TAMEDMALA: CRITICALLY SCALED SCORE TAMING FOR ROBUST LANGEVIN SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

The Metropolis-adjusted Langevin algorithm (MALA) can be highly efficient when the score provides a reliable local proposal direction, but excessively large scores can instead produce proposals that are almost always rejected. We introduce a proposal-scale-coupled family of score tamers,

$$\mathcal{T}_{\delta,c,p,\gamma}(g) = \frac{g}{(1+\delta^{\gamma}\|g\|^{p}/c^{p})^{1/p}},$$

combined with an exact Metropolis–Hastings correction. For fixed-dimensional product blocks, we derive the leading Hastings expansion and show a scaling transition governed by the coupling exponent γ: for $\gamma < 1$, taming worsens the proposal-variance scaling, whereas for $\gamma \geq 1$ the classical $d^{-1/3}$ MALA scaling is preserved. At the critical coupling $\gamma = 1$, we obtain an explicit nonnegative asymptotic efficiency penalty on regular targets, clarifying that taming should act as a robustness safeguard rather than improve well-behaved MALA. We further show that $p = 2$ uniquely balances saturated deterministic drift with proposal noise. The resulting coordinate-wise TamedMALA achieves essentially MALA-level stationary efficiency across six independent tune-and-evaluate repetitions, while maintaining reliable recovery from deliberately pathological score states where MALA exhibits rejection-induced sticking. A short-chain tuner selects the additional taming scale automatically with stable stationary performance.

1 INTRODUCTION

The Metropolis-adjusted Langevin algorithm (MALA) uses the score $g(x) = \nabla \log \pi(x)$ to bias Gaussian proposals toward regions of high target density. In the classical product-target limit, this first-order information improves the proposal-variance scaling from the random-walk order to $\delta_d = \Theta(d^{-1/3})$, with optimal limiting acceptance probability approximately 0.574 (Roberts & Rosenthal, 1998). The same scaling persists well beyond the original iid setting (Pillai et al., 2012). This efficiency, however, relies on score magnitude being a useful local signal. When the score becomes extremely large, a nominal Langevin drift can instead generate an implausibly large proposal that is almost surely rejected. This phenomenon is visible both in classical ergodicity analyses and in practical far-from-typical-set behavior (Roberts & Tweedie, 1996; Livingstone & Zanella, 2022).

Controlling large Langevin drifts is not a new idea. Roberts and Tweedie introduced MALTA by truncating the drift, and later work developed adaptive versions of bounded-drift MALA (Roberts & Tweedie, 1996; Atchadé, 2006). Smoothing and proximal constructions regularize the target or replace gradients by proximity maps (Skare et al., 2000; Pereyra, 2016). In the unadjusted setting, TULA uses smooth taming to stabilize Euler discretizations with superlinear drift (Brosse et al., 2019); Metropolis-adjusted versions of that tamer have subsequently been evaluated as robust MALA baselines (Livingstone & Zanella, 2022). The Barker proposal takes a different route and redesigns the gradient proposal to improve robustness to tuning and target heterogeneity (Livingstone & Zanella, 2022). These methods make clear that robustness can be purchased, but also expose a central design question: *how strongly may the score be regularized before the favorable MALA scaling itself is altered?*

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.

We study the proposal-scale-coupled family

$$\mathcal{T}_{\delta,c,p,\gamma}(g) = \frac{g}{\left(1 + \delta^{\gamma} \|g\|^{p} / c^{p}\right)^{1/p}}, \qquad \gamma > 0,\ p \geq 1,\ c > 0, \tag{1}$$

applied coordinate-wise or in fixed-dimensional blocks and followed by the exact Metropolis–Hastings correction. The coupling exponent γ determines how quickly the regularization disappears as the proposal variance δ shrinks. Our main theoretical result identifies a sharp boundary: for $0 < \gamma < 1$, taming changes the high-dimensional proposal-variance exponent; for $\gamma > 1$, it is asymptotically invisible at leading order; and at $\gamma = 1$ it changes the leading Hastings constant while preserving the classical $d^{-1/3}$ scaling. This motivates the default method

$$\boxed{\mathcal{T}_{\delta,c,2,1}(g_i) = \frac{g_i}{\sqrt{1 + \delta g_i^2 / c^2}}}$$

which we call **TamedMALA**.

The paper makes four contributions. First, we derive the taming-induced term in the exact forward/reverse Hastings expansion and the resulting limiting acceptance law. Once the competing orders ε^3 and $\varepsilon^{2\gamma+1}$ are established, the crossover at $\gamma = 1$ follows algebraically. Second, at the critical coupling $\gamma = 1$, we derive an explicit nonnegative asymptotic efficiency penalty relative to MALA on regular product targets. Third, we show that $p = 2$ is uniquely scale-balanced: even under an arbitrarily large score, the maximal tamed deterministic displacement remains on the same $O(\sqrt{\delta})$ scale as the Gaussian proposal noise. Fourth, we test the resulting sampler in both stationary and pathological regimes. TamedMALA matches MALA-level stationary efficiency while substantially outperforming conventional smooth taming baselines; off stationarity, it retains prompt recovery when raw MALA sticks under pathological scores.

Our empirical claim is therefore a robustness–efficiency tradeoff. TamedMALA is designed to behave approximately like MALA when the score is moderate and to cap the influence of score magnitude when it ceases to be a trustworthy proposal-scale signal.

2 Related Work and Positioning

Truncation and taming. MALTA bounds the Langevin drift to restore stability in regimes where ordinary MALA can fail (Roberts & Tweedie, 1996); adaptive bounded-drift variants were studied by Atchadé (2006). TULA instead uses smooth taming for superlinear drifts (Brosse et al., 2019), and Metropolis-adjusted versions have been used as robust MALA baselines (Livingstone & Zanella, 2022). The conventional coordinate-wise tamer $g_i/(1 + (\delta/2)|g_i|)$ is the $p = 1, \gamma = 1, c = 2$ member of (1). We study the broader proposal-scale family and how γ, p, c enter the Hastings error and high-dimensional limit.

Regularized and robust proposals. Smoothing and proximal constructions regularize rapidly varying or nonsmooth targets (Skare et al., 2000; Pereyra, 2016); recent proximal-MALA theory also shows that suitable regularization can preserve classical MALA scaling (Pillai, 2024; Crucinio et al., 2025). Barker proposals take a different route, redesigning the gradient proposal for robustness to scale mismatch and heterogeneity (Livingstone & Zanella, 2022). Our benchmark therefore includes MALTA, conventional smooth taming, and Barker as complementary baselines.

Optimal scaling and tuning. Our asymptotic analysis builds on classical MALA scaling (Roberts & Rosenthal, 1998; Pillai et al., 2012) and uses exact MH symmetry to organize the one-block Hastings error, following the general perspective of Dobson et al. (2026). Automatic tuning is used as a practical deployment mechanism: selected parameters are frozen and evaluated on independent chains; full baseline search details are given in Appendix E.

3 Method

Let $\pi(x) \propto e^{-V(x)}$ on $\mathbb{R}^D$, with score $g(x) = \nabla \log \pi(x) = -\nabla V(x)$. Partition the score into blocks $g^{(b)} \in \mathbb{R}^m$ of fixed size m. For proposal variance $\delta > 0$, taming scale $c > 0$, shape $p \geq 1$,

and proposal-scale coupling $\gamma > 0$, define

$$\mathcal{T}_{\delta,c,p,\gamma}\left(g^{(b)}\right) = \frac{g^{(b)}}{\left(1 + \delta^{\gamma} \left\|g^{(b)}\right\|^{p} / c^{p}\right)^{1/p}}. \tag{2}$$

The proposal is

$$Y = X + \frac{\delta}{2}\mathcal{T}_{\delta,c,p,\gamma}(g(X)) + \sqrt{\delta}\, Z, \qquad Z \sim N(0, I_D), \tag{3}$$

where the tamer is applied blockwise. We use the exact Gaussian forward and reverse proposal densities in the Metropolis–Hastings ratio, so the resulting chain has invariant distribution π whenever the usual MH construction is well defined.

The theory is stated for fixed m as the number of blocks grows. This includes coordinate-wise taming ($m = 1$) and small physically meaningful blocks. It deliberately does *not* cover an unnormalized radial tamer formed from the norm of the full increasing-dimensional score.

3.1 Default TamedMALA

Our default sampler fixes

$$\gamma = 1, \qquad p = 2, \qquad m = 1.$$

Thus

$$\widetilde{g}_i(x) = \frac{g_i(x)}{\sqrt{1 + \delta g_i(x)^2 / c^2}}, \qquad Y_i = X_i + \frac{\delta}{2}\widetilde{g}_i(X) + \sqrt{\delta} Z_i. \tag{4}$$

A single scalar c is used across coordinates; scale heterogeneity should be handled by ordinary preconditioning rather than by coordinate-specific regularization strengths.

3.2 Automatic tuning of (δ, c)

The practical tuner maximizes a short-chain estimate of $Q_{0.25}$ rank-normalized bulk ESS per retained proposal-gradient evaluation rather than targeting an acceptance rate. It begins with a broad multiplicative search in δ and c, expands when the selected point lies on a boundary, and then refines geometrically. Tuning chains are discarded and final comparisons use fresh chains and independent proposal streams. Baseline-specific search spaces and compute accounting are reported in Appendix E.

4 Theory

We analyze a product of fixed-dimensional blocks

$$\Pi_d(Q) = \prod_{j=1}^{d} \pi(q_j), \qquad q_j \in \mathbb{R}^m,$$

with m fixed and $d \to \infty$.

Assumption 4.1 (Regular block target)**.** *Write $\pi(x) \propto e^{-V(x)}$. We assume $V \in C^4(\mathbb{R}^m)$; derivatives through order four have polynomial growth; $X \sim \pi$ has moments of every finite order needed below; and integration by parts is valid for the polynomial-growth vector fields used in the proof.*

Set $\varepsilon = \sqrt{\delta}$, $v = \nabla V$, $H = \nabla^2 V$, and

$$C_{p,c}(v) = \frac{\|v\|^{p} v}{p c^{p}}.$$

For ordinary MALA, the leading one-block negative log Hastings coefficient is

$$\alpha_0(x, z) = \frac{1}{4}\langle z, H(x)v(x)\rangle - \frac{1}{12}\nabla^3 V(x)[z, z, z]. \tag{5}$$

The leading taming coefficient is

$$\beta_{p,c}(x, z) = \langle z, C_{p,c}(v(x))\rangle. \tag{6}$$

4.1 A PHASE TRANSITION IN THE PROPOSAL-SCALE COUPLING

A Taylor expansion of the tamer gives

$$\mathcal{T}_{\varepsilon^2,c,p,\gamma}(v) = v - \varepsilon^{2\gamma} C_{p,c}(v) + O(\varepsilon^{4\gamma}),$$

and the corresponding one-block Hastings error has the form

$$\Delta_{\varepsilon,\gamma} = \varepsilon^3 \alpha_0(X,Z) + \varepsilon^{2\gamma+1} \beta_{p,c}(X,Z) + R_{\varepsilon,\gamma}, \tag{7}$$

with

$$\|R_{\varepsilon,\gamma}\|_{L^q} = O\big(\varepsilon^4 + \varepsilon^{2\gamma+2} + \varepsilon^{4\gamma+1}\big)$$

for some $q > 3$. Establishing (7) is the main perturbative step: the exact Metropolis–Hastings forward/reverse calculation cancels lower-order contributions and identifies both the taming coefficient and its remainder order. Once these two leading powers are known, the location of their crossover at $\gamma = 1$ is immediate.

Define

$$\nu_\gamma = \min\{3, 2\gamma + 1\}$$

and the leading coefficient

$$\alpha_\gamma = \begin{cases} \beta_{p,c}, & 0 < \gamma < 1, \\ \alpha_0 + \beta_{p,c}, & \gamma = 1, \\ \alpha_0, & \gamma > 1. \end{cases} \qquad \Sigma_\gamma = \mathbb{E}[\alpha_\gamma^2].$$

Theorem 4.2 (High-dimensional scaling phase transition)**.** *Under Assumption 4.1, fix $\gamma > 0$, finite $p \geq 1$, $c > 0$, and suppose $\Sigma_\gamma > 0$. If*

$$\delta_d = \ell^2 d^{-1/\nu_\gamma},$$

then the total negative log Hastings ratio converges as

$$\Delta_d \Longrightarrow N\left(\frac{1}{2}\ell^{2\nu_\gamma}\Sigma_\gamma,\, \ell^{2\nu_\gamma}\Sigma_\gamma\right),$$

and the limiting mean acceptance probability is

$$A_\gamma(\ell) = 2\Phi\left(-\frac{1}{2}\ell^{\nu_\gamma}\sqrt{\Sigma_\gamma}\right). \tag{8}$$

Equivalently,

$$\boxed{\delta_d = \begin{cases} \Theta\big(d^{-1/(2\gamma+1)}\big), & 0 < \gamma < 1, \\ \Theta(d^{-1/3}), & \gamma \geq 1. \end{cases}} \tag{9}$$

The theorem distinguishes the roles of the two errors in (7). For $\gamma < 1$, taming enters before the intrinsic MALA error and worsens the dimension exponent. For $\gamma > 1$, it enters after the MALA term and is asymptotically invisible at leading order. The critical value

$$\boxed{\gamma = 1}$$

is the unique coupling at which taming changes the leading Hastings constant without changing the $d^{-1/3}$ exponent.

The same calculation gives the optimal limiting acceptance. If $u = \ell^{\nu_\gamma}\sqrt{\Sigma_\gamma}/2$, then the leading squared-jump speed is proportional to $u^{2/\nu_\gamma} 2\Phi(-u)$, so the optimum solves

$$2\Phi(-u) = \nu_\gamma u \phi(u).$$

For every $\gamma \geq 1$, $\nu_\gamma = 3$ and the classical 0.574 rule is recovered.

4.2 TAMING IS NOT FREE ON REGULAR TARGETS

At $\gamma = 1$, let $\Sigma_{p,c} = \mathbb{E}[(\alpha_0 + \beta_{p,c})^2]$ and $\Sigma_0 = \mathbb{E}[\alpha_0^2]$.

Proposition 4.3 (Positive regular-target penalty)**.** *Under Assumption 4.1,*

$$\Sigma_{p,c} - \Sigma_0 = \frac{1}{p^2 c^{2p}} \mathbb{E} \|v\|^{2p+2} + \frac{1}{2pc^p} \mathbb{E}\left[\|v\|^p \|H\|_F^2 + p \|v\|^{p-2} \|Hv\|^2 \right] \geq 0. \tag{10}$$

Consequently, if the inequality is strict, the ratio of optimally tuned limiting squared-jump speeds is

$$\frac{\mathrm{Speed}^\star_{p,c}}{\mathrm{Speed}^\star_{\mathrm{MALA}}} = \left(\frac{\Sigma_0}{\Sigma_{p,c}} \right)^{1/3} < 1.$$

As $c \to \infty$, ordinary MALA is recovered.

For regular targets with the usual nondegenerate MALA coefficient $\Sigma_0 > 0$, the critical and supercritical limits are also nondegenerate; exceptional cases are discussed in Appendix B. This result sharpens the interpretation of the method. On the smooth, regular, stationary product regime where MALA already behaves ideally, finite taming cannot improve the leading asymptotic local efficiency. A useful tamer must justify itself in regimes excluded by those assumptions: large local scores, singular score geometry, finite-step instability, or atypical states.

4.3 $p = 2$ IS UNIQUELY DRIFT–NOISE BALANCED

For $r = \|g\|$, the tamed score magnitude is

$$m_{\delta,p}(r) = \frac{r}{(1 + \delta r^p / c^p)^{1/p}}.$$

Taming becomes order one at

$$r_{\mathrm{act}} \asymp c\, \delta^{-1/p}.$$

The raw deterministic MALA displacement becomes comparable to the Gaussian noise when $r \asymp \delta^{-1/2}$. Hence these two score scales coincide only for $p = 2$.

Moreover, as $r \to \infty$,

$$m_{\delta,p}(r) \to c\delta^{-1/p},$$

so

$$\frac{\|\text{maximal deterministic displacement}\|}{\sqrt{\delta}} \leq \frac{c}{2} \delta^{1/2 - 1/p}. \tag{11}$$

For $p < 2$, the right-hand side diverges as $\delta \to 0$; for $p > 2$, it vanishes; and for $p = 2$,

$$\boxed{\left\| \frac{\delta}{2} \mathcal{T}_{\delta,c,2,1}(g) \right\| \leq \frac{c}{2} \sqrt{\delta}.} \tag{12}$$

Thus arbitrarily large scores cannot force deterministic motion to a larger asymptotic order than the proposal noise. This is a scale criterion, not an efficiency ordering over p.

Proof overview. The detailed derivations are collected in Appendix A. The key steps are: (i) expand the tamer and the exact forward/reverse Gaussian Hastings ratio; (ii) identify the leading one-block coefficient (7); (iii) use the exact MH swap identity to show that the one-block mean is asymptotically one half of its variance; and (iv) apply a triangular-array CLT after choosing $d\delta_d^{\nu_\gamma} = O(1)$.

5 EXPERIMENTS

The experiments are designed to separate three questions: whether the theory predicts the dimensional scaling; whether critical taming preserves stationary efficiency; and whether it provides robustness precisely where score magnitude becomes pathological.

5.1 Experimental protocol

Unless otherwise stated, the benchmark dimension is 128. We compare MALA, MALTA, the conventional global and coordinate-wise Brosse-style taming baselines (TMALA/TMALAc), the Barker proposal, and TamedMALA. The primary metric is the lower-quartile rank-normalized bulk ESS across a fixed diagnostic panel, divided by retained-sampling proposal-gradient evaluations (Vehtari et al., 2021). Every method is tuned on short chains and then evaluated on fresh, independent chains; the final benchmark data are never used for parameter selection. The default tuning stage uses 2 coarse chains (300 warm-up, 1000 retained) followed by 4 refinement chains (1000 warm-up, 4000 retained); final evaluation uses 8 chains with 5000 warm-up and 30000 retained samples. We repeat the complete tune-and-evaluate pipeline six times with independent seeds.

We separate targets by evidentiary role. Gaussian, quartic, and sextic are regular product targets within the class of Assumption 4.1 and serve as finite-dimensional checks of the theory. Ginzburg–Landau, the planar singular target $\pi(x) \propto |x_1|^2 e^{-\|x\|^2/2}$, and the Poisson random-effects posterior deliberately violate product, smoothness, or independence assumptions; they are empirical generalization/stress tests only, not validations of Theorem 4.2. The appendix adds further targets. No final benchmark chain failed and no final benchmark proposal was recorded as a numerical rejection.

5.2 Scaling phase transition

To test Theorem 4.2 without assuming an optimal acceptance rule, we fix $p = 2, c = 1$, sweep γ, and independently find the proposal variance giving acceptance 0.5 at each dimension. Gaussian, quartic, and sextic targets admit exact stationary sampling, avoiding burn-in contamination. We fit

$$\log \delta_d^\star = a + \widehat{\beta}(\gamma) \log d$$

on the large-dimension tail. Figure 1a is intended as a finite-dimensional sanity check on the phase law, not as a numerical proof of the asymptotic theorem. Below $\gamma = 1$, the fitted slopes become systematically steeper as predicted; above the boundary they flatten toward the MALA value $-1/3$. The sextic target has the largest finite-size offset near the critical boundary, but by $\gamma \in \{1.3, 1.5\}$ its fitted slope is approximately -0.34.

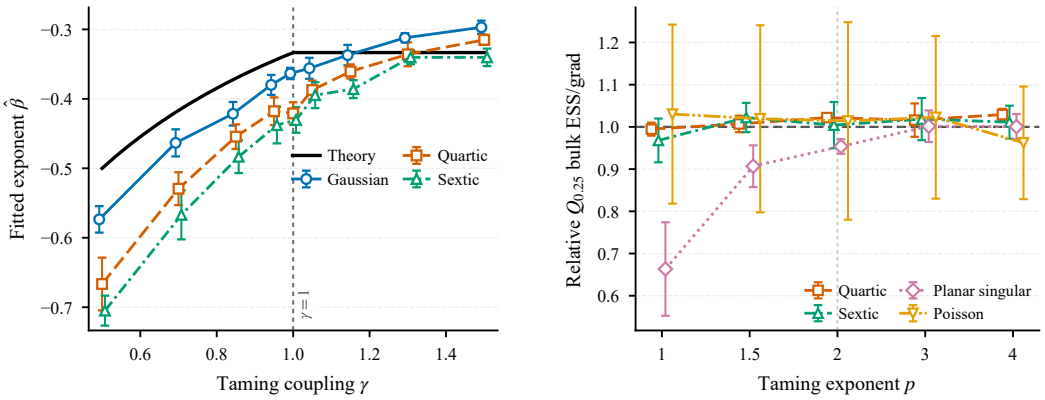
5.3 Choosing p

The theory singles out $p = 2$ by drift–noise scale balance, not by a universal ESS optimum. To isolate the effect of p, the ablation changes only p and does *not* retune the remaining parameters. For each target and each of the six production benchmark repetitions, we import the MALA proposal variance and the coordinate-wise $p = 2$ TamedMALA pair (δ, c) selected by the headline benchmark. We then evaluate $p \in \{1, 1.5, 2, 3, 4\}$ with that TamedMALA (δ, c) held fixed across p, and normalize by an independently evaluated MALA chain using its imported production δ. Quartic and sextic chains are initialized exactly from stationarity.

Figure 1b reports the resulting six-repetition sensitivity analysis. On the regular quartic and sextic targets, all finite-p mean efficiencies remain close to MALA, ranging from 0.958 to 1.023 times the MALA reference. The planar singular target is more discriminating: relative efficiency rises from 0.756 ± 0.123 at $p = 1$ to 0.958 ± 0.029 at $p = 2$, and remains near MALA for larger p. The Poisson estimates are noisier across repetitions. Thus the ablation does not support an ESS-optimality claim for $p = 2$; rather, it shows that the scale-balanced choice incurs little stationary-efficiency cost while avoiding the larger $p < 2$ loss on the singular target. Full values are reported in Appendix 4.

5.4 Stationary benchmark

Table 1 shows that TamedMALA retains MALA-level stationary efficiency while substantially improving over conventional smooth taming. Its geometric mean efficiency relative to MALA across the six main targets is 0.999; MALTA is similarly competitive. The small quartic point-estimate advantage of TamedMALA over MALA is a finite-dimensional effect within replicate variation and is consistent with Proposition 4.3. The practical distinction appears in the off-stationary experiment below.



(a) Scaling phase transition.

(b) p-sensitivity at frozen production parameters.

Figure 1: **Theory validation and p-sensitivity. (a)** Empirical large-d fits of $\log \delta_d^\star = a + \widehat{\beta} \log d$ at fixed acceptance 0.5, compared with the phase law in Theorem 4.2. Error bars are the reported 95% uncertainties of the fitted slopes. **(b)** $Q_{0.25}$ bulk ESS per retained-sampling gradient relative to MALA when only p is varied; each TamedMALA curve reuses the corresponding headline $p = 2$ values of (δ, c), and points show mean $\pm$ sample SD over six production-parameter repetitions. The horizontal reference is MALA and the vertical guide marks $p = 2$.

Table 1: Stationary sampling efficiency across six independent end-to-end tune-and-evaluate repetitions. Entries are $100 \times Q_{0.25}$ rank-normalized bulk ESS per retained-sampling gradient evaluation (mean $\pm$ sample SD; higher is better). The first block contains regular product targets covered by the scaling assumptions; the second block contains out-of-class empirical stress tests. TamedMALA denotes coordinate-wise $p = 2, \gamma = 1$.

Target	MALA	MALTA	TMALA	TMALAc	Barker	**TamedMALA**
Regular product targets (within theory class)						
Gaussian	$\mathbf{9.285 \pm 0.067}$	9.210 ± 0.130	1.271 ± 0.027	3.677 ± 0.032	4.196 ± 0.092	9.225 ± 0.125
Quartic	2.733 ± 0.072	$\mathbf{2.787 \pm 0.040}$	1.065 ± 0.014	1.930 ± 0.034	1.924 ± 0.051	2.784 ± 0.036
Sextic	$\mathbf{1.315 \pm 0.015}$	1.298 ± 0.054	0.787 ± 0.016	1.105 ± 0.051	1.080 ± 0.026	1.296 ± 0.034
Out-of-class empirical stress tests						
Ginzburg–L	2.317 ± 0.046	$\mathbf{2.370 \pm 0.032}$	0.899 ± 0.039	1.672 ± 0.039	1.646 ± 0.045	2.266 ± 0.040
Planar sing.	$\mathbf{8.584 \pm 0.108}$	8.578 ± 0.068	1.249 ± 0.021	3.598 ± 0.046	4.080 ± 0.061	8.383 ± 0.220
Poisson RE	2.608 ± 0.186	$\mathbf{2.777 \pm 0.097}$	2.488 ± 0.111	2.565 ± 0.223	2.076 ± 0.053	2.705 ± 0.152

5.5 Robustness outside stationarity

Stationary ESS alone does not test the motivation for score taming. Following the spirit of the far-from-typical-set experiments used for TULA (Brosse et al., 2019), we freeze the parameters selected by the stationary benchmark and perturb only one coordinate into a pathological-score regime. All other coordinates are exact stationary draws.

We parameterize severity by the dimensionless raw-MALA drift/noise ratio

$$\rho_0 = \frac{\sqrt{\delta_{\mathrm{MALA}}}}{2} |g_1(x_0)|, \tag{13}$$

with $\rho_0 \in \{1, 2, 4, 8, 16, 32, 64\}$. We measure *prompt stable recovery*: a run succeeds only after ten consecutive states satisfy the safe-score criterion $\rho \leq 1$, and this must occur within 500 proposal-gradient evaluations. We run 100 chains per condition for each of the three independently tuned production parameter sets; there is no robustness-specific retuning.

On sextic, MALA still recovers reliably at $\rho_0 = 8$, but its prompt recovery probability falls to 0.003 ± 0.006 at $\rho_0 = 16$ and to zero at $\rho_0 = 32, 64$. Its pre-recovery acceptance simultaneously collapses from about 0.51 at severity 8 to 2.9×10^{-4} at severity 16. MALTA, TMALAc, and

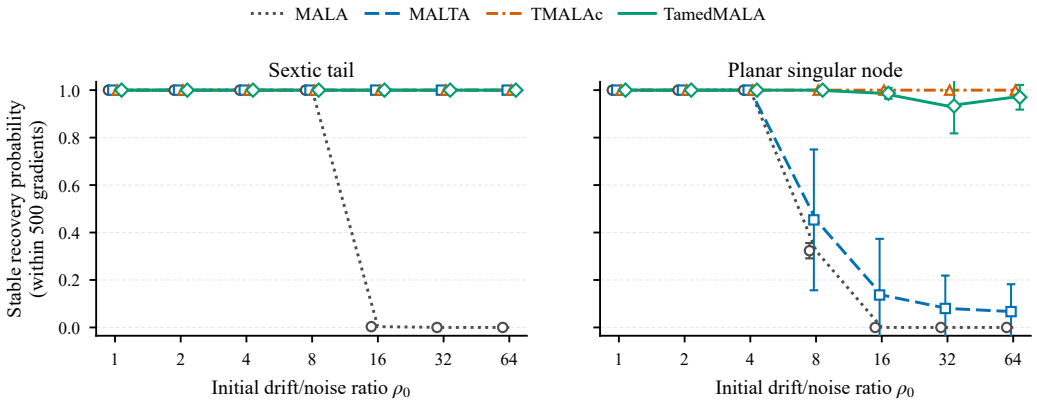


Figure 2: **Prompt recovery from pathological score magnitudes.** Left: sextic tail. Right: planar singular node. The horizontal axis is the initial raw-MALA drift/noise ratio ρ_0; the vertical axis is the probability of ten consecutive safe-score states within 500 proposal gradients. Points show mean $\pm$ sample SD across the three production parameter sets. Curves are drawn at the true severity values; markers and error bars are shifted slightly in log-severity only to prevent overlap.

TamedMALA recover in every run through $\rho_0 = 64$, showing that all three explicit large-score controls resolve the pure superlinear-tail failure.

The planar singular target is more discriminating. At $\rho_0 = 8$, prompt recovery is 0.323 ± 0.032 for MALA and 0.453 ± 0.297 for MALTA, whereas TMALAc and TamedMALA both recover in every run. From severity 16 onward MALA does not recover within the budget; MALTA falls to 0.137 ± 0.237, 0.080 ± 0.139, and 0.067 ± 0.115 at severities $16, 32, 64$. TMALAc remains at 1.0, while TamedMALA remains high at 0.987 ± 0.023, 0.933 ± 0.115, and 0.970 ± 0.052. TMALAc is therefore the strongest pure recovery baseline in this experiment; TamedMALA's empirical advantage is the combination of comparable pathological recovery with the much stronger stationary efficiency in Table 1.

Across all robustness runs, there are no numerical rejections or hard numerical failures. The measured failure mode is rejection-induced algorithmic sticking, not floating-point overflow.

5.6 Can the extra parameter be tuned automatically?

The additional scale c is a potential practical weakness if performance depends on manual selection. We therefore validate the fixed tuner in three ways: independently re-evaluated (δ, c) landscapes, three independent headline tuning seeds, and explicit tuning-budget sensitivity.

Figure 3 shows broad high-efficiency regions rather than a fragile isolated optimum. On Gaussian, selected configurations are effectively MALA-like; on sextic, all headline runs select $c = 8$, but the stationary objective remains broad toward weaker taming. We therefore make the narrow claim that the extra parameter can be selected automatically without sacrificing stationary performance, not that stationary tuning identifies a uniquely robustness-optimal c. A low/default/high budget audit gives no systematic gain from more than doubling the default search budget; detailed selections, costs, and validation scores are reported in Appendix E.

6 Discussion and Limitations

The theory and experiments together support a deliberately limited claim. Critical score taming can preserve the favorable MALA dimensional exponent and MALA-like stationary efficiency while preventing catastrophic large-score proposal behavior. The regular-target penalty in Proposition 4.3 is important: it explains why a successful sampler should approach MALA, rather than beat it, when taming is unnecessary.

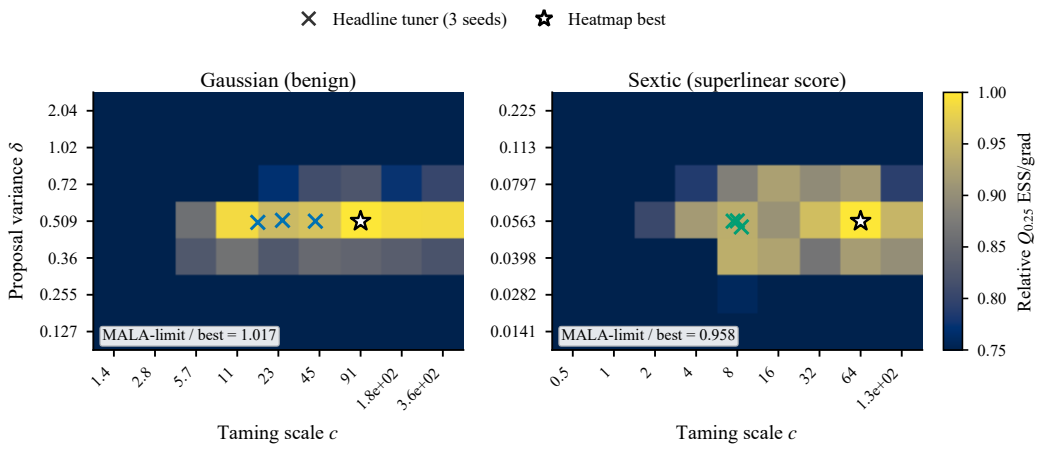


Figure 3: **Stationary tuning landscapes for the extra scale c.** Left: benign Gaussian target. Right: sextic target. Each heatmap cell is an independent evaluation of $Q_{0.25}$ bulk ESS/gradient, normalized by the best finite displayed cell. Crosses mark the three headline-tuner selections and the star marks the independently evaluated heatmap optimum; the inset text reports the best displayed MALA-limit efficiency relative to the heatmap best. The three crosses are slightly displaced horizontally only to avoid exact overlap.

Several limitations remain. First, only the Gaussian, quartic, and sextic product targets are used as within-assumption finite-dimensional checks of the scaling theory. The interacting, singular, and Bayesian rows are deliberately out-of-class empirical tests and are not evidence for Theorem 4.2. Second, singular-score targets are deliberately outside Assumption 4.1. The frozen-parameter p-ablation shows a marked stationary loss for $p < 2$ on the planar singular target while $p \geq 2$ remains near MALA, which is consistent with the drift–noise argument but does not constitute a boundary-layer theorem for singular targets. Third, c is tuned for stationary efficiency, not explicitly for worst-case recovery. The sextic tuning landscape shows why these objectives can differ: extreme-score states are rare under stationarity, so stationary ESS contains little information about how strongly to regularize those states. Finally, we have not attempted to prove that our simple grid-and-refinement tuner is preferable to generic Bayesian optimization or other black-box optimizers; the tuner is a practical mechanism for removing manual hyperparameter selection, not a contribution in optimization.

A natural extension is to combine stationary-efficiency tuning with an explicit robustness criterion, for example selecting the strongest taming among configurations statistically indistinguishable on stationary ESS. Such a rule would change the practical algorithm and is left for future work.

7 Conclusion

MALA's use of score magnitude is both its strength and a potential source of fragility. We introduced a proposal-scale-coupled score-taming family and identified a critical coupling, $\gamma = 1$, at which taming remains visible in the leading Metropolis–Hastings error without worsening MALA's $d^{-1/3}$ proposal-variance scaling. At this boundary, finite taming carries an explicit regular-target asymptotic penalty, making robustness—rather than universal efficiency improvement—the appropriate motivation. The scale-balanced choice $p = 2$ caps the saturated deterministic drift at the same order as proposal noise.

Empirically, coordinate-wise $p = 2$ TamedMALA preserves MALA-level stationary efficiency across heterogeneous targets, while avoiding the rejection-induced sticking of ordinary MALA under deliberately pathological score magnitudes. The additional taming scale can be selected automatically with a modest short-chain budget and does not create fragile stationary performance. These results suggest that proposal-scale-aware taming is a simple way to add a robustness safeguard to exact Langevin sampling without discarding the high-dimensional behavior that makes MALA attractive.

REFERENCES

Yves F. Atchadé. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in Applied Probability*, 8(2):235–254, 2006. doi: 10.1007/s11009-006-8550-0.

Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. The tamed unadjusted langevin algorithm. *Stochastic Processes and their Applications*, 129(10):3638–3663, 2019. doi: 10.1016/j.spa.2018.10.002.

Francesca R. Crucinio, Alain Durmus, Pablo Jiménez, and Gareth O. Roberts. Optimal scaling results for moreau–yosida metropolis-adjusted langevin algorithms. *Bernoulli*, 31(3):1889–1907, 2025. doi: 10.3150/24-BEJ1790.

P. Dobson, J. M. Sanz-Serna, and K. C. Zygalakis. Optimal scaling of mcmc algorithms: Exploiting the symmetry of the metropolis–hastings formula. *arXiv preprint arXiv:2607.00586*, 2026.

Samuel Livingstone and Giacomo Zanella. The barker proposal: Combining robustness and efficiency in gradient-based mcmc. *Journal of the Royal Statistical Society: Series B*, 84(2):496–523, 2022. doi: 10.1111/rssb.12482.

Marcelo Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26(4):745–760, 2016. doi: 10.1007/s11222-015-9567-4.

Natesh S. Pillai. Optimal scaling for the proximal langevin algorithm in high dimensions. *Journal of Machine Learning Research*, 25(404):1–32, 2024.

Natesh S. Pillai, Andrew M. Stuart, and Alexandre H. Thiéry. Optimal scaling and diffusion limits for the langevin algorithm in high dimensions. *The Annals of Applied Probability*, 22(6):2320–2356, 2012. doi: 10.1214/11-AAP828.

Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B*, 60(1):255–268, 1998. doi: 10.1111/1467-9868.00123.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. doi: 10.2307/3318418.

Øivind Skare, Fred Espen Benth, and Arnoldo Frigessi. Smoothed langevin proposals in metropolis–hastings algorithms. *Statistics & Probability Letters*, 49(4):345–354, 2000. doi: 10.1016/S0167-7152(00)00067-5.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc. *Bayesian Analysis*, 16(2):667–718, 2021. doi: 10.1214/20-BA1221.

A PROOF DETAILS

This appendix records the core derivations used in Section 4. The notation follows the main text.

A.1 ONE-BLOCK EXPANSION

Let $d_y = y - x$. For proposal drift b_ε, the exact negative log Hastings ratio can be written as

$$\Delta_\varepsilon = V(y) - V(x) + \frac{1}{2}\langle d_y, b_\varepsilon(x) + b_\varepsilon(y)\rangle + \frac{\varepsilon^2}{8}\left(\|b_\varepsilon(y)\|^2 - \|b_\varepsilon(x)\|^2\right). \tag{14}$$

For ordinary MALA, Taylor expansion of V, ∇V, and $\|\nabla V\|^2$ gives

$$\Delta_{0,\varepsilon} = \varepsilon^3\left[\frac{1}{4}\langle Z, Hv\rangle - \frac{1}{12}\nabla^3 V[Z,Z,Z]\right] + O_{L^q}(\varepsilon^4).$$

For the tamer,

$$\mathcal{T}_{\varepsilon^2,c,p,\gamma}(v) = v - \varepsilon^{2\gamma} C_{p,c}(v) + O_{L^q}(\varepsilon^{4\gamma}).$$

Substitution into (14), together with $Y - X = \varepsilon Z + O_{L^q}(\varepsilon^2)$, gives

$$\Delta_{\varepsilon,\gamma} = \varepsilon^3\alpha_0 + \varepsilon^{2\gamma+1}\beta_{p,c} + O_{L^q}\left(\varepsilon^4 + \varepsilon^{2\gamma+2} + \varepsilon^{4\gamma+1}\right).$$

A.2 METROPOLIS–HASTINGS SYMMETRY

For a one-step proposal started at stationarity, the MH swap identity implies

$$\mathbb{E}[f(\Delta_\varepsilon)] = \mathbb{E}[e^{-\Delta_\varepsilon} f(-\Delta_\varepsilon)]$$

for integrable f. Taking $f(u) = u$ and expanding $1 - e^{-u}$ yields, whenever $\Delta_\varepsilon/\varepsilon^\nu \to \alpha$ in L^3,

$$\mathbb{E}[\Delta_\varepsilon] = \frac{1}{2}\mathbb{E}[\Delta_\varepsilon^2] + o(\varepsilon^{2\nu}) = \frac{1}{2}\Sigma\varepsilon^{2\nu} + o(\varepsilon^{2\nu}),$$

and

$$\mathrm{Var}(\Delta_\varepsilon) = \Sigma\varepsilon^{2\nu} + o(\varepsilon^{2\nu}), \qquad \Sigma = \mathbb{E}[\alpha^2].$$

This is the step that converts the order of the leading one-block Hastings coefficient into the high-dimensional scaling exponent.

A.3 PROOF OF THEOREM 4.2

The total log Hastings error is the sum of d independent one-block contributions. With

$$\nu_\gamma = \min\{3, 2\gamma + 1\}, \qquad \delta_d = \ell^2 d^{-1/\nu_\gamma}, \qquad \varepsilon_d = \ell d^{-1/(2\nu_\gamma)},$$

we have

$$d\varepsilon_d^{2\nu_\gamma} = \ell^{2\nu_\gamma}.$$

The preceding symmetry calculation therefore gives convergence of the total mean and variance to

$$\frac{1}{2}\ell^{2\nu_\gamma}\Sigma_\gamma, \qquad \ell^{2\nu_\gamma}\Sigma_\gamma.$$

The uniform L^3 remainder bound gives a Lyapunov numerator

$$O(d\varepsilon_d^{3\nu_\gamma}) = O(d^{-1/2}),$$

so the triangular-array CLT applies. The Gaussian acceptance identity

$$G \sim N(s^2/2, s^2) \quad \Longrightarrow \quad \mathbb{E}[1 \wedge e^{-G}] = 2\Phi(-s/2)$$

then gives (8).

A.4 Proof of the positive penalty

At $\gamma = 1$,

$$\alpha_{p,c} = \alpha_0 + \langle Z, C \rangle , \qquad C = \frac{\|v\|^p v}{pc^p}.$$

Hence

$$\Sigma_{p,c} - \Sigma_0 = 2\mathbb{E}[\alpha_0 \langle Z, C \rangle] + \mathbb{E}\|C\|^2 .$$

The last term is

$$\mathbb{E}\|C\|^2 = \frac{1}{p^2 c^{2p}} \mathbb{E}\|v\|^{2p+2} .$$

Gaussian fourth-moment contraction gives

$$\mathbb{E}_Z[\alpha_0 \langle Z, C \rangle] = \frac{1}{4} \langle C, Hv \rangle - \frac{1}{4} \langle C, \nabla(\Delta V) \rangle .$$

Applying integration by parts under $\pi \propto e^{-V}$ to the vector field HC yields

$$\mathbb{E} \langle C, Hv - \nabla(\Delta V) \rangle = \mathbb{E}[H : \nabla C].$$

Since

$$D_v C(v) = \frac{1}{pc^p} \left(\|v\|^p I + p \|v\|^{p-2} vv^\top \right),$$

we obtain

$$H : \nabla C = \frac{1}{pc^p} \left(\|v\|^p \|H\|_F^2 + p \|v\|^{p-2} \|Hv\|^2 \right),$$

which proves (10).

B Nondegeneracy of the limiting variance

For $0 < \gamma < 1$,

$$\Sigma_\gamma = \mathbb{E}[\beta_{p,c}^2] = \frac{1}{p^2 c^{2p}} \mathbb{E}\|v\|^{2p+2} ,$$

which is positive unless the score vanishes almost surely. For $\gamma > 1$, $\Sigma_\gamma = \Sigma_0$; at $\gamma = 1$, Proposition 4.3 gives $\Sigma_{p,c} \geq \Sigma_0$. Thus the usual MALA nondegeneracy condition $\Sigma_0 > 0$ is sufficient for all $\gamma \geq 1$. If $\Sigma_0 = 0$, higher-order terms must be examined.

C ADDITIONAL STATIONARY BENCHMARK RESULTS

Table 2: Full stationary benchmark. Entries are $100 \times Q_{0.25}$ bulk ESS per retained-sampling gradient (mean $\pm$ sample SD across six repetitions).

Target	MALA	MALTA	TMALA	TMALAc	Barker	TamedMALA
Gaussian	$9.285 \pm .067$	$9.210 \pm .130$	$1.271 \pm .027$	$3.677 \pm .032$	$4.196 \pm .092$	$9.225 \pm .125$
Aniso. Gaussian	$0.215 \pm .004$	$0.216 \pm .002$	$0.216 \pm .004$	$0.217 \pm .002$	$0.214 \pm .002$	$0.214 \pm .005$
Hyperbolic	$4.020 \pm .071$	$4.080 \pm .087$	$0.810 \pm .026$	$2.352 \pm .104$	$2.749 \pm .119$	$4.026 \pm .063$
Quartic	$2.733 \pm .072$	$2.787 \pm .040$	$1.065 \pm .014$	$1.930 \pm .034$	$1.924 \pm .051$	$2.784 \pm .036$
Sextic	$1.315 \pm .015$	$1.298 \pm .054$	$0.787 \pm .016$	$1.105 \pm .051$	$1.080 \pm .026$	$1.296 \pm .034$
Double well	$9.853 \pm .111$	$9.893 \pm .112$	$2.458 \pm .054$	$6.270 \pm .115$	$4.275 \pm .049$	$9.765 \pm .215$
Ginzburg–Landau	$2.317 \pm .046$	$2.370 \pm .032$	$0.899 \pm .039$	$1.672 \pm .039$	$1.646 \pm .045$	$2.266 \pm .040$
Planar singular	$8.584 \pm .108$	$8.578 \pm .068$	$1.249 \pm .021$	$3.598 \pm .046$	$4.080 \pm .061$	$8.383 \pm .220$
Curved singular	$7.736 \pm .792$	$8.186 \pm .249$	$1.208 \pm .034$	$3.406 \pm .130$	$3.881 \pm .179$	$8.132 \pm .231$
Poisson RE	$2.608 \pm .186$	$2.777 \pm .097$	$2.488 \pm .111$	$2.565 \pm .223$	$2.076 \pm .053$	$2.705 \pm .152$

Table 3: Planar-singular target diagnostics. Entries are $100\times$ bulk ESS per retained-sampling gradient evaluation (mean $\pm$ sample SD across six end-to-end replicates).

Observable	MALA	MALTA	TMALAc	Barker	**TamedMALA**
$\lvert x_1 \rvert$	17.801 ± 3.526	19.974 ± 1.729	8.190 ± 0.263	8.779 ± 0.179	$\mathbf{20.502 \pm 0.586}$
x_1	0.381 ± 0.003	0.381 ± 0.010	0.368 ± 0.004	0.324 ± 0.003	$\mathbf{0.385 \pm 0.008}$
$\mathrm{sign}(x_1)$	0.246 ± 0.002	0.245 ± 0.006	0.243 ± 0.003	0.210 ± 0.001	$\mathbf{0.248 \pm 0.005}$

D ADDITIONAL p-ABLATION RESULTS

Table 4: Stationary p-ablation with production parameters frozen. Entries are $Q_{0.25}$ bulk ESS/gradient relative to an independently evaluated MALA chain using the corresponding production proposal variance (mean $\pm$ sample SD across six production-parameter repetitions). Within each target and repetition, the same headline TamedMALA (δ, c) is used for every finite p.

Target	$p = 1$	$p = 1.5$	$p = 2$	$p = 3$	$p = 4$
Quartic	0.987 ± 0.038	1.022 ± 0.055	1.021 ± 0.045	1.019 ± 0.053	1.023 ± 0.044
Sextic	0.958 ± 0.039	0.996 ± 0.044	0.990 ± 0.039	0.999 ± 0.047	0.991 ± 0.037
Planar singular	0.756 ± 0.123	0.939 ± 0.050	0.958 ± 0.029	0.965 ± 0.074	0.994 ± 0.022
Poisson RE	1.040 ± 0.263	1.035 ± 0.243	1.056 ± 0.267	1.030 ± 0.238	0.954 ± 0.209

E BASELINE TUNING PROTOCOL AND EVALUATION SEPARATION

All headline samplers use the same target-specific diagnostic panel, the same primary objective $Q_{0.25}$ bulk ESS per retained-sampling gradient, and the same broad-then-refine search architecture. The parameters exposed by each proposal are:

For every target–sampler pair, the search starts from a target-scale proposal variance, evaluates a broad multiplicative grid, and geometrically refines the best region; MALTA additionally searches its truncation threshold D, while the two TamedMALA variants additionally search c. Tuning chains are then discarded and the frozen parameters are evaluated on fresh chains with independent proposal random streams. Consequently, using the same ESS/gradient quantity as tuning objective and reported metric does not leak final-evaluation data into model selection. This protocol does *not* equalize total search compute across methods with different hyperparameter dimension; the manuscript therefore makes no equal-tuning-cost claim.

Algorithm 1 Joint short-chain tuning used for TamedMALA.

1: Evaluate a broad multiplicative grid in (δ, c) by $Q_{0.25}$ bulk ESS/gradient.
2: **while** the selected point lies on a search boundary **do**
3: Expand the corresponding axis and evaluate the new candidates.
4: **end while**
5: Refine geometrically around the current maximizer.
6: Freeze $(\widehat{\delta}, \widehat{c})$, discard tuning chains, and run fresh evaluation chains.

Table 5: Parameters tuned in the stationary benchmark. TMALA and TMALAc have the fixed Brosse-form taming strength implied by their definitions, so only the proposal variance is searched.

Sampler	Tuned parameters
MALA	δ
MALTA	δ, D
TMALA	δ
TMALAc	δ
Barker	proposal scale δ
TamedMALA-global	δ, c
TamedMALA-coordinate	δ, c

F TUNER BUDGET ACCOUNTING

The budget labels used internally correspond to actual score-evaluation costs, not qualitative notions of compute. Counting all warm-up and retained proposals used during the parameter search, the realized budgets were:

Table 6: Realized tuner cost in proposal-gradient evaluations, mean $\pm$ sample SD across three seeds. Adaptive bracketing causes small seed-dependent variation in the number of evaluated candidates.

Target	Low	Default	High
Gaussian	$(1.34 \pm 0.17) \times 10^5$	$(6.29 \pm 0.60) \times 10^5$	$(1.34 \pm 0.08) \times 10^6$
Sextic	$(1.26 \pm 0.11) \times 10^5$	$(5.76 \pm 0.46) \times 10^5$	$(1.41 \pm 0.03) \times 10^6$

Under a common independent long-chain validation, the low/default/high tuner settings achieve relative efficiencies

	low	default	high
Gaussian	0.978 ± 0.017	0.992 ± 0.011	0.995 ± 0.004
Sextic	0.892 ± 0.048	0.959 ± 0.003	0.956 ± 0.045.

Thus the deliberately small budget begins to hurt on sextic, whereas increasing the default budget by more than a factor of two yields no systematic gain.

AI USE STATEMENT

Generative AI tools, primarily ChatGPT, were used during the development of this work. They assisted with mathematical derivations, implementation and debugging of experimental code, and editing of the manuscript. AI-generated mathematical arguments, code, and research suggestions were reviewed and checked by the authors before being incorporated into the work.