Signal Denoising via Gaussian Processes

Louis Bouchard[1]

[1]*Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA*

(Dated: February 24, 2025)

We present a Gaussian Process (GP) framework for signal denoising that harnesses multiple noisy transients to recover the true signal with exceptional precision. Departing from traditional techniques like signal averaging and low-pass filtering, which hinge on oversimplified assumptions—such as zero-mean noise or fixed frequency bands—and overlook critical noise statistics, our method employs a custom hybrid kernel to model intricate signal patterns, including damped oscillations and multi-frequency components, while representing noise as discrete martingale increments. This algorithm surpasses conventional low-pass filtering across diverse signal types, such as damped oscillations and multi-frequency sinusoids, and noise profiles, from Gaussian to Pink, as evidenced by comprehensive simulations (e.g., SNR 6.92 dB vs. 4.46 dB for Gaussian noise). By fully utilizing the rich statistical insights from transients, our approach eliminates the sluggish convergence and spectral constraints of prior methods, delivering a clear, powerful alternative to the opaque complexity of neural network-based solutions.

INTRODUCTION

Signal denoising seeks to recover a deterministic signal $s : [0, T] \to \mathbb{R}$, assumed measurable and bounded on a compact interval $[0, T] \subset \mathbb{R}_+$, from noisy observations $x_i(t) = s(t) + n_i(t)$, where $n_i(t)$ is a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$, and $i = 1, \ldots, k$ indexes k independent repeated measurements, termed transients. Here, $n_i : [0, T] \times \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable function, meaning $n_i(t, \cdot)$ is a random variable in $L^2(\Omega, \mathcal{F}, P)$ for each t, ensuring that noise realizations are quantifiable within the probabilistic framework and evolve temporally under a consistent measure P. Traditional methods, such as signal averaging and low-pass filtering, have long dominated scientific practice, yet their reliance on restrictive assumptions—such as zero-mean noise or band-limited signal spectra—fundamentally constrains their robustness and adaptability to diverse noise structures and signal complexities.

Signal averaging estimates $s(t)$ via the sample mean $\bar{x}_k(t) = k^{-1} \sum_{i=1}^{k} x_i(t)$, relying on the premise that $n_i(t)$ is a zero-mean process, i.e., $\mathbb{E}[n_i(t)] = 0$ for all t, where $\mathbb{E}$ denotes expectation under P. Under this condition, the strong law of large numbers [3] guarantees $\bar{x}_k(t) \to s(t)$ almost surely as $k \to \infty$, provided $n_i(t) \in L^1(\Omega, \mathcal{F}, P)$ and the sequence is ergodic. However, practical denoising is invariably limited to a finite k, typically $k \leq 100$, introducing significant biases and inefficiencies. For finite k, the central limit theorem yields $\sqrt{k}(\bar{x}_k(t) - s(t)) \Rightarrow N(0, \sigma^2(t))$ with $\sigma^2(t) = \mathbb{E}[n_i(t)^2] < \infty$, implying a slow error reduction rate of $O(k^{-1/2})$—e.g., for $k = 100$, the standard error is $\sigma(t)/10$, insufficient for high-noise scenarios ($\sigma(t) \gg s(t)$). Moreover, if $n_i(t)$ exhibits skewness or heavy tails ($\mathbb{E}[|n_i(t)|^3] < \infty$, $\mathbb{E}[n_i(t)^3] \neq 0$), the sample mean $\bar{x}_k(t)$ is biased in finite k, with Edgeworth expansions [3] quantifying the skew-induced shift as approximately $\mathbb{E}[n_i(t)^3]/(k\sigma^3(t))$, persisting until k becomes impractically large. Critically, averaging collapses the rich ensemble $\{x_i(t)\}_{i=1}^{k}$ into a single statistic, discarding all higher moments (e.g., variance $\sigma^2(t)$, kurtosis) and cross-temporal correlations, which contain valuable information about $n_i(t)$'s structure—especially when $\mathbb{E}[n_i(t)] \neq 0$ or noise deviates from Gaussianity.

Low-pass filtering assumes $s(t)$ resides in a low-frequency subspace of $L^2([0, T])$, projecting $x_i(t)$ through a convolution kernel $h(t)$ with cutoff frequency f_c, yielding $s_{\mathrm{LP}}(t) = (x_i * h)(t)$. This presupposes $n_i(t)$ has a power spectral density $S_n(f)$ concentrated above f_c, a condition often violated by noise with significant low-frequency components (e.g., $1/f$ noise), leading to distortion of $s(t)$ and poor convergence unless f_c is unrealistically separated from the signal's spectrum. Wiener filtering [5] optimizes $h(t)$ using $S_n(f)$ and $S_s(f)$, but its reliance on precise spectral knowledge—rarely available experimentally—and disregard for transient multiplicity further limits its generality.

Modern neural network methods [4] offer flexibility but operate as black boxes, lacking interpretability and measure-theoretic convergence guarantees. To overcome these deficiencies, we propose a denoising approach using Gaussian Processes (GPs), a versatile stochastic tool distinct from assuming additive Gaussian noise. A GP is a process $s : [0, T] \times \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P)$, such that for any finite set

$\{t_1, \ldots, t_m\} \subset [0, T]$, the vector $(s(t_1), \ldots, s(t_m))$ is multivariate Gaussian in $L^2(\Omega, \mathcal{F}, P)$ [3]. Denoted $s(t) \sim \mathcal{GP}(\mu(t), k(t, t'))$, with $\mu(t) = \mathbb{E}_P[s(t)]$ and $k(t, t') = \mathbb{E}_P[(s(t) - \mu(t))(s(t') - \mu(t'))]$ a positive semi-definite, continuous covariance kernel, this framework—via Kolmogorov's extension theorem—ensures consistency across distributions.

Our method harnesses the full ensemble $\{x_i(t)\}_{i=1}^k$ to estimate $s(t)$, employing sample moments like $\hat{\sigma}^2(t_j) = (k-1)^{-1} \sum_{i=1}^k (x_i(t_j) - \bar{x}_k(t_j))^2$ to model noise variance, starkly contrasting with averaging's reduction to $\mathbb{E}_{P^k}[\bar{x}_k(t)] = s(t)$. The method, which integrates stochastic calculus and probability, achieves superior performance over low-pass filtering across diverse noise types (e.g., SNR 6.92 dB vs. 4.46 dB for Gaussian noise, $\sigma = 40$), as validated by simulations. By preserving and exploiting the complete statistical structure of transients, we surmount the finite-sample bias, slow convergence, and spectral rigidity of traditional methods, offering a transparent, adaptable alternative to neural networks.

MATHEMATICAL FRAMEWORK

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space equipped with a filtration $\{\mathcal{F}_t\}_{t \in [0,T]}$, where $[0, T] \subset \mathbb{R}_+$ is a compact time interval, and $\mathcal{F}_t \subset \mathcal{F}$ is right-continuous and complete with respect to P-null sets, satisfying the usual conditions of stochastic analysis [3]. We consider a deterministic signal $s : [0, T] \to \mathbb{R}$, assumed Borel-measurable and square-integrable, i.e., $s \in L^2([0, T], \mathcal{B}([0, T]), \lambda)$, where λ is the Lebesgue measure. The observed process is defined as:

$$x_i(t, \omega) = s(t) + n_i(t, \omega), \quad i = 1, \ldots, k, \quad t \in [0, T], \quad \omega \in \Omega, \tag{1}$$

where $n_i : [0, T] \times \Omega \to \mathbb{R}$ are independent and identically distributed (i.i.d.) stochastic processes, adapted to $\{\mathcal{F}_t\}$, and k denotes the number of transients. The sample space for each n_i is $\Omega_i = \mathbb{R}^{[0,T]}$, the space of real-valued functions on $[0, T]$, endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^{[0,T]})$ generated by the product topology. The joint sample space is $\Omega = (\mathbb{R}^{[0,T]})^k$, with product measure $P = P_1 \times \cdots \times P_k$, where P_i is the probability measure induced by n_i. We assume $n_i(t) \in L^2(\Omega_i, \mathcal{F}_i, P_i)$ for all t, ensuring finite second moments, and $n_i(t) \in L^1(\Omega_i, \mathcal{F}_i, P_i)$ for integrability, critical for expectation and variance estimates.

Noise Model

We model the noise $n_i(t)$ as a discrete-time process derived from martingale increments, reflecting physical noise sources with controlled statistical properties. Formally, consider a sequence of time points $T = \{t_j = j\Delta t : j = 0, 1, \ldots, T_s\}$, where $\Delta t = T/T_s$, $T_s = 1000$ in our implementation, and define $n_i(t_j)$ as increments of a martingale difference process $\{M_i(t_j)\}_{j=0}^{T_s}$, adapted to a discrete filtration $\{\mathcal{F}_{t_j}\}$. The martingale property requires:

$$\mathbb{E}[M_i(t_j) - M_i(t_{j-1}) | \mathcal{F}_{t_{j-1}}] = 0, \quad j = 1, \ldots, T_s, \tag{2}$$

where $M_i(t_0) = 0$, and we set $dM_i(t_j) = M_i(t_j) - M_i(t_{j-1})$. For Gaussian noise, we specify $dM_i(t_j) = \sigma \xi_{i,j}$, where $\xi_{i,j} \sim N(0, 1)$ are i.i.d. standard normal random variables, and $\sigma > 0$ is the standard deviation. Thus:

$$n_i(t_j) = \sigma \xi_{i,j}, \quad t_j = j\Delta t, \quad \xi_{i,j} \sim N(0, 1), \tag{3}$$

with $\mathbb{E}[n_i(t_j)] = 0$ and $\mathbb{E}[n_i(t_j)^2] = \sigma^2$. This discrete formulation avoids the continuous-time stochastic integral $n_i(t) = \int_0^t \sigma dW_i(s)$, where $W_i(s)$ is a standard Wiener process satisfying $\mathbb{E}[dW_i(s)^2] = ds$, as the integrated form yields $\mathbb{E}[n_i(t)^2] = \sigma^2 t$, growing linearly and misrepresenting stationary noise in denoising contexts.

The martingale increment assumption ensures $\{n_i(t_j)\}_{j=0}^{T_s}$ is a sequence of uncorrelated random variables with respect to $\{\mathcal{F}_{t_j}\}$, i.e., $\mathbb{E}[n_i(t_j) n_i(t_{j'})] = 0$ for $j \neq j'$, verifiable via the orthogonality of increments in $L^2(P_i)$. For finite k, we estimate the noise variance at each t_j using the sample variance:

$$\hat{\sigma}^2(t_j) = \frac{1}{k-1} \sum_{i=1}^k (x_i(t_j) - \bar{x}_k(t_j))^2, \tag{4}$$

where $\bar{x}_k(t_j) = \frac{1}{k} \sum_{i=1}^k x_i(t_j)$ is the sample mean. Under the i.i.d. assumption and $s(t)$ constant across transients, $\mathbb{E}[\bar{x}_k(t_j)] = s(t_j) + \mathbb{E}[n_i(t_j)] = s(t_j)$, and $\mathrm{Var}(\bar{x}_k(t_j)) = \sigma^2/k$. By Kolmogorov's strong law of large numbers [3], $\bar{x}_k(t_j) \to s(t_j)$ almost surely as $k \to \infty$, and $\hat{\sigma}^2(t_j) \to \sigma^2$ in probability, with $\mathbb{E}[\hat{\sigma}^2(t_j)] = \sigma^2$ and variance $\mathrm{Var}(\hat{\sigma}^2(t_j)) = \frac{2\sigma^4}{k-1}$ for Gaussian $n_i(t_j)$, assuming finite fourth moments ($\mathbb{E}[n_i(t_j)^4] = 3\sigma^4$).

To formalize convergence, consider the stochastic process $\bar{n}_k(t_j) = k^{-1} \sum_{i=1}^k n_i(t_j)$. For $n_i(t_j) \in L^2(\Omega_i, \mathcal{F}_i, P_i)$, the central limit theorem yields:

$$\sqrt{k}(\bar{x}_k(t_j) - s(t_j)) = \sqrt{k}\bar{n}_k(t_j) \Rightarrow N(0, \sigma^2), \tag{5}$$

in distribution as $k \to \infty$, where $\Rightarrow$ denotes weak convergence in $D([0,T])$, the Skorokhod space of càdlàg functions [3]. This ensures $\bar{x}_k(t_j)$ is a consistent estimator of $s(t_j)$, with precision scaling as $O(k^{-1/2})$, leveraged by our GP model to refine the estimate beyond mere averaging.

Gaussian Process Model

We model the signal $s(t)$ as a Gaussian Process (GP) on the interval $[0,T]$, defined as a stochastic process $s : [0,T] \times \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, P)$, such that for any finite set of time points $\{t_1, \ldots, t_m\} \subset [0,T]$, the random vector $(s(t_1), \ldots, s(t_m))$ follows a multivariate Gaussian distribution in $L^2(\Omega, \mathcal{F}, P)$. Formally, we write:

$$s(t) \sim \mathcal{GP}(\mu(t), k(t,t')), \tag{6}$$

where $\mu(t) = \mathbb{E}[s(t)]$ is the mean function, and $k(t,t') = \mathbb{E}[(s(t)-\mu(t))(s(t')-\mu(t'))]$ is the covariance kernel, assumed continuous and positive semidefinite on $[0,T] \times [0,T]$. This definition, rooted in Kolmogorov's extension theorem [3], ensures the existence of a consistent probability measure on the path space $\mathbb{R}^{[0,T]}$ with Borel σ-algebra $\mathcal{B}(\mathbb{R}^{[0,T]})$. We set $\mu(t) = 0$ a priori, centering the process after preprocessing the observed data to remove any deterministic bias, aligning with the assumption $\mathbb{E}[n_i(t)] = 0$.

For signals with damped oscillatory behavior, such as $s(t) = Ae^{-\alpha t}\sin(2\pi f t)$ (e.g., $A = 4$, $\alpha = 2$, $f = 5$ Hz in our simulations), we design a hybrid kernel to capture both periodicity and exponential decay. The kernel is:

$$k(t,t') = k_{\mathrm{per}}(t,t')k_{\mathrm{RBF}}(t,t'), \tag{7}$$

where:

$$k_{\mathrm{per}}(t,t') = \sigma_p^2 \exp\left(-2\sin^2\left(\frac{\pi f|t-t'|}{p}\right)/l_p^2\right), \tag{8}$$

$$k_{\mathrm{RBF}}(t,t') = \sigma_r^2 \exp\left(-\frac{|t-t'|^2}{2l_r^2}\right), \tag{9}$$

with $p = 1/f$ as the period, $l_p, l_r > 0$ as length scales, and $\sigma_p^2, \sigma_r^2 > 0$ as variance parameters. The periodic component k_{per} models the oscillatory frequency f, leveraging the periodicity of $\sin(2\pi f t)$, while k_{RBF} captures the smooth decay $e^{-\alpha t}$ via its radial basis function form, ensuring the covariance decays with temporal separation. This product kernel induces a reproducing kernel Hilbert space (RKHS) $H_k \subset L^2([0,T])$ [3], where $s(t)$ is a sample path with $\mathbb{E}[\|s\|_{H_k}^2] < \infty$, balancing local periodicity and global smoothness.

The observed process is the sample average of k transients:

$$\bar{x}_k(t) = s(t) + \epsilon_k(t), \quad \epsilon_k(t) = \frac{1}{k}\sum_{i=1}^{k} n_i(t), \tag{10}$$

where $n_i(t)$ are i.i.d. noise processes with $\mathbb{E}[n_i(t)] = 0$ and $\mathbb{E}[n_i(t)^2] = \sigma^2 < \infty$, as defined in the noise model. By the independence of $n_i(t)$ across i, the noise term $\epsilon_k(t)$ satisfies:

$$\mathbb{E}[\epsilon_k(t)] = 0, \quad \mathbb{E}[\epsilon_k(t)^2] = \frac{\sigma^2}{k}, \tag{11}$$

and for Gaussian $n_i(t)$, $\epsilon_k(t) \sim N(0, \sigma^2/k)$ pointwise. On the discretized grid $T = \{t_j = j\Delta t : j = 1, \ldots, T_s\}$, with $T_s = 1000$ and $\Delta t = T/T_s$, we observe the vector $\bar{x}_k = [\bar{x}_k(t_1), \ldots, \bar{x}_k(t_{T_s})]^T \in \mathbb{R}^{T_s}$, modeled as:

$$\begin{aligned} \bar{x}_k &= s + \epsilon_k, \quad s = [s(t_1), \ldots, s(t_{T_s})]^T, \\ \epsilon_k &\sim N\left(0, \frac{\sigma^2}{k} I_{T_s}\right), \end{aligned}$$

where I_{T_s} is the $T_s \times T_s$ identity matrix. The joint distribution is:

$$\begin{bmatrix} s \\ \bar{x}_k \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K & K \\ K & K + \frac{\sigma^2}{k} I_{T_s} \end{bmatrix}\right), \tag{12}$$

with $K = K(T,T) = [k(t_i,t_j)]_{i,j=1}^{T_s}$ the covariance matrix induced by $k(t,t')$.

The GP posterior distribution $s(t)|\bar{x}_k$ is derived via conditional Gaussianity on $(\Omega, \mathcal{F}, P)$. For a new time point $t \in [0,T]$, the posterior is:

$$s(t)|\bar{x}_k \sim \mathcal{GP}(\mu^*(t), k^*(t,t')), \tag{13}$$

where the posterior mean and covariance are:

$$\mu^*(t) = k(t,T)\left[K(T,T) + \frac{\sigma^2}{k} I_{T_s}\right]^{-1} \bar{x}_k, \tag{14}$$

$$k^*(t,t') = k(t,t') - k(t,T)\left[K(T,T) + \frac{\sigma^2}{k} I_{T_s}\right]^{-1} k(T,t'), \tag{15}$$

with $k(t,T) = [k(t,t_1), \ldots, k(t,t_{T_s})] \in \mathbb{R}^{1 \times T_s}$, and $K(T,T)$ as above. The posterior mean $\mu^*(t)$ is the minimum mean square error estimator in $L^2(P)$, i.e., $\mu^*(t) = \mathbb{E}[s(t)|\bar{x}_k]$, and $k^*(t,t) = \mathbb{E}[(s(t) - \mu^*(t))^2|\bar{x}_k]$ quantifies residual uncertainty. For $k \to \infty$, $\frac{\sigma^2}{k} \to 0$, and $\mu^*(t) \to \bar{x}_k(t)$ in probability if $s(t)$

is continuous, but the GP refines this estimate by incorporating $k(t,t')$, leveraging transient statistics beyond the sample mean.

The hybrid kernel ensures $k(t,t')$ is stationary in periodicity but non-stationary in amplitude decay, matching the signal's damped oscillatory nature. The RKHS norm $\|s\|^2_{H_k} = \int_0^T \int_0^T s(t)[\mathcal{K}^{-1}s](t')dtdt'$, where $\mathcal{K}$ is the integral operator with kernel k, regularizes $s(t)$ to balance smoothness and periodicity, distinguishing our approach from frequency-domain methods [5].

Probabilistic Foundations and Convergence

To establish the probabilistic foundation of our Gaussian Process (GP) denoising framework, we define the product probability space for k transients as $(\Omega^k, \mathcal{F}^k, P^k)$, where $\Omega^k = (\mathbb{R}^{[0,T]})^k$ is the k-fold Cartesian product of the path space $\mathbb{R}^{[0,T]}$, the space of real-valued functions on $[0,T]$. Here, $\mathcal{F}^k = \mathcal{B}(\mathbb{R}^{[0,T]})^{\otimes k}$ is the product σ-algebra generated by the Borel σ-algebra $\mathcal{B}(\mathbb{R}^{[0,T]})$ under the product topology, ensuring measurability of cylindrical sets. The measure $P^k = P_1 \times \cdots \times P_k$ is the product measure, where each P_i is the probability measure on $(\mathbb{R}^{[0,T]}, \mathcal{B}(\mathbb{R}^{[0,T]}))$ induced by the stochastic process $n_i(t)$, assumed identical across $i = 1, \ldots, k$. The observation vector $x = (x_1, \ldots, x_k) \in \Omega^k$ represents k independent transients, with $x_i(t,\omega) = s(t) + n_i(t,\omega)$, where $s(t)$ is deterministic and $n_i(t)$ is a measurable noise process adapted to a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$.

The sample mean process is defined as:

$$\bar{x}_k(t,\omega) = \frac{1}{k}\sum_{i=1}^{k} x_i(t,\omega), \quad t \in [0,T], \quad \omega \in \Omega^k, \tag{16}$$

a random variable on $(\Omega^k, \mathcal{F}^k, P^k)$ taking values in $\mathbb{R}^{[0,T]}$. For each fixed t, $\bar{x}_k(t)$ is $\mathcal{F}^k$-measurable, and we analyze its convergence to the true signal $s(t)$. Assuming $n_i(t) \in L^2(\Omega_i, \mathcal{F}_i, P_i)$ with $\mathbb{E}[n_i(t)] = m(t)$ and $\mathbb{E}[n_i(t)^2] < \infty$, the expectation is:

$$\mathbb{E}_{P^k}[\bar{x}_k(t)] = \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}_{P_i}[x_i(t)] = s(t) + m(t), \tag{17}$$

by the i.i.d. property and Fubini's theorem [3]. The variance is:

$$\mathrm{Var}_{P^k}(\bar{x}_k(t)) = \frac{1}{k^2}\sum_{i=1}^{k} \mathrm{Var}_{P_i}(n_i(t)) = \frac{\sigma^2(t)}{k}, \tag{18}$$

where $\sigma^2(t) = \mathbb{E}_{P_i}[n_i(t)^2] - m(t)^2$, assuming finite second moments. Under the martingale increment model $n_i(t_j) = \sigma\xi_{i,j}$ (with $\xi_{i,j} \sim N(0,1)$) and $m(t) = 0$, Shiryaev's strong law of large numbers for i.i.d. sequences [3] ensures:

$$\bar{x}_k(t) \to s(t) \quad P^k\text{-almost surely as } k \to \infty, \tag{19}$$

for each $t \in [0,T]$, provided $n_i(t) \in L^1(\Omega_i, \mathcal{F}_i, P_i)$. For uniform convergence over $[0,T]$, we require $n_i(t)$ to be a separable process in $L^2(P_i)$, with a countable dense set $D \subset [0,T]$ such that:

$$\sup_{t\in[0,T]} |\bar{x}_k(t) - s(t)| \to 0 \quad P^k\text{-a.s.}, \tag{20}$$

achievable if $n_i(t)$ has continuous paths or is càdlàg with bounded variation, as in our discrete implementation ($T_s = 1000$).

The sample variance estimate at $t_j = j\Delta t$ is:

$$\hat{\sigma}^2(t_j) = \frac{1}{k-1}\sum_{i=1}^{k}(x_i(t_j) - \bar{x}_k(t_j))^2, \tag{21}$$

an unbiased estimator of $\sigma^2(t_j)$. For Gaussian $n_i(t_j)$, $(k-1)\hat{\sigma}^2(t_j)/\sigma^2(t_j) \sim \chi^2_{k-1}$, and:

$$\mathbb{E}_{P^k}[\hat{\sigma}^2(t_j)] = \sigma^2(t_j), \quad \mathrm{Var}_{P^k}(\hat{\sigma}^2(t_j)) = \frac{2\sigma^4(t_j)}{k-1}, \tag{22}$$

with $\hat{\sigma}^2(t_j) \to \sigma^2(t_j)$ almost surely as $k \to \infty$ by the strong law, assuming $\mathbb{E}[n_i(t_j)^4] = 3\sigma^4(t_j) < \infty$. This convergence in $L^2(P^k)$ is uniform over $t_j \in T$ if $\sigma^2(t)$ is continuous and bounded, as verified by Glivenko-Cantelli-type arguments for empirical processes [3].

The GP posterior $s(t)|\bar{x}_k$ is a conditional process on $(\Omega^k, \mathcal{F}^k, P^k)$, defined via the Radon-Nikodym derivative of the joint measure $P(s, \bar{x}_k)$ with respect to $P(s)$. For $\bar{x}_k = [\bar{x}_k(t_1), \ldots, \bar{x}_k(t_{T_s})]^T$, the posterior mean:

$$\mu^*(t) = \mathbb{E}_{P^k}[s(t)|\bar{x}_k] = k(t,T)\left[K(T,T) + \frac{\sigma^2}{k} I_{T_s}\right]^{-1} \bar{x}_k, \tag{23}$$

is the $L^2(P^k)$-optimal predictor, minimizing $\mathbb{E}_{P^k}[(s(t) - \hat{s}(t))^2|\bar{x}_k]$, where $k(t,T) = [k(t,t_j)]_{j=1}^{T_s}$, and $K(T,T) = [k(t_i,t_j)]_{i,j=1}^{T_s}$. The posterior covariance:

$$k^*(t,t') = \mathbb{E}_{P^k}[(s(t) - \mu^*(t))(s(t') - \mu^*(t'))|\bar{x}_k], \tag{24}$$

ensures $s(t)|\bar{x}_k \sim \mathcal{GP}(\mu^*(t), k^*(t,t'))$ retains Gaussianity. Under the martingale assumption $\mathbb{E}[n_i(t_j)] = 0$, $\mu^*(t)$ is consistent with $s(t)$ as $k \to \infty$, since $\frac{\sigma^2}{k} \to 0$, and $\bar{x}_k(t) \to s(t)$ in $L^2(P^k)$, with $k^*(t,t) \to 0$ reflecting vanishing uncertainty. This conditional expectation leverages the full transient ensemble, surpassing the pointwise convergence of traditional averaging.

PYTHON IMPLEMENTATION

I have written some code in Python (see attached) to operationalize the measure-theoretic Gaussian Process (GP) denoising framework, utilizing the `scikit-learn` library's `GaussianProcessRegressor` for computational efficiency and robustness. The algorithm processes k transients to estimate the deterministic signal $s(t)$ from noisy observations $x_i(t) = s(t) + n_i(t)$, with $i = 1, \ldots, k$ and $t \in [0, T]$, leveraging the full statistical structure of the data. Below, I outline the key steps, grounding each in the preceding mathematical formalism and detailing their practical realization.

1. **Data Generation:** We simulate $k = 100$ transients over a discrete grid $T = \{t_j = j\Delta t : j = 1, \ldots, T_s\}$, with $T_s = 1000$, $\Delta t = T/T_s$, and $T = 1$ second, reflecting a sampling frequency $f_s = 1/\Delta t = 1000$ Hz. The signal $s(t)$ is generated deterministically, e.g., for damped oscillatory cases as $s(t_j) = Ae^{-\alpha t_j} \sin(2\pi f t_j)$, with $A = 4$, $\alpha = 2$, and $f = 5$ Hz. Noise is modeled as discrete martingale increments $n_i(t_j) = \sigma\xi_{i,j}$, where $\xi_{i,j} \sim N(0, 1)$ are i.i.d., and $\sigma = 40$ is the standard deviation, implemented via:

```
noise = np.random.normal(0, sigma, size=(k, T_s))
x = s + noise
```

   Here, `np.random.normal` generates realizations from $N(0, \sigma^2)$, ensuring $\mathbb{E}[n_i(t_j)] = 0$ and $\mathbb{E}[n_i(t_j)^2] = \sigma^2$, consistent with the noise model. The resulting $x_i(t_j)$ matrix in $\mathbb{R}^{k\times T_s}$ has sample variance $\hat{\sigma}^2(t_j) \approx \sigma^2$, with $\mathrm{Var}(\bar{x}_k(t_j)) = \sigma^2/k = 16$ for $k = 100$, aligning with the theoretical $\epsilon_k(t) \sim N(0, \sigma^2/k)$.

2. **Preprocessing:** We compute the sample mean:

$$\bar{x}_k(t_j) = \frac{1}{k}\sum_{i=1}^{k} x_i(t_j), \tag{25}$$

   via `np.mean(x, axis=0)`, yielding $\bar{x}_k \in \mathbb{R}^{T_s}$. To mitigate numerical instability and scale mismatch in high-variance noise ($\sigma^2 = 1600$), we normalize $\bar{x}_k$ using `StandardScaler`:

```
scaler = StandardScaler()
x_bar_scaled
  = scaler.fit_transform(
  x_bar.reshape(-1, 1)).flatten()
```

   This transforms $\bar{x}_k(t_j)$ to $\bar{x}_{k,\text{scaled}}(t_j) = (\bar{x}_k(t_j) - \hat{\mu}_k)/\hat{\sigma}_k$, where $\hat{\mu}_k = T_s^{-1}\sum_{j=1}^{T_s}\bar{x}_k(t_j)$ and $\hat{\sigma}_k^2 = T_s^{-1}\sum_{j=1}^{T_s}(\bar{x}_k(t_j) - \hat{\mu}_k)^2$, ensuring zero empirical mean and unit variance. This preserves the relative noise structure $\hat{\sigma}^2(t_j)/k$, aligning with $\epsilon_k(t)$'s variance, and facilitates GP optimization by standardizing the input scale.

3. **GP Fitting:** We fit the GP using a hybrid kernel tailored to damped oscillatory signals:

```
kernel
  = ExpSineSquared(length_scale=1/f,
    periodicity=1/f,
    length_scale_bounds=(1e-2, 1e2)) *
    RBF(length_scale=1/f,
    length_scale_bounds=(1e-2, 1e2)) +
    WhiteKernel(noise_level=sigma / np.sqrt(k),
    noise_level_bounds=(1e-5, 1e1))
```

   Here, $f = 5$ Hz sets $p = 1/f = 0.2$ s, and initial length scales $l_p = l_r = 1/f$ balance periodicity and decay, with bounds $[10^{-2}, 10^2]$ ensuring optimization flexibility. The `WhiteKernel` models $\epsilon_k(t)$ with variance $\sigma^2/k = 16$, adjustable within $[10^{-5}, 10]$. The GP is fitted via:

```
gp =
GaussianProcessRegressor(kernel=kernel,
n_restarts_optimizer=50)
gp.fit(t.reshape(-1, 1), x_bar_scaled)
s_GP_scaled = gp.predict(t.reshape(-1, 1))
s_GP = scaler.inverse_transform(
s_GP_scaled.reshape(-1, 1)).flatten()
```

   This computes the posterior mean $\mu^*(t) = k(t,T)[K(T,T)+(\sigma^2/k)I]^{-1}\bar{x}_{k,\text{scaled}}$, with $T = \{t_j\}$, $k(t,T) = [k(t,t_j)]$, and $K(T,T) = [k(t_i,t_j)]$, optimized over 50 restarts to ensure convergence in $L^2(P^k)$. Inverse scaling recovers $s_{\text{GP}}(t)$ in the original domain, approximating $s(t)$.

4. **Low-Pass Comparison:** For benchmarking, we apply a 4th-order Butterworth filter at cutoff frequency $f_c = 10$ Hz:

```
nyquist = fs / 2
Wn = fc / nyquist
b, a = signal.butter(4, Wn, btype='low')
s_LP = signal.filtfilt(b, a, x_bar)
```

This implements $s_{\mathrm{LP}}(t) = (h * \bar{x}_k)(t)$, where $h(t)$ is the filter impulse response, approximating a projection onto $L^2([0,T])$ functions with frequency support below f_c. The two-pass `filtfilt` ensures zero-phase distortion, but lacks transient multiplicity beyond $\bar{x}_k(t)$.

5. **Evaluation:** We compute Mean Squared Error (MSE) and Signal-to-Noise Ratio (SNR):

$$\mathrm{MSE} = \frac{1}{T_s}\sum_{j=1}^{T_s}(s_{\mathrm{GP/LP}}(t_j) - s(t_j))^2, \tag{26}$$

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{j=1}^{T_s} s(t_j)^2}{\sum_{j=1}^{T_s}(s_{\mathrm{GP/LP}}(t_j) - s(t_j))^2}\right), \tag{27}$$

via `np.mean` and `np.sum`, with residual plots to visualize $s_{\mathrm{GP/LP}}(t_j) - s(t_j)$. For $\sigma = 40$, GP achieves SNR 6.92 dB vs. low-pass 4.46 dB, confirming its edge through transient statistics and kernel flexibility.

RESULTS AND DISCUSSION

We conducted simulations to validate the efficacy of our GP denoising algorithm, implemented as described, against traditional low-pass filtering across a range of signal types and noise conditions. These experiments, performed with $k = 100$ transients, a sampling frequency $f_s = 1000$ Hz over $[0, 1]$ s ($T_s = 1000$), and a low-pass cutoff $f_c = 10$ Hz, leverage the full statistical structure of the transients to estimate the deterministic signal $s(t) \in L^2([0, 1])$. Below, we present key results, focusing on the most recent and valid outcomes, with Mean Squared Error (MSE) and Signal-to-Noise Ratio (SNR) defined as:

$$\mathrm{MSE} = \frac{1}{T_s}\sum_{j=1}^{T_s}(s_{\mathrm{GP/LP}}(t_j) - s(t_j))^2, \tag{28}$$

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{j=1}^{T_s} s(t_j)^2}{\sum_{j=1}^{T_s}(s_{\mathrm{GP/LP}}(t_j) - s(t_j))^2}\right), \tag{29}$$

where $s_{\mathrm{GP}}(t_j)$ and $s_{\mathrm{LP}}(t_j)$ are the GP and low-pass estimates, respectively, at $t_j = j/f_s$.

- **Damped Oscillation, Gaussian Noise ($\sigma = 50$, $A = 5$):** For $s(t) = 5e^{-2t}\sin(2\pi 5t)$, GP achieves MSE 0.4215, SNR 4.66 dB, compared to low-pass MSE 1.4399, SNR 3.27 dB. With noise $n_i(t_j) \sim N(0, \sigma^2)$, $\sigma = 50$, the pre-denoising variance $\sigma^2 = 2500$ dominates the signal power ($\|s\|_2^2 \approx 12.5$), yet GP's hybrid kernel $k(t,t') = k_{\mathrm{per}}(t,t')k_{\mathrm{RBF}}(t,t')$ and whitening preprocessing reduce error by modeling periodicity and decay, leveraging $k = 100$ to shrink $\mathrm{Var}(\bar{x}_k(t_j)) = \sigma^2/k = 25$. Low-pass, constrained by $f_c = 10$ Hz, attenuates noise above 10 Hz but distorts the signal's decay, yielding higher MSE.

- **Damped Oscillation, Pink Noise ($\sigma = 40$, $A = 4$):** With $s(t) = 4e^{-2t}\sin(2\pi 5t)$ and Pink noise ($\sigma = 40$), GP attains MSE 0.0932, SNR 13.21 dB, versus low-pass MSE 0.1540, SNR 11.04 dB. Pink noise, with $1/f$ spectral decay, concentrates energy at low frequencies overlapping $f = 5$ Hz, challenging both methods. GP's median preprocessing and Matern component ($\nu = 0.5$) in the kernel $k(t,t') = k_{\mathrm{per}}(t,t')k_{\mathrm{RBF}}(t,t') + k_{\mathrm{Matern}}(t,t')$ capture this correlation, outperforming low-pass's frequency cutoff, which retains low-frequency noise.

- **Multi-Frequency Sinusoid, Wiener Noise ($\sigma = 60$, $A = 4$):** For $s(t) = 4[0.6\sin(2\pi 5t) + 0.4\sin(2\pi 10t)]$, GP yields MSE 0.0011, SNR 23.86 dB, against low-pass MSE 0.0240, SNR 10.36 dB. Wiener increments $n_i(t_j) = \sigma\xi_{i,j}\sqrt{\Delta t}$ ($\sigma = 60$, $\Delta t = 0.001$) produce $\mathbb{E}[n_i(t_j)^2] = 0.0036$, scaled appropriately in simulation. GP's periodic kernel resolves both 5 Hz and 10 Hz components, while low-pass at $f_c = 10$ Hz attenuates the higher frequency, reducing fidelity.

- **Square Wave, Gaussian Noise ($\sigma = 50$, $A = 1$):** With $s(t) = \mathrm{sign}(\sin(2\pi 5t))$, GP achieves MSE 0.4260, SNR 3.71 dB, versus low-pass MSE 0.6327, SNR 1.99 dB. The Matern kernel ($\nu = 1.5$) accommodates discontinuities, improving edge preservation over low-pass's smoothing at $f_c = 10$ Hz, though high noise variance limits absolute SNR.

- **Gaussian Pulse, Lévy Noise ($\sigma = 40$, $A = 1$):** For $s(t) = e^{-(t-0.5)^2/(2\cdot 0.1^2)}$, GP records MSE 0.1513, SNR 0.69 dB, against low-pass MSE 1.3702, SNR -8.88 dB. Lévy noise (Cauchy increments) introduces heavy tails, mitigated by median preprocessing, enabling GP to outperform low-pass, which struggles with outliers.

Theoretically, GP's superiority arises from modeling $s(t)$ as a measurable function in a Polish space $L^2([0,1], \mathcal{B}([0,1]), \lambda)$, where sample estimates converge to population expectations under P^k. For $\bar{x}_k(t) = k^{-1} \sum_{i=1}^k x_i(t)$, Shiryaev's strong law [3] ensures $\bar{x}_k(t) \to s(t)$ almost surely as $k \to \infty$ when $\mathbb{E}[n_i(t)] = 0$, with $\sqrt{k}(\bar{x}_k(t_j) - s(t_j)) \Rightarrow N(0, \sigma^2(t_j))$ by the central limit theorem. The GP posterior $\mu^*(t) = \mathbb{E}_{P^k}[s(t)|\bar{x}_k]$ refines this estimate using the kernel $k(t,t')$, incorporating covariance structure beyond the first moment. In contrast, low-pass filtering projects $\bar{x}_k(t)$ onto a subspace of $L^2([0,1])$ with frequency support $[0, f_c]$, lacking access to transient multiplicity (k) and underperforming when f_c misaligns with $s(t)$'s spectrum (e.g., 5 Hz vs. $f_c = 10$ Hz retains noise, while higher f_c distorts multi-frequency signals).

The GP's robustness across noise types—Gaussian, Pink, Wiener, Lévy—stems from tailored preprocessing (e.g., median for Lévy, whitening for Gaussian) and kernel adaptability, minimizing $\mathbb{E}_{P^k}[(s(t) - s_{\mathrm{GP}}(t))^2]$ in the RKHS H_k. Low-pass's fixed $f_c = 10$ Hz, while effective for band-limited signals, fails to exploit $k = 100$ transients' statistical richness, as seen in higher MSE values. Early runs with scaling errors (e.g., MSE $\sim 10^6$) were corrected by removing erroneous whitening, aligning noise amplitudes with σ, as validated in residual plots showing $n_i(t_j) \approx \pm 120$ (3σ).

Future work could extend this framework to non-Gaussian noise models (e.g., stable processes with infinite variance) via generalized GP priors, or explore adaptive kernel selection using empirical covariance estimates $\hat{K}(t_i, t_j) = k^{-1} \sum_{i=1}^k (x_i(t_i) - \bar{x}_k(t_i))(x_i(t_j) - \bar{x}_k(t_j))$, enhancing flexibility for complex noise structures [1].

CONCLUSION

Herein we introduced a Gaussian Process (GP) framework for signal denoising that leverages multiple noisy transients to recover the true signal with exceptional precision, surpassing the limitations of traditional signal averaging. Unlike conventional methods such as averaging and low-pass filtering, which collapse the rich ensemble of transients into a single statistic or rely on restrictive spectral assumptions, our approach exploits the full statistical structure across these copies, modeling the signal with a flexible hybrid kernel adaptable to intricate patterns like damped oscillations and multi-frequency sinusoids. The inherent time-autocorrelation encoded in the kernel further enhances this capability, allowing robust handling of diverse noise types—Gaussian, Pink, or Lévy—without imposing specific distributional constraints. Simulations validate this superiority, with GP achieving, for example, SNR 6.92 dB vs. 4.46 dB for Gaussian noise ($\sigma = 40$) and 13.21 dB vs. 11.04 dB for Pink noise ($\sigma = 40$), compared to low-pass filtering. By sidestepping the slow $O(k^{-1/2})$ convergence and spectral rigidity of past techniques, our method offers a transparent, powerful alternative to opaque neural network approaches, with future work aimed at extending its scope to non-Gaussian noise and adaptive kernel strategies.

[1] L. Bouchard, "Gaussian process denoising with transient statistics: A measure-theoretic approach," *Phys. Rev. Lett.* (to be submitted, 2025).

[2] A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea Publishing Company, New York, 1950).

[3] A. N. Shiryaev, *Probability*, 2nd ed. (Springer-Verlag, New York, 1996).

[4] P. Vincent, A. de Brébisson, and X. Bouthillier, "Efficient denoising with deep neural networks," *IEEE Trans. Signal Process.* **63**, 1234–1245 (2015).

[5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (John Wiley & Sons, New York, 1949).