51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Linear System Identification via Atomic Norm Regularization

Parikshit Shah, Badri Narayan Bhaskar, Gongguo Tang and Benjamin Recht
University of Wisconsin-Madison

Abstract— **This paper proposes a new algorithm for linear system identification from noisy measurements. The proposed algorithm balances a data fidelity term with a norm induced by the set of single pole filters. We pose a convex optimization problem that approximately solves the atomic norm minimization problem and identifies the unknown system from noisy linear measurements. This problem can be solved efficiently with standard, free software. We provide rigorous statistical guarantees that explicitly bound the estimation error (in the $\mathcal{H}_2$-norm) in terms of the stability radius, the Hankel singular values of the true system and the number of measurements. These results in turn yield complexity bounds and asymptotic consistency. We provide numerical experiments demonstrating the efficacy of our method for estimating linear systems from a variety of linear measurements.**

Keywords System identification. Atomic norms. Hankel operators. Optimization.

I. Introduction

Identifying dynamical systems from noisy observation of their input-output behavior is of fundamental importance in systems and control theory. Often times models derived from physical first principles are not available to the control engineer, and computing a surrogate model from data is essential to the design of a control system. System identification from data is thus ubiquitous in problem domains ranging from process engineering, dynamic modeling of mechanical and aerospace systems, and systems biology. Though there are a myriad of approaches and excellent texts on the subject (see, for example [10]), there is still no universally agreed upon approach for this problem. One reason is that quantifying the interplay between system parameters, measurement noise, and model mismatch tends to be challenging.

This paper draws novel connections between contemporary high-dimensional statistics, operator theory, and linear systems theory to prove consistent estimators of linear systems from small measurement sets. In particular, building on recent studies of *atomic norms* in estimation theory [3], [1], we propose a penalty function which encourages estimated models to have small McMillan degree.

System identification has enjoyed renewed interest due to the realization that many data analysis approaches from statistics and machine learning are applicable to this domain. As an example, modeling noisy impulse response observations can be studied in a kernelized Gaussian process framework, and the related system identification problem can be analyzed from a Bayesian perspective [17], [18]. Regularization approaches have also been recently considered in the context of indentification of FIR models [4], as well as Laguerre expansions [22]. A related family of system identification techniques use finite sample Hankel matrices to estimate dynamical system models, using either singular value decompositions (e.g, [21], [12]) or semidefinite programming [6], [9], [20], [7]. In all of these techniques, no statistical guarantees were given about the quality of estimation with finite noisy data, and it was difficult to determine how sensitive these methods were to the hidden system parameters or measurement noise. Moreover, since these problems were dealing with finite, truncated Hankel matrices, it is never certain if the size of the Hankel matrix is sufficient to reveal the true McMillan degree. Moreover, the techniques based on semidefinite programming are challenging to scale to very large problems, as their complexity grows superlinearly with the number of measurements.

In contrast, the atomic norm regularizer proposed in this paper is not only equivalent to the sum of the Hankel singular values (the Hankel nuclear norm), but is also well approximated by a finite dimensional, ℓ_1 minimization problem. We show that solving least-squares problems regularized by our atomic norm is consistent, and scales gracefully with the stability radius, the McMillan degree of the system to be identified, and the number of measurements. Our numerical experiments validate these theoretical underpinnings, and show that our method has great promise to provide concrete estimates on the hard limits of estimating linear systems.

A. Notation

We adopt standard notation; $\mathbb{D}$ and $\mathbb{S}$ will denote respectively the open unit ball and the unit circle in the complex plane $\mathbb{C}$. $\mathcal{H}_2$ and $\mathcal{H}_\infty$ will denote the Hardy spaces of functions analytic outside $\mathbb{D}$, with the norms

$$\|f\|_{\mathcal{H}_2} = \tfrac{1}{2\pi}\int_0^{2\pi} |f(e^{i\theta})|^2 d\theta \quad \text{and} \quad \|f\|_{\mathcal{H}_\infty} = \sup_{z\in\mathbb{S}} |f(z)|$$

respectively. $\ell_2([a,b])$ will denote the set of square summable sequences on the integers in $[a,b]$.

II. Atomic Decompositions of Transfer Functions

We restrict our attention to SISO systems in this paper, as this will simplify the presentation. It is possible to extend our techniques to MIMO systems, we will briefly discuss this later. Suppose we wish to estimate a SISO, LTI system with transfer function $G_\star(z)$ from a finite collection of measurements $y = \Phi(G_\star)$. The set of all transfer functions is an infinite dimensional space, so reconstructing $G_\star$ from this data is ill-posed. In order to make it well posed, a common regularization approach constructs a penalty function $\mathrm{pen}(\cdot)$ that encourages "low-complexity" models and solves the optimization problem

$$\text{minimize}_G \ \|\Phi(G) - y\|_2^2 + \mu \,\mathrm{pen}(G)\,. \tag{1}$$

This formulation uses the parameter μ to balance between model complexity and fidelity to the data. The least-squares cost can be modified to other convex loss functions if knowledge about measurement noise is available (as in [20], [13]), though in general it is less clear how to design a good penalty function.

In many applications, we know that the true model can be decomposed as a linear combination of very simple building blocks. For instance, sparse vectors can be written as short linear combinations of vectors from some discrete dictionary and low-rank matrices can be written as a sum of a few rank-one factors. In [3], Chandraskearan et al. proposed a universal heuristic for constructing regularizers based on such prior information. If we assumed that

$$G_\star = \sum_{i=1}^{r} c_i a_i\,, \ \text{for some } a_i \in \mathcal{A}, c_i \in \mathbb{C}\,,$$

where $\mathcal{A}$ is a set of "atoms" normalized to have unit norm and r is relatively small, then the appropriate penalty function is the gauge function (or the Minkowski functional) induced by the atomic set $\mathcal{A}$:

$$\begin{aligned}\|G\|_{\mathcal{A}} :&= \inf\left\{t \ : \ G \in t \ \mathrm{conv}(\mathcal{A})\right\} \\ &= \inf\left\{\sum_{a\in\mathcal{A}} |c_a| \ : \ G = \sum_{a\in\mathcal{A}} c_a a\right\}.\end{aligned} \tag{2}$$

In [3], it is shown that minimizing the atomic norm subject to compressed measurements yields the tightest known bounds for recovering many models from linear measurements. In [1], the atomic norm regularizer was studied in the context of denoising problems and was found to produce consistent estimates at nearly optimal estimation error rates for many classes of atoms.

To apply these atomic norm techniques to system identification, we must first determine the appropriate set of atoms. For discrete time LTI systems with small McMillan degree, we can always decompose any finite dimensional, strictly proper system $G(z)$ as:

$$G(z) = \sum_{i=1}^{s} \frac{c_i}{z - a_i}\,.$$

via a partial fraction expansion. Hence, it follows that our set of atoms should be single-pole transfer functions. We propose the following atomic set for linear systems

$$\mathcal{A} = \left\{\varphi_w(z) = \frac{1-|w|^2}{z-w} \ : \ w \in \mathbb{D}\right\}.$$

The numerator is normalized so that the Hankel norm of each atom is 1. See the discussion in Section III for precisely why this normalization is desirable.

The atomic norm penalty function associated with these atoms is

$$\|G(z)\|_{\mathcal{A}} = \inf\left\{\sum_{w\in\mathbb{D}} |c_w| \ : \ G(z) = \sum_{w\in\mathbb{D}} \frac{c_w(1-|w|^2)}{z-w}\right\}, \tag{3}$$

where the summation implies that only a countable number of terms have nonzero coefficients c_w. This expression finds the decomposition of $G(z)$ into a linear combination of single pole systems such that the ℓ_1 norm, weighted by the norms of the single poles, is as small as possible.

With this penalty function in hand, we now turn to analyzing its utility. In Section III, we first show that for most systems of interest $\|G\|_{\mathcal{A}}$ is a well-defined, bounded quantity. Moreover, we will show that the atomic norm is equivalent to the nuclear norm of the Hankel operator associated with G. Hence, the models that are preferred by our penalty function will have low-rank Hankel operators, and thus low McMillan degrees.

In Section IV, we turn to computation, demonstrating practical algorithms for approximating atomic norm regularization problems for several classes of measurements. We will show that with finite data, our atomic norm minimization problem is well-approximated by a finite-dimensional ℓ_1 norm regularization problem. In particular, using specialized algorithms adapted to the solution of an ℓ_1-norm regularized least-squares problem, we can solve atomic norm regularization problems in time competitive with respect to techniques that regularize with the nuclear norm and SVD-based subspace identification methods.

Finally, we analyze the statistical performance of atomic norm minimization in Section V. We show that our algorithm is asymptotically consistent over several measurement ensembles of interest. We focus on sampling the transfer function on the unit circle and present $\mathcal{H}_2$ error bounds in terms of the stability radius, Hankel singular values, $\mathcal{H}_\infty$ norm, and McMillan degree of the system to be estimated.

III. The Hankel Nuclear Norm and Atomic Norm Minimization

A. Preliminaries: the Hankel operator

Recall that the *Hankel operator*, Γ_G, of the transfer function G is defined as the mapping from the past to the future under the transfer function G. Given a signal u supported on $(-\infty, -1]$, the output under G is given by $g * u$ where "$*$" denotes convolution and g is the impulse response of G:

$$G(z) = \sum_{k=1}^{\infty} g_k z^{-k} .$$

Γ_G is then simply the projection of $g * u$ onto $[0, \infty)$. An introduction to Hankel operators in control theory can be found in [23, Chapter 7].

The *Hankel norm* of G is the operator norm of Γ_G considered as an operator mapping $\ell_2(-\infty, -1]$ to $\ell_2[0, \infty)$. The *Hankel nuclear norm* of G is the nuclear norm (aka the trace norm or Schatten 1-norm) of Γ_G. To be precise, an operator T is in the *trace class* S_1 if the trace of $(T^*T)^{1/2}$ is finite. This implies that T is a compact operator and admits a singular value decomposition

$$T(f) = \sum_{i=1}^{\infty} \sigma_i \langle V_i, f \rangle U_i .$$

The sequence σ_i are called the *Hankel singular values* of T. Moreover, the Schatten 1-norm of T is given by

$$\|T\|_1 = \operatorname{trace}\left((T^*T)^{1/2}\right) = \sum_{i=1}^{\infty} \sigma_i .$$

B. The atomic and Hankel nuclear norms are equivalent

The rank of the Hankel operator is equal to the McMillan degree of the linear system defined by G. Rank minimization is notoriously computationally challenging (see [19] for a discussion), and direct penalization of the rank of the Hankel operator is likely intractable. A common alternative to minimizing rank is minimizing the sum of singular values of the associated operator. In our context, this would consist of minimizing the Schatten 1-norm of the Hankel operator. For rational transfer functions, we can compute this Hankel nuclear norm via a balanced realization [23]. On the other hand, while the maximal Hankel singular value can be written variationally as an LMI, we are not aware of any such semidefinite programming formulations for the Hankel nuclear norm.

The following theorem provides a path towards minimizing the Hankel nuclear norm, via minimization of the atomic norm $\|G(z)\|_{\mathcal{A}}$ as a proxy. Indeed, from the view of Banach space theory, the atomic norm is *equivalent* to the Hankel nuclear norm.

Theorem 3.1: Let $G \in \mathcal{H}_2$. Then Γ_G is trace class if and only if there exists a sequence $\{\lambda_k\} \in \ell_1$ and a sequence $\{w_k\}$ with $w_k \in \mathbb{D}$ such that

$$g(z) = \sum_{i=1}^{\infty} \lambda_k \frac{1 - |w_k|^2}{z - w_k} . \tag{4}$$

Moreover, we have the following chain of inequalities

$$\tfrac{\pi}{8} \|G\|_{\mathcal{A}} \leq \|\Gamma_G\|_1 \leq \|G\|_{\mathcal{A}} \tag{5}$$

where $\|G\|_{\mathcal{A}}$ is given by (2).

Proof Outline Theorem 3.1 follows by carefully combining several different results from operator theory. Peller first showed that transfer functions with trace class Hankel operators formed a *Besov space* [15]. Peller's argument can be found in his book [16]. The atomic decomposition of such operators is due to Coifman and Rochberg [5]. The norm bounds (5) were proven by Bonsall and Walsh [2]. There they show that the $\frac{\pi}{8}$ is the best possible lower bound. They also show that if $\|\Gamma_g\|_1 \leq C\|g\|_{\mathcal{A}}$ for all g, then C must be at least $\frac{1}{2}$, so the chain of inequalities is nearly optimal. A concise presentation of the full argument can be found in [14]. A modern perspective using the theory of reproducing kernels can be found in [24].

Theorem 3.1 asserts that a transfer function has a finite atomic norm if and only if the sum of its Hankel singular values is finite. In particular, this means that every rational transfer function has a finite atomic norm. More importantly, the atomic norm is equivalent to the Hankel nuclear norm. Thus if we can approximately solve atomic norm-minimization, we can approximately solve Hankel nuclear norm minimization and vice-versa. We now turn to such computational considerations.

IV. Algorithms for atomic norm minimization

From here on, let us assume that the $G_\star$ that we seek to estimate has all of its poles of magnitude at most ρ (we will call ρ the *stability radius*, and treat it as a known parameter). Let $\mathbb{D}_\rho$ denote the set of all

complex numbers with norm at most ρ. Note that if $G_\star$ has stability radius ρ then

$$\|G\|_{\mathcal{A}} := \inf \left\{ \sum_{w \in \mathbb{D}_\rho} |c_w| \ : \ G(z) = \sum_{w \in \mathbb{D}_\rho} \frac{c_w(1-|w|^2)}{z-w} \right\}.$$

That is, we can restrict our set of atoms to only be those single pole systems with stability radius equal to ρ. For the remainder of this manuscript, we assume that $\mathcal{A}$ only consists of such single pole systems.

In what follows, we focus our attention on linear measurement maps. Let $\mathcal{L}_k : \mathcal{H} \mapsto \mathbb{C}$ be a linear functional that serves as a measurement operator for the system $G(z)$, so that noisy measurements of the form $y_k = \mathcal{L}_k(G) + \omega_k$, $k = 1, \ldots, n$ are recorded. Many maps of interest can be phrased as linear functionals of the transfer function,

1) Samples of the frequency response $\mathcal{L}_k(G) := G(e^{i\theta_k})$ for $k = 1, \ldots, n$. From a control theoretic perspective, this measurement operator corresponds to measuring the gain and phase of the linear system at different frequencies.
2) Samples of the impulse response, $\mathcal{L}_k(G) := g_{i_k}$ for $k = 1, \ldots, n$ and $i_k \in [1, \infty)$.
3) Convolutions of the impulse response with a pseudorandom signal u_k: $\mathcal{L}_k(G) := \sum_{j=1}^{\infty} g_j u_{k-j}$.

In all of these cases, we consider the problem

$$\text{minimize}_G \, \tfrac{1}{2} \sum_{i=1}^{n} |\mathcal{L}_i(G) - y_i|^2 + \mu \|G\|_{\mathcal{A}} \,. \qquad (6)$$

This problem is equivalent to the constrained, semi-infinite programming problem

$$\begin{array}{ll} \text{minimize}_{x,G} & \tfrac{1}{2}\sum_{k=1}^{n} |x_k - y_k|^2 + \mu \sum_{w \in \mathbb{D}_\rho} |c_w| \\ \text{subject to} & x_k = \mathcal{L}_k(G) \ \text{ for } k = 1, \ldots, n \\ & G = \sum_{w \in \mathbb{D}_\rho} \frac{c_w(1-|w|^2)}{z-w} \end{array}$$

Eliminating the equality constraint gives yet another equivalent formulation

$$\begin{array}{ll} \text{minimize}_{x,c} & \tfrac{1}{2}\sum_{k=1}^{n} |x_k - y_k|^2 + \mu \sum_{w \in \mathbb{D}_\rho} |c_w| \\ \text{subject to} & x_k = \sum_{w \in \mathbb{D}_\rho} c_w \mathcal{L}_k \left(\frac{1-|w|^2}{z-w} \right). \end{array} \qquad (7)$$

Note that in this final formulation, our decision variable is x, a finite dimensional vector, and c_w, the coefficients of the atomic decomposition. The infinite dimensional variable G has been eliminated. Let us define a norm on $\mathbb{R}^n$ based on the formulation (7)

$$\|x\|_{\mathcal{L}(\mathcal{A})} = \inf \left\{ \sum_{w \in \mathbb{D}_\rho} |c_w| \ : \ x_i = \sum_{w \in \mathbb{D}_\rho} c_w \mathcal{L}_i \left(\frac{1-|w|^2}{z-w} \right) \right\}.$$

Then we see that problem (6) is equivalent to the denoising problem

$$\text{minimize}_x \, \tfrac{1}{2} \|x - y\|_2^2 + \mu \|x\|_{\mathcal{L}(\mathcal{A})} \,. \qquad (8)$$

Note that the first term is simply the squared Euclidean distance between y and x in $\mathbb{R}^n$. The second term is an atomic norm on $\mathbb{R}^n$ induced by the linear map of the set of transfer functions via the measurement operator $\mathcal{L}$. In order to tractably solve (6), we thus only need focus on computational schemes for computing or approximating $\|x\|_{\mathcal{L}(\mathcal{A})}$. The following proposition asserts that we can approximate this finite dimensional atomic norm via a sufficiently fine discretization of the unit disk.

Proposition 4.1: Let $\mathbb{D}_\rho^{(\epsilon)}$ be a finite subset of the unit disc such that for any $w \in \mathbb{D}_\rho$ there exists a $v \in \mathbb{D}_\rho^{(\epsilon)}$ satisfying $|w - v| \leq \epsilon$. Define

$$\|x\|_{\mathcal{L}(\mathcal{A}_\epsilon)} = \inf \left\{ \sum_{w \in \mathbb{D}_\rho^{(\epsilon)}} |c_w| \ : \ x_i = \sum_{w \in \mathbb{D}_\rho^{(\epsilon)}} c_w \mathcal{L}_i \left(\frac{1-|w|^2}{z-w} \right) \right\}.$$

Then there exists a constant $C_\epsilon \in [0, 1]$ such that

$$C_\epsilon \|x\|_{\mathcal{L}(\mathcal{A}_\epsilon)} \leq \|x\|_{\mathcal{L}(\mathcal{A})} \leq \|x\|_{\mathcal{L}(\mathcal{A}_\epsilon)} \,.$$

The set $\mathbb{D}_\rho^{(\epsilon)}$ is called an ϵ-net for the set $\mathbb{D}_\rho$. We show in the extended version of this paper that when $\mathcal{L}_k(G) = G(e^{i\theta_k})$, C_ϵ is at least $(1 - \frac{16\rho\epsilon}{\pi(1-\rho)})$. Other measurement ensembles can be treated similarly.

When we replace $\|x\|_{\mathcal{L}(\mathcal{A})}$ with its discretized counterpart $\|x\|_{\mathcal{L}(\mathcal{A})}$ in (8),

$$\text{minimize}_x \, \tfrac{1}{2} \|x - y\|_2^2 + \mu \|x\|_{\mathcal{L}(\mathcal{A}_\epsilon)}$$

is equivalent to

$$\text{minimize}_c \, \tfrac{1}{2} \|Mc - y\|_2^2 + \mu \sum_{w \in \mathbb{D}_\rho^{(\epsilon)}} |c_w| \qquad (9)$$

where

$$M_{ij} = \mathcal{L}_i \left(\tfrac{1-|w_j|^2}{z-w_j} \right)$$

and j indexes the set $\mathbb{D}_\rho^{(\epsilon)}$. That is M is an $n \times |\mathbb{D}_\rho^{(\epsilon)}|$ matrix. Problem (9) is a weighted ℓ_1 regularization problem with real or complex data depending on specific problem. We call (9) *Discretized Atomic Soft Thresholding* (DAST), as coined in [1].

The DAST problem can be solved very efficiently with general purpose packages such as YALMIP [11] or CVX [8]. DAST yields an approximate solution to problem (6), and, as we will see, yields a statistically consistent estimate provided the parameter ϵ is adjusted to meet the desired numerical accuracy.

V. Statistical Bounds

Let $\mathcal{L}_k : \mathcal{H} \mapsto \mathbb{C}$ be a linear functional that serves as a measurement operator for the system $G(z)$. In this section, let us suppose that we obtain noisy measurements of the form

$$y_k = \mathcal{L}_k \left(G(z) \right) + \omega_k \qquad k = 1, \ldots, n$$

where ω_k is a noise sequence consisting of independent, identically distributed random variables. We will specialize our results to the case where $\mathcal{L}$ returns samples from the frequency response at uniformly spaced frequencies: $z_k = e^{\frac{2\pi i k}{m}}$. Extension to other measurement ensembles will be explored in an extended version of the paper.

Our goal in this section is to prove that solving the DAST optimization problem yields a good approximation to the transfer function we are probing. The following theorem provides a precise statistical guarantee on the performance of our algorithm. We omit the proof in this paper due to space constraints.

Theorem 5.1: Let $G_\star$ be a strictly proper transfer function with bounded Hankel nuclear norm. Suppose the noise sequence ω_k is i.i.d. Gaussian with mean zero and variance σ^2. Choose $\delta \in (0,1)$ and set $\epsilon = \frac{\pi(1-\rho)\delta}{16\rho}$. Let $\mathbb{D}_\rho^{(\epsilon)}$ be as in Proposition 4.1 and let $\hat{c}$ be the optimal solution of (9) with

$$\mu = 2\sigma\sqrt{n\log\left(\frac{11\rho^2}{\delta(1-\rho)}\right)}.$$

Set $\hat{G}(z) = \sum_{w\in\mathbb{D}_\rho^{(\epsilon)}} \hat{c}_w \frac{1-|w|^2}{z-w}$. Then if the set of vectors $\{\mathcal{L}(\varphi_a) \in \mathbb{R}^n \ : \ a \in \mathbb{D}_\rho^{(\epsilon)}\}$ spans $\mathbb{R}^n$, we have

$$\|\hat{G}(z) - G_\star(z)\|_{\mathcal{H}_2}^2 \leq$$
$$186\frac{1+\rho}{1-\rho}\left(\sqrt{\sigma^2\log\left(\frac{11\rho^2}{(1-\rho)\epsilon}\right)\frac{\|\Gamma_{G_\star}\|_1^2}{n(1-\delta)^2}} + \frac{4\|\Gamma_{G_\star}\|_1^2}{\pi n(1-\delta)^2}\right)$$

with probability $1 - e^{-o(n)}$.

Corollary 5.2: Under the conditions stated in Theorem 5.1, there is a quantity C depending on ρ and σ such that for sufficiently large n

$$\|\hat{G}(z) - G_\star(z)\|_{\mathcal{H}_2}^2 \ \leq C\|\Gamma_{G_\star}\|_1 n^{-\frac{1}{2}}$$

with probability exceeding $1 - e^{-o(n)}$.

Let us make a few remarks on the main result. First of all, the right hand side is a parameter of the number of samples, the Hankel nuclear norm of the true system, and the stability radius of the true system. Also, if the McMillan degree of $G_\star(z)$ is d, then we can upper bound the Hankel nuclear norm by the product of the McMillan degree and the Hankel norm of $G_\star$: $\|\Gamma_{G_\star}\|_1 \leq d\|\Gamma_{G_\star}\|$. Second, note that as n tends to infinity, the right hand side tends to zero. In particular, this means that our discretized algorithm is consistent, and we can quantify the worst case convergence rate.

VI. Numerical Experiments

In this section we validate the proposed framework via some preliminary numerical experiments conducted in MATLAB. In many of the experiments where the solution of convex optimization problems was required, the software package CVX [8] was used. Throughout our experiments, the discretization of the unit circle was held to approximately 2000 points.

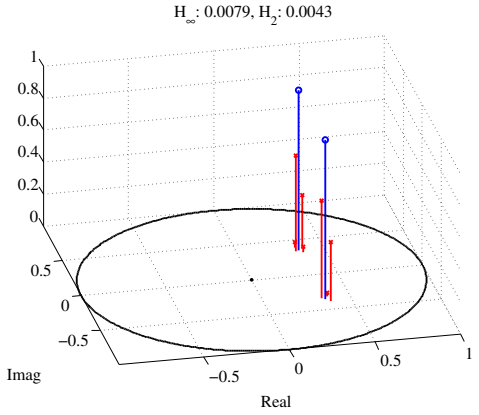

Fig. 1: The locations marked with a circle represent the locations of the poles (in the complex plane) of a second order discrete time LTI system. The locations marked with a cross correspond to poles recovered by DAST.

In the first experiment we consider a stable system G with two poles. We make $m = 80$ noisy observations of the frequency response by evaluating the transfer functions $G(z_j)$ at regularly spaced frequencies $z_j = e^{i\theta_j}$ on the complex unit circle. The noise is additive i.i.d. zero-mean Gaussian with a variance of $\sigma^2 = 10^{-4}$. We reconstruct $\hat{G}(z)$ by DAST as proposed in section IV. Our algorithm recovers a system of degree 6 which achieves an $\mathcal{H}_2$ performance error of .0043 and $\mathcal{H}_\infty$ error of .0079. The locations of the true and recovered poles are depicted graphically in Fig. 1.

In Fig. 2, we compare our algorithm to a widely used method known as *subspace identification* [10, Chapter 10]. A second order system, starting from an initial condition of $x[0] = 0$ is excited by a random input $u[t]$ corresponding to an i.i.d. sequence of zero-mean, unit-variance Gaussian random variables for m time steps. We record the output $y[t]$ of the system for m time steps. From this input-output relationship, we use DAST and subspace identification to attempt to reconstruct the unknown system. We plot the estimation error in the $\mathcal{H}_2$ norm as m is increased from 10 time units to 120 time units. As is evident, the performance of DAST is superior to that of subspace identification when m is small, i.e. of the order of 10 to 50 measurements.

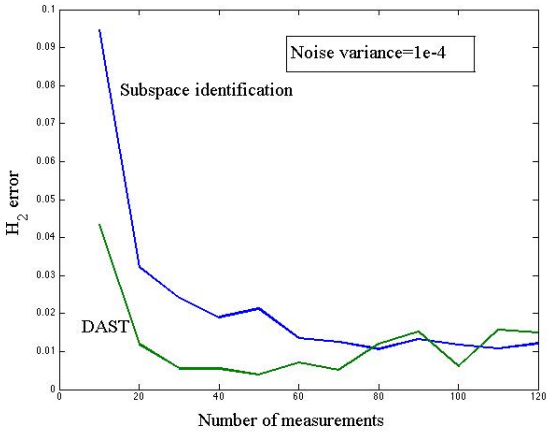

Fig. 2: $\mathcal{H}_2$ estimation error for DAST and subspace identification methods.

Another aspect that we emphasize is that in these experiments, subspace identification was assisted with the knowledge of the true system order. If the wrong model order was used, the performance of subspace identification worsened noticeably. By contrast, DAST does not need knowledge of the true system order.

VII. Conclusion

By using the atomic norm framework of [3], we were able to posit a reasonable regularizer for linear systems, understand the computational demands of such a regularizer, and analyze its statistical performance. Since it is closely connected to the Hankel nuclear norm but is computationally more practical, we believe that our atomic norm will be useful in a variety of practical implementations and also in theoretical analysis.

While we focused on the single-input single-output (SISO) case in this paper, we expect that these techniques would extend to the multi-input multi-output (MIMO) case. The translation to multi-input single-ouput systems is straight forward, one must simply choose the following atomic set:

$$\mathcal{A} = \left\{ \frac{(1-|w|^2)b^*}{z-w} \;:\; w \in \mathbb{D}, b \in \mathbb{S}^{p-1} \right\}.$$

Such a model can be fit (after discretizing the disk), by solving a block ℓ_1 minimization model. After fitting a model for each output, a MIMO model could be acquired by applying model reduction to parallel MISO fits.

References

[1] B. Bhaskar and B. Recht. Atomic norm denoising with applications to line spectral estimation. In *Proceedings of the 49th Annual Allerton Conference*, 2011.

[2] F. F. Bonsall and D. Walsh. Symbols for trace class Hankel operators with good estimates for norms. *Glasgow Mathematical Journal*, 28:47–54, 1986.

[3] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. Submitted to *Foundations of Computational Mathematics*. Preprint available at `arxiv.org/1012.0621`, 2010.

[4] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and Gaussian processes - revisited. In *IFAC World Congress*, pages 2303–2308, 2011.

[5] R. R. Coifman and R. Rochberg. Representation theorems for holomorphic and harmonic functions in L^p. *Asterisque*, 77:11–66, 1980.

[6] Maryam Fazel, Haitham Hindi, and Stephen Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, 2001.

[7] Maryam Fazel, Ting Kei Pong, Defeng Sun, and Paul Tseng. Hankel matrix rank minimization with applications in system identification and realization. Submitted for publication. Preprint available at `http://faculty.washington.edu/mfazel/`, 2011.

[8] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`, May 2010.

[9] Zhang Liu and Lieven Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2009.

[10] Lennart Ljung. *System Identification. Theory for the user*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1998.

[11] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[12] Peter Van Overschee and Bart De Moor. N4SID: Subspace algorithms for the identification of combined deterministic–stochastic systems. *Automatica*, 30:75–93, 1994.

[13] Fernando G. Paganini. A set-based approach for white noise modeling. *IEEE Transactions on Automatic Control*, 41(10):1453–1465, 1996.

[14] Jonathan R. Partington. *An Introduction to Hankel Operators*. Cambridge University Press, 1989.

[15] Vladimir V. Peller. Nuclearity of hankel operators. Technical Report E-I-79, LOMI Preprints, Leningrad, 1979.

[16] Vladimir V. Peller. *Hankel Operators and Their Applications*. Springer Monographs in Mathematics. Springer, Berlin, 2003.

[17] G. Pillonetto, A. Chiuso, and G. de Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1):81–93, 2010.

[18] G. Pillonetto, A. Chiuso, and G. de Nicolao. Prediction error identification of linear systems. *Automatica*, 47(2):291–305, 2011.

[19] Benjamin Recht, Maryam Fazel, and Pablo Parrilo. Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[20] Roy S. Smith. Nuclear norm minimization methods for frequency domain subspace identification. In *Proceedings of the American Control Conference*, 2012.

[21] Michel Verhaegen and Patrick Dewilde. Subspace model identification. *Int. J. of Control*, 56(5):1187–1210, 1992.

[22] B. Wahlberg. System identification using La guerre models. *IEEE Transactions on Automatic Control*, 36(5):551–562, 1991.

[23] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, New Jersey, 1995.

[24] Kehe Zhu. *Operator Theory in Function Spaces*. American Mathematical Society, Providence, 2007.