Frequency Domain Gaussian Process Models for H^∞ Uncertainties

Alex Devonport ALEX_DEVONPORT@BERKELEY.EDU
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Peter Seiler PSEILER@UMICH.EDU
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor

Murat Arcak ARCAK@BERKELEY.EDU
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Editors: N. Matni, M. Morari, G. J. Pappas

Abstract

Complex-valued Gaussian processes are used in Bayesian frequency-domain system identification as prior models for regression. If each realization of such a process were an H_∞ function with probability one, then the same model could be used for probabilistic robust control, allowing for robustly safe learning. We investigate sufficient conditions for a general complex-domain Gaussian process to have this property. For the special case of processes whose Hermitian covariance is stationary, we provide an explicit parameterization of the covariance structure in terms of a summable sequence of nonnegative numbers.

Keywords: Gaussian processes; system identification.

1. Introduction

With the general popularity of Gaussian process models in machine learning and in particular their growing adoption in data-driven control, there have been recent advances in using Gaussian process models as nonparametric Bayesian estimators in system identification. Initially this was done in the time domain, with works like Pillonetto and De Nicolao (2010) and Chen et al. (2012) using Gaussian processes to identify the impulse response of a linear time-invariant (LTI) stable system. Subsequent works consider frequency-domain regression, such as Lataire and Chen (2016) which uses a modified complex Gaussian process regression model to estimate transfer functions from discrete Fourier transform (DFT) data, and Stoddard et al. (2019) which considers a similar regression approach to estimate the generalized frequency response of nonlinear systems. These methods also have close ties to some non-probabilistic estimation methods, such as analytic interpolation (Singh and Sznaier (2020); Takyar and Georgiou (2010)) and kernel-based interpolation (Khosravi and Smith (2021, 2019)). At the heart of these Bayesian techniques is the *prior model*, a probabilistic dynamical model of an uncertain system that represents one's knowledge of the system prior to collecting any data.

Probabilistic dynamical models for uncertain systems are also used extensively in probabilistic robust control, such as probabilistic μ analysis (Khatri and Parrilo (1998); Balas et al. (2012); Biannic et al. (2021)), disk margins (Somers et al. (2022)), and the methods reviewed in Calafiore and

Dabbene (2007). In probabilistic robust control, each possible realization of the probabilistic uncertainty must be interpretable as a system of the type being modeled; otherwise, robustness guarantees involving ensembles of uncertainties would not be meaningful. This is a strong interpretability requirement compared to Bayesian system identification, where typically only the regression mean needs to be interpretable.

Since both Bayesian system identification and probabilistic robust control use probabilistic uncertainty models, applying both techniques to the same model is a promising strategy for safely learning an unknown or uncertain control system, since the Bayesian uncertainty of the learned model could be used to construct a probabilistic robustness guarantee for a suitably chosen controller. However, this is only possible if the nonparametric uncertainty model used for system identification satisfies the stronger interpretability requirement of probabilistic robust control. The contribution of this paper is to provide conditions under which a Gaussian process model satisfies this stronger requirement. These conditions are expressed in terms of the covariance of the process; as such, This represents a development in the frequency domain of similar efforts to establish almost-sure stability of time-domain Gaussian process models, such as Proposition 5.4 in Pillonetto and De Nicolao (2010) and the notion of stable kernels in Pillonetto et al. (2022). Specifically, we provide conditions under which realizations of a complex Gaussian process of a complex variable correspond to the z-transform of an LTI, causal, BIBO stable, and real system with probability one. Since an LTI, causal, and BIBO stable system is characterized by a z-transform that resides in the Hardy space H_∞, we refer to such processes as H_∞ Gaussian processes. Having conditions expressed in terms of the frequency-domain covariance functions allows one to design frequency-domain covariance functions directly, as opposed to the approach used by prior works in Bayesian system identification, where frequency-domain covariances must be derived from the z-transform (or Laplace transform) of a time-domain stochastic impulse response. In cases where prior knowledge is given in frequency-domain terms, being able to construct the frequency-domain covariance is more practical.

In addition to the general conditions, we provide a complete characterization (Theorem 10) of the covariance structure of a special class of H_∞ Gaussian process, namely those whose Hermitian covariance is stationary. Each Hermitian stationary H_∞ process is parameterized by a summable sequence of nonnegative reals, which lead to computationally tractable closed forms for certain choices of sequences. Since stationary processes are a popular choice for GP regression priors, this characterization makes it possible to construct useful and computationally convenient priors for Bayesian system identification that are also fully interpretable as probabilistic dynamical models.

To verify the utility of H_∞ GP models for Bayesian transfer function estimation, we apply the technique to two second-order systems using a mixture of a Hermitian stationary H_∞ processes constructed with Theorem 10 and an H_∞ process designed to model resonance peaks. Contrary to other recent work in Bayesian system identification, we choose to use the *strictly linear estimator* for our Gaussian process models instead of the *widely linear estimator*. Although the widely linear estimate is superior for general processes, we find that for H_∞ Gaussian process models the strictly linear estimator works nearly as well while being simpler and more stable to compute than the widely linear estimator.

The rest of the paper is organized as follows. Section 3 introduces the system setup, reviews background information on complex-valued random variables and stochastic processes, and introduces the classes of complex Gaussian processes that we study in this paper. Section 4 provides the conditions and characterizations described in the last paragraph, and represents the main technical contribution of this work. Section 5 reviews widely linear and strictly linear complex estimators for complex Gaussian process regression and presents numerical examples of Bayesian system identification. We omit the full proofs of of Theorem 5, Proposition 8, and Theorem 10 in this paper in favor of "proof sketches" for brevity. The full proofs are in an extended paper (Devonport et al. (2022)) available online.

2. Notation

For a complex vector or matrix X, X^* denotes the complex conjugate and X^H denotes the conjugate transpose. We denote the exterior of the unit disk as $E = \{Z \in \mathbb{C} : |z| > 1\}$, and its closure as $\bar{E} = \{Z \in \mathbb{C} : |z| \geq 1\}$. L_2 is the Hilbert space of functions $f : \mathbb{C} \to \mathbb{C}$ such that $\int_{-\pi}^{\pi} |f(Re^{j\Omega})|^2 d\Omega < \infty$, equipped with the inner product $\langle f, g\rangle_2 = \int_{-\pi}^{\pi} f(e^{j\Omega})g^*(e^{j\Omega})d\Omega$. H_2 is the Hilbert space of functions $f : \mathbb{C} \to \mathbb{C}$ that are bounded and analytic for all $z \in E$ and $\int_{-\pi}^{\pi} |f(Re^{j\Omega})|^2 d\Omega < \infty$, equipped with the inner product $\langle f, g\rangle_2 = \int_{-\pi}^{\pi} f(e^{j\Omega})g^*(e^{j\Omega})d\Omega$. It is a vector subspace of L_2. H_∞ is the Banach space of functions $f : \bar{E} \to \mathbb{C}$ that are bounded and analytic for all $z \in E$ and $\sup_{\Omega\in[-\pi,\pi]} |f(e^{j\Omega})| < \infty$, equipped with the norm $\|f\|_\infty = \sup_{\Omega\in[-\pi,\pi]} |f(e^{j\Omega})|$. ℓ^1 is the space of absolutely summable sequences, that is sequences $\{a_n\}_{n=0}^{\infty}$ such that $\sum_{n=0}^{\infty} |a_n| < \infty$. $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance Σ; likewise, $\mathcal{CN}\left(\mu, \Sigma, \tilde{\Sigma}\right)$ denotes a complex Gaussian distribution with mean μ, Hermitian covariance Σ, and complementary covariance $\tilde{\Sigma}$.

3. Preliminaries

The object of this paper is to construct nonparametric statistical models for causal, LTI, BIBO stable systems in the frequency domain. Since our main focus will be the probabilistic aspects of the model, we restrict our attention to the simplest dynamical case: a single-input single-output system in discrete time. Thus, our dynamical systems are frequency-domain multiplier operators $H_f : L_2 \to L_2$ whose output is defined pointwise as $(H_f u)(\omega) = f(\omega)u(\omega)$, where $f : \mathbb{C} \to \mathbb{C}$ is the system's transfer function. Thanks to the bijection $H_f \leftrightarrow f$, we generally mean the function f when we refer to "the system".

Since our aim is to construct a probabilistic model for the system that is not restricted to a finite number of parameters, we must work directly with random complex functions of a complex variable: this is a special type of complex stochastic process that we call a z-domain process.

Definition 1 *Let $(\Xi, F, \mathbb{P})$ denote a probability space. A z-domain stochastic process with domain $D \subseteq \mathbb{C}$ is a function $f : \Xi \times D \to \mathbb{C}$.*

Note that each value of $\xi \in \Xi$ yields a function $f_\xi = f(\xi, \cdot) : D \to \mathbb{C}$, which is called either a "realization" or a "sample path" of f. If we take ξ to be selected at random according to the probability law $\mathbb{P}$, then f_ξ represents a "random function" in the frequentist sense. Alternatively, if we have a prior belief about the likelihood of some f_ξ over others, we may encode this belief in a Bayesian sense using the measure $\mathbb{P}$. We drop the dependence of f on ξ from the notation outside of definitions, as it will be clear when $f(z)$ refers to the random variable $f(\cdot, z)$ or when f stands for a realization f_ξ.

Definition 2 *A Gaussian z-domain process is a z-domain process f such that, for any $n > 0$, the random vector $(f(z_1), \dots, f(z_n))$ is complex multivariate Gaussian for all $(z_1, \dots, z_n) \in D^n$.*

A complex Gaussian process is more than two real-valued Gaussian processes added together, as the real and imaginary parts may depend on each other. Unlike a real Gaussian process, which is characterized by its mean $m(t) = \mathbb{E}[x(t)]$ and covariance $k(t,s) = \mathbb{E}[x(t)x(s)]$, a complex Gaussian process f is characterized by three functions: its mean $m(z) = \mathbb{E}[f(z)]$, its *Hermitian covariance* $k(z,w) = \mathbb{E}[f(z)f^*(w)]$ and its *complementary covariance* $\tilde{k}(z,w) = \mathbb{E}[f(z)f(w)]$.

3.1. H_∞ Gaussian Processes

Consider a deterministic input-output operator H_g with transfer function function $g : D \to \mathbb{C}$. The condition that H_g belong to the operator space H^∞ of LTI, causal, and BIBO stable systems is that g belong to the function space H_∞. Now suppose we wish to construct a random operator H_f using the realizations of a z-domain process f as its transfer function: the analogous condition is that the realizations of f lie in H_∞ with probability one.

Definition 3 *A z-domain process is called an H_∞ process when the set $\{\xi \in \Xi : f_\xi \in H_\infty\}$ has measure one under $\mathbb{P}$.*

Less formally, an H_∞ process is a z-domain process f such that $\mathbb{P}(f \in H_\infty) = 1$. Having $f_\xi \in H_\infty$ implies that $\bar{E} \subseteq D$: we usually take $D = \bar{E}$. If we also require that H_g give real outputs to real inputs in the time domain, g must satisfy the conjugate symmetry relation $g(z^*) = g^*(z)$ for all $z \in D$. The analogous condition for H_f is to require that f satisfy the condition with probability one.

Definition 4 *A z-domain process f is called conjugate symmetric when the set $\{\xi \in \Xi : f_\xi(z^*) = f_\xi^*(z),\ \forall z \in D\}$ has measure one under $\mathbb{P}$.*

Combining definitions 2, 3, and 4, we arrive at our main object of study: conjugate-symmetric H_∞ Gaussian processes.

Example 1 ("Cozine" process) *The random transfer function*

$$f(z) = \frac{X - a(X\cos(\omega_0) - Y\sin(\omega_0))z^{-1}}{1 - 2a\cos(\omega_0)z^{-1} + a^2 z^{-2}}, \tag{1}$$

where $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $a \in (0,1)$, $\omega_0 \in [0,\pi]$, is a z-domain Gaussian process. From the form of the transfer function, we see that H is bounded on the unit circle, analytic on E, and conjugate symmetric with probability one, from which it follows that f is a conjugate symmetric H_∞ process. Since f corresponds to the z-transform of an exponentially decaying discrete cosine with random magnitude and phase, we call it a "cozine" process. *Its Hermitian and complementary covariances are*

$$\begin{aligned} k(z,w) &= \frac{1 - a\cos(\omega_0)(z^{-1} + (w^*)^{-1}) + a^2(zw^*)^{-1}}{(1 - 2a\cos(\omega_0)z^{-1} + a^2z^{-2})(1 - 2a\cos(\omega_0)(w^*)^{-1} + a^2(w^*)^{-2})}, \\ \tilde{k}(z,w) &= \frac{1 - a\cos(\omega_0)(z^{-1} + w^{-1}) + a^2(zw)^{-1}}{(1 - 2a\cos(\omega_0)z^{-1} + a^2z^{-2})(1 - 2a\cos(\omega_0)w^{-1} + a^2w^{-2})}. \end{aligned} \tag{2}$$

As a Bayesian prior for an H^∞ system, this process represents a belief that the transfer function exhibits a resonance peak (of unknown magnitude) at ω_0. Knowing ω_0 in advance is a strong belief, but it can be relaxed by taking a hierarchical model where ω_0 enters as a hyperparameter. When used as a prior, the hierarchical model represents the less determinate belief that there is a resonance peak *somewhere*, whose magnitude can be made arbitrarily small if no peak is evident in the data.

The construction in Example 1, where properties of conjugate symmetry and BIBO stability can be checked directly, may be extended to random transfer functions of any finite order. However, the technique does not carry to the infinite-order H_∞ processes required for nonparametric Bayesian system identification, or more generally for applications that do not place an *a priori* restriction on the order of the system. We are therefore motivated to find conditions under which a z-domain process is a conjugate-symmetric H_∞ Gaussian process expressed directly in terms of k and $\tilde{k}$.

4. Constructing H_∞ Gaussian Processes

The following result provides the general test to determine if f is an H_∞ Gaussian process, by establishing with probability one that $f_\xi \in H_2$ and $\|f_\xi\|_\infty < \infty$.

Theorem 5 *Let f be a z-domain Gaussian process with mean zero and continuous Hermitian covariance k and complementary covariance $\tilde{k}$. Let $k_r = \frac{1}{2}\operatorname{Re}[k + \tilde{k}]$, $k_i = \frac{1}{2}\operatorname{Re}[k - \tilde{k}]$ denote the covariance functions of the real and imaginary parts of f respectively. Then f is an H_∞ process under the following conditions:*

1. *There exist positive, finite constants C_r, C_i, α_r, α_i, δ_r, δ_i, such that k_r and k_i, restricted to the unit circle, satisfy the following continuity conditions:*

$$\begin{aligned} k_r(e^{j\theta}, e^{j\theta}) + k_r(e^{j\phi}, e^{j\phi}) - 2k_r(e^{j\theta}, e^{j\phi}) &\le \frac{C_r}{|\log|\theta - \phi||^{1+\alpha_r}} \quad \forall |\theta - \phi| < \delta_r \\ k_i(e^{j\theta}, e^{j\theta}) + k_i(e^{j\phi}, e^{j\phi}) - 2k_i(e^{j\theta}, e^{j\phi}) &\le \frac{C_i}{|\log|\theta - \phi||^{1+\alpha_i}} \quad \forall |\theta - \phi| < \delta_i. \end{aligned} \tag{3}$$

2. *Let $\{z_n\}_{n=1}^\infty$ be a countable dense sequence of points in E. For $n \in \mathbb{N}$, define the Gramian matrices $K_r^n, K_i^n, R^n \in \mathbb{R}^{n\times n}$ as $(K_r^n)_{jl} = k_r(z_j, z_l)$, $(K_i^n)_{jl} = k_i(z_j, z_l)$, and $(R^n)_{jl} = r(z_j, z_l)$, where $r(z_j, z_l) = z_i z_j^*/(z_i z_j^* - 1)$. K_r^n, K_i^n, and R^n satisfy*

$$\sup_{n\in\mathbb{N}} \operatorname{trace} K_r^n (R^n)^{-1} < \infty \quad \textit{and} \quad \sup_{n\in\mathbb{N}} \operatorname{trace} K_i^n (R^n)^{-1} < \infty. \tag{4}$$

Proof sketch Discrete-time H_2 is an RKHS with kernel $r(z, w) = zw^*/(zw^* - 1)$. (This fact is proven in Devonport et al. (2022).) Condition (4) then ensures by Driscoll's zero-one theorem (Driscoll (1973)) that the sample paths of f inhabit H_2 with probability one. Condition (3) ensures by (Adler and Taylor, 2007a, Theorem 1.4.1) that the restriction of f to the unit circle is bounded with probability one. Since an H_∞ function is precisely an H_2 function whose values on the unit circle are bounded, it follows that f inhabits H_∞ with probability one. ■

Remark 6 *According to Driscoll's theorem, the probability that $f \in H_2$ is either zero or one. (Zero occurs when either supremum in condition (4) is infinite.) Similarly, the realizations of a Gaussian process are bounded with probability zero or one (Landau and Shepp (1970)). This means that the realizations of a z-domain Gaussian process are either almost surely H_∞ functions or almost surely not.*

Remark 7 *Condition (4) is necessary and sufficient for f to inhabit H_2 with probability one. On the other hand, condition (3) is sufficient but not necessary for $\|f\|_\infty$ to be bounded. Indeed, necessary and sufficient conditions for a stochastic process to be almost surely bounded are generally not available even for real-valued Gaussian processes except in special cases. Fortunately, covariance functions in practice often satisfy a stronger condition that implies (3) ((Adler and Taylor, 2007b, eq. 2.5.17)), namely that $k(s,t) = k(s,s) - q(s-t) + O(|s-t|^{2+\delta})$, for small $|s-t|$, where q is a positive definite quadratic form and $\delta > 0$.*

The general condition for a process to be conjugate symmetric is given by the following result.

Proposition 8 *Let f be a z-domain Gaussian process with domain D, covariance k, and complementary covariance $\tilde{k}$. Then f is conjugate-symmetric if and only if k and $\tilde{k}$ satisfy the conditions*

$$k(z,z) = k(z^*,z^*), \qquad k(z,z) = \tilde{k}(z,z^*) \tag{5}$$

for all $z \in D$.

Proof sketch Under (5), the joint distribution for $(f^*(z), f(z^*))$ is a degenerate complex Gaussian distribution where both components are perfectly correlated with the same variance, and thus equal with probability one. If (5) doesn't hold, this cannot be true for all z. ■

Together, Theorem 5 and Proposition 8 give sufficient conditions on the covariance functions of a general mean-zero z-domain Gaussian process in order for it to be a conjugate-symmetric H_∞ Gaussian process. While Conditions (3) and (5) can be verified in practice, Condition (4) generally cannot. We are therefore motivated to find special cases of z-domain Gaussian processes for which (4) can be replaced by a more tractable condition. The broadest such case that we have found is where, in addition to satisfying Conditions (3) and (5), the Hermitian covariance function is stationary when restricted to the unit circle.

Definition 9 *A z-domain Gaussian process is* Hermitian stationary *when its Hermitian covariance function satisfies $k(e^{j\theta}, e^{j\phi}) = k(e^{j(\theta-\phi)}, 1)$ for all $\theta, \phi \in [-\pi, \pi)$.*

Using a stationary process as a prior is common practice in machine learning and control-theoretic applications of Gaussian process models. Stationary processes are useful for constructing regression priors that do not introduce unintended biases in their belief about the frequency response: since $f(e^{j\theta})$ has the same Hermitian variance across the entire unit circle, a sample path from a Hermitian stationary H_∞ process is just as likely to exhibit low-pass behavior as it is high-pass or band-pass.[1] We can obtain a "partially informative" prior by adding an H_∞ process encoding strong beliefs in one frequency range (such as the presence of a resonance peak) to an H_∞ process encoding weaker beliefs across all frequencies. The sum, also an H_∞ process, encodes a combination of these beliefs.

Under the additional condition of Hermitian stationarity, it turns out that the H_∞ process is characterized by a sequence of nonnegative constants.

Theorem 10 *Let f be a Hermitian stationary, conjugate-symmetric z-domain Gaussian process with continuous Hermitian covariance k and complementary covariance $\tilde{k}$. Then f is an H_∞ process if and only if k and $\tilde{k}$ have the form*

$$k(z,w) = \sum_{n=0}^{\infty} a_n^2 (zw^*)^{-n}, \qquad \tilde{k}(z,w) = \sum_{n=0}^{\infty} a_n^2 (zw)^{-n}, \tag{6}$$

1. To be truly "noninformative" in the sense of introducing unwanted biases, the complementary covariance should be stationary. However, this is not possible while satisfying (5).

where $\{a_n\}_{n=0}^{\infty}$ *is a nonnegative real* ℓ^1 *sequence. Furthermore,* f *may be expanded as*

$$f(z) = \sum_{n=0}^{\infty} a_n w_n z^{-n}, \tag{7}$$

where $w_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

Proof sketch That (7) leads to (6) follows from direct calculation and the independence of the w_n. Under the summability condition on the a_n, the impulse response of f is absolutely summable with probability one, implying BIBO stability. Thus (7) is BIBO stable (and hence in H_∞) with probability one.

To see that a Hermitian stationary, conjugate-symmetric H_∞ Gaussian process must have (6), (7) and that the summability condition holds, start with an expansion of f into the basis $\{z^{-n}\}_{n=0}^{\infty}$ of H_∞. Hermitian stationarity implies that the coefficients are uncorrelated, turning the basis function expansion into (7), from which (6) follows. Since the process is H_∞, its impulse response must be absolutely summable with probability one, which is only true if the summability condition on the a_n holds. ■

Theorem 10 provides a useful tool for constructing conjugate-symmetric H_∞ Gaussian processes: all we need to do is select a summable sequence of nonnegative numbers. For example, we use Theorem 10 to construct the following regression prior for the next section.

Example 2 (Geometric H_∞ process) *Take* $a_n^2 = \alpha^n$ *with* $\alpha \in (0,1)$*; this yields a conjugate-symmetric* H_∞ *Gaussian process with Hermitian covariance* $k_\alpha(z,w) = \sum_{n=0}^{\infty} \alpha^n (zw^*)^{-n} = \frac{zw^*}{zw^*-\alpha}$ *and complementary covariance* $\tilde{k}_\alpha(z,w) = \frac{zw}{zw-\alpha}$.

5. Gaussian Process Regression in the Frequency Domain

Let $H_g \in H^\infty$ denote the system whose transfer function $g \in H_\infty$ we wish to identify. While not stochastic, g is unknown, and we represent both our uncertainty and our prior beliefs in a Bayesian fashion with an H_∞ Gaussian process f with Hermitian and complementary covariances k and $\tilde{k}$. To model our prior beliefs, the distribution of f should give greater probability to functions we believe are likely to be similar to g, and should assign probability zero to functions we know that g cannot be. As an example of the latter, the fact that $P(f \in H_\infty) = 1$ encodes our belief that $g \in H_\infty$, which demonstrates the importance of H_∞ Gaussian processes for prior model design.

We suppose that our data consists of n noisy frequency-domain point estimates $y_i = g(z_i) + e_i$, where $e_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_e^2\right)$, $z_i \in \bar{E}$. If our primary form of data is a time-domain trace of input and output values, we first convert this data into an *empirical transfer function estimate* (ETFE). There are several well-established methods to construct ETFEs from time traces, such as Blackman-Tukey spectral analysis, windowed filter banks, or simply dividing the DFT of the output trace by the DFT of the input trace. In our numerical examples, we will use windowed filter banks.

Our approach is essentially the same procedure as standard Gaussian process regression as described in Rasmussen and Williams (2006) extended to the complex case. We take the mean of the prior model to be zero without loss of generality. To estimate the transfer function at a new point z, we note that $g(z)$ is related to $(y_1, \ldots, y_n)$ under the prior model as

$$\begin{bmatrix} g(z) \\ y \end{bmatrix} \sim \mathcal{CN}\left(0, \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy}^H & K_{yy} \end{bmatrix}, \begin{bmatrix} \tilde{K}_{xx} & \tilde{K}_{xy} \\ \tilde{K}_{xy}^H & \tilde{K}_{yy} \end{bmatrix}\right); \tag{8}$$

where $y \in \mathbb{C}^n$, $K_{yy} \in \mathbb{C}^{n\times n}$, $K_{xy} \in \mathbb{C}^{n\times 1}$, and $K_{xx} \in \mathbb{C}$ are defined componentwise as

$$(y)_i = y_i, \quad (K_{yy})_{ij} = k(z_i, z_j) + \sigma_e^2 \delta_{ij}, \quad (K_{xy})_{ij} = k(z, z_i), \quad K_{xx} = k(z, z) + \sigma_e^2, \tag{9}$$

and the components of the complementary covariance matrix are defined analogously. The minimum-error linear estimator $\hat{g}(z)$ for $g(z)$ given the data and its predictive Hermitian variance σ_g^2 are (Schreier and Scharf, 2010, §5.3)

$$\hat{g}(z) = K_{xy}^H K_{yy}^{-1} y, \qquad \sigma_g^2(z) = k(z, z) - K_{xy}^H K_{yy}^{-1} k_{xy}. \tag{10}$$

These expressions are identical to the posterior mean and variance of a real Gaussian process regression model (cf. Equation (2.19) in Rasmussen and Williams (2006)) except that K_{xx}, K_{xy}, and K_{yy} are complex-valued.

For general complex-valued Gaussian regression, the *widely linear* estimator, which incorporates y^* and the complementary covariance, is an improvement over the strictly linear estimator. The degree of improvement is measured by the matrix $P = K_{yy} - \tilde{K}_{yy}(K_{yy}^*)^{-1}\tilde{K}_{yy}^*$, which is the error variance of linearly estimating y^* from y under the prior model. In particular, when $P = 0$ the strictly and widely linear estimators coincide (Schreier and Scharf, 2010, §5.4.1). In our experiments, we find that the strictly linear estimator performs well for conjugate-symmetric H_∞ GP priors, and that P is nearly singular and very small in norm compared to K_{yy} and $\tilde{K}_{yy}$, implying that its performance is close to the widely linear estimator.[2] For these reasons, we use the strictly linear estimator in the regression examples below. We believe the strictly linear estimator works well for conjugate-symmetric H_∞ process because of the prior assumption of causality and conjugate symmetry. We discuss this in more detail in Devonport et al. (2022).

For $z \in D$ and $\eta > 0$, define the *confidence ellipsoid* $\mathcal{E}_\eta(z) = \{w \in \mathbb{C} : |w - \hat{g}(z)|^2 \leq \eta^2 \sigma_g^2(z)\}$. By Markov's inequality, we know that $f(z) \in \mathcal{E}_\eta(z)$ with probability $\geq 1 - 1/\eta^2$. This implies bounds on the real and imaginary parts by projecting the confidence ellipsoid onto the real and imaginary axes: from these we can construct probabilistic bounds on the magnitude and phase of $f(z)$ via interval arithmetic, which we will see in the numerical examples.

Let $\theta \in \Theta$ denote the hyperparameters of a covariance function k_θ, so that K_{yy} becomes a function of θ: then the log marginal likelihood of the data under the posterior for the strictly linear estimator is $L(\theta) = -\frac{1}{2}\left(y^H K_{yy}(\theta)^{-1} y + \log\det K_{yy}(\theta) + n \log 2\pi\right)$. Keeping the data y and input locations z_i fixed, $L(\theta)$ measures the probability of observing data y when the prior covariance function is k_θ. By maximizing L with respect to θ, we find the covariance among k_θ, $\theta \in \Theta$ that best explains the observations.[3]

5.1. Examples: Identifying Second-order Systems

We apply the strictly-linear H_∞ Gaussian process regression method described above to the problem of identifying two second-order systems. The first test system is a second-order system that exhibits a resonance peak. The system is specified in continuous time, with canonical second-order transfer function $g(s) = \frac{\omega_0^2}{s^2 + 2\xi\omega_0 s + \omega_0^2}$, where $\omega_0 = 20\pi$ rad/s, and $\xi = 0.1$, and converted to the discrete-time transfer function $g(z)$ using a zero-order hold discretization with a sampling frequency of

2. The widely linear regression equations, which we show in Devonport et al. (2022), require inverting P. In this case, the widely linear estimator is numerically unstable compared to the strictly linear estimator.

3. Although it seems contradictory to choose prior parameters based on posterior data, it can be justified as an empirical-Bayes approximation to a hierarchical model with θ as hyperparameter.

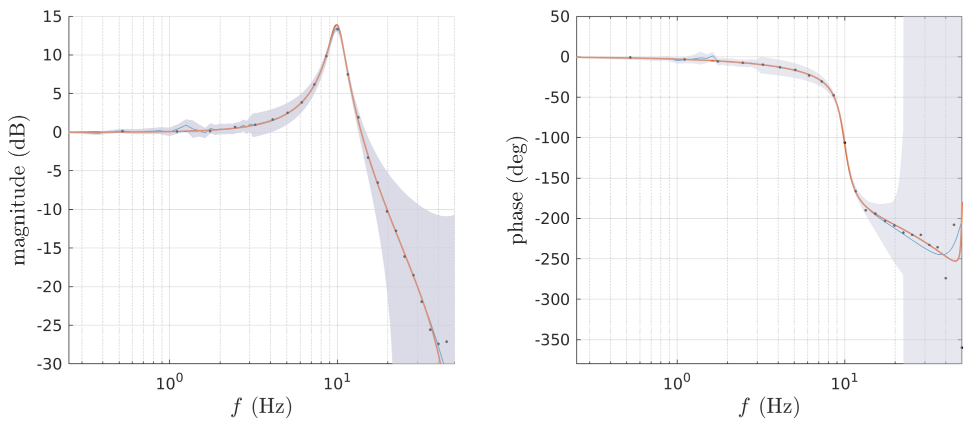

Figure 1: Bode plot of the second-order resonant system (orange), and its estimate (blue) using H_∞ Gaussian process regression from an empirical transfer function estimate (black points) with $\eta = 3$ confidence ellipsoid bounds (grey).

$f_s = 100$ Hz. We suppose that we know *a priori* that there is a resonance peak, but not about its location or half-width, and we have no other strong information about the frequency response. For this prior belief, an appropriate prior model is a weighted mixture of a cozine process and a Hermitian stationary process. In particular, we use the family of H_∞ processes with covariance functions $k(z,w) = \sigma_g^2 k_g(z,w) + \sigma_c^2 k_c(z,w), \tilde{k}(z,w) = \sigma_g^2 \tilde{k}_g(z,w) + \sigma_c^2 \tilde{k}_c(z,w)$, where k_g is the covariance of the geometric H_∞ process defined in Example 2, and k_c is the covariance of the cozine process, and likewise for the complementary covariance. σ_g^2 and σ_c^2 are weights that determine the relative importance of the two parts of the model. This family of covariances has five hyperparameters: $k_g \in [0,\infty)$, $\alpha \in (0,1)$, $k_c \in [0,\infty)$, $\omega_0 \in [0,\pi]$, and $a \in (0,1)$.

We suppose that an input trace $u(n)$ of Gaussian white noise with variance $\sigma_u^2 = 1/f_s$ is run through H_g yielding an output trace $y(n)$; our observations comprise these two traces, corrupted by additive Gaussian white noise of variance $\sigma^2 = 10^{-4}/f_s$. To obtain an empirical transfer function estimate, we run both observation traces through a bank of 25 windowed 1000-tap DFT filters. The impulse responses of the filter bank are $h_i(n) = e^{j\omega_i n} w(n)$ for $i = 1,\ldots,25$, with Gaussian window $w(n) = \exp(-\frac{1}{2}(\sigma_w(n-500)/1000)^2)$ for $n = 0,\ldots,999$, and $w(n) = 0$ otherwise, with window half-width $\sigma_w = 0.25$. Let u_i, y_i denote the outputs of filter h_i with inputs u, y respectively: $y_i(n)/u_i(n)$ gives a running estimate of $g(e^{j\omega_i})$, whose value after 1000 time steps we take as our observation at $z_i = e^{j\omega_i}$. Figure 1 shows the regression from the strictly linear estimator (10) after tuning the covariance hyperparameters via maximum likelihood, along with predictive error bounds based on $\eta = 3$ confidence ellipsoids.

The second is a second-order allpass filter. This system is specified in discrete time with the transfer function $g(z) = \frac{|z_0|^2 - 2\,\mathrm{Re}[z_0] + 1}{1 - 2\,\mathrm{Re}[z_0] + |z_0|^2}$, where $z_0 = 0.5e^{\pm j\pi/4}$ are the system's poles, with sampling frequency $f_s = 100$ Hz. For this system we assume that we do not have *a priori* information on the structure of the frequency response, so we use a Hermitian stationary H_∞ process as the prior model. In particular, we take the family of geometric H_∞ process, indexed by hyperparameter $\alpha \in (0,1)$. To construct the empirical transfer function estimate, we use the same data model and filter bank as the previous example. Figure 2 shows the strictly linear regression after tuning the covariance hyperparameters, again with predictive error bounds from $\eta = 3$ confidence ellipsoids.

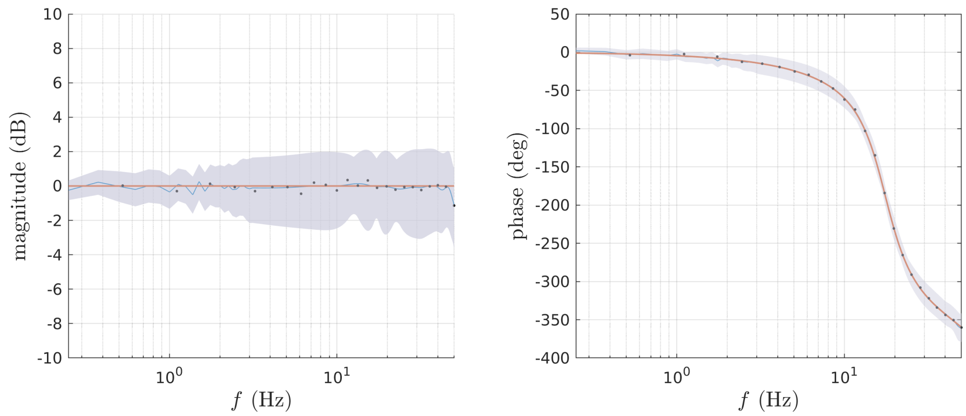

Figure 2: Bode plot of the second-order allpass system (orange), and its estimate (blue) using H_∞ Gaussian process regression from an empirical transfer function estimate (black points) with $\eta = 3$ confidence ellipsoid bounds (grey).

6. Conclusion

The H_∞ processes constructed using the results of this paper, particularly Theorem 10, are effective priors for Bayesian nonparametric identification of transfer functions. Furthermore, the strictly linear estimator, which is suboptimal for general complex Gaussian process priors, provides transfer function estimates that are close to optimal for conjugate-symmetric H_∞ priors. We have numerical evidence that suggests that as the number of frequency data points increases, the covariance becomes *maximally improper*, a case in which the strictly linear is indeed optimal. We will investigate this conjecture in future work.

The applications presented in this paper use H_∞ Gaussian process as statistically interpretable regression priors, but do not consider questions of probabilistic robustness. We intend to follow this work with a similar investigation into the robustness properties of H_∞ models, such as probabilistic bounds on the H_∞ norm, and integral quadratic constraints that hold with high probability for an H_∞ process with given mean and covariance functions.

Acknowledgments

This work was supported by the grants ONR N00014-18-1-2209, AFOSR FA9550-18-1-0253. NSF CNS-2111688.

References

Robert J Adler and Jonathan E Taylor. *Random fields and geometry*, volume 80. Springer, 2007a.

Robert J Adler and Jonathan E Taylor. *Applications of random fields and geometry: applications and case studies*. Unpublished manuscript, 2007b. URL https://robert.net.technion.ac.il/publications/.

Gary Balas, Peter Seiler, and Andrew Packard. Analysis of an UAV flight control system using probabilistic μ. In *AIAA Guidance, Navigation, and Control Conference*, page 4989, 2012.

Jean-Marc Biannic, Clément Roos, Samir Bennani, Fabrice Boquet, Valentin Preda, and Bénédicte Girouart. Advanced probabilistic μ-analysis techniques for AOCS validation. *European Journal of Control*, 62:120–129, 2021.

Giuseppe C Calafiore and Fabrizio Dabbene. Probabilistic robust control. In *2007 American Control Conference*, pages 147–158. IEEE, 2007.

Tianshi Chen, Henrik Ohlsson, and Lennart Ljung. On the estimation of transfer functions, regularizations and Gaussian processes—revisited. *Automatica*, 48(8):1525–1535, 2012.

Alex Devonport, Peter Seiler, and Murat Arcak. Frequency domain Gaussian process models for H^∞ uncertainties, 2022. URL https://arxiv.org/abs/2211.15923.

Michael F Driscoll. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 26(4):309–316, 1973.

Sven Khatri and Pablo A Parrilo. Guaranteed bounds for probabilistic μ. In *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No. 98CH36171)*, volume 3, pages 3349–3354. IEEE, 1998.

Mohammad Khosravi and Roy S Smith. Kernel-based identification of positive systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1740–1745. IEEE, 2019.

Mohammad Khosravi and Roy S Smith. Kernel-based identification with frequency domain side-information. *arXiv preprint arXiv:2111.00410*, 2021.

Henry J Landau and Lawrence A Shepp. On the supremum of a Gaussian process. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 369–378, 1970.

John Lataire and Tianshi Chen. Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, 72:217–229, 2016.

Gianluigi Pillonetto and Giuseppe De Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1):81–93, 2010.

Gianluigi Pillonetto, Tianshi Chen, Alessandro Chiuso, Giuseppe De Nicolao, and Lennart Ljung. *Regularized system identification: Learning dynamic models from data*. Springer Nature, 2022.

Carl Edward Rasmussen and Christopher K.I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.

Peter J Schreier and Louis L Scharf. *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.

Rajiv Singh and Mario Sznaier. A Loewner matrix based convex optimization approach to finding low rank mixed time/frequency domain interpolants. In *2020 American Control Conference (ACC)*, pages 5169–5174. IEEE, 2020.

Franca Somers, Sovanna Thai, Clément Roos, Jean-Marc Biannic, Samir Bennani, Valentin Preda, and Francesco Sanfedino. Probabilistic gain, phase and disk margins with application to AOCS validation. *IFAC-PapersOnLine*, 55(25):1–6, 2022.

Jeremy G Stoddard, Georgios Birpoutsoukis, Johan Schoukens, and James S Welsh. Gaussian process regression for the estimation of generalized frequency response functions. *Automatica*, 106:161–167, 2019.

Mir Shahrouz Takyar and Tryphon T Georgiou. Analytic interpolation with a degree constraint for matrix-valued functions. *IEEE transactions on automatic control*, 55(5):1075–1088, 2010.