuclear norm algorithm:

Initialize: $k = 0, W_1^0 = W_2^0 = I$.

- (1) Solve the weighted NN problem in (9) to get A^{k+1} .
- (2) Compute the SVD: $W_1^k A^{k+1} W_2^k = U \Sigma V^T$, and set $Y^{k+1} = (W_1^k)^{-1} U \Sigma U^T (W_1^k)^{-1}$ and $Z^{k+1} = (W_2^k)^{-1} V \Sigma V^T (W_2^k)^{-1}$.

(3) Set $W_1^k = (Y^k + \delta I)^{-1/2}$ and $W_2^k = (Z^k + \delta I)^{-1/2}$.

There are various ways to solve the plain and weighted nuclear norm STLS formulations, including interior-point methods (Toh et al., 1999) and iterative thresholding (Cai et al., 2010). In the next section we focus on augmented Lagrangian methods (ALM) (Bertsekas, 1982) which allow fast convergence without using computationally expensive second-order information.

3. Fast computation via ALM

While the weighted nuclear norm problem in (9) can be solved via an interior point method, it is computationally expensive even for modest size data because of the need to compute Hessians. We develop an effective first-order approach for STLS based on the augmented Lagrangian multiplier (ALM) method (Bertsekas, 1982; Lin et al., 2010). Consider a general equality constrained optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ such that } h(\mathbf{x}) = 0. \quad (10)$$

ALM first defines an augmented Lagrangian function:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \mu) = f(\mathbf{x}) + \boldsymbol{\lambda}^T h(\mathbf{x}) + \frac{\mu}{2} \|h(\mathbf{x})\|_2^2 \quad (11)$$

The augmented Lagrangian method alternates optimization over \mathbf{x} with updates of $\boldsymbol{\lambda}$ for an increasing sequence of μ_k . The motivation is that either if $\boldsymbol{\lambda}$ is near the optimal dual solution for (10), or, if μ is large enough, then the solution to (11) approaches the global minimum of (10). When f and h are both continuously differentiable, if μ_k is an increasing sequence, the solution converges Q -linearly to the optimal one (Bertsekas, 1982). The work of (Lin et al., 2010) extended the analysis to allow objective functions involving nuclear-norm terms. The ALM method iterates the following steps:

Augmented Lagrangian Multiplier method

- (1) $\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}_k, \mu_k)$
- (2) $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu_k h(\mathbf{x}_{k+1})$
- (3) Update $\mu_k \rightarrow \mu_{k+1}$ (we use $\mu_k = a^k$ with $a > 1$).

Next, we derive an ALM algorithm for nuclear-norm STLS and extend it to use reweighted nuclear norms based on a solution of the Sylvester's equations.

3.1. ALM for nuclear-norm STLS

We would like to solve the problem:

$$\min \|A\|_* + \alpha \|E\|_F^2, \text{ such that } \quad (12)$$

$$\bar{A} = A + E, \text{ and } \mathcal{L}(E) = \mathbf{b}$$

To view it as (10) we have $f(x) = \|A\|_* + \alpha \|E\|_F^2$ and $h(x) = \{\bar{A} - A - E, \mathcal{L}(E) - \mathbf{b}\}$. Using Λ as our matrix Lagrangian multiplier, the augmented Lagrangian is:

$$\min_{E: \mathcal{L}(E) = \mathbf{b}} \|A\|_* + \alpha \|E\|_F^2 + \text{tr}(\Lambda^T (\bar{A} - A - E)) + \frac{\mu}{2} \|\bar{A} - A - E\|_F^2. \quad (13)$$

Instead of a full optimization over $\mathbf{x} = (E, A)$, we use coordinate descent which alternates optimizing over each matrix variable holding the other fixed. We do not wait for the coordinate descent to converge at each ALM step, but rather update Λ and μ after a single iteration, following the inexact ALM algorithm in (Lin et al., 2010)³. Finally, instead of relaxing the constraint $\mathcal{L}(E) = \mathbf{b}$, we keep the constrained form, and follow each step by a projection (Bertsekas, 1982).

The minimum of (13) over A is obtained by the singular value thresholding operation (Cai et al., 2010):

$$A_{k+1} = \mathcal{S}_{\mu^{-1}}(\bar{A} - E_k + \mu_k^{-1} \Lambda_k) \quad (14)$$

where $\mathcal{S}_\gamma(Z)$ soft-thresholds the singular values of $Z = USV^T$, i.e. $\tilde{S} = \max(S - \gamma, 0)$ to obtain $\tilde{Z} = U\tilde{S}V^T$.

The minimum of (13) over E is obtained by setting the gradient with respect to E to zero, followed by a projection⁴ onto the affine space defined by $\mathcal{L}(E) = \mathbf{b}$:

$$\tilde{E}_{k+1} = \frac{1}{2\alpha + \mu_k} (\Lambda_k + \mu_k(\bar{A} - A))$$

$$\text{and } E_{k+1} = \Pi_{E: \mathcal{L}(E) = \mathbf{b}} \tilde{E}_{k+1} \quad (15)$$

3.2. ALM for re-weighted nuclear-norm STLS

To use the log-determinant heuristic, i.e., the re-weighted nuclear norm approach, we need to solve the weighted nuclear norm subproblems:

$$\min \|W_1 A W_2\|_* + \alpha \|E\|_F^2 \text{ where } \quad (16)$$

$$\bar{A} = A + E, \text{ and } \mathcal{L}(E) = \mathbf{b}$$

There is no known analytic thresholding solution for the weighted nuclear norm, so instead we follow (Liu et al., 2010) to create a new variable $D = W_1 A W_2$ and add this definition as an additional linear constraint:

$$\min \|D\|_* + \alpha \|E\|_F^2 \text{ where } \quad (17)$$

$$\bar{A} = A + E, D = W_1 A W_2, \text{ and } \mathcal{L}(E) = \mathbf{b}$$

³This is closely related to the popular alternating direction of multipliers methods (Boyd et al., 2011).

⁴For many constraints of interest this projection is highly efficient: when the constraint fixes some entries $E_{ij} = 0$, projection simply re-sets these entries to zero. Projection onto Toeplitz structure simply takes an average along each diagonal, e.t.c

Algorithm 1 ALM for weighted NN-STLS

Input: \bar{A} , W_1 , W_2 , α
repeat
 • Update D via soft-thresholding:
 $D_{k+1} = \mathcal{S}_{\mu_k^{-1}}(W_1 A W_2 - 1/\mu_k \Lambda_2^k)$.
 • Update E as in (15).
 • Solve Sylvester system for A in (19).
 • Update $\Lambda_1^{k+1} = \Lambda_1^k + \mu_k(\bar{A} - A - E)$,
 $\Lambda_2^{k+1} = \Lambda_2^k + \mu_k(D - W_1 A W_2)$ and $\mu_k \rightarrow \mu_{k+1}$.
until convergence

Now we have two Lagrangian multipliers Λ_1 and Λ_2 and the augmented Lagrangian is

$$\min_{E: \mathcal{L}(E)=b} \|D\|_* + \alpha \|E\|_F^2 + \text{tr}(\Lambda_1^T (\bar{A} - A - E)) + \text{tr}(\Lambda_2^T (D - W_1 A W_2)) + \frac{\mu}{2} \|\bar{A} - A - E\|_F^2 + \frac{\mu}{2} \|D - W_1 A W_2\|_F^2 \quad (18)$$

We again follow an ALM strategy, optimizing over D, E, A separately followed by updates of Λ_1, Λ_2 and μ . Note that (Deng et al., 2012) considered a strategy for minimizing re-weighted nuclear norms for matrix completion, but instead of using exact minimization over A , they took a step in the gradient direction. We derive the exact update, which turns out to be very efficient via a Sylvester equation formulation. The updates over D and over E look similar to the unweighted case. Taking a derivative with respect to A we obtain a linear system of equations in an unusual form: $-\Lambda_1 - W_y \Lambda_2 W_z - \mu(\bar{A} - A - E) - \mu W_1(D - W_1 A W_2) W_2 = 0$. Rewriting it, we obtain:

$$A + W_1^2 A W_2^2 = \frac{1}{\mu_k} (\Lambda_1 + W_1 \Lambda_2 W_2) + (\bar{A} - E) + W_1 D W_2 \quad (19)$$

we can see that it is in the form of Sylvester equation arising in discrete Lyapunov systems (Kailath, 1980):

$$A + B_1 A B_2 = C \quad (20)$$

where A is the unknown, and B_1, B_2, C are coefficient matrices. An efficient solution is described in (Bartels & Stewart, 1972). These ALM steps for reweighted nuclear norm STLS are summarized in Algorithm 1.

To obtain the full algorithm for STLS, we combine the above algorithm with steps of re-weighting the nuclear norm and the binary search over α as described in Section 2.1. We use it for experiments in Section 5. A faster algorithm that avoids the need for a binary search will be presented in a future publication.

4. Accuracy analysis for STLS

In context of matrix completion and robust PCA, the nuclear norm relaxation has strong theoretical accu-

racy guarantees (Recht et al., 2010; Chandrasekaran et al., 2011). We now study accuracy guarantees for the STLS problem via the nuclear norm and the reweighted nuclear norm approaches. The analysis is conducted in the plain TLS setting, where the optimal solution is available via the SVD, and it gives valuable insight into the accuracy of our approach for the much harder STLS problem. In particular, we quantify the dramatic benefit of using reweighting. In this section we study a simplification of our STLS algorithm, where we set the regularization parameter α once and do not update it through the iterations. The full adaptive approach from Section 2.1 is analyzed in the addendum to this paper where we show that it can in fact recover the exact SVD solution for plain TLS.

We first consider the problem $\min \|A - \bar{A}\|_F^2$ such that $\text{rank}(A) \leq N - 1$. For the exact solution via the SVD, the minimum approximation error is simply the square of the last singular value $Err_{SVD} = \|\hat{A}_{SVD} - \bar{A}\|_F^2 = \sigma_N^2$. The nuclear-norm approximation will have a higher error. We solve $\min \|A - \bar{A}\|_F^2 + \alpha \|A\|_*$ for the smallest choice of α that makes A rank-deficient. A closed form solution for A is the soft-thresholding operation with $\alpha = \sigma_N$. It subtracts α from all the singular values, making the error $Err_{nn} = N\sigma_N^2$. While it is bounded, this is a substantial increase from the SVD solution. Using the log-det heuristic, we obtain much tighter accuracy guarantees even when we fix α , and do not update it during re-weighting. Let $a_i = \frac{\sigma_i}{\sigma_N}$, the ratio of the i -th and the smallest singular values. In the appendix using ‘log-thresholding’ we derive that

$$Err_{rw-nn} \approx \sigma_N^2 \left(1 + \frac{1}{2} \sum_{i < N} (a_i - \sqrt{a_i^2 - 1})^2 \right) \leq \sigma_N^2 \left(1 + \frac{1}{2} \sum_{i < N} \frac{1}{a_i^2} \right) \quad (21)$$

For larger singular values the approximation is much more accurate than for the smallest ones. In contrast, for the plain nuclear norm approach, the errors are equally bad for the largest and smallest singular values. Considering that natural signals (and singular value spectra) often exhibit fast decay (exponential or power-law decay), we can quantify the improvement. Suppose that the singular values have exponential decay, $\sigma_i = \sigma_N a^{N-i}$, with $a > 1$, or power-law decay $\sigma_i = \sigma_N (N - i + 1)^p$. The approximation errors are $Err_{exp} = \sigma_N^2 \left(1 + \frac{1}{2} \sum_{i < N} (a^i - \sqrt{a^{2i} - 1})^2 \right)$ and $Err_p = \sigma_N^2 \left(1 + \frac{1}{2} \sum_{i < N} (i^p - \sqrt{i^{2p} - 1})^2 \right)$ respectively. For exponential decay, if $N = 100$, and $a = 1.1$, the approximation error is $1.84 \sigma_N^2$ for our approach, and $N\sigma_N^2 = 100 \sigma_N^2$, for the nuclear norm relaxation. This is a dramatic improvement in approximation, that strongly supports using the log-det

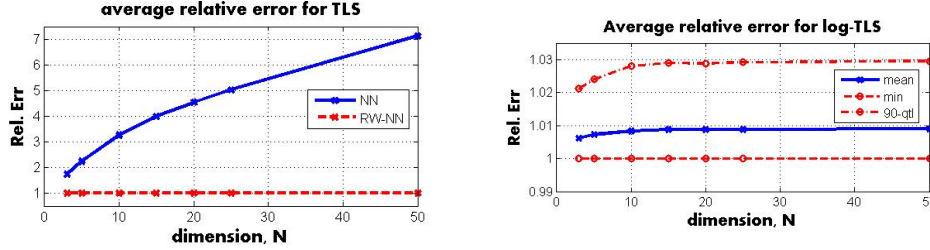


Figure 1. RW-NN for plain TLS: (a) relative error in Frobenius norm for NN and full RW-NN (avg. over 100 trials) w.r.t. SVD. (b) Min, average, and 90%-quantile relative error for the simplified non-adaptive log-det STLS relative to SVD.

heuristic over the nuclear norm for approximating matrix rank!

5. Experimental Results

Our first experiment considers plain TLS, where we know the optimal solution via the SVD. We evaluate the accuracy of the nuclear norm (NN) and two flavors of the reweighted nuclear norm algorithm: the full adaptive one described in Section 2.1, which we will refer to as (RW-NN), and the simplified approach with fixed α as described in Section 4 (log-det).

We simulate random i.i.d. Gaussian matrices A of size $N \times N$, use a maximum of 3 re-weightings in RW-NN, and update the ALM parameter μ_k as $\mu_k = 1.05^k$. We plot the relative error of NN-TLS with respect to exact TLS via SVD, i.e. the norm of error (w.o. squaring) for NN-TLS divided by the norm of error for TLS. We compare it to the relative error for full RW-NN TLS (again with respect to exact TLS) in Figure 1 (a). The results are averaged over 100 independent trials.

The NN solution, as we expect, is a factor of \sqrt{N} worse than TLS in Frobenius norm. The full RW-NN always recovers the optimal TLS solution, i.e. the relative error is exactly 1, as we establish in the addendum. The simplified non-adaptive log-det STLS in Figure 1 (b) is almost as good as the adaptive: the average error is only about 1% higher than exact TLS, dramatically better than \sqrt{N} for plain NN. These empirical results agree with our theoretical analysis in Section 4.

Next, we compare NN and RW-NN for a structured TLS problem with a pattern of entries in E fixed at 0 (entries are fixed independently with probability 0.5). This is a practically important case where the entries of E fixed at zero represent exact measurements while allowing other entries to have noisy measurements. The solution of plain TLS via SVD ignores the constraints and is infeasible for this structured problem. We still compute the relative error with respect to exact TLS to quantify the increase in error needed to obey the imposed structure. Again, in Figure 2 (a) we can see that the RW-NN solution provides much bet-

ter accuracy than NN, and not far worse than 1, the infeasible lower-bound given by plain TLS.

Next we consider Toeplitz structured errors, which means that the matrix E is constant on the diagonals:

$$\begin{bmatrix} e_1 & e_2 & e_3 & \dots \\ e_0 & e_1 & e_2 & \dots \\ e_{-1} & e_0 & e_1 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (22)$$

Toeplitz structure arises in time-series modeling, analysis of linear systems, and system identification, as the convolution operation can be represented as a multiplication by a Toeplitz matrix (Kailath, 1980). We simulate the Toeplitz entries at the start of each diagonal as i.i.d. Gaussian. The Toeplitz structure is quite restrictive, with only $M + N - 1$ degrees of freedom instead of $O(MN)$. However, we reach a similar conclusion as we had before: RW-NN solution provides much better accuracy than NN, much closer to the infeasible lower-bound via plain TLS. We show the results for STLS with Toeplitz structure in Figure 2(b).

Finally, we compare our re-weighted nuclear norm approach to the latest widely used non-convex solver for STLS, SLRA (Markovsky & Udevich, 2014). The success of non-convex solvers depends heavily on a good initialization. We consider a problem with a block-diagonal structure where some entries are corrupted by large outliers. The weights on these entries are set to be very small, so an ideal STLS solver should find the solution while minimizing the influence of the outliers. Figure 3 shows that for moderate levels of outliers, both RW-NN and the non-convex SLRA approach find very accurate solutions. However, for larger levels of noise, while RW-NN continues to have good performance, the accuracy of SLRA plummets, presumably due to the difficulty of finding a good initialization. For this setting we know the exact solution without outliers, and we measure accuracy by correlation (i.e. cosine of subspace angles) of the recovered and the exact STLS nullspaces, averaged over 100 trials.

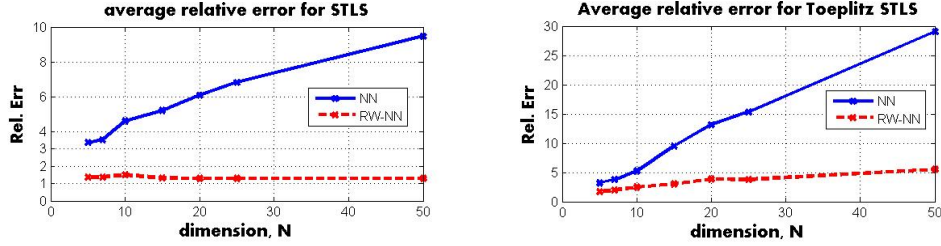


Figure 2. Relative error in Frobenius norm for NN and RW-NN (avg. over 100 trials). The SVD solution is infeasible, providing a lower bound 1 on the error. (a) STLS with some observed entries (b) STLS with Toeplitz structure.

5.1. Quantification of cellular heterogeneity

We now demonstrate the utility of STLS for solving a broad and important class of problems arising in biology, namely inferring heterogeneity in biological systems. Most biological systems (such as human cancers, tissues, the human microbiome and other microbial communities) are mixtures of cells in different physiological states or even different cell types. While the primary biomedical interest is in characterizing the different cell types and physiological states, experimental approaches can typically measure only the population average across physiological states. Our aim is to combine these readily available population-average measurements and use STLS to infer the distribution of cells across distinct physiological states.

We consider a cell culture containing cells in K distinct physiological states, such as phases of cell growth or division cycles (Slavov et al., 2011). As the growth rate of the culture changes, the fraction of cells in each physiological state changes. This fractional change is of primary interest but it is often too expensive or even technologically impossible to measure directly. Since the cells cannot be easily separated we consider the general case when we know M indicator genes (such as cyclins) that are either present or absent in K distinct physiological states, $S \in \mathbb{R}^{M \times K}$. Existing methods for high-throughput measurements of mRNA levels, such as DNA microarrays and RNA-seq, can quantify relative changes of mRNA levels across different conditions but cannot accurately quantify the ratios between different mRNAs, i.e., depending on chemical composition and the physical properties, each RNA has its own normalization scaler accounting for biases such as GC (guanine-cytosine) content. To avoid such biases we explicitly scale the measured relative expression levels $X \in \mathbb{R}^{M \times N}$ by an unknown positive diagonal matrix $Z = \text{diag}(z)$. The goal is to find $U \in \mathbb{R}^{K \times N}$, the fraction of cells across the K physiological states for each of N different conditions, such as different steady-state growth rates. Mathematically the problem is:

$$X = ZSU, \quad (23)$$

where we aim to recover the decomposition *up-to scal-*

ing knowing X and S only. We now study conditions for identifiability without noise, and extend it to a structured TLS problem in presence of noise.

Linear Algebraic solution We define $\lambda = [\frac{1}{z_1}, \dots, \frac{1}{z_M}]$, and $\Lambda = \text{diag}(\lambda)$. Thus $\Lambda = Z^{-1}$. We now have to find Λ and U :

$$\Lambda X = SU, \quad (24)$$

and both unknowns enter the equations linearly. We transpose both sides and move everything to one side to get: $U^T S^T - X^T \Lambda = 0$. Now let us stack columns of U^T , i.e. rows of U into a vector, $\mathbf{u} = \text{vec}(U^T)$. Then $\text{vec}(U^T S^T) = (S \otimes I)\mathbf{u}$, where \otimes stands for the Kronecker product. Similarly defining a block-diagonal matrix $\text{blkdiag}(X^T)$, with columns of X^T (i.e. rows of X) in diagonal blocks. This way $X^T \Lambda = \text{blkdiag}(X^T)\lambda$. Combining this together we have:

$$[(S \otimes I), -\text{blkdiag}(X^T)] \begin{bmatrix} \mathbf{u} \\ \lambda \end{bmatrix} = 0 \quad (25)$$

Any vector in the nullspace of $A \triangleq [(S \otimes I) - \text{blkdiag}(X^T)]$ is a solution to this problem. If we have a single vector in the nullspace of A , then we have a unique solution up-to scaling.

Noisy case: structured Total Least Squares approach When the observation matrix X is corrupted by noise, it is no longer low-rank. The structured matrix A in (25) will only have a trivial null-space. Furthermore, the simple approach of setting the smallest singular value to zero will not work because it ignores the structure of the compound matrix A . The errors in X correspond to errors in the block-diagonal portions of the right part of A . Other entries are known exactly. This is precisely the realm of structured total least squares (STLS) that we explored in Section 2.

We will now experimentally apply our reweighted nuclear norm approach for STLS for the cell heterogeneity quantification problem to demonstrate the inference of the fractions of cells in different physiological states. We use experimentally measured levels of 14

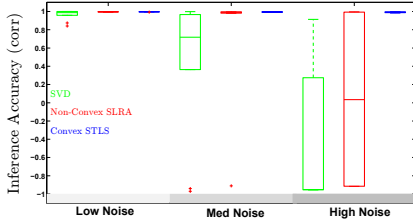


Figure 3. Comparison of convex STLS (RW-NN) with non-convex SLRA and SVD on problems with large outliers. At low-noise all the solvers are accurate, but only RW-NN remains accurate at high noise. We measure accuracy by correlation with the exact solution. We present the distribution of correlations over 100 trials as a boxplot.

genes, five expressed in HOC phase, six expressed in LOC phase, and three in both phases, across 6 exponentially growing yeast cultures at different growth rates. The resulting A matrix in (25) is 84×26 . Our algorithm infers the fraction of cells in HOC and in LOC phase, up to a scalar factor, in close agreement with expectations from physical measurements in synchronized cultures (Slavov et al., 2011; Slavov & Botstein, 2011). Thus we can extend the observed trend to asynchronous cultures where this fraction is very hard to measure experimentally. Such analysis can empower research on cancer heterogeneity that is a major obstacle to effective cancer therapies. This modest size experiment provides a proof of concept and we are pursuing applications to more complex biological systems.

6. Appendix: Error analysis for re-weighted STLS

To gain insight into the re-weighted nuclear norm we consider the diagonal case first, where $A = \text{diag}(\mathbf{x})$. The diagonal matrix case penalizing rank of A is equivalent to the vector problem penalizing sparsity of \mathbf{x} , so we use the vector notation for simplicity. As both the Frobenius and nuclear norms are unitarily invariant⁵, the analysis directly extends to the non-diagonal case.

The log heuristic for sparsity solves the following problem: $\min \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \sum_i \log(\delta + |x_i|)$, for a very small $\delta > 0$. This is a separable problem with a closed form solution for each coordinate⁶ (contrast this with the

⁵ Taking the SVD $A = USV^T$ we have $\|USV\|_F^2 = \|S\|_F^2$ and $\|USV\|_* = \|S\|_*$ since U, V are unitary.

⁶ For δ small enough, the global minimum is always at 0, but if $y > 2\sqrt{\alpha}$ there is also a local minimum with a large domain of attraction between 0 and y . Iterative linearization methods with small enough stepsize starting at y will converge to this local minimum.

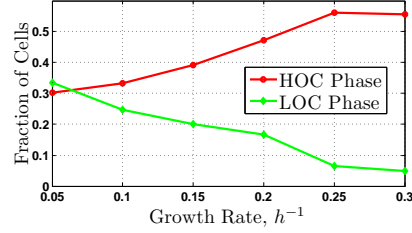


Figure 4. STLS infers accurate fractions of cells in different physiological phases from measurements of population-average gene expression across growth rate.

soft-thresholding operation):

$$x_i = \begin{cases} \frac{1}{2} \left((y_i - \delta) + \sqrt{(y_i - \delta)^2 - 4(\alpha - y_i\delta)} \right), & y_i > 2\sqrt{\alpha} \\ \frac{1}{2} \left((y_i + \delta) - \sqrt{(y_i + \delta)^2 - 4(\alpha + y_i\delta)} \right), & y_i < -2\sqrt{\alpha} \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

Assuming that δ is negligible, then we have:

$$x_i \approx \begin{cases} \frac{1}{2}(y_i + \sqrt{y_i^2 - 4\alpha}), & \text{if } y_i > 2\sqrt{\alpha} \\ \frac{1}{2}(y_i - \sqrt{y_i^2 - 4\alpha}), & \text{if } y_i < -2\sqrt{\alpha} \\ 0, & \text{otherwise,} \end{cases} \quad (27)$$

and we chose α to annihilate the smallest entry in \mathbf{x} , i.e. $\alpha = \frac{1}{4} \min_i y_i^2$. Sorting the entries in $|\mathbf{y}|$ in increasing order, with $y_0 = y_{\min}$, and defining $a_i = \frac{|y_i|}{|y_0|}$, we have $a_i \geq 1$ and the error in approximating the i -th entry, for $i > 0$ is

$$Err_i = |x_i - y_i|^2 = \frac{y_0^2}{2} \left(a_i - \sqrt{a_i^2 - 1} \right)^2 \leq \frac{y_0^2}{2a_i^2}. \quad (28)$$

Also, by our choice of α , we have $Err_0 = y_0^2$ for $i = 0$. The approximation error quickly decreases for larger entries. In contrast, for ℓ_1 soft-thresholding, the errors of approximating large entries are as bad as the ones for small entries. This analysis extends directly to the log-det heuristic for relaxing matrix rank.

7. Conclusions

We considered a convex relaxation for a very rich class of structured TLS problems, and provided theoretical guarantees. We also developed an efficient first-order augmented Lagrangian multipliers algorithm for reweighted nuclear norm STLS, which can be applied beyond TLS to matrix completion and robust PCA problems. We applied STLS to quantifying cellular heterogeneity from population average measurements. In future work we will study STLS with sparse and group sparse solutions, and explore connections to robust LS (El Ghaoui & Lebret, 1997).

References

- Bartels, R. H. and Stewart, G. W. Solution of the matrix equation $AX + XB = C$. *Communications of the ACM*, 15(9):820–826, 1972.
- Beck, A. and Ben-Tal, A. A global solution for the structured total least squares problem with block circulant matrices. *SIAM Journal on Matrix Analysis and Applic.*, 27(1):238–255, 2005.
- Bertsekas, D. P. *Constrained Optim. and Lagrange Multiplier Methods*. Academic Press, 1982.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- Cai, J., Candes, E. J., and Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optim.*, 20(4):1956–1982, 2010.
- Candes, E. J., Wakin, M. B., and Boyd, S. P. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applic.*, 14(5):877–905, 2008.
- Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., and Willsky, A. S. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optim.*, 21(2), 2011.
- Deng, Y., Dai, Q., Liu, R., Zhang, Z., and Hu, S. Low-rank structure learning via log-sum heuristic recovery. *arXiv preprint arXiv:1012.1919*, 2012.
- El Ghaoui, L. and Lebret, H. Robust solutions to least-squares problems with uncertain data. *SIAM J. on Matrix Analysis and Applic.*, 18(4):1035–1064, 1997.
- Fazel, M., Hindi, H., and Boyd, S. P. A rank minimization heuristic with application to minimum order system approximation. In *IEEE American Control Conference*, 2001.
- Golub, G. H. and Van Loan, C. F. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893, 1980.
- Kailath, T. *Linear systems*. Prentice-Hall, 1980.
- Khajehnejad, A., Xu, W., Avestimehr, S., and Hassibi, B. Weighted ℓ_1 minimization for sparse recovery with prior information. In *IEEE Int. Symposium on Inf. Theory, 2009.*, pp. 483–487, 2009.
- Lin, Z., Chen, M., and Ma, Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.
- Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., and Ma, Y. Robust recovery of subspace structures by low-rank representation. *arXiv preprint arXiv:1010.2955*, 2010.
- Markovsky, I. and Usevich, K. Software for weighted structured low-rank approximation. *J. Comput. Appl. Math.*, 256:278–292, 2014.
- Markovsky, I. and Van Huffel, S. Overview of total least-squares methods. *Signal processing*, 87(10): 2283–2302, 2007.
- Markovsky, I., Willems, J. C., Van Huffel, S., De Moor, B., and Pintelon, R. Application of structured total least squares for system identification and model reduction. *Automatic Control, IEEE Trans. on*, 50(10):1490–1500, 2005.
- Mohan, K. and Fazel, M. Reweighted nuclear norm minimization with application to system identification. In *American Control Conference*, 2010.
- Needell, D. Noisy signal recovery via iterative reweighted l1-minimization. In *Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009*, pp. 113–117. IEEE, 2009.
- Recht, B., Fazel, M., and Parrilo, P. A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
- Slavov, N. and Botstein, D. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular bio. of the cell*, 22(12), 2011.
- Slavov, N., Macinskas, J., Caudy, A., and Botstein, D. Metabolic cycling without cell division cycling in respiring yeast. *Proceedings of the National Academy of Sciences*, 108(47), 2011.
- Srebro, N. and Jaakkola, T. Weighted low-rank approximations. In *Int. Conf. Machine Learning (ICML)*, 2003.
- Toh, K. C., Todd, M. J., and Tütüncü, R. H. SDPT3 – a Matlab software package for semidefinite programming, version 1.3. *Optim. Method. Softw.*, 11(1–4):545–581, 1999.
- Zhu, H., Giannakis, G. B., and Leus, G. Weighted and structured sparse total least-squares for perturbed compressive sampling. In *IEEE Int. Conf. Acoustics, Speech and Signal Proc.*, 2011.