COMPLEX PHASE RETRIEVAL VIA CONDITIONAL QUANTILES: FINITE-SAMPLE RECOVERY WITHOUT NOISE MOMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Phase retrieval becomes statistically fragile when response noise is heavy-tailed or asymmetric: mean-based objectives can be dominated by extreme observations or even lose their population interpretation, while median-based methods fix attention on a single response quantile. We instead formulate complex phase retrieval through a prescribed conditional quantile,

$$Q_\tau(Y \mid a) = |a^* x_\star|^2,$$

and propose robust quantile Wirtinger flow (RQWF). RQWF combines a clipped spectral estimate of the signal direction, quantile-based radial calibration, and bounded-score Wirtinger refinement, so that initialization and optimization target the same response functional. Under complex Gaussian sensing, we prove end-to-end finite-sample recovery assuming only local regularity of the noise distribution around the target quantile, with no noise-moment requirement. The resulting recovery rate is $\widetilde{O}\big(\sqrt{\frac{n}{N}}\big)$, and a shifted-Cauchy submodel gives a matching dimension–sample lower bound up to logarithmic factors. The analysis also quantifies the bias introduced by smoothing and allows the same update sample to be reused throughout optimization. On a compute-matched Cauchy benchmark, RQWF-P lowers median recovery error relative to every tested complete SRPR/AdaIPL baseline in all 12 conditions and relative to Median-TWF in 10 of 12; prescribed non-median fits also improve selected asymmetric-noise cases. Student-t gains depend on the comparator, while Gaussian and near-noiseless settings expose limits of the advantage. Together, the results show that conditional quantiles provide a tractable statistical target for complex phase retrieval precisely in regimes where moment-based formulations become unreliable.

1 INTRODUCTION

Diffraction measurements record intensity but lose phase. Recovering a complex signal from these measurements is central to coherent imaging and related inverse problems (Fannjiang & Strohmer, 2020). Complex phase retrieval models this task as recovering $x_\star \in \mathbb{C}^n$ from measurements associated with $|a^* x_\star|^2$, up to a global phase. Here a^* denotes the conjugate transpose of the sensing vector a. Spectral initialization and first-order refinement provide efficient recovery under suitable sensing conditions (Candes et al., 2015; Chen & Candes, 2015). Dense asymmetric or heavy-tailed response noise raises a further question: which feature of the response distribution identifies the signal?

We study a prescribed conditional response quantile:

$$Q_\tau(Y \mid a) = |a^* x_\star|^2, \qquad \tau \in (0,1). \tag{1}$$

This target remains meaningful when the response has no finite mean. Unlike median fitting, it also accommodates a specified non-median quantile. The choice of τ is part of the statistical model; changing it generally changes the target. It is specified before observing reconstruction errors, not selected afterwards to favor a method.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our main result establishes finite-sample recovery of this target without noise-moment assumptions. Under independent complex Gaussian sensing and independent response noise with a positive, locally Lipschitz density near zero, Robust Quantile Wirtinger Flow (RQWF) attains

$$d(\widehat{x}, x_\star) = O_{\mathbb{P}}\left(\frac{1}{\kappa}\sqrt{\frac{n\log N}{N}}\right), \qquad \frac{n\log N}{N} \longrightarrow 0,$$

with fixed model parameters. Here d measures error up to global phase, κ lower-bounds the noise density at the target quantile, and N counts all samples used for initialization and refinement. The noise law is unrestricted outside the regularity neighborhood and may contain atoms there. A shifted-Cauchy submodel yields an $\Omega(\sqrt{n/N})$ lower bound at fixed noise scale. The bounds therefore match in their dimension–sample dependence up to logarithmic factors on that submodel.

RQWF estimates the signal direction from clipped responses and calibrates its radius through a scalar quantile fit on an independent sample. It then refines the estimate using a bounded residual score. The initialization avoids a mean-based scale formula, while the refinement limits the contribution of extreme residuals. Section 3 gives the complete algorithm; Appendix A.1 records its specification and numerical scope.

Establishing this guarantee requires control of initialization, calibration, and smoothing. We show that response clipping preserves a positive spectral gap without noise moments and that direction error enters radial calibration at second order. Pseudo-Huber smoothing shifts the population estimating equation, generally introducing bias of order $O(\delta^2 \log(1/\delta))$. A uniform local gradient bound controls sampling error even when refinement reuses the same observations.

Relative to existing quantile-loss methods for real sparse phase retrieval (Shen et al., 2022) and finite-moment theory for complex heavy-tailed recovery (Huang et al., 2025), our result establishes recovery of a prescribed complex response-quantile target without noise moments. The guarantee covers fixed and growing dense dimensions under local density regularity. A proposed sparse extension and its unresolved proof requirements appear separately in Appendix B.3.

Experiments examine heavy-tailed, Gaussian, low-sampling, and asymmetric settings. They show both favorable and unfavorable regimes for empirical RQWF variants. Section 5 separates a structure-aligned RQWF-T implementation from the practical RQWF-P solver through a componentwise bridge and complete-pipeline, matched-operator-budget comparisons. The results support Cauchy robustness and conditional benefits of non-median fitting, while exposing low-sampling and optimization costs. Section 6 adds held-out intensity prediction on real optical measurements with injected noise. This separate practical study does not certify theorem-admissible tuning or complex-signal recovery.

2 Related Work

Classical approaches recover phase through convex formulations such as PhaseLift, PhaseCut, and PhaseMax (Candes et al., 2013; Waldspurger et al., 2015; Goldstein & Studer, 2017), or through initialized first-order methods (Candes et al., 2015; Chen & Candes, 2015; Wang et al., 2017). Landscape analysis provides another route under clean complex Gaussian sensing (Sun et al., 2018). These results establish recovery mechanisms for the quadratic measurement map; our analysis must additionally account for a quantile-defined target and moment-free response noise.

Robust methods modify the objective or the estimating procedure. Median-TWF uses medians for truncation (Zhang et al., 2018), while absolute-loss and smoothed-absolute-loss formulations admit a median interpretation at the population level (Duchi & Ruan, 2019; Zheng & Xue, 2024). Other approaches use ℓ_p penalties, inexact proximal updates, or robust alternating minimization (Qian et al., 2017; Zheng et al., 2024b; Kim & Lee, 2025). Residual quantiles can also determine step sizes (Zheng et al., 2024a). Robust-gradient methods address heavy tails together with adversarial contamination (Buna & Rebeschini, 2025); robust-PCA techniques provide computationally tractable spectral initialization in this setting (Das & Batra, 2026). These contamination guarantees differ from recovery under our independent response-noise model.

Check loss directly targets a conditional response quantile (Koenker & Bassett, 1978). Smoothing theory for linear quantile regression (Tan et al., 2022; He et al., 2023) does not by itself resolve quadratic phase geometry or initialization. Shen et al. combine check loss with an $L_{1/2}$ penalty for

Table 1: Quantile roles and guarantee scope.

Method	Quantile role	Scope
Median-TWF	Median filtering	Outliers + bounded noise
Duchi–Ruan / SRPR	ℓ_1 / smoothed ℓ_1	Sparse-corruption guarantees
AdaIPL	Residual step-size	Local convergence; outliers
Shen et al.	Check loss	Real sparse; numerical evidence
Huang et al.	None (mean target)	Complex; finite $q > 2$ moment
RQWF	Prescribed $Q_\tau(Y \mid a)$	Complex; local density, no moments

real sparse phase retrieval (Shen et al., 2022). Their numerical results establish a relevant precursor, but do not supply the complex end-to-end guarantee considered here. Thresholded WF obtains sparse statistical rates under sub-exponential noise (Cai et al., 2016), while Huang et al. analyze complex least-squares recovery under finite $q > 2$ noise moments (Huang et al., 2025). RQWF instead estimates a prescribed response quantile with no noise-moment requirement.

Table 1 separates quantile-based robustification from quantile-defined estimation. Our distinction is not check loss alone, but a complex, moment-free guarantee joining initialization, smoothing bias, and statistical recovery. Appendix C compares the precise models, sample requirements, and optimization guarantees; these are not uniform dominance claims.

3 Statistical Model and RQWF

Notation. Write $[N] = \{1, \dots, N\}$, a^* for conjugate transpose, and $\|\cdot\|_2$ for Euclidean norm. For $J \subseteq [N]$, let $m_J = |J|$; $\nabla_{\bar{z}}$ denotes the Wirtinger derivative. The notation $\mathcal{CN}(0, I_n)$ denotes circular complex Gaussian sensing with covariance I_n. The parameters c, δ, η, T are the clipping level, smoothing bandwidth, step size, and update count, respectively. Recovery error is measured modulo global phase by (3).

RQWF estimates the signal at a prescribed conditional response quantile. For independent observations, write

$$Y_i = |a_i^* x_\star|^2 + \varepsilon_i, \qquad Q_\tau(\varepsilon_i \mid a_i) = 0, \qquad \tau \in (0, 1). \tag{2}$$

By translation equivariance, this is the conditional-quantile model in (1). The target is a conditional median when $\tau = 1/2$ and a non-median response quantile otherwise. Both cases permit asymmetric noise; neither requires finite noise moments. The recovery theorem additionally requires identically distributed additive errors independent of all sensing vectors (Assumption 1); the conditional-quantile identity alone does not imply that guarantee.

Because the measurements are invariant to global phase, we use

$$d(z, x_\star) = \min_{\theta \in [0, 2\pi)} \|z - e^{\mathrm{i}\theta} x_\star\|_2. \tag{3}$$

Section 4 imposes the sensing and local density conditions needed for recovery. These conditions are stronger than the quantile model alone.

For the residual $u_i(z) = Y_i - |a_i^* z|^2$, define

$$\rho_{\tau,\delta}(u) = (\tau - \tfrac{1}{2})u + \tfrac{1}{2}\sqrt{u^2 + \delta^2}, \qquad \delta > 0, \tag{4}$$

$$\psi_{\tau,\delta}(u) = \rho'_{\tau,\delta}(u) = \tau - \tfrac{1}{2} + \frac{u}{2\sqrt{u^2 + \delta^2}}, \qquad |\psi_{\tau,\delta}(u)| \le \max(\tau, 1 - \tau). \tag{5}$$

The score is bounded in the response residual. The loss is convex in that residual, but its composition with the quadratic intensity map is generally nonconvex in z. Smoothing introduces a population bias, which is included in the recovery bound rather than treated as exact quantile calibration.

Before using the observations, partition their indices $\{1, \dots, N\}$ into disjoint sets D, V, U, independently of the data values. Write $m_J = |J|$ for $J = D, V, U$, so that $N = m_D + m_V + m_U$. The observations are independent, hence the three resulting data blocks are independent. The sets D

and V are used only for initialization; U is reserved for refinement. The theorem permits any fixed positive split proportions satisfying its stated size conditions, rather than requiring equal thirds.

On D, clip only the response weights and take a unit leading eigenvector v_0 of the resulting weighted covariance. This estimates direction without requiring response moments, but the unit norm of v_0 does not estimate signal energy. On the independent block V, fit the squared radius s by a convex scalar quantile loss. Radial calibration and refinement use the original signed responses, not clipped responses or amplitudes. Both initialization estimates are independent of U; no new data are drawn during refinement. Algorithm 1 gives the full steps.

On the refinement sample, let

$$F_{\tau,\delta}(z) = \frac{1}{m_U} \sum_{i \in U} \rho_{\tau,\delta}(Y_i - |a_i^* z|^2), \qquad m_U = |U|.$$

Its Wirtinger derivative is

$$G_{m_U}(z) = \nabla_{\bar{z}} F_{\tau,\delta}(z) = -\frac{1}{m_U} \sum_{i \in U} \psi_{\tau,\delta}(Y_i - |a_i^* z|^2) a_i a_i^* z. \tag{6}$$

The update is

$$z_{t+1} = z_t - \eta G_{m_U}(z_t), \qquad \widehat{x} = z_T. \tag{7}$$

Algorithm 1 Theoretical RQWF (RQWF-T)

Input: Data $\{(a_i, Y_i)\}_{i=1}^N$; data-independent splits D, V, U; target τ; fixed $c, \delta, \eta > 0$ and update count T.

1. **Direction.** Set $w_c(y) = \min\{\max(y, 0), c\}$ and compute a unit leading eigenvector v_0 of $\widehat{M}_D = m_D^{-1} \sum_{i \in D} w_c(Y_i) a_i a_i^*$.
2. **Radius.** With v_0 fixed, solve

$$\widehat{s} \in \arg\min_{s \geq 0} \frac{1}{m_V} \sum_{i \in V} \rho_{\tau,\delta}(Y_i - s|a_i^* v_0|^2), \qquad z_0 = \sqrt{\widehat{s}}\, v_0. \tag{8}$$

3. **Refinement.** For $t = 0, \ldots, T-1$, apply (7) with G_{m_U} from (6) and the same U. Return $\widehat{x} = z_T$.

The same sample U is reused across iterations. Although the residual score is bounded, the Gaussian design factor $a_i a_i^* z$ remains unbounded. Under the real Euclidean identification of $\mathbb{C}^n$, the gradient is $2G_{m_U}$; this fixes our step-size convention. There is no projection of the iterates onto the unit sphere.

A dense update costs $O(m_U n)$ arithmetic. The theory covers fixed clipping and bandwidth, independent initialization samples, and ungated refinement. Appendix B specifies numerical tolerances; Appendix F describes the empirical variants. We write RQWF-T for this fixed-parameter theoretical specification and RQWF-P for empirical variants; these labels do not imply that an implementation satisfies the theorem's tuning or sample-size conditions.

4 Finite-Sample Recovery Theory

Algorithm 1 achieves a $\widetilde{O}_{\mathbb{P}}(\sqrt{n/N})$ phase-error rate under local density regularity, without noise moments. The main theorem counts all three splits and separates optimization, sampling, and smoothing errors.

Assumption 1. The pairs are independent, $a_i \sim \mathcal{CN}(0, I_n)$, $\|x_\star\|_2 = 1$, and the identically distributed errors are independent of all sensing vectors. Their cdf satisfies $F(0) = \tau$. For fixed $R > 0$, the noise law has a density on a neighborhood of $[-R, R]$, with $f(0) \geq \kappa > 0$, $0 \leq f \leq M$, and $|f(t) - f(s)| \leq \ell |t - s|$ on $[-R, R]$, where $\ell > 0$. Outside this neighborhood, the noise law is unrestricted and may even contain atoms. No noise moment or tail regularity is required.

Write $M_R = M + 2/R$, $\ell_R = \ell + 4M/R + 12/R^2$ and $r_R = \min\{1/2, \kappa/(432\ell_R)\}$. These constants translate local density information into bounds on the population score; their derivation is in Appendix A.5.

Theorem 2. *Under Assumption 1, fix $c > 0$ and split fractions $m_J \geq pN$ for $J = D, V, U$, where $0 < p \leq 1/3$ and $N = m_D + m_V + m_U$. There is $C_\star$, depending only on κ, M, ℓ, R, c, p, such that for $0 < \zeta < 1/4$ and $N \geq C_\star n \log(N/\zeta)$, Algorithm 1 with $\delta = N^{-1/2}$ and $0 < \eta \leq \kappa/(1024 M_R^2)$ satisfies, with probability at least $1 - \zeta$, simultaneously for all $T \geq 0$,*

$$d(z_T, x_\star) \leq r_R(1 - \eta\kappa/4)^{T/2} + \frac{C}{\kappa}\left[\sqrt{\frac{n\log(N/\zeta)}{N}} + \frac{n\log(N/\zeta)}{N} + \frac{\ell \log(3R\sqrt{N}) + R^{-2}}{N}\right]. \tag{9}$$

Here C depends only on the fixed split fractions. The choice of $C_\star$ also ensures $\delta \leq \min\{1, R/2, \kappa/(4\ell)\}$.

This is the local-density assembly result proved in Appendices A.4 and A.5. The technical statements in Appendix A.1 give explicit sufficient conditions. The sample threshold $C_\star$ can deteriorate with the density bounds, clipping level, and initialization spectral gap. The admissible step size uses declared density regularity bounds; this is not an adaptive tuning rule for an unknown noise law. For fixed admissible η, the deterministic schedule

$$T_N = \left\lceil \frac{4\log[N \max\{1, \kappa\}^2]}{\eta\kappa} \right\rceil$$

makes the optimization term at most $1/(\kappa\sqrt{N})$. Certified numerical approximations retain the rate (Appendix B).

Smoothing bias. Let $F_\delta = F * K_\delta$ with $K_\delta(u) = \delta^2/[2(u^2 + \delta^2)^{3/2}]$. Then the population score at the signal is $\mathbb{E}G_{m_U}(x_\star) = \beta_\delta x_\star$, where $\beta_\delta = F_\delta(0) - \tau$ and

$$|\beta_\delta| \leq \frac{\ell\delta^2}{2}\log(3R/\delta) + \frac{\delta^2}{4R^2}.$$

This generally nonzero bias explains the last term in (9). If the density is C^2 near zero, its leading term is $\frac{1}{2}f'(0)\delta^2\log(1/\delta)$; symmetric median noise cancels it exactly. Appendix A.2 supplies the centered-risk identification and sharp expansion.

Dimension regimes. Fix the noise regularity, clipping, and split fractions, and use $\widehat{x} = z_{T_N}$. For $n = O(1)$ and $N \to \infty$, Theorem 2 directly gives $d(\widehat{x}, x_\star) = O_{\mathbb{P}}(\sqrt{\log N/N})$. For $n = n_N \to \infty$ with $n \log N/N \to 0$, it gives

$$d(\widehat{x}, x_\star) = O_{\mathbb{P}}\left(\sqrt{\frac{n\log N}{N}}\right) = \widetilde{O}_{\mathbb{P}}\left(\sqrt{\frac{n}{N}}\right) = o_{\mathbb{P}}(1).$$

Both are consequences of the same finite-sample bound. The growing-dimensional regime permits increasing dense dimension with vanishing error, under a sample requirement $N \gtrsim n\log(N/\zeta)$.

Proposed sparse extension. Recovery with $N \ll n$ needs an estimator that uses sparsity. The present theorem concerns dense RQWF. Appendix B.3 gives a separate support-screening and restricted-RQWF proposal, its beta-min requirement, and the remaining proof steps. No sparse recovery or sparse lower-bound theorem is asserted.

Cauchy lower bound.

Theorem 3 (Cauchy submodel)**.** *Let $n \geq 33$, $0 < r_0 \leq 1/4$, and $\varepsilon = \sigma[C_0 - \tan\{\pi(\tau - 1/2)\}]$, where C_0 is standard Cauchy and $\sigma > 0$. For N independent measurements with complex Gaussian sensing and errors independent of sensing vectors, there is a universal $c_0 > 0$ such that*

$$\inf_{\widehat{x}} \sup_{\substack{\|x\|=1 \\ d(x,e_1)\leq r_0}} \mathbb{P}_x\left\{d(\widehat{x}, x) \geq c_0 \min\left(r_0, \sigma\sqrt{\frac{n-1}{N}}\right)\right\} \geq \frac{1}{2}. \tag{10}$$

The infimum includes all measurable estimators, even if σ and τ are known.

Here $f(0) = \sin^2(\pi\tau)/(\pi\sigma)$. For a fixed interior quantile, the stochastic dimension–sample dependence in (9) therefore matches this submodel's lower rate up to logarithms. The matching statement concerns this submodel and dimension–sample scaling, with fixed quantile and noise scale.

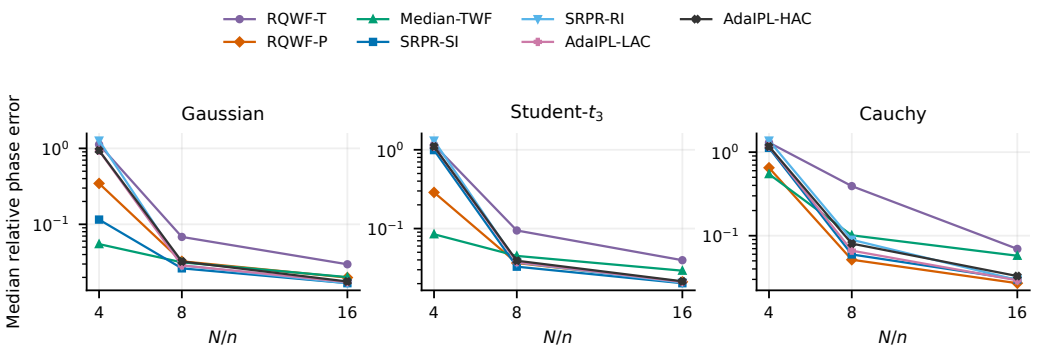


Figure 1: Complete-pipeline complex baselines at $n = 256$, $\sigma = 0.05$, and matrix-product budget 2,400. Points are medians over 20 paired problems. All 42 cell medians and further uncertainty/budget plots are retained in Appendix D.1; this display is one declared slice.

5 Experiments

5.1 Fixed-scale protocol

Our primary matched-budget study contains 840 paired synthetic problems (42 cells, 20 problems per cell), 15 methods/variants, and 12,600 fits. RQWF-T follows the independent $D/V/U$ structure with fractions $0.4/0.2/0.4$, fixed clipping $c = 4$, common $\delta = N^{-1/2}$, and ungated refinement on the same U. RQWF-P reuses all data with empirical clipping, gates, and annealing. RQWF-T's experimental step $\eta = 0.02$ is not certified to satisfy Theorem 2; structural correspondence is not theorem validation. Sensing vectors are independent $\mathcal{CN}(0, I_n)$ and signals have unit norm. Responses retain their sign and noise has fixed population scale, without sample-SNR normalization. We report relative phase error

$$\operatorname{dist}(z, x_\star) = d(z, x_\star)/\|x_\star\|_2, \tag{11}$$

where d is defined in (3). The grid uses $n \in \{128, 256\}$, $N/n \in \{4, 8, 16\}$, and $\varepsilon = \sigma Z$ for Gaussian, Student-t_3, or standard Cauchy Z, with $\sigma \in \{0.05, 0.20\}$, $\tau = 1/2$, and six noiseless controls. Equal scale does not imply equal variance or difficulty. The primary endpoint is a predeclared budget of 2,400 full-data matrix-vector product equivalents; 600 and 1,200 are secondary checkpoints. Initialization, Gram construction, line searches, and IPL inner solves count toward this budget, but it is not equal FLOPs or wall time. There is no truth-based stopping or best-checkpoint selection (Appendix D.1).

5.2 Complete-pipeline baselines and robustness

We implement complex extensions of SRPR (Zheng & Xue, 2024) with spectral (SI) and random (RI) initialization, both with smooth MAPG followed by nonsmooth IPL, and AdaIPL (Zheng et al., 2024a) with LAC and HAC inner stopping. These are paper-based reimplementations, not official-code reproductions. Their real, sparse-corruption theory is not transferred to the present complex, dense-noise setting. Every accepted IPL step satisfies a recorded primal–dual gap criterion; unfinished inner work consumes budget without changing the returned outer iterate. All 840 problems reach at least one accepted IPL step by the primary endpoint for each of these four baselines. Median-TWF provides an additional reference.

RQWF-P has lower cell-median error than each SRPR/AdaIPL version in all 12 Cauchy cells, and than Median-TWF in 10/12. These are descriptive cell counts, not significance claims. Its advantage is not universal: against SRPR-SI it wins only 2/12 Gaussian and 2/12 Student-t_3 cells. At $n = 256$, $N/n = 8$, Cauchy scale 0.05, median errors are 0.0516 (P), 0.0597 (SRPR-SI), and 0.0662 (AdaIPL-LAC). Low sampling remains difficult: at $n = 256$, $N/n = 4$ without noise, P has median error 0.439 while Median-TWF reaches 6.93×10^{-14}. All 12,600 fits are finite; this does not imply accurate recovery.

Table 2: Cumulative T-to-P changes at a matrix-product budget of 2,400. Each row compares with the preceding variant, starting from T. Counts are out of 42 settings, with median recovery error computed over 20 paired trials per setting. The last column requires the pointwise 95% paired-bootstrap CI for the median difference (after minus before) to lie entirely below zero; no multiple-comparison adjustment is applied.

Successive change	Lower median	CI below zero
Reuse all data	42/42	42/42
Empirical clipping	21/42	1/42
Median radius	18/42	1/42
50 power steps	13/42	2/42
P effective step	26/42	8/42
Add gates	10/42	7/42
Fixed scaled bandwidth	34/42	27/42
Add annealing (P)	9/42	6/42

Table 3: Matched versus median fitting: median relative phase errors and matched-fit wins out of 50 paired problems.

Noise	τ	N/n	RQWF-P		RQWF-T	
			Matched / median	Wins	Matched / median	Wins
ALD	0.25	8	0.2038 / 0.2507	50/50	0.6811 / 0.8131	50/50
ALD	0.25	32	0.0561 / 0.0891	50/50	0.1095 / 0.1534	50/50
ALD	0.75	8	0.1664 / 0.1959	50/50	0.6880 / 0.4405	0/50
ALD	0.75	32	0.0559 / 0.0846	50/50	0.1063 / 0.1333	50/50
Lognormal	0.25	8	0.0407 / 0.0624	50/50	0.4373 / 0.4921	45/50
Lognormal	0.25	32	0.0126 / 0.0254	50/50	0.0254 / 0.0441	50/50
Lognormal	0.75	8	0.1117 / 0.0788	0/50	0.6979 / 0.2744	0/50
Lognormal	0.75	32	0.0391 / 0.0376	6/50	0.0770 / 0.0551	0/50

5.3 Bridging the theoretical and practical implementations

Starting from T, we successively change data reuse, clipping, radial scale, eigensolver, effective step, gates, fixed bandwidth, and annealing to obtain P. Each edge changes one declared component; P without gates is an additional control. At budget 2,400, reusing all observations reduces median error in all 42 cells. Adding gates improves 10/42, and adding annealing improves 9/42; thus these refinements are not uniformly beneficial (Table 2). Effects depend on the path: reuse also changes affordable update counts, and annealing changes both smoothing and effective step. The bridge explains an empirical gap; it does not extend the split-sample theorem to P.

5.4 Prescribed non-median targets

An independently frozen 4,500-problem suite (Appendix D.3) includes asymmetric-Laplace (ALD) and shifted-lognormal noise with known zero τ-quantile, $\tau \in \{0.25, 0.75\}$, at fixed scale 0.10. For $n = 256$ and $N/n \in \{8, 32\}$, matched and median fits share data and initialization randomness and use the predeclared 1,000-update endpoint. P's other tuning is identical between fits; T's radial fit changes with τ. Table 3 reports every cell for both versions.

For P, matching improves all 50 pairs in every ALD and lower-target lognormal cell, but has larger median error in both upper-target lognormal cells. For the latter case at ratio 32, matched fitting nevertheless improves test coverage (0.720 versus 0.619, target 0.75) and centered target check loss (0.00304 versus 0.00833). Correctly specifying the target does not guarantee smaller finite-budget phase error. Sample-size, density, small-noise, and optimization diagnostics remain in Appendix D.3. Its earlier budget excludes Gram work and is not pooled with the new study.

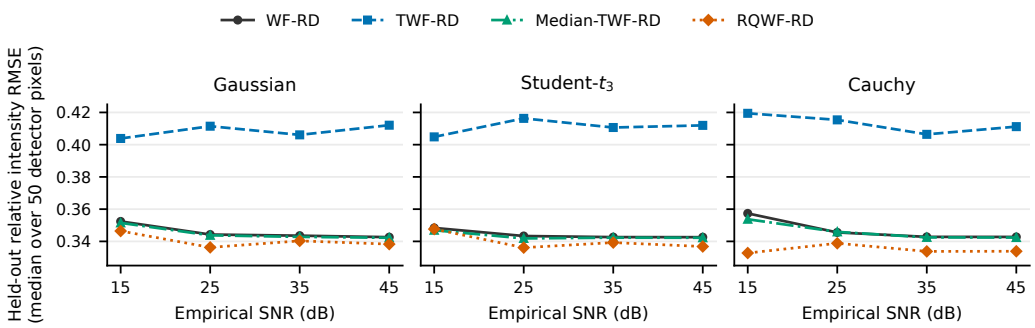


Figure 2: Held-out prediction on real optical measurements with injected training noise. Each curve summarizes 50 fixed detector pixels. The suffix RD denotes the dataset-specific archived implementation, not the synthetic T/P solver or an official baseline. These runs are not compute-matched.

6 Practical Validation

We evaluate recorded optical measurements from the Rice transmission-matrix dataset (Metzler et al., 2017), using its amplitude-SLM 16×16 setting. Each of 50 fixed detector pixels defines a prediction task with 256 input coordinates, 2,048 training patterns, and 1,024 held-out patterns. Gaussian, Student-t_3, or Cauchy draws are added to training intensities and rescaled to empirical SNR 15, 25, 35, or 45 dB. Held-out recorded intensities receive no additional noise; they are not noise-free truth. The endpoint is relative intensity RMSE, aggregated by the median across pixels, because true complex transmission vectors are unknown.

RQWF-RD has lower median prediction error than WF-RD and TWF-RD in all 12 conditions, and than Median-TWF-RD in 11/12. At Cauchy 15 dB, the medians are 0.3328, 0.3574, 0.4195, and 0.3537, respectively; the reduction relative to WF-RD is 6.9% (49/50 paired pixels improve). This is descriptive evidence for robustness to injected noise on measured data, not a theorem test or a calibrated non-median target experiment. The recipes differ from the synthetic implementations, use unequal iteration ceilings and early termination, and remain real-valued from real initialization. Accordingly, they assess restricted intensity prediction, not identification of a general complex transmission vector. Appendix E gives all results, algorithm definitions, data checks, and these limitations. No SRPR/AdaIPL comparison is claimed on this dataset.

7 Conclusion

Conditional response quantiles make complex phase retrieval well defined without response-noise moments. RQWF combines clipped directional initialization, quantile radial calibration, and bounded-score refinement. Under independent Gaussian sensing and local density regularity, its finite-sample bound covers fixed and growing dense dimensions. A shifted-Cauchy lower bound matches the dimension–sample dependence up to logarithms. Sparse recovery and density-adaptive tuning remain extensions. Fixed-scale experiments support Cauchy robustness of RQWF-P and show that prescribed non-median targets can improve recovery, but not uniformly. They also expose a gap between the practical solver and split-sample RQWF-T at moderate sampling ratios. A componentwise bridge attributes consistent gains to data reuse, but not to every practical refinement. The optical-measurement study adds held-out prediction evidence under injected noise, with a separate implementation and statistical scope. These observations motivate sharper initialization and tuning analysis, rather than a universal ranking.

Reproducibility Statement

The appendices provide proofs, numerical-accuracy conditions, and source-to-result audit details. The new 840-problem bridge/baseline suite stores frozen code/configuration, shared-data hashes, checkpoint estimates, work counts, and 18,437 accepted IPL certificates (Appendix D.1). The

independent 4,500-problem suite and its smooth-only SRPR-s comparator remain separately identified in Appendix D.3. For the measurement study, archived R code, 2,400 fit records, and 273,985 trace rows reproduce the supplied summary; rerunning the fits additionally requires the public MAT measurement files (Appendix E). Historical archives and auxiliary linear-model plots remain explicitly separate from the current comparisons.

AI USE STATEMENT

Generative AI assisted literature checking, mathematical derivation and proof drafting, experimental design, numerical code implementation and testing, synthetic-data simulation and result analysis, figure and table preparation, and manuscript editing. Exact Gaussian-moment calculations and deterministic numerical checks audited selected proof identities; they are not an independent certification of the proofs. Numerical tables and result curves derive from recorded simulation outputs and optical-measurement prediction logs, including the new matched-budget suite and separately identified archives. Implementation unit checks are not treated as statistical recovery evidence.

REFERENCES

Alex Buna and Patrick Rebeschini. Robust gradient descent for phase retrieval. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 2080–2088, 2025.

T. Tony Cai, Xiaodong Li, and Zongming Ma. Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow. *The Annals of Statistics*, 44(5):2221–2251, 2016. doi: 10.1214/16-AOS1443. URL https://arxiv.org/abs/1506.03382.

Emmanuel J. Candes, Thomas Strohmer, and Vladislav Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013. doi: 10.1002/cpa.21432. URL https://arxiv.org/abs/1109.4499.

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Yuxin Chen and Emmanuel Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Advances in Neural Information Processing Systems*, 28, 2015.

Frédéric Chyzak and Frank Nielsen. A closed-form formula for the Kullback–Leibler divergence between Cauchy distributions. *arXiv preprint arXiv:1905.10965*, 2019. URL https://arxiv.org/abs/1905.10965.

Santanu Das and Jatin Batra. Tractable gaussian phase retrieval with heavy tails and adversarial corruption with near-linear sample complexity. *arXiv preprint arXiv:2601.18245*, 2026. URL https://arxiv.org/abs/2601.18245.

John C. Duchi and Feng Ruan. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. *Information and Inference: A Journal of the IMA*, 8(3):471–529, 2019. doi: 10.1093/imaiai/iay015.

Albert Fannjiang and Thomas Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29: 125–228, 2020.

Tom Goldstein and Christoph Studer. Convex phase retrieval without lifting via phasemax. In *International Conference on Machine Learning*, pp. 1273–1281. PMLR, 2017.

Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023. doi: 10.1016/j.jeconom.2021.07.010.

Gao Huang, Song Li, and Deanna Needell. Stable phase retrieval: Optimal rates in poisson and heavy-tailed models. *arXiv preprint arXiv:2510.00551*, 2025. URL https://arxiv.org/abs/2510.00551.

Seonho Kim and Kiryung Lee. Robust phase retrieval by alternating minimization. *IEEE Transactions on Signal Processing*, 73:40–54, 2025. doi: 10.1109/TSP.2024.3515008.

Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. doi: 10.2307/1913643.

Christopher A. Metzler, Manoj K. Sharma, Sudarshan Nagesh, Richard G. Baraniuk, Oliver Cossairt, and Ashok Veeraraghavan. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *IEEE International Conference on Computational Photography (ICCP)*, 2017. doi: 10.1109/ICCPHOT.2017.7951483. URL https://dsp.rice.edu/research/transmissionmatrices/.

Cheng Qian, Xiao Fu, Nicholas D. Sidiropoulos, Lei Huang, and Junhao Xie. Inexact alternating optimization for phase retrieval in the presence of outliers. *IEEE Transactions on Signal Processing*, 65(22):6069–6082, 2017. doi: 10.1109/TSP.2017.2740166.

Si Shen, Jiayao Xiang, Huijuan Lv, and Ailing Yan. $L_{1/2}$-regularized quantile method for sparse phase retrieval. *Open Journal of Applied Sciences*, 12(12):2135–2151, 2022. doi: 10.4236/ojapps.2022.1212147.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018. URL https://arxiv.org/abs/1602.06664.

Kean Ming Tan, Lan Wang, and Wen-Xin Zhou. High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society: Series B*, 84(1):205–233, 2022. doi: 10.1111/rssb.12485.

Irene Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149:47–81, 2015.

Gang Wang, Georgios Giannakis, Yousef Saad, and Jie Chen. Solving most systems of random quadratic equations. *Advances in Neural Information Processing Systems*, 30, 2017.

Huishuai Zhang, Yuejie Chi, and Yingbin Liang. Median-truncated nonconvex approach for phase retrieval with outliers. *IEEE Transactions on information Theory*, 64(11):7287–7310, 2018.

Zhong Zheng and Lingzhou Xue. Smoothed robust phase retrieval. *arXiv preprint arXiv:2409.01570*, 2024. URL https://arxiv.org/abs/2409.01570.

Zhong Zheng, Necdet Serhat Aybat, Shiqian Ma, and Lingzhou Xue. Adaptive algorithms for robust phase retrieval. *arXiv preprint arXiv:2409.19162*, 2024a. URL https://arxiv.org/abs/2409.19162.

Zhong Zheng, Shiqian Ma, and Lingzhou Xue. A new inexact proximal linear algorithm with adaptive stopping criteria for robust phase retrieval. *IEEE Transactions on Signal Processing*, 72:1081–1093, 2024b. doi: 10.1109/TSP.2024.3365933.

A Proofs

A.1 Auxiliary statements

Algorithm 1 in the main text defines the estimator analyzed in Theorem 2. The split is chosen independently of the data values. Its clipping level, bandwidth, and step size are fixed before refinement; D and V are used only for initialization, while the same independent update sample U is reused across iterations. The spectral eigenvector is defined only up to phase, which does not affect the scalar radius problem or phase-invariant recovery. Appendix B gives certified numerical tolerances for the eigenvector and scalar radial problems.

The main theorem uses only Assumption 1. This appendix retains a global-density reference version to organize the existing proof calculations, followed by the local-density transfer. Global regularity is a proof device here, not an additional hypothesis of Theorem 2.

Assumption 4 (Design, noise, and scale)**.** The pairs are independent, $a_i \sim \mathcal{CN}(0, I_n)$, and $y_i = |a_i^* x_\star|^2 + \varepsilon_i$, where $\|x_\star\|_2 = 1$. The identically distributed errors are independent of the sensing vectors, $F(0) = \tau$, and their density satisfies

$$f(0) \geq \kappa > 0, \qquad 0 \leq f(t) \leq M, \qquad |f(t) - f(s)| \leq \ell |t - s| \quad (s, t \in \mathbb{R}),$$

for finite M and $\ell > 0$. No moment of ε_i is required.

The lower bound is at the target, not on the whole real line. Assumption 4 gives a convenient first set of constants; Theorem 8 subsequently replaces its global density conditions by local ones. Fixed-scale shifted Cauchy noise is admissible. Unit energy specifies the parameter class; Appendix B gives the conversion to other signal scales.

The centered check risk is well defined even if the uncentered risk is infinite. With $\Delta = |a^* z|^2 - |a^* x_\star|^2$ and $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$,

$$\mathcal{E}(z) = \mathbb{E}[\rho_\tau(\varepsilon - \Delta) - \rho_\tau(\varepsilon)] = \mathbb{E} \int_0^\Delta [F(t) - \tau]\, dt \geq 0. \tag{12}$$

Indeed the loss difference is bounded by $|\Delta|$, which is integrable. Assumption 4 makes zero the unique noise quantile; Gaussian intensity sensing then identifies the phase orbit. For smoothing, let $F_\delta = F * K_\delta$, where $K_\delta(t) = \delta^2/[2(t^2 + \delta^2)^{3/2}]$, and put $\beta_\delta = F_\delta(0) - \tau$.

Proposition 5 (Bias and its sharp order)**.** *For $0 < \delta \leq 1$, define*

$$\bar{\beta}_\delta = \frac{\ell \delta^2}{2} \log(3/\delta) + \frac{\delta^2}{4}. \tag{13}$$

Then $|\beta_\delta| \leq \bar{\beta}_\delta$ and $\mathbb{E} G_m(x_\star) = \beta_\delta x_\star$. If, additionally, f is C^2 near zero, then $\beta_\delta = \frac{1}{2} f'(0) \delta^2 \log(1/\delta) + O(\delta^2)$. For symmetric noise and $\tau = 1/2$, $\beta_\delta = 0$ exactly.

Thus a generic $O(\delta^2)$ bias claim is false for this smoothing kernel. Appendix A.2 also shows a corresponding displacement of the population stationary point when $f'(0) \neq 0$.

Throughout, $C \geq 1$ denotes a sufficiently large universal constant. All sample sizes are positive integers, $0 < \zeta < 1/4$, and iteration counts are nonnegative integers. Tuning parameters are deterministic. For an update sample of size m_U, set

$$\Lambda_U = n \log\left(\frac{C m_U [n + \log(m_U/\zeta)]^2}{\delta} \right) + \log(C/\zeta), \quad t_U = C\left(\sqrt{\Lambda_U / m_U} + \Lambda_U / m_U \right), \tag{14}$$

$$r = \min\{1/2, \kappa/(432\ell)\}, \qquad \mu = \kappa/4, \quad L = 8M, \quad b = 2\bar{\beta}_\delta + t_U. \tag{15}$$

Theorem 6 (Finite-sample local recovery)**.** *Under Assumption 4, fix $0 < \zeta < 1/4$, $0 < \delta \leq \min\{1, \kappa/(2\ell)\}$ and $0 < \eta \leq \mu/(4L^2)$. With probability at least $1 - \zeta$ over U, simultaneously for every z_0 with $\max\{d(z_0, x_\star), \sqrt{3} b/\mu\} \leq r$, the ungated updates $z_{t+1} = z_t - \eta G_{m_U}(z_t)$ remain in that ball and, for every $T \geq 0$,*

$$d(z_T, x_\star) \leq (1 - \eta\mu)^{T/2} d(z_0, x_\star) + \frac{\sqrt{3}}{\mu}(2\bar{\beta}_\delta + t_U). \tag{16}$$

The same update data may be reused at every iteration.

The proof combines population phase-aligned curvature with a gradient deviation bound uniform on $\|z\| \leq 3/2$; it does not require an empirical Hessian bound independent of δ. For initialization, set $c = c_y$ and $S \sim \mathrm{Exp}(1)$. The population clipped spectral gap is

$$\gamma_c = \mathbb{E}\{S[F(c - S) - F(-S)]\} \geq \kappa b_c^3 e^{-2b_c} > 0, \qquad b_c = \min\{c/4, \kappa/(4\ell), 1\}. \tag{17}$$

Clipped matrix concentration controls the leading direction without a noise moment. Independent scalar calibration then has error $O(\kappa^{-1}[\bar{\beta}_\delta + (M + \ell)\sin^2 \angle(v_0, x_\star) + \sqrt{\log(1/\zeta)/m_V} + \log(1/\zeta)/m_V])$. The squared directional term is essential; treating the noisy marginal response quantile as a radius estimate is not sufficient.

Corollary 7 (Recovery from data, with all three splits counted)**.** *Fix the constants in Assumption 4, $c > 0$, and split fractions $m_J \geq pN$ for $J = D, V, U$, where $0 < p \leq 1/3$ and $N = m_D + m_V + m_U$. There is a constant $C_\star$, depending on these fixed quantities, such that if $N \geq C_\star n \log(N/\zeta)$, the spectral–radial initializer* (8)*, $\delta = N^{-1/2}$, and the steps of Theorem 6 satisfy, with probability at least $1 - \zeta$,*

$$d(z_T, x_\star) \leq r(1 - \eta\mu)^{T/2} + \frac{C}{\kappa}\left[\sqrt{\frac{n\log(N/\zeta)}{N}} + \frac{n\log(N/\zeta)}{N} + \frac{\ell\log(3\sqrt{N})+1}{N}\right]. \tag{18}$$

Here C may depend on the fixed split fractions.

Appendix A.4 gives explicit sufficient finite-sample inequalities, rather than hiding the initializer and basin conditions in $C_\star$. Constants can deteriorate with noise scale, density regularity, and clipping; this is not a uniform low-SNR or low-density guarantee. For a target optimization tolerance $\epsilon_{\mathrm{opt}} < r$, $T \geq 2(\eta\mu)^{-1}\log(r/\epsilon_{\mathrm{opt}})$ suffices for the first term.

Theorem 8 (Local-density extension)**.** *Retain the design, independence, quantile, and unit-energy conditions of Assumption 4. Instead of its global density conditions, suppose the noise law has a density on a neighborhood of $[-R, R]$, $R > 0$, with $f(0) \geq \kappa$, $0 \leq f \leq M$, and $|f(t) - f(s)| \leq \ell|t - s|$ for $s, t \in [-R, R]$. Outside this interval its law is unrestricted, including possible atoms. Define*

$$M_R = M + 2/R, \qquad \ell_R = \ell + 4M/R + 12/R^2, \tag{19}$$

$$\bar\beta_{\delta,R} = \frac{\ell\delta^2}{2}\log(3R/\delta) + \frac{\delta^2}{4R^2}. \tag{20}$$

For $0 < \delta \leq \min\{1, R/2, \kappa/(4\ell)\}$, the conclusions of Theorems 6 and 15 hold with $M, \ell, \bar\beta_\delta$ replaced by $M_R, \ell_R, \bar\beta_{\delta,R}$ in their constants and sample conditions, and with their bandwidth restriction replaced by the one above. The spectral gap remains positive: in (17)*, use $b_{c,R} = \min\{c/4, R/2, \kappa/(4\ell), 1\}$ in place of b_c.*

For fixed local regularity and split fractions, this again gives the dimension–sample rate $\sqrt{n\log(N/\zeta)/N}$ with $\delta = N^{-1/2}$ and sufficiently large N. The proof in Appendix A.5 controls the scalar score at large shifts without assuming a global density bound. Its constants deteriorate as R shrinks; it is not a uniform guarantee for arbitrarily narrow regular neighborhoods.

A.2 Proofs of Proposition 5 and Lemma 10

All expectations in the proof concern the model in Assumption 4. Inner products are u^*v, and $\mathcal{G}_\delta(z) = \mathbb{E}G_m(z)$ denotes the population Wirtinger gradient, not the real gradient. The latter equals $2\mathcal{G}_\delta$.

For fixed a, the difference of check losses is the integral, from zero to Δ, of $\mathbf{1}\{\varepsilon \leq t\} - \tau$. The integrand is bounded, so conditional integration gives (12). Since $f(0) > 0$ and f is continuous, the integral is strictly positive for every $\Delta \neq 0$, including shifts outside the interval where a density lower bound is available. For Gaussian intensity sensing, $\Delta = a^*Ha$ with $H = zz^* - x_\star x_\star^*$ and $\mathbb{E}\Delta^2 = (\mathrm{tr}\, H)^2 + \mathrm{tr}(H^2)$. Thus zero excess risk implies $H = 0$, or equality of the phase orbits. The centered smoothed risk $\mathbb{E}[\rho_{\tau,\delta}(\varepsilon - \Delta) - \rho_{\tau,\delta}(\varepsilon)]$ is also finite. On a bounded parameter ball its derivative is dominated by a constant times $\|a\|^2$, so differentiation under expectation is valid, without $\mathbb{E}|\varepsilon| < \infty$.

The kernel K_δ integrates to one, has cdf $H_\delta(t) = 1/2 + t/(2\sqrt{t^2 + \delta^2})$, and $\int |t|K_\delta(t)\,dt = \delta$. Let

$$m_\delta(s) = F_\delta(s) - \tau, \qquad f_\delta = f * K_\delta.$$

Symmetry of K_δ and the residual convention imply $\mathbb{E}_\varepsilon[-\psi_{\tau,\delta}(\varepsilon - s)] = m_\delta(s)$. Consequently

$$\mathcal{G}_\delta(z) = \mathbb{E}_a[m_\delta(\Delta)aa^*z], \quad \mathcal{G}_\delta(x_\star) = \beta_\delta x_\star, \quad 0 \leq f_\delta \leq M, \quad \mathrm{Lip}(f_\delta) \leq \ell, \quad f_\delta(0) \geq \kappa - \ell\delta. \tag{21}$$

The last inequality uses the kernel's first absolute moment, not a noise moment.

We now prove Proposition 5. For $u > 0$ put $R_\delta(u) = \{1 - u/\sqrt{u^2 + \delta^2}\}/2$. Then

$$\beta_\delta = \int_0^\infty R_\delta(u)[f(u) - f(-u)]\,du, \qquad 0 \leq R_\delta(u) \leq \min\{1/2, \delta^2/(4u^2)\}.$$

On $(0,1)$ use $|f(u)-f(-u)| \le 2\ell u$; on $(1,\infty)$ use $\int_1^\infty [f(u)+f(-u)]\,du \le 1$. Since

$$\int_0^1 uR_\delta(u)\,du = \frac{1-\sqrt{1+\delta^2}+\delta^2\,\mathrm{arsinh}(1/\delta)}{4} \le \frac{\delta^2}{4}\log(3/\delta) \quad (0<\delta\le 1),$$

the asserted upper bound follows. Symmetry makes the bias integral zero.

If f is C^2 near zero, fix a small constant $R>0$ such that $f(u)-f(-u)=2f'(0)u+O(u^2)$ for $0<u<R$. The error integral is $O(\delta^2)$ because $\int_0^R u^2 R_\delta(u)\,du = O(\delta^2)$. The tail above R is also $O(\delta^2)$. Finally $\int_0^R uR_\delta(u)\,du = \delta^2\log(1/\delta)/4 + O(\delta^2)$. This proves the expansion, including its coefficient.

Corollary 9 (Population displacement)**.** *Under the additional smoothness in Proposition 5, for all sufficiently small δ there is a stationary point $z_\delta = r_\delta x_\star$ near $x_\star$, with $r_\delta > 0$, and*

$$r_\delta - 1 = -\frac{f'(0)}{8f(0)}\delta^2\log(1/\delta) + O(\delta^2).$$

Thus the logarithmic bias can be a parameter displacement, not only an upper bound on the score.

Proof. For $S=|a^*x_\star|^2 \sim \mathrm{Exp}(1)$, define $H(s)=\mathbb{E}[m_\delta((s-1)S)S]$. Rotational invariance gives $\mathcal{G}_\delta(rx_\star) = rH(r^2)x_\star$. We have $H(1)=\beta_\delta$, $H'(1)=2f_\delta(0)$, and $|H'(s)-H'(1)| \le 6\ell|s-1|$. Hence H is strictly increasing on a fixed neighborhood of one for all small δ. The intermediate value theorem gives a unique root there, $s_\delta - 1 = O(|\beta_\delta|)$, and Taylor's formula gives $s_\delta - 1 = -\beta_\delta/[2f_\delta(0)] + O(\beta_\delta^2)$. Taking its positive square root and using $|f_\delta(0)-f(0)| \le \ell\delta$ proves the claim. No assertion about other stationary points is needed. □

Lemma 10 (Phase-aligned local geometry)**.** *For δ, r, μ, L in Theorem 6, let h be the difference between z and its nearest phase representative of $x_\star$ and let $e=\|h\| \le r$. Then*

$$\mathrm{Re}\langle \mathcal{G}_\delta(z), h\rangle \ge \mu e^2 - 2\bar\beta_\delta e, \qquad \|\mathcal{G}_\delta(z)\| \le Le + 2\bar\beta_\delta. \tag{22}$$

Proof. Phase equivariance permits writing $z = x_\star + h$ with $\alpha = x_\star^* h \in \mathbb{R}$. Put $u=a^*x_\star$, $v=a^*h$, $B=\mathrm{Re}(\bar u v)$ and $Q=|v|^2$. Then $\Delta = 2B+Q$ and $\mathrm{Re}\langle aa^*z, h\rangle = B+Q$. Gaussian fourth moments give

$$\mathbb{E}B^2 = \tfrac{1}{2}e^2 + \tfrac{3}{2}\alpha^2, \qquad \mathbb{E}BQ = 2\alpha e^2, \qquad \mathbb{E}Q^2 = 2e^4.$$

For example, write $v=\alpha u + w$, with independent proper complex Gaussian w of variance $e^2-\alpha^2$; the identities follow by expansion, $\mathbb{E}|u|^4=2$, and vanishing odd or unbalanced complex moments. In particular,

$$\mathbb{E}[\Delta(B+Q)] = e^2 + 3\alpha^2 + 6\alpha e^2 + 2e^4 = e^2 - e^4 + 3(\alpha+e^2)^2 \ge (1-e^2)e^2. \tag{23}$$

Global Lipschitz regularity yields $m_\delta(\Delta) = \beta_\delta + f_\delta(0)\Delta + R$ with $|R| \le \ell\Delta^2/2$, even for unbounded Gaussian shifts. To bound its contribution, put $P=|u||v|$. For $e \le 1$,

$$\begin{aligned}\mathbb{E}[\Delta^2|B+Q|] &\le \mathbb{E}[4P^3 + 8P^2Q + 5PQ^2 + Q^3] \\ &\le 24e^3 + 48e^4 + 30e^5 + 6e^6 \le 108e^3.\end{aligned}$$

Each bound follows from Hölder's inequality and $\mathbb{E}|u|^6=6$, $\mathbb{E}|v|^6=6e^6$, without independence of u,v. The constant bias term is at least $-|\beta_\delta|\|z\|e \ge -2\bar\beta_\delta e$. As $f_\delta(0) \ge \kappa/2$, $e \le 1/2$, and $54\ell e \le \kappa/8$, the curvature coefficient is at least $3\kappa/8 - \kappa/8 = \mu$.

For the norm inequality, use $\mathcal{G}_\delta(z) = \beta_\delta z + \mathbb{E}[(m_\delta(\Delta)-\beta_\delta)aa^*z]$. Since $|m_\delta(\Delta)-\beta_\delta| \le M|\Delta|$, testing against any unit complex vector w and applying Cauchy–Schwarz gives

$$|\langle w, \mathcal{G}_\delta(z) - \beta_\delta z\rangle| \le M(\mathbb{E}\Delta^2)^{1/2}(\mathbb{E}|a^*w|^2|a^*z|^2)^{1/2} \le 2Me(2+e)(1+e) \le 8Me.$$

Here $|\Delta| \le |a^*h||a^*(z+x_\star)|$, and $\mathbb{E}|a^*v_1|^2|a^*v_2|^2 = \|v_1\|^2\|v_2\|^2 + |v_1^*v_2|^2$. Adding $|\beta_\delta|\|z\| \le 2\bar\beta_\delta$ proves the result. □

A.3 Proof of Theorem 6

Lemma 11 (Uniform bounded-score gradient deviation)**.** *For independent pairs and Gaussian sensing as in Assumption 4, $0 < \zeta < 1/4$, and $0 < \delta \le 1$, with probability at least $1 - \zeta$,*

$$\sup_{\|z\| \le 3/2} \|G_{m_U}(z) - \mathcal{G}_\delta(z)\| \le t_U, \tag{24}$$

where t_U is defined in (14)*.*

Proof. Write $m = m_U$ and $R = 3/2$. For fixed $\|z\| \le R$ and unit v, the scalar projection of a single summand is $X = -\psi_{\tau,\delta}(y - |a^*z|^2)\,\mathrm{Re}(v^*aa^*z)$. For every integer $k \ge 1$,

$$\mathbb{E}|X|^k \le [\mathbb{E}|a^*v|^{2k}\mathbb{E}|a^*z|^{2k}]^{1/2} \le k!R^k.$$

The centered summand is subexponential with a universal scale. Bernstein's inequality bounds its sample-mean deviation probability above t by $2\exp[-cm\min(t^2, t)]$. This argument uses boundedness of the score, not independence of the score and the design and not moments of y.

Separately, the Gaussian norm tail and a union bound give

$$\max_{i \le m} \|a_i\|^2 \le B = C[n + \log(m/\zeta)]$$

except on an event of probability $\zeta/2$. For $g_i(z) = -\psi_{\tau,\delta}(y_i - |a_i^*z|^2)a_ia_i^*z$, the score derivative is at most $1/(2\delta)$, so

$$\|g_i(z) - g_i(w)\| \le \left(\|a_i\|^2 + \frac{R^2}{\delta}\|a_i\|^4\right)\|z - w\| \quad (\|z\|, \|w\| \le R).$$

The empirical map is therefore CB^2/δ-Lipschitz on the norm event. Its population counterpart has Lipschitz constant at most $n + R^2n(n+1)/\delta \le CB^2/\delta$.

Use an ϵ-net of the radius-R ball in $\mathbb{R}^{2n}$, $\epsilon = \delta/(CmB^2)$, of size at most $(1 + 2R/\epsilon)^{2n}$, and a $1/2$-net of its unit sphere of size at most 5^{2n}. A union bound for the unconditioned scalar Bernstein inequalities on these deterministic nets, with a sufficiently large constant in (14), makes every net deviation at most $C[\sqrt{\Lambda_U/m} + \Lambda_U/m]$ except with probability $\zeta/2$. A sphere net controls the vector norm within factor two. Intersecting with the Gaussian norm event and extending the parameter net costs at most C/m, absorbed in t_U. In particular, we have not conditioned a concentration inequality on a row-truncation event and then assumed the rows remain independent. This proves (24). The factor $2n$ from the real coordinates of $\mathbb{C}^n$ is absorbed in the universal constants. □

Proposition 12 (Deterministic global-phase contraction)**.** *Choose θ_z minimizing $\|z - e^{\mathrm{i}\theta}x_\star\|_2$ and put $h = z - e^{\mathrm{i}\theta_z}x_\star$, $e = \|h\| = d(z, x_\star)$. Suppose, for every point with phase-aligned distance at most r, that*

$$\mathrm{Re}\langle d_0(z), h\rangle \ge \mu e^2 - be, \qquad \|d_0(z)\| \le Le + b, \qquad 0 < \mu \le L. \tag{25}$$

With $e_0 = d(z_0, x_\star)$, if $0 < \eta \le \mu/(4L^2)$ and $\max\{e_0, \sqrt{3}b/\mu\} \le r$, then every iterate of $z^+ = z - \eta d_0(z)$ remains in the ball and has distance at most $(1 - \eta\mu)^{T/2}e_0 + \sqrt{3}b/\mu$ after T steps.

Proof. Re-aligning after an update can only decrease the distance. Therefore

$$d(z^+, x_\star)^2 \le (1 - 2\eta\mu + 2\eta^2L^2)e^2 + 2\eta be + 2\eta^2b^2 \tag{26}$$

$$\le (1 - \eta\mu)e^2 + 3\eta b^2/\mu. \tag{27}$$

The second line uses $2be \le (\mu/2)e^2 + 2b^2/\mu$, $2\eta^2L^2 \le \eta\mu/2$, and $\eta\mu \le 1/4$. Induction bounds the squared distance by $\max\{e_0^2, 3b^2/\mu^2\} \le r^2$. Summing the geometric recursion gives $d(z_T, x_\star)^2 \le (1 - \eta\mu)^T e_0^2 + 3b^2/\mu^2$; taking square roots completes the proof. □

To prove Theorem 6, work on the single event (24). Lemma 10 implies (25) with $d_0 = G_{m_U}$ and $b = 2\bar{\beta}_\delta + t_U$. Since $r \le 1/2$, the whole phase ball lies in $\|z\| \le 3/2$. Also $\kappa \le M$ implies $\mu \le L$. Proposition 12 proves the assertion simultaneously for every qualifying starting point and every iteration count. The uniform event is why dependence of the iterates on U is harmless; no fresh sample per update or union bound over T is required.

A.4 Proofs of Theorem 15 and Corollary 7

Lemma 13 (Quantitative spectral gap and direction)**.** *Let $w_c(y) = \min\{\max(y, 0), c\}$, $c > 0$, and $\widehat{M}_D = m_D^{-1} \sum_{i \in D} w_c(y_i) a_i a_i^*$. Its population matrix has the form $M_D = \alpha_c I + \gamma_c x_\star x_\star^*$, with gap* (17)*. For*

$$t_D = Cc \left(\sqrt{\frac{n + \log(C/\zeta)}{m_D}} + \frac{n + \log(C/\zeta)}{m_D} \right), \tag{28}$$

$\|\widehat{M}_D - M_D\|_{\mathrm{op}} \le t_D$ *with probability at least* $1 - \zeta$*. If* $t_D \le \gamma_c/4$*, a unit leading eigenvector* v_0 *obeys*

$$\omega := \sqrt{1 - |v_0^* x_\star|^2} \le 2t_D/\gamma_c, \qquad d(v_0, x_\star) \le \sqrt{2}\,\omega.$$

Proof. Let $g_c(s) = \mathbb{E}_\varepsilon w_c(s + \varepsilon)$. Conditioning on $S = |a^* x_\star|^2$, independence of the orthogonal Gaussian component gives $\alpha_c = \mathbb{E} g_c(S)$ and $\gamma_c = \mathbb{E}[(S - 1) g_c(S)]$. The bounded function g_c has derivative $g_c'(s) = F(c - s) - F(-s)$. Integration by parts against the exponential density gives $\gamma_c = \mathbb{E}[S g_c'(S)]$; both boundary terms $s e^{-s} g_c(s)$ vanish. For $s \in [b_c, 2b_c]$, the interval $[-s, c - s]$ contains $[-b_c, b_c]$. On that latter interval, $f(t) \ge \kappa - \ell b_c \ge \kappa/2$. Thus $g_c'(s) \ge \kappa b_c$ there and

$$\gamma_c \ge \kappa b_c \int_{b_c}^{2b_c} s e^{-s}\, ds \ge \kappa b_c^3 e^{-2b_c} > 0.$$

For fixed unit v, the quadratic form summand satisfies $|w_c(y)|\,|a^* v|^2 \le c |a^* v|^2$ and hence has kth moment at most $c^k k!$. Scalar Bernstein and a $1/4$-net of the unit sphere in $\mathbb{R}^{2n}$, of size at most 9^{2n}, prove (28). For a Hermitian matrix E, such a net satisfies $\|E\|_{\mathrm{op}} \le 2 \max_v |v^* E v|$, by approximating its extremal unit vector and bounding the error by $2(1/4)\|E\|_{\mathrm{op}}$. No response moment is used.

Finally, the leading empirical eigenvalue is at least $\alpha_c + \gamma_c - t_D$. Projecting its eigenvector equation onto $x_\star^\perp$ yields $(\gamma_c - t_D)\omega \le t_D$. This gives the stated angle bound. The phase-distance formula $d(v_0, x_\star)^2 = 2 - 2\sqrt{1 - \omega^2} \le 2\omega^2$ completes the proof. □

Lemma 14 (Radial calibration and second-order direction error)**.** *Let v be a unit vector independent of V, and $\omega = \sqrt{1 - |v^* x_\star|^2}$. Put*

$$t_V = C \left(\sqrt{\frac{\log(C/\zeta)}{m_V}} + \frac{\log(C/\zeta)}{m_V} \right), \qquad a_0 = \min\{1/2, \kappa/(24\ell)\}, \tag{29}$$

$$q_v = \frac{4}{\kappa}[\bar{\beta}_\delta + (M + 2\ell)\omega^2 + t_V]. \tag{30}$$

If $0 < \delta \le \min\{1, \kappa/(2\ell)\}$, $\omega \le \kappa/(16\sqrt{3}\ell)$, and $q_v \le a_0$, then, conditionally on v, with probability at least $1 - \zeta$, the minimizer $\widehat{s}$ of (8) *is unique and*

$$|\widehat{s} - 1| \le q_v, \qquad d(\sqrt{\widehat{s}}\, v, x_\star) \le \sqrt{2}\,\omega + q_v.$$

The conditional assertion applies to each v satisfying the conditions.

Proof. Write $b_a = |a^* v|^2$, $u_a = |a^* x_\star|^2$ and define the population and empirical radial derivatives

$$H(s) = \mathbb{E}[m_\delta(s b_a - u_a) b_a], \qquad \widehat{H}(s) = -\frac{1}{m_V} \sum_{i \in V} \psi_{\tau,\delta}(y_i - s b_i) b_i.$$

Both are well defined by bounded scores, regardless of the raw risk. Gaussian moments give

$$\mathbb{E} b_a^k = k!, \quad \mathbb{E}[(b_a - u_a) b_a] = \omega^2, \quad \mathbb{E}[(b_a - u_a)^2 b_a] = 4\omega^2, \quad \mathbb{E}(b_a - u_a)^2 = 2\omega^2. \tag{31}$$

For completeness, rotate $x_\star = \sqrt{1 - \omega^2}\, v + \omega w$ with $w \perp v$. Conditional on b_a, the first two moments of u_a are $(1 - \omega^2) b_a + \omega^2$ and $(1 - \omega^2)^2 b_a^2 + 4(1 - \omega^2)\omega^2 b_a + 2\omega^4$. Substitution and $\mathbb{E} b_a^k = k!$ prove the mixed identities.

Taylor's formula for m_δ now gives

$$|H(1)| \le |\beta_\delta| + M\omega^2 + \tfrac{\ell}{2}(4\omega^2) \le B_0 := \bar{\beta}_\delta + (M + 2\ell)\omega^2.$$

Also, differentiating under expectation is justified by $f_\delta \le M$, and Cauchy–Schwarz gives

$$\begin{aligned} H'(s) &= \mathbb{E}[f_\delta(sb_a - u_a)b_a^2] \\ &\ge 2f_\delta(0) - \ell\{|s-1|\mathbb{E}b_a^3 + \mathbb{E}|b_a - u_a|b_a^2\} \\ &\ge \kappa - \ell(6|s-1| + 4\sqrt{3}\,\omega) \ge \kappa/2 \quad \text{for } |s-1| \le a_0. \end{aligned}$$

At each of the two points $1 \pm q_v$, the summand in $\widehat{H}$ is bounded in absolute value by b_i, which is exponential. Conditionally on v, scalar Bernstein therefore bounds both deviations $|\widehat{H} - H|$ by t_V with probability $1 - \zeta$. At these endpoints the population derivative has the corresponding sign and magnitude at least $(\kappa/2)q_v - B_0 = B_0 + 2t_V$. The empirical derivatives consequently have opposite signs.

Almost surely $b_i > 0$ and $\widehat{H}'(s) = m_V^{-1}\sum_i K_\delta(y_i - sb_i)b_i^2 > 0$. The finite-sample objective is strictly convex and coercive on $[0, \infty)$ because $\tau < 1$ and its large-s slope tends to $(1-\tau)m_V^{-1}\sum_i b_i > 0$. It has a unique root of $\widehat{H}$ between the positive endpoints $1 - q_v$ and $1 + q_v$, which is its minimizer. Finally $|\sqrt{\widehat{s}} - 1| \le |\widehat{s} - 1|$, and the triangle inequality gives the asserted phase error. □

We now assemble the initialization and refinement bounds. Evaluate t_D, t_V, t_U in (28), (29), and (14) with failure argument $\zeta/3$. Define deterministic quantities

$$d_D = \frac{2t_D}{\gamma_c}, \qquad q = \frac{4}{\kappa}[\bar{\beta}_\delta + (M + 2\ell)d_D^2 + t_V], \qquad b = 2\bar{\beta}_\delta + t_U.$$

Theorem 15 (Finite-sample assembly)**.** *Under Assumption 4, fix $0 < \zeta < 1/4$ and independent splits D, V, U, fixed $c > 0$, $0 < \delta \le \min\{1, \kappa/(2\ell)\}$, and $0 < \eta \le \mu/(4L^2)$. Sufficient conditions are*

$$\begin{gathered} t_D \le \gamma_c/4, \qquad d_D \le \kappa/(16\sqrt{3}\ell), \qquad q \le a_0, \\ \sqrt{2}\,d_D + q \le r, \qquad \sqrt{3}\,b/\mu \le r. \end{gathered} \tag{32}$$

With probability at least $1 - \zeta$, the exact spectral–radial initializer and all subsequent updates satisfy

$$d(z_T, x_\star) \le (1 - \eta\mu)^{T/2}(\sqrt{2}\,d_D + q) + \sqrt{3}\,b/\mu \quad (T \ge 0).$$

Proof. On the spectral event of Lemma 13, $\omega \le d_D$. Conditional on any such direction, the radial event of Lemma 14 fails with probability at most $\zeta/3$. Its radius q_v is bounded by q; hence $d(z_0, x_\star) \le r$. Integrating the conditional failure probability and intersecting with the uniform update event costs at most ζ altogether. The remaining inequality in (32) supplies the invariant statistical radius in Theorem 6. □

To prove Corollary 7, hold κ, M, ℓ, c, p fixed. The spectral gap in (17) is then a fixed positive number. Let $h_N = n\log(N/\zeta)/N$ and $\delta = N^{-1/2}$. When $N \ge C_\star n\log(N/\zeta)$, increasing $C_\star$ if necessary ensures $\delta \le \min\{1, \kappa/(2\ell)\}$, $n \le N$ and $\log(1/\zeta) \le N$. Since $m_J \ge pN$,

$$t_D = O(c[\sqrt{h_N} + h_N]), \qquad t_V = O(\sqrt{h_N} + h_N), \qquad \Lambda_U = O(n\log(N/\zeta)).$$

The last relation follows by bounding $n + \log(m_U/\zeta) \le CN$ in the logarithm of (14). Also $\bar{\beta}_\delta = \ell\log(3\sqrt{N})/(2N) + 1/(4N)$. All quantities on the left of (32), apart from their fixed positive denominators, can therefore be made small enough by choosing $C_\star$ sufficiently large. Theorem 15 proves the corollary and explicitly counts the direction and radial observations. No choice of δ depends on the unknown signal or realized noise. The conclusion gives sufficient, conservative sample sizes, not sharp phase-transition constants.

A.5 Proofs of Theorems 2 and 8

The local-density extension below, combined with the assembly in Appendix A.4, proves Theorem 2. We prove Theorem 8 under its local assumptions, not under the global density hypothesis of Assumption 4. Write P_ε for the noise probability measure. Its restriction near $[-R, R]$ has the stated Lipschitz density; elsewhere it need not be absolutely continuous. Keep

$$F_\delta(s) = \int H_\delta(s-u)\, dP_\varepsilon(u), \qquad f_\delta(s) = \int K_\delta(s-u)\, dP_\varepsilon(u), \qquad m_\delta(s) = F_\delta(s) - \tau.$$

For fixed $\delta > 0$, these integrals are smooth as needed and $0 \le f_\delta \le 1/(2\delta)$ globally, since K_δ is a bounded smooth density. This bandwidth-dependent upper bound is sufficient for differentiation under expectation, but is not used as the constant in the recovery rate.

Lemma 16 (Effective score regularity)**.** *Under the assumptions and bandwidth restriction of Theorem 8, for every $s \in \mathbb{R}$,*

$$f_\delta(0) \ge \kappa/2, \qquad |\beta_\delta| \le \bar{\beta}_{\delta,R}, \tag{33}$$

$$|m_\delta(s) - \beta_\delta| \le M_R|s|, \tag{34}$$

$$|m_\delta(s) - \beta_\delta - f_\delta(0)s| \le \tfrac{1}{2}\ell_R s^2, \tag{35}$$

$$f_\delta(s) \ge f_\delta(0) - \ell_R|s|. \tag{36}$$

These statements do not assert that f_δ has a global upper bound or a global Lipschitz constant independent of δ.

Proof. For $|s| \le R/2$, split the convolution into $|u| \le R$ and $|u| > R$. The local part is bounded by M, whereas in the second part $|s-u| \ge R/2$. The kernel satisfies, for $v \ne 0$,

$$K_\delta(v) \le \frac{\delta^2}{2|v|^3}, \qquad |K_\delta'(v)| \le \frac{3\delta^2}{2|v|^4}.$$

Since the tail is a subprobability measure and $\delta \le R/2$,

$$\sup_{|s|\le R/2} f_\delta(s) \le M + \frac{4\delta^2}{R^3} \le M + \frac{1}{R} =: \overline{M}. \tag{37}$$

For the derivative of the local part, integration by parts is valid because a Lipschitz function is absolutely continuous:

$$\int_{-R}^{R} K_\delta'(s-u)f(u)\, du = K_\delta(s+R)f(-R) - K_\delta(s-R)f(R) + \int_{-R}^{R} K_\delta(s-u)f'(u)\, du.$$

The integral is bounded by ℓ in absolute value. The two boundary terms total at most $8M\delta^2/R^3$. Differentiating the tail convolution costs at most $24\delta^2/R^4$. Thus

$$\sup_{|s|\le R/2} |f_\delta'(s)| \le \ell + \frac{8M\delta^2}{R^3} + \frac{24\delta^2}{R^4} \le \ell + \frac{2M}{R} + \frac{6}{R^2} =: \overline{\ell}. \tag{38}$$

No density or moment bound on the tail was used. At zero, nonnegativity of the tail and $f(u) \ge \kappa - \ell|u|$ locally give

$$f_\delta(0) \ge \kappa \int_{-R}^{R} K_\delta(u)\, du - \ell \int_{-R}^{R} |u| K_\delta(u)\, du \ge \frac{\kappa R}{\sqrt{R^2+\delta^2}} - \ell\delta \ge \kappa(2/\sqrt{5} - 1/4) > \kappa/2.$$

To bound the bias, use the absence of an atom at zero and decompose

$$\beta_\delta = \int_0^R R_\delta(u)[f(u) - f(-u)]\, du + \int_{|u|>R} [H_\delta(-u) - \mathbf{1}\{u \le 0\}]\, dP_\varepsilon(u).$$

The tail term has magnitude at most $\delta^2/(4R^2)$. The local term is bounded by $2\ell \int_0^R uR_\delta(u)\,du$, where

$$\int_0^R uR_\delta(u)\,du = \frac{R^2 - R\sqrt{R^2+\delta^2} + \delta^2 \operatorname{arsinh}(R/\delta)}{4} \le \frac{\delta^2}{4}\log(3R/\delta).$$

This proves (33).

For $|s| \le R/2$, integrating (37) and (38) gives (34)–(36), because $\overline{M} \le M_R$ and $\bar{\ell} \le \ell_R$. For $|s| > R/2$, boundedness of a cdf instead gives $|m_\delta(s) - \beta_\delta| = |F_\delta(s) - F_\delta(0)| \le 1 \le 2|s|/R$. Moreover,

$$\begin{aligned}|m_\delta(s) - \beta_\delta - f_\delta(0)s| &\le 1 + \overline{M}|s| \\ &\le (4/R^2 + 2\overline{M}/R)s^2 \le \tfrac{1}{2}\ell_R s^2.\end{aligned}$$

Finally $f_\delta(s) \ge 0$ and $\ell_R|s| \ge \ell_R R/2 \ge \overline{M} \ge f_\delta(0)$. This proves the global one-sided lower bound (36) and completes the proof. □

Proof of Theorem 8. The identities defining the population score remain valid for the noise measure P_ε: they integrate a bounded score, not the uncentered noise loss. Centered risk is still finite and identified, since the density is positive on a neighborhood of zero.

In Lemma 10, the only uses of global noise regularity were: a lower bound on $f_\delta(0)$, a quadratic Taylor remainder for m_δ, and a linear bound on $m_\delta(s) - m_\delta(0)$. These are exactly (33)–(35). Consequently its proof, including the Gaussian moment constants, applies with

$$r_R = \min\{1/2, \kappa/(432\ell_R)\}, \qquad \mu = \kappa/4, \qquad L_R = 8M_R, \qquad b_R = 2\bar{\beta}_{\delta,R} + t_U.$$

Lemma 11 needs only independent pairs, Gaussian sensing, bounded scores, and the deterministic score derivative bound $1/(2\delta)$. It is unchanged by atoms or irregular densities away from zero. The deterministic recursion therefore proves the local recovery statement with $0 < \eta \le \mu/(4L_R^2)$.

For the spectral initializer, $g_c(s) = \mathbb{E}w_c(s+\varepsilon)$ is bounded and Lipschitz even for a mixed noise law. Its derivative is $F(c-s) - F(-s)$ for almost every s; possible endpoint atoms only affect a Lebesgue-null set of shifts. Integration against the continuous exponential density and integration by parts are therefore still valid. For $s \in [b_{c,R}, 2b_{c,R}]$, the interval $[-s, c-s]$ contains $[-b_{c,R}, b_{c,R}] \subset [-R, R]$, on which $f \ge \kappa/2$. Thus

$$\gamma_c = \mathbb{E}\{S[F(c-S) - F(-S)]\} \ge \kappa b_{c,R}^3 e^{-2b_{c,R}} > 0.$$

The clipped matrix concentration and eigenspace perturbation arguments do not use a noise density and remain unchanged.

For radial calibration, (35) and the same mixed Gaussian moments give

$$|H(1)| \le \bar{\beta}_{\delta,R} + (M_R + 2\ell_R)\omega^2.$$

For the slope, it is important to use only the one-sided inequality (36), not an unavailable global Lipschitz bound:

$$H'(s) \ge 2f_\delta(0) - \ell_R(6|s-1| + 4\sqrt{3}\,\omega) \ge \kappa - \ell_R(6|s-1| + 4\sqrt{3}\,\omega).$$

Differentiation is justified by the global bound $f_\delta \le 1/(2\delta)$ and $\mathbb{E}b_a^2 = 2$ for each fixed bandwidth. Hence the radial conditions become $\omega \le \kappa/(16\sqrt{3}\ell_R)$ and $q_v \le \min\{1/2, \kappa/(24\ell_R)\}$, with the stated substitutions in q_v. Endpoint concentration, strict convexity, and coercivity require only bounded scores and positive K_δ, so the rest of the radial proof is unchanged. The three events now assemble exactly as in Theorem 15, with total failure probability at most ζ.

For fixed κ, M, ℓ, R, c and split fractions, the effective constants and gap are fixed positive numbers. With $\delta = N^{-1/2}$,

$$\bar{\beta}_{\delta,R} = \frac{\ell \log(3R\sqrt{N})}{2N} + \frac{1}{4R^2N}.$$

Increasing the constant in $N \ge C_\star n\log(N/\zeta)$ ensures the new bandwidth restriction and all five basin conditions. For example the resulting all-iteration bound is

$$d(z_T, x_\star) \le r_R(1-\eta\mu)^{T/2} + \frac{C}{\kappa}\left[\sqrt{\frac{n\log(N/\zeta)}{N}} + \frac{n\log(N/\zeta)}{N} + \frac{\ell\log(3R\sqrt{N}) + R^{-2}}{N}\right], \quad (39)$$

where C may depend on the fixed split fractions and $C_\star$ also on the local regularity and clipping parameters. □

If f is C^2 near zero, the proof of the sharp expansion in Proposition 5 uses only the local integral above; its tail is still $O(\delta^2)$. Thus its leading coefficient is unchanged. For the population displacement, let $H(s) = \mathbb{E}[m_\delta((s-1)S)S]$ as before. The scalar bounds give $|H(s)-\beta_\delta-2f_\delta(0)(s-1)| \le 3\ell_R(s-1)^2$ and $H'(s) \ge \kappa - 6\ell_R|s-1|$. Also

$$|f_\delta(0) - f(0)| \le \ell\delta + \frac{M\delta^2}{2R^2} + \frac{\delta^2}{2R^3} = O(\delta).$$

The same root argument proves Corollary 9. For signal norm ρ, the local interval transforms as $\widetilde{R} = R/\rho^2$; the effective constants obey $\widetilde{M}_{\widetilde{R}} = \rho^2 M_R$ and $\widetilde{\ell}_{\widetilde{R}} = \rho^4 \ell_R$. Thus the scale conversion and the numerical-error certificates extend with the same substitutions.

For an admissible example with remote atoms, the law

$$P_\varepsilon = 0.6\ \mathrm{Unif}[-1,1] + 0.1\,\delta_{-2} + 0.3\,\delta_2$$

has $F(0) = 0.4$ and density 0.3 on a neighborhood of $[-1/2, 1/2]$, but no global density. Here δ_u denotes a point mass, not the smoothing bandwidth. The theorem applies with $R = 1/2$, $\kappa = M = 0.3$, and any positive Lipschitz upper bound ℓ. Replacing each point mass at ± 2 by $\sum_{k\ge 1} 2^{-k}\delta_{\pm 2^k}$ preserves that quantile and local density while making $\mathbb{E}|\varepsilon| = \infty$. This is an independently sampled fixed noise law with the specified quantile. It is not a guarantee for adversarially selected replacements or for contamination that changes the target quantile.

A.6 Proof of Theorem 3

Proof of Theorem 3. Let P_x be the joint law of one sensing vector and response. The sensing distribution is the same under every parameter. Conditional on a, the response laws are Cauchy distributions with common scale σ and a location difference $|a^*x|^2 - |a^*x'|^2$; the shift fixing the τ quantile cancels. The equal-scale Cauchy divergence formula (Chyzak & Nielsen, 2019) and $\log(1+t) \le t$ imply, for unit signals,

$$\mathrm{KL}(P_x^{\otimes N}, P_{x'}^{\otimes N}) \le \frac{N}{4\sigma^2}\mathbb{E}(|a^*x|^2 - |a^*x'|^2)^2 \tag{40}$$

$$= \frac{N}{2\sigma^2}(1 - |x^*x'|^2). \tag{41}$$

This calculation uses no Cauchy moment.

Put $d_1 = n - 1 \ge 32$. Greedily packing the binary cube $\{-1,1\}^{d_1}$ with Hamming distance greater than $\lfloor d_1/8 \rfloor$ yields M_0 words with

$$\log M_0 \ge d_1[\log 2 - h(1/8)] > 0.3d_1, \qquad h(t) = -t\log t - (1-t)\log(1-t).$$

Indeed each greedy deletion removes at most $\sum_{j \le d_1/8}\binom{d_1}{j} \le \exp[d_1 h(1/8)]$ words. The binomial bound follows, for $p = 1/8$, by multiplying the sum by $p^{pd_1}(1-p)^{(1-p)d_1}$ and bounding it by $\sum_j \binom{d_1}{j} p^j(1-p)^{d_1-j} = 1$. Normalize each retained word to a unit vector v_j and set

$$r_1 = \frac{1}{10}\min\{r_0, \sigma\sqrt{d_1/N}\}, \qquad x_j = (\sqrt{1-r_1^2}, r_1 v_j) \in \mathbb{R}^n \subset \mathbb{C}^n.$$

Then $d(x_j, e_1) \le \sqrt{2} r_1 \le r_0$. All pairwise real inner products are at least $1 - 2r_1^2 > 0$, so phase distance equals ordinary Euclidean distance between these signals. The Hamming separation gives $d(x_j, x_k) \ge r_1/\sqrt{2}$ for $j \ne k$. By (41) with $x' = e_1$,

$$\frac{1}{M_0}\sum_j \mathrm{KL}(P_{x_j}^{\otimes N}, P_{e_1}^{\otimes N}) \le \frac{N r_1^2}{2\sigma^2} \le 0.005 d_1.$$

To state the information step explicitly, choose the index J uniformly and let Z contain all observed pairs. Its mutual information satisfies $I(J;Z) \le M_0^{-1}\sum_j \mathrm{KL}(P_{x_j}^{\otimes N}, Q)$ for any reference law Q:

subtracting $I(J;Z)$ leaves the nonnegative divergence of the mixture from Q. For any decoder with error probability p_e, conditioning also on its error indicator gives $H(J \mid Z) \le \log 2 + p_e \log M_0$. Hence

$$p_e \ge 1 - \frac{0.005d_1 + \log 2}{0.3d_1} > \frac{1}{2}.$$

This is the usual finite-set Fano inequality, with its constants made explicit.

Decode any estimator $\widehat{x}$ by its nearest packed phase orbit. If $d(\widehat{x}, x_J) < r_1/(2\sqrt{2})$, the triangle inequality forces correct decoding. Thus the average, and hence maximal, probability of error at least $r_1/(2\sqrt{2})$ exceeds $1/2$. Taking $c_0 = 1/(20\sqrt{2})$ proves (10). Finally, the shifted Cauchy cdf gives $F_\varepsilon(0) = \tau$ and $f_\varepsilon(0) = 1/[\pi\sigma(1+\tan^2\{\pi(\tau - 1/2)\})] = \sin^2(\pi\tau)/(\pi\sigma)$. The same probability bound also implies an expected-distance lower bound of one half the displayed threshold. □

B Numerical Guarantees

B.1 Proofs of Proposition 17 and Lemma 18

For a unit direction v independent of V, let $\widehat{H}_v$ be the empirical radial derivative in Lemma 14. Its constrained first-order residual is

$$\mathcal{R}_v(s) = \begin{cases} |\widehat{H}_v(s)|, & s > 0, \\ \max\{0, -\widehat{H}_v(0)\}, & s = 0. \end{cases} \tag{42}$$

The boundary case matters: a constrained minimizer at zero need not have zero derivative.

Proposition 17 (Certified inexact recovery)**.** *Fix deterministic budgets $\epsilon_{\mathrm{eig}}, \epsilon_{\mathrm{rad}}, \epsilon_{\mathrm{grad}} \ge 0$. The unit direction $\widetilde{v}$ must depend only on D and, if used, randomness independent of all three samples. Suppose the solver outputs satisfy, almost surely,*

$$\lambda_1(\widehat{M}_D) - \widetilde{v}^* \widehat{M}_D \widetilde{v} \le \epsilon_{\mathrm{eig}}, \qquad \widetilde{s} \ge 0, \qquad \mathcal{R}_{\widetilde{v}}(\widetilde{s}) \le \epsilon_{\mathrm{rad}}. \tag{43}$$

Start from $\widetilde{z}_0 = \sqrt{\widetilde{s}}\,\widetilde{v}$ and assume every computed update direction differs from G_{m_U} at that iterate by at most ϵ_{grad}. Define

$$\widetilde{d}_D = \frac{2t_D}{\gamma_c} + \sqrt{\frac{2\epsilon_{\mathrm{eig}}}{\gamma_c}}, \qquad \widetilde{b} = 2\bar{\beta}_\delta + t_U + \epsilon_{\mathrm{grad}},$$
$$\widetilde{q} = \frac{4}{\kappa}[\bar{\beta}_\delta + (M + 2\ell)\widetilde{d}_D^2 + t_V + \epsilon_{\mathrm{rad}}]. \tag{44}$$

Under the assumptions of Theorem 15, with d_D, q, b replaced by $\widetilde{d}_D, \widetilde{q}, \widetilde{b}$ in its five conditions, with probability at least $1 - \zeta$,

$$d(\widetilde{z}_T, x_\star) \le (1 - \eta\mu)^{T/2}(\sqrt{2}\,\widetilde{d}_D + \widetilde{q}) + \sqrt{3}\,\widetilde{b}/\mu \quad (T \ge 0). \tag{45}$$

The conclusion also holds with all local-density substitutions of Theorem 8.

Proof. For $n \ge 2$, on the spectral event the empirical eigengap is at least $\gamma_c - 2t_D \ge \gamma_c/2$. Expansion in the empirical eigenbasis bounds the sine angle of $\widetilde{v}$ from an exact leading direction by $\sqrt{2\epsilon_{\mathrm{eig}}/\gamma_c}$. For rank-one unit projectors the operator distance equals the sine angle; its triangle inequality therefore gives $\omega = \sqrt{1 - |\widetilde{v}^* x_\star|^2} \le \widetilde{d}_D$. For $n = 1$ the direction error is zero, so this bound is automatic.

Conditional on $\widetilde{v}$, the two endpoints $1 \pm \widetilde{q}$ are deterministic and radial Bernstein concentration remains valid. Their empirical derivative magnitudes are at least

$$\frac{\kappa\widetilde{q}}{2} - [\bar{\beta}_\delta + (M + 2\ell)\omega^2] - t_V \ge \bar{\beta}_\delta + (M + 2\ell)\widetilde{d}_D^2 + t_V + 2\epsilon_{\mathrm{rad}} > \epsilon_{\mathrm{rad}},$$

with opposite signs. Since $1 - \widetilde{q} \ge 1/2$ and $\widehat{H}_{\widetilde{v}}$ is increasing, the residual condition excludes $\widetilde{s} = 0$ and all points outside these endpoints. Thus $|\widetilde{s} - 1| \le \widetilde{q}$ and $d(\widetilde{z}_0, x_\star) \le \sqrt{2}\,\widetilde{d}_D + \widetilde{q}$. The gradient error changes the two regularity inequalities by at most $\epsilon_{\mathrm{grad}} e$ and ϵ_{grad}, respectively. The same deterministic recursion, intersected with the three statistical events, gives (45). The local-density proof uses only the scalar inequalities already proved in Lemma 16, so the substitutions remain valid. □

A fixed eigensolver iteration count is not a certificate. Neither is a small eigenvector residual alone: for $\widehat{M} = \mathrm{diag}(2, 1)$, $v = (0, 1)^*$ has zero eigenvector residual but Rayleigh deficit one. One sufficient computable upper bound is obtained from a unitary basis Q: with $B = Q^* \widehat{M}_D Q$,

$$U_Q = \max_j \left(B_{jj} + \sum_{k \neq j} |B_{jk}| \right) \geq \lambda_1(\widehat{M}_D).$$

Indeed, apply the eigenvector equation of B at a coordinate of largest modulus. Checking $U_Q - \widetilde{v}^* \widehat{M}_D \widetilde{v} \leq \epsilon_{\mathrm{eig}}$ therefore suffices. A numerically near-eigenbasis can make this upper bound tight. These are exact-arithmetic certificates; roundoff in orthogonality, matrix products, and residual evaluation must be included in the error budget for a rigorous floating-point certificate.

Lemma 18 (Data-computable radial calibration)**.** *For any finite real responses, $b_i = |a_i^* v|^2 > 0$, $\delta > 0$, and $0 < \tau < 1$, the constrained radial problem has a unique minimizer. For every $\epsilon_{\mathrm{rad}} > 0$, it can be solved to residual* (42) *at most ϵ_{rad} in finitely many scalar evaluations, without knowing $x_\star$, its norm, or the proof basin.*

Proof. The strictly increasing derivative $\widehat{H}_v$ satisfies

$$0 < \widehat{H}_v'(s) \leq \widehat{L}_V := \frac{1}{2\delta m_V} \sum_{i \in V} b_i^2, \qquad \lim_{s \to \infty} \widehat{H}_v(s) = (1 - \tau) m_V^{-1} \sum_i b_i > 0.$$

If $\widehat{H}_v(0) \geq 0$, return zero, whose constrained residual is zero. Otherwise the unique minimizer is the positive root. The score vanishes at

$$u_\tau = \frac{\delta(1 - 2\tau)}{2\sqrt{\tau(1 - \tau)}}.$$

Consequently the finite, observable upper endpoint

$$S_V = 1 + \max\left\{0, \max_{i \in V} \frac{y_i - u_\tau}{b_i}\right\}$$

has $y_i - S_V b_i < u_\tau$ for all i and hence $\widehat{H}_v(S_V) > 0$. Bisection on $[0, S_V]$ preserves the root bracket. After k refinements its midpoint s_k satisfies $|\widehat{H}_v(s_k)| \leq \widehat{L}_V S_V \, 2^{-k-1}$. Thus $k \geq \lceil \log_2^+(\widehat{L}_V S_V / \epsilon_{\mathrm{rad}}) \rceil$ suffices, where $\log_2^+(u) = \max\{0, \log_2 u\}$. If an evaluated point has zero derivative, return it immediately. All quantities are finite for the stated inputs. □

Gaussian sensing and a direction independent of V give $b_i > 0$ almost surely. Doubling a positive trial endpoint until its derivative becomes positive is another terminating way to construct the bracket. The bound above is data-dependent; it is not a noise-uniform runtime bound, particularly when b_i is very small or a response is extreme.

B.2 Proof of Corollary 19

Corollary 19 (A finite-iteration statistical schedule)**.** *Fix the regularity parameters, clipping level, and split fractions as in Corollary 7, or its local-density extension. For sufficiently large $N \geq C_\star n \log(N/\zeta)$, use*

$$\delta = N^{-1/2}, \quad \epsilon_{\mathrm{eig}} \leq N^{-1}, \quad \epsilon_{\mathrm{rad}}, \epsilon_{\mathrm{grad}} \leq N^{-1/2}, \quad T = \left\lceil \frac{\log[N \max\{1, \kappa\}^2]}{\eta \mu} \right\rceil, \tag{46}$$

with fixed admissible η. If the certificates in Proposition 17 hold, with probability at least $1 - \zeta$ the output obeys

$$d(\widetilde{z}_T, x_\star) \leq \frac{C}{\kappa} \left[\sqrt{\frac{n \log(N/\zeta)}{N}} + \frac{n \log(N/\zeta)}{N} + B_N \right], \tag{47}$$

where $B_N = [\ell \log(3\sqrt{N}) + 1]/N$ in the global-density case and $B_N = [\ell \log(3R\sqrt{N}) + R^{-2}]/N$ in the local-density case. Here C may depend on the fixed split fractions, and $C_\star$ also on the fixed model and tuning parameters.

Proof. The positive spectral gap is fixed. The additional direction error is $O(N^{-1/2})$ and the two other numerical tolerances are $O(N^{-1/2})$. As in Appendix A.4, increasing $C_\star$ makes all five inexact basin conditions hold. Moreover $(1-\eta\mu)^{T/2} \le [\max\{1,\kappa\}\sqrt{N}]^{-1}$. The initial error is at most the applicable basin radius, which is at most $1/2$. Its optimization contribution is therefore at most $1/(\kappa\sqrt{N})$, absorbed by the leading term in (47). The gradient tolerance contributes at most $4\sqrt{3}/(\kappa\sqrt{N})$ and is absorbed in the same way. The remaining terms follow from the same concentration and bias bounds. □

For a fixed, theory-aligned experimental interface, use proper $\mathcal{CN}(0, I_n)$ sensing (real and imaginary variances $1/2$), unit-energy test signals, and a predeclared population-centered noise law: $\varepsilon = \sigma[Z - Q_\tau(Z)]$, with fixed σ, not an empirical quantile or a realized-sample SNR normalization. Use fixed index splits, for example thirds, and the same deterministic c, δ, η across repetitions of a cell; v uses only D. Compute the radial solution on V, then exactly T ungated updates on U with response residual $y - |a^* z|^2$. Choose admissible η using declared model regularity bounds (M_R for the local-density result), not the true reconstruction error. Log numerical certificates and report any failures; do not discard them and reinterpret the theorem as a success probability conditional on passing. The conservative schedule (46) is a sufficient reference, not a recommended practical iteration count for every baseline. For equal-budget empirical comparisons, predeclare a different budget and retain the optimization term in (45).

For $\rho = \|x_\star\| > 0$, introduce $\widetilde{x} = x_\star/\rho$, $\widetilde{z} = z/\rho$, $\widetilde{y} = y/\rho^2$, and $\widetilde{\varepsilon} = \varepsilon/\rho^2$. The normalized tuning parameters and density bounds are

$$\widetilde{c} = c/\rho^2, \quad \widetilde{\delta} = \delta/\rho^2, \quad \widetilde{f}(t) = \rho^2 f(\rho^2 t), \quad \widetilde{\kappa} = \rho^2\kappa, \quad \widetilde{M} = \rho^2 M, \quad \widetilde{\ell} = \rho^4 \ell.$$

The score is invariant under simultaneous scaling of its residual and bandwidth. Thus $G_{m,\delta}(\rho\widetilde{z}) = \rho\widetilde{G}_{m,\widetilde{\delta}}(\widetilde{z})$, and the same η implements the normalized update. The clipped spectral matrix scales by ρ^2 without changing its eigenvectors, and the radial objective is multiplied by ρ^2 after $s = \rho^2\widetilde{s}$. Consequently the preceding results apply with the tilded constants and yield relative error $d(z, x_\star)/\rho$. In particular the leading density multiplier becomes $1/(\rho^2\kappa)$. This is an analytical change of variables; choosing tuning parameters using an unknown true ρ is not authorized by it. Uniform guarantees over an energy range require the transformed constants to be controlled throughout that range.

The guarantee covers the fixed-tuning, split-sample, ungated estimator. Data-adaptive clipping, bandwidth selection, gates, and oracle stopping require separate analysis.

B.3 Sparse extension (unproved)

This sparse extension is a proposal, not a proved theorem. The dense result does not imply a sparse rate by replacing n with $s \log n$. A credible first extension is exact support screening followed by an independent, support-restricted application of RQWF. It trades a stronger beta-min assumption and a larger initialization budget for a short proof route.

Algorithm 2 Proposed sparse RQWF

Input: Four independent splits S, D, V, U; known support size s; fixed clipping and RQWF tuning.

1. **Screen.** Compute $q_j = |S|^{-1} \sum_{i \in S} w_c(y_i)|a_{ij}|^2$. Let $\widehat{I}$ contain the s largest q_j.
2. **Initialize on the selected support.** Apply the direction and radial steps of Algorithm 1 to $a_{i,\widehat{I}}$ using D and V.
3. **Refine on the selected support.** Apply the same bounded-score updates using U in $\mathbb{C}^{\widehat{I}}$; set every other coordinate to zero.

To compare the screening budget with the conditional estimation rate, write $I_\star = \mathrm{supp}(x_\star)$ and assume $|I_\star| = s$ for this proposal. The existing clipped spectral population identity gives $\mathbb{E}q_j = \alpha_c + \gamma_c |x_{\star j}|^2$. For $\beta_{\min} = \min_{j \in I_\star} |x_{\star j}|$, the active–inactive gap is at least $\gamma_c \beta_{\min}^2$. Each summand is bounded by $c|a_{ij}|^2$, so the intended coordinate concentration scale is

$$Cc\left[\sqrt{\frac{\log(n/\zeta)}{|S|}} + \frac{\log(n/\zeta)}{|S|}\right].$$

Exact top-s screening requires this scale to be smaller than half the gap. The resulting proposed sufficient screening budget is

$$|S| \gtrsim \left(\frac{c^2}{\gamma_c^2 \beta_{\min}^4} + \frac{c}{\gamma_c \beta_{\min}^2} \right) \log(n/\zeta).$$

For balanced nonzero coordinates, $\beta_{\min}^2 \asymp 1/s$, and fixed c, γ_c, the leading cost is $s^2 \log(n/\zeta)$. Signals with arbitrarily small nonzero coordinates have no such uniform screening guarantee.

Conditional on correct screening, the three remaining independent splits have the same model on a fixed s-dimensional support. The natural target is therefore

$$d(\widehat{x}, x_\star) = \widetilde{O}_{\mathbb{P}}\left(\frac{1}{\kappa} \sqrt{\frac{s}{N_{\text{fit}}}} \right), \qquad N_{\text{fit}} = |D| + |V| + |U|,$$

subject to all dense theorem conditions with dimension s. This is a *proposed conditional rate*, not a completed sparse theorem. The full sample budget is $N_{\text{total}} = |S| + N_{\text{fit}}$. A regime $N_{\text{total}} \ll n$ would also require $s^2 \log n \ll n$ for this screening strategy. Known s, exact support size, and beta-min are additional assumptions; unknown sparsity and approximate sparsity need further analysis.

To complete the proposed sparse proof, first formalize simultaneous coordinate concentration and top-s separation, including ties and the failure allocation. Next condition on the independent screening sample and transfer every initializer, calibration, update, and numerical condition to the recovered support, counting all four splits. Finally combine the screening failure with the conditional recovery failure and state the resulting finite-sample constants. The $s \log(en/s)$ estimation complexity without exact support recovery would require a different restricted geometry/concentration argument and a justified sparse update.

The sparse lower bound also remains open. Up to scale and radius truncation, the intended shifted-Cauchy lower rate is $\Omega(\sqrt{s \log(en/s)/N_{\text{total}}})$. A proof would construct separated supports in a local unit-sphere chart, control their phase distance, and apply the Cauchy KL bound (41) with a sparse packing entropy. This packing and its admissible (s, n) range are not established here. An unconditional lower bound over all sparse signals would not directly match a beta-min, exact-support upper bound: these are different parameter classes, and the comparison must say so.

C Theoretical Comparison

We distinguish three questions: what functional identifies the signal, what noise model is covered, and whether the guarantee concerns an estimator, an initialization-to-solution algorithm, or local optimization. Throughout this comparison, n is the signal dimension, N the sample size, s the sparsity, and d_t the signal error up to sign or phase. Constants and admissible parameter ranges remain those of the cited results; displayed orders do not identify their constants or failure probabilities.

C.1 Clean and corrupted models

Table 4 separates the results grouped in the short main-text table. A corruption model with an exact or bounded-noise inlier majority is different from a dense random-noise model in which every observation is perturbed.

Table 4: Clean and corruption-robust theory. Bounds are stated for the analyzed setting, not for our complex experimental adaptations.

Method	Analyzed model	Sample / algorithm guarantee
WF (Candes et al., 2015)	Complex Gaussian sensing; noiseless intensities.	$N \gtrsim n \log n$; spectral initialization and geometric convergence (Theorem 3.3).
Median-TWF (Zhang et al., 2018)	Real Gaussian theory; a small constant outlier fraction and dense noise w bounded in ℓ_∞.	$N \gtrsim n \log n$; initialization plus $d_t \lesssim (1-\rho)^t \|x_\star\| + \|w\|_\infty / \|x_\star\|$ (Theorems 1–2). A complex extension is stated.
Duchi–Ruan (2019)	Real/complex clean models; the outlier theorem uses real Gaussian sensing and corruption models M1/M2.	$N \gtrsim n$ under the respective assumptions; spectral–radial initialization and quadratic prox-linear convergence with exact subproblems (Theorems 2–3).
SRPR (Zheng & Xue, 2024)	Real Gaussian landscape theory; sparse arbitrary corruption, with a bounded dense-noise extension.	$N \gtrsim n \log n$ with bandwidth-dependent constants; local minimizer localization (Theorems 5–6), and clean local linear gradient convergence (Theorem 1).
AdaIPL (Zheng et al., 2024a)	Real sub-Gaussian sensing, small-ball/sharpness conditions, and sparse corrupted responses.	$N \gtrsim n$ gives residual-quantile control under fixed model constants. Given a local start, LAC/HAC yield geometric outer convergence and inner-work bounds (Section 6).

Median-TWF's bounded-noise floor depends on $\|w\|_\infty$ rather than a shrinking random-noise estimation term. Duchi–Ruan's outlier Theorem 3 permits corruption fractions below $1/\pi$ under M1 and $1/4$ under M2, subject to its sampling conditions; this is not a dense Cauchy-noise theorem. Its real/complex clean result must not be read as a complex counterpart of that outlier theorem. SRPR further localizes minimizers within $O(\delta\alpha/(1-\alpha))$ when the corruption fraction is α and inliers are exact. With bounded dense noise γ, the radius becomes $O(\max\{\sqrt{\delta\gamma}, \delta\alpha/(1-\alpha)\})$ under its small-noise and bandwidth conditions. These are landscape statements, not a moment-free statistical rate or a guarantee for every practical initialization. AdaIPL uses a quantile of absolute residuals to choose a step size for an ℓ_1 objective; that quantile does not specify $Q_\tau(Y \mid a)$.

C.2 Dense noise and quantile models

Table 5 makes the moment and signal-class distinctions explicit. In particular, a sparse rate cannot be compared with a dense rate merely by replacing n with s.

As a quantile-loss precursor, Shen et al. already use asymmetric check loss for phase retrieval. Our contribution is not the first use of quantiles or check loss, but a dense complex finite-sample guarantee joining a prescribed response target, initialization, calibration, smoothing, and refinement. It does not subsume their sparse objective or establish our proposed sparse extension.

Table 5: Statistical scope and target. σ_1 denotes a bound on the sub-exponential noise norm and $\sigma_q = \|\varepsilon\|_{L_q}$.

Method	Signal, noise, and target	Statistical guarantee / scope
Shen et al. (2022)	Real sparse signal; check loss with an $L_{1/2}$ penalty.	ADMM and numerical recovery under bounded/Laplace noise; no complex end-to-end statistical recovery rate is provided.
Thresholded WF (Cai et al., 2016)	Real s-sparse signal; Gaussian sensing and independent, centered sub-exponential noise.	Error $O(\sigma_1\|x_\star\|^{-1}\sqrt{s\log n/N})$ after contraction; initialization requires $N \gtrsim (1+\sigma_1^2/\|x_\star\|^4)s^2\log(Nn)$ (Theorem 3.1).
Buna–Rebeschini (2025)	Real signal; fourth-moment-bounded noise and adversarial contamination of sensing vectors and responses.	Robust initialization/descent; statistical and contamination errors. Handles a corruption model not covered by our theorem.
Huang et al. (2025)	Complex sub-Gaussian sensing; $\mathbb{E}(\varepsilon \mid a) = 0$, $\sigma_q < \infty$, $q > 2$; design-dependent noise allowed.	$N \gtrsim n$; NCVX-LS/CVX-LS statistical bounds, including $O(\sigma_q\|x_\star\|^{-1}\sqrt{n/N})$ in the high-energy regime (Theorem 4); minimax lower bounds.
RQWF	Complex Gaussian sensing; independent additive noise, $F(0) = \tau$, positive locally Lipschitz density; no noise moments.	$N \gtrsim n\log(N/\zeta)$; split initialization and refinement give $O_{\mathbb{P}}(\kappa^{-1}\sqrt{n\log N/N})$ for fixed model parameters; Cauchy submodel lower bound.

For comparison with complex finite-moment theory, let $r = \|x_\star\| > 0$. The NCVX-LS bound of Huang et al. (2025, Theorem 4) is

$$d(\widehat{x}, x_\star) \lesssim \min\left\{\frac{\sigma_q}{r}\sqrt{\frac{n}{N}},\ \sqrt{\sigma_q}\left(\frac{n}{N}\right)^{1/4}\right\}.$$

Its finite-moment mean model allows broader sensing and design-dependent noise; the displayed result is for an estimator, not a convergence theorem for an arbitrary nonconvex solver. RQWF instead controls an explicit algorithm for an independent-noise quantile model, including Cauchy laws excluded by $q > 2$. The tradeoff is Gaussian sensing, local density assumptions, model-dependent tuning, and an extra logarithmic factor. Thus neither theorem subsumes the other, and the difference is not a faster dimension–sample rate for RQWF.

The RQWF guarantee and its limitations can be summarized as follows. With fixed admissible model parameters and $\delta = N^{-1/2}$, Theorem 2 separates three contributions:

$$d(z_T, x_\star) \lesssim \underbrace{e^{-c_{\mathrm{opt}}T}}_{\text{optimization}} + \underbrace{\frac{1}{\kappa}\sqrt{\frac{n\log(N/\zeta)}{N}}}_{\text{sampling}} + \underbrace{\frac{\ell\log(3R\sqrt{N}) + R^{-2}}{\kappa N}}_{\text{smoothing}},$$

where the smaller $n\log(N/\zeta)/N$ term is absorbed under the theorem's sample condition and $c_{\mathrm{opt}} > 0$ depends on the admissible step and density constants. The guarantee starts from data: $N = m_D + m_V + m_U$ counts directional initialization, independent radial calibration, and refinement. The same U is reused at every update; full reuse across initialization and refinement is not covered.

Key proof steps are a moment-free clipped spectral gap, second-order directional error in quantile calibration, and a uniform gradient bound for reused update data. Pseudo-Huber smoothing additionally creates $O(\delta^2\log(1/\delta))$ bias under asymmetry.

Theorem 3 matches the dimension–sample dependence on a shifted-Cauchy submodel at fixed scale and interior τ, up to logarithms. It does not prove uniform minimax optimality over every locally regular noise law. The recovery theorem does not cover adversarially corrupted sensing vectors, arbitrary dependent noise, unknown-density adaptive tuning, practical annealing/gates, or sparse recovery.

D SYNTHETIC EXPERIMENTS

D.1 BRIDGE AND BASELINE IMPLEMENTATIONS

This study is separate from the fixed-scale v2 suite in Appendix D.3. It uses 42 complete cells: two dimensions $n = 128, 256$, three ratios $N/n = 4, 8, 16$, three symmetric noise families at two fixed scales $\sigma = 0.05, 0.20$, and six noiseless controls. Each cell contains 20 independent problems shared by all 15 methods/variants. All signals are complex unit vectors and sensing is proper complex Gaussian. The principal endpoint is 2,400 charged matrix-product equivalents; 600 and 1,200 are predeclared intermediate checkpoints. The pilot uses a different seed namespace and is excluded. Parameters are frozen before production and no reconstruction error enters a solver or selects an endpoint.

The endpoints reproduce the existing T/P implementations at identical updates, data, and initialization seeds. Starting from T, the bridge is:

1. B1: replace disjoint D, V, U by all observations in every stage.
2. B2: replace fixed cap 4 by the empirical 0.95 quantile of y_+, capped at $9\,\mathrm{median}(y_+)/\log 2$.
3. B3: replace the smoothed quantile radial solve by squared radius $s_0 = \mathrm{median}(y_+)/\log 2$.
4. B4: replace the full leading eigensolve by 50 power iterations, with shared random initialization across power-based variants.
5. B5: replace step 0.02 by $2\delta\mu/s_0$, $\mu = 0.25$, where s_0 denotes the frozen initial squared norm.
6. B6: add P's normalized-amplitude gate $[0.3, 5]$ and 0.95 centered-residual gate, retaining its declared fallback rules.
7. B7: replace $\delta = N^{-1/2}$ by the fixed value $0.25 s_0$.
8. P: replace this bandwidth by $s_0 \max\{10^{-8}, 0.25/(1 + t/150)^2\}$.

The nine nodes are complemented by P without gates and by Median-TWF, SRPR-SI, SRPR-RI, AdaIPL-LAC, and AdaIPL-HAC, giving 15 methods. B7 is also P without annealing. All nodes retain the same fitted quantile $\tau = 1/2$. T uses a radial KKT tolerance of 10^{-9} and a full Hermitian eigensolve with a floating-point residual check; the latter is not an interval-arithmetic certificate of the theorem's spectral condition.

Table 2 summarizes all 42 cells, including clean controls. Its pointwise confidence intervals are not adjusted for multiple comparisons.

The reuse edge changes the sample information available to initialization, radial calibration, and refinement as a block. It also changes per-update cost under the work cap. The annealing edge changes both bandwidth and effective step. Thus this is a sequential implementation bridge, not an order-independent factorial attribution or a proof that the empirical variant satisfies the theory. Gates and annealing frequently increase the median error along this path, and those results are retained.

SRPR-SI and SRPR-RI implement the two-stage structure in Zheng & Xue (2024): pseudo-Huber smoothing with $\delta = 0.25$ and monotone accelerated proximal-gradient updates, followed by nonsmooth inexact prox-linear (IPL) refinement. SI uses the smallest eigenvector of the covariance restricted to responses below their median, followed by the nonnegative weighted-median radial minimizer, following Duchi & Ruan (2019). RI uses an isotropic complex direction and an independent $\mathrm{Unif}(0, 4)$ radius. Both acceleration terms and the reference-gradient candidate are retained. Smooth backtracking starts at 0.05, shrinks by 0.5, and uses Armijo coefficient 10^{-4}. The smooth stage receives a predeclared 300 matrix-product equivalents after initialization, then IPL uses the remaining allowance. This switch rule and line-search initialization are implementation choices for the present budget study, not claims about unspecified original code. The same pipeline is used for clean controls, although the original SRPR procedure need not refine noiseless data with IPL.

AdaIPL uses the adaptive step and LAC/HAC criteria of Zheng et al. (2024a). With $L = 2\|A\|_2^2/N$, its step is $t = \min\{1/L, G q_{0.5}(|\,|Az|^2 - y|)\}$, using the median order statistic and $G = 8(100)/(L^2\|z_0\|_2^2)$. The IPL stage of SRPR uses $t = 1/L$. Both acceptance parameters are

0.24. These are explicitly complex reimplementations; the original real/sparse-corruption theorems and published benchmark results are not reproduced or asserted here.

For complex vectors viewed as real Euclidean vectors, define

$$Bd = \frac{2}{N}\Re\{\overline{Az} \odot Ad\}, \qquad B^*\lambda = \frac{2}{N}A^*((Az) \odot \lambda), \qquad b = \frac{y - |Az|^2}{N}.$$

Each inner solve minimizes $H(d) = \|Bd - b\|_1 + \|d\|_2^2/(2t)$, whose dual is $D(\lambda) = -t\|B^*\lambda\|_2^2/2 - \lambda^T b$ with $\|\lambda\|_\infty \leq 1$. Projected accelerated gradient uses the conservative Lipschitz bound $4t \max_i |(Az)_i|^2 \|A\|_2^2/N^2$ and returns $d = -tB^*\lambda$. Acceptance requires

$$H(d) - D(\lambda) \leq \begin{cases} 0.24[H(0) - H(d)] & \text{LAC}, \\ 0.24\|d\|_2^2/(2t) & \text{HAC}, \end{cases}$$

up to a logged roundoff allowance $10^{-12}\max\{1, H(0)\}$. Every accepted certificate is saved. A budget-limited inner solve does not commit an uncertified outer step. At budget 600, zero accepted IPL steps occur in 839/840 SRPR-SI, 779/840 SRPR-RI, 154/840 AdaIPL-LAC, and 270/840 AdaIPL-HAC runs. At 2,400 these counts are all zero; median accepted step counts are respectively 2, 3, 8, and 6. The comparison therefore measures finite-budget behavior, not all methods at convergence.

D.2 Budget accounting and results

A forward or adjoint product on m rows costs m/N equivalents. Constructing a weighted Gram matrix costs nm/N, including T's initialization and each baseline's separate SI and spectral-norm work. Every Jacobian product, rejected line-search trial, and incomplete inner iteration is charged. Eigensolves and scalar arithmetic are timed but are not represented by this metric: equal allowance means neither equal FLOPs nor equal runtime. Shared-machine timing is descriptive only. The older v2 budget excludes Gram construction, so identical numerical budget labels across suites are not comparable.

Checkpoints retain the last committed iterate available within each cap. Numerical failures would receive infinite error without filtering; none occurs in the 12,600 fits. Within each cell, the same 20 problem indices are resampled jointly across methods for 2,000 bootstrap draws. Intervals for medians and differences of medians are pointwise, not simultaneous. The median-win counts in the main text are descriptive summaries. All 37,800 checkpoint errors are recomputed from saved estimates, all work caps are checked, and all 18,437 accepted IPL certificates satisfy their logged criterion. Five representative datasets regenerated from their seeds reproduce their data hashes. Independent code checks cover the complex real-adjoint identity, finite-difference gradients, a tiny dual optimum solved by active-set enumeration, endpoint parity, checkpoint noninterference, and clean real/complex recovery. These are implementation checks, not proof certification.

Table 6: Complete primary-budget medians, $n = 128$. G, t_3, C, and Clean denote Gaussian, Student-t_3, Cauchy, and noiseless data. All methods share 20 problems per row. T/P abbreviate RQWF-T/P; SRPR/AdaIPL are the complex extensions defined above.

N/n	Noise / scale	T	P	Med.-TWF	SRPR-SI	SRPR-RI	Ada-LAC	Ada-HAC
4	C / 0.05	1.19	0.629	0.281	1.09	1.26	1.16	1.16
4	C / 0.2	1.48	0.857	1.03	1.32	1.4	1.36	1.36
4	G / 0.05	1.1	0.234	0.0563	0.101	1.24	0.858	0.868
4	G / 0.2	1.17	0.453	0.233	0.74	1.2	0.995	0.998
4	Clean / 0	1.1	0.258	7.96e-15	0.00224	1.22	0.826	0.837
4	t_3 / 0.05	1.12	0.29	0.0911	0.115	1.27	0.851	0.852
4	t_3 / 0.2	1.2	0.709	0.348	0.703	1.29	1.01	1.01
8	C / 0.05	0.37	0.0515	0.104	0.058	0.0614	0.0642	0.0706
8	C / 0.2	1	0.213	0.411	0.227	0.516	0.299	0.305
8	G / 0.05	0.0629	0.0321	0.0316	0.0267	0.0273	0.0287	0.0296
8	G / 0.2	0.308	0.128	0.12	0.102	0.106	0.117	0.122
8	Clean / 0	0.000727	1e-15	3e-16	2.34e-12	1.14e-11	1.3e-10	1.84e-05
8	t_3 / 0.05	0.094	0.0365	0.0464	0.0334	0.0341	0.0354	0.0384
8	t_3 / 0.2	0.488	0.146	0.183	0.132	0.141	0.15	0.15
16	C / 0.05	0.0672	0.0259	0.0575	0.0288	0.0289	0.0285	0.0304
16	C / 0.2	0.307	0.113	0.229	0.118	0.121	0.123	0.124
16	G / 0.05	0.0301	0.0204	0.0202	0.0173	0.0172	0.018	0.0181
16	G / 0.2	0.127	0.0781	0.0804	0.0644	0.0642	0.0677	0.0685
16	Clean / 0	0.0016	3.25e-16	2.14e-16	1.7e-12	1.69e-12	1.75e-12	0.00536
16	t_3 / 0.05	0.0371	0.0208	0.0291	0.0202	0.0202	0.0201	0.0206
16	t_3 / 0.2	0.169	0.0861	0.117	0.08	0.0805	0.0829	0.0832

Table 7: Complete primary-budget medians, $n = 256$; conventions as in Table 6. Large finite errors are retained.

N/n	Noise / scale	T	P	Med.-TWF	SRPR-SI	SRPR-RI	Ada-LAC	Ada-HAC
4	C / 0.05	1.3	0.65	0.546	1.13	1.38	1.17	1.17
4	C / 0.2	1.54	0.93	1.21	1.39	1.45	1.36	1.35
4	G / 0.05	1.12	0.345	0.0549	0.115	1.27	0.934	0.939
4	G / 0.2	1.23	0.565	0.228	0.939	1.39	1.08	1.09
4	Clean / 0	1.16	0.439	6.93e-14	1.04	1.31	1.18	1.18
4	t_3 / 0.05	1.15	0.288	0.0847	0.991	1.31	1.1	1.1
4	t_3 / 0.2	1.18	0.619	0.341	1.01	1.35	1.14	1.14
8	C / 0.05	0.392	0.0516	0.101	0.0597	0.0894	0.0662	0.0804
8	C / 0.2	1.1	0.213	0.42	0.226	0.943	0.363	0.368
8	G / 0.05	0.0682	0.0326	0.0316	0.0263	0.0287	0.0288	0.0324
8	G / 0.2	0.307	0.131	0.124	0.103	0.12	0.121	0.123
8	Clean / 0	0.00261	1.21e-15	3.92e-16	1.66e-10	2.8e-09	7.25e-09	0.00322
8	t_3 / 0.05	0.0945	0.0365	0.045	0.0328	0.0372	0.0376	0.0391
8	t_3 / 0.2	0.441	0.148	0.18	0.134	0.174	0.159	0.163
16	C / 0.05	0.0695	0.027	0.0576	0.0299	0.0295	0.0292	0.0329
16	C / 0.2	0.298	0.11	0.226	0.115	0.119	0.122	0.123
16	G / 0.05	0.0298	0.0199	0.0202	0.0168	0.0168	0.0174	0.0177
16	G / 0.2	0.131	0.0811	0.0797	0.0669	0.0675	0.0706	0.071
16	Clean / 0	0.00311	3.72e-16	2.62e-16	1.69e-12	1.71e-12	1.78e-12	0.00523
16	t_3 / 0.05	0.0397	0.0211	0.0292	0.0202	0.0206	0.021	0.0214
16	t_3 / 0.2	0.169	0.0855	0.115	0.0792	0.0818	0.0827	0.0852

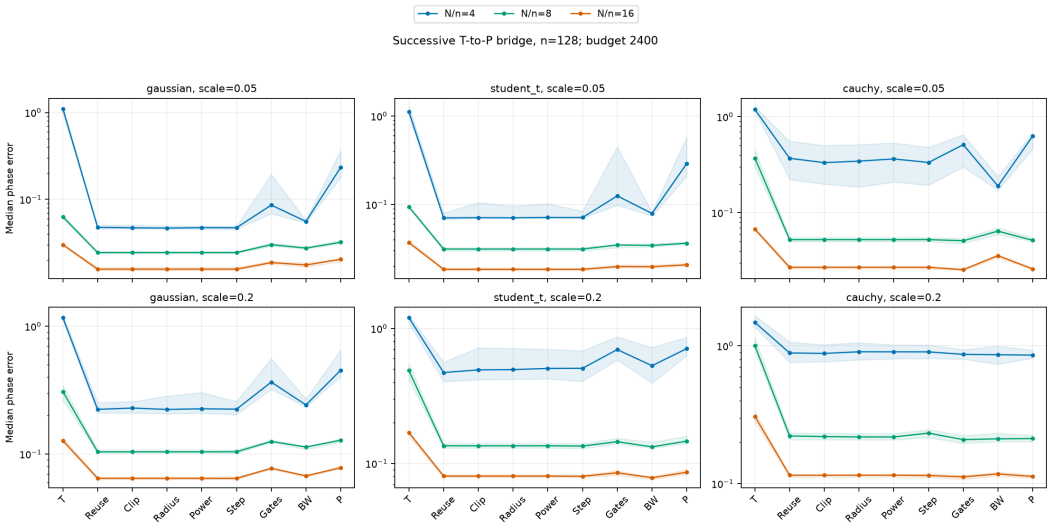


Figure 3: Complete sequential bridge at $n = 128$, budget 2,400. Bands are pointwise 95% bootstrap intervals for cell medians. BW is the fixed scaled-bandwidth node B7.

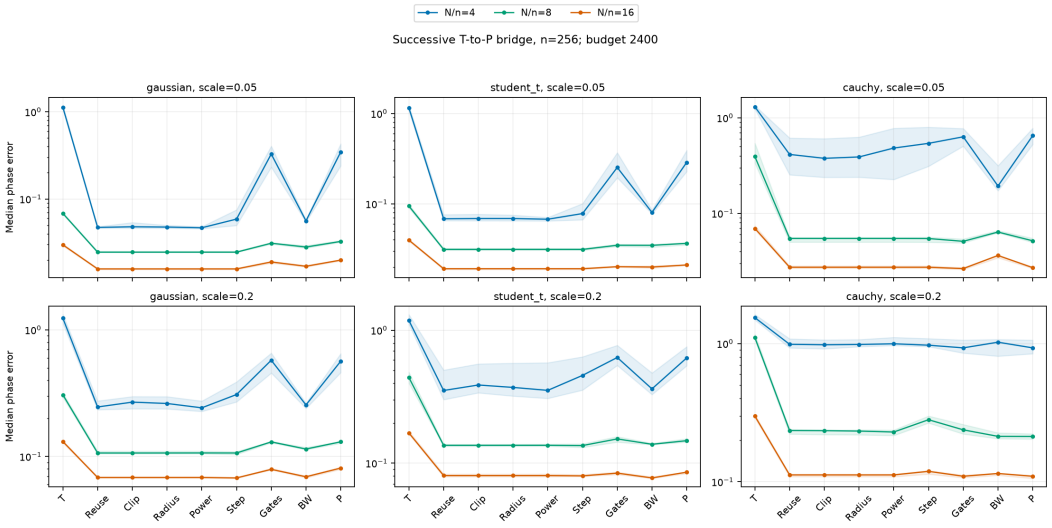


Figure 4: Complete sequential bridge at $n = 256$; same conventions as above.

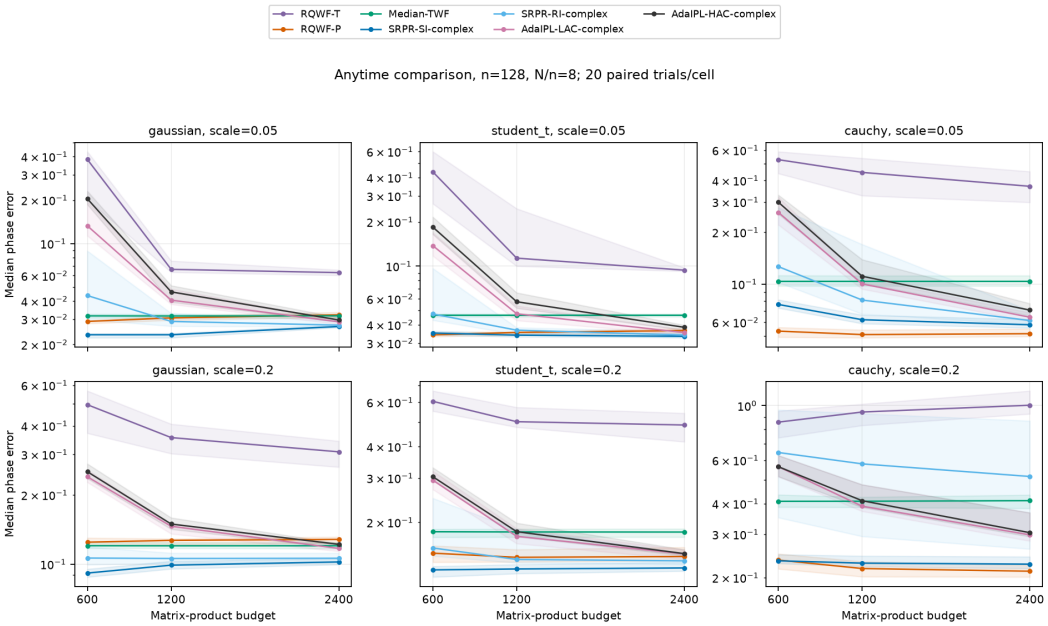


Figure 5: Budget–error curves at $n = 128$, $N/n = 8$. Points use the last committed iterate, not the best error seen so far. Bands are pointwise 95% bootstrap intervals. Other ratios remain in the released cell summaries.

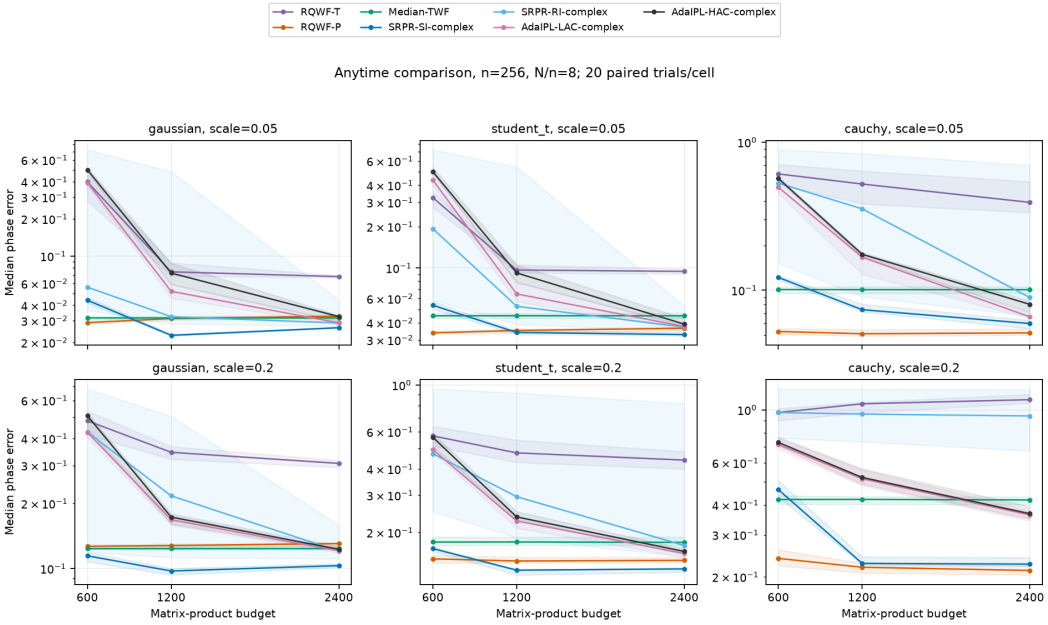


Figure 6: Budget–error curves at $n = 256$, $N/n = 8$; same conventions as above.

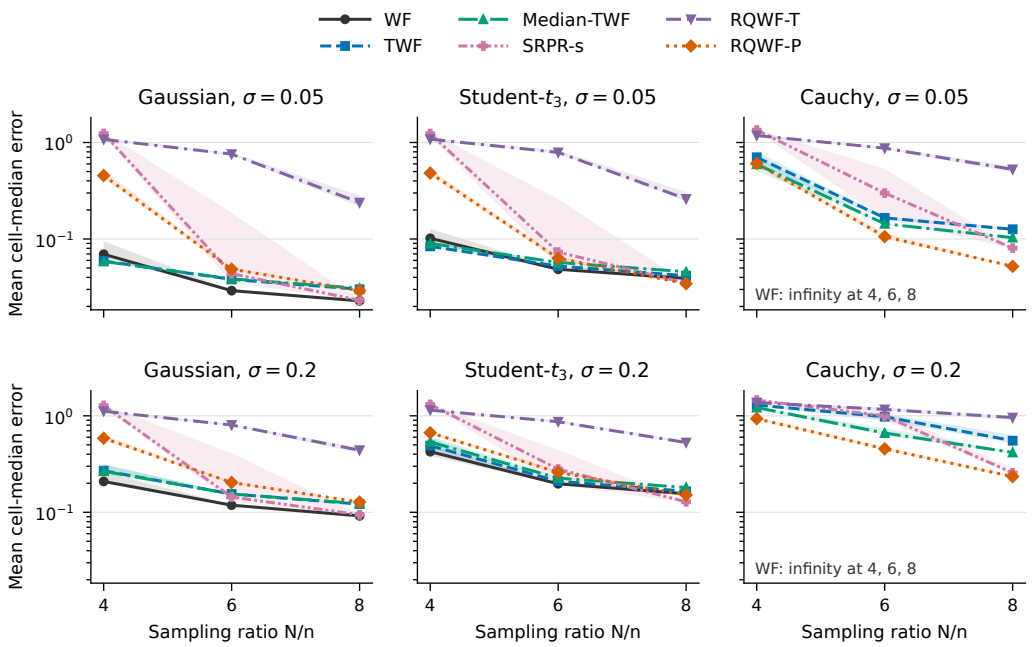


Figure 7: Earlier fixed-scale v2 grid at its optimization budget of 600, which excludes Gram construction. Each point averages three dimension-specific cell medians (50 paired problems each); shading gives pointwise 95% stratified paired-bootstrap intervals. All 54 dense-noise cells and all methods are included; infinite WF summaries are annotated. SRPR-s is a smooth-only complex adaptation, not the complete pipeline.

D.3 Fixed-scale protocol and diagnostics

The fixed-scale v2 suite contains 90 complete cells, each with 50 paired problems. The dense-noise grid has 54 cells (2,700 problems, six methods); the noiseless/small-noise controls have 12 cells (600 problems, six methods); quantile matching has eight cells (400 problems, two versions and two fitted quantiles); sample-size and density diagnostics each have eight cells (400 problems, two versions). Thus there are 4,500 problems and 23,000 fits. Results are not pooled with the historical archives below. They are also not pooled with the new matched-budget study in Appendix D.1, whose work accounting includes Gram construction. This earlier suite retains the smooth-only SRPR-s comparator and the original 600 optimization-call cap.

All sensing rows are independent proper complex Gaussians, and the true signal is an isotropic complex unit vector. Noise is independent of sensing. For asymmetric Laplace noise, draw $u \sim \mathrm{Unif}(0,1)$ and set $\varepsilon = \sigma \log(u/\tau)/(1-\tau)$ for $u < \tau$, and $\varepsilon = -\sigma \log((1-u)/(1-\tau))/\tau$ otherwise. Then $Q_\tau(\varepsilon) = 0$ and $f(0) = \tau(1-\tau)/\sigma$. Lognormal noise is $\sigma[\exp(Z) - \exp\{\Phi^{-1}(\tau)\}]$, $Z \sim N(0,1)$, with $f(0) = \phi(q)/(\sigma e^q)$, $q = \Phi^{-1}(\tau)$. No fitted quantile is chosen using reconstruction error. Quantile and scaling studies use 1,000 updates, with a predeclared 500-update checkpoint, not the better of two truth errors.

The RQWF-T implementation forms the clipped Hermitian matrix on D, computes its full eigendecomposition, selects the largest eigenvalue, and checks decomposition residual and orthogonality. These floating-point diagnostics are not interval-arithmetic certificates. Radial bisection on V uses KKT tolerance 10^{-9}. The shared bandwidth is $N^{-1/2}$ and step is 0.02. The independent $D/V/U$ fractions are $0.4/0.2/0.4$, with fixed cap 4 and no gates or annealing. The theorem's step and basin conditions are not certified at the tested ratios. RQWF-P uses 50 spectral power steps, 0.95 empirical clipping capped at $9\,\mathrm{median}(y_+)/\log 2$, median-based scale, amplitude gates $[0.3, 5]$, 0.95 residual trimming, update coefficient 0.25, and annealing factor $\max\{10^{-8}, 0.25/(1+t/150)^2\}$ multiplying the frozen scale. WF, TWF, and Median-TWF use 50 power steps and the declared complex adaptations of their updates; all settings are in the archived configuration.

The optimization budget sums the number of rows processed by each forward or adjoint operation divided by total N. Initialization, radial projections, rejected line-search trials, and incomplete last steps consume budget; the returned iterate is the last complete one. Full matrix construction, eigendecomposition, and scalar solving are timed separately. Offline diagnostic calls are also separately counted and included in total time, but not in the 600 optimization-call cap. Shared-machine parallel execution prevents strong wall-time speed claims. For RQWF-P, logged smooth loss and gradient are ungated diagnostics at the current bandwidth, not a fixed objective for its gated, annealed update.

The smooth-only SRPR-s comparator uses complex isotropic random direction and an independent $\mathrm{Unif}(0, 4)$ radius, smoothing 0.25, and monotone accelerated proximal-gradient steps with both extrapolation terms retained. Backtracking uses real gradient equal to twice the Wirtinger derivative, initial step 0.05, shrink factor 0.5, and Armijo coefficient 10^{-4}. This is a declared complex smooth-stage adaptation of the random-initialized method in Zheng & Xue (2024), not an exact reproduction of its unspecified implementation choices. Spectral initialization and the subsequent IPL corruption stage are absent. Neither this comparator nor the historical locally implemented SRPR variant establishes superiority to the complete published pipeline.

For failure accounting, nonfinite or overflow runs, including the norm safeguard 10^{100}, receive error $+\infty$. Large finite errors remain in every summary. The 90th percentile is an order statistic (inverse empirical CDF), avoiding interpolation through infinity. All 672 numerical failures occur in WF under Cauchy noise: 265/450 at scale 0.05 and 407/450 at scale 0.20. Within-cell intervals resample the same 50 problem keys jointly across methods/quantiles, 2,000 times. Figure 7 independently resamples within each dimension and averages the three cell medians. Intervals are pointwise, not simultaneous comparisons; an observed 50/50 win count is not a population guarantee. Raw records retain problem/seed identifiers, shared data hashes, source and configuration hashes, initialization, operator counts, numerical diagnostics, failure reasons, and saved trajectories. The frozen source snapshot, not an uncommitted-worktree label alone, identifies the executed implementation. Every reported median and 90th percentile was recomputed from all raw method records, and pairing was checked by per-problem data hashes.

An independent 4,096-row test sample per problem is shared across its four fits. Coverage estimates $\mathbb{P}(Y \le |a^*\widehat{x}|^2)$. Centered target check loss averages $\rho_\tau(Y - |a^*\widehat{x}|^2) - \rho_\tau(Y - |a^*x_\star|^2)$; centering uses the true signal only for evaluation, never optimization. Table 8 reports means across the 50 problems. Closer target coverage, lower target check loss, and lower phase error are distinct criteria, as the upper-target lognormal cases demonstrate.

Table 8: Independent-test metrics for every quantile-matching cell. P/T denote RQWF-P/T; loss differences are relative to the true signal.

Noise	τ	N/n	Version	Test coverage		Loss difference
				Matched	Median	Matched / median
ALD	0.25	8	P	0.291	0.396	0.0416 / 0.0757
ALD	0.25	8	T	0.371	0.464	0.2179 / 0.3418
ALD	0.25	32	P	0.261	0.384	0.0047 / 0.0215
ALD	0.25	32	T	0.294	0.408	0.0157 / 0.0444
ALD	0.75	8	P	0.689	0.604	0.0316 / 0.0521
ALD	0.75	8	T	0.627	0.583	0.2207 / 0.1570
ALD	0.75	32	P	0.742	0.623	0.0047 / 0.0183
ALD	0.75	32	T	0.716	0.610	0.0145 / 0.0325
Lognormal	0.25	8	P	0.299	0.394	0.0078 / 0.0189
Lognormal	0.25	8	T	0.366	0.455	0.1622 / 0.2271
Lognormal	0.25	32	P	0.258	0.380	0.0010 / 0.0060
Lognormal	0.25	32	T	0.285	0.403	0.0035 / 0.0131
Lognormal	0.75	8	P	0.670	0.592	0.0206 / 0.0194
Lognormal	0.75	8	T	0.588	0.539	0.2620 / 0.1250
Lognormal	0.75	32	P	0.720	0.619	0.0030 / 0.0083
Lognormal	0.75	32	T	0.716	0.624	0.0095 / 0.0111

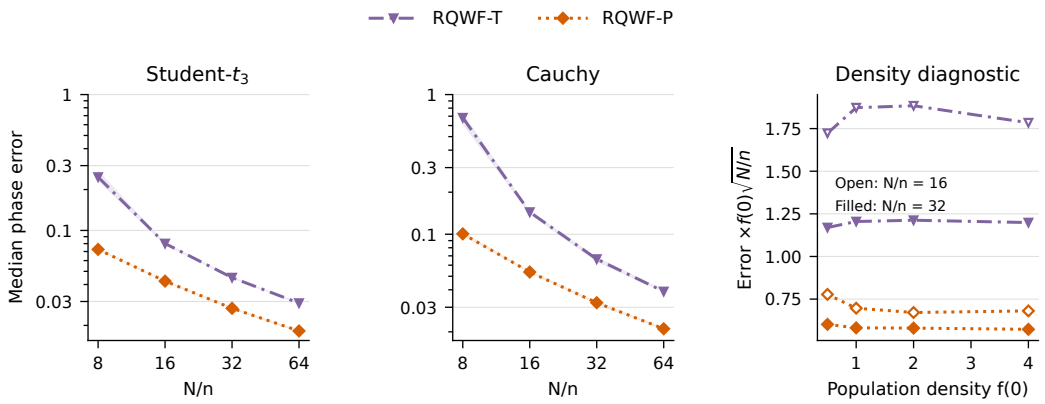


Figure 8: Squared-intensity scaling diagnostics at the fixed 1,000-update endpoint, $n = 128$. Left/middle: fixed scale 0.10, 50 problems per point, with pointwise 95% bootstrap intervals. Right: Cauchy scales $\sigma = 1/(\pi f(0))$ for $f(0) = 0.5, 1, 2, 4$; normalized cell medians. These are empirical diagnostics, not fitted or certified theoretical rates.

At ratio 8, RQWF-T's Student-t_3 median improves from 0.428 to 0.247 between 500 and 1,000 updates, with 48/50 problems improving by over 10%. Under Cauchy noise at ratio 16 it improves from 0.161 to 0.144 (27/50). At ratio 64, the respective 1,000-update median gradient norms are 1.22×10^{-8} and 7.88×10^{-7}, versus 0.0259 and 0.0284 at ratio 8. A small gradient is not a global-optimality certificate. The low-ratio results cannot be interpreted as purely statistical error; the normalized density plot likewise does not isolate density from all other distributional constants. Complete 500/1,000 diagnostics and per-cell intervals accompany the raw source-indexed results.

D.4 FULL GRIDS AND CONTROLS

Each table retains every dimension, ratio, scale, and method. Parentheses contain 90th-percentile error; superscripts count numerical failures among 50 runs, already included as $+\infty$ in both error summaries.

Table 9: Fixed-scale Gaussian: median relative error (90th percentile), 50 trials per cell. Superscripts give failed runs; failures contribute $+\infty$. SRPR-s is the smooth-stage complex RI adaptation.

n	N/n	σ	WF	TWF	Median-TWF	SRPR-s	RQWF-T	RQWF-P
128	4	0.05	0.0553 (0.892)	0.0586 (0.95)	0.0575 (0.677)	1.08 (1.36)	1.01 (1.27)	0.389 (0.79)
128	6	0.05	0.029 (0.0311)	0.0377 (0.0404)	0.0383 (0.0405)	0.0308 (0.21)	0.693 (1.05)	0.0488 (0.074)
128	8	0.05	0.0229 (0.0248)	0.0297 (0.0323)	0.0306 (0.0333)	0.0231 (0.0255)	0.226 (0.471)	0.0291 (0.0321)
256	4	0.05	0.0634 (0.584)	0.0601 (0.641)	0.0592 (0.762)	1.26 (1.47)	1.06 (1.22)	0.461 (0.824)
256	6	0.05	0.0293 (0.0312)	0.0384 (0.0406)	0.039 (0.0409)	0.0382 (1.21)	0.781 (1.03)	0.0485 (0.0709)
256	8	0.05	0.0228 (0.0241)	0.0301 (0.0324)	0.0308 (0.0328)	0.0231 (0.0261)	0.241 (0.467)	0.0291 (0.031)
512	4	0.05	0.09 (1.02)	0.0586 (0.771)	0.0587 (0.894)	1.37 (1.48)	1.15 (1.26)	0.513 (0.767)
512	6	0.05	0.0291 (0.0304)	0.0383 (0.0396)	0.0385 (0.04)	0.0624 (1.26)	0.799 (0.945)	0.0489 (0.0662)
512	8	0.05	0.0229 (0.0239)	0.0306 (0.032)	0.0311 (0.0326)	0.0232 (0.0291)	0.248 (0.447)	0.0291 (0.0303)
128	4	0.2	0.192 (0.656)	0.247 (0.882)	0.232 (0.883)	1.17 (1.42)	1.05 (1.17)	0.504 (0.797)
128	6	0.2	0.12 (0.13)	0.156 (0.169)	0.155 (0.169)	0.126 (0.853)	0.744 (0.989)	0.196 (0.247)
128	8	0.2	0.0923 (0.0973)	0.121 (0.128)	0.123 (0.132)	0.0938 (0.0994)	0.406 (0.697)	0.128 (0.139)
256	4	0.2	0.203 (0.738)	0.268 (0.859)	0.258 (0.849)	1.3 (1.41)	1.11 (1.25)	0.624 (0.771)
256	6	0.2	0.117 (0.126)	0.154 (0.162)	0.154 (0.164)	0.125 (1.11)	0.837 (1.07)	0.206 (0.243)
256	8	0.2	0.0915 (0.0968)	0.121 (0.126)	0.123 (0.128)	0.0942 (0.101)	0.444 (0.712)	0.127 (0.135)
512	4	0.2	0.231 (0.909)	0.294 (0.921)	0.309 (0.844)	1.36 (1.48)	1.18 (1.33)	0.628 (0.801)
512	6	0.2	0.118 (0.124)	0.155 (0.163)	0.155 (0.162)	0.181 (1.32)	0.829 (1.03)	0.209 (0.243)
512	8	0.2	0.0915 (0.0958)	0.122 (0.127)	0.124 (0.129)	0.0951 (0.194)	0.469 (0.669)	0.128 (0.134)

Table 10: Fixed-scale Student-t_3: median relative error (90th percentile), 50 trials per cell. Superscripts give failed runs; failures contribute $+\infty$. SRPR-s is the smooth-stage complex RI adaptation.

n	N/n	σ	WF	TWF	Median-TWF	SRPR-s	RQWF-T	RQWF-P
128	4	0.05	0.0912	0.084	0.0906	1.13	1.04	0.456
			(0.401)	(0.313)	(0.391)	(1.43)	(1.19)	(0.81)
128	6	0.05	0.0476	0.0518	0.0564	0.046	0.735	0.0612
			(0.0552)	(0.0566)	(0.0641)	(0.0576)	(1.1)	(0.0879)
128	8	0.05	0.039	0.0417	0.0458	0.0352	0.249	0.0347
			(0.0455)	(0.0446)	(0.0491)	(0.039)	(0.566)	(0.0374)
256	4	0.05	0.0901	0.0822	0.0878	1.21	1.07	0.445
			(0.197)	(0.876)	(0.713)	(1.47)	(1.19)	(0.684)
256	6	0.05	0.0487	0.0526	0.0577	0.0495	0.797	0.0644
			(0.0546)	(0.0568)	(0.0632)	(1.21)	(1.02)	(0.0937)
256	8	0.05	0.0394	0.0417	0.0458	0.0353	0.258	0.0344
			(0.0444)	(0.0441)	(0.0489)	(0.0396)	(0.63)	(0.0369)
512	4	0.05	0.124	0.0853	0.0919	1.37	1.15	0.546
			(1.06)	(0.959)	(0.928)	(1.47)	(1.3)	(0.793)
512	6	0.05	0.0494	0.0527	0.0572	0.124	0.838	0.064
			(0.0541)	(0.0547)	(0.0597)	(1.24)	(1.02)	(0.0759)
512	8	0.05	0.0387	0.0415	0.0457	0.0346	0.274	0.034
			(0.0429)	(0.0429)	(0.0479)	(0.0376)	(0.492)	(0.0361)
128	4	0.2	0.328	0.381	0.393	1.25	1.11	0.63
			(0.724)	(0.91)	(0.882)	(1.42)	(1.33)	(0.846)
128	6	0.2	0.192	0.208	0.227	0.178	0.815	0.247
			(0.252)	(0.229)	(0.252)	(0.737)	(1.08)	(0.397)
128	8	0.2	0.154	0.166	0.181	0.128	0.517	0.153
			(0.188)	(0.181)	(0.2)	(0.146)	(0.745)	(0.176)
256	4	0.2	0.381	0.458	0.519	1.32	1.14	0.625
			(0.86)	(0.859)	(0.825)	(1.45)	(1.3)	(0.783)
256	6	0.2	0.201	0.21	0.226	0.184	0.905	0.256
			(0.238)	(0.221)	(0.242)	(1.3)	(1.17)	(0.333)
256	8	0.2	0.157	0.165	0.179	0.128	0.519	0.15
			(0.175)	(0.174)	(0.195)	(0.137)	(0.859)	(0.163)
512	4	0.2	0.567	0.634	0.695	1.37	1.19	0.749
			(0.991)	(0.987)	(1.03)	(1.49)	(1.31)	(0.959)
512	6	0.2	0.199	0.212	0.226	0.478	0.884	0.276
			(0.221)	(0.221)	(0.244)	(1.33)	(1.08)	(0.328)
512	8	0.2	0.155	0.165	0.182	0.129	0.549	0.15
			(0.181)	(0.171)	(0.189)	(0.203)	(0.755)	(0.158)

Table 11: Fixed-scale Cauchy: median relative error (90th percentile), 50 trials per cell. Superscripts give failed runs; failures contribute $+\infty$. SRPR-s is the smooth-stage complex RI adaptation.

n	N/n	σ	WF	TWF	Median-TWF	SRPR-s	RQWF-T	RQWF-P
128	4	0.05	∞^{25}	0.497	0.497	1.28	1.12	0.547
			(∞)	(0.945)	(0.999)	(1.5)	(1.32)	(0.866)
128	6	0.05	0.741^{17}	0.157	0.146	0.117	0.804	0.11
			(∞)	(0.269)	(0.174)	(0.152)	(1)	(0.25)
128	8	0.05	0.623^{12}	0.119	0.102	0.0807	0.439	0.0518
			(∞)	(0.184)	(0.113)	(0.0906)	(0.712)	(0.0603)
256	4	0.05	∞^{29}	0.704	0.527	1.33	1.19	0.638
			(∞)	(1.21)	(1.19)	(1.44)	(1.3)	(0.905)
256	6	0.05	∞^{26}	0.169	0.143	0.166	0.886	0.103
			(∞)	(0.266)	(0.157)	(1.34)	(1.13)	(0.179)
256	8	0.05	∞^{30}	0.128	0.103	0.0796	0.575	0.0528
			(∞)	(0.168)	(0.116)	(0.0853)	(0.843)	(0.0572)
512	4	0.05	∞^{46}	0.908	0.761	1.42	1.23	0.633
			(∞)	(1.36)	(1.25)	(1.5)	(1.39)	(0.786)
512	6	0.05	∞^{42}	0.17	0.142	0.614	0.932	0.104
			(∞)	(0.84)	(0.157)	(1.27)	(1.13)	(0.162)
512	8	0.05	∞^{38}	0.132	0.105	0.0813	0.565	0.0516
			(∞)	(0.171)	(0.114)	(0.0857)	(0.83)	(0.0565)
128	4	0.2	∞^{37}	1.17	1.11	1.36	1.33	0.905
			(∞)	(1.44)	(1.34)	(1.55)	(1.52)	(1.12)
128	6	0.2	∞^{38}	0.868	0.609	0.715	1.1	0.435
			(∞)	(1.28)	(1.09)	(1.35)	(1.38)	(0.705)
128	8	0.2	∞^{38}	0.492	0.405	0.239	0.898	0.231
			(∞)	(0.793)	(0.478)	(0.273)	(1.11)	(0.288)
256	4	0.2	∞^{49}	1.32	1.24	1.48	1.37	0.94
			(∞)	(1.47)	(1.43)	(1.57)	(1.5)	(1.1)
256	6	0.2	∞^{48}	0.907	0.653	0.974	1.16	0.441
			(∞)	(1.27)	(1.18)	(1.32)	(1.35)	(0.68)
256	8	0.2	∞^{47}	0.554	0.423	0.259	0.977	0.233
			(∞)	(1.15)	(0.736)	(1.17)	(1.22)	(0.314)
512	4	0.2	∞^{50}	1.39	1.3	1.53	1.4	0.958
			(∞)	(1.48)	(1.45)	(1.61)	(1.5)	(1.16)
512	6	0.2	∞^{50}	1.16	0.727	1.3	1.24	0.489
			(∞)	(1.37)	(1.3)	(1.44)	(1.38)	(0.749)
512	8	0.2	∞^{50}	0.61	0.423	0.272	1	0.24
			(∞)	(1.34)	(0.467)	(1.02)	(1.15)	(0.273)

For the clean and small-noise controls, Table 12 retains all 12 negative-control cells. All six methods returned finite estimates on these 600 problems. Large errors at low sampling and residual optimization error remain visible; finite execution is not equated with successful recovery.

Table 12: Negative controls: median relative phase error, 50 paired problems per cell, optimization budget 600. No numerical failures occurred.

n	N/n	WF	TWF	Median-TWF	SRPR-s	RQWF-T	RQWF-P
Noiseless							
128	4	0.0356	0.00122	5.6×10^{-4}	1.15	1.03	0.421
128	6	8.5×10^{-4}	3.7×10^{-8}	1.5×10^{-8}	0.00652	0.696	0.00942
128	8	5.7×10^{-5}	2.4×10^{-11}	8.6×10^{-12}	5.3×10^{-4}	0.175	2.9×10^{-4}
256	4	0.0532	0.00343	0.0022	1.33	1.06	0.498
256	6	0.00107	6.0×10^{-8}	2.1×10^{-8}	0.0239	0.751	0.00994
256	8	6.0×10^{-5}	2.5×10^{-11}	9.1×10^{-12}	8.9×10^{-4}	0.197	3.1×10^{-4}
512	4	0.093	0.011	0.00569	1.34	1.13	0.474
512	6	0.00101	5.8×10^{-8}	2.0×10^{-8}	0.0386	0.823	0.013
512	8	7.1×10^{-5}	3.2×10^{-11}	1.2×10^{-11}	0.00177	0.231	3.6×10^{-4}
Gaussian scale 10^{-4}							
256	4	0.0429	0.00295	0.00163	1.24	1.07	0.5
256	6	0.00102	7.6×10^{-5}	7.7×10^{-5}	0.0346	0.742	0.0117
256	8	7.7×10^{-5}	6.2×10^{-5}	6.2×10^{-5}	7.9×10^{-4}	0.265	3.5×10^{-4}

The fixed smoothing, sample split, and conservative initialization of RQWF-T differ substantially from the practical solver, so these controls do not isolate a single causal mechanism for their performance gap. Conversely, the noiseless model does not have the positive local noise density assumed by the main theorem. Its role here is a computational negative control, not a test of that theorem's assumptions.

E Real-Data Experiments

E.1 Data and evaluation

The measurement archive uses the `AmpSLM_16x16` subset of Rice's coherent transmission-matrix collection (Metzler et al., 2017). Its loader reads `XH_train.mat` and `YH_squared_train.mat` as unsigned 8-bit values, transposes the recorded input patterns, and divides the input values by 255. The input dimension is 256. The first 2,048 patterns are used for fitting and the next 1,024 for evaluation; this is a fixed sequential split, not a randomized acquisition split. Fifty fixed detector indices are obtained by rounding 50 equally spaced values from 512 to 65,024. They share the same input patterns and acquisition, and are not assumed to be independent replications.

For each detector and each of 12 noise/SNR conditions, a Gaussian, Student-t_3, or Cauchy vector e is generated and transformed to

$$\widetilde{e} = \frac{\|y_{\text{fit}}^{\text{meas}}\|_2}{10^{\text{SNR}/20}} \frac{e}{\|e\|_2}, \qquad y_{\text{fit}} = y_{\text{fit}}^{\text{meas}} + \widetilde{e}.$$

The SNR grid is 15, 25, 35, and 45 dB. The per-realization normalization couples the injected coordinates and does not preserve the fixed-scale i.i.d. noise model used by the theorem or the new synthetic study. All four methods share each perturbed training vector. The deterministic seed is $20260923 + 17p + 1009s + 100003j$, where p is detector index, s is SNR in dB, and $j = 1, 2, 3$ indexes the noise family above. Measured validation responses are not additionally perturbed, but their native acquisition noise and possible model mismatch remain.

For an estimate z, the reported endpoint is

$$E_{\text{val}}(z) = \frac{\left\| |A_{\text{val}} z|^2 - y_{\text{val}}^{\text{meas}} \right\|_2}{\|y_{\text{val}}^{\text{meas}}\|_2},$$

with the implementation's 10^{-12} denominator floor. We summarize the median across the 50 detector pixels, not phase error, image PSNR, or error relative to an independently measured complex transmission vector. The archive's word "clean" denotes absence of the additional simulated noise; it does not imply noise-free physical observations.

E.2 Implementation details

The suffix RD is introduced only to distinguish the archived algorithms from the synthetic T/P recipes and from canonical published baselines. No fit is changed or rerun in the integration. Each method uses 40 power iterations initialized at $\mathbf{1}/\sqrt{n}$. WF-RD uses positive response weights; the other methods cap these at their empirical 0.95 quantile. With direction v, the squared initial radius is $\sum_i |a_i^T v|^2 w_i / \sum_i |a_i^T v|^4$ with numerical floors. This weighted least-squares scale is not T's quantile radial solve or P's median scale.

Write $u = Az$, $r = |u|^2 - y$, and $b_0 = \max\{\text{median}(y), 1\}$. The common gradient form is $A^*(q \odot u)/N$, where

$$q_i = \begin{cases} r_i & \text{WF-RD,} \\ r_i \mathbf{1}\{|r_i| \le 5 \,\text{mean}\, |r|\} & \text{TWF-RD,} \\ r_i \mathbf{1}\{|r_i| \le 12 \,\text{median}\, |r|\} & \text{Median-TWF-RD,} \\ -b_0 \psi_{1/2,\delta_t}(-r_i) & \text{RQWF-RD,} \end{cases} \qquad \delta_t = b_0 \max\{0.02, 0.25/(1+t/50)^2\}.$$

The TWF-RD labels therefore designate local residual-trimmed WF recipes, not verified reproductions of canonical TWF or Median-TWF. Backtracking starts at $0.1/b_0$ for the three reference recipes and $0.5/b_0$ for RQWF-RD, halving up to 12 times. The reference recipes accept against squared training residual, whereas RQWF-RD accepts against the unsmoothed median check loss. Failure to accept terminates fitting. Only training criteria drive the update and termination in the supplied code. Iteration ceilings are 2,500 for WF-RD/RQWF-RD and 1,000 for the other two; neither operator work nor time is matched. Recorded median attempted iteration counts are 2,500, 62.5, 504.5, and 1,298, respectively. The TWF-RD recipe reaches its ceiling in none of the 600 fits, so its prediction error must be interpreted with this early termination behavior.

The input matrix and power initialization are real, and the updates preserve the real subspace. Declaring the arrays complex does not remove this invariant. Consequently the study evaluates a restricted intensity predictor and cannot demonstrate identification of a general complex transmission vector. Moreover real sensing entails conjugation ambiguity for a general complex signal. The physical quantile model is not calibrated: $\tau = 1/2$ specifies the robust training loss, not a verified quantile of the acquisition process. Native-noise-only controls, independent acquisitions, and matched-compute SRPR/AdaIPL comparisons are absent from this archive.

E.3 Results and validation

Table 13 retains all 12 conditions and all four methods, including the Student-t_3, 15 dB condition where Median-TWF-RD has lower median error than RQWF-RD. The latter has lower medians than WF-RD and TWF-RD in 12/12 conditions and than Median-TWF-RD in 11/12. Its Cauchy medians remain near 0.33–0.34 across SNR levels, so the results do not establish consistent signal recovery as injected noise vanishes. We report descriptive medians and paired detector wins, without independence-based significance claims across the 50 pixels.

Table 13: All measured-data prediction results: median held-out relative intensity RMSE across 50 fixed detector pixels. RD methods are defined above and are not compute-matched.

Noise	SNR (dB)	WF-RD	TWF-RD	Med.-TWF-RD	RQWF-RD
Gaussian	15	0.35238	0.40385	0.35147	0.34643
Gaussian	25	0.34426	0.41145	0.34375	0.33628
Gaussian	35	0.34355	0.40608	0.34279	0.34031
Gaussian	45	0.34263	0.41207	0.34224	0.33830
Student-t_3	15	0.34836	0.40483	0.34691	0.34763
Student-t_3	25	0.34334	0.41631	0.34183	0.33613
Student-t_3	35	0.34265	0.41064	0.34229	0.33926
Student-t_3	45	0.34261	0.41196	0.34224	0.33683
Cauchy	15	0.35736	0.41945	0.35372	0.33278
Cauchy	25	0.34556	0.41538	0.34590	0.33882
Cauchy	35	0.34281	0.40640	0.34248	0.33379
Cauchy	45	0.34276	0.41119	0.34238	0.33388

The supplied archive contains R source, 2,400 fit records (600 paired pixel/condition problems), 273,985 trace rows, 48 method-level summary rows, metadata, and its original plot. Checks establish unique complete keys, identical per-condition noise metadata across methods, declared dimensions and ceilings, finite metrics, monotone trace indices, and SNR consistency from the logged norms. All three supplied validation metrics are reaggregated: the largest absolute discrepancy from the supplied summary is 5.56×10^{-16}. The original raw MAT files and fitted vectors are not included in this archive. Thus these checks establish consistency of the supplied logs, not independent regeneration of the measurements or refitting of the algorithms. Original source files and their hashes are retained separately from the redrawn manuscript figure and derived tables.

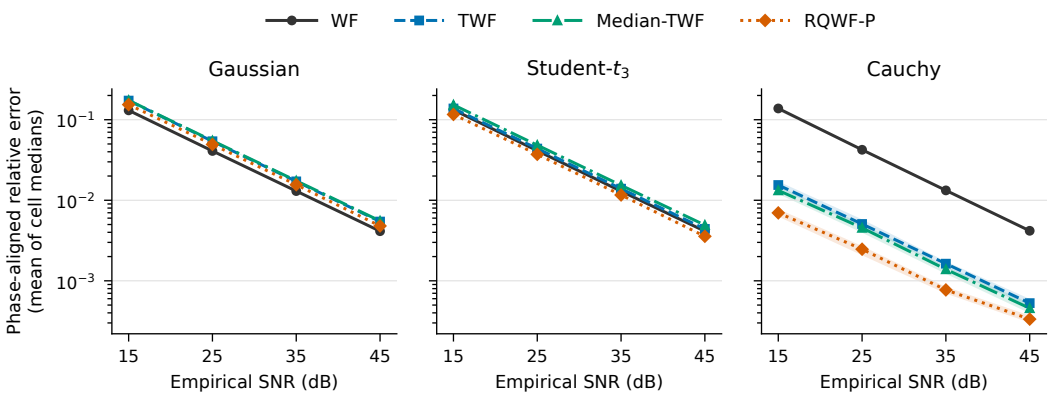


Figure 9: Historical SNR-normalized recovery; unequal iteration ceilings. These observations are separate from the fixed-scale main comparison.

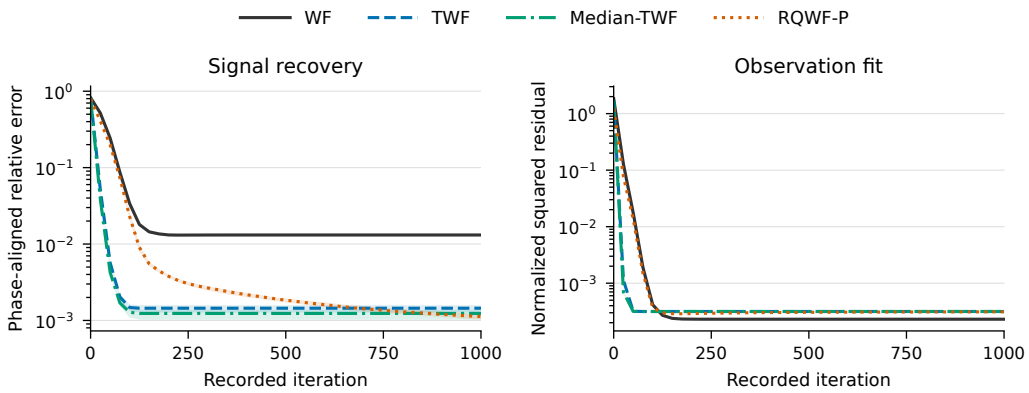


Figure 10: Historical Cauchy trajectories: common iterations, not common operator work. Squared residual is an observation-fit diagnostic, not every method's training objective.

F Supplementary Results

F.1 Historical experiments

This section concerns earlier, separately configured experiments, not the fixed-scale v2 suite in the main text. Most archived intensity suites set

$$\varepsilon = \xi \frac{\|(|a_i^* x_\star|^2)_{i=1}^m\|_2}{\|\xi\|_2 \, 10^{\mathrm{SNR}/20}}. \tag{48}$$

This empirical energy-SNR normalization introduces a shared random scale depending on clean responses, unlike the independent fixed-scale noise in our theorem and new experiments. Cross-law rankings are protocol-specific.

Tables 14 and 15 preserve the original numerical summaries, now represented by Figures 9 and 10 in this appendix. The figure script reproduces every entry to the reported six decimal places. Table 14 averages nine cell medians at 25 dB, each from 50 trials; WF/RQWF-P use 2,500-step ceilings and TWF/Median-TWF use 1,000. Table 15 pools 450 paired Cauchy problems at iteration 1,000. Its normalized squared residual is $\sum_i (y_i - |a_i^* z_k|^2)^2 / \sum_i |a_i^* x_\star|^4$. These are accuracy checkpoints, not equal-compute comparisons.

For the bootstrap intervals in Figure 9, each replicate draws 50 trial keys with replacement within each of the nine fixed cells, shares the draw across the four methods, and averages the nine resulting medians. Cells are not resampled or weighted by their error. For Figure 10, each replicate draws

Table 14: Recovery error at 25 dB.

Noise	WF	TWF	Median-TWF	RQWF-P
Gaussian	0.040946	0.054026	0.054850	0.049350
Student-t_3	0.041246	0.043837	0.048351	0.037168
Cauchy	0.042356	0.005074	0.004543	0.002468

Table 15: Cauchy recovery at 1,000 iterations.

Method	Phase error	Normalized squared residual
WF	0.013134	0.000231
TWF	0.001446	0.000316
Median-TWF	0.001236	0.000316
RQWF-P	0.001124	0.000312

450 problem keys with replacement from the pooled study and retains each selected problem's entire recorded trajectory, both diagnostics, and all four methods. Thus pairing is preserved across methods, time, and diagnostics. Each figure uses its own NumPy PCG64 stream initialized at seed 20260914 and 2,000 replicates; the bands are the 2.5th and 97.5th percentiles at each plotted point, not simultaneous bands or a multiple-comparison significance test.

The script checks $(n, m, m/n, \mathrm{SNR}, \mathrm{trial}, \mathrm{seed})$, shared noise metadata, unique method/time records, and finite outcomes. The common noise slice has 1,800 problems per noise law, with no missing method records. All 450 trajectory problems have every common 25-step record through 1,000. The duplicated method-name column agrees exactly. The noise archive copies differ only in text serialization and contain identical parsed records; the full trajectory copies are byte-identical. Only one source per study is analyzed. The two figures use existing observations only; bootstrap resampling is not a new recovery simulation.

F.2 Archive coverage and limitations

The following discussion concerns historical implementations, not the new structure-aligned pilot in Appendix F.3. RQWF-P denotes a family, not a single frozen solver. The intensity grids and trajectory archive use their supplied configurations; the repository follow-up studies use a separate configurable implementation. Their result files and configurations must be identified separately. The legacy code name `run_rqwf_theory` does not designate RQWF-T: its clipping threshold is empirical, radial and update bandwidths depend on data and can differ, and its finite power iteration and bounded radial search do not provide the certificates in Appendix B. Its optional truth-based stopping also differs from a fixed iteration count.

The production archives use all-data spectral initialization, amplitude gates, optional residual gates, and/or annealing. In the archived intensity RQWF-P, $q_{\mathrm{trim}} = 1$ disables residual-quantile trimming while retaining amplitude gates. A noisy marginal response quantile is not an exact signal-radius formula. Data-selected clipping and bandwidth, a fixed number of power iterations, and iterate-dependent gates require separate analysis; they are not covered by Algorithm 1's theorem.

The additional Student-t_3 comparison with SRPR contains 2,820 paired problems and 14,100 method records, including dimensions up to 2,048. It is not a complete five-dimension grid: $n = 2048$, $m/n \in \{12, 16\}$ are absent. All methods use truth-based early stopping at phase error 0.045, checked once every 25 iterations. At $m/n = 4$, archived SRPR reaches that threshold in 351/360 problems (97.50%), compared with 207/360 (57.50%) for RQWF-P. The archived SRPR baseline is a local pseudo-Huber gradient method with backtracking, not a verified reproduction of the full published solver. It therefore does not support a claim of superiority to SRPR.

Table 16 counts parameter/noise configurations with trial keys after within-suite deduplication. The counts do not imply independence across suites; another archive copy is not another problem. The original four-method intensity grid uses 50 trials per cell. Noiseless and Gaussian experiments have ratios 3–8; the heavy-tail grid has ratios 6, 7, 8. Noisy grids use 15, 25, 35, and 45 dB. WF and RQWF-P have 2,500-step ceilings, whereas TWF and Median-TWF have 1,000-step ceilings. The

Table 16: Archive coverage.

Suite	Problems	Method records	Audit status
Noiseless intensity	900	3,600	Complete supplied grid
Gaussian intensity	3,600	14,400	Complete supplied grid
Student-t_3 intensity	1,800	7,200	Complete supplied grid
Cauchy intensity	1,800	7,200	Complete supplied grid
Cauchy plus positive outliers	1,350	5,400	Complete supplied grid
Noiseless plus positive outliers	1,775	7,100	Partial extension
Student-t_3 with SRPR	2,820	14,100	Missing cells; 43 errors
Cauchy loss trajectories	450	1,800	Duplicate ZIP copy removed
Repository follow-up studies	190	620	Small matched studies

separate SRPR suite changes RQWF-P's ceiling to 1,000 and uses threshold 0.045 for truth-based stopping. These protocols must not be pooled as one algorithmic comparison.

The two supplied full trajectory files have identical SHA-256 hashes; they are one study. The noiseless-outlier "full" and "fractions" files overlap on 3,600 method records with exactly matching errors. Their union has 7,100 records. Despite the directory name, there are no 30% contamination records; the 20% condition has only 25 of 50 trials at $n = 512, m/n = 8$. The SRPR suite has 38 of 40 planned dimension/ratio cells: 100 trials per cell at $n \leq 512$ and 30 at larger n. Its 43 error records are all WF runs at ratios 2, 2.5, or 3.

Response contamination is a separate stress test. The supplied outlier generator adds positive uniform values, scaled by signal energy, at randomly selected indices. It is not an adversary choosing arbitrary responses or sensing vectors. With dense Cauchy noise normalized to 35 dB before adding outliers, the 10% condition gives success at phase error 0.045 in 449/450 RQWF-P trials, 399/450 Median-TWF trials, 291/450 TWF trials, and 0/450 WF trials. All four methods have zero success at 50% and 100% contamination. The noiseless-outlier union has RQWF-P success 365/425 at the incomplete 20% condition. Because these runs stop using truth-based error, successful final errors are stopping outcomes, not fully optimized precision measurements.

F.3 Pilot study

A separate implementation, RQWF-T/1.0, uses disjoint data-independent $D/V/U$ blocks, fixed response cap $c = 4$, a shared bandwidth $\delta = N^{-1/2}$ for radial calibration and refinement, radial bisection, and ungated updates on the same U, without unit-sphere projection or truth-based stopping. Implementation checks include 45 complex directional finite differences at $\tau \in \{0.25, 0.5, 0.75\}$; the maximum scaled discrepancy is 1.079×10^{-7}. These are implementation checks, not a proof or a statistical-rate validation.

The pilot uses unit signals, $n = 64$, $N/n \in \{4, 8\}$, two paired seeds per cell, and independent Gaussian, Student-t_3, or Cauchy errors with fixed population scale $\sigma = 0.05$ and $\tau = 1/2$. There is no realized-SNR normalization. RQWF-T/1.0 uses split fractions $0.4/0.2/0.4$, step $\eta = 0.02$, and 150 updates; RQWF-P/1.0 is a separately configured empirical comparator. Table 17 reports all six cells, without selecting favorable outcomes. All 24 method runs returned finite estimates; this does not mean successful recovery.

Table 17: Fixed-scale pilot.

Noise	N/n	RQWF-T/1.0	RQWF-P/1.0
Gaussian	4	1.0617	0.7360
Gaussian	8	0.5912	0.0306
Student-t_3	4	0.9045	0.3610
Student-t_3	8	0.6713	0.0419
Cauchy	4	0.9045	0.6147
Cauchy	8	0.7648	0.0839

Entries are median relative phase errors over two seeds, not a well-powered ranking. Equivalent operator budgets are recorded but not matched. The pilot also fails the main theorem's step-size condition: since $\kappa \le f(0) \le M_R$, its sufficient bound obeys $\kappa/(1024M_R^2) \le 1/[1024f(0)] < 0.000154$ for these three laws, whereas the implemented step is 0.02. The eigensolver checks a relative eigenvector residual, not the leading-eigenvalue deficit required by Proposition 17. Radial bisection also allows interval-width stopping, so the returned KKT residual must be checked separately. Thus the pilot documents a working computational path with substantial remaining recovery error; it establishes neither the theorem's finite-sample conditions nor its statistical rate.

F.4 Linear-model diagnostics

We examine density and sample-size effects in an auxiliary complex-linear quantile model,

$$q_x(a) = \mathrm{Re}(a^* x),$$

implemented through the equivalent real design $X = [\mathrm{Re}(A), -\mathrm{Im}(A)]$. The noise follows a population-quantile-centered Student-t_3 distribution. We use $n \in \{128, 256, 512\}$, sampling ratios $m/n \in \{4, 6, 8, 12\}$, target densities $\kappa_\tau \in \{0.5, 1, 2, 4\}$, and quantile levels $\tau \in \{0.25, 0.50, 0.75\}$. Each cell contains 30 paired trials, with WF, TWF, Median-WF, and RQWF evaluated on the same simulated problems. Because the response is complex-linear rather than a squared intensity, this study is an auxiliary diagnostic and not a direct numerical verification of the phase-retrieval theorem.

Figure 11 reports the scaling diagnostic at $\tau = 0.25$. The horizontal axis is $(1/\kappa_\tau)\sqrt{n/m}$, and the vertical axis is the median relative parameter error over 30 paired trials per cell. The panels correspond to $n = 128, 256, 512$ and retain all tested density and sampling settings. The alignment across dimensions is approximate and is used only as a scaling diagnostic.

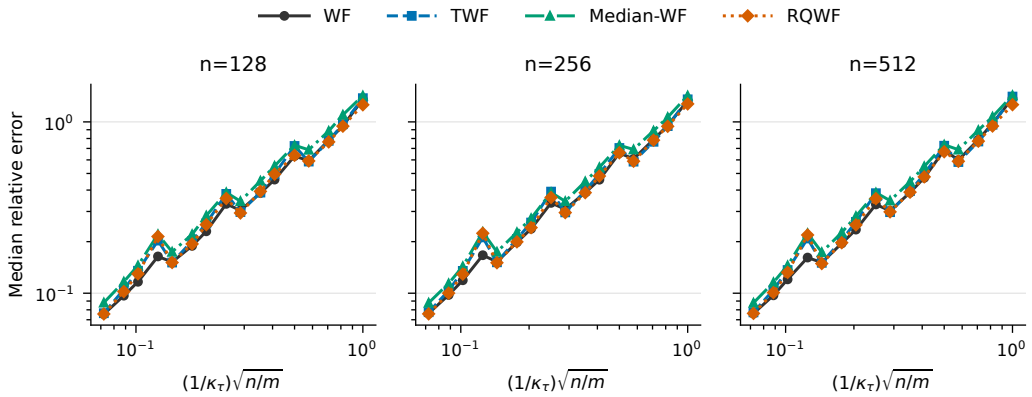


Figure 11: Linear-model scaling at $\tau = 0.25$.

Figure 12 reports the density grid at $\tau = 0.25$. Rows correspond to $n = 128, 256, 512$, columns correspond to $m/n = 4, 6, 8, 12$, and each curve varies κ_τ over $\{0.5, 1, 2, 4\}$. Errors decrease as the sampling ratio increases, while smaller target density generally corresponds to a harder estimation problem. No cells are removed or cropped.

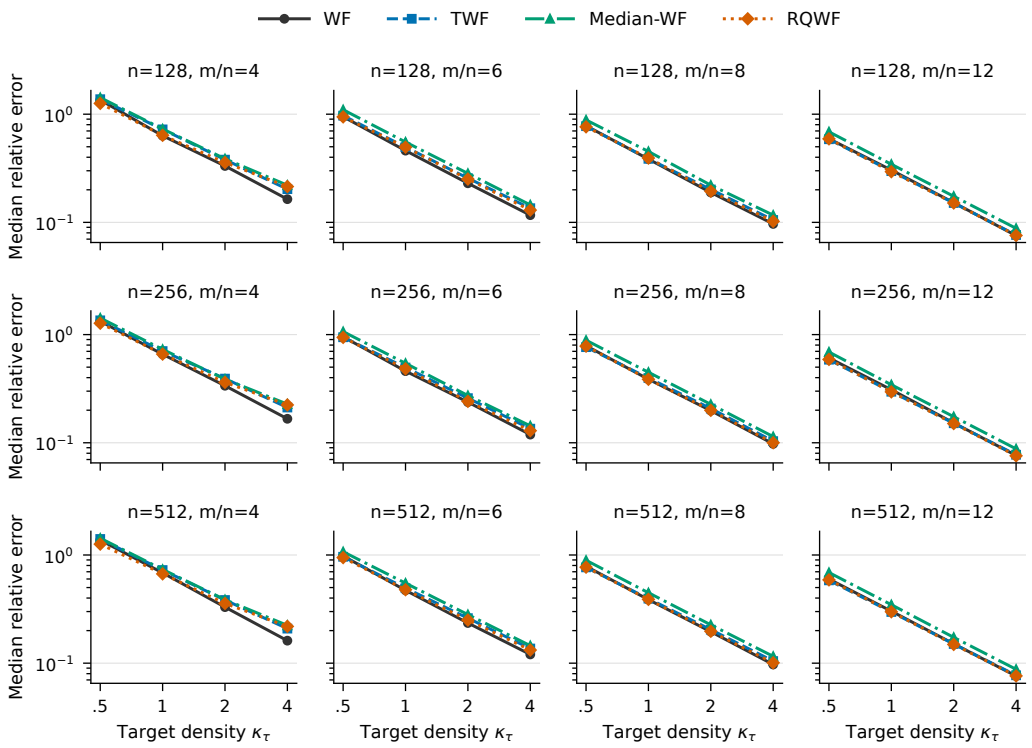


Figure 12: Linear-model density grid at $\tau = 0.25$.

Figure 13 reports the scaling diagnostic at $\tau = 0.50$. Median relative parameter error is plotted against $(1/\kappa_\tau)\sqrt{n/m}$ for $n = 128, 256, 512$. Every plotted value is the median of 30 paired trials. RQWF and Median-WF target the same conditional quantile in this setting, and their paired outputs coincide in the supplied implementation.

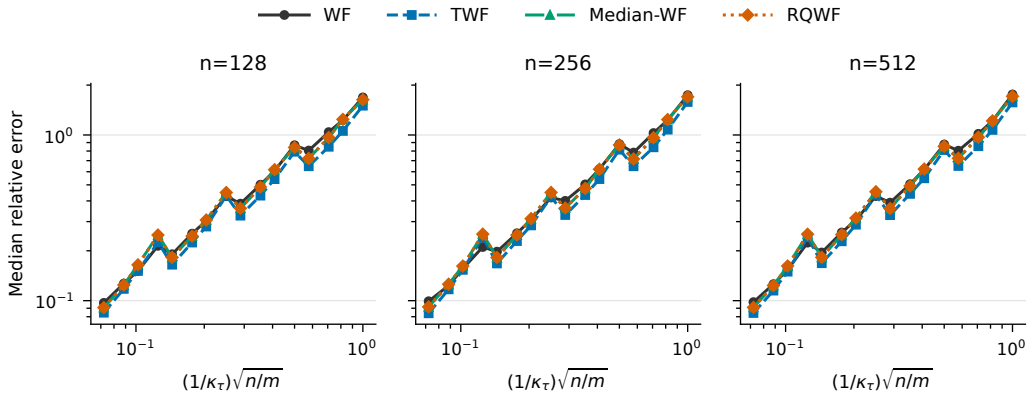


Figure 13: Linear-model scaling at $\tau = 0.50$.

Figure 14 reports the density grid at $\tau = 0.50$. Each panel fixes $(n, m/n)$ and reports the median relative parameter error as κ_τ varies. The cellwise display confirms that the aggregate trend is not determined by one dimension or sampling ratio, although density, finite-sample variation, and optimization error have not fully separated.

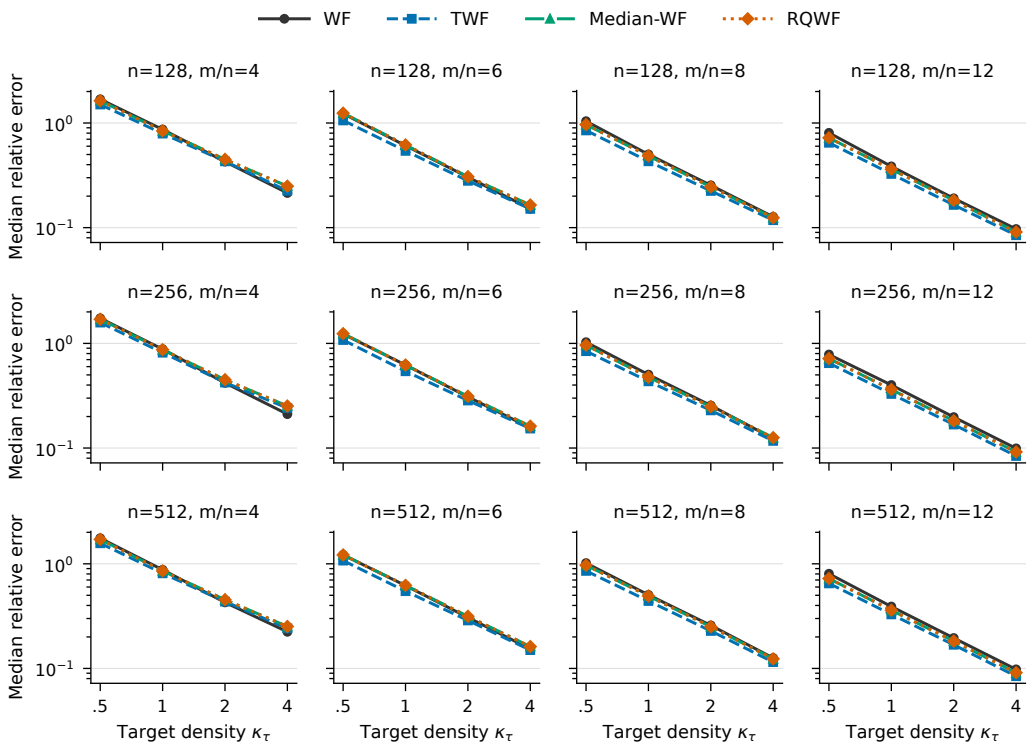


Figure 14: Linear-model density grid at $\tau = 0.50$.

Figure 15 reports the scaling diagnostic at $\tau = 0.75$. The upper-quantile experiment uses the same dimensions, sampling ratios, density levels, and 30 paired trials per cell. The rescaled curves exhibit similar sample-size trends across dimensions, but their discrepancies remain visible and are retained in the interpretation.

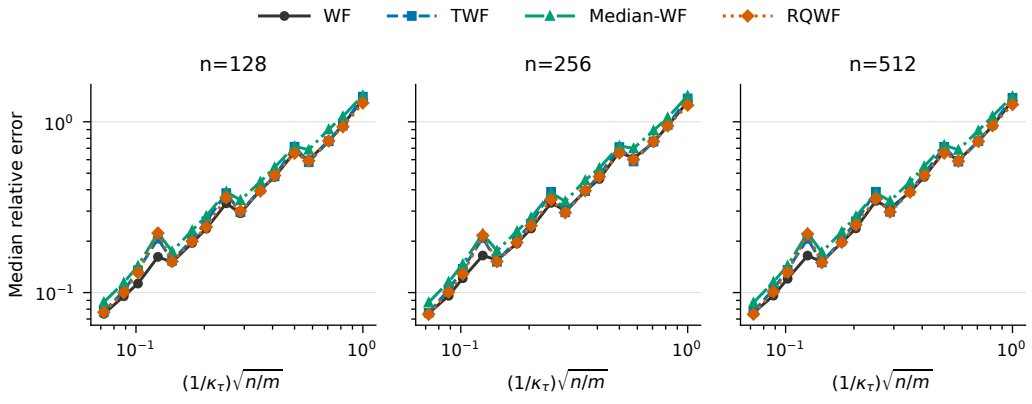


Figure 15: Linear-model scaling at $\tau = 0.75$.

Figure 16 reports the density grid at $\tau = 0.75$. The panels show all 12 dimension–sampling-ratio combinations and all four target-density levels. Methods are compared on identical paired trials within each cell. Error decreases with increasing sample size, but the density dependence is not identical across methods and cells.

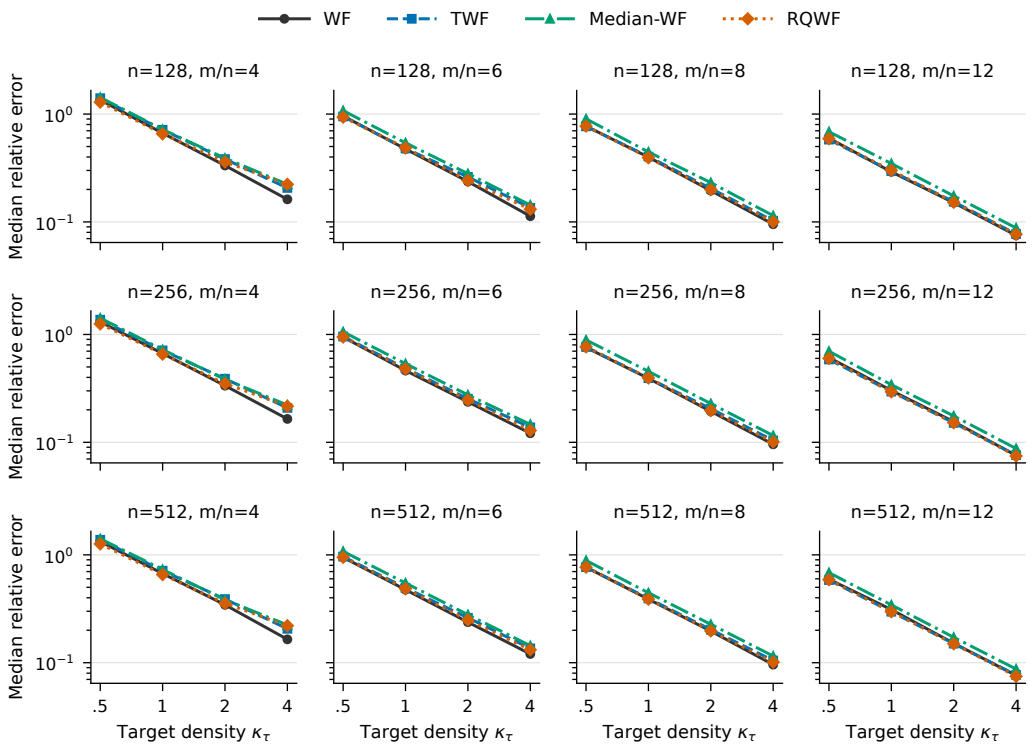


Figure 16: Linear-model density grid at $\tau = 0.75$.

Taken together, the six plots support a decreasing-error trend with increasing sample size and show a persistent dependence on the target density. They do not establish an exact $m^{-1/2}$ law: the fixed-condition results remain affected by finite-sample and optimization error. The collapse plots summarize the joint trend, while the complete-grid plots retain the individual dependence on n, m/n, κ_τ, and τ.

Across the full empirical program, the matched-budget bridge/baseline suite shows RQWF-P's Cauchy advantage and identifies data reuse as a key practical gain; the real-measurement study adds held-out intensity-prediction improvements under injected noise. The independent fixed-scale suite shows that prescribed non-median fits can improve selected asymmetric-noise cases. Together, these results support the practical relevance of conditional quantiles. They do not, however, validate RQWF-T's finite-sample rate: the new suite supplies independent fixed-scale noise, split-sample structure, and numerical diagnostics, but does not certify admissible tuning or every finite-sample sufficient condition. Algorithmic correspondence, admissible tuning, and verification of the sufficient conditions remain separate requirements. Density–sample-size diagnostics for the theorem must use squared-intensity sensing; no current study extends its guarantee to P or RD or establishes general complex-signal recovery from real optical data.