

Linear Gaussian SURE Bias for Haar Soft Thresholding under Smooth Noise

Alper Ferudun*

11 October 2026

Abstract

We give a fixed, smooth, sub-Gaussian noise distribution for which the bias of the Gaussian Stein unbiased risk estimate does not vanish after division by dimension, even for ordinary Haar soft thresholding. For every dyadic dimension $d \geq 2$, threshold $1/8$ and signal constant on each finest coordinate pair, the bias is at most $-273d/1604$. The noise coordinates are independent, identically distributed and standardized; their density is an everywhere positive mixture of two Gaussians. The bound covers both an unchanged and a thresholded constant coefficient. At the zero signal, the normalized bias has an exact absolutely convergent scale series with a strictly negative limit and an explicit truncation bound. The construction uses the positive proportion of localized Haar rows, not an independence assumption on the transformed coefficients. It concerns fixed-threshold bias and does not contradict positive results under strong log-concavity or resolve adaptive threshold selection.

1 Model and main result

Let $X = \theta + Y \in \mathbb{R}^d$, where θ is deterministic and the coordinates of Y are independent, centered and have variance one. For an estimator $\hat{\theta} = X + h(X)$ define

$$\hat{r}_h = d + 2 \operatorname{div} h(X) + \|h(X)\|^2, \quad e_d = \mathbb{E} \hat{r}_h - \mathbb{E} \|\hat{\theta} - \theta\|^2. \quad (1)$$

The statistic in (1) is the unit-variance Gaussian SURE formula, used here without assuming Gaussian noise. Its bias is the quantity under study, not an assertion of unbiasedness.

Goldstein's AIM Problem 1.4 [1] asks for non-Gaussian SURE bias identities, dimension-normalized bias, and applications to wavelet shrinkage. Fathi, Goldstein, Reinert and Saumard [2, Section 4] develop the Stein-kernel and zero-bias framework for this question. The underlying risk expansion and general bias identities are prior work. Their positive results impose additional hypotheses, and their wavelet discussion identifies the difficulty caused by loss of independence after transformation. Our purpose is narrower: an explicit smooth-noise Haar counterexample with a uniform quantitative bound.

For $d = 2^J$, let H_d be the real orthogonal Haar matrix. It has one constant row $d^{-1/2}(1, \dots, 1)$ and, for each $m = 2, 4, \dots, d$, d/m detail rows on disjoint consecutive blocks of length m . Within each such block the row is $m^{-1/2}$ on its first half and $-m^{-1/2}$ on its second half. In particular, its $d/2$ finest rows are

$$\frac{e_{2k-1} - e_{2k}}{\sqrt{2}}, \quad 1 \leq k \leq d/2. \quad (2)$$

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

Write $c_t(z) = \text{clip}_t(z) = \max\{-t, \min\{z, t\}\}$ and $s_t(z) = z - c_t(z)$ for soft thresholding at $t > 0$. Choose a set S of rows to threshold, and put

$$q_j(z) = -\mathbf{1}_{\{j \in S\}} c_t(z_j), \quad h(x) = H_d^T q(H_d x). \quad (3)$$

Thus each selected coefficient is replaced by s_t of that coefficient; unselected coefficients are unchanged. The usual choice is all detail rows, leaving the constant coefficient untouched.

Theorem 1.1. *Let $d = 2^J$, $J \geq 1$, and let*

$$Y_i = \frac{\xi_i + \eta G_i}{\sqrt{1 + \eta^2}}, \quad \delta = \frac{\eta^2}{1 + \eta^2}, \quad (4)$$

where $\eta > 0$, the ξ_i are fair independent signs, and the G_i are independent standard normal variables, independent of the signs. Suppose S contains every finest row (2), and $\theta_{2k-1} = \theta_{2k}$ for every k . Then

$$\frac{e_d}{d} \leq 2t \frac{|S|}{d} - \frac{1}{2} + \frac{\delta}{2t^2} \leq 2t - \frac{1}{2} + \frac{\delta}{2t^2}. \quad (5)$$

In particular, the fixed choices $\eta = 1/20$ and $t = 1/8$ give

$$\boxed{e_d \leq -\frac{273}{1604} d.} \quad (6)$$

The common noise density is positive and C^∞ , and $\mathbb{E}e^{uY_i} \leq e^{u^2/2}$ for every real u .

There is no bound on the sizes of the pairwise-constant signal values. Both the usual details-only convention and thresholding all coefficients satisfy the hypotheses. The law, threshold and constant do not depend on dimension. Thus dimension growth alone does not make this bias negligible, even without atoms or heavy tails.

2 Risk identity and the localized cancellation

Lemma 2.1. *Let H be any real orthogonal matrix, and let Y have an absolutely continuous law, mean zero and covariance I_d . Use (3) with H in place of H_d . Set $\mu = H\theta$ and $\epsilon = HY$. Then*

$$e_d = -2 \sum_{j \in S} \mathbb{P}(|\mu_j + \epsilon_j| < t) + 2 \sum_{j \in S} \mathbb{E}[\epsilon_j c_t(\mu_j + \epsilon_j)]. \quad (7)$$

Consequently,

$$e_d \leq -2 \sum_{j \in S} \mathbb{P}(|\mu_j + \epsilon_j| < t) + 2t|S|. \quad (8)$$

Proof. The correction q is Lipschitz and has diagonal derivative with entries $-\mathbf{1}_{\{j \in S\}} \mathbf{1}_{\{|z_j| < t\}}$ almost everywhere. Absolute continuity makes threshold boundaries null events. The chain rule and invariance of trace under orthogonal conjugation give

$$\text{div } h(X) = - \sum_{j \in S} \mathbf{1}_{\{|\mu_j + \epsilon_j| < t\}}.$$

Since $\|h(X)\| \leq t\sqrt{|S|}$ and Y has finite second moments, all relevant expectations exist. Expanding the squared error yields

$$\mathbb{E}\|\hat{\theta} - \theta\|^2 = d + \mathbb{E}\|h(X)\|^2 + 2\mathbb{E}\langle Y, h(X) \rangle.$$

Orthogonality changes the last inner product into $\langle \epsilon, q(\mu + \epsilon) \rangle$. Subtraction from (1) proves (7). Every ϵ_j is centered with variance one, so $\mathbb{E}|\epsilon_j| \leq 1$. Since $|c_t| \leq t$, the cross term is at most $2t$ per selected row, which gives (8). No transformed-coordinate independence is used. $\square$

Proof of Theorem 1.1. The variables in (4) are iid, centered and standardized. Their law is an equally weighted mixture of two Gaussians with distinct means and positive variance, so its density is positive and smooth. Moreover,

$$\mathbb{E}e^{uY_i} = \cosh\left(\frac{u}{\sqrt{1+\eta^2}}\right) \exp\left(\frac{\eta^2 u^2}{2(1+\eta^2)}\right) \leq e^{u^2/2},$$

using $\cosh v \leq e^{v^2/2}$.

Fix a finest row. Its signal coefficient is zero and its noise coefficient is $(Y_{2k-1} - Y_{2k})/\sqrt{2}$. The event $\xi_{2k-1} = \xi_{2k}$ has probability $1/2$. Conditional on this event the coefficient is normal with mean zero and variance δ . Chebyshev's inequality therefore gives

$$\mathbb{P}\left(\left|\frac{Y_{2k-1} - Y_{2k}}{\sqrt{2}}\right| < t\right) \geq \frac{1}{2} \left(1 - \frac{\delta}{t^2}\right). \quad (9)$$

The inequality is still valid when its right side is negative. In (8), retain the $d/2$ finest rows and discard the other nonpositive probability terms. This gives

$$e_d \leq -\frac{d}{2} \left(1 - \frac{\delta}{t^2}\right) + 2t|S|,$$

which is (5). Finally $\eta = 1/20$ gives $\delta = 1/401$, and

$$2t - \frac{1}{2} + \frac{\delta}{2t^2} = -\frac{1}{4} + \frac{32}{401} = -\frac{273}{1604} \quad (t = 1/8).$$

□

3 An exact scale-series limit

At the zero signal the whole bias, not just an upper bound, admits a simple expression. For a dyadic block length m , let $K_m \sim \text{Bin}(m, 1/2)$ and let $G \sim N(0, 1)$ be independent. Define

$$V_m = \frac{2K_m - m}{\sqrt{m(1+\eta^2)}} + \sqrt{\delta}G, \quad b_m = -2\mathbb{P}(|V_m| < t) + 2\mathbb{E}[V_m c_t(V_m)]. \quad (10)$$

These are one-dimensional marginal quantities, not a specification of an independent family of wavelet coefficients.

Theorem 3.1. *Under (4), take $\theta = 0$ and threshold all detail rows but not the constant row. For $d = 2^J$,*

$$\frac{e_{2^J}}{2^J} = \sum_{\ell=1}^J 2^{-\ell} b_{2^\ell}. \quad (11)$$

For $\eta = 1/20$ and $t = 1/8$, the series $L = \sum_{\ell \geq 1} 2^{-\ell} b_{2^\ell}$ converges absolutely and

$$\left|\frac{e_d}{d} - L\right| \leq \frac{2}{d}, \quad L \leq -\frac{273}{1604} < 0. \quad (12)$$

If the constant row is also thresholded, the limit is the same and the error bound furnished by this argument is $4/d$.

Proof. A detail coefficient on a block of length m is a signed normalized sum of m noise coordinates. Sign changes preserve the joint law of the independent fair signs, and the normalized sum of the Gaussian parts is standard normal and independent of those signs. The coefficient thus has the law V_m in (10). There are d/m detail coefficients at that scale. Linearity of expectation in (7) proves (11), without any independence assumption across scales or rows.

Since $\mathbb{E}V_m^2 = 1$ and $0 \leq V_m c_t(V_m) \leq t|V_m|$,

$$-2 \leq b_m \leq 2t.$$

For $t = 1/8$ this gives $|b_m| \leq 2$. The geometric tail is at most $2 \sum_{\ell > J} 2^{-\ell} = 2/d$, proving absolute convergence and the first part of (12). Theorem 1.1, applied for every J , gives the negative upper bound on L . The constant coefficient has the same marginal law V_d ; thresholding it adds b_d/d to (11), of absolute value at most $2/d$. $\square$

Each b_m is a finite binomial mixture of Gaussian integrals. Thus (11) permits deterministic numerical evaluation as well as the analytic bound (12). The supplied program gives, for illustration, $e_{2048}/2048 \approx -0.5996787086$ in the details-only convention. This floating-point value is not a certified enclosure of L and is not needed for either theorem.

4 Scope and reproducibility

The example is not log-concave. Writing $a = (1 + \eta^2)^{-1/2}$, its density is a constant multiple of $\exp\{-y^2/(2\delta)\} \cosh(ay/\delta)$, so

$$(\log f)''(0) = -\frac{1}{\delta} + \frac{a^2}{\delta^2} = \frac{(1 + \eta^2)(1 - \eta^2)}{\eta^4} = 159999 \quad (\eta = 1/20).$$

Consequently the strong-log-concavity assumptions of [2, Theorem 4.1] are not met. Our fixed-threshold result does not assert failure of every SURE-calibrated threshold, disprove a minimax wavelet theorem, or answer all conditional questions in [1]. Gaussian-mixture-specific risk estimators have also been studied [3]; here the statistic is explicitly the unmodified Gaussian formula (1).

The obstruction is localization: a positive proportion of Haar rows always involve only two original coordinates. These coefficients do not become Gaussian merely because the total dimension grows. Dependence between transformed coordinates is not, by itself, the argument. The proof does not extend to arbitrary increasingly delocalized orthogonal transforms.

The accompanying Python program uses exact arithmetic in $\mathbb{Q}(\sqrt{2})$ to check Haar normalization and 1,104 pointwise risk identities for both threshold conventions and zero and nonzero pairwise-constant signals. Normal and optimized Python runs agree. Separate Gaussian-moment calculations and piecewise quadrature provide numerical illustrations only. The all-dimensional proofs are the arguments above, not an extrapolation from finite tests. The checks also retain the nonunit-variance failure of an unscaled SURE formula and a threshold atom for unsmoothed Rademacher noise.

AI tools assisted with derivation, literature search, code and manuscript preparation. This preprint is self-audited and unrefereed; the computational checks are neither independent human peer review nor formal verification. The general SURE framework is credited to the cited authors, and no absolute-priority claim is made for this explicit construction.

References

- [1] L. Goldstein, Problem 1.4, *Stein's method and applications in high-dimensional statistics*, American Institute of Mathematics problem list, <http://aimpl.org/steinhd/1/> (accessed 11 October 2026).
- [2] M. Fathi, L. Goldstein, G. Reinert and A. Saumard, *Relaxing the Gaussian assumption in Shrinkage and SURE in high dimension*, arXiv:2004.01378v3, 29 March 2022, <https://arxiv.org/abs/2004.01378v3>.
- [3] J. Sadasivan and C. S. Seelamantula, *An unbiased risk estimator for Gaussian mixture noise distributions—Application to speech denoising*, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), <https://doi.org/10.1109/ICASSP.2016.7472531>.