MIMO System Identification by Randomized Active-Set Methods

Jared Miller[1], Rajiv Singh[2], Mario Sznaier[1]

***Abstract*— MIMO (multi-input multi-output) system identification is a particular instance of a parsimonious model selection problem. If the observed data is assumed to arise from a stable and low order plant, then the representing model should also be stable and have few poles in its realization. These constraints are challenging to impose in an L_1 or nuclear norm framework, especially when observations are non-uniformly sampled. This paper implements MIMO identification by randomized active-set methods, as realized by Fully Corrective Frank-Wolfe (FCFW). Reweighting pole-group penalties allow for further system sparsification while monotonically decreasing the regularized fitting error. Efficacy of the approach is shown on two examples.**

I. INTRODUCTION

Identification of minimal order dynamic models that fit the data in a certain prediction sense is generally a hard non-convex problem. Typical relaxations are based on nuclear norm minimization (leading to subspace based approaches) that may be ill-conditioned and do not guarantee model stability [1]. Incorporating requirements such as stability or limits on bandwidth or overshoot into the objective formalism also leads to substantial increase in the computational complexity. Atomic set based formulations for SISO (single-input single-output) system identification have been proposed that seek to soften or overcome the non-convexity and computational complexity issues [2]. Atomic formulations can be solved using sparsity-inducing Frank Wolfe methods.

A randomized SISO system identification based on noisy time and frequency responses with the possibility of missing data was presented by Yilmaz, Bekiroglu *et. al.* in [3], [4], [5]. Their algorithm samples the unit disk to find candidate poles, and this stochastic Frank-Wolfe process converges in expectation to the true solution of a convex surrogate objective. Plant characteristics such as overshoot can be encoded by adding sector bounds to the sampling region. They use a Forward Frank-Wolfe algorithm that never drops poles when added to the system. Their algorithm therefore performs a denoising of input data by ensuring the estimated response arose from a stable system, but the resulting systems are typically of very high order (*e.g.* ≈ 1000 poles). A follow-up paper treated the MIMO case through a fixed set of atoms using an ℓ_1/ℓ_∞ penalty, but did not perform randomization or Frank-Wolfe over this new atomic domain [6].

[1]J. Miller and M. Sznaier are with the ECE Department, Northeastern University, Boston, MA 02115. (Emails: miller.jare@husky.neu.edu, msznaier@coe.neu.edu).

[2]R. Singh is with The MathWorks, Inc., 1 Apple Hill Drive, Natick, MA 01760 USA (e-mail: rsingh@mathworks.com)

J. Miller and M. Sznaier were partially supported by NSF grants CNS–1646121, CMMI–1638234, IIS–1814631 and ECCS—1808381 and AFOSR grant FA9550-19-1-0005.

This paper presents a two-stage algorithm (Alg. 2) that monotonically decreases fitting error to perform active-set MIMO identification. The contributions of this paper over [3], [4], [5], [6] are:

- The use of a two-stage Fully Corrective Frank-Wolfe Algorithm with linear convergence rate over the sublinear greedy Forward Frank-Wolfe scheme in [3]
- Incorporation of prior frequency domain knowledge for a MIMO system
- A sparsification algorithm that maintains feasibility and monotonically reduces fitting error and model order

Fig. 1 illustrates the reduced order and fitting error of models found by our work (Alg. 2) over the approach in [5].

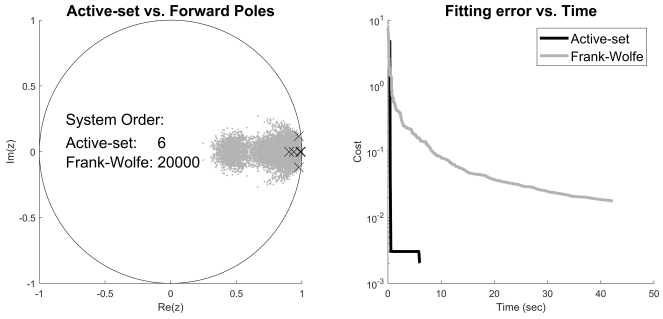

Fig. 1. SISO identification by the proposed two-stage Active-set (Fully Corrective Frank-Wolfe Algorithm) vs. the Forward algorithm proposed in [5]. Forward only Frank-Wolfe running for 10000 iterations leads to a very high order system (gray), and each iteration adds 2 complex-conjugate poles. The proposed Active Set approach results in a 6^{th} order system with a maximum intermediate order of 8. The right pane shows a faster decrease in error in FCFW vs. Forward algorithm.

The paper is organized as follows: Section II reviews preliminaries such as the Frank-Wolfe algorithm and the approach proposed in [5]. Section III presents the two-stage MIMO identification algorithm. Section IV demonstrates this algorithm on a SISO and MIMO identification problem using mixed domain data. Section V concludes the paper.

II. PRELIMINARIES

A. Notation

$\mathbb{R}, \mathbb{C}$	Real and Complex numbers
$\mathbb{D}_\rho$	Closed disc of radius ρ in $\mathbb{C}$
j	$\sqrt{-1}$
$\dagger$	Conjugate Transpose
G_{ij}	Transfer function from input j to output i
$\text{conv}(\mathcal{A})$	Convex Hull of Set $\mathcal{A}$
z^{-1}	Delay Operator
T_u	Toeplitz Matrix of a vector u
$\otimes$	Kronecker Product
$\odot$	Hadamard Product

B. Atomic Norms

Let $\mathcal{H}$ be a Banach space with inner product $\langle \cdot, \cdot \rangle$, and $\{a_i\}_{i=1}^I \in \mathcal{H}$ be a set of I vectors called 'atoms'. The collection of $\{a_i\}$ is an 'atomic set' $\mathcal{A}$, and has a polyhedral convex hull if I is finite. Elements $x \in \text{cone}(\mathcal{A}) \subset \mathcal{H}$ are conic combinations of $\{a_i\}$: there exists nonnegative weights $c_i \geq 0$ such that $x = \sum_{i=1}^I c_i a_i$. An atomic 'norm' $\|\cdot\|_{\mathcal{A}}$ can be defined over $\mathcal{A}$ for elements $x \in \text{cone}(\mathcal{A})$:

$$\|x\|_{\mathcal{A}} = \inf_t t : x \in t \, \text{conv}\mathcal{A} = \inf_{c_i \geq 0} \left\{ \sum_i c_i : x = \sum_i c_i a_i \right\}$$

The quantity $\|x\|_{\mathcal{A}}$ measures the size of x with respect to $\mathcal{A}$ based on how much the convex hull $\mathcal{A}$ must stretch to contain x. The particular choice of c_i with respect to $\mathcal{A}$ to represent an element x is referred to as the 'loading'. If the optimal representation of x has weights $c_i \geq 0$, then the active set (support) $\mathcal{S} = \text{supp}(x)$, $\mathcal{S} \subseteq \mathcal{A}$ is the set of atoms such that $c_i > 0$. Atomic norms are only true norms if $\mathcal{A}$ is centrally symmetric: $a \in \mathcal{A}$ iff $-a \in \mathcal{A}$ $\forall a \in \mathcal{A}$ [7]. Atomic norms are a convex surrogate for minimizing the number of active atoms $\|c\|_0$ such that $x = \sum_i c_i a_i$.

C. Frank Wolfe

Frank-Wolfe (conditional gradient) algorithms are first-order methods to solve constrained optimization problems. For a convex $f(x)$, such a problem is below:

$$\bar{x} = \underset{x}{\text{argmin}} \, f(x) \qquad \|x\|_{\mathcal{A}} \leq \tau \tag{1}$$

Refer to [7] for more detail about the Frank-Wolfe process. Algorithm 1 implements Forward Frank-Wolfe with stepsizes α^* chosen by the line minimization rule where x is within an atomic ball of radius τ.

Algorithm 1 Frank-Wolfe with Line Minimization Rule

$x^0 = \mathbf{0}$
for $k = 0 \ldots T$ **do**
  $a_t = \text{argmax}_{a \in \mathcal{A}} \langle -\nabla f(x^k), a \rangle$ **LMO**
  $\alpha^* = \text{argmin}_{\alpha \in [0,1]} f(x^k + \alpha \tau a_t)$
  $x^{k+1} = x_k + \alpha \tau a_t$
end for

The key step in Frank-Wolfe process is the Linear Minimization Oracle (LMO) to choose new atoms. The new atom a_t produces a surrogate duality gap $DG = \langle -\nabla f(x), a_t \rangle$ which upper bounds suboptimality: $f(x) - f(\bar{x}) \leq DG$. Randomized Frank-Wolfe schemes to approximate the LMO have an expected guarantee of convergence [8].

Fully Corrective Frank-Wolfe (FCFW) is a Frank-Wolfe variant which optimizes a loading $c_i \geq 0, \sum_i c_i \leq \tau$ over the active set of atoms. The LMO selects a new atom, and a correction step optimizes the loading over the augmented set of atoms. If $\mathcal{A}$ is a finite set, the constrained-optimal point $\bar{x}$ will be reached within floating-point precision in finite time. FCFW also has a linear convergence rate in terms of duality gap [9], but performing correction steps may be as difficult as the solving the original problem. If $f(x)$ in Equation (1) is quadratic, Active Set techniques for solving quadratic programs can be used in the correction step to perform FCFW [10].

D. Frank-Wolfe for SISO System Identification

A SISO system can be represented by a scalar transfer function $G(z) = Y(z)/X(z)$. A stable SISO $G(z)$ can be approximated to arbitrary precision by:

$$\hat{G}(z) = \sum_{k=1}^{N_1} \frac{c_k^{\text{exp}}}{z - p_k} + 2 \sum_{k'=1}^{N_2} \frac{c_{k'}^{\cos}(z - a_{k'}) + c_{k'}^{\sin} b_{k'}}{z^2 - 2a_{k'}z + a_{k'}^2 + b_{k'}^2} \tag{2}$$

under the constraints that each $p_k, a_{k'}^2 + b_{k'}^2 \in [0, \rho)$ and $c_k^{\text{exp}}, c_{k'}^{\cos}, c_{k'}^{\sin} \in \mathbb{R}$. A pole is 'active' for $\hat{G}(z)$ if it appears in Eq. (2), and the active real and complex poles are p_k and $a_{k'} \pm j b_{k'}$. The system order of this approximation is $N_1 + 2N_2$. Repeated poles can be approximated by perturbations [11]. This forms an atomic decomposition of the transfer function $G(z)$. Yilmaz et. al. characterized the atoms comprising any stable discrete time transfer function (up to approximation) [5]. Given a set of poles $S \in \mathbb{D}_\rho$, a set of transfer functions based on S are:

$$\mathcal{A}_1(S) = \left\{ \pm \alpha_p^1 \left(\frac{1}{z - p} + \frac{1}{z - \bar{p}} \right) \right\} \tag{3a}$$

$$\mathcal{A}_2(S) = \left\{ \pm \alpha_p^2 \left(\frac{-j}{z - p} + \frac{j}{z - \bar{p}} \right) \right\} \tag{3b}$$

$$\mathcal{A}_3(S) = \left\{ \pm \alpha_p^3 \left(\frac{1}{z - p} \right) \right\} \tag{3c}$$

$$\mathcal{A}_4(S) = \{\pm 1\} \tag{3d}$$

$$\mathcal{A}(S) = \mathcal{A}_1(S) \cup \mathcal{A}_2(S) \cup \mathcal{A}_3(S) \cup \mathcal{A}_4(S) \tag{3e}$$

The atomic set of transfer functions based on S is $\mathcal{A}(S)$. $\alpha_p^{1,2,3}$ are normalization constants to maintain a unit Hankel norm among the transfer functions. The time-domain impulse response g is the inverse Z-transform of $G(z)$. If u is an applied discrete time-domain input (vector of length N_s), then the discrete time output is $y = T_u g$ where T_u is a Toeplitz matrix of u. Atomic norm-based low order stable system identification can proceed by solving the following optimization problem given a recorded input u and recorded output y:

$$E_{min} = \min_{\|g\|_{\mathcal{A}} \leq \tau} \|y - T_u g\|_2^2 \tag{4}$$

The regularization parameter τ is a proxy for system order. An increase in τ generally leads to a rise in system approximated order. Frank-Wolfe, where S is all poles in $\mathbb{D}_\rho$, requires solving a NP-hard polynomial optimization problem at each LMO. Tractable choices of S for Frank-Wolfe include a fine gridding (ϵ-net) that discretely samples the disk $\mathbb{D}_\rho$ [2]. Randomized sampling of $\mathbb{D}_\rho$ converges to the true solution of (4) in expectation [5], but the solution system is usually of high order. Prior knowledge of system characteristics such as rise time and overshoot can restrict poles in S to sector bounds.

E. MIMO Model Structure

Let $G(z)$ be a discrete time MIMO plant with N_u inputs and N_y outputs. Each input-output rational transfer function $[G(z)]_{ij} = N_{ij}(z)/P_{ij}(z)$ has poles and zeros corresponding to the roots of $N_{ij}(z)$ and $P_{ij}(z)$. The poles of $G(z)$ are the roots of the least common multiple $P(z) = \text{LCM}_{ij}(P_{ij}(z))$ over all input-output responses with multiplicities. The order of $G(z)$ is defined as the degree of $P(z)$. Multiple input-output systems $G(z)_{ij}$ can therefore share the same poles $P(z)$ without increasing system order.

In a system identification framework, output measurements of $G(z)$ are recorded from applied inputs as corrupted by a noise process. This measurement model is:

$$(y_i)_n = (g_{ij} * u_j)_n + (\eta_i)_n \tag{5}$$

where $\{u_j\}_{j=1}^{N_u}$ are the channel inputs, $*$ is convolution, $\{y_i\}_{i=1}^{N_y}$ are the measured noisy outputs, and η_i is the noise process. The input u_j must be persistently exciting in order to stimulate all modes of the system, which in this paper is obtained by white noise. The measurement $(\eta_i)_n$ is likewise modeled as additive white Gaussian noise (AWGN). The system is observed over N_s samples, and the time index n ranges from $n = 1 \dots N_s$. This identification task is conducted under the prior that $G(z)$ is a low order system, and therefore that $P(z)$ has a low degree.

III. ACTIVE SET MIMO

System identification through FCFW can proceed on MIMO models with variations from the SISO procedure.

A. MIMO Atomic Domains

MIMO systems exhibit a group sparsity structure in the choice of poles, and cannot simply be treated as a set of independent SISO transfer functions $[G(z)]_{ij}$.

$$\hat{G}_{ij}(z) = \sum_{k=1}^{N_1} \frac{c_{ijk}^{\exp}}{z - p_k} + 2\sum_{k'=1}^{N_2} \frac{c_{ijk'}^{\cos}(z - a_{k'}) + c_{ijk'}^{\sin} b_{k'}}{z^2 - 2a_{k'}z + a_{k'}^2 + b_{k'}^2}$$

In a SISO system, a pole p_k is present in the identified transfer function $\hat{G}(z)$ if the respective loading term $c_k \neq 0$. In the MIMO case, p_k is present if $\exists (i,j) \mid c_{ijk}^{\exp} \neq 0$. Likewise, the complex pole $a \pm jb$ is present if $\exists (i,j) \mid (c_{ijk}^{\cos}, c_{ijk}^{\sin}) \neq (0,0)$. This grouping structure can be carried to an atomic framework based on the pole-atoms $\mathcal{A}(S)$. Let $\{u_j\}_{j=1}^{N_u}$ be per-channel inputs and $\{y_i\}_{i=1}^{N_y}$ contain corresponding corrupted output over N_s samples. The low-order MIMO optimization problem is therefore:

$$E_{min} = \min_{\|\hat{G}\|_{\mathcal{A}} \leq \tau} \sum_{i=1}^{N_y} \|y_i - \sum_{j=1}^{N_u} T_{u_j} \hat{g}_{ij}\|_2^2 \tag{6}$$

B. Least Squares Formulation

The MIMO optimization model in (6) can be further manipulated into a structured and constrained quadratic program. Let $P \in \mathbb{R}^{N_s \times N_p}$ be a dictionary matrix of pole N_s-length pole in $\mathcal{A}(S)$. The N_s-length SISO response to an input u is $y = T_u P c$. The MIMO model has coefficients $c_{ij} = c_{ijk}^{\exp} \cup c_{ijk'}^{\cos} \cup c_{ijk'}^{\sin}$ for each input-output pair (i,j), and the output at channel i is $y_i = \sum_{j=1}^{N_u} T_{u_j} P c_{ij}$.

A typical Least Squares problem has the form $f(c) = \frac{1}{2}\|Ac - b\|_2^2$ for a data matrix A and an answer vector b. In the MIMO case $b = [y_1^T, y_2^T \dots y_{N_y}^T]^T$. If coefficients c are arranged as $c = [c_{1,1}^T, c_{1,2}^T \dots c_{1,N_u}^T \dots c_{N_y,N_u}^T]^T$ and the input is $T_u = [T_{u_1} \dots T_{u_{N_u}}]$, then the data matrix A for a set of poles S has the form:

$$A_S^{time} = I_{N_y} \otimes (T_u(I_{N_u} \otimes P)) \tag{7}$$

An atomic penalty can be defined on the coefficients c by using a mixed ℓ_1/ℓ_∞ penalty. Based on the atomic set $\mathcal{A}(S)$ of pole responses, define a partition of coefficients $g_{poles} = g_{real} \cup g_{comp}$. The real exponential responses and optionally constant signals $\mathcal{A}_3(S) \cup \mathcal{A}_4(S)$ are in coefficient group $g_{real} = \cup_{ij} c_{ijk}^{\exp}$, and the complex exponential impulse responses in $\mathcal{A}_1(S) \cup \mathcal{A}_2(S)$ are in coefficient group $g_{comp} = \cup_{ij} c_{ijk'}^{\cos}\ c_{ijk'}^{\sin}$. A group-sparsity inducing atomic penalty can be formulated based on this structure:

$$\|c\|_{\mathcal{A}} = \sum_{k=1}^{N_1} \|c_{ijk}^{\exp}\|_\infty + 2\sum_{k'=1}^{N_2} \|c_{ijk'}^{\cos}\ c_{ijk'}^{\sin}\|_\infty \tag{8}$$

The operation $\|c_{ijk}^{\exp}\|_\infty$ yields the maximum coefficient across all input-output pairs for real pole k. In this paper, the ℓ_1/ℓ_∞ norm is used to promote group sparsity with penalties $\|c_{ij}\|_\infty$ [12]. The problem in [6] uses a mixed ℓ_1/ℓ_∞ penalty group together complex poles. The MIMO optimization problem is therefore:

$$E_{min} = \min_{\|c\|_{\mathcal{A}} \leq \tau} \|A_S^{time} c - b^{time}\|_2^2 \tag{9}$$

Gradients and residuals of $f_S(c) = \frac{1}{2}\|A_S^{time} c - b^{time}\|_2^2$ can be evaluated quickly by exploiting the high degree of structure in A_S^{time}. The operators and gradient are:

$$A_S^{time} c_i = \sum_{j=1}^{N_u} T_{u_j} P c_{ij} \tag{10a}$$

$$A_S^{time^T} b_{ij} = P^T T_{u_j}^T b_i \tag{10b}$$

$$\nabla f_S(c) = A_S^{time^T}(A_S^{time}(c) - b) \tag{10c}$$

C. Frequency Response

Information about the plant's frequency response G_{ij} can be incorporated into optimization problem (6). These responses may arise from sources - direct experimentation using sine-sweeps and data collection using a spectrum analyzer, prior information (e.g., manufacturer supplied charts), or even constructed by empirical estimated on periodic responses (such as by using `etfe` function of MATLAB® System Identification Toolbox™ [13]). The objective function in optimization may be enriched by adding frequency regularization for a set of weighting functions W_{ij}:

$$\begin{aligned} E_{min} = \min_{\|\hat{G}\|_{\mathcal{A}} \leq \tau} & \sum_i \|y_i - \hat{y}_i\|_2^2 \\ & + \sum_{i,j} \|W_{ij} \odot (G_{ij} - \hat{G}_{ij})\|_2^2 \end{aligned} \tag{11}$$

E_{min} thus represents a trade-off between the time-domain and frequency-domain descriptions of the energy in the error signal. The frequency response $\hat{G}_{ij}$ is the transfer function $\hat{G}$ evaluated at targeted frequencies ($z \to e^{j\omega}$). The weighting functions $W_{ij}(\omega)$ act as regularizers and can encourage agreement in specific bands of the frequency spectrum. Given a set of poles S with frequency responses $F \in \mathbb{C}^{N_f \times N_p}$, the frequency data operators and gradients are:

$$b_{ij}^{freq} = W_{ij} \odot G_{ij} \tag{12a}$$

$$A_S^{freq} c_{ij} = W_{ij} \odot (F c_{ij}) \tag{12b}$$

$$A_S^{freq^T} b_{ij} = \text{Re}(F^\dagger (\bar{W}_{ij} \odot b_{ij})) \tag{12c}$$

The time-frequency answer vector and data operators are:

$$b = [b^{time};\ b^{freq}] \tag{13a}$$

$$A_S c = [A_S^{time} c;\ A_S^{freq} c] \tag{13b}$$

$$A_S^T b = A_S^{time^T} b^{time} + A_S^{freq^T} b^{freq} \tag{13c}$$

D. Active Set approaches

MIMO system identification (with or without frequency information) can be solved through any suitable atomic norm minimization routine. FCFW methods are one possible approach to solve these optimization problems. The MIMO atomic set $\mathcal{A}(S)$ contains a non-overlapping block structure in $g_{poles} = g_{real} \cup g_{comp}$. The linear minimization oracle ($\text{LMO}(c) = \text{argmax}_{a \in \mathcal{A}} \langle -\nabla f(c), a \rangle$) for the ℓ_1/ℓ_∞ penalty in equation (8) can be computed efficiently. For each pole p_k and $a_{k'} \pm j b_{k'}$ compute:

$$n_k^{real} = \|\nabla f(c)_{ijk}^{\text{exp}}\|_\infty \tag{14a}$$

$$n_{k'}^{comp} = \|\nabla f(c)_{ijk'}^{\cos}\ \nabla f(c)_{ijk'}^{\sin}\|_\infty / 2 \tag{14b}$$

If $\max_k n_k^{real} > \max_{k'} n_{k'}^{comp}$, a first order system atom is added to the active atomic set $\mathcal{S} \subset \mathcal{A}$. This subsystem is:

$$s_{ij}^{real} = \frac{\alpha_k}{z - p_k} \text{sign}(-\nabla f(c)_{ijk}^{\text{exp}}) \tag{15}$$

where α_k is a normalization constant. In the alternative case where a complex pole is chosen, a second order atom $s_{k'} = \alpha_{k'} \text{sign}(-\nabla f(c)_{k'}^{\cos}\ -\nabla f(c)_{k'}^{\sin})$ will be added instead.

After adding a new atom, optimize the loading over all atoms $\mathcal{S}_{aug}$ by solving an active-set quadratic program [10]. This allows for model order to be dropped if the loading over any pole-subsystem is zero.

Pole randomization may be integrated into the MIMO identification framework, similar to the SISO case [5]. Let an initial set of poles S be randomly picked from the unit circle and used to identify a system $G(z)$. The recovered system $\hat{G}$ will have poles $S_{active} \subseteq S$. Randomly sampling new poles S_{new} and solving the same MIMO optimization problem for the augmented set $S_{aug} = [S_{active}\ S_{new}]$ will result in a cost $E_{aug} \leq E_{min}$. This process results in a non-strictly decreasing sequence of cost functions that converges in expectation to the infinite-dimensional optima over $S = \mathbb{D}_\rho$ [5]. Computation time between iterations can be reduced by exploiting warm-starts in the active set routine.

E. Sparsification

The ℓ_1/ℓ_∞ penalty in Equation (8) is a convex approximation to finding the minimum-order-system. One sparsification approach is reweighting the ℓ_1/ℓ_∞ in a similar fashion as the Reweighted L_1 norm [14]. This reweighting process is a greedy method that does not add any new poles while monotonically decreasing fitting error and sparsifying the system. [6] also performs reweighting on pole-groups in its soft atomic penalty to promote sparsity. Because their objective functions change between iterations, models generated are therefore incomparable outside of system order. In contrast, reweighting the hard atomic constraint allows for models to be compared due to the decreasing fitting error. Reweighting is illustrated in Figure 2, where a noise-corrupted 3-input 2-output 4th order system is approximated based on a time-frequency penalty.

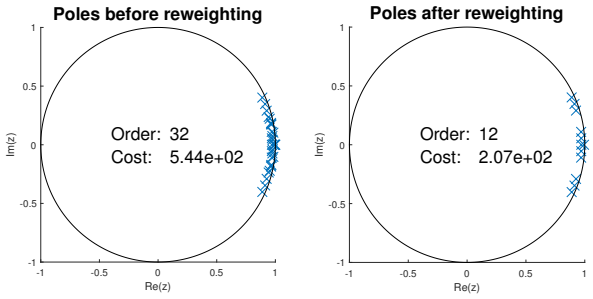

Fig. 2. Identification with prior knowledge of pole bounds: $|z| \in [0.97, 1)$

After solving the MIMO optimization problem for a solution c with active groups c_k, compute group norms $\|c_k\|_\infty$ to find weights on the next round (for a small $\epsilon > 0$).

$$w_k^{real} = \frac{1}{\|c_k\|_\infty + \epsilon} \qquad w_{k'}^{comp} = \frac{2}{\|c_{k'}^{cos}\ c_{k'}^{sin}\|_\infty + \epsilon} \tag{16}$$

$$\tilde{w} = \frac{\tau w}{\sum_{k=1}^{N_1} w_k^{real} \|c_k\|_\infty + \sum_{k'=1}^{N_2} w_{k'}^{comp} \|c_{k'}^{cos}\ c_{k'}^{sin}\|_\infty} \tag{17}$$

Equation (17) normalizes the weights to τ such that the point c maintains feasibility on the old and new atomic ball. The new atomic norm is:

$$\|c\|_{\mathcal{A},\tilde{w}} = \sum_{k=1}^{N_1} \tilde{w}_k^{real} \|c_{ijk}^{\text{exp}}\|_\infty + \sum_{k'=1}^{N_2} \tilde{w}_{k'}^{comp} \|c_{ijk}^{\cos}\ c_{ijk}^{\sin}\|_\infty \tag{18}$$

The reweighting process penalizes modes with low $\|c_k\|_\infty$, and will converge to an accumulation point based on the same analysis as in [14].

F. Final Algorithm

Algorithm 2 details a combined MIMO fitting routine.

Figure 3 demonstrates Algorithm 2 on a 3-input 2-output system with AWGN input. The monotonic decrease in fitting error throughout execution are observed in orange, and system order (blue) is reduced in the Reweight stage.

Algorithm 2 MIMO fitting $(y, u, G_{ij}) \rightarrow \hat{G}_{ij}$

$c = 0,\ S = \varnothing$
while Random Iterations **do**
  Randomly pick poles S_{new}
  $S \leftarrow S \cup S_{new}$
  Use FCFW to minimize $\|A_S c - b\|_2^2$ over $\|c\|_{\mathcal{A}} \leq \tau$
  $S \leftarrow \text{supp}(c)$
end while
while Reweight Iterations **do**
  Compute weights $\tilde{w}$ given c and groups g_{poles}
  Use FCFW to minimize $\|A_S c - b\|_2^2$ over $\|c\|_{\mathcal{A},\tilde{w}} \leq \tau$
  $S \leftarrow \text{supp}(c)$
end while
Generate $\hat{G}$ from (S, c)

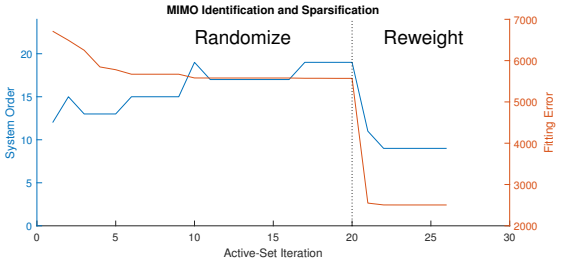

Fig. 3. Fitting error and System Order during Alg. 2 execution

IV. Illustrative Examples

A. Data Based Model Reduction: SISO example

This examples emulates a data based model reduction application for a flexible structure such as a steel plate. We simulate step and frequency responses for 400 seconds and over the frequency range [5, 500] mHz. The first two flexible modes of the structure (roughy around frequencies 14 mHz and 50 mHz) are of practical interest where we need good model accuracy. We use a weighting function that is proportional to the inverse of the amplitude of the measured frequency response, with small weights at frequencies $\geq$ 80 mHz. The prior knowledge is the location of the first two modes and that they are lightly damped. We use this knowledge to limit the pole sampling region to a circle sector defined by radius $\in (0.9, 1)$ and angle $\in (-30^o, 30^o)$. Using $\tau = 1e-4$, 1000 atoms in each iteration with 20 randomization and 10 reweighting iterations, we get a model of order 54. An Hankel Singular Value plot of the resulting model suggests further scope for reduction, leading finally to a model of order 6. Results were also generated using the randomized Frank Wolfe algorithm of [5], which yielded a model with 38 poles although the implied complexity was lower than ideal; HSVD suggested an order of 4. The fit of this model to the measured frequency and step responses are shown in figures 4 and 5 respectively.

B. MIMO Identification of Two Tank System

This system consists of the cascaded of two water tanks as shown schematically in figure 6 [15]. The upper tank (tank 1) is fed by hot and cold water via computer-controlled valves. The lower tank (tank 2) is fed by water from an exit at the bottom of tank 1. An overflow maintains a constant level in tank 2. A cold water bias stream also feeds tank 2 and enables the tanks to have different steady-state temperatures. Inputs are the hot and cold water flowing into tank 1, while the outputs are the water temperatures in the two tanks.

The apriori knowledge of the system is as follows:

- The system is stable but has some long time constants.
- We expect the step response to be non-oscillatory.
- There is a large and frequency dependent uncertainty in the system owing to the mixing and heating losses. Uncertainty is low at low frequencies ($\omega < 1$ mHz) and a linear model can provide a good approximation, and grows significantly at high frequencies ($\omega > 4.6$ mHz).

For identification we simulate two experiments. First, the hot and cold water flows are stepped up one at a time and the corresponding temperatures determined. Second, the input variables are varied in periodic fashion and the resulting responses are collected. The noise introduced in the simulated data is in accordance with the expected uncertainty. We identify a 6^{th} order model using the following settings: $\tau = 20$, sampling 1000 atoms in each randomization iteration. The selection of poles is limited to the real segment

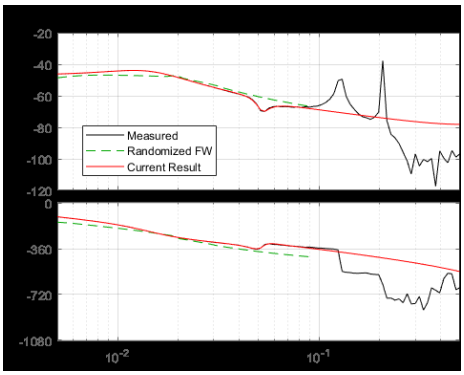

Fig. 4. Model fit to FRF. Note that frequencies above 80 mHz were ignored. Randomized FW result uses model of order 38 while the current result is based on order of 6

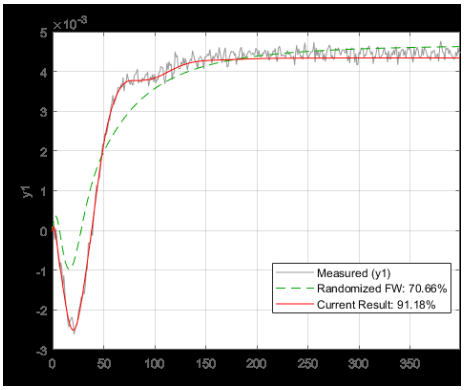

Fig. 5. Model fit to the step response. The reported percent fit uses the NRMSE [13] measure (Percent fit = (1-NRMSE)*100). The Randomized FW result misses certain higher frequencies

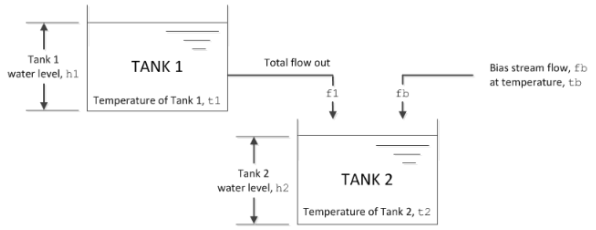

Fig. 6. Schematic diagram of a two-tank system

$\in (0.7, 1)$ since we expect the poles to be real and relatively low damped. The weighting profile is shown in figure 7.

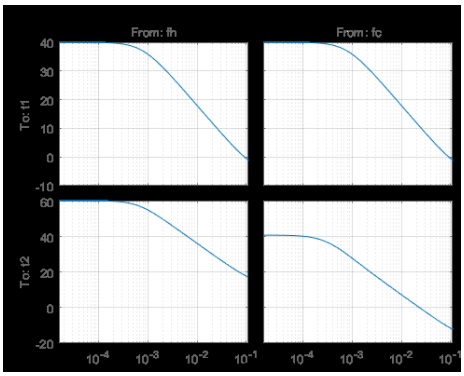

Fig. 7. Weighting function for the frequency response penalty

After 20 randomization and 4 reweighting iterations, a 14^{th} order model is obtained, which is further reduced to 6^{th} order based on Hankel singular value analysis of the model. The Randomized FW method failed to produce good results and are not reported. The obtained results are as shown in figures 8 and 9.

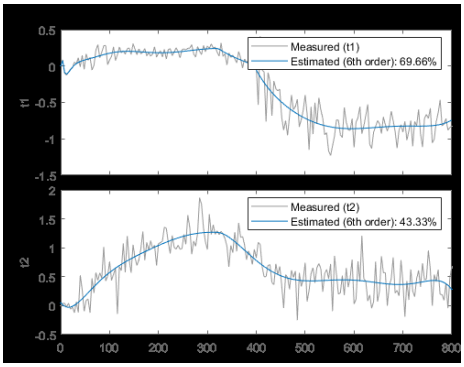

Fig. 8. Time response comparison

V. CONCLUSION

We conclude the our algorithm offers an attractive approach for mixed domain identification of MIMO systems. It allows easy incorporation of system priors related to system dynamics and data reliability. It allows us to implement a notion of group sparsity which helps us identify systems of minimal McMillan degree. Examples demonstrate that the FCFW-algorithm is faster and yields better results than algorithms in [5]. Future work includes control synthesis.

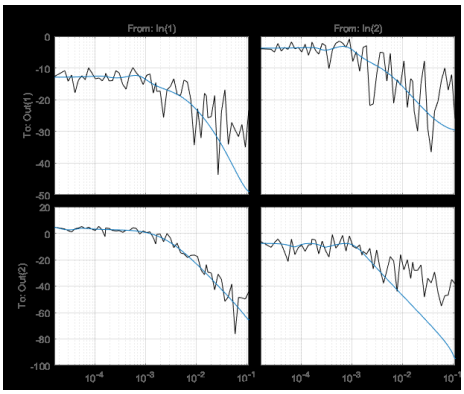

Fig. 9. Frequency response comparison

REFERENCES

[1] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, "Regularized linear system identification using atomic, nuclear and kernel-based norms: The role of the stability constraint," *Automatica*, vol. 69, pp. 137–149, 2016.

[2] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, "Linear system identification via atomic norm regularization," in *51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 6265–6270.

[3] B. Yilmaz, C. Lagoa, and M. Sznaier, "An efficient atomic norm minimization approach to identification of low order models," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 5834–5839.

[4] K. Bekiroglu, B. Yilmaz, C. Lagoa, and M. Sznaier, "Parsimonious model identification via atomic norm minimization," in *2014 European Control Conference (ECC)*. IEEE, 2014, pp. 2392–2397.

[5] B. Yılmaz, K. Bekiroglu, C. Lagoa, and M. Sznaier, "A randomized algorithm for parsimonious model identification," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 532–539, 2017.

[6] K. Bekiroglu, C. Lagoa, and M. Sznaier, "Low-order model identification of mimo systems from noisy and incomplete data," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 4029–4034.

[7] M. Jaggi, "Revisiting frank-wolfe: Projection-free sparse convex optimization," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13, 2013, pp. I–427–I–435.

[8] E. Frandi, R. Nanculef, S. Lodi, C. Sartori, and J. A. Suykens, "Fast and scalable lasso via stochastic frank-wolfe methods with a convergence guarantee," *arXiv preprint arXiv:1510.07169*, 2015.

[9] S. Lacoste-Julien and M. Jaggi, "On the global linear convergence of frank-wolfe optimization variants," in *Advances in Neural Information Processing Systems*, 2015, pp. 496–504.

[10] M. Vinyes and G. Obozinski, "Fast column generation for atomic norm regularization," 2017.

[11] P. Van Dooren, K. A. Gallivan, and P.-A. Absil, "H_2-optimal model reduction with higher-order poles," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2738–2753, 2010.

[12] P. Zhao, G. Rocha, and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *Annals of statistics*, vol. 37, no. 6A, pp. 3468–3497, 2009.

[13] L. Ljung, *System Identification Toolbox for use with MATLAB, Version 9.7*, 9th ed., Natick, MA, USA, 2017.

[14] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[15] R. S. Smith and J. Doyle, "The two tank experiment: A benchmark control problem," *1988 American Control Conference*, pp. 2026–2031, 1988.