

# A Simple Parameter inference with transformers

**Adam Kleman**  
Czech Technical University in Prague

## Abstract

This work investigates parameter inference from synthetic periodic signals using an encoder-only Transformer model. The task is formulated as an inverse problem, where discretized time-domain observations are used to infer the underlying signal frequencies. Starting from a single-frequency sinusoid, the problem is extended to mixtures of two frequencies and to signals with an added nonlinear interaction term. Both deterministic regression and probabilistic inference based on normalizing flows are considered. Experimental results show that the Transformer reliably recovers frequencies in simple settings, while performance degrades with increasing noise and signal complexity. The probabilistic formulation captures this behavior through increased predictive uncertainty, providing additional insight in ambiguous regimes. These results demonstrate the suitability of Transformer-based models for controlled inverse signal problems and highlight the role of uncertainty-aware inference under noise.

## 1 Introduction

Estimating parameters from observed signals is a classical problem in signal processing<sup>1</sup>, with frequency estimation from periodic signals being a common example. Traditional methods based on spectral analysis work well under ideal conditions, but their performance often degrades in the presence of noise, short observation windows, or more complex signal structures.

In recent years, machine learning approaches have addressed tasks as supervised inverse problems, learning a direct mapping from observed data to underlying parameters. Transformer architectures have proven to be effective for sequential data modeling<sup>2</sup> and have recently been explored for inverse problems and parameter inference tasks<sup>3</sup>. In addition, probabilistic modeling techniques, such as normalizing flows, allow

uncertainty-aware inference beyond point estimates.

In this work, we investigate the use of an encoder-only Transformer to infer frequencies from synthetic periodic signals. Starting from a single-frequency sinusoid, we extend the task to mixtures of two frequencies and to signals with an added nonlinear interaction term. We further analyze deterministic and probabilistic formulations in order to study the behavior of the model under noise and ambiguity.

## 2 Related Work

Parameter estimation from periodic signals has been studied for a long time in signal processing. Frequency recovery is most often solved using spectral methods or simple interpolation techniques<sup>1</sup>. These approaches usually give good baseline results, but can degrade under noise, short observation windows, or model mismatch.

More recently, machine learning methods have been used to approach parameter estimation as a supervised inverse problem, where the goal is to learn a mapping from discretized signal samples to the underlying parameters. Transformer-based models have appeared in several scientific inference problems and have been applied to parameter estimation tasks involving noisy or overlapping signals<sup>4</sup>.

In order to represent uncertainty and not only a single predicted value, normalizing flows are often used for conditional density estimation, since they allow modeling more complex non-Gaussian distributions<sup>5</sup>.

When estimating more than one frequency at the same time, the predicted parameters do not have a natural order, making it necessary to use permutation-invariant losses or evaluation procedures<sup>6</sup>.

### 3 Dataset

All datasets used in this work are synthetically generated, allowing full control over signal parameters. The time variable  $t$  is defined in a fixed interval and uniformly discretized into a one-dimensional vector of samples, which serves as the input to the model. Unless stated otherwise, the number of time samples is fixed and the size of the dataset  $N$  is of the order of  $10^3$ , with higher values used in selected experiments. The time interval is chosen as  $t \in [0, 2\pi]$ , so for some signals it will be even less than one full period. Frequencies are sampled from a predefined interval, and additive Gaussian noise can be applied to study robustness under different noise levels.

#### 3.1 Single-frequency baseline

As a baseline, we consider a single periodic signal of the form  $y(t) = A \cdot \sin(\omega \cdot t)$ , where  $A$  is the amplitude and  $\omega$  is the frequency (Fig. 1a). For our experiment, let us fix the amplitude to  $A = 1$ , which reduces inferring the single parameter  $\omega$ .

#### 3.2 Two-frequency linear mixture

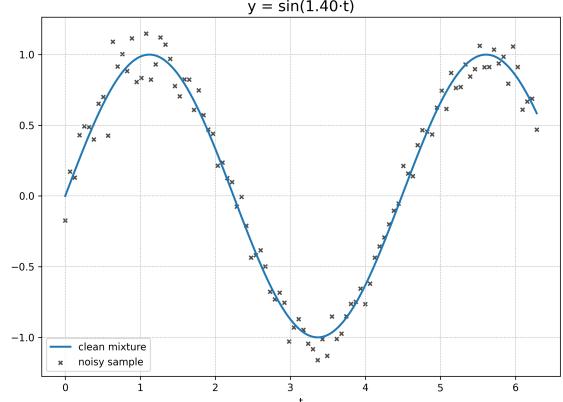
The data set is extended to signals composed of two sinusoidal components,  $y(t) = \sin(\omega_1 \cdot t) + \sin(\omega_2 \cdot t)$ . In this case, the target consists of two frequencies  $\omega_1$  and  $\omega_2$  (Fig. 1b). Since the order of the frequencies is not physically meaningful, the output is treated as an unordered pair  $\{\omega_1, \omega_2\}$ . This introduces permutation ambiguity and makes the inverse problem more challenging, especially when the value of frequency is higher.

#### 3.3 Nonlinear interaction term

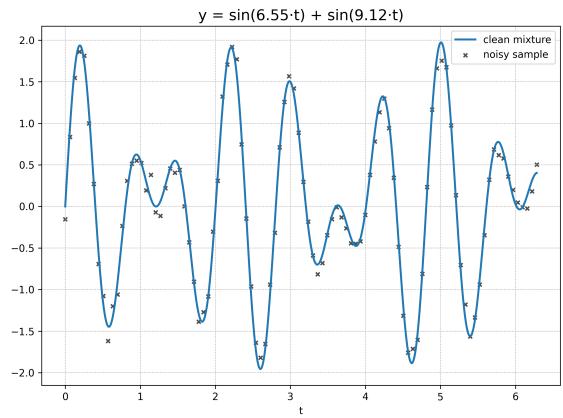
In a further extension, a nonlinear interaction term is added to the signal  $y(t) = \sin(\omega_1 \cdot t) + \sin(\omega_2 \cdot t) + \sin(\omega_1 \cdot t) \sin(\omega_2 \cdot t)$  (Fig. 1c). The additional product term increases the complexity of the signal and introduces stronger coupling between the two frequencies. Although the model architecture remains unchanged, this modification significantly increases ambiguity in the inverse mapping and leads to greater uncertainty in the inferred parameters, particularly in the presence of noise.

## 4 Methods

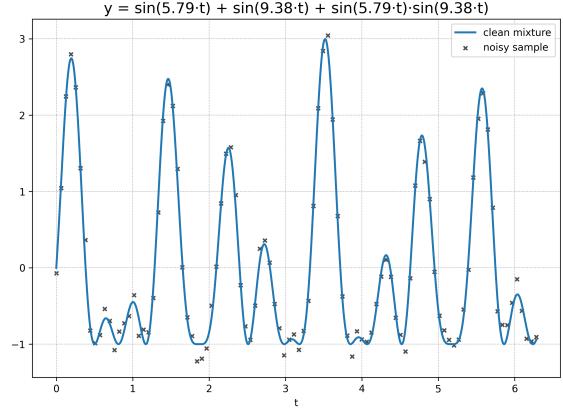
The model architecture is based on an encoder-only Transformer, which maps the input time series to a latent representation. Two output



(a) Single-frequency signal.



(b) Two-frequency linear mixture.



(c) Two-frequency nonlinear mixture.

Figure 1: Signal examples with different complexity with additional Gaussian noise ( $\sigma = 0.1$ )

formulations are considered: deterministic regression and probabilistic inference.

#### 4.1 Transformer encoder

The input signal is represented as a one-dimensional sequence of time samples. Each sample is first projected into a higher-

dimensional embedding space using a linear layer. Positional encoding is added to preserve information about the temporal order of the samples. The embedded sequence is then processed by a stack of Transformer encoder layers<sup>2</sup>, consisting of multi-head self-attention and feed-forward blocks. To obtain a fixed-size representation of the entire signal, the encoder output is averaged over the time dimension.

## 4.2 Deterministic regression

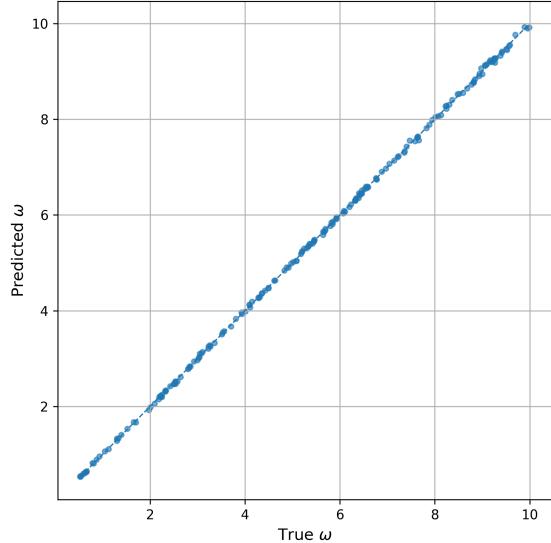
In the deterministic setting, the pooled Transformer representation is passed through a fully connected output head that predicts the signal parameters directly. For the single-frequency case, the model outputs a scalar frequency value. For the two-frequency case, the model outputs two values that correspond to the estimated frequencies. The model is trained using a mean squared error loss between the predicted and true parameters. This approach provides a strong baseline, but does not explicitly model the prediction uncertainty.

## 4.3 Probabilistic inference with normalizing flows

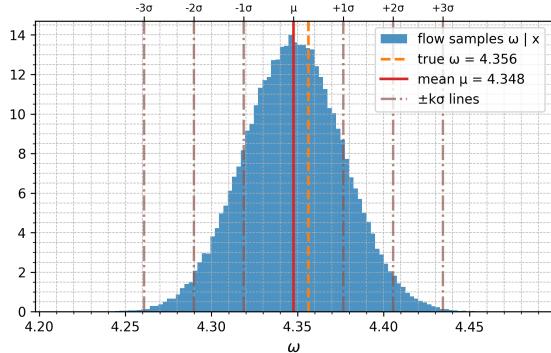
To model uncertainty in the inferred parameters, a probabilistic output head based on normalizing flows is used. The Transformer representation serves as a conditioning input to the flow, which defines a conditional probability distribution over the target parameters. The flow is constructed as a sequence of invertible transformations applied to a simple base distribution. In our implementation, the flow is constructed using masked auto-regressive transformations<sup>7</sup>, with permutations applied between layers to improve expressivity.

Training is performed by maximizing the log-likelihood of the true parameters under the predicted distribution, or equivalently by minimizing the negative log-likelihood. During inference, multiple samples can be drawn from the flow to obtain both a point estimate and a measure of uncertainty.

When predicting two frequencies, the order of the output is not physically meaningful. To account for this, permutation ambiguity is handled during training and evaluation by aligning predictions with targets using the best matching permutation. This ensures that the loss does not depend on an arbitrary ordering of the frequency parameters and allows a fair evaluation of the model performance.



(a) Predicted mean frequency versus true frequency for the test set.



(b) Posterior distribution  $p(\omega | x)$  for a single test signal. The true frequency, posterior mean, and  $\pm k\sigma$  intervals are shown.

Figure 2: Probabilistic inference results for single-frequency test example using a normalizing-flow output head. The results demonstrate both accurate point estimates and meaningful uncertainty quantification.

## 5 Experiments and Results

The experiments are designed to study the accuracy of inference, the robustness to noise, and the effect of increasing signal complexity. Model performance is evaluated using mean squared error (MSE) and mean absolute error (MAE) between predicted and true parameters. For probabilistic models, point estimates are obtained by averaging samples from the predicted distribution.

Probabilistic flow-based heads achieve lower error than deterministic regression (Tab. 1).

## 5.1 Single-frequency inference

The first set of experiments focuses on the simplest signal consisting of a single sinusoidal component. The Transformer model is trained to infer the frequency  $\omega$  from discretized signal samples. In this setting, the probabilistic model consistently learns an accurate mapping between the input signal and the target frequency (Fig. 2). The predicted values closely follow the ground-truth frequencies, with low MAE and MSE across the test set. Increasing the size of the data set or the discretization of the time improves the performance and accuracy of the model. However, even with a relatively small dataset of  $N = 1000$  samples, the model achieves good predictive performance.

## 5.2 Two-frequency linear mixtures

The second experiment considers signals composed of a linear combination of two sinusoids. The task is to infer the unordered pair of frequencies  $\{\omega_1, \omega_2\}$ . Compared to the single-frequency case, this setting is significantly more challenging. Although the model is able to recover both frequencies in most cases, errors increase with higher frequency values (Fig. 3). This behavior reflects the intrinsic ambiguity of the inverse problem rather than a failure of the model. Scatter plots in the frequency space show that predicted frequency pairs generally follow the true distribution, although with increased spread (Fig. 4).

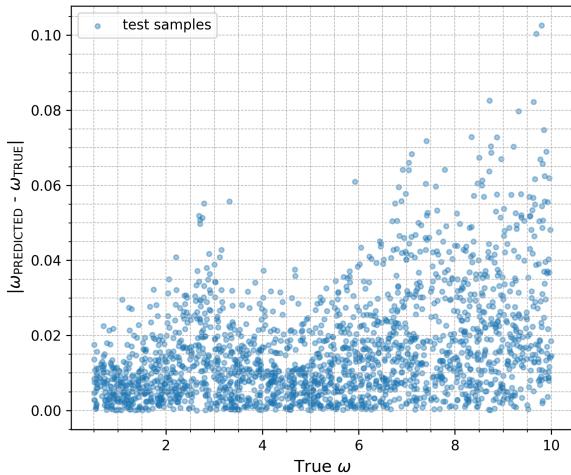


Figure 3: Absolute error between predicted and true frequency for each test sample. ( $N = 2000$ ,  $\omega \in [0.5, 10]$ ,  $t \in [0, 2\pi]$ )

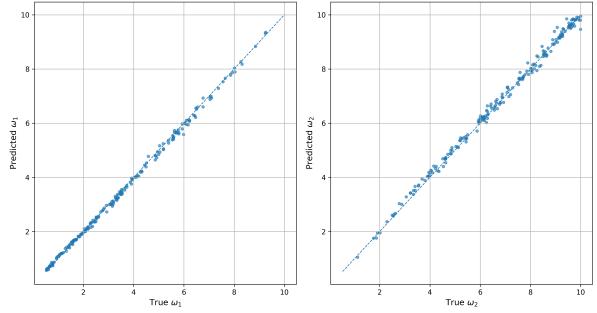


Figure 4: Predicted mean frequencies versus true frequencies for the two-frequency linear mixture, using a flow-based output head. Frequencies are sorted to resolve permutation ambiguity.

Table 1: Comparison of different output heads using the same Transformer encoder.

Head	MSE	MAE
Deterministic, $1 \times \omega$	0.006342	0.063750
Flow, $1 \times \omega$	0.001052	0.025040
Flow, $2 \times \omega$ (Lin.)	0.015823	0.097533
Flow, $2 \times \omega$ (Nonlin.)	0.013789	0.087584

## 5.3 Nonlinear interaction term

In the final experiment, a nonlinear interaction term is added to the two-frequency signal. This modification further increases the difficulty of the inference task. Although the model architecture is unchanged, probabilistic models show higher errors compared to the linear mixture case. Also, captures increased ambiguity through wider posterior distributions. These results suggest that the added nonlinear term introduces additional coupling between the frequencies, making the inverse mapping less identifiable.

## 5.4 Effect of noise

To evaluate robustness, additive Gaussian noise is applied to the input signals with increasing noise variance. As expected, the performance of the model degrades monotonically as the noise level increases. For small noise values, the model remains relatively stable and predictions remain accurate. At higher noise levels, errors increase and predictions become less reliable (Tab. 2). In the probabilistic setting, increased noise leads to broader predicted distributions, indicating greater uncertainty in the inferred parameters.

Table 2: Effect of Gaussian noise ( $\sigma$ ) on double-frequency linear mixture inference.

$\sigma$	Test MSE	MAE
0.10	0.015823	0.097533
0.20	0.017569	0.097023
0.30	0.029101	0.126703
0.40	0.043599	0.148789
0.50	0.068242	0.192128
0.60	0.089586	0.218073
0.70	0.140084	0.258257
0.80	0.378684	0.377969
0.90	0.602784	0.426835
1.00	0.684471	0.485811

## 6 Discussion

The experimental results highlight several important aspects of parameter inference from periodic signals using Transformer-based models. First, the strong performance in the single-frequency case confirms that the Transformer encoder can effectively extract frequency-related features from discretized time-domain signals. However, increasing the size of the data set or the time resolution only partially reduces the prediction error, indicating that some limitations are inherent to the inverse problem rather than to data scarcity alone.

For signals composed of two frequencies, the main source of error arises from permutation ambiguity and reduced identifiability when the frequencies are close. In such cases, multiple parameter configurations can produce similar observations, making precise inference difficult even for expressive models. The probabilistic approach based on normalizing flows provides a more informative representation in this regime, as increased uncertainty is reflected by broader predictive distributions rather than overly confident point estimates.

The addition of a nonlinear interaction term further increases the ambiguity without requiring architectural changes, suggesting that the proposed model is structurally robust but limited by the information content of the signal. These results suggest that uncertainty-aware inference is particularly important for inverse problems affected by noise and intrinsic ambiguity.

## 7 Conclusion

This work investigated parameter inference from synthetic periodic signals using an encoder-only Transformer architecture. Start-

ing from a single-frequency sinusoid, the model was shown to reliably recover the underlying frequency from discretized time-domain observations. The task was then extended to mixtures of two frequencies and to signals with an additional nonlinear interaction term, which significantly increased the difficulty of the inverse problem.

The experimental results showed that prediction errors increase with higher noise levels and increased signal complexity. The probabilistic formulation based on a normalizing-flow output head provided additional insight in these challenging settings by capturing uncertainty through broader predictive distributions. Overall, the results demonstrate that Transformer-based models are a viable tool for solving controlled inverse signal problems and highlight the importance of uncertainty-aware inference in settings with noise.

## References

- [1] Minglong Cheng, Guoqing Jia, Weidong Fang, Huiyue Yi, and Wuxiong Zhang. Accurate sinusoidal frequency estimation algorithm for internet of things based on phase angle interpolation using frequency shift. *Applied Sciences*, 12(12):6232, 2022.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.
- [3] Tian Zhou, Yifan Chen, and Hongwei Sun. Transformer-based solvers for inverse problems, 2024.
- [4] Lucia Papalini, Federico De Santi, Massimiliano Razzano, Ik Siong Heng, and Elena Cuoco. Can transformers help us perform parameter estimation of overlapping signals in gravitational wave detectors?, 2025.
- [5] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference, 2019.
- [6] Masataro Asai. Set cross entropy: Likelihood-based permutation invariant loss function for probability distributions, 2018.
- [7] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation, 2017.