



TALAGRAND INEQUALITY AND TIME SERIES IMPUTATION

O. Bernfield

The Harley School, Rochester, New York

A. Iosevich

Department of Mathematics, University of Rochester, Rochester, New York

J. Iosevich

Rochester Institute of Technology, Rochester, New York

S. Lyudovik

University of Rochester, Rochester, New York

A. Mayeli

Department of Mathematics, CUNY Graduate Center, New York

A. Nappo

Villanova University, Villanova, Pennsylvania

M. Zhang

University of Rochester, Rochester, New York

S. Zhou

University of Rochester, Rochester, New York

Received: 9/28/25, Accepted: 9/2/26, Published: 9/28/26

Abstract

The purpose of this paper is to study the imputation of time series. More precisely, let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ and suppose that the values $\{f(x)\}_{x \in M}$ are missing. Under what reasonable assumptions can we recover the missing values with high accuracy? Our work builds on earlier results where the imputation problem was studied using Logan's L^1 minimization idea combined with Bourgain–Talagrand type inequalities. We improve and extend those results. Our final result revisits Logan's classical noise removal procedure and uses Talagrand's theorem to obtain a strong bound if the noise function is suitably random. We also conduct a series of numerical experiments that demonstrate the presence of structure in large values of a forecastable time series.

1. Introduction

The main object studied in this paper is a time series $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, where $\mathbb{Z}_N = \{0, 1, 2, \dots, N-1\}$ denotes the integers modulo N and serves as a convenient model

for signals of length N . We define the Fourier transform of f by

$$\widehat{f}(m) = N^{-\frac{1}{2}} \sum_{x \in \mathbb{Z}_N} f(x) e^{-2\pi i x m / N}, \quad m \in \mathbb{Z}_N.$$

Suppose that the values $\{f(x) : x \in M\}$ are missing, where $M \subset \mathbb{Z}_N$. The main purpose of this paper is to provide algorithms for recovering these values and to discuss the theoretical and applied justification and limitations of the methods. The algorithms considered here seek a function $g : \mathbb{Z}_N \rightarrow \mathbb{C}$ that approximately agrees with f on M^c and minimizes the L^1 -norm of its Fourier transform.

1.1. Previous Results

The results in this subsection were established in [1], building on the methods developed in [7, 6].

For a positive integer n , write $[n] = \{0, 1, \dots, n-1\}$.

Definition 1. Let $0 < p < 1$. A random set $S \subset [n]$ is called *generic* if each element of $[n]$ is selected independently with probability p .

We shall use the following concentration inequality due to Talagrand.

Theorem 1 ([8]). *Let $(\varphi_j)_{j=1}^n$ be an orthonormal system in $L^2(\mathbb{Z}_N)$ such that $\|\varphi_j\|_{L^\infty} \leq K$ for $1 \leq j \leq n$. There exist a constant $\gamma_0 \in (0, 1)$ and a subset $I \subset \{1, \dots, n\}$ with $|I| \geq \gamma_0 n$ such that, for every $a = (a_i) \in \mathbb{C}^n$,*

$$\left(\sum_{i \in I} |a_i|^2 \right)^{\frac{1}{2}} \leq C_T K (\log(n) \log \log(n))^{\frac{1}{2}} \left\| \sum_{i \in I} a_i \varphi_i \right\|_{L^1},$$

where $C_T > 0$ is a universal constant.

Thus, the synthesis operator

$$T_I : \ell^2(I) \longrightarrow L^1(\mathbb{Z}_N), \quad (a_i) \longmapsto \sum_{i \in I} a_i \varphi_i,$$

has lower bound

$$\frac{1}{C_T K (\log(n) \log \log(n))^{\frac{1}{2}}}.$$

Remark 1. The point of Theorem 1 is that, even when the entire orthonormal system has complicated behavior, it contains a large subsystem with useful concentration properties. The argument in [1] gives the following probabilistic form suited to the present setting.

Theorem 2 ([1]). *There exists $\gamma_0 \in (0, 1)$ such that the following holds. Let M be a generic subset of $\mathbb{Z}_N$ satisfying $|M| \leq \gamma_0 N / \log(N)$. Then, with probability $1 - o_N(1)$, every function $h : \mathbb{Z}_N \rightarrow \mathbb{C}$ supported in M satisfies*

$$\left(\frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|^2 \right)^{\frac{1}{2}} \leq C_T (\log(N) \log \log(N))^{\frac{1}{2}} \frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|, \quad (1)$$

where $C_T > 0$ depends only on γ_0 .

Remark 2. The Talagrand constant C_T in Theorem 2 depends on γ_0 . Determining the precise nature of this dependence is an interesting problem.

We use normalized L^p -norms when the underlying measure is the probability measure μ on $\mathbb{Z}_N$ given by $\mu(\{x\}) = 1/N$:

$$\|u\|_{L^p(\mu)} := \left(\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |u(x)|^p \right)^{\frac{1}{p}}.$$

Thus, Equation (1) can be written as

$$\|\widehat{h}\|_{L^2(\mu)} \leq C_T (\log(N) \log \log(N))^{\frac{1}{2}} \|\widehat{h}\|_{L^1(\mu)}.$$

When the underlying measure is counting measure on $\mathbb{Z}_N$, we write

$$\|u\|_p := \left(\sum_{x \in \mathbb{Z}_N} |u(x)|^p \right)^{\frac{1}{p}}.$$

Definition 2. Let $0 \leq \epsilon < N$, let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, and let $S \subset \mathbb{Z}_N$. We say that $\widehat{f}$ is ϵ -concentrated on S if

$$\|\widehat{f}\|_{L^1(S^c)} \leq \frac{\epsilon}{N} \|\widehat{f}\|_1.$$

Theorem 3 ([1]). *Let $\delta \geq 0$, let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a nonempty generic subset of $\mathbb{Z}_N$ with $|M| \leq \gamma_0 N / \log(N)$ and γ_0 is as in Theorem 2. Let*

$$g \in \operatorname{argmin}_u \|\widehat{u}\|_1 \quad \text{subject to} \quad \|u - f\|_{L^1(M^c)} \leq \delta N^{-1} \|f\|_{L^1(M^c)}. \quad (2)$$

Suppose that $\widehat{f}$ is ϵ -concentrated on a set $S \subset \mathbb{Z}_N$ satisfying

$$|S| < \frac{1}{16C_T^2} \frac{N}{\log(N) \log \log(N)}.$$

Let $h = f - g$. Then, with probability $1 - o_N(1)$, either $h = 0$ or

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq (4\epsilon + 5\delta) \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \quad (3)$$

Remark 3. Theorem 3 applies with high probability when the missing set M is generic and satisfies the stated cardinality condition. The probabilistic mechanism, including the uniformity over functions supported in M , is established in [1].

Remark 4. Equation (3) can be rewritten as

$$\frac{\frac{1}{|M|} \sum_{x \in M} |h(x)|}{\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|} \leq 4\epsilon + 5\delta.$$

It would be desirable to replace the denominator by the mean of $|f|$ on M and thereby express the result in terms of the *weighted mean absolute error*, defined by

$$\text{MAE}_{\text{weighted}}(f, h) := \frac{\sum_{x \in M} |h(x)|}{\sum_{x \in M} |f(x)|}.$$

Although these two means need not be equal, the following theorem shows that they are close with high probability for a generic set M , provided that $\|f\|_\infty$ is controlled.

The following consequence of Hoeffding's inequality makes this observation precise.

Theorem 4. *Let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ be nonzero, and let M be a generic subset of $\mathbb{Z}_N$ with inclusion probability p . Then, for every $t > 0$,*

$$\begin{aligned} \Pr \left(M \neq \emptyset \text{ and } \left| \frac{1}{|M|} \sum_{x \in M} |f(x)| - \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)| \right| \geq t \right) \\ \leq 2 \left(1 - p + p \exp \left(-\frac{2t^2}{\|f\|_\infty^2} \right) \right)^N. \end{aligned} \quad (4)$$

Proof. Condition on the event $|M| = k$, where $1 \leq k \leq N$. Conditional on this event, M is uniformly distributed among the k -element subsets of $\mathbb{Z}_N$. Hoeffding's inequality for sampling without replacement [5] gives

$$\Pr \left(\left| \frac{1}{k} \sum_{x \in M} |f(x)| - \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)| \right| \geq t \mid |M| = k \right) \leq 2 \exp \left(-\frac{2kt^2}{\|f\|_\infty^2} \right).$$

Summing over k and using the binomial distribution of $|M|$ yields

$$\begin{aligned} \Pr \left(M \neq \emptyset \text{ and } \left| \frac{1}{|M|} \sum_{x \in M} |f(x)| - \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)| \right| \geq t \right) \\ \leq 2 \sum_{k=1}^N \binom{N}{k} p^k (1-p)^{N-k} \exp \left(-\frac{2kt^2}{\|f\|_\infty^2} \right) \\ \leq 2 \left(1 - p + p \exp \left(-\frac{2t^2}{\|f\|_\infty^2} \right) \right)^N. \end{aligned}$$

□

For fixed $t > 0$, the right-hand side of Equation (4) tends to zero whenever $Np \rightarrow \infty$.

Remark 5. Theorem 4 is less informative when $\|f\|_\infty$ is much larger than $\|f\|_{L^1(\mu)}$, as happens when f has an aberrant spike. For example, suppose that $s_1 \leq |f(x)| \leq s_2$ for every $x \neq x_0$ and that $|f(x_0)| \gg s_2$. When p is small, a generic set M is unlikely to contain x_0 , and the sample mean of $|f|$ on M can differ substantially from the mean on $\mathbb{Z}_N$.

1.2. Our Results

Our first result retains the L^1 -loosening in Equation (2) and improves the error bound in Equation (3).

Theorem 5. *Let $\delta \geq 0$, let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a nonempty generic subset of $\mathbb{Z}_N$ with $|M| \leq \gamma_0 N / \log(N)$ and γ_0 is as in Theorem 2. Let*

$$g \in \operatorname{argmin}_u \|\widehat{u}\|_1 \quad \text{subject to} \quad \|u - f\|_{L^1(M^c)} \leq \delta N^{-1} \|f\|_{L^1(M^c)}. \quad (5)$$

Suppose that $\widehat{f}$ is ϵ -concentrated on a set $S \subset \mathbb{Z}_N$ satisfying

$$|S| < \frac{1}{16C_T^2} \frac{N}{\log(N) \log \log(N)}.$$

Let $h = f - g$. Then, with probability $1 - o_N(1)$, either $h = 0$ or

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq \left(4\epsilon \frac{|S|}{N - \epsilon} + \left(4 \frac{|S|}{N} + 1 \right) \delta \right) \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \quad (6)$$

Our next result is a variant of Theorem 5 in which the loosening of the classical Logan minimization condition is imposed in the L^2 -norm rather than the L^1 -norm. Relative to Theorem 3, the term 4ϵ is replaced by $4\epsilon|S|/(N - \epsilon)$, while 5δ is replaced by $(4(|S|/N)^{\frac{1}{2}} + 1)\delta$. Numerical experiments indicate that this modification often yields improved results on real-life data sets [2].

Theorem 6. *Let $\delta \geq 0$, let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a nonempty generic subset of $\mathbb{Z}_N$ with $|M| \leq \gamma_0 N / \log(N)$ and γ_0 is as in Theorem 2. Let*

$$g \in \operatorname{argmin}_u \|\widehat{u}\|_1 \quad \text{subject to} \quad \|u - f\|_{L^2(M^c)} \leq \delta N^{-1} \|f\|_{L^2(M^c)}. \quad (7)$$

Suppose that $\widehat{f}$ is ϵ -concentrated on a set $S \subset \mathbb{Z}_N$ satisfying

$$|S| < \frac{1}{16C_T^2} \frac{N}{\log(N) \log \log(N)}.$$

Let $h = f - g$. Then, with probability $1 - o_N(1)$, either $h = 0$ or

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq \left(4\epsilon \frac{|S|}{N - \epsilon} + \left(4 \left(\frac{|S|}{N} \right)^{\frac{1}{2}} + 1 \right) \delta \right) \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \quad (8)$$

Our final result applies Logan's method to noise removal. We first recall the following theorem of Donoho and Stark.

Theorem 7 ([4]). *Let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ have Fourier transform supported in $S \subset \mathbb{Z}_N$. Suppose that the received signal is $r = f + n$, where $n : \mathbb{Z}_N \rightarrow \mathbb{C}$ is supported in $T \subset \mathbb{Z}_N$. If $|T||S| < N/2$, then f can be recovered by*

$$\operatorname{argmin}_u \|r - u\|_1 \quad \text{subject to} \quad \hat{u} \text{ is supported in } S.$$

Theorem 7 gives an exact recovery guarantee under deterministic support-size conditions. Using Talagrand's probabilistic technology, we obtain the following result for random trigonometric noise.

Theorem 8. *Let $f, n : \mathbb{Z}_N \rightarrow \mathbb{C}$, and suppose that the received signal is $r = f + n$. Assume that $\hat{n}$ is supported in a generic set $S \subset \mathbb{Z}_N$ with $|S| \leq \gamma_0 N / \log(N)$, where γ_0 is as in Theorem 2, and assume that f is supported in $T \subset \mathbb{Z}_N$. If*

$$|T| < \frac{N}{4C_T^2 \log(N) \log \log(N)},$$

then, with probability $1 - o_N(1)$, the minimizer

$$m \in \operatorname{argmin}_u \|r - u\|_1 \quad \text{subject to} \quad \hat{u} \text{ is supported in } S, \quad (9)$$

satisfies $m = n$, and hence $f = r - m$. More generally, let $0 \leq \epsilon < 1$ and suppose that

$$\|f\|_{L^1(T^c)} \leq \epsilon \|f\|_1, \quad (10)$$

and

$$|T| < \frac{N}{16C_T^2 \log(N) \log \log(N)}.$$

Then, with probability $1 - o_N(1)$, the recovered signal $r - m$ satisfies

$$\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x) - (r(x) - m(x))| \leq 4\epsilon \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|.$$

2. Proofs of Theorems

Proof of Theorem 5. For a set $A \subset \mathbb{Z}_N$, let 1_A denote its indicator function. Set $h = f - g$. Since f is feasible in Equation (5), the minimality of g gives $\|\hat{g}\|_1 \leq \|\hat{f}\|_1$.

On the other hand,

$$\begin{aligned}\|\widehat{g}\|_1 &= \|\widehat{f} - \widehat{h}\|_{L^1(S)} + \|\widehat{f} - \widehat{h}\|_{L^1(S^c)} \\ &\geq \|\widehat{f}\|_{L^1(S)} - \|\widehat{f}\|_{L^1(S^c)} + \|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)} \\ &\geq \left(1 - \frac{2\epsilon}{N}\right) \|\widehat{f}\|_1 + \|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)}.\end{aligned}$$

Consequently,

$$\|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)} \leq \frac{2\epsilon}{N} \|\widehat{f}\|_1. \quad (11)$$

Write $a = 1_M h$ and $b = 1_{M^c} h$. Equation (5) implies

$$\begin{aligned}\|\widehat{b}\|_{L^1(S)} &\leq |S| \|\widehat{b}\|_\infty \leq |S| N^{-\frac{1}{2}} \|b\|_1 \\ &\leq |S| N^{-\frac{3}{2}} \delta \|f\|_{L^1(M^c)} \leq |S| N^{-\frac{1}{2}} \delta \|f\|_{L^1(\mu)},\end{aligned}$$

and

$$\begin{aligned}\|\widehat{b}\|_1 &\leq N \|\widehat{b}\|_\infty \leq N^{\frac{1}{2}} \|b\|_1 \\ &\leq N^{-\frac{1}{2}} \delta \|f\|_{L^1(M^c)} \leq N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}.\end{aligned}$$

On the event of probability $1 - o_N(1)$ given by Theorem 2, since a is supported in M , Equation (1) and the assumption on $|S|$ give

$$\begin{aligned}\|\widehat{a}\|_{L^1(S)} &\leq |S|^{\frac{1}{2}} \|\widehat{a}\|_2 \\ &\leq |S|^{\frac{1}{2}} N^{-\frac{1}{2}} C_T \sqrt{\log(N) \log \log(N)} \|\widehat{a}\|_1 \\ &\leq \frac{1}{4} \|\widehat{a}\|_1.\end{aligned}$$

Since $a = h - b$, it follows that

$$\begin{aligned}\|\widehat{h}\|_{L^1(S)} &\leq \|\widehat{a}\|_{L^1(S)} + \|\widehat{b}\|_{L^1(S)} \\ &\leq \frac{1}{4} \|\widehat{h}\|_1 + \left(\frac{1}{4} + \frac{|S|}{N}\right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)} &= \|\widehat{h}\|_1 - 2\|\widehat{h}\|_{L^1(S)} \\ &\geq \frac{1}{2} \|\widehat{h}\|_1 - \frac{1}{2} \left(4 \frac{|S|}{N} + 1\right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}.\end{aligned}$$

Combining this estimate with Equation (11), we obtain

$$\|\widehat{h}\|_1 \leq \frac{4\epsilon}{N} \|\widehat{f}\|_1 + \left(4 \frac{|S|}{N} + 1\right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}. \quad (12)$$

The concentration assumption also gives

$$\begin{aligned} \left(1 - \frac{\epsilon}{N}\right) \|\widehat{f}\|_1 &\leq \|\widehat{f}\|_{L^1(S)} \leq |S| \|\widehat{f}\|_\infty \leq |S| N^{-\frac{1}{2}} \|f\|_1, \\ \text{and hence } \|\widehat{f}\|_1 &\leq \frac{|S| N^{\frac{1}{2}}}{N - \epsilon} \|f\|_1. \end{aligned} \quad (13)$$

Substituting Equation (13) into Equation (12) and using $\|f\|_{L^1(\mu)} = N^{-1} \|f\|_1$, we find that

$$\|\widehat{h}\|_1 \leq \left(\frac{4\epsilon|S|}{N - \epsilon} + \left(4 \frac{|S|}{N} + 1\right) \delta \right) N^{-\frac{1}{2}} \|f\|_1. \quad (14)$$

Finally, Fourier inversion gives $|h(x)| \leq N^{-1/2} \|\widehat{h}\|_1$ for every x . Averaging this inequality over M and applying Equation (14) proves Equation (6). $\square$

Proof of Theorem 6. Set $h = f - g$. The argument at the beginning of the preceding proof, which uses only minimality and ϵ -concentration, gives

$$\|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)} \leq \frac{2\epsilon}{N} \|\widehat{f}\|_1. \quad (15)$$

Again write $a = 1_M h$ and $b = 1_{M^c} h$. By Equation (7), Parseval's identity, and the inequality $\|f\|_2 \leq \|f\|_1$,

$$\begin{aligned} \|\widehat{b}\|_{L^1(S)} &\leq |S|^{\frac{1}{2}} \|\widehat{b}\|_2 = |S|^{\frac{1}{2}} \|b\|_2 \\ &\leq |S|^{\frac{1}{2}} \delta N^{-1} \|f\|_{L^2(M^c)} \leq |S|^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}, \end{aligned}$$

and

$$\begin{aligned} \|\widehat{b}\|_1 &\leq N^{\frac{1}{2}} \|\widehat{b}\|_2 = N^{\frac{1}{2}} \|b\|_2 \\ &\leq N^{-\frac{1}{2}} \delta \|f\|_{L^2(M^c)} \leq N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}. \end{aligned}$$

On the event supplied by Theorem 2, the same estimate for a as in the preceding proof gives $\|\widehat{a}\|_{L^1(S)} \leq \frac{1}{4} \|\widehat{a}\|_1$. Consequently,

$$\|\widehat{h}\|_{L^1(S)} \leq \frac{1}{4} \|\widehat{h}\|_1 + \left(\frac{1}{4} + \left(\frac{|S|}{N} \right)^{\frac{1}{2}} \right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)},$$

and hence

$$\begin{aligned} \|\widehat{h}\|_{L^1(S^c)} - \|\widehat{h}\|_{L^1(S)} &\geq \frac{1}{2} \|\widehat{h}\|_1 \\ &\quad - \frac{1}{2} \left(4 \left(\frac{|S|}{N} \right)^{\frac{1}{2}} + 1 \right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}. \end{aligned}$$

Combining this with Equation (15), we obtain

$$\|\widehat{h}\|_1 \leq \frac{4\epsilon}{N} \|\widehat{f}\|_1 + \left(4 \left(\frac{|S|}{N}\right)^{\frac{1}{2}} + 1\right) N^{\frac{1}{2}} \delta \|f\|_{L^1(\mu)}. \quad (16)$$

Combining Equation (16) with Equation (13) and $\|f\|_{L^1(\mu)} = N^{-1} \|f\|_1$ now yields

$$\|\widehat{h}\|_1 \leq \left(\frac{4\epsilon|S|}{N-\epsilon} + \left(4 \left(\frac{|S|}{N}\right)^{\frac{1}{2}} + 1\right) \delta\right) N^{-\frac{1}{2}} \|f\|_1. \quad (17)$$

As before, Fourier inversion gives $|h(x)| \leq N^{-1/2} \|\widehat{h}\|_1$. Averaging over M and applying Equation (17) proves Equation (8). $\square$

Proof of Theorem 8. Let m be a minimizer in Equation (9), and set $h = m - n$. Since n is feasible in Equation (9), minimality gives

$$\|r - m\|_1 \leq \|r - n\|_1 = \|f\|_1.$$

Moreover, $r - m = f - h$, and $\widehat{h}$ is supported in S . Applying Theorem 2 to $\widehat{h}$ and using the fact that the square of the unitary Fourier transform is reflection, we obtain, with probability $1 - o_N(1)$,

$$\|h\|_2 \leq C_T \sqrt{\log(N) \log \log(N)} N^{-\frac{1}{2}} \|h\|_1. \quad (18)$$

First suppose that f is supported in T . Then

$$\begin{aligned} \|f - h\|_1 &= \|f - h\|_{L^1(T)} + \|h\|_{L^1(T^c)} \\ &\geq \|f\|_1 + \|h\|_{L^1(T^c)} - \|h\|_{L^1(T)}. \end{aligned}$$

By the Cauchy–Schwarz inequality and Equation (18),

$$\|h\|_{L^1(T)} \leq |T|^{\frac{1}{2}} \|h\|_2 \leq |T|^{\frac{1}{2}} C_T \sqrt{\log(N) \log \log(N)} N^{-\frac{1}{2}} \|h\|_1.$$

The first assumption on $|T|$ implies that this is strictly less than $\frac{1}{2} \|h\|_1$ whenever $h \neq 0$. Hence

$$\|h\|_{L^1(T^c)} - \|h\|_{L^1(T)} = \|h\|_1 - 2\|h\|_{L^1(T)} > 0.$$

If $h \neq 0$, it follows that $\|r - m\|_1 = \|f - h\|_1 > \|f\|_1$, contradicting minimality. Thus, $h = 0$, so $m = n$ and $f = r - m$.

Now assume the weaker condition in Equation (10). The same decomposition gives

$$\begin{aligned} \|f - h\|_1 &= \|f - h\|_{L^1(T)} + \|h - f\|_{L^1(T^c)} \\ &\geq \|f\|_1 - 2\|f\|_{L^1(T^c)} + \|h\|_{L^1(T^c)} - \|h\|_{L^1(T)} \\ &\geq \|f\|_1 - 2\epsilon\|f\|_1 + \|h\|_{L^1(T^c)} - \|h\|_{L^1(T)}. \end{aligned}$$

Minimality therefore implies

$$\|h\|_{L^1(T^c)} - \|h\|_{L^1(T)} \leq 2\epsilon\|f\|_1. \quad (19)$$

Under the stronger assumption on $|T|$, Equation (18) gives

$$\|h\|_{L^1(T)} \leq \frac{1}{4}\|h\|_1.$$

Consequently,

$$\|h\|_{L^1(T^c)} - \|h\|_{L^1(T)} = \|h\|_1 - 2\|h\|_{L^1(T)} \geq \frac{1}{2}\|h\|_1.$$

Combining this with Equation (19) yields

$$\|h\|_1 \leq 4\epsilon\|f\|_1. \quad (20)$$

Since $f - (r - m) = h$, dividing Equation (20) by N gives the asserted average error bound. $\square$

3. Numerical Experiments

One important aspect of our recovery algorithms is that they rely on the set of “missing” values being generic, and specifically, that the “missing” values are chosen randomly. In those cases, we are able to get theoretical bounds on the error produced by our algorithms, as given in the previous sections. However, it is interesting to consider the case when the missing values are not random. In particular, we can observe what happens when many values of the same magnitude are removed or lost, such as the greatest values, the lowest values, and the middle values.

The figures below compare the true series (red) with imputed results obtained using an L^1 optimizer (green) and an L^1 δ -optimizer with L^2 -loosening (blue) for a subset of the U.S. inflation data in [3]. The L^1 optimizer is

$$g \in \underset{u}{\operatorname{argmin}} \|\hat{u}\|_1 \quad \text{subject to} \quad u(x) = f(x) \text{ for } x \notin M.$$

The L^1 δ -optimizer with L^2 -loosening is the algorithm in Theorem 6.

Figure 1 shows what happens when we take out a random third of the values, the upper third of the values, the lower third of the values, and the middle third of the values, respectively. As is evident from the graphs, the error of imputation increases dramatically when we take out the upper third or the lower third of the values, in which case the imputation underestimates or overestimates the missing values, respectively. Taking out the middle third results in a more reasonable error,

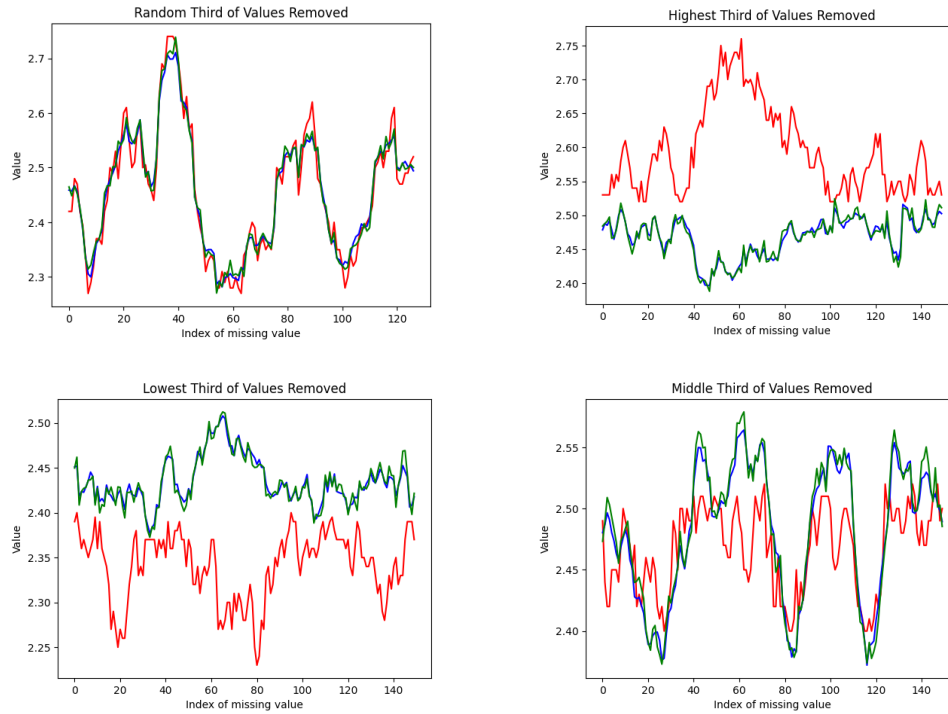


Figure 1: Imputed values with various sections of data removed.

though still much worse than the error resulting from taking out the random third of the values.

This suggests that the upper, lower, and middle thirds of the values have some structure that is important to the accuracy of the imputation. In particular, this structure does not exist in random sets, so when the missing values are random, there is far less error. Additionally, the upper third and lower third of values seem to have more structure than the middle third of values since the error resulting from removing the top and lower third is higher than that which results from removing the middle third.

Next, Figure 2 shows how the recovery error increases as the proportion p of removed data grows. The x -axis records the proportion of values removed, and the y -axis shows the mean absolute error in the recovered values. The blue and green curves correspond to removing the highest values and imputing the values using the L^1 optimizer and $L^1 \delta$ -optimizer, respectively (although the blue and green lines are almost identical, so the green line may be difficult to see), and red and orange curves correspond to removing values uniformly at random and imputing

the values using the L^1 optimizer and L^1 δ -optimizer, respectively (although the red and orange lines are almost identical, so the orange line may be difficult to see).

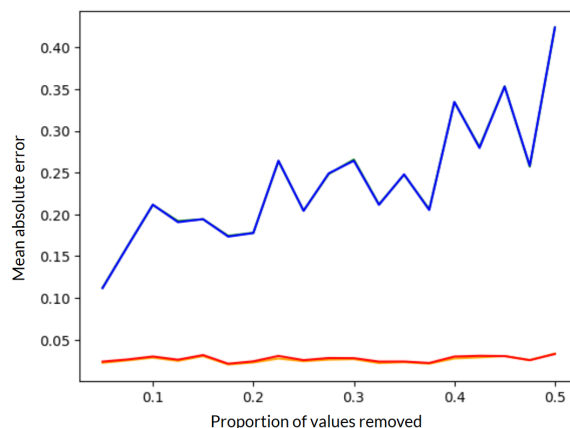


Figure 2: Mean absolute recovery error as a function of the proportion of removed data.

As one can see from the graph, the error is not only much higher when the highest values are removed than when random values are removed, but it also increases significantly more when the proportion of values removed increases. In particular, as we increase the proportion of removed values from about 10% to about 50%, the amount of absolute error approximately quadruples when the highest values are removed, whereas it stays approximately the same when random values are removed.

This suggests once again that the highest values have structure that is crucial to accurate recovery, and as increasingly more of this structure is lost, the recovery algorithm becomes increasingly less accurate. However, when random values are removed, it seems that this structure is mostly preserved, so the accuracy in recovery does not change significantly.

4. Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements. A. I. was supported in part by the National Science Foundation under NSF DMS-2154232. A. M. was supported in part by an AMS-Simons Research Enhancement Grant and by the National Science Foundation under NSF DMS-2453769.

References

- [1] W. Burstein, A. Iosevich, A. Mayeli, and H. Nathan, Fourier minimization and imputation of time series, preprint, [arXiv:2506.19226](https://arxiv.org/abs/2506.19226), 2025.
- [2] W. Burstein, A. Iosevich, A. Mayeli, and H. Nathan, Fourier minimization and time series imputation—experimental results, in preparation, 2025.
- [3] S. Chimmuniyan, US inflation rate, Kaggle dataset, <https://www.kaggle.com/datasets/samithsachidanandan/us-inflation-rate>.
- [4] D. L. Donoho and P. B. Stark, Uncertainty principles and signal recovery, *SIAM J. Appl. Math.* **49** (3) (1989), 906–931.
- [5] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (301) (1963), 13–30.
- [6] A. Iosevich, B. Kashin, I. Limonova, and A. Mayeli, Subsystems of orthogonal systems and the recovery of sparse signals in the presence of random losses, *Russian Math. Surveys* **79** (6) (2024), 1095–1097.
- [7] A. Iosevich and A. Mayeli, Uncertainty principles, restriction, Bourgain’s Λ_q theorem, and signal recovery, *Appl. Comput. Harmon. Anal.* **76** (2025), 101734.
- [8] M. Talagrand, Selecting a proportion of characters, *Israel J. Math.* **108** (1998), 173–191.