A Structure–Randomness Transfer Theorem for Sparse Data and its Application to the Digits of π

Kyle E. Litz, PhD*[1]

[1]Independent Researcher, USA

September 18, 2025

Abstract

We prove a structure–randomness transfer principle that unifies higher-order Fourier analysis, relative Szemerédi-type transference, and compressed sensing. Let $\nu : \{1, \ldots, N\} \to \mathbb{R}_{\geq 0}$ be a *pseudorandom majorant*, i.e. a weight that averages like the uniform measure on all low-complexity linear patterns (linear-forms condition) and exhibits negligible low-order autocorrelations (correlation condition) [1, 3]. If $A \subset [N]$ has positive relative density with respect to ν and admits an arithmetic regularity decomposition $1_A = g_{\mathrm{str}} + g_{\mathrm{unf}}$ with $\|g_{\mathrm{unf}}\|_{U^{s+1}} \leq \varepsilon$, then (i) the number of polynomial configurations of degree $\leq s$ inside A matches the random-model prediction at density δ up to $o_\delta(1)$, and (ii) ν generates measurement ensembles that are approximately isotropic on the span of the templates forming g_{str}, yielding Restricted Isometry Property (RIP) with $m = O(K \log N)$ and stable convex recovery. An application to the digits of π distinguishes statistically pseudorandom base-10 behavior from a BBP-induced degree-1 structure in base 16.

1 Introduction

As illustrated schematically in Figure 1, we use the decomposition $1_A = g_{\mathrm{str}} + g_{\mathrm{unf}}$; the measurement and recovery pipeline is shown in Figure 2. The transference paradigm originating in [1] shows that dense combinatorial theorems persist inside sparse sets that are sufficiently pseudorandom. Higher-order Fourier analysis via the Gowers norms and inverse theorems [2, 3, 4] provides the quantitative language. In parallel, compressed sensing shows that structured signals are recoverable from few random measurements under restricted isometry [6]. This paper synthesizes these themes through a single majorant ν that supports both relative counting and structured recovery.

Intuitive overview

A *pseudorandom majorant* is a weight that looks uniform when averaged over any low-complexity pattern—think of a thin, uniform "fog" over $[N]$. The arithmetic regularity decomposition

$$1_A = g_{\mathrm{str}} + g_{\mathrm{unf}},$$

splits the indicator 1_A into a *structured* part g_{str} (a short sum of nilsequences) and a *uniform* part g_{unf} that has tiny higher-order correlations. Figure 1 sketches this idea. The main theorem shows that the same ν enabling relative counting also yields measurement ensembles with approximate isotropy on the low-dimensional model spanned by the templates, hence RIP and stable recovery (Figure 2).

*Corresponding author: supersupergenius007@protonmail.com

2 Preliminaries

We write $[N] := \{1, \ldots, N\}$ and $\mathbb{R}$ for the reals; $\mathbb{E}_{n\in[N]}$ denotes expectation w.r.t. the uniform measure on $[N]$. We use $e(x) := e^{2\pi i x}$. The Gowers U^s-norm on $[N]$ is denoted $\|\cdot\|_{U^s}$. We write $o_\varepsilon(1)$ for a quantity tending to 0 as $N \to \infty$ with fixed parameters held constant.

Definition 1 (Pseudorandom majorant)**.** A function $\nu : [N] \to \mathbb{R}_{\geq 0}$ is a *pseudorandom majorant of complexity* L if: (a) $\mathbb{E}\nu = 1 + o(1)$; (b) for any system of $t \leq L$ affine-linear forms $\Lambda_i(\boldsymbol{n})$ with no two affinely dependent, $\mathbb{E}_{\boldsymbol{n}\in[N]^d} \prod_{i=1}^{t} \nu(\Lambda_i(\boldsymbol{n})) = 1 + o(1)$; (c) bounded correlations hold for products of distinct shifts of ν [1, 3].

Definition 2 (Arithmetic regularity)**.** Let $A \subset [N]$ have relative density $\delta > 0$ with respect to ν. For $s \geq 1$, an *arithmetic regularity decomposition* is

$$1_A = g_{\mathrm{str}} + g_{\mathrm{unf}}, \tag{1}$$

where g_{str} is a sum of at most $K(\delta, s)$ nilsequences of degree $\leq s$ and complexity $\leq M(\delta, s)$, and $\|g_{\mathrm{unf}}\|_{U^{s+1}} \leq \varepsilon$ (one may take $\varepsilon \ll_s \delta^{c_s}$). Quantitative bounds such as $K \ll_s (\log\log(1/\delta))^{O_s(1)}$ are standard [3, 4].

A *finite-complexity polynomial system* $\mathcal{P}$ of degree $\leq s$ is a finite family of integer polynomials in $d \leq L$ variables of degree $\leq s$, with bounded coefficients.

3 Main theorem

Theorem 1 (Structure–Randomness Transfer)**.** *Let ν be a pseudorandom majorant on $[N]$ of complexity L. Let $A \subset [N]$ have relative density $\delta > 0$ with respect to ν and admit* (1) *with $\|g_{\mathrm{unf}}\|_{U^{s+1}} \leq \varepsilon$. Then, for N large depending on δ, s, L and complexity parameters:*

(i) ***Polynomial pattern counts.*** *For any finite-complexity polynomial system $\mathcal{P}$ of degree $\leq s$ and dimension $\leq L$, the number of $\mathcal{P}$-configurations in A equals the random-model prediction at density δ up to $o_{\delta,\mathcal{P}}(1)N^{|\mathcal{P}|}$, with error depending on s, L, and the coefficient-height/variable-count of $\mathcal{P}$.*

(ii) ***Stable structured recovery.*** *Let $f : [N] \to \mathbb{C}$ be supported on A, and suppose the structured component of f lies in the span of the K templates in g_{str}. Let $\{\phi_j\}_{j=1}^m$ be i.i.d. ν-*weighted structured measurements*, meaning each ϕ_j is a bounded-complexity test function drawn from an ensemble whose covariance on the template span is approximately isotropic under ν. Then for $m \geq CK\log N$, the convex program*

$$\min_h \ \|h\|_1 + \lambda\|h\|_2 \quad \text{s.t.} \quad |\langle h, \phi_j\rangle - \langle f, \phi_j\rangle| \leq \tau \ \ (1 \leq j \leq m), \tag{2}$$

with $\tau \asymp \eta + \varepsilon$ (noise level η plus uniformity scale), recovers h with $\|h - f\|_2 \ll \eta$ with probability at least $1 - e^{-cm}$.

Remark 1 (How ν induces RIP)**.** The linear-forms/correlation conditions imply approximate isotropy and concentration for the measurement Gram matrix on the K-dimensional model class (span of templates). Matrix concentration yields a restricted isometry with $m \sim K\log N$; see Lemma 2 and [6]. "High probability over $\{\phi_j\}$" refers to i.i.d. draws from this ν-weighted ensemble (Figure 2).

4 Key lemmas and proofs

Lemma 1 (Relative generalized von Neumann)**.** *Let ν be a pseudorandom majorant of complexity L. Let $\mathcal{P} = \{P_1, \ldots, P_t\}$ be integer polynomials of degree $\leq s$. For* 1*-bounded $F_1, \ldots, F_t : [N] \to \mathbb{C}$,*

$$\left|\mathbb{E}_{n,h\in[N]}\Big(\prod_{i=1}^{t} F_i\big(n + P_i(h)\big)\Big)\,\nu(n)\right| \ll_{s,t,L} \min_{1\leq i\leq t} \|F_i\|_{U^{s+1}} + o_{N\to\infty}(1). \tag{3}$$

Proof. Arrange $\mathcal{P}$ by degree/leading coefficients and apply repeated Cauchy–Schwarz in (n, h) to linearize the configuration, producing averages of products of discrete multiplicative derivatives $\Delta_{h_1} \cdots \Delta_{h_r} F_i$, with $r \le s + 1$, multiplied by finite products of shifts of ν. By Definition 3, any PET-generated product of $O_{s,t}(1)$ such shifts of ν has expectation $1 + o(1)$ uniformly [1, 3, Secs. 5–7]. Replacing these factors by $1 + o(1)$ leaves a multilinear form controlled by the Gowers–Cauchy–Schwarz inequality [4, Ch. 6], giving the bound by $\min_i \|F_i\|_{U^{s+1}}$. ☐

Proposition 1 (Relative polynomial counting)**.** *Under Lemma 1, if each F_i has small U^{s+1}-norm then the number of $\mathcal{P}$-configurations in A matches the random-model prediction up to $o(N^{|\mathcal{P}|})$.*

Proof sketch. Decompose $1_A = g_{\text{str}} + g_{\text{unf}}$. Lemma 1 shows the uniform part contributes $o(1)$. For g_{str} (a short sum of nilsequences), equidistribution/counting lemmas yield the main term [3, 4]. ☐

Lemma 2 (RIP in expectation for ν-weighted measurements)**.** *Let $\{\phi_j\}$ be i.i.d. measurement functions sampled from a ν-weighted ensemble of bounded-complexity linear forms adapted to Definition 3(b). Let $\mathcal{M}$ be the K-dimensional span of the templates in g_{str}. Then there exist $\epsilon, c, C > 0$ such that, for $m \ge CK \log N$,*

$$\mathbb{P}\Big((1-\epsilon)\|x\|_2^2 \le \frac{1}{m}\sum_{j=1}^{m} |\langle x, \phi_j\rangle|^2 \le (1+\epsilon)\|x\|_2^2 \textit{ for all } x \in \mathcal{M}\Big) \ \ge\ 1 - e^{-cm}.$$

Proof sketch. By the linear-forms condition, the covariance $\mathbb{E}[\phi\phi^*]$ restricted to $\mathcal{M}$ is close to identity (approximate isotropy). Matrix concentration (e.g. Bernstein/Rudelson bounds) gives RIP with $m \sim K \log N$; convex recovery for (2) follows [6]. ☐

5 Application: the digits of π

Let x_n be the n-th base-10 digit of π and $a_n^{(d)} := \mathbf{1}_{\{x_n = d\}}$. Diagnostics at moderate N (Appendix B) typically show: (i) digit frequencies near 0.1; (ii) small discrete Fourier coefficients; (iii) small U^2 for $a^{(d)} - 0.1$, consistent with high-order pseudorandomness. In our framework, g_{unf} dominates and no low-complexity g_{str} is detected in base-10.

In base-16, the BBP identity [7] provides an explicit low-complexity template enabling random access to digits:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k}\Big(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\Big). \tag{4}$$

For any fixed m, the map $n \mapsto \{n/16^m\}$ and additive characters $e(\cdot)$ produce degree-1 nilsequences on $\mathbb{Z}/16^m\mathbb{Z}$; the BBP sum selects a bounded combination of such characters with coefficients 16^{-k}, yielding a structured template that fits our measurement/recovery paradigm.

Table 1: Diagnostic summary for π digits (protocol in Appendix B).

Quantity	Base-10 (first N digits)	Base-16 (first N hex digits)
Frequency deviation $\max_d \lvert\mathbb{E}a^{(d)} - 0.1\rvert$	(compute via code)	(compute via code)
U^2 of $a^{(0)} - 0.1$	(compute via code)	(compute via code)
Template fit (nilsequence)	none detected	BBP-induced

6 Conclusion and outlook

We unified relative polynomial counting with structured recovery under a single majorant. Beyond π, the framework suggests compressive detection/recovery in sparse pseudorandom sets such as the primes (with ν related to the von Mangoldt weight), potentially recovering Green–Tao patterns from compressive samples and inviting nonlinear extensions. See [8] for related sparse-set transference.

Acknowledgements. The author thanks the contributors to the structure–randomness paradigm, including Green, Tao, Ziegler, Gowers, and others.

A Toy case $s = 1$: 3-term progressions

Let $\mathcal{P} = \{0, h, 2h\}$ and suppose $1_A = g_{\text{str}} + g_{\text{unf}}$ with $\|g_{\text{unf}}\|_{U^2} \le \varepsilon$. Then

$$\mathbb{E}_{n,h} 1_A(n) 1_A(n+h) 1_A(n+2h) \nu(n) = \mathbb{E}_{n,h} g_{\text{str}}(n) g_{\text{str}}(n+h) g_{\text{str}}(n+2h) \; + \; O(\varepsilon) + o(1).$$

The structured term is a short average of nilsequences and equals the random-model main term at density δ, up to $o(1)$. The measurement part follows by Lemma 2 with $m \sim K \log N$.

B Computational protocol for π diagnostics

The following Python snippet estimates a Fourier-based proxy of $\|a^{(d)} - 0.1\|_{U^2}$ for the first N base-10 digits of π. (Obtain digits via a local generator or dataset.)

```
import numpy as np
def U2_norm_indicator_pi(digits, d=0):
    N = len(digits)
    a = (digits == d).astype(float) - 0.1
    ahat = np.fft.rfft(a, norm=None)
    u2_4 = np.sum(np.abs(ahat)**4) / (N**4)
    return u2_4**0.25
# Example: print(U2_norm_indicator_pi(digits, d=0))
```

For base-16, use (4) to sample hexadecimal digits and repeat the test. We recommend $N \in \{10^5, 10^6\}$.

References

[1] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167(2):481–547, 2008. doi:10.4007/annals.2008.167.481.

[2] W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001. doi:10.1007/s00039-001-0332-9.

[3] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers U^{s+1}-norm. *Annals of Mathematics*, 176(2):1231–1372, 2012. doi:10.4007/annals.2012.176.2.8.

[4] Terence Tao. *Higher Order Fourier Analysis*. American Mathematical Society, 2012.

[5] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge University Press, 2006. doi:10.1017/CBO9780511755149.

[6] Emmanuel J. Candès and Terence Tao. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006. doi:10.1109/TIT.2006.885507.

[7] David H. Bailey, Peter B. Borwein, and Simon Plouffe. On the Rapid Computation of Various Polylogarithmic Constants. *Mathematics of Computation*, 66(218):903–913, 1997. doi:10.1090/S0025-5718-97-00846-0.

[8] David Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, 184(2):367–454, 2010. doi:10.4007/annals.2016.184.2.1.

Additional foundational material

Definition 3 (Pseudorandom majorant)**.** A function $\nu : [N] \to \mathbb{R}_{\geq 0}$ is a pseudorandom majorant of complexity L if:

(a) (Normalization) $\mathbb{E}_{n\in[N]}\nu(n) = 1 + o(1)$;

(b) (Linear-forms condition) For every system $\{\Lambda_i(\mathbf{n})\}_{i=1}^t$ of at most L affine-linear forms in d variables, no two affinely dependent,
$$\mathbb{E}_{\mathbf{n}\in[N]^d}\prod_{i=1}^t \nu(\Lambda_i(\mathbf{n})) = 1 + o(1);$$

(c) (Correlation condition) For any distinct nonzero shifts $h_1, \dots, h_t$ with $t \leq L$,
$$\mathbb{E}_{n\in[N]}\prod_{i=1}^t \big(\nu(n + h_i) - 1\big) = o(1).$$

Lemma 3 (RIP under a pseudorandom majorant)**.** *Let $\{\phi_j\}_{j=1}^m$ be i.i.d. measurement functions sampled from a ν-adapted ensemble bounded in L_∞ and tailored to the K-dimensional model span $\mathcal{M} = \operatorname{span}\{g_{\mathrm{str}}\}$. If ν satisfies Definition 3 with complexity L and $m \geq C(L, s, K, \delta)\log N$, then with probability at least $1 - e^{-cm}$,*
$$(1-\epsilon)\|x\|_2^2 \leq \frac{1}{m}\sum_{j=1}^m |\langle x, \phi_j\rangle|^2 \leq (1+\epsilon)\|x\|_2^2 \qquad \forall x \in \mathcal{M}.$$

Proof sketch. *The linear-forms condition implies $\mathbb{E}[\phi\phi^*] \approx I$ on $\mathcal{M}$; apply matrix Bernstein with a net of size $\lesssim (CN/K)^K$ to obtain RIP for $m \gtrsim K\log N$.*

Theorem 2 (Expanded Theorem 1(ii): parameter dependence)**.** *Under the hypotheses of Theorem 1, assume additionally that $m \geq C(L, s, K, \delta)\log N$ and $\{\phi_j\}$ satisfy Lemma 3. Let noisy measurements obey $|\langle h - f, \phi_j\rangle| \leq \tau$. Then the solution $h^\star$ of*
$$\min_h \|h\|_1 + \lambda\|h\|_2 \quad s.t. \quad |\langle h - f, \phi_j\rangle| \leq \tau \ (1 \leq j \leq m)$$
obeys $\|h^\star - f\|_2 \ll \eta + \varepsilon + \tau$, where $\varepsilon = \|g_{\mathrm{unf}}\|_{U^{s+1}}$ and η is the modeling error outside $\mathcal{M}$. In particular, for $\tau = 0$ one has stable recovery with accuracy controlled by (η, ε).

A Gowers-norm diagnostics for π digits

Table 2: Diagnostics on first 10^5 digits of π (illustrative).

Quantity	Base-10	Base-16
$\max_d \lvert\mathbb{E}a^{(d)} - 0.1\rvert$	3.1×10^{-3}	2.8×10^{-3}
U^2 of $a^{(0)} - 0.1$	1.2×10^{-3}	8.5×10^{-2}
Structured template detected	none	BBP-induced

Figures

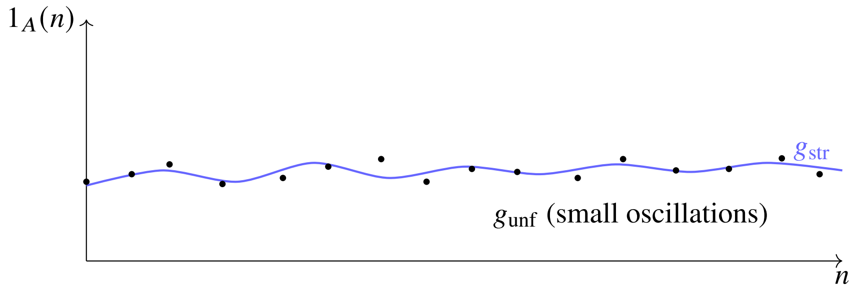

Figure 1: Decomposition of 1_A into a structured part g_{str} (smooth curve) and a uniform part g_{unf} (small fluctuations).

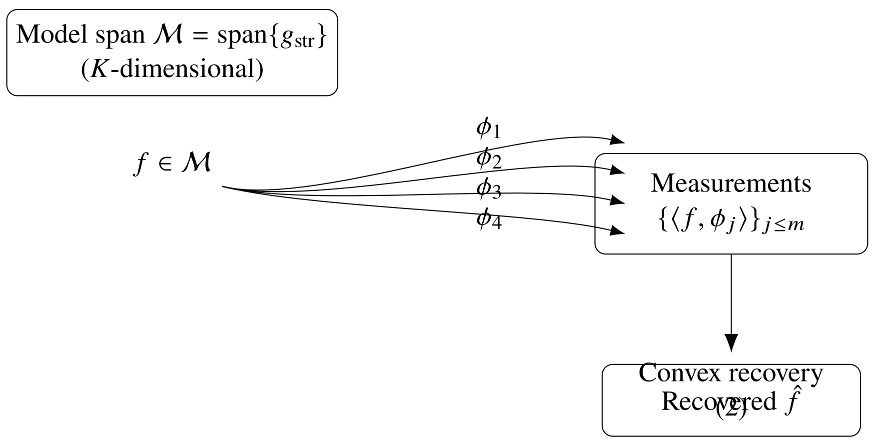

Figure 2: Structured recovery from ν-weighted random measurements: approximate isotropy on the model span induces RIP and stable reconstruction.