Finite Sample Identification: From Frequency to Time Domain

Anastasios Tsiamis [1] **Mohamed Abdalmoaty** [1] **Roy S. Smith** [1] **John Lygeros** [1]

Abstract

We review frequency domain system identification under finite samples and study its implications to time domain identification. We focus on an open loop setting where the excitation input is periodic and consider the Empirical Transfer Function Estimate (ETFE), i.e., we estimate the frequency response at certain desired evenly-spaced frequencies from input-output data. Under sub-Gaussian colored noise and certain stability assumptions, we can establish finite-sample guarantees for the ETFE. The estimates are concentrated around the true values with an error rate of the order of $\tilde{\mathcal{O}}(\sqrt{M/N_{\mathrm{tot}}})$, where N_{tot} is the total number of samples and M is the number of desired frequencies. Given the ETFE, we show that by tuning the number of frequencies M, we can recover the time domain impulse responses of the system at a rate of $\tilde{\mathcal{O}}(N_{\mathrm{tot}}^{-1/3})$ for general irrational systems and $\tilde{\mathcal{O}}(N_{\mathrm{tot}}^{-1/2})$ for state-space systems.

1. Introduction

We study frequency domain identification, where the goal is learning the frequency response of an unknown linear system from input-output data. The problem has been studied extensively in the classical system identification literature (Ljung, 1999; Schoukens et al., 2004; Pintelon & Schoukens, 2012). The estimation error guarantees (on its distribution) are typically asymptotic, and, thus, are valid when the number of samples grows to infinity.

Instead, here, we adopt a finite-sample point of view, inspired by the extensive work over the past years (Goldenshluger, 1998; Faradonbeh et al., 2018; Simchowitz et al., 2018; Oymak & Ozay, 2021; Sarkar et al., 2021; Tsiamis & Pappas, 2019; Wagenmaker & Jamieson, 2020; Tu et al., 2022; Ziemann & Tu, 2022; Ziemann et al., 2022). Since in reality, all data are finite, such a shift of focus is necessary to obtain a more complete understanding of learning for dynamics and control systems. Detailed related work and a tutorial on the subject can be found in (Tsiamis et al., 2023; Ziemann et al., 2023). With the exception of Wagenmaker & Jamieson (2020), most of the aforementioned works deal exclusively with identification in the time domain.

Frequency domain and time domain identification have many similarities–ignoring initial conditions, transients, or leakage effects, the two domains are equivalent (Schoukens et al., 2004). Still, working in one domain may offer some advantages over the other (Schoukens et al., 2004). For example, the frequency domain approach allows a unified treatment of discrete and continuous time systems, simplifies the analysis of systems with delays, and offers a more explicit way of designing the input excitation.

In this extended abstract, we first review the recent result of Tsiamis et al. (2024), where finite sample guarantees are derived for the well-established Empirical Transfer Function Estimate (ETFE) (Ljung, 1999) under open-loop periodic excitation. The result states that under certain stability conditions, the ETFE error decays with a rate of $\sqrt{M/N_{\mathrm{tot}}}$, where N_{tot} is the total number of samples. The parameter M is the number of selected frequencies at which we estimate the frequency response (resolution). Based on the ETFE, we can also recover the impulse responses of the system in the time domain by properly tuning the number of selected frequencies M. We obtain a slow rate of $\tilde{\mathcal{O}}(N_{\mathrm{tot}}^{-1/3})$ for general irrational systems and a faster rate of $\tilde{\mathcal{O}}(N_{\mathrm{tot}}^{-1/2})$ for state-space (rational) systems.

Notation. For any vector x, let $\|x\|$ denote the Euclidean norm. For any matrix S, let $\|S\|_{\mathrm{op}}$ denote the operator (spectral) norm. A universal constant is a constant that is independent of the problem at hand, e.g., the system or the algorithm. For any integer M, let $[M] \triangleq 0, \dots, M-1$.

2. Problem setting

Consider an *unknown*, linear, discrete-time, time-invariant system of the form

$$
\begin{aligned}
y_t &= \bar{y}_t + v_t, \\
\bar{y}_t &= G(\mathrm{q})u_t,\; G(\mathrm{q}) \triangleq \sum_{s=0}^{\infty} g_s \mathrm{q}^{-s},
\end{aligned} \tag{1}
$$

[1]Automatic Control Laboratory, ETH Zürich, Zürich, Switzerland. Correspondence to: Anastasios Tsiamis <atsiamis@ethz.ch>.

where $t \in \mathbb{Z}$ is the time, $u_t \in \mathbb{R}^{d_u}$ is the input, $y_t \in \mathbb{R}^{d_y}$ is the output, $\mathrm{q}^{-s} u_t = u_{t-s}$ is the backward shift operator, and $g_t \in \mathbb{R}^{d_y \times d_u}$, $t \geq 0$ is the impulse response. The noiseless output $\bar{y}_t$ is perturbed by some random noise process $v_t \in \mathbb{R}^{d_y}$.

Assumption 2.1 (Noise)**.** The noise process v_t is filtered sub-Gaussian white noise, that is,

$$v_t = H(\mathrm{q}) e_t, \quad H(\mathrm{q}) \triangleq \sum_{s=0}^{\infty} h_s \mathrm{q}^{-s}, \tag{2}$$

where $h_t \in \mathbb{R}^{d_y \times d_e}$ are the *unknown* filter coefficients. Let $e_t \in \mathbb{R}^{d_e}$ be i.i.d. zero mean, with covariance $\mathbb{E} e_t e_t^\top = \sigma_e^2 I_{d_e}$, and K^2-sub-Gaussian for some $K > 0$.

The noise model can capture general process noise, including measurement noise–see for example Section 4.1. We assume throughout that the input is bounded, which reflects practical physical constraints.

Assumption 2.2 (Input Bound)**.** All inputs are bounded

$$\|u_t\| \leq D_u, \text{ for all } t \in \mathbb{Z}$$

for some $D_u > 0$ independent of t.

To guarantee a well-defined estimation problem, we consider the following stability conditions.

Assumption 2.3 (Strict Stability)**.** The input-output impulse response is strictly stable (Ljung, 1999), that is,

$$\|G\|_\star \triangleq \sum_{t=0}^{\infty} t \|g_t\|_{\mathrm{op}} < \infty. \tag{3}$$

The auto-correlation function of the noise $R_t \triangleq \mathbb{E} v_s v_{s-t}^\top$ is also strictly stable

$$\|R\|_\star \triangleq \sum_{t=0}^{\infty} t \|R_t\|_{\mathrm{op}} < \infty. \tag{4}$$

Strict stability guarantees that the derivative of the frequency response $\partial G(e^{j\omega})/\partial \omega$ is uniformly bounded over all frequencies. This, in turn, implies that the response $G(e^{j\omega})$ is Lipschitz. Strict stability along with Assumption 2.2 guarantees that the transient phenomena have a limited effect on the estimation procedure.

We start all identification experiments at time $t = 0$. Hence, the initial conditions are determined by all past signals $u_{-1}, u_{-2}, \dots$ and $e_{-1}, e_{-2}, \dots$, which are nonzero in general, and unknown.

2.1. Empirical Transfer Function Estimate

The goal of frequency domain identification is to estimate the frequency response $G(e^{j\omega})$, given input-output data. We assume access to d_u experiments of length N, that is, data $(u_0^{(i)}, y_0^{(i)}, \dots, u_{N-1}^{(i)}, y_{N-1}^{(i)})$, for $i = 1, \dots, d_u$. This brings the total number of samples to $N_{\mathrm{tot}} \triangleq d_u N$. We assume that the trajectories are statistically independent.

Given any signal $z = \{z_t\}_{t \in [N]}$, let

$$Z_k \triangleq \mathcal{F}_k^N(z) \triangleq \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} z_t e^{-j \frac{2\pi k}{N} t}, \; k \in [N]$$

denote its N-point Discrete Fourier Transform (DFT), evaluated at k. Let $Y_k^{(i)}, U_k^{(i)}$ be the N-point DFTs of $y_t^{(i)}$ and $u_t^{(i)}$ respectively for the $i-$th experiment, $i = 1, \dots, d_u$. Let $Y_k \in \mathbb{C}^{d_y \times d_u}, U_k \in \mathbb{C}^{d_u \times d_u}$ denote the stacked DFTs for all experiments

$$Y_k \triangleq \begin{bmatrix} Y_k^{(1)} & \cdots & Y_k^{(d_u)} \end{bmatrix}, U_k \triangleq \begin{bmatrix} U_k^{(1)} & \cdots & U_k^{(d_u)} \end{bmatrix}. \tag{5}$$

Then, an estimate of $G(e^{j\omega})$ at frequency $\omega_k = 2\pi k/N$, for $k = 0, \dots, N-1$, can be obtained using the *ETFE*

$$\hat{G}_k \triangleq Y_k U_k^{-1}, \tag{6}$$

provided that U_k is invertible; the estimate is undefined if not. Since the number of frequencies N scales with the number of data, it is generally impossible to estimate the responses at all frequencies consistently (without assuming structure) (Ljung, 1999). Instead, we can learn the responses at a smaller frequency set. Given a frequency-resolution parameter $M < N$, we focus on estimating $G(e^{j\omega})$ at $\omega = 2\pi\ell/M$, for $\ell \in [M]$.

2.2. Excitation Method

Since we only need to estimate the frequency responses at $2\pi\ell/M$, $\ell \in [M]$, it is sufficient to excite the system at only these frequencies (Ljung, 1999). Assuming that M divides N, it is sufficient for the DFT U_k of the input to be non-zero at only $2\pi k/N = 2\pi\ell/M$ or $k = \ell N/M$. The latter condition is satisfied if and only if the excitation input is periodic with a period equal to M. We also need invertibility of U_k at $k = \ell N/M$. We assume the following.

Assumption 2.4 (Excitation)**.** Let the input signals be periodic with period M such that $u_{t+M}^{(i)} = u_t^{(i)}$, for $t \geq 0$ and every experiment $i = 1, \dots, d_u$. Assume that M divides N with $N_p \triangleq N/M$. Consider *one period* of the input signals and let the respective M-point DFTs be

$$\tilde{U}_\ell^{(i)} = \mathcal{F}_\ell^M(u^{(i)}) = \frac{1}{\sqrt{M}} \sum_{t=0}^{M-1} u_t^{(i)} e^{-\frac{2\pi\ell}{M} t}, i = 1, \dots, d_u$$

for $\ell \in [M]$, with respective stacked DFTs

$$\tilde{U}_\ell \triangleq \begin{bmatrix} \tilde{U}_\ell^{(1)} & \cdots & \tilde{U}_\ell^{(d_u)} \end{bmatrix}.$$

Assume that for all $\ell \in [M]$ the stacked DFTs satisfy

$$\sigma_{u,\ell}^2 I_{d_u} \preceq \tilde{U}_\ell \tilde{U}_\ell^*, \tag{7}$$

for some $0 < \sigma_{u,\ell}^2$ such that $\sum_{\ell=0}^{M-1} \sigma_{u,\ell}^2 \leq M D_u^2$, where D_u is the input upper bound of Assumption 2.2.

Such assumptions are standard when dealing with experiment design in the frequency domain. For example, Assumption 2.4 is satisfied by design (with uniform $\sigma_{u,\ell}^2$ across $\ell \in [M] - \{0\}$) when pseudorandom binary sequence (PRBS) signals are used and we excite one input at a time (Ljung, 1999, Ch. 13).

3. Finite-sample guarantees for the ETFE

In this section, we review the main result of (Tsiamis et al., 2024). Following the convention of (5), we define the stacked DFTs of the noises and the noiseless outputs as

$$V_k \triangleq \begin{bmatrix} V_k^{(1)} & \cdots & V_k^{(d_u)} \end{bmatrix}, \bar{Y}_k \triangleq \begin{bmatrix} \bar{Y}_k^{(1)} & \cdots & \bar{Y}_k^{(d_u)} \end{bmatrix}$$

Then, for every frequency $\omega_k = 2\pi k/N$ we have

$$Y_k = \bar{Y}_k + V_k = G(e^{j\omega_k})U_k + T_{k,N} + V_k, \tag{8}$$

where $T_{k,N} = \bar{Y}_k - G(e^{j\omega_k})U_k$ accounts for transient and time-aliasing phenomena since the DFT of $\{\bar{y}_t^{(i)}\}_{t=0}^{N-1}$ is different from $\{G(e^{j\omega_k})U_k^{(i)}\}_{k=0}^{N-1}$ for finite N. This transiet error term persists even in the absence of stochastic noise but vanishes as N grows to infinity.

The estimation error is equal to

$$\hat{G}_k - G(e^{j\omega_k}) = T_{k,N} U_k^{-1} + V_k U_k^{-1}, \tag{9}$$

where the input matrix U_k is invertible, and we only look at the frequencies $k = \ell N_p$, $\ell \in [M]$. Let $\Phi_{v,N}(k) \triangleq \mathbb{E} V_k^{(i)} (V_k^{(i)})^*$ be the aliased power spectrum of the process v_t at frequency k, where due to independence, the experiment index i does not affect the definition. Define the signal-to-noise ratio (SNR) at $k = \ell N_p$, $\ell \in [M]$ as

$$\mathrm{SNR}_{k,N} \triangleq \frac{\sigma_{u,\ell}}{\sqrt{\|\Phi_{v,N}(k)\|_{\mathrm{op}}}}, \tag{10}$$

where $\|\Phi_{v,N}(k)\|_{\mathrm{op}}$ is interpreted as the matrix norm for fixed k. We obtain the following finite-sample guarantees.

Theorem 3.1 (ETFE)**.** *Consider the ETFE* (6) *and fix a failure probability $\delta > 0$. Under Assumptions 2.1-2.4, with probability at least $1 - \delta$ for all $k = \ell N_p$, $\ell \in [M]$*

$$\|G(e^{j\omega_k}) - \hat{G}_k\|_{\mathrm{op}} \leq \frac{2\|G\|_\star D_u \sqrt{M}}{\sigma_{u,\ell} N} \tag{11}$$
$$+ \frac{\sqrt{M}}{\sqrt{N}} \mathrm{SNR}_{k,N}^{-1} \left(\sqrt{d_y} + c \frac{K^2}{\sigma_e^2} \sqrt{d_u + \log M/\delta} \right)$$

where c is a universal constant, $\|G\|_\star$ is defined in (3)*, and D_u is the maximum input norm.*

The first term of the right-hand side captures the transient error $T_{k,N} U_k^{-1}$ and is bounded using strict stability, input boundedness, and the properties of DFT. The second one captures the error $V_k U_k^{-1}$ due to stochastic noise and is bounded based on the Hanson-Wright inequality (Hanson & Wright, 1971; Vershynin, 2018). Assumption 2.4 guarantees that the input matrix U_k is invertible. The full proof can be found in (Tsiamis et al., 2024).

Recall that the total number of samples is equal to $N_{\mathrm{tot}} = d_u N$. As we increase the number of samples N_{tot} while keeping M constant, the former term decays at a faster rate of $1/N_{\mathrm{tot}}$ compared to the latter's $1/\sqrt{N_{\mathrm{tot}}}$. Hence, the non-asymptotic rate is of the order of

$$\mathcal{O}\left(\frac{\sqrt{M}}{\sqrt{N_{\mathrm{tot}}}} \sqrt{d_u} \left(\sqrt{d_u + \log M/\delta} + \sqrt{d_y} \right) \right).$$

The rate is similar to the ones for non-asymptotic parametric identification in time-domain (Ziemann et al., 2023); the optimal rate in that line of work is typically of the order $\sqrt{d}$ for some d scaling with the number of unknown parameters. Here, we have a similar scaling of $\sqrt{M}(\sqrt{d_u} + \sqrt{d_y})$ times an additional $\sqrt{d_u}$ dimensional dependence. This is an artifact of imposing a strict input norm bound in Assumption 2.2. If we allow σ_u, D_u to scale with $\sqrt{d_u}$ (as is the case for white-noise inputs in the time-domain (Ziemann et al., 2023)), we can remove this extra term.

A benefit of frequency-domain identification is that it provides specialized guarantees for every frequency of interest by breaking down the SNR into SNRs for every frequency. This offers direct insights on which frequencies to focus on and how to design the excitation inputs.

4. Application to time domain identification

We can recover the impulse responses or Markov parameters $g_t, t \in [M]$ of the system, using the frequency domain estimates $\hat{G}_{\ell N_p}, \ell \in [M]$ (McKelvey & Akçay, 1994). It is sufficient to take the inverse DFT (non-normalized version)

$$\hat{g}_t \triangleq \frac{1}{M} \sum_{\ell=0}^{M-1} \hat{G}_{\ell N_p} e^{j \frac{2\pi\ell}{M} t}, \; t \in [M]. \tag{12}$$

Note that using a finite frequency grid inevitably introduces time-aliasing. Under perfect knowledge of the frequency responses, the inverse DFT gives us

$$\tilde{g}_t \triangleq \frac{1}{M} \sum_{\ell=0}^{M-1} G(e^{j \frac{2\pi\ell}{M} t}) e^{j \frac{2\pi k}{N} t}, \; t \in [M], \tag{13}$$

where the responses $\tilde{g}_t$ are the *time-aliased* responses. Using standard properties of DFT, we can rewrite

$$\tilde{g}_t = g_t + \sum_{\tau=1}^{\infty} g_{t+\tau M}, \; t \in [M], \tag{14}$$

which justifies the term time-aliased

We can use the tail sum of the impulse response norms

$$\mathcal{C}^g_M \triangleq \sum_{t=M}^{\infty} \|g_t\|_{\mathrm{op}} \tag{15}$$

to control the time-aliasing error. Invoking the triangle inequality, we immediately obtain the following result.

Theorem 4.1 (Discrete-Time Impulse Response)**.** *Recall the notation of Theorem 3.1. Fix a failure probability $\delta > 0$. Under Assumptions 2.1-2.4, with probability at least $1-\delta$ uniformly for all $t \in [M]$*

$$\|g_t - \hat{g}_t\|_{\mathrm{op}} \le \frac{2\|G\|_\star D_u \sqrt{M}}{\sigma_u N} + \mathcal{C}^g_M + \frac{\sqrt{M}}{\sqrt{N}}\mathrm{SNR}_N^{-1}\left(\sqrt{d_y} + c\frac{K^2}{\sigma_e^2}\sqrt{d_u + \log M/\delta}\right) \tag{16}$$

where c is a universal constant and

$$\mathrm{SNR}_N^{-1} \triangleq \frac{1}{M}\sum_{\ell\in[M]} \mathrm{SNR}^{-1}_{\ell N_p, N},\ \sigma_u^{-1} \triangleq \frac{1}{M}\sum_{\ell\in[M]} \sigma^{-1}_{u,\ell}$$

are the average SNR and excitations among the frequencies of interest.

The rate is similar to the one of Theorem 3.1 with two notable differences. First, we have the additional time-aliasing error, which decays with the tail sum $\mathcal{C}^g_M$. Second, since the inverse DFT averages the frequency responses over all frequencies, the bound depends on the average-case excitation.

Under Assumption 2.3, the time-aliasing error approaches zero with a rate at least M^{-1} as the number of frequencies M grows. In particular, we have

$$\|g_t - \tilde{g}_t\|_{\mathrm{op}} \le \mathcal{C}^g_M \le \|G\|_\star M^{-1}.$$

In this case, we can balance the aliasing term M^{-1} and the dominant estimation term $M^{1/2}N^{-1/2}$ by selecting $M = \Theta(N^{-1/3})$. This gives us a rate of the order of $\tilde{\mathcal{O}}(N^{-1/3})$. This slow rate agrees with the result of (Goldenshluger, 1998) for strictly stable systems and is common in non-prametric settings (Tsybakov, 2008).

4.1. State-space models

In certain settings, where the decay rate of the impulse responses is faster, the above rate can be improved. For example, a special case is estimating Finite Impulse Response (FIR) models. In this case, we have $C^g_M = 0$ if we picked $M > n$, where n is the order of the FIR. More generally, we can obtain faster rates for stable state-space (rational) models (Chen et al., 1993), Consider a linear system in state-space form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + H_1 e_t \\ y_t &= Cx_t + Du_t + H_2 e_t, \end{aligned} \tag{17}$$

where $A \in \mathbb{R}^{d_x\times d_x}, B \in \mathbb{R}^{d_x\times d_u}, C \in \mathbb{R}^{d_y\times d_u}, D \in \mathbb{R}^{d_u\times d_u}$ are the state parameters with d_x the state dimension, while $H_1 \in \mathbb{R}^{d_x\times d_e}, H_2 \in \mathbb{R}^{d_y\times d_e}$ determine the noise second-order statistics. Hence, we have

$$g_0 = D,\quad g_t = CA^{t-1}B,\, t\ge 1,$$

with frequency response

$$G(e^{j\omega}) = D + C(e^{j\omega}I - A)^{-1}B.$$

The definition of the noise impulse and frequency responses is similar.

In the case of state-space systems, Assumption 2.3 is equivalent to exponential stability. Namely, the matrix A has all eigenvalues inside the unit circle or

$$\|A^t\|_{\mathrm{op}} \le c_A\rho^t, t\ge 0 \tag{18}$$

for some $c_A > 0$, $0 < \rho < 1$. As a result, the time-aliasing error decays much faster, i.e., exponentially fast with M. In this case, we can balance ρ^M and $M^{1/2}N^{-1/2}$ by selecting $M = \Theta(\log N)$. This gives us a faster rate of $\tilde{\mathcal{O}}(N^{-1/2})$, which reflects typical rates for parametric problems (Ziemann et al., 2023; Oymak & Ozay, 2021).

Corollary 4.2 (State-Space)**.** *Consider system (17). Let Assumptions 2.1-2.4 hold. Select $M = \beta\log N$ with $\beta \ge (-\log\rho)^{-1}$, where ρ is defined in (18). Fix a failure probability $\delta > 0$. Then, with probability at least $1-\delta$ uniformly for all $t\in[M]$*

$$\|g_t - \hat{g}_t\|_{\mathrm{op}} \le \frac{2\|G\|_\star D_u\sqrt{M}}{\sigma_u N} + \frac{c_A}{N(1-\rho)} + \frac{\sqrt{M}}{\sqrt{N}}\mathrm{SNR}_N^{-1}\left(\sqrt{d_y} + c\frac{K^2}{\sigma_e^2}\sqrt{d_u+\log M/\delta}\right) \tag{19}$$

The condition on β is sufficient to guarantee that the time-aliasing term decays faster than $N^{-1/2}$. The result of Corollary 4.2 could be combined with a realization procedure, e.g., the Ho-Kalman one (Oymak & Ozay, 2021), to obtain guarantees for recovering the state space parameters.

5. Future work

An interesting direction is studying (9) from the perspective of non-parametric least squares (Wainwright, 2019; Ziemann et al., 2022). Another interesting direction is studying continuous time frequency domain identification.

A. Supplementary material

For completeness, we include the definition of sub-Gaussian random vectors.

Definition A.1. A random vector $e \in \mathbb{R}^{d_e}$ is K^2-sub-Gaussian if and only if for any $\xi \in \mathbb{R}^{d_e}$

$$\mathbb{E}\exp\left(\xi^\top e_t\right) \le \exp\left(\frac{K^2\|\xi\|^2}{2}\right). \tag{20}$$

Proof of Theorem 4.1

By the triangle inequality

$$\|\hat{g}_t - g_t\|_{\text{op}} \le \|\hat{g}_t - \tilde{g}_t\|_{\text{op}} + \|\tilde{g}_t - g_t\|_{\text{op}}.$$

For the second term, by (14), we have

$$\|\tilde{g}_t - g_t\|_{\text{op}} \le \sum_{\tau=1}^{\infty} \|g_{t+\tau M}\|_{\text{op}} \le C_M^g.$$

Note that (14) follows from the properties of DFT– see (McKelvey & Akçay, 1994).

For the first term, we have,

$$\begin{aligned}\|\hat{g}_t - \tilde{g}_t\|_{\text{op}} &\le \frac{1}{M}\sum_{\ell=0}^{M-1} \|G(e^{j\frac{2\pi\ell}{M}t}) - \hat{G}_{N_p\ell}\|_{\text{op}}|e^{j\frac{2\pi k}{N}t}| \\ &\le \frac{1}{M}\sum_{\ell=0}^{M-1} \|G(e^{j\frac{2\pi\ell}{M}t}) - \hat{G}_{N_p\ell}\|_{\text{op}} \end{aligned} \tag{21}$$

The proof follows by invoking Theorem 3.1. ∎

Proof of Corollary 4.2

Under the condition

$$M \ge (-\log\rho)^{-1}\log N$$

it follows that

$$\rho^M \le N^{-1}.$$

This, in turn, implies

$$C_M^g \le c_A \frac{1}{N(1-\rho)}$$

The result follows immediately from Theorem 4.1.

References

Chen, C.-W., Juang, J.-N., and Lee, G. Frequency domain state-space system identification. In *1993 American Control Conference*, pp. 3057–3061. IEEE, 1993.

Faradonbeh, M. K. S., Tewari, A., and Michailidis, G. Finite Time Identification in Unstable Linear Systems. *Automatica*, 96:342–353, 2018.

Goldenshluger, A. Nonparametric estimation of transfer functions: rates of convergence and adaptation. *IEEE Transactions on Information Theory*, 44(2):644–658, 1998.

Hanson, D. L. and Wright, F. T. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3): 1079–1083, 1971.

Ljung, L. *System Identification: Theory for the User*. Prentice Hall, 1999.

McKelvey, T. and Akçay, H. An efficient frequency domain state-space identification algorithm. In *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, volume 4, pp. 3359–3364. IEEE, 1994.

Oymak, S. and Ozay, N. Revisiting Ho-Kalman based system identification: robustness and finite-sample analysis. *IEEE Transactions on Automatic Control*, 2021.

Pintelon, R. and Schoukens, J. *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

Sarkar, T., Rakhlin, A., and Dahleh, M. A. Finite time LTI system identification. *Journal of Machine Learning Research*, 22(26):1–61, 2021.

Schoukens, J., Pintelon, R., and Rolain, Y. Time domain identification, frequency domain identification. Equivalencies! Differences? In *Proceedings of the 2004 American Control Conference*, volume 1, pp. 661–666. IEEE, 2004.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification. *arXiv preprint arXiv:1802.08334*, 2018.

Tsiamis, A. and Pappas, G. J. Finite Sample Analysis of Stochastic System Identification. In *IEEE 58th Conference on Decision and Control (CDC)*, 2019.

Tsiamis, A., Ziemann, I., Matni, N., and Pappas, G. J. Statistical learning theory for control: A finite-sample perspective. *IEEE Control Systems Magazine*, 43(6):67–97, 2023.

Tsiamis, A., Abdalmoaty, M., Smith, R. S., and Lygeros, J. Finite sample frequency domain identification. *arXiv preprint arXiv:2404.01100*, 2024.

Tsybakov, A. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008.

Tu, S., Frostig, R., and Soltanolkotabi, M. Learning from many trajectories. *arXiv preprint arXiv:2203.17193*, 2022.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Wagenmaker, A. and Jamieson, K. Active learning for identification of linear dynamical systems. In *Conference on Learning Theory*, pp. 3487–3582. PMLR, 2020.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Ziemann, I. and Tu, S. Learning with little mixing. *Advances in Neural Information Processing Systems*, 35: 4626–4637, 2022.

Ziemann, I., Tsiamis, A., Lee, B., Jedra, Y., Matni, N., and Pappas, G. J. A Tutorial on the Non-Asymptotic Theory of System Identification. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 8921–8939. IEEE, 2023.

Ziemann, I. M., Sandberg, H., and Matni, N. Single trajectory nonparametric learning of nonlinear dynamics. In *conference on Learning Theory*, pp. 3333–3364. PMLR, 2022.