A MATRIX OPTIMIZATION PERSPECTIVE ON BATCH BAYESIAN OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Batch Bayesian optimization (BO) selects candidates with dependent outcomes, motivating coordinated updates in joint acquisition optimization. We study this problem via a matrix representation of the batch, relating posterior dependence to gradient modes that combine candidate movements with input directions. For radial kernels, we characterize single-point EI and chained batch-UCB curvature through a shared metric and rank-bounded corrections. We derive an exact local comparison and a sharp perturbation bound identifying when Muon's normalization of active singular modes improves quadratic gain, motivating Muon-BO. We establish a regret bound for chained batch-UCB with explicit acquisition optimization error, account for clipping in projected Muon updates, and connect inner iterations to regret under additional assumptions. Experiments on GP draws, synthetic benchmarks, and a rover task examine the theory and demonstrate practical benefits.

1 INTRODUCTION

Bayesian optimization (BO) uses a predictive model to select evaluations of an expensive black-box function (Sacks et al., 1989; Jones et al., 1998), with a batch format frequently implemented to run multiple evaluations in parallel (Rasmussen & Williams, 2006; Srinivas et al., 2010; Desautels et al., 2014; Wilson et al., 2018). Each round selects a batch of q points in a d-dimensional space, which naturally formulates a matrix decision variable $\mathbf{X} \in \mathbb{R}^{q\times d}$, with one candidate per row. Under a Gaussian-process (GP) model, these outcomes are generally dependent: a candidate with high marginal uncertainty may contribute little additional information when another candidate already covers that uncertainty. Batch selections therefore assess candidates together, as incorporated in multipoint expected improvement and GP-BUCB, which use the joint predictive distribution and conditional uncertainty, respectively (Ginsbourger et al., 2010; Chevalier & Ginsbourger, 2013; Desautels et al., 2014).

This dependence also matters when optimizing the acquisition. For two candidates near a promising region, moving both toward higher predicted values and adjusting their relative positions can affect exploitation and conditional uncertainty differently. Joint solvers such as multistart L-BFGS-B already account for candidate interactions by optimizing the full batch acquisition (Byrd et al., 1995; Balandat et al., 2020). The choice of update thus determines how movement is distributed across both candidates and shared input directions, which motivates a matrix view of the optimization. The batch matrix $\mathbf{X}$ representation makes this structure explicit, with rows indexing candidates and columns indexing input coordinates. The singular modes of its gradient $\mathbf{G} = \nabla\alpha(\mathbf{X})$ pair candidate combinations with input directions, with singular values measuring their first-order strengths. Gradient ascent scales each mode by its singular value, and if a dominant mode encounters strong adverse curvature, the resulting restriction on step size can leave weaker, flatter modes with little movement. Muon replaces each positive singular value by one, giving equal weight to active modes and increasing the relative movement along weaker ones (Jordan et al., 2024; Liu et al., 2025), which can strengthen within-batch rearrangements when they lie in weaker active modes. This redistribution can also reduce the alignment with the gradient, trading first-order gain for more balanced movement across active modes. It therefore improves the optimally scaled local quadratic gain when the reduction in directional curvature is sufficient to offset this loss. In this work, we study the Muon-style polar update as an inner solver for joint batch acquisition optimization, characterize when acquisition curvature favors Muon's allocation of movement across candidate combinations, and connect acquisition optimization accuracy to BO regret through chained batch-UCB.

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 contributions are summarized as follows:

- **Dependence and acquisition geometry.** We express how posterior covariance derivatives enter the batch gradient and interpret its singular modes as coordinated candidate movements. For radial kernels with a constant prior mean, we characterize the single-point EI Hessian (Theorem 5.1) and each chained batch-UCB Hessian block (Corollary 5.2) as a scalar multiple of the shared kernel metric plus a rank-bounded correction.

- **Curvature conditions for polar updates.** We compare optimally scaled gradient and polar steps on the local quadratic model and establish a sharp perturbation bound for their gain margin (Propositions 5.3–5.4). These results identify the curvature alignment required for a strict polar-step advantage and its robustness to residual curvature.

- **Optimization accuracy and regret.** We establish a cumulative regret bound for jointly optimized chained batch-UCB with the GP-BUCB rate form plus an additive acquisition optimization error (Theorem 4.1). For projected Muon, we bound a corrected ascent measure (Theorem 6.1) and, under additional basin and restart assumptions, express the solver's contribution to the regret bound in terms of inner iterations (Corollary A.11).

- **Empirical evaluation.** We examine acquisition curvature and optimization behavior on GP-draw objectives and evaluate BO performance on synthetic benchmarks and a 60-dimensional rover task. Comparisons across acquisitions and dimensions assess where the benefits of Muon persist.

2 Related Work

Dependent batch utilities and joint optimization. Multipoint expected improvement and parallel knowledge gradient account for candidate dependence through the joint predictive distribution (Chevalier & Ginsbourger, 2013; Wu & Frazier, 2016), while local penalization provides another mechanism for coordinating batch selection (González et al., 2016). Analytic gradients and stochastic-gradient methods enable joint maximization of multipoint expected improvement (Marmin et al., 2015; Wang et al., 2020), and Monte Carlo methods extend gradient-based optimization to broader acquisition families (Balandat et al., 2020). Reformulations such as qLogEI further improve numerical stability (Ament et al., 2023). These approaches establish both dependence-aware batch utilities and methods for optimizing them. We study how the matrix geometry of the joint update interacts with acquisition curvature.

Batch BO regret and optimization error. GP-BUCB establishes a regret bound for greedy batch construction (Desautels et al., 2014) and proves a batch analogue of the GP-UCB rate (Srinivas et al., 2010). GP-UCB-PE and BBKB give guarantees for other sequential-within-batch constructions (Contal et al., 2013; Calandriello et al., 2020), while DPP and parallel Thompson-sampling analyses concern randomized selection rules (Kathuria et al., 2016; Nava et al., 2022; Takeno & Iwazaki, 2026). Inexact acquisition maximization has also been studied for UCB and Thompson sampling (Kim et al., 2025). We analyze jointly optimized chained batch-UCB, making the acquisition value error explicit in the regret bound and connecting it to Muon inner iterations.

Matrix preconditioning and polar updates. Muon exploits the singular structure of matrix gradients by normalizing the active singular values (Jordan et al., 2024; Liu et al., 2025), with connections to spectral-norm steepest descent and trust-region methods (Bernstein & Newhouse, 2024; Kovalev, 2025). Recent work analyzes its convergence and curvature-dependent behavior (Shen et al., 2025; Wang et al., 2026). We apply this update geometry to batch acquisition optimization, derive a sharp local comparison with the gradient direction, and account for coordinate-box projection in the inner-iteration analysis.

High-dimensional BO. Trust regions (TuRBO) (Eriksson et al., 2019), sparse priors (SAASBO) (Eriksson & Jankowiak, 2021), and embeddings (REMBO) (Wang et al., 2016) address the challenges of BO in high-dimensional search by modifying the surrogate function and changing the outer loop. Our work concerns the acquisition solver for a given posterior and feasible region, making its focus complementary to these outer-loop approaches.

3 PRELIMINARIES

Problem setup. We consider maximizing an expensive black-box function $f : \Omega \to \mathbb{R}$ over the box $\Omega = \prod_{j=1}^{d}[l_j, u_j]$, where $l_j < u_j$, and assume that a maximizer $\mathbf{x}^\star \in \arg\max_{\mathbf{x}\in\Omega} f(\mathbf{x})$ exists. Evaluations are collected in batches of q points, with independent Gaussian observation noise of variance $\sigma_n^2 > 0$. In theory, $N = qT$ counts T rounds of q evaluations after the initial design of size n_{init}. Experimental budgets include initialization, and a bound for total regret also includes the initialization regret. The theory uses the maximization convention; the implementation equivalently minimizes $-f$. We model f with a Gaussian process (GP) surrogate with constant prior mean m and radial kernel $k(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|_{\mathbf{\Lambda}})$, $\|\mathbf{v}\|_{\mathbf{\Lambda}} = (\mathbf{v}^\mathsf{T}\mathbf{\Lambda}^{-1}\mathbf{v})^{1/2}$, $\mathbf{\Lambda} = \operatorname{diag}(\ell_1^2, \ldots, \ell_d^2)$, and $k(\mathbf{x}, \mathbf{x}) = a^2 \in (0, 1]$. The surrogate uses the observation noise variance σ_n^2, and Appendix A.2.2 discusses an exception. Observed data determine the posterior mean $\mu(\cdot)$ and covariance. For the regret analysis, the kernel and noise variance remain fixed across rounds, following Srinivas et al. (2010) and Desautels et al. (2014). We write $\sigma^2(\mathbf{x} \mid S)$ for the posterior variance after additionally conditioning on noisy observations at locations in a multiset S. Finally, let γ_N denote the maximal information gain of N noisy observations where repeated locations are allowed (Srinivas et al., 2010).

Posterior dependence and acquisition geometry. Let $\mathcal{D}_t$ denote the observations available at round t. For a candidate matrix $\mathbf{X} \in \mathbb{R}^{q\times d}$ with rows $\mathbf{x}_i^\mathsf{T}$, the GP posterior with fixed hyperparameters gives $\mathbf{f}_\mathbf{X} \mid \mathcal{D}_t \sim \mathcal{N}\big(\boldsymbol{\mu}_t(\mathbf{X}), \mathbf{\Sigma}_t(\mathbf{X})\big)$ and $[\mathbf{\Sigma}_t(\mathbf{X})]_{ij} = c_t(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{f}_\mathbf{X} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_q))^\mathsf{T}$, $\boldsymbol{\mu}_t(\mathbf{X}) = (\mu_t(\mathbf{x}_1), \ldots, \mu_t(\mathbf{x}_q))^\mathsf{T}$, and c_t is the posterior covariance (Rasmussen & Williams, 2006). Thus $\mathbf{X}$ contains the candidate locations, while $\mathbf{\Sigma}_t(\mathbf{X})$ describes their predictive dependence under the GP surrogate. Batch acquisitions use this joint predictive law to assess candidates together; Section 4 makes this dependence explicit for chained batch-UCB.

At a fixed posterior, suppose $\alpha(\mathbf{X}) = F_t(\boldsymbol{\mu}_t(\mathbf{X}), \mathbf{\Sigma}_t(\mathbf{X}))$ is differentiable. Define $\mathbf{b}$ and symmetric $\mathbf{W}$ by $dF_t = \mathbf{b}^\mathsf{T} d\boldsymbol{\mu} + \langle \mathbf{W},\, d\mathbf{\Sigma}\rangle$. The chain rule gives

$$\nabla_{\mathbf{x}_i}\alpha = b_i \nabla\mu_t(\mathbf{x}_i) + 2\sum_{j=1}^{q} W_{ij}\nabla_1 c_t(\mathbf{x}_i, \mathbf{x}_j), \tag{1}$$

where ∇_1 differentiates the first covariance argument, and symmetry gives the factor two, including on the diagonal. For chained batch-UCB, $b_i = 1$ and $\mathbf{W}$ differentiates the conditional-uncertainty bonus, including its exploration multiplier. This identity shows how posterior covariance enters candidate gradients; further differentiation yields cross-candidate Hessian blocks that need not vanish. Vectorization preserves these interactions, while the matrix representation organizes updates by candidate combinations and shared input directions.

Polar update and Muon. A batch is a matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_q)^\mathsf{T} \in \Omega^q \subset \mathbb{R}^{q\times d}$, and throughout $q \le d$. For $\mathbf{G} \in \mathbb{R}^{q\times d}$ with reduced singular value decomposition $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$, the polar factor is $\operatorname{Polar}(\mathbf{G}) = \mathbf{U}\mathbf{V}^\mathsf{T}$, and $\operatorname{Polar}(0) = 0$. Each singular mode $\mathbf{u}_r\mathbf{v}_r^\mathsf{T}$, formed from the corresponding columns of $\mathbf{U}$ and $\mathbf{V}$, pairs a combination of candidates with a shared input direction, moving candidate i along $u_{r,i}\mathbf{v}_r$. The polar update replaces each positive singular value by one, increasing the relative movement along weaker active modes; whether this redistribution improves local progress depends on the acquisition's directional curvature (Section 5). The inner solver we analyze is the momentum-free *Muon* update in Algorithm 1. Throughout, *Muon* names this algorithm, *the polar step* names its update, and Muon-BO denotes the batch-BO loop that uses it. For an acquisition α and initial iterate $\mathbf{X}_0$, define $\Delta_0 = \sup_{\Omega^q}\alpha - \alpha(\mathbf{X}_0)$. Appendix A.1 collects the facts about the classical polar factor used throughout (Jordan et al., 2024; Kovalev, 2025; Pethick et al., 2025).

Algorithm 1 Muon inner maximizer (one restart)

1: **Input:** $\mathbf{X}_0 \in \Omega^q$, steps T_{in}, step sizes $\{\eta_k\}$
2: **for** $k = 0, \ldots, T_{\text{in}} - 1$ **do**
3: $\quad\mathbf{G}_k \leftarrow \nabla\alpha(\mathbf{X}_k)$
4: $\quad\mathbf{X}_{k+1} \leftarrow \Pi_\Omega\big(\mathbf{X}_k + \eta_k \operatorname{Polar}(\mathbf{G}_k)\big)$, where Π_Ω is the row-wise projection onto Ω
5: **end for**
6: **Return** the iterate $\mathbf{X}_0, \ldots, \mathbf{X}_{T_{\text{in}}}$ with the best acquisition value

4 Regret of Jointly Optimized Batches

We first define the batch acquisition used in our regret analysis. At round t, let μ_t and σ_t denote the GP posterior mean and standard deviation given all observations collected so far. For a batch $\mathbf{X}$ with rows $\mathbf{x}_1, \dots, \mathbf{x}_q$, define

$$\mathcal{U}_t(\mathbf{X}) \;=\; \sum_{i=1}^{q} \Big[\mu_t(\mathbf{x}_i) + \beta_t^{1/2}\, \sigma_t\big(\mathbf{x}_i \,\big|\, \mathbf{x}_{1:i-1}\big)\Big]. \tag{2}$$

Here, $\sigma_t(\mathbf{x}_i \mid \mathbf{x}_{1:i-1})$ is the posterior standard deviation after additionally conditioning on noisy observations at the preceding batch locations, with $\sigma_t(\cdot \mid \emptyset) = \sigma_t(\cdot)$. This variance update requires only the locations of those points, and the posterior mean μ_t remains unchanged. GP-BUCB (Desautels et al., 2014) selects each point greedily by maximizing its corresponding confidence bound. We instead sum them into one acquisition function and optimize all batch points jointly. We call this objective *chained batch-UCB*. The conditional variances can be computed through Schur complements, and their gradients are available analytically (Appendix C). Under positive observation noise, $\mathcal{U}_t$ remains twice continuously differentiable on Ω^q whenever the kernel $\mathbf{x} \mapsto k(\mathbf{x}, \mathbf{x}')$ is, even when batch points coincide (Appendix A.2.1). To derive a regret guarantee, we require the confidence bound to remain valid after conditioning on pending locations:

Assumption 1 (Hallucinated confidence)**.** *There exist nondecreasing $\beta_t > 0$ and an event $\mathcal{E}$ with $\Pr(\mathcal{E}) \geq 1 - \delta$ on which, for all rounds t, all $\mathbf{x} \in \Omega$, and all multisets $S \subset \Omega$ with $|S| \leq q - 1$,*

$$|\mu_t(\mathbf{x}) - f(\mathbf{x})| \;\leq\; \beta_t^{1/2}\, \sigma_t(\mathbf{x} \,|\, S).$$

Here, hallucination means updating the posterior covariance at pending locations without observing their outcomes. Assumption 1 can be justified using the confidence construction of Desautels et al. (2014). With C bounding the hallucinated information gain over the relevant S, their Proposition 1 gives $\sigma_t(\mathbf{x} \,|\, S) \geq e^{-C}\sigma_t(\mathbf{x})$, so an $\mathbf{x}$-uniform sequential GP-UCB band (Srinivas et al., 2010) with parameter θ_t implies Assumption 1 with $\beta_t = e^{2C}\theta_t$. The initialization procedure of Desautels et al. (2014) controls the post-initialization inflation factor independently of q, at an initialization cost that generally grows with q. Appendix A.2.2 provides details and discusses the GP-prior setting, where the regret bound includes an additional $O(q)$ term.

Theorem 4.1 (Regret of jointly optimized chained batch-UCB)**.** *Let Assumption 1 hold, $k(\mathbf{x}, \mathbf{x}) \leq 1$, fix the GP hyperparameters across rounds, and let feedback be batch-synchronous (the observations of round t are available before round $t + 1$). Suppose at every round $t \leq T$ the evaluated batch $\hat{\mathbf{X}}_t = (\hat{\mathbf{x}}_{t,1}, \dots, \hat{\mathbf{x}}_{t,q})^\mathsf{T} \in \Omega^q$ is an ϵ_t-approximate joint maximizer for some $\epsilon_t \geq 0$: $\mathcal{U}_t(\hat{\mathbf{X}}_t) \geq \sup_{\mathbf{X} \in \Omega^q} \mathcal{U}_t(\mathbf{X}) - \epsilon_t$. Then, on $\mathcal{E}$, the cumulative regret $R_N = \sum_{t \leq T} \sum_{i \leq q} [f(\mathbf{x}^\star) - f(\hat{\mathbf{x}}_{t,i})]$ over $N = qT$ evaluations satisfies*

$$R_N \;\leq\; \sqrt{C_1 N \beta_T \gamma_N} \;+\; \sum_{t=1}^{T} \epsilon_t, \qquad C_1 = \frac{8}{\log(1 + \sigma_n^{-2})}.$$

The value gap ϵ_t is the statistical interface between the joint solver and the regret analysis. Exact maximization removes this term and recovers the GP-BUCB rate. Approximate maximization adds at most the sum of the per-round value gaps to the stated upper bound. Theorem 4.1 therefore provides a guarantee for joint batch optimization with an explicit solver-error term, rather than establishing a sharper rate or better realized regret than greedy batch construction (Remark A.7). This also identifies what the inner solver must deliver, while allowing its update rule to be chosen according to acquisition geometry. Figure 1 illustrates this result on objectives drawn from a GP prior, presenting cumulative regret R_N against the evaluation budget N for several joint acquisition (UCB) solvers and greedy GP-BUCB. After the initial design, the joint methods exhibit similar cumulative-regret growth to the greedy comparator and broadly follow the $\sqrt{N\gamma_N}$ reference guide.

5 The Curvature of Batch Acquisitions

Theorem 4.1 identifies acquisition value accuracy as the relevant output of the joint solver. We now examine when the geometry induced by a GP acquisition favors the polar update. Equation (1) shows

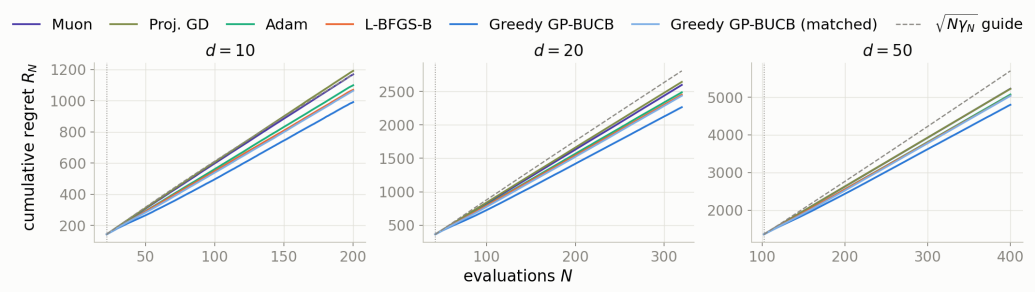


Figure 1: Cumulative regret R_N against evaluations N from n_{init} (dotted vertical line) on GP-draw objectives with Matérn-5/2 kernel. The surrogate uses the true hyperparameters, with $\beta_t \equiv 4$, $n_{\text{init}} = 2d + 2$, and $q = 8$. Each panel shows one dimension $d \in \{10, 20, 50\}$, with mean R_N over 50 draws (25 at $d = 50$). Curves are (2) solved by Muon, projected GD, Adam and multistart L-BFGS-B from the same $R = 8$ uniform start batches, and greedy GP-BUCB with matched and unlimited budgets. The dashed line refers to the $\sqrt{N\gamma_N}$ rate.

how statistical dependence enters first derivatives. We characterize the single-point EI Hessian and each chained batch-UCB Hessian block as a shared kernel metric plus a rank-bounded correction (Theorem 5.1 and Corollary 5.2). This structure alone does not imply a Muon advantage, as the generic Frobenius-smoothness rates give the same $\sqrt{q}$ factor when expressed in nuclear-norm stationarity (Shen et al., 2025). Our comparison therefore examines the curvature along the actual gradient and polar directions. Propositions 5.3–5.4 identify when the directional curvature makes polar normalization improve the optimally scaled local quadratic gain.

Theorem 5.1 (Structure of the EI Hessian, radial kernels)**.** *Let the GP have constant prior mean m and kernel $k(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|_{\mathbf{\Lambda}})$ with kernel sections $\mathbf{x} \mapsto k(\mathbf{x}, \mathbf{x}_i)$ twice continuously differentiable (RBF, or Matérn-ν with $\nu \geq 5/2$), n observations, and $\sigma(\mathbf{x}) > 0$. Write* EI *in the minimization convention of the implementation,* $\mathrm{EI}(\mathbf{x}) = u\,\Phi(z) + \sigma(\mathbf{x})\phi(z)$ *with $u = f^\star - \mu(\mathbf{x}) - \xi$ and $z = u/\sigma(\mathbf{x})$ (for maximization of f replace $f^\star - \mu$ by $\mu - f^\star$; the decomposition and the range statement for $\mathbf{R}(\mathbf{x})$ below are unchanged, and the coefficient becomes $C(\mathbf{x}) = +\Phi(z)c_\mu(\mathbf{x}) + \phi(z)c_\sigma(\mathbf{x})$, i.e. the sign of the Φ-term flips in the closed forms). Then, at every $\mathbf{x}$,*

$$\nabla^2 \mathrm{EI}(\mathbf{x}) \;=\; C(\mathbf{x})\,\mathbf{\Lambda}^{-1} \;+\; \mathbf{R}(\mathbf{x}),$$

where $\mathbf{R}(\mathbf{x})$ is symmetric with $\operatorname{range}\mathbf{R}(\mathbf{x}) \subseteq V_x := \operatorname{span}\{\mathbf{\Lambda}^{-1}(\mathbf{x} - \mathbf{x}_1), \ldots, \mathbf{\Lambda}^{-1}(\mathbf{x} - \mathbf{x}_n)\}$, so $\operatorname{rank}\mathbf{R}(\mathbf{x}) \leq \min(n, d)$, and $C(\mathbf{x}) = -\Phi(z)\,c_\mu(\mathbf{x}) + \phi(z)\,c_\sigma(\mathbf{x})$ is in closed form (coefficients in App. A.3). For the RBF kernel $\kappa(r) = a^2 e^{-r^2/2}$ (any $\mathbf{\Lambda}$), with $z = (f^\star - \mu(\mathbf{x}) - \xi)/\sigma(\mathbf{x})$,

$$C(\mathbf{x}) = \Phi(z)\,\big(\mu(\mathbf{x}) - m\big) + \phi(z)\,\frac{a^2 - \sigma^2(\mathbf{x})}{\sigma(\mathbf{x})}.$$

Corollary 5.2 (Structure of the chained batch-UCB)**.** *Under the kernel and constant-prior-mean conditions of Theorem 5.1 and $\sigma_n^2 > 0$, every $d \times d$ block of $\nabla^2 \mathcal{U}_t(\mathbf{X})$ (diagonal and off-diagonal) decomposes as a scalar multiple of $\mathbf{\Lambda}^{-1}$ plus a residual matrix whose row and column spaces lie in the span of the $\mathbf{\Lambda}^{-1}$-scaled pairwise differences among the active points (the n data points and the rows of $\mathbf{X}$). In particular the residual matrix of each block has rank at most $\min(n + q - 1,\, d)$.*

The decomposition identifies how the shared kernel metric and differences among active points enter acquisition curvature. The rank bounds apply to the residual blocks and do not imply low effective rank of the full batch Hessian or a polar-step advantage. To determine when polar normalization helps, Proposition 5.3 compares the optimally scaled quadratic gains along the gradient and polar directions using their directional curvatures. Proposition 5.4 then gives a sharp perturbation bound for the gain margin, identifying when this advantage persists uniformly under bounded residual curvature.

Proposition 5.3 (One-step gains on the local model)**.** *Fix $\mathbf{X}$ with $\mathbf{G} = \nabla\alpha(\mathbf{X}) \neq 0$, let $\mathcal{H}$ be the negative-Hessian operator acting on $q \times d$ increments, and let $m(\mathbf{D}) = \langle \mathbf{G}, \mathbf{D}\rangle - \frac{1}{2}\langle \mathbf{D}, \mathcal{H}[\mathbf{D}]\rangle$ be the local model. For any $\mathbf{Z}$ with $\langle \mathbf{Z}, \mathcal{H}[\mathbf{Z}]\rangle > 0$ the exact line-search gain is $\max_{\eta \in \mathbb{R}} m(\eta\mathbf{Z}) = \frac{\langle \mathbf{G}, \mathbf{Z}\rangle^2}{2\langle \mathbf{Z}, \mathcal{H}[\mathbf{Z}]\rangle}$. If both $\langle \mathbf{G}, \mathcal{H}[\mathbf{G}]\rangle > 0$ and $\langle \mathbf{Q}, \mathcal{H}[\mathbf{Q}]\rangle > 0$ for $\mathbf{Q} = \mathrm{Polar}(\mathbf{G})$, then with $\rho = \operatorname{rank}\mathbf{G}$, $r_{\text{eff}}(\mathbf{G}) = \|\mathbf{G}\|_*^2/\|\mathbf{G}\|_F^2 \in [1, \rho]$, and $S(\mathbf{Z}) = \langle \mathbf{Z}, \mathcal{H}[\mathbf{Z}]\rangle/\|\mathbf{Z}\|_F^2$,*

$$\frac{\textit{gain along } Q}{\textit{gain along } G} = \frac{r_{\text{eff}}(\mathbf{G})}{\rho} \cdot \frac{S(\mathbf{G})}{S(\mathbf{Q})}.$$

Proposition 5.4 (Sharp alignment margin). *Using the notation of Proposition 5.3, set* $\bar{r} = r_{\mathrm{eff}}(\mathbf{G})/\rho \in (0,1]$, $\bar{\mathbf{G}} = \mathbf{G}/\|\mathbf{G}\|_F$ *and* $\bar{\mathbf{Q}} = \mathbf{Q}/\sqrt{\rho}$, *so* $\|\bar{\mathbf{G}}\|_F = \|\bar{\mathbf{Q}}\|_F = 1$ *and* $\langle \bar{\mathbf{G}}, \bar{\mathbf{Q}} \rangle = \sqrt{\bar{r}}$. *Write* $\mathcal{H} = \mathcal{H}_0 + \mathcal{P}$ *with* $\mathcal{H}_0$ *and* $\mathcal{P}$ *self-adjoint and* $\|\mathcal{P}\|_{F\to F} \le \tau$, *where* $\|\cdot\|_{F\to F}$ *is the operator norm induced by the Frobenius norm, and define* $\Gamma_0 = \bar{r}\,\langle \bar{\mathbf{G}}, \mathcal{H}_0[\bar{\mathbf{G}}] \rangle - \langle \bar{\mathbf{Q}}, \mathcal{H}_0[\bar{\mathbf{Q}}] \rangle$ *and* $\Gamma = \bar{r}\,S(\mathbf{G}) - S(\mathbf{Q})$. *Then*

$$|\Gamma - \Gamma_0| \;\le\; \tau\sqrt{(1-\bar{r})(1+3\bar{r})}, \tag{3}$$

and the bound is attained by a self-adjoint perturbation of rank at most two. When both $\langle \mathbf{G}, \mathcal{H}[\mathbf{G}] \rangle > 0$ *and* $\langle \mathbf{Q}, \mathcal{H}[\mathbf{Q}] \rangle > 0$, *the polar direction has the strictly larger exact line-search gain on the local model exactly when* $\Gamma > 0$. *If* $\mathcal{H}_0 \succeq h_0 \mathcal{I}$ *for the identity operator* $\mathcal{I}$ *and some* $h_0 > \tau$, *then*

$$\Gamma_0 \;>\; \tau\sqrt{(1-\bar{r})(1+3\bar{r})} \tag{4}$$

is necessary and sufficient for this strict advantage to hold uniformly over all self-adjoint $\mathcal{P}$ *with* $\|\mathcal{P}\|_{F\to F} \le \tau$.

Proposition 5.3 identifies a tradeoff between gradient alignment and directional curvature: when both directional curvatures are positive, polar normalization improves the optimally scaled quadratic gain exactly when $S(\mathbf{Q}) < \bar{r}S(\mathbf{G})$. For chained batch-UCB, the structured negative-Hessian operator incorporates the shared kernel metric and candidate coupling, while Proposition 5.4 determines whether its favorable margin withstands bounded residual curvature under the stated positivity assumption. These conditions do not follow from posterior dependence or the block-rank bounds alone; under $\mathcal{H} = c\mathcal{I}$ with $c > 0$, the gain ratio is $\bar{r} \le 1$. Appendix B.2 examines the decomposition on Hessian snapshots along Muon-driven runs on GP draws, where the spectra show a flat bulk at the empirically estimated block coefficients with few outlying eigenvalues.

6 Muon as the Inner Solver

The local curvature comparison explains which acquisition geometries favor the polar direction. To apply the polar update in BO, we must also account for the feasible domain Ω^q and the accuracy reached after a finite inner budget. Algorithm 1 takes a polar step and projects it onto the coordinate box. We analyze this procedure for a generic smooth acquisition, including chained batch-UCB. The first result bounds progress with an explicit clipping correction, and the subsequent value conversion requires additional assumptions. Appendix A.5 extends the analysis to Monte Carlo gradients.

Assumption 2 (Frobenius smoothness). α *is continuously differentiable on* Ω^q *and* $\|\nabla\alpha(\mathbf{X}) - \nabla\alpha(\mathbf{Y})\|_F \le L_F\|\mathbf{X} - \mathbf{Y}\|_F$ *for all* $\mathbf{X}, \mathbf{Y} \in \Omega^q$.

Theorem 6.1 (Projected Muon). *Let Assumption 2 hold and run Algorithm 1 from* $\mathbf{X}_0 \in \Omega^q$ *with constant step* $\eta > 0$. *Let* $\mathcal{A}_k$ *be the set of coordinates clipped by the projection at step* k *and* $E_k = \sum_{(i,j)\in\mathcal{A}_k} |G_k^{ij}||Q_k^{ij}|$ *(here* $\mathbf{Q}_k = \mathrm{Polar}(\mathbf{G}_k)$*). Then, for every* $T_{\mathrm{in}} \ge 1$,

$$\frac{1}{T_{\mathrm{in}}} \sum_{k=0}^{T_{\mathrm{in}}-1} \Big(\|\nabla\alpha(\mathbf{X}_k)\|_* - E_k \Big) \le \frac{\Delta_0}{T_{\mathrm{in}}\eta} + \frac{L_F\, q\, \eta}{2},$$

and Δ_0 *may be replaced throughout by any upper bound* $\Delta \ge \Delta_0$. *If no clipping occurs (so that* $E_k \equiv 0$*) and* $\Delta_0, L_F > 0$, *then the step size* $\eta = \sqrt{2\Delta_0/(L_F q T_{\mathrm{in}})}$ *yields*

$$\frac{1}{T_{\mathrm{in}}} \sum_k \|\nabla\alpha(\mathbf{X}_k)\|_* \le \sqrt{\frac{2L_F q \Delta_0}{T_{\mathrm{in}}}}.$$

Under the same no-clipping condition, if additionally $\|\nabla\alpha(\mathbf{X}) - \nabla\alpha(\mathbf{Y})\|_* \le L_*\|\mathbf{X} - \mathbf{Y}\|_2$ *on* Ω^q *with* $L_* > 0$ *(mixed-norm smoothness condition), the step size* $\eta = \sqrt{2\Delta_0/(L_* T_{\mathrm{in}})}$ *gives* $\sqrt{2L_*\Delta_0/T_{\mathrm{in}}}$. *When* $L_* < qL_F$, *this improves the preceding bound by a factor* $\sqrt{qL_F/L_*}$.

Theorem 6.1 bounds a corrected ascent measure that accounts for box projection through E_k. Without clipping, $\langle \mathbf{G}, \mathrm{Polar}(\mathbf{G}) \rangle = \|\mathbf{G}\|_*$ gives the first-order ascent gain, and substantial clipping can make $\|\mathbf{G}_k\|_* - E_k$ nonpositive and the bound uninformative. This measure therefore does not certify constrained stationarity. Appendix A.4 also presents a smooth reparameterization that avoids

explicit projection while preserving Assumption 2 (Proposition A.9). The factor $\sqrt{qL_F/L_*}$ compares two upper bounds for Muon, while Propositions 5.3–5.4 separately identify conditions for a local advantage over the gradient direction. To connect inner iterations to regret, we additionally assume a local Polyak–Łojasiewicz (PŁ) inequality, which converts a small gradient into a small acquisition value gap within a basin (Karimi et al., 2016). Under the basin and restart assumptions stated in Appendix A.4, Corollary A.11 combines Theorem 6.1 with Theorem 4.1 to bound the solver's regret contribution in terms of inner iterations.

Figure 2 illustrates the theoretical predictions on GP-draw objectives. The left panel evaluates the quantity bounded by Theorem 6.1 on frozen batch-UCB surfaces, using the theorem's step-size formula with empirical estimates of its constants. The dotted curves are corresponding plug-in bound estimates, which are not certified upper bounds because the sampled estimates need not bound the required constants globally. The observed quantity roughly follows the $T_{\text{in}}^{-1/2}$ decay. The mid panel reports the raw nuclear norm without the clipping correction, which also approximately follows the $T_{\text{in}}^{-1/2}$ rate. The right panel shows cumulative regret from the full BO procedure, normalized by its value at $T_{\text{in}} = 5$, using Muon with the step size prescribed by Corollary A.11 with empirical estimates of its constants. Regret generally decreases as the inner budget grows, with a more pronounced initial plateau at higher dimensions. This trend is qualitatively consistent with the decreasing solver-error term in the corollary's regret bound.

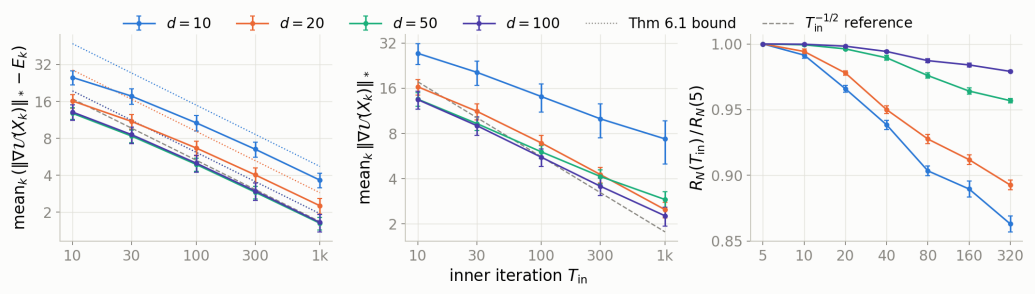


Figure 2: Left panel shows the trajectory-averaged $\|\nabla\mathcal{U}(\mathbf{X}_k)\|_* - E_k$ versus T_{in} on a log-log scale for Muon, evaluated on six frozen mid-run batch-UCB surfaces per dimension from GP draws ($q = 8$). The dotted curves show the corresponding plug-in bound estimates $\sqrt{2L_F q\Delta_0/T_{\text{in}}}$ (coincident for $d = 50$ and $d = 100$), with L_F and Δ_0 estimated on each surface, and the dashed line marks a $T_{\text{in}}^{-1/2}$ reference. The middle panel presents the trajectory-averaged raw nuclear norm $\|\nabla\mathcal{U}(\mathbf{X}_k)\|_*$ from the same runs, without the clipping correction. This quantity is not directly bounded by the theorem and is presented for reference. Right panel displays cumulative regret R_N, normalized by its value at $T_{\text{in}} = 5$, versus the inner budget T_{in} for Muon with the step size prescribed by Corollary A.11 over 30 GP draws.

7 Experiments

We evaluate Muon as an inner solver for batch BO on Levy, Rastrigin, and Ackley (Surjanovic & Bingham, 2013) at $d \in \{10, 20, 50, 100\}$, followed by a 60-dimensional rover trajectory-planning task (Wang et al., 2018; Eriksson et al., 2019). We compare Muon with projected gradient ascent (projected GD), Adam (Kingma & Ba, 2015), multistart L-BFGS-B following the BoTorch `optimize_acqf` recipe (Balandat et al., 2020), and compute-matched greedy GP-BUCB. All comparisons in this section use batch-UCB (2) with $\beta_t \equiv 4$. Further experimental details are reported in Appendix B.4. The fixed-step methods use step size $\eta = s\,\bar{w}/\sqrt{d}$, where $\bar{w}$ is the average box width and s is tuned on frozen acquisition surfaces for each dimension by the acquisition value (Appendix C.2). For the benchmark experiments, we run 20 seeds at $d \le 50$ and 10 at $d = 100$ and report mean $\pm$ standard deviation over seeds. Appendix B.4 also includes results for qLogEI (Ament et al., 2023) with joint optimization and sequential-greedy construction, as well as unlimited greedy GP-BUCB and random screening. We present (best-so-far) simple regret in the main text and report the cumulative regret R_N curves in Appendices B.4.2–B.4.4, and Appendix B.4.6 reports the full numerical simple-regret results for all methods. Appendix B.4.7 examines the effect of batch size q.

7.1 LEVY, RASTRIGIN, AND ACKLEY

Levy is a multimodal benchmark with many local minima sitting on a smooth bowl, and its difficulty grows with d as the number of local minima grows. The top row of Figure 3 plots the best simple regret so far against N for the five methods. At $d = 10$, their final performance is similar relative to the variability across seeds: Muon achieves 20.4 ± 16, compared with 27.1 ± 16 for L-BFGS-B, the next-best method by mean regret. Starting from $d = 20$, Muon gains an advantage early and maintains it through the evaluation budget. Its final regret is 50.5 ± 21 at $d = 20$, compared with 84.1 ± 25 for compute-matched greedy GP-BUCB, and 131 ± 16 at $d = 50$, compared with 182 ± 72 for Adam. The largest gap occurs at $d = 100$, where Muon achieves 51.5 ± 20, compared with 323 ± 149 for compute-matched greedy GP-BUCB, while projected GD makes little progress beyond the initial design. These results demonstrate the practical benefit of Muon on Levy.

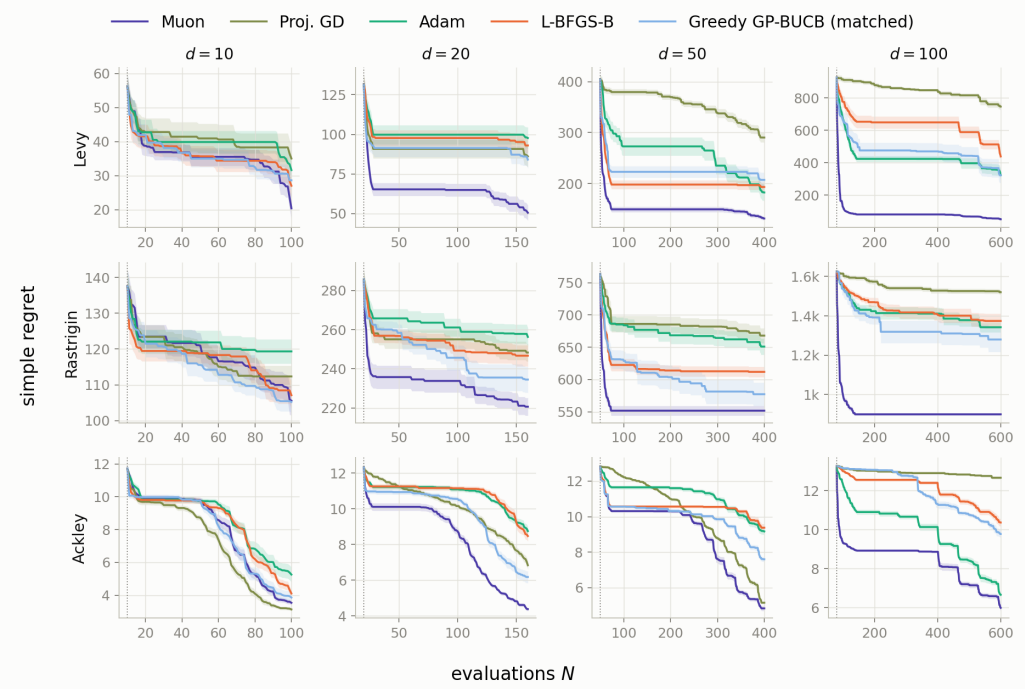


Figure 3: (Best-so-far) simple regret on Levy (top), Rastrigin (mid), and Ackley (bottom) using batch-UCB with $q = 8$, plotted against the number of evaluations N from the initial design n_{init} (dotted line) onward. Panels correspond to $d = 10, 20, 50, 100$. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

Rastrigin is a separable, multimodal benchmark with regularly spaced local minima. The middle row of Figure 3 plots simple regret against the evaluation budget N. Similarly, at $d = 10$, all five methods end within one standard deviation, with Muon achieving 106 ± 18 and compute-matched greedy GP-BUCB reaching 105 ± 13, and none outperforms the random screen (103 ± 15, Table 5), which shows no clear practical separation among the tested inner solvers at this dimension. Muon achieves the lowest mean regret among the five plotted methods at $d \geq 20$, with 221 ± 20 and 552 ± 34 at $d = 20$ and $d = 50$, compared with 234 ± 30 and 577 ± 80 for compute-matched greedy GP-BUCB. The gap widens at $d = 100$, where Muon approaches its final regret within the first 15 rounds, while the other plotted methods are still descending at the end of the budget.

Ackley has a narrow central basin surrounded by a relatively flat, oscillatory landscape. The bottom row of Figure 3 plots the simple regret against N. At $d = 10$, projected GD achieves the lowest mean regret (3.14 ± 0.37), compared with 3.54 ± 0.49 for Muon as the second-best method. For $d \geq 20$, Muon starts outperforming other methods. At $d = 20$ and $d = 50$, Muon achieves 4.38 ± 0.41 and 4.83 ± 0.63, respectively. The closest methods are compute-matched greedy GP-BUCB at $d = 20$ (6.18 ± 1.5) and projected GD at $d = 50$ (5.15 ± 0.31). At $d = 100$, Muon achieves 5.99 ± 0.47, compared with 6.66 ± 0.38 for Adam, while projected GD makes limited progress. These results

illustrate the benefits of Muon as an inner acquisition solver, particularly in the higher-dimensional configurations tested here.

7.2 Rover trajectory planning

The Rover task (Wang et al., 2018; Eriksson et al., 2019) optimizes a trajectory through a two-dimensional obstacle field. The trajectory is specified by 30 waypoints, giving a 60-dimensional search space $[0, 1]^{60}$, and the reward is 5 minus the trajectory cost. We evaluate Muon as an inner solver for the chained batch-UCB acquisition and compare the performance to the same methods used in the benchmark experiment above. Detailed setup is reported in Appendix B.4.5. Figure 4 plots best-so-far reward and cumulative regret. At $N = 600$, projected GD, Muon, and L-BFGS-B achieve similar final rewards relative to the variability across seeds, with 1.93 ± 1.04, 1.83 ± 0.87, and 1.62 ± 0.69, respectively. All three attain higher mean rewards than unlimited greedy GP-BUCB (0.33 ± 0.92), compute-matched greedy (0.27 ± 1.24), and Adam (0.20 ± 1.21). Unlimited greedy also requires approximately 18 times as many gradient evaluations as the joint solvers. Muon achieves the lowest mean cumulative regret ($7{,}130 \pm 400$), compared to $7{,}520 \pm 400$ for projected GD. These results demonstrate that Muon remains competitive on the rover task, with a comparable final reward and a lower mean cumulative regret.

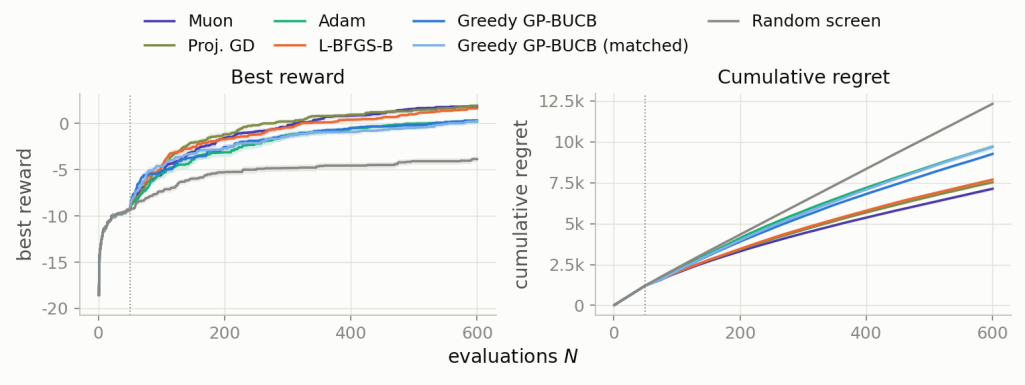


Figure 4: Best-so-far reward (left; higher is better) and cumulative regret (right; lower is better) on Rover-60 using batch-UCB with $q = 8$, $N = 600$, and $n_{\text{init}} = 50$, plotted against the number of evaluations. Curves show means over 20 seeds, with bands indicating $\pm$ one standard error. All gradient-based joint solvers use the same BoTorch-style starting batches, with 10 restarts and an iteration cap of 200 per restart.

8 Conclusion and Discussion

We study batch BO through the matrix geometry of joint acquisition optimization. Posterior dependence motivates coordinated candidate movements, and our curvature analysis identifies when polar normalization improves local quadratic gain. We establish a regret bound with explicit joint optimization error and connect projected Muon iterations to this error under additional assumptions. Experiments show that Muon is competitive across many tested configurations.

Candidate dependence and the Hessian decomposition motivate the matrix perspective, but dependence and rank-bounded curvature corrections alone do not imply a polar-step advantage, as it additionally requires favorable alignment between the gradient and acquisition curvature. Our regret analysis also assumes fixed GP hyperparameters and inherits GP-BUCB's confidence inflation. Extending it to hybrid models (Luo et al., 2024) and acquisitions such as qEI (Chevalier & Ginsbourger, 2013) and qLogEI (Ament et al., 2023), and reducing the inflation (Remark A.6), remain open directions.

Finally, box projection can destroy ascent, and our corrected ascent measure does not certify constrained stationarity. Converting inner iterations into acquisition accuracy requires additional basin, PŁ, and restart assumptions (Corollary A.11). A future extension could connect feasible update directions to both curvature and a constrained-stationarity measure, so that the local design interpretation remains useful near the boundary.

REPRODUCIBILITY STATEMENT

The main-text results are proved in Appendix A. Additional theory results are stated with their assumptions and proved in Appendix A. The settings in which Assumption 1 is available are discussed in Appendix A.2.2. The experimental protocol (surrogate, acquisitions, budgets, refit schedule, seeds and the shared initial design, the BoTorch settings copied for L-BFGS-B and sequential greedy, the compute-matching rule for greedy, and the step-size protocol with the exact step constants used) is given in Appendix C, and the inner wall-clock times for the main benchmark comparisons are reported in Appendix C.3. The implementation is NumPy/SciPy. The runner and analysis scripts for every experiment, the per-run results with their seeds, and the step tables and the scripts that produce every figure and table of this paper are included in the supplementary materials.

ETHICS STATEMENT

This work studies optimization algorithms on synthetic benchmarks, on functions drawn from a Gaussian-process prior, and on a simulated rover planning task. It involves no human subjects, no personal data and no application with a foreseeable harmful use; we see no ethical concerns beyond those common to optimization research.

AI USE STATEMENT

In this work, we used a large language model as a writing and coding assistant. It drafted and edited manuscript text from the authors' outline and results, wrote and revised analysis and plotting scripts, and checked LaTeX. We have not used generative AI tools for generating research ideas, for proving the theorems, or for producing experimental results. The theorems and proofs were written and audited by the authors, and every reported number is produced by the released experiment scripts from the recorded runs. We have reviewed all AI-assisted text and code and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Jeremy Bernstein and Laker Newhouse. Old optimizer, new norm: An anthology. *arXiv preprint arXiv:2409.20325*, 2024.

Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

Daniele Calandriello, Luigi Carratino, Alessandro Lazaric, Michal Valko, and Lorenzo Rosasco. Near-linear time Gaussian process optimization with adaptive batching and resparsification. In *International Conference on Machine Learning (ICML)*, 2020.

Clément Chevalier and David Ginsbourger. Fast computation of the multi-points expected improvement with applications in batch selection. In *Learning and Intelligent Optimization (LION 2013)*, volume 7997 of *Lecture Notes in Computer Science*, pp. 59–69. Springer, 2013. doi: 10.1007/978-3-642-44973-4_7.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 844–853, 2017.

Emile Contal, David Buffoni, Alexandre Robicquet, and Nicolas Vayatis. Parallel Gaussian process optimization with upper confidence bound and pure exploration. In *Joint European Conference*

on Machine Learning and Knowledge Discovery in Databases, volume 8188 of *Lecture Notes in Computer Science*, pp. 225–240. Springer, 2013.

Thomas Desautels, Andreas Krause, and Joel W. Burdick. Parallelizing exploration-exploitation tradeoffs in Gaussian process bandit optimization. *Journal of Machine Learning Research*, 15: 4053–4103, 2014.

Colin Doumont, Donney Fan, Natalie Maus, Jacob R. Gardner, Henry Moss, and Geoff Pleiss. We still don't understand high-dimensional Bayesian optimization. In *Artificial Intelligence and Statistics (AISTATS)*, 2026.

David Eriksson and Martin Jankowiak. High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In *Uncertainty in artificial intelligence*, pp. 493–503. PMLR, 2021.

David Eriksson, Michael Pearce, Jacob Gardner, Ryan D. Turner, and Matthias Poloczek. Scalable global optimization via local Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Michael P. Friedlander and Mark Schmidt. Hybrid deterministic-stochastic methods for data fitting. *SIAM Journal on Scientific Computing*, 34(3):A1380–A1405, 2012. doi: 10.1137/110830629.

David Ginsbourger, Rodolphe Le Riche, and Laurent Carraro. Kriging is well-suited to parallelize optimization. In Yoel Tenne and Chi-Keong Goh (eds.), *Computational Intelligence in Expensive Optimization Problems*, volume 2 of *Adaptation, Learning, and Optimization*, pp. 131–162. Springer, 2010. doi: 10.1007/978-3-642-10701-6_6.

Javier González, Zhenwen Dai, Philipp Hennig, and Neil D. Lawrence. Batch Bayesian optimization via local penalization. In *Artificial Intelligence and Statistics (AISTATS)*, pp. 648–657. PMLR, 2016.

Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998. doi: 10.1023/A:1008306431147.

Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024. URL `https://kellerjordan.github.io/posts/muon/`.

Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition. In *Joint European conference on machine learning and knowledge discovery in databases*, pp. 795–811. Springer, 2016.

Tarun Kathuria, Amit Deshpande, and Pushmeet Kohli. Batched Gaussian process bandit optimization via determinantal point processes. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Hwanwoo Kim, Chong Liu, and Yuxin Chen. Bayesian optimization with inexact acquisition: Is random grid search sufficient? *arXiv preprint arXiv:2506.11831*, 2025.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Dmitry Kovalev. Understanding gradient orthogonalization for deep learning via non-Euclidean trust-region optimization. *arXiv preprint arXiv:2503.12645*, 2025.

Tim Tsz-Kit Lau, Qi Long, and Weijie Su. PolarGrad: A class of matrix-gradient optimizers from a unifying preconditioning perspective. *arXiv preprint arXiv:2505.21799*, 2025.

Jingyuan Liu, Jianlin Su, et al. Muon is scalable for LLM training. *arXiv preprint arXiv:2502.16982*, 2025.

Hengrui Luo, Younghyun Cho, James W Demmel, Xiaoye S Li, and Yang Liu. Hybrid parameter search and dynamic model selection for mixed-variable Bayesian optimization. *Journal of Computational and Graphical Statistics*, 33(3):855–868, 2024.

Sébastien Marmin, Clément Chevalier, and David Ginsbourger. Differentiating the multipoint expected improvement for optimal batch design. In *Machine Learning, Optimization, and Big Data (MOD 2015)*, volume 9432 of *Lecture Notes in Computer Science*, pp. 37–48. Springer, 2015. doi: 10.1007/978-3-319-27926-8_4.

Shuntaro Nagashima and Hideaki Iiduka. Improved convergence rates of Muon optimizer for nonconvex optimization. *arXiv preprint arXiv:2601.19400*, 2026.

Elvis Nava, Mojmír Mutný, and Andreas Krause. Diversified sampling for batched Bayesian optimization with determinantal point processes. In *Artificial Intelligence and Statistics (AISTATS)*, 2022.

Thomas Pethick, Wanyun Xie, Kimon Antonakopoulos, Zhenyu Zhu, Antonio Silveti-Falls, and Volkan Cevher. Training deep learning models with norm-constrained LMOs. In *International Conference on Machine Learning (ICML)*, 2025. arXiv:2502.07529.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989. doi: 10.1214/ss/1177012413.

Naoki Sato, Hiroki Naganuma, and Hideaki Iiduka. Convergence bound and critical batch size of Muon optimizer. *arXiv preprint arXiv:2507.01598*, 2025.

Wei Shen, Ruichuan Huang, Minhui Huang, Cong Shen, and Jiawei Zhang. On the convergence analysis of Muon. *arXiv preprint arXiv:2505.23737*, 2025.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *International Conference on Machine Learning (ICML)*, 2010.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, 2012. doi: 10.1109/TIT.2011.2182033.

Sonja Surjanovic and Derek Bingham. Virtual library of simulation experiments: Test functions and datasets, 2013. URL `http://www.sfu.ca/~ssurjano`.

Shion Takeno and Shogo Iwazaki. Improved regret analysis for parallel Gaussian process bandit optimization. *arXiv preprint arXiv:2608.16492*, 2026.

Jialei Wang, Scott C. Clark, Eric Liu, and Peter I. Frazier. Parallel Bayesian global optimization of expensive functions. *Operations Research*, 68(6):1850–1865, 2020. doi: 10.1287/opre.2019.1966.

Shuche Wang, Fengzhuo Zhang, Jiaxiang Li, Dirk Bergemann, and Zhuoran Yang. Why Muon outperforms Adam: A curvature perspective. *arXiv preprint arXiv:2606.04662*, 2026.

Zi Wang, Clement Gehring, Pushmeet Kohli, and Stefanie Jegelka. Batched large-scale Bayesian optimization in high-dimensional spaces. In *Artificial Intelligence and Statistics (AISTATS)*, 2018.

Ziyu Wang, Frank Hutter, Masrour Zoghi, David Matheson, and Nando de Freitas. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55:361–387, 2016.

James T. Wilson, Frank Hutter, and Marc Peter Deisenroth. Maximizing acquisition functions for Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Jian Wu and Peter I. Frazier. The parallel knowledge gradient method for batch Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

A PROOFS

This appendix contains the complete proofs, in main-text order: the classical polar facts used throughout (Appendix A.1), the regret theory of Section 4 (Appendix A.2), the curvature results of Section 5 (Appendix A.3), the solver results of Section 6 (Appendix A.4), and the stochastic extension of the solver theory, whose statement and proof are given in Appendix A.5.

A.1 POLAR FACTS AND ASCENT INEQUALITIES

Throughout the appendix we freely use the singular value decomposition and von Neumann's trace inequality: for $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{q\times d}$ with singular values $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \dots$ and $\sigma_1(\mathbf{B}) \geq \sigma_2(\mathbf{B}) \geq \dots$,

$$\langle \mathbf{A},\, \mathbf{B}\rangle = \operatorname{trace}(\mathbf{A}^\mathsf{T}\mathbf{B}) \;\leq\; \sum_i \sigma_i(\mathbf{A})\,\sigma_i(\mathbf{B}). \tag{5}$$

The following facts about the polar factor are classical (Jordan et al., 2024; Kovalev, 2025; Bernstein & Newhouse, 2024; Pethick et al., 2025); they are used in Section 3 and in every proof below.

Lemma A.1 (Polar facts; classical)**.** *Let $\mathbf{G} \in \mathbb{R}^{q\times d}$ have rank $\rho \geq 1$ and reduced SVD $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$, and set $\mathbf{Q} = \operatorname{Polar}(\mathbf{G}) = \mathbf{U}\mathbf{V}^\mathsf{T}$. Then: (i) $\mathbf{Q} \in \arg\max\{\langle \mathbf{G},\, \mathbf{Z}\rangle : \|\mathbf{Z}\|_2 \leq 1\}$ and $\langle \mathbf{G},\, \mathbf{Q}\rangle = \|\mathbf{G}\|_*$; (ii) if $\rho = q$, $\mathbf{Q}$ minimizes $\|\mathbf{G} - \mathbf{Z}\|_F$ over matrices with orthonormal rows; (iii) $\|\mathbf{Q}\|_2 = 1$ and $\|\mathbf{Q}\|_F^2 = \rho$.*

A.1.1 PROOF OF LEMMA A.1

Write $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ with $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_\rho)$, $\sigma_1 \geq \dots \geq \sigma_\rho > 0$, and $\mathbf{U} \in \mathbb{R}^{q\times\rho}$, $\mathbf{V} \in \mathbb{R}^{d\times\rho}$ having orthonormal columns ($\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}_\rho$). Let $\mathbf{Q} = \mathbf{U}\mathbf{V}^\mathsf{T}$.

(i). For any $\mathbf{Z}$ with $\|\mathbf{Z}\|_2 \leq 1$ we have $\sigma_i(\mathbf{Z}) \leq 1$ for all i, so by (5), $\langle \mathbf{G},\, \mathbf{Z}\rangle \leq \sum_i \sigma_i(\mathbf{G})\sigma_i(\mathbf{Z}) \leq \sum_i \sigma_i(\mathbf{G}) = \|\mathbf{G}\|_*$. The candidate $\mathbf{Q}$ is feasible, since its nonzero singular values are those of $\mathbf{U}\mathbf{V}^\mathsf{T}$, namely ρ ones (because $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{V}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{V}^\mathsf{T}$ is the orthogonal projector onto the column space of $\mathbf{V}$, whose nonzero eigenvalues are ρ ones). It attains the bound:

$$\langle \mathbf{G},\, \mathbf{Q}\rangle = \operatorname{trace}\left(\mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathsf{T}\,\mathbf{U}\mathbf{V}^\mathsf{T}\right) = \operatorname{trace}\left(\mathbf{V}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}\right) = \operatorname{trace}\left(\mathbf{\Sigma}\mathbf{V}^\mathsf{T}\mathbf{V}\right) = \operatorname{trace}\mathbf{\Sigma} = \|\mathbf{G}\|_*,$$

using $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_\rho$, the cyclic property of the trace, and $\mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}_\rho$.

(ii). Let $\mathbf{Z}$ satisfy $\mathbf{Z}\mathbf{Z}^\mathsf{T} = \mathbf{I}_q$ (orthonormal rows; this set is nonempty iff $q \leq d$, which $\rho = q$ forces). Then $\|\mathbf{Z}\|_F^2 = \operatorname{trace}(\mathbf{Z}\mathbf{Z}^\mathsf{T}) = q$, so

$$\|\mathbf{G} - \mathbf{Z}\|_F^2 = \|\mathbf{G}\|_F^2 - 2\langle \mathbf{G},\, \mathbf{Z}\rangle + \|\mathbf{Z}\|_F^2 = \|\mathbf{G}\|_F^2 + q - 2\langle \mathbf{G},\, \mathbf{Z}\rangle :$$

minimizing $\|\mathbf{G} - \mathbf{Z}\|_F$ over the set is equivalent to maximizing $\langle \mathbf{G},\, \mathbf{Z}\rangle$ over it. Every such $\mathbf{Z}$ has $\|\mathbf{Z}\|_2 = 1$, so by (i), $\langle \mathbf{G},\, \mathbf{Z}\rangle \leq \|\mathbf{G}\|_*$; and when $\rho = q$ the matrix $\mathbf{Q} = \mathbf{U}\mathbf{V}^\mathsf{T}$ has $\mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{U}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{U}^\mathsf{T} = \mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{I}_q$ ($\mathbf{U}$ is then square orthogonal), so $\mathbf{Q}$ is feasible and attains the maximum.

(iii). Established in (i): the nonzero singular values of $\mathbf{Q}$ are ρ ones, hence $\|\mathbf{Q}\|_2 = 1$ and $\|\mathbf{Q}\|_F^2 = \rho$. □

A.1.2 ASCENT INEQUALITIES

Lemma A.2. *Let $\mathbf{X}, \mathbf{X} + \mathbf{D} \in \Omega^q$. Under Assumption 2,*

$$\alpha(\mathbf{X} + \mathbf{D}) \;\geq\; \alpha(\mathbf{X}) + \langle \nabla\alpha(\mathbf{X}),\, \mathbf{D}\rangle - \tfrac{L_F}{2}\|\mathbf{D}\|_F^2.$$

Under the mixed-norm smoothness $\|\nabla\alpha(\mathbf{X}) - \nabla\alpha(\mathbf{Y})\|_* \leq L_*\|\mathbf{X} - \mathbf{Y}\|_2$,

$$\alpha(\mathbf{X} + \mathbf{D}) \;\geq\; \alpha(\mathbf{X}) + \langle \nabla\alpha(\mathbf{X}),\, \mathbf{D}\rangle - \tfrac{L_*}{2}\|\mathbf{D}\|_2^2.$$

Proof. Since Ω^q is convex, the segment $\{\mathbf{X} + t\mathbf{D} : t \in [0,1]\}$ lies in Ω^q, so the smoothness assumptions apply along it. By the fundamental theorem of calculus applied to $t \mapsto \alpha(\mathbf{X} + t\mathbf{D})$,

$$\alpha(\mathbf{X} + \mathbf{D}) - \alpha(\mathbf{X}) - \langle \nabla\alpha(\mathbf{X}),\, \mathbf{D}\rangle = \int_0^1 \langle \nabla\alpha(\mathbf{X} + t\mathbf{D}) - \nabla\alpha(\mathbf{X}),\; \mathbf{D}\rangle\, dt.$$

Under Assumption 2, Cauchy–Schwarz bounds the integrand:

$$\left|\langle \nabla\alpha(\mathbf{X}+t\mathbf{D}) - \nabla\alpha(\mathbf{X}),\, \mathbf{D}\rangle\right| \le \|\nabla\alpha(\mathbf{X}+t\mathbf{D}) - \nabla\alpha(\mathbf{X})\|_F \cdot \|\mathbf{D}\|_F \le L_F\, t\, \|\mathbf{D}\|_F^2,$$

and $\int_0^1 L_F t\, \|\mathbf{D}\|_F^2 dt = \frac{L_F}{2}\|\mathbf{D}\|_F^2$. Under mixed-norm smoothness, the Hölder pairing of the nuclear and spectral norms (a special case of (5): $|\langle \mathbf{A},\, \mathbf{B}\rangle| \le \sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}) \le \|\mathbf{B}\|_2 \sum_i \sigma_i(\mathbf{A}) = \|\mathbf{A}\|_* \|\mathbf{B}\|_2$) bounds the integrand by $\|\nabla\alpha(\mathbf{X}+t\mathbf{D}) - \nabla\alpha(\mathbf{X})\|_* \|\mathbf{D}\|_2 \le L_* t \|\mathbf{D}\|_2^2$, and the same integration applies. □

A.2 Proofs for Section 4

A.2.1 Well-posedness of the chained batch-UCB

We record the argument behind the differentiability remark of Section 4. For $\sigma_n^2 > 0$ and a kernel with $k(\mathbf{x},\mathbf{x}) = a^2 > 0$ whose sections $\mathbf{x} \mapsto k(\mathbf{x},\mathbf{x}')$ are C^2 (for a radial kernel this is the statement that $\mathbf{u} \mapsto \kappa(\|\mathbf{u}\|_{\mathbf{\Lambda}})$ is C^2, hence k is C^2 jointly; it holds for the RBF and Matérn-$5/2$ kernels used here), every chained conditional variance is strictly positive on all of Ω^q: the posterior variance of $f(\mathbf{x})$ given *noisy* observations $\mathbf{y}_S = \mathbf{f}_S + \boldsymbol{\varepsilon}$ never vanishes, since if it did, $f(\mathbf{x}) = \mathbf{c}^\mathsf{T}\mathbf{y}_S + b$ almost surely for some $\mathbf{c}, b$, so $f(\mathbf{x}) - \mathbf{c}^\mathsf{T}\mathbf{f}_S - b = \mathbf{c}^\mathsf{T}\boldsymbol{\varepsilon}$ with the two sides independent, hence both almost surely constant; this forces $\mathbf{c} = 0$ and then $f(\mathbf{x})$ constant a.s., contradicting $k(\mathbf{x},\mathbf{x}) = a^2 > 0$. Hence $\mathcal{U}_t$ is twice continuously differentiable on Ω^q, including at collapsed batches: the kernel supplies the derivatives, and positive noise keeps the square root smooth where a noiseless variance would vanish.

Two further properties of (2) are used implicitly. First, repeated rows earn only a noise-limited bonus. If row i repeats a location at which $m \ge 1$ of the earlier rows already sit, its chained variance is the posterior variance after m hallucinated noisy observations at that location. With s^2 the variance after conditioning on the available data and all earlier rows at other locations, assimilating the m repeats last gives $s^2\sigma_n^2/(\sigma_n^2 + ms^2) \le \sigma_n^2/m$, so the bonus of the repeated row is at most $\beta_t^{1/2}\sigma_n/\sqrt{m}$, whatever σ_t is there. Each extra copy keeps the mean term but adds a diminishing exploration bonus, so this does not rule out a collapsed batch. This is also why Assumption 1 is stated for multisets S, and the bound on C over multisets in Appendix A.2.2 covers exactly these repeated locations.

Second, $\mathcal{U}_t$ depends on the order of the rows, since row i is conditioned on rows $1, \dots, i-1$. We fix the order in which the rows of $\mathbf{X}$ are listed and use it consistently: it defines the prefixes $\hat{\mathbf{x}}_{t,<i}$ in Step 3 of the proof of Theorem 4.1 and the global order of the N evaluated points in Lemma A.4. The convention costs nothing in the bound, because the supremum in the theorem ranges over all of Ω^q and hence over every ordering of every set of q points, and it says nothing about how the batch is evaluated; the q points are queried in parallel as an unordered set.

A.2.2 Justification of Assumption 1

Throughout, the GP kernel and noise variance are fixed; posterior variances are then deterministic functionals of observation locations. Two ingredients:

(i) Uniform unhallucinated confidence. Assumption 1 is stated in the exact, $\mathbf{x}$-uniform form, with θ_t the sequential constant evaluated at one plus the number of observations available at the start of round t (the index $fb[t] + 1$ of Desautels et al., 2014), so that β_t is nondecreasing. Throughout, as in all of the cited confidence bounds, each batch is chosen as a measurable function of the history at the start of its round and of randomness independent of f and of the observation noise (as any solver acting on $\mathcal{U}_t$ is, and ϵ_t-optimality alone would not exclude a selector that inspects f). That form is available in two of the three settings of Srinivas et al. (2010; 2012).

In the finite-domain Bayesian setting (Srinivas et al., 2012, Thm. 1 and Lem. 5.1), which applies when the box Ω of Section 3 is replaced by a finite candidate set, a union bound over the candidates and over t gives constants θ_t and an event of probability $\ge 1 - \delta$ on which $|\mu_t(\mathbf{x}) - f(\mathbf{x})| \le \theta_t^{1/2}\sigma_t(\mathbf{x})$ for all t and all candidates $\mathbf{x}$. In the RKHS setting ($\|f - m\|_k$ bounded, and the bound is on $f - m$ because constants need not lie in the RKHS), the same $\mathbf{x}$-uniform band holds on the whole box with a θ_t that grows with $\|f - m\|_k^2$ and with γ_t times a polylogarithm of t/δ. Theorem 6 of Srinivas et al. (2012), which supports their Theorem 3, establishes such a band for conditionally zero-mean noise bounded almost surely in magnitude by the model's σ_n, and this assumption excludes Gaussian noise.

For conditionally sub-Gaussian noise, including Gaussian noise, a corresponding uniform confidence bound follows from Chowdhury & Gopalan (2017, Theorem 2).

The latter band is for the posterior computed with the regularization parameter $\lambda = 1 + \eta$ ($\eta = 2/(n_{\text{init}} + N)$ accounts for the full observation horizon, including initialization) in place of the noise variance, and the actual noise enters only through its sub-Gaussian proxy R in θ_t. So if that route is taken, the surrogate is the λ-regularized posterior, i.e. λ plays the role of σ_n^2 in (2), Assumption 1, Proposition 1 of Desautels et al. (2014), C_1 and γ_N (all of which are statements about the surrogate), while the true Gaussian observation noise, of whatever variance, affects only θ_t through R. For the box Ω of Section 3, the exact form is therefore justified when $\|f - m\|_k < \infty$, with bounded martingale-difference noise (Thm. 6 of Srinivas et al. (2012)) or, for Gaussian noise, with the surrogate using the regularization of the band (Thm. 2 of Chowdhury & Gopalan (2017)). It is not justified by these sources for an f sampled from the GP prior, whose paths have infinite RKHS norm almost surely (Srinivas et al., 2012), and that case is covered only by the following variant.

In the continuous-domain Bayesian setting (Srinivas et al., 2012, Thm. 2), under the sample-path derivative condition of that theorem, the band is not $\mathbf{x}$-uniform: it holds at the selected points (Srinivas et al., 2012, Lem. 5.5), by the non-anticipation just stated, so that a union bound over the q points of a round adds $2\log q$ to θ_t, and at a discretized comparator $[\mathbf{x}^\star]_t$ with an additive slack $1/t^2$ (Srinivas et al., 2012, Lems. 5.6–5.7). In that setting, Steps 1 and 3 of the proof of Theorem 4.1 go through with the comparator batch built from $[\mathbf{x}^\star]_t$, at the cost of q/t^2 per round, so the theorem holds with an additional additive $O(q\sum_t t^{-2}) = O(q)$ term in R_N. Indexing the discretization by evaluation rather than by round, as Desautels et al. (2014, Case 2, Lems. 13–17) do (row i of the comparator batch of round t is then $[\mathbf{x}^\star]_\tau$ with $\tau = q(t-1) + i$, and β_t is taken at the round's last index), the slack sums to $\sum_\tau \tau^{-2} \le \pi^2/6 < 2$, the additive constant in their Theorem 2. We state Assumption 1 and Theorem 4.1 in the exact form and record this variant here. (All of these results are stated for a zero prior mean; a known constant mean m is removed by centering the observations, which is how the implementation proceeds.)

(ii) Variance-ratio control under hallucination. Desautels et al. (2014, Prop. 1) show that for any conditioning multiset S,

$$\frac{\sigma_t(\mathbf{x})}{\sigma_t(\mathbf{x} \,|\, S)} \;=\; \exp\big(I(f(\mathbf{x}); \mathbf{y}_S \,|\, \text{data}_t)\big).$$

Hence if $C \ge \sup_{t,\mathbf{x},|S|\le q-1} I(f(\mathbf{x}); \mathbf{y}_S | \text{data}_t)$, then on the event of (i), for every such S,

$$|\mu_t(\mathbf{x}) - f(\mathbf{x})| \le \theta_t^{1/2}\sigma_t(\mathbf{x}) \le \theta_t^{1/2} e^C\, \sigma_t(\mathbf{x} \,|\, S) = \beta_t^{1/2}\sigma_t(\mathbf{x} \,|\, S), \qquad \beta_t = e^{2C}\theta_t$$

(the factor e^C multiplies a standard deviation once, hence appears squared in β_t). Note the S-uniformity is free: (ii) is a deterministic statement about variances, so no additional union bound over S is needed.

Bounding C over multisets. For a general multiset $S = \{\mathbf{s}_1, \dots, \mathbf{s}_m\}$, $m \le q-1$ (repetitions allowed), the chain rule for mutual information gives

$$\begin{aligned} I(f(\mathbf{x}); \mathbf{y}_S \,|\, \text{data}) = \sum_{j=1}^{m} I\big(f(\mathbf{x}); y_{\mathbf{s}_j} \,\big|\, y_{\mathbf{s}_{<j}}, \text{data}\big) \;&\le\; \sum_{j=1}^{m} \tfrac{1}{2}\log\big(1 + \sigma_n^{-2}\sigma_t^2(\mathbf{s}_j \,|\, \mathbf{s}_{<j})\big) \\ &\le\; (q-1)\max_{\mathbf{x}'} \tfrac{1}{2}\log\big(1 + \sigma_n^{-2}\sigma_t^2(\mathbf{x}')\big), \end{aligned}$$

where the middle inequality bounds each conditional mutual information by the information provided by the corresponding noisy observation ($I(f(\mathbf{x}); y_{\mathbf{s}_j}|\cdot) \le I(f(\mathbf{s}_j); y_{\mathbf{s}_j}|\cdot) = \frac{1}{2}\log(1 + \sigma_n^{-2}\sigma_t^2(\mathbf{s}_j|\mathbf{s}_{<j}))$; data-processing, as $y_{\mathbf{s}_j}$ depends on f only through $f(\mathbf{s}_j)$), and the last uses that conditioning shrinks variance. For m repeats at a single point this recovers $\frac{1}{2}\log(1 + m\sigma_t^2/\sigma_n^2) \le m \cdot \frac{1}{2}\log(1 + \sigma_t^2/\sigma_n^2)$, by Bernoulli's inequality $(1+s)^m \ge 1 + ms$. After the uncertainty-sampling initialization of Desautels et al. (2014, Thm. 5), posterior variances, and hence this bound, are driven below any fixed constant independent of q; alternatively, Desautels et al. (2014, Lem. 4) bound the maximal information gain of any $q-1$ points after T^{init} uncertainty-sampling evaluations by $(q-1)\gamma_{T^{\text{init}}}/T^{\text{init}}$, which is exactly a bound on C over multisets (their proof bounds each marginal gain in a sequence of $q-1$ actions after the greedy initialization, so repeated locations are covered). The bound persists at every later round: by data processing ($\mathbf{y}_S$ depends on f only

through $\mathbf{f}_S$), $I(f(\mathbf{x}); \mathbf{y}_S \mid \mathrm{data}_t) \le I(\mathbf{f}_S; \mathbf{y}_S \mid \mathrm{data}_t) = \frac{1}{2}\log\det(\mathbf{I} + \sigma_n^{-2}\mathbf{K}_S^{(t)})$, where $\mathbf{K}_S^{(t)}$ is the posterior covariance of $\mathbf{f}_S$ at round t, and conditioning on further observations can only shrink $\mathbf{K}_S^{(t)}$ in the Loewner order, hence this bound (the argument in the proof of Lemma A.4); the pointwise quantity $I(f(\mathbf{x}); \mathbf{y}_S \mid \cdot)$ itself is not monotone under conditioning, which is why the total gain is used. For $q \ge 2$, making this at most a prescribed C requires $\gamma_{T^{\mathrm{init}}}/T^{\mathrm{init}} \le C/(q-1)$, i.e. an initialization length $T^{\mathrm{init}} = T^{\mathrm{init}}(q, k, \sigma_n, C)$ that grows with q (finite whenever γ_T is sublinear); only the post-initialization multiplier e^{2C}, not the initialization budget, is independent of q; for $q = 1$ no observation is hallucinated, $C = 0$ and no initialization is needed.

A.2.3 Two auxiliary lemmas

Lemma A.3 (Variance-to-information conversion)**.** *If $0 \le v \le 1$ and $\sigma_n^2 > 0$, then $v \le \dfrac{\log(1+\sigma_n^{-2}v)}{\log(1+\sigma_n^{-2})}$.*

Proof. The function $g(u) = u/\log(1+u)$ is nondecreasing on $(0,\infty)$: $g'(u)$ has the sign of $\log(1+u) - u/(1+u)$, which is positive because $\log(1+u) > u/(1+u)$ for $u > 0$ (integrate $1/(1+t) \ge 1/(1+u)$ on $[0,u]$, strictly on the interior). Applying monotonicity with $u = \sigma_n^{-2}v \le \sigma_n^{-2}$: $\sigma_n^{-2}v/\log(1+\sigma_n^{-2}v) \le \sigma_n^{-2}/\log(1+\sigma_n^{-2})$, and rearranging gives the claim; $v = 0$ holds trivially. □

Lemma A.4 (Chained variances aggregate to information gain)**.** *Under batch-synchronous feedback, order the $N = qT$ evaluated points by round and, within a round, by row index, and let $\hat\sigma_{t,i} = \sigma_t(\hat{\mathbf{x}}_{t,i} \mid \hat{\mathbf{x}}_{t,<i})$. Then*

$$\sum_{t=1}^{T}\sum_{i=1}^{q} \hat\sigma_{t,i}^2 \;\le\; \frac{2}{\log(1+\sigma_n^{-2})}\,\gamma_N.$$

Proof. Under batch feedback, when round t begins, all points of rounds $< t$ have been observed; thus $\hat\sigma_{t,i}^2$ is the posterior variance of $f(\hat{\mathbf{x}}_{t,i})$ conditioned on the initial data and observations at *all* predecessors of $\hat{\mathbf{x}}_{t,i}$ in the stated global order (real observations for earlier rounds, hallucinated for the within-round prefix, identical for variances, which depend on locations only). The log-determinant identity underlying the information-gain chain rule (Srinivas et al., 2012, Lemma 5.3) therefore gives, for each realized sequence of locations,

$$\frac{1}{2}\log\det(\mathbf{I} + \sigma_n^{-2}\mathbf{K}_{\mathrm{post}}) = \frac{1}{2}\sum_{t,i}\log\left(1 + \sigma_n^{-2}\hat\sigma_{t,i}^2\right) \le \gamma_N,$$

where $\mathbf{K}_{\mathrm{post}}$ is the covariance at the N locations under the GP posterior conditioned on data_0, and data_0 denotes any observations made before round 1 (e.g. during initialization; $\mathrm{data}_0 = \emptyset$ is allowed). The last inequality holds because γ_N is defined over all sequences of N locations, repetitions included, and because the Gaussian information gain at those locations can only decrease under conditioning on data_0: the posterior covariance satisfies $\mathbf{K}_{\mathrm{post}} \preceq \mathbf{K}_{\mathrm{prior}}$, and $\log\det(\mathbf{I} + \sigma_n^{-2}\mathbf{K})$ is monotone in $\mathbf{K}$ for the Loewner order. Each $\hat\sigma_{t,i}^2 \le k(\hat{\mathbf{x}}_{t,i}, \hat{\mathbf{x}}_{t,i}) \le 1$, so Lemma A.3 applies with $v = \hat\sigma_{t,i}^2$:

$$\sum_{t,i}\hat\sigma_{t,i}^2 \le \frac{1}{\log(1+\sigma_n^{-2})}\sum_{t,i}\log\left(1+\sigma_n^{-2}\hat\sigma_{t,i}^2\right) = \frac{\log\det(\mathbf{I}+\sigma_n^{-2}\mathbf{K}_{\mathrm{post}})}{\log(1+\sigma_n^{-2})} \le \frac{2\gamma_N}{\log(1+\sigma_n^{-2})}.$$ □

A.2.4 Proof of Theorem 4.1

Work on the event $\mathcal{E}$ of Assumption 1. Fix a round t and abbreviate $\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{t,i}$, $\hat\sigma_i = \sigma_t(\hat{\mathbf{x}}_i \mid \hat{\mathbf{x}}_{1:i-1})$.

Step 1 (upper bound at q copies of the optimum). The batch $\mathbf{X}^\star = (\mathbf{x}^\star, \dots, \mathbf{x}^\star)^\mathsf{T}$ is feasible ($\mathbf{x}^\star \in \Omega$). Its chained variances are, by definition of (2), $\sigma_t(\mathbf{x}^\star \mid S_i)$ with $S_i = \{\mathbf{x}^\star\}^{i-1}$ ($i-1$ copies; $S_1 = \emptyset$). Each S_i is a multiset of size $i - 1 \le q - 1$, so Assumption 1 applies at $\mathbf{x} = \mathbf{x}^\star$ with $S = S_i$ and gives, for every i,

$$f(\mathbf{x}^\star) \;\le\; \mu_t(\mathbf{x}^\star) + \beta_t^{1/2}\sigma_t(\mathbf{x}^\star \mid S_i).$$

Summing over $i = 1, \ldots, q$:

$$q\, f(\mathbf{x}^\star) \;\le\; \sum_{i=1}^{q} \Big[\mu_t(\mathbf{x}^\star) + \beta_t^{1/2}\sigma_t(\mathbf{x}^\star \mid S_i)\Big] \;=\; \mathcal{U}_t(\mathbf{X}^\star).$$

Step 2 (approximate joint optimality). Since $\mathbf{X}^\star \in \Omega^q$,

$$\mathcal{U}_t(\mathbf{X}^\star) \;\le\; \sup_{\mathbf{X}\in\Omega^q} \mathcal{U}_t(\mathbf{X}) \;\le\; \mathcal{U}_t(\hat{\mathbf{X}}_t) + \epsilon_t.$$

Step 3 (lower bound at the selected batch). For each i, apply Assumption 1 at $\mathbf{x} = \hat{\mathbf{x}}_i$ with the multiset $S = \{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{i-1}\}$ ($|S| = i - 1 \le q - 1$; this is the same prefix, in the same row order, that defines $\hat{\sigma}_i$):

$$f(\hat{\mathbf{x}}_i) \;\ge\; \mu_t(\hat{\mathbf{x}}_i) - \beta_t^{1/2}\hat{\sigma}_i.$$

Summing over i: $\sum_i f(\hat{\mathbf{x}}_i) \ge \sum_i \mu_t(\hat{\mathbf{x}}_i) - \beta_t^{1/2}\sum_i \hat{\sigma}_i$.

Step 4 (per-round regret). Combining Steps 1–3 and expanding $\mathcal{U}_t(\hat{\mathbf{X}}_t) = \sum_i \mu_t(\hat{\mathbf{x}}_i) + \beta_t^{1/2}\sum_i \hat{\sigma}_i$ by definition:

$$\sum_{i=1}^{q} \big[f(\mathbf{x}^\star) - f(\hat{\mathbf{x}}_i)\big] \;\le\; \Big[\mathcal{U}_t(\hat{\mathbf{X}}_t) + \epsilon_t\Big] - \Big[\sum_i \mu_t(\hat{\mathbf{x}}_i) - \beta_t^{1/2}\sum_i \hat{\sigma}_i\Big] \;=\; 2\beta_t^{1/2}\sum_{i=1}^{q}\hat{\sigma}_i \;+\; \epsilon_t.$$

Step 5 (aggregation over rounds). Summing Step 4 over $t = 1, \ldots, T$, using $\beta_t \le \beta_T$ (monotonicity), then Cauchy–Schwarz over the N terms and Lemma A.4 gives:

$$\begin{aligned} R_N \;&\le\; 2\beta_T^{1/2}\sum_{t,i}\hat{\sigma}_{t,i} + \sum_t \epsilon_t \;\le\; 2\beta_T^{1/2}\sqrt{N\sum_{t,i}\hat{\sigma}_{t,i}^2} + \sum_t \epsilon_t \\ &\le\; 2\beta_T^{1/2}\sqrt{\frac{2N\gamma_N}{\log(1+\sigma_n^{-2})}} + \sum_t \epsilon_t, \end{aligned}$$

where $8/\log(1+\sigma_n^{-2}) = C_1$ gives $R_N \le \sqrt{C_1 N\beta_T\gamma_N} + \sum_t \epsilon_t$. □

Remark A.5. Joint optimality is used exactly once (Step 2), against one explicitly constructed comparator batch. The device is genuinely batch-level: a greedily built batch is not in general a near-maximizer of $\mathcal{U}_t$, so Steps 1–2 are unavailable to the greedy analysis, just as the greedy proof's per-point argmax step (Desautels et al., 2014, Lem. 10) is unavailable here; the two arguments are complementary, with identical rates.

Remark A.6 (Sharper constants). Our $\beta_t = e^{2C}\theta_t$ inflation and the initialization device are inherited from Desautels et al. (2014). Takeno & Iwazaki (2026) recently removed the analogous multiplicative batch-size factor for (sequentially selected) parallel Thompson sampling via an elliptical-potential counting argument, without initialization. Their technique controls how often the posterior standard deviation based on delayed observations substantially exceeds that based on all preceding locations. Takeno & Iwazaki (2026) note, however, that it does not directly tighten GP-BUCB, whose β_t is already inflated by the batch size, and that a UCB-based method achieving their bound would have to select diverse points without that inflation. Since $\mathcal{U}_t$ inherits the same inflation, whether the batch-level comparator of Steps 1–2 above changes this is open, and the structure of Theorem 4.1 is unaffected either way.

Remark A.7 (What Theorem 4.1 certifies). Let $r_t = \sum_{i=1}^{q}[f(\mathbf{x}^\star) - f(\hat{\mathbf{x}}_{t,i})]$ denote the regret of round t. The per-round bound underlying Theorem 4.1 is

$$r_t \le 2\beta_t^{1/2}\sum_{i=1}^{q}\sigma_t(\hat{\mathbf{x}}_{t,i} \mid \hat{\mathbf{x}}_{t,1:i-1}) + \epsilon_t.$$

As in the GP-BUCB analysis, the conditional variances are controlled through information gain. This yields the same regret-rate form $\sqrt{N\beta_T\gamma_N}$, with an additional term accounting for joint acquisition optimization error. The acquisition values admit a separate comparison. Since the greedy batch $\mathbf{X}_t^g$ is feasible for the joint problem, an ϵ_t-approximate joint maximizer satisfies

$$\mathcal{U}_t(\hat{\mathbf{X}}_t) \ge \mathcal{U}_t(\mathbf{X}_t^g) - \epsilon_t.$$

Thus, exact joint optimization achieves an acquisition value at least as large as that of greedy construction under the same posterior and acquisition parameters. This does not imply lower realized regret. The relationship between acquisition value and observed performance is instance-dependent. The theorem establishes a regret guarantee for jointly optimized batches with explicit solver error; it does not establish an improvement in the regret rate or constants over greedy GP-BUCB.

A.3 Proofs for Section 5

Two remarks on the statements of Section 5. The constant-mean hypothesis of Theorem 5.1 cannot be dropped, since a non-constant prior mean contributes a Hessian whose range need not lie V_x; and the theorem bounds the rank of $\mathbf{R}(\mathbf{x})$, not its magnitude, which Appendix B.2 measures. Proposition 5.3 compares fixed directions, so quasi-Newton methods, which rescale by an approximate inverse Hessian, fall outside it; the polar step obtains the matched geometry at $O(qd\min(q,d))$ per step with no line search or memory and remains valid under resampled stochastic gradients (Wilson et al., 2018; Balandat et al., 2020).

A.3.1 Proof of Proposition 5.3

For fixed $\mathbf{Z}$ with $c := \langle \mathbf{Z}, \mathcal{H}[\mathbf{Z}]\rangle > 0$, the scalar function $\eta \mapsto m(\eta\mathbf{Z}) = \eta\langle \mathbf{G}, \mathbf{Z}\rangle - \frac{c}{2}\eta^2$ is a downward parabola with vertex $\eta^\star = \langle \mathbf{G}, \mathbf{Z}\rangle/c$ and maximal value $m(\eta^\star\mathbf{Z}) = \langle \mathbf{G}, \mathbf{Z}\rangle^2/c - \frac{c}{2}\langle \mathbf{G}, \mathbf{Z}\rangle^2/c^2 = \langle \mathbf{G}, \mathbf{Z}\rangle^2/(2c)$; note this value is invariant to rescaling $\mathbf{Z}$. Substituting $\mathbf{Z} = \mathbf{G}$: $\langle \mathbf{G}, \mathbf{G}\rangle^2/(2\langle \mathbf{G}, \mathcal{H}[\mathbf{G}]\rangle) = \|\mathbf{G}\|_F^4/(2S(\mathbf{G})\|\mathbf{G}\|_F^2) = \|\mathbf{G}\|_F^2/(2S(\mathbf{G}))$. Substituting $\mathbf{Z} = \mathbf{Q} = \mathrm{Polar}(\mathbf{G})$: $\langle \mathbf{G}, \mathbf{Q}\rangle = \|\mathbf{G}\|_*$ by Lemma A.1(i) and $\langle \mathbf{Q}, \mathcal{H}[\mathbf{Q}]\rangle = S(\mathbf{Q})\|\mathbf{Q}\|_F^2 = S(\mathbf{Q})\,\rho$ by Lemma A.1(iii), giving $\|\mathbf{G}\|_*^2/(2S(\mathbf{Q})\rho)$. Dividing,

$$\frac{\|\mathbf{G}\|_*^2/(2S(\mathbf{Q})\rho)}{\|\mathbf{G}\|_F^2/(2S(\mathbf{G}))} = \frac{\|\mathbf{G}\|_*^2}{\|\mathbf{G}\|_F^2}\cdot\frac{S(\mathbf{G})}{S(\mathbf{Q})\,\rho} = \frac{r_{\mathrm{eff}}(\mathbf{G})}{\rho}\cdot\frac{S(\mathbf{G})}{S(\mathbf{Q})}. \qquad \square$$

A.3.2 Proof of Proposition 5.4

Identify matrices with vectors under any fixed Frobenius-isometric vectorization, so that $\mathcal{H}_0$ and $\mathcal{P}$ are represented by symmetric matrices $\mathbf{H}_0$ and $\mathbf{P}$ with $\|\mathbf{P}\|_2 \le \tau$. Write $\mathbf{u} = \mathrm{vec}(\bar{\mathbf{G}})$, $\mathbf{v} = \mathrm{vec}(\bar{\mathbf{Q}})$ and $\mathbf{K} = \bar{r}\,\mathbf{u}\mathbf{u}^\mathsf{T} - \mathbf{v}\mathbf{v}^\mathsf{T}$. Since $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ and $\mathbf{u}^\mathsf{T}\mathbf{v} = \sqrt{\bar{r}}$, the two possibly nonzero eigenvalues of $\mathbf{K}$ have sum $\bar{r} - 1$ and product $-\bar{r}(1-\bar{r})$. For $0 < \bar{r} < 1$ they have opposite signs, and consequently

$$\|\mathbf{K}\|_* = \sqrt{(\bar{r}-1)^2 + 4\bar{r}(1-\bar{r})} = \sqrt{(1-\bar{r})(1+3\bar{r})}.$$

For $\bar{r} = 1$, $\mathbf{u} = \mathbf{v}$ and $\mathbf{K} = 0$, giving the same formula. Since $S(\mathbf{G}) = \langle \bar{\mathbf{G}}, \mathcal{H}[\bar{\mathbf{G}}]\rangle$ and $S(\mathbf{Q}) = \langle \bar{\mathbf{Q}}, \mathcal{H}[\bar{\mathbf{Q}}]\rangle$, we have $\Gamma - \Gamma_0 = \mathrm{trace}(\mathbf{P}\mathbf{K})$, and $|\,\mathrm{trace}(\mathbf{P}\mathbf{K})| \le \|\mathbf{P}\|_2\,\|\mathbf{K}\|_* \le \tau\|\mathbf{K}\|_*$ by the duality of the spectral and nuclear norms, which is (3). The choices $\mathbf{P} = \pm\tau\,\mathrm{sign}(\mathbf{K})$, with the matrix sign equal to zero on the null space of $\mathbf{K}$, attain both endpoints; they are symmetric, of rank at most two, and of spectral norm τ when $\mathbf{K} \neq 0$.

Whenever both directional curvatures are positive, Proposition 5.3 gives $\mathrm{gain}(\mathbf{Q})/\,\mathrm{gain}(\mathbf{G}) = \bar{r}\,S(\mathbf{G})/S(\mathbf{Q})$, which exceeds one exactly when $\Gamma > 0$. If $\mathcal{H}_0 \succeq h_0\mathcal{I}$ with $h_0 > \tau$, every admissible $\mathcal{P}$ leaves $\mathcal{H} \succeq (h_0 - \tau)\mathcal{I} \succ 0$, so both directional curvatures are positive and the equivalence above applies. If (4) holds, then $\Gamma \ge \Gamma_0 - \tau\|\mathbf{K}\|_* > 0$ for every admissible $\mathcal{P}$. If it fails, choose the admissible operator $\mathcal{P}$ whose matrix representation is $\mathbf{P} = -\tau\,\mathrm{sign}(\mathbf{K})$; it gives $\Gamma = \Gamma_0 - \tau\|\mathbf{K}\|_* \le 0$, so the strict advantage fails for that perturbation. This proves that (4) is necessary and sufficient. The coefficient in (3) is therefore sharp for the stated operator-norm uncertainty class; this does not assert attainability within a specified family of GP posteriors. $\square$

Then we use the acquisition decomposition of Corollary 5.2, applied to the negative Hessian, so that block (i,j) of $\mathcal{H}$ is $c_{ij}\mathbf{\Lambda}^{-1} + \mathbf{R}_{ij}$ with $\mathbf{C} = (c_{ij})$ symmetric, and let $b_{ij} = \|\mathbf{R}_{ij}\|_2$ and $\mathbf{B}_R = (b_{ij})_{i,j\le q}$. For an increment with rows $\mathbf{d}_j^\mathsf{T}$, each output row of $\mathcal{P}[\mathbf{D}]$ has norm at most $\sum_j b_{ij}\|\mathbf{d}_j\|$. Hence $\|\mathcal{P}\|_{F\to F} \le \|\mathbf{B}_R\|_2$ is a usable residual bound. If $\mathbf{C} \succ 0$, the structured part $\mathcal{H}_0[\mathbf{D}] = \mathbf{C}\mathbf{D}\mathbf{\Lambda}^{-1}$ satisfies $\mathcal{H}_0 \succeq \lambda_{\min}(\mathbf{C})\lambda_{\min}(\mathbf{\Lambda}^{-1})\,\mathcal{I}$. Neither inequality implies a favorable margin without additional information about the acquisition. When the residual bound is large, the direct margin Γ, using the full Hessian in just the two directions, may still be informative.

An exact example with candidate coupling. Structured quadratic models are also used to study Muon in Wang et al. (2026), and the following construction illustrates the shared-metric structure relevant here. Let $q = d \geq 3$, choose orthogonal $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{q\times q}$ and positive $s_1, \ldots, s_q$, and set $\mathbf{A} = \mathbf{U}\mathrm{diag}(s_i)\mathbf{U}^\mathsf{T}$, $\mathbf{G} = \mathbf{U}\mathrm{diag}(s_i)\mathbf{V}^\mathsf{T}$ and

$$\alpha(\mathbf{D}) = \langle \mathbf{G},\, \mathbf{D}\rangle - \tfrac{1}{2}\,\mathrm{trace}(\mathbf{D}^\mathsf{T}\mathbf{A}\mathbf{D}).$$

At $\mathbf{D} = 0$, $\mathcal{H}[\mathbf{D}] = \mathbf{A}\mathbf{D}$ has blocks $A_{ij}\mathbf{I}_d$; choosing $\mathbf{U}$ to make $\mathbf{A}$ dense gives cross-candidate curvature. This is the shared-metric form with $\mathbf{\Lambda} = \mathbf{I}_d$ and zero residual. The polar direction is $\mathbf{Q} = \mathbf{U}\mathbf{V}^\mathsf{T}$, which equals the quadratic maximizer $\mathbf{A}^{-1}\mathbf{G}$. Its optimal step size is one, while the optimal gradient step size is $\sum_i s_i^2 / \sum_i s_i^3$. Therefore,

$$\frac{\mathrm{gain}(\mathbf{Q})}{\mathrm{gain}(\mathbf{G})} = \frac{(\sum_i s_i)(\sum_i s_i^3)}{(\sum_i s_i^2)^2},$$

which is at least one by the Cauchy–Schwarz inequality, with equality if and only if all s_i are equal. For $s_1 = \sqrt{q-1}$ and $s_2 = \cdots = s_q = 1$, this equals $\big(\sqrt{q-1} + 2 + 1/\sqrt{q-1}\big)/4$, which exceeds one and grows as $\sqrt{q}/4$. Scaling the linear term by a positive small constant scales both optimal displacements and makes them feasible inside a box around zero without changing the gain ratio. The construction demonstrates the mechanism within a compatible quadratic structure; it is not a claim that a particular GP acquisition realizes this quadratic. For comparison, $\mathcal{H} = c\mathcal{I}$ gives gain ratio $\bar{r} \leq 1$ for the same gradient, showing why alignment information is essential.

Suppose the negative Hessian is M_H-Lipschitz in the $F \to F$ norm on the two line segments considered below; this is an additional local regularity assumption. For $\mathbf{Z} \in \{\mathbf{G}, \mathbf{Q}\}$ with positive directional curvature, put $\eta_Z = \langle \mathbf{G},\, \mathbf{Z}\rangle / \langle \mathbf{Z},\, \mathcal{H}[\mathbf{Z}]\rangle$ and $\mathbf{D}_Z = \eta_Z \mathbf{Z}$. If both segments $[\mathbf{X}, \mathbf{X} + \mathbf{D}_Z]$ lie in the box, Taylor's theorem gives

$$\begin{aligned}\alpha(\mathbf{X} + \mathbf{D}_Q) - \alpha(\mathbf{X} + \mathbf{D}_G) \;\geq\; & \frac{\|\mathbf{G}\|_F^2\,\Gamma}{2S(\mathbf{G})S(\mathbf{Q})} \\ & - \frac{M_H}{6}\big(\|\mathbf{D}_Q\|_F^3 + \|\mathbf{D}_G\|_F^3\big).\end{aligned} \tag{6}$$

Indeed, each Taylor remainder is bounded in absolute value by $M_H\|\mathbf{D}_Z\|_F^3/6$, and subtracting the two quadratic gains gives the first term. Thus a positive right-hand side certifies an actual acquisition improvement. The sharp claim in Proposition 5.4 concerns the quadratic margin; (6) is a sufficient finite-step condition. The optimally scaled rays isolate the update directions for analysis; they are not the prescribed steps of Algorithm 1, and the statement does not cover clipped steps or prove a multi-round regret separation.

A.3.3 Proof of Theorem 5.1

All quantities are evaluated at a fixed $\mathbf{x}$ with $\sigma(\mathbf{x}) > 0$. Write $\boldsymbol{\omega}_i = \mathbf{x} - \mathbf{x}_i$, $\mathbf{w}_i = \mathbf{\Lambda}^{-1}\boldsymbol{\omega}_i \in V_x$, $r_i = \|\boldsymbol{\omega}_i\|_{\mathbf{\Lambda}}$, $k_i = k(\mathbf{x}, \mathbf{x}_i)$, $\mathbf{J} = [\nabla k_1, \ldots, \nabla k_n] \in \mathbb{R}^{d\times n}$, $\mathbf{A} = (\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}$ (well defined since $\sigma_n^2 > 0$), $\hat{\boldsymbol{\alpha}} = \mathbf{A}(\mathbf{y} - m\mathbf{1})$, $\mathbf{b} = \mathbf{A}\mathbf{k}_n(\mathbf{x})$.

Step 0 (kernel sections). Let $g(\mathbf{x}) = \kappa(r)$ with $r = \|\boldsymbol{\omega}\|_{\mathbf{\Lambda}} = (\boldsymbol{\omega}^\mathsf{T}\mathbf{\Lambda}^{-1}\boldsymbol{\omega})^{1/2}$, $\boldsymbol{\omega} = \mathbf{x} - \mathbf{x}_i$. For $\boldsymbol{\omega} \neq 0$,

$$\nabla r = \frac{\mathbf{\Lambda}^{-1}\boldsymbol{\omega}}{r} = \frac{\mathbf{w}}{r}, \qquad \nabla^2 r = \frac{\mathbf{\Lambda}^{-1}}{r} - \frac{\mathbf{w}\mathbf{w}^\mathsf{T}}{r^3} \quad \text{(differentiate } \nabla r \text{ by the quotient rule)},$$

so by the chain rule,

$$\nabla k_i = \kappa'(r_i)\,\frac{\mathbf{w}_i}{r_i},\; \nabla^2 k_i = \kappa''(r_i)\frac{\mathbf{w}_i\mathbf{w}_i^\mathsf{T}}{r_i^2} + \kappa'(r_i)\Big(\frac{\mathbf{\Lambda}^{-1}}{r_i} - \frac{\mathbf{w}_i\mathbf{w}_i^\mathsf{T}}{r_i^3}\Big) = \; a_i(\mathbf{x})\,\mathbf{\Lambda}^{-1} + h_i(\mathbf{x})\,\mathbf{w}_i\mathbf{w}_i^\mathsf{T}, \tag{7}$$

with

$$a_i(\mathbf{x}) = \frac{\kappa'(r_i)}{r_i}, \qquad h_i(\mathbf{x}) = \frac{1}{r_i^2}\Big(\kappa''(r_i) - \frac{\kappa'(r_i)}{r_i}\Big).$$

For the RBF kernel $\kappa(r) = a^2 e^{-r^2/2}$: $\kappa'(r) = -a^2 r e^{-r^2/2}$ and $\kappa''(r) = a^2(r^2-1)e^{-r^2/2}$, so $a_i = -k_i$ and $h_i = \big(a^2(r_i^2-1)e^{-r_i^2/2} + a^2 e^{-r_i^2/2}\big)/r_i^2 = k_i$. For the Matérn-5/2 kernel $\kappa(r) = a^2(1+\sqrt{5}r + \frac{5}{3}r^2)e^{-\sqrt{5}r}$: differentiating,

$$\kappa'(r) = a^2 e^{-\sqrt{5}r}\Big(\sqrt{5} + \tfrac{10}{3}r - \sqrt{5}\big(1+\sqrt{5}r+\tfrac{5}{3}r^2\big)\Big) = -\tfrac{5}{3}a^2\, r\,(1+\sqrt{5}r)\, e^{-\sqrt{5}r},$$

so $a_i = -\frac{5}{3}a^2(1+\sqrt{5}r_i)e^{-\sqrt{5}r_i}$, and differentiating once more, $\kappa''(r) = -\frac{5}{3}a^2(1+\sqrt{5}r - 5r^2)e^{-\sqrt{5}r}$, giving $\kappa'' - \kappa'/r = \frac{25}{3}a^2 r^2 e^{-\sqrt{5}r}$ and $h_i = \frac{25}{3}a^2 e^{-\sqrt{5}r_i}$.

The coincident case $\boldsymbol{\omega}_i = 0$. Equation (7) was derived for $\boldsymbol{\omega}_i \neq 0$; at $\mathbf{x} = \mathbf{x}_i$ it holds with no dyad, for every C^2 radial section. Indeed, in the scaled coordinates $\tilde{\boldsymbol{\omega}} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\omega}$ the section is $\tilde{g}(\tilde{\boldsymbol{\omega}}) = \kappa(\|\tilde{\boldsymbol{\omega}}\|)$, which is invariant under every orthogonal map $\tilde{\boldsymbol{\omega}} \mapsto O\tilde{\boldsymbol{\omega}}$; since $\tilde{g}$ is C^2, $\nabla\tilde{g}(0)$ is an $O(d)$-invariant vector and $\nabla^2\tilde{g}(0)$ an $O(d)$-invariant symmetric form, so $\nabla\tilde{g}(0) = 0$ and $\nabla^2\tilde{g}(0) = c_i\mathbf{I}$ for a scalar c_i. Transforming back, $\nabla k_i(\mathbf{x}_i) = 0$ and $\nabla^2 k_i(\mathbf{x}_i) = c_i\boldsymbol{\Lambda}^{-1}$, which is (7) with $a_i(\mathbf{x}_i) = c_i$ and $h_i\mathbf{w}_i\mathbf{w}_i^{\mathsf{T}} = 0$ (both $\nabla k_i = 0$ and the vanishing dyad are trivially in $V_x \otimes V_x$). The explicit limits confirm the scalar: RBF $c_i = -a^2$, Matérn-5/2 $c_i = -\frac{5}{3}a^2$. Hence, at every $\mathbf{x}$, every gradient ∇k_i is a multiple of $\mathbf{w}_i \in V_x$, and every dyad lies in $V_x \otimes V_x$.

Step 1 (posterior mean). With constant prior mean m, $\mu(\mathbf{x}) = m + \mathbf{k}_n(\mathbf{x})^{\mathsf{T}}\hat{\boldsymbol{\alpha}}$ with $\hat{\boldsymbol{\alpha}} = \mathbf{A}(\mathbf{y} - m\mathbf{1})$ constant in $\mathbf{x}$ (this is where the constant-mean hypothesis enters: a non-constant prior mean would add its own gradient and Hessian, which need not lie in V_x), so $\nabla\mu = \sum_i \hat{\alpha}_i \nabla k_i \in V_x$ and, by (7),

$$\nabla^2\mu = \sum_i \hat{\alpha}_i \nabla^2 k_i = \Big(\underbrace{\textstyle\sum_i \hat{\alpha}_i a_i(\mathbf{x})}_{=:\, c_\mu(\mathbf{x})}\Big)\boldsymbol{\Lambda}^{-1} + \sum_i \hat{\alpha}_i h_i(\mathbf{x})\, \mathbf{w}_i\mathbf{w}_i^{\mathsf{T}}.$$

Step 2 (posterior variance and standard deviation). $\sigma^2(\mathbf{x}) = \kappa(0) - \mathbf{k}_n(\mathbf{x})^{\mathsf{T}}\mathbf{A}\mathbf{k}_n(\mathbf{x})$ with $\kappa(0)$ constant. Since $\nabla\big[\mathbf{k}_n^{\mathsf{T}}\mathbf{A}\mathbf{k}_n\big] = 2\mathbf{J}\,\mathbf{A}\mathbf{k}_n = 2\mathbf{J}\mathbf{b}$ (the factor 2 from symmetry of $\mathbf{A}$), $\nabla\sigma^2 = -2\mathbf{J}\mathbf{b} \in V_x$ (each column of $\mathbf{J}$ lies in V_x). Differentiating $\mathbf{x} \mapsto \mathbf{J}(\mathbf{x})\mathbf{b}(\mathbf{x})$ by the product rule, with $\nabla\mathbf{b} = \nabla(\mathbf{A}\mathbf{k}_n) = \mathbf{A}\mathbf{J}^{\mathsf{T}}$,

$$\nabla^2\sigma^2 = -2\Big[\sum_j b_j \nabla^2 k_j + \mathbf{J}\,\mathbf{A}\,\mathbf{J}^{\mathsf{T}}\Big] = \Big(\underbrace{-2\textstyle\sum_j b_j a_j(\mathbf{x})}_{=:\, c_{\sigma^2}(\mathbf{x})}\Big)\boldsymbol{\Lambda}^{-1} - 2\sum_j b_j h_j \mathbf{w}_j\mathbf{w}_j^{\mathsf{T}} - 2\mathbf{J}\mathbf{A}\mathbf{J}^{\mathsf{T}},$$

where $\mathbf{J}\mathbf{A}\mathbf{J}^{\mathsf{T}}$ has range in V_x (its columns are combinations of columns of $\mathbf{J}$). From $\sigma = (\sigma^2)^{1/2}$, $\nabla\sigma = \nabla\sigma^2/(2\sigma) \in V_x$ and

$$\nabla^2\sigma = \frac{\nabla^2\sigma^2}{2\sigma} - \frac{(\nabla\sigma^2)(\nabla\sigma^2)^{\mathsf{T}}}{4\sigma^3} = \Big(\underbrace{\frac{c_{\sigma^2}(\mathbf{x})}{2\sigma(\mathbf{x})}}_{=:\, c_\sigma(\mathbf{x})}\Big)\boldsymbol{\Lambda}^{-1} + (\text{terms with range in } V_x).$$

Step 3 (EI and its derivatives). With $u = f^\star - \mu - \xi$, $z = u/\sigma$, $\mathrm{EI} = u\,\Phi(z) + \sigma\,\phi(z)$. Using $\Phi' = \phi$ and $\phi'(z) = -z\phi(z)$:

$$\frac{\partial \mathrm{EI}}{\partial\mu} = -\Phi(z) + u\phi(z)\frac{\partial z}{\partial\mu} + \sigma\phi'(z)\frac{\partial z}{\partial\mu} = -\Phi(z) + \big(u - \sigma z\big)\phi(z)\cdot\Big(-\frac{1}{\sigma}\Big) = -\Phi(z),$$

since $u - \sigma z = 0$; and

$$\frac{\partial \mathrm{EI}}{\partial\sigma} = u\phi(z)\frac{\partial z}{\partial\sigma} + \phi(z) + \sigma\phi'(z)\frac{\partial z}{\partial\sigma} = \phi(z) + \big(u - \sigma z\big)\phi(z)\cdot\Big(-\frac{z}{\sigma}\Big) = \phi(z).$$

Hence $\nabla\mathrm{EI} = -\Phi(z)\nabla\mu + \phi(z)\nabla\sigma$, and differentiating once more, using $\Phi' = \phi$, $\phi'(z) = -z\phi(z)$,

$$\nabla^2\mathrm{EI} = -\Phi(z)\,\nabla^2\mu + \phi(z)\,\nabla^2\sigma \;-\; \phi(z)\,\big[\nabla z\,\nabla\mu^{\mathsf{T}} + z\,\nabla z\,\nabla\sigma^{\mathsf{T}}\big].$$

Since $\nabla z = (-\nabla\mu - z\nabla\sigma)/\sigma$, i.e. $\nabla\mu = -\sigma\nabla z - z\nabla\sigma$, the bracket equals $\nabla z(-\sigma\nabla z - z\nabla\sigma)^{\mathsf{T}} + z\nabla z\nabla\sigma^{\mathsf{T}} = -\sigma\,\nabla z\nabla z^{\mathsf{T}}$, so

$$\nabla^2\mathrm{EI} = -\Phi(z)\,\nabla^2\mu + \phi(z)\,\nabla^2\sigma + \sigma\,\phi(z)\,\nabla z\,\nabla z^{\mathsf{T}},$$

which is manifestly symmetric, and $\nabla z \in V_x$ because $\nabla\mu, \nabla\sigma \in V_x$. The outer-product term has both factors in V_x.

Assembly. Collecting the $\mathbf{\Lambda}^{-1}$ components of Steps 1–2 through the coefficients $-\Phi(z)$ and $\phi(z)$,

$$\nabla^2 \mathrm{EI}(\mathbf{x}) = \underbrace{\big[-\Phi(z)c_\mu(\mathbf{x}) + \phi(z)c_\sigma(\mathbf{x})\big]}_{C(\mathbf{x})}\,\mathbf{\Lambda}^{-1} \;+\; \mathbf{R}(\mathbf{x}),$$

where $\mathbf{R}$ collects all remaining terms; each has range in V_x, and $\mathbf{R}$ is symmetric (difference of two symmetric matrices), so its row space equals its column space and lies in V_x, whence $\mathrm{rank}\mathbf{R} \le \dim V_x \le \min(n, d)$.

RBF specialization. With $a_i = -k_i$: $c_\mu = \sum_i \hat{\alpha}_i(-k_i) = -\mathbf{k}_n(\mathbf{x})^\mathsf{T}\hat{\boldsymbol{\alpha}} = -(\mu(\mathbf{x}) - m)$ and $c_{\sigma^2} = -2\sum_j b_j(-k_j) = 2\mathbf{k}_n^\mathsf{T}\mathbf{b} = 2\mathbf{k}_n^\mathsf{T}\mathbf{A}\mathbf{k}_n = 2\big(a^2 - \sigma^2(\mathbf{x})\big)$, so $c_\sigma = (a^2 - \sigma^2)/\sigma$ and $C(\mathbf{x}) = \Phi(z)(\mu(\mathbf{x}) - m) + \phi(z)(a^2 - \sigma^2)/\sigma$ (the computation does not use the form of $\mathbf{\Lambda}$). Under the maximization convention $\partial\mathrm{EI}/\partial\mu = +\Phi(z)$ and $\partial\mathrm{EI}/\partial\sigma = \phi(z)$ by the same computation, so $C(\mathbf{x}) = +\Phi(z)c_\mu + \phi(z)c_\sigma$ and the closed form reads $-\Phi(z)(\mu(\mathbf{x}) - m) + \phi(z)(a^2 - \sigma^2)/\sigma$; $\mathbf{R}(\mathbf{x})$ itself changes, but its range statement, and hence the rank bound, do not. □

A.3.4 Batch structure

Corollary A.8. *For the repulsive-EI batch acquisition* $\alpha(\mathbf{X}) = \sum_i \mathrm{EI}(\mathbf{x}_i) - \lambda\sum_{i<j} r(\mathbf{x}_i - \mathbf{x}_j)$ *with* $r(\mathbf{u}) = e^{-\|\mathbf{u}\|^2/(2\varrho^2)}$, $\varrho > 0$,

$$\nabla^2\alpha(\mathbf{X}) = \bigoplus_{i=1}^{q}\big[C(\mathbf{x}_i)\mathbf{\Lambda}^{-1} + \mathbf{R}(\mathbf{x}_i)\big] \;-\; \lambda\mathbf{P}(\mathbf{X}), \qquad \|\mathbf{P}(\mathbf{X})\|_2 \le \frac{2(q-1)}{\varrho^2},$$

where $\mathbf{P}$ *is the Hessian of the pairwise repulsion.*

Proof. The separable sum contributes the block diagonal, blockwise by Theorem 5.1. For the repulsion: with $s = \|\mathbf{u}\|^2/\varrho^2$, $\nabla^2 r(\mathbf{u}) = \big(\mathbf{u}\mathbf{u}^\mathsf{T}/\varrho^4 - \mathbf{I}/\varrho^2\big)e^{-s/2}$, whose eigenvalues are $(s-1)e^{-s/2}/\varrho^2$ (along $\mathbf{u}$) and $-e^{-s/2}/\varrho^2$ (multiplicity $d-1$). On $s \ge 0$, $\max|s-1|e^{-s/2}$ is 1 (at $s = 0$; the interior critical point $s = 3$ gives $2e^{-3/2} \approx 0.45$), so $\|\nabla^2 r\|_2 \le 1/\varrho^2$ everywhere. The matrix $\mathbf{P}$ has diagonal blocks $\sum_{j\neq i}\nabla^2 r(\mathbf{x}_i - \mathbf{x}_j)$ and off-diagonal blocks $-\nabla^2 r(\mathbf{x}_i - \mathbf{x}_j)$. For any symmetric block matrix $\mathbf{M} = (\mathbf{M}_{ij})$, $\|\mathbf{M}\|_2 \le \max_i\sum_j\|\mathbf{M}_{ij}\|_2$ (block Gershgorin: for a unit eigenvector $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_q)$ with $\mathbf{M}\mathbf{v} = \theta\mathbf{v}$, pick i maximizing $\|\mathbf{v}_i\|$ and bound $|\theta|\,\|\mathbf{v}_i\| = \|\sum_j \mathbf{M}_{ij}\mathbf{v}_j\| \le \sum_j\|\mathbf{M}_{ij}\|_2\|\mathbf{v}_i\|$). Here each block row sums to at most $(q-1)/\varrho^2 + (q-1)/\varrho^2$. □

A.3.5 Proof of Corollary 5.2

Fix the round and write the active points as $\mathcal{P} = \{\mathbf{x}_1^{\mathrm{data}}, \ldots, \mathbf{x}_n^{\mathrm{data}}\} \cup \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ and $W = \mathrm{span}\{\mathbf{\Lambda}^{-1}(\mathbf{v} - \mathbf{v}') : \mathbf{v}, \mathbf{v}' \in \mathcal{P}\}$; since $\mathbf{v} - \mathbf{v}' = (\mathbf{v} - \mathbf{x}_1) - (\mathbf{v}' - \mathbf{x}_1)$, W is spanned by the $n + q - 1$ generators $\mathbf{\Lambda}^{-1}(\mathbf{v} - \mathbf{x}_1)$, $\mathbf{v} \in \mathcal{P} \setminus \{\mathbf{x}_1\}$, so $\dim W \le \min(n + q - 1, d)$. Every quantity entering $\mathcal{U}_t(\mathbf{X}) = \sum_i[\mu_t(\mathbf{x}_i) + \beta_t^{1/2}\sqrt{v_i(\mathbf{X})}]$ is, up to the additive constant m in μ_t (which has no derivatives; here the constant-prior-mean hypothesis is used), a C^2 function of the kernel evaluations $\{k(\mathbf{v}, \mathbf{v}') : \mathbf{v}, \mathbf{v}' \in \mathcal{P}\}$ (the v_i are Schur complements of kernel matrices over $\mathcal{P}$ shifted by $\sigma_n^2\mathbf{I}$, strictly positive by Section 4, so the square roots and matrix inverses are smooth, and the kernel evaluations are C^2 in the rows of $\mathbf{X}$ by hypothesis). Consider any second partial derivative of $\mathcal{U}_t$ with respect to entries of rows $\mathbf{x}_m$ and $\mathbf{x}_{m'}$ (possibly $m = m'$). By the chain rule it is a finite sum of terms, each a product of (smooth scalar functions of kernel evaluations) with either (i) one second derivative $\partial^2 k(\mathbf{v}, \mathbf{w})/\partial\mathbf{x}_m\partial\mathbf{x}_{m'}$ of a single kernel evaluation, or (ii) two first derivatives $\partial k/\partial\mathbf{x}_m$ and $\partial k/\partial\mathbf{x}_{m'}$. By (7) (including the coincident case treated in Step 0 of the previous proof) and its mixed-argument analogue ($\partial^2 k(\mathbf{x}_m, \mathbf{v})/\partial\mathbf{x}_m\partial\mathbf{x}_m$ has the $a\mathbf{\Lambda}^{-1} + h\,\mathbf{w}\mathbf{w}^\mathsf{T}$ form with $\mathbf{w} = \mathbf{\Lambda}^{-1}(\mathbf{x}_m - \mathbf{v})$; $\partial^2 k(\mathbf{x}_m, \mathbf{x}_{m'})/\partial\mathbf{x}_m\partial\mathbf{x}_{m'}$ equals its negative), every type-(i) term is a scalar multiple of $\mathbf{\Lambda}^{-1}$ plus a dyad with both factors in W; every type-(ii) term is an outer product of two gradients, each of which is a multiple of some $\mathbf{\Lambda}^{-1}(\mathbf{v} - \mathbf{v}') \in W$. Summing preserves the form "scalar·$\mathbf{\Lambda}^{-1}$ + (matrix with row and column spaces in W)", which is the claim; the rank bound is $\dim W$. □

A.4 Proofs for Section 6

A.4.1 Proof of Theorem 6.1

Fix k and abbreviate $\mathbf{G} = \mathbf{G}_k = \nabla\alpha(\mathbf{X}_k)$, $\mathbf{Q} = \mathbf{Q}_k = \mathrm{Polar}(\mathbf{G})$, $\mathbf{Y} = \mathbf{X}_k + \eta\mathbf{Q}$, $\mathbf{X}_{k+1} = \Pi_\Omega(\mathbf{Y})$, $\mathbf{D} = \mathbf{X}_{k+1} - \mathbf{X}_k$. If $\mathbf{G} = 0$ then $\mathbf{Q} = 0$, the step is null, and the claimed per-step inequality holds trivially; assume $\mathbf{G} \neq 0$.

Step 1: the clipped-move structure. The box projection acts coordinatewise: X_{k+1}^{ij} is Y^{ij} clipped to $[l_j, u_j]$. Since $\mathbf{X}_k \in \Omega^q$ (true at $k = 0$ by hypothesis and preserved by projection), exactly three cases arise for each coordinate (i, j):

Case	Condition	Consequence
no clip	$Y^{ij} \in [l_j, u_j]$	$D^{ij} = \eta Q^{ij}$
upper clip	$Y^{ij} > u_j$	$\eta Q^{ij} > u_j - X_k^{ij} \geq 0$, so $Q^{ij} > 0$ and $D^{ij} = u_j - X_k^{ij} \in [0, \eta Q^{ij})$
lower clip	$Y^{ij} < l_j$	$\eta Q^{ij} < l_j - X_k^{ij} \leq 0$, so $Q^{ij} < 0$ and $D^{ij} = l_j - X_k^{ij} \in (\eta Q^{ij}, 0]$

Define $\mathbf{S} = \eta\mathbf{Q} - \mathbf{D}$. By the table, $\mathbf{S}$ is supported on the clipped set $\mathcal{A}_k$, and on it S^{ij} has the same sign as Q^{ij} with $|S^{ij}| \leq \eta|Q^{ij}|$; moreover $|D^{ij}| \leq \eta|Q^{ij}|$ for *every* coordinate.

Step 2: the inner product. Using Lemma A.1(i) and then bounding $\langle \mathbf{G}, \mathbf{S}\rangle$ coordinatewise over its support,

$$\langle \mathbf{G}, \mathbf{D}\rangle = \langle \mathbf{G}, \eta\mathbf{Q}\rangle - \langle \mathbf{G}, \mathbf{S}\rangle = \eta\|\mathbf{G}\|_* - \sum_{(i,j)\in\mathcal{A}_k} G^{ij}S^{ij} \;\geq\; \eta\|\mathbf{G}\|_* - \sum_{(i,j)\in\mathcal{A}_k} |G^{ij}|\,|S^{ij}|$$
$$\geq\; \eta\|\mathbf{G}\|_* - \eta \sum_{(i,j)\in\mathcal{A}_k} |G^{ij}||Q^{ij}| \;=\; \eta\big(\|\mathbf{G}\|_* - E_k\big).$$

Step 3: the step length. From $|D^{ij}| \leq \eta|Q^{ij}|$ coordinatewise, $\|\mathbf{D}\|_F \leq \eta\|\mathbf{Q}\|_F = \eta\sqrt{\mathrm{rank}\mathbf{G}} \leq \eta\sqrt{q}$ by Lemma A.1(iii).

Step 4: ascent and telescoping. $\mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{D} \in \Omega^q$, so Lemma A.2 (Frobenius case) gives

$$\alpha(\mathbf{X}_{k+1}) \;\geq\; \alpha(\mathbf{X}_k) + \eta\big(\|\mathbf{G}_k\|_* - E_k\big) - \frac{L_F q}{2}\eta^2.$$

Summing over $k = 0, \ldots, T_{\mathrm{in}} - 1$ and using $\sum_k[\alpha(\mathbf{X}_{k+1}) - \alpha(\mathbf{X}_k)] = \alpha(\mathbf{X}_{T_{\mathrm{in}}}) - \alpha(\mathbf{X}_0) \leq \Delta_0$,

$$\eta \sum_{k=0}^{T_{\mathrm{in}}-1} \big(\|\mathbf{G}_k\|_* - E_k\big) \;\leq\; \Delta_0 + \frac{L_F q}{2}\eta^2\, T_{\mathrm{in}};$$

dividing by $T_{\mathrm{in}}\eta > 0$ yields the display; since only the upper bound $\alpha(\mathbf{X}_{T_{\mathrm{in}}}) - \alpha(\mathbf{X}_0) \leq \Delta_0$ was used, any $\Delta \geq \Delta_0$ may stand in for Δ_0.

Step 5: step-size choice. With $E_k \equiv 0$ and $\Delta_0, L_F > 0$ the right side is $g(\eta) = \Delta_0/(T_{\mathrm{in}}\eta) + L_F q\eta/2$, minimized at $\eta^\star = \sqrt{2\Delta_0/(L_F q T_{\mathrm{in}})}$ where both terms equal $\sqrt{L_F q \Delta_0/(2T_{\mathrm{in}})}$, giving $g(\eta^\star) = \sqrt{2L_F q\Delta_0/T_{\mathrm{in}}}$. For the mixed-norm interior case, $\mathbf{D} = \eta\mathbf{Q}$ exactly, $\|\mathbf{D}\|_2 = \eta\|\mathbf{Q}\|_2 \leq \eta$, and Lemma A.2 (mixed case) gives $\alpha(\mathbf{X}_{k+1}) \geq \alpha(\mathbf{X}_k) + \eta\|\mathbf{G}_k\|_* - \frac{L_*}{2}\eta^2$; the same telescoping and the choice $\eta = \sqrt{2\Delta_0/(L_* T_{\mathrm{in}})}$ give $\sqrt{2L_*\Delta_0/T_{\mathrm{in}}}$. □

A.4.2 Smooth unconstraining

The box of Section 6 can be removed before optimizing by a smooth reparameterization, in the spirit of the inverse stereographic map that Doumont et al. (2026) use against boundary-seeking. The following proposition, referenced in Section 6, is stated and proved for completeness; the experiments use the projection of Algorithm 1.

Proposition A.9 (Smooth unconstraining)**.** *Let $s : \mathbb{R} \to (0, 1)$ be a C^2 bijection with $s' > 0$ and s', s'' bounded (e.g. the logistic function), and define $T(\mathbf{Z})^{ij} = l_j + (u_j - l_j)s(Z^{ij})$. If α satisfies Assumption 2 with $G_\infty = \sup_{\Omega^q}\|\nabla\alpha\|_F < \infty$, then $\tilde{\alpha} = \alpha \circ T$ satisfies Assumption 2 on all of $\mathbb{R}^{q\times d}$ with $\tilde{L}_F \leq c_1^2 L_F + c_2 G_\infty$, where $\bar{w} = \max_j(u_j - l_j)$, $c_1 = \bar{w}\sup|s'|$, $c_2 = \bar{w}\sup|s''|$; interior stationary points of α correspond to stationary points of $\tilde{\alpha}$.*

A.4.3 Proof of Proposition A.9

Write $\mathbf{A} = \nabla\alpha(T(\mathbf{Z}))$, $\mathbf{A}' = \nabla\alpha(T(\mathbf{Z}'))$. By the chain rule, coordinatewise, $\nabla\tilde{\alpha}(\mathbf{Z})^{ij} = A^{ij}\,T'(\mathbf{Z})^{ij}$ with $T'(\mathbf{Z})^{ij} = (u_j - l_j)\,s'(Z^{ij})$. The scalar function s gives, for every coordinate,

$$|T'(\mathbf{Z})^{ij}| \le c_1, \qquad |T'(\mathbf{Z})^{ij} - T'(\mathbf{Z}')^{ij}| = (u_j - l_j)\,|s'(Z^{ij}) - s'(Z'^{ij})| \le c_2\,|Z^{ij} - Z'^{ij}|,$$

the second using the mean value theorem on s' (whose derivative s'' is bounded), and likewise $|T(\mathbf{Z})^{ij} - T(\mathbf{Z}')^{ij}| \le c_1|Z^{ij} - Z'^{ij}|$, whence $\|T(\mathbf{Z}) - T(\mathbf{Z}')\|_F \le c_1\|\mathbf{Z} - \mathbf{Z}'\|_F$. Split

$$\nabla\tilde{\alpha}(\mathbf{Z}) - \nabla\tilde{\alpha}(\mathbf{Z}') = \underbrace{(\mathbf{A} - \mathbf{A}') \odot T'(\mathbf{Z})}_{\text{(I)}} + \underbrace{\mathbf{A}' \odot \big(T'(\mathbf{Z}) - T'(\mathbf{Z}')\big)}_{\text{(II)}}.$$

For (I): $\|(\mathbf{A} - \mathbf{A}') \odot T'(\mathbf{Z})\|_F \le \big(\max_{ij}|T'(\mathbf{Z})^{ij}|\big)\|\mathbf{A} - \mathbf{A}'\|_F \le c_1 \cdot L_F\|T(\mathbf{Z}) - T(\mathbf{Z}')\|_F \le c_1^2 L_F\|\mathbf{Z} - \mathbf{Z}'\|_F$. For (II): coordinatewise, $|(\mathrm{II})^{ij}| \le |A'^{ij}| \cdot c_2|Z^{ij} - Z'^{ij}|$, so

$$\begin{aligned}
\|(\mathrm{II})\|_F^2 &= \sum_{ij} |(\mathrm{II})^{ij}|^2 \\
&\le c_2^2 \sum_{ij} |A'^{ij}|^2\,|Z^{ij} - Z'^{ij}|^2 \\
&\le c_2^2 \Big(\max_{ij}|A'^{ij}|^2\Big) \sum_{ij} |Z^{ij} - Z'^{ij}|^2 \\
&\le c_2^2\,G_\infty^2\,\|\mathbf{Z} - \mathbf{Z}'\|_F^2,
\end{aligned}$$

using $\max_{ij}|A'^{ij}| \le \|\mathbf{A}'\|_F \le G_\infty$. Adding the two bounds gives $\tilde{L}_F \le c_1^2 L_F + c_2 G_\infty$. Finally, since $s' > 0$ everywhere, $\nabla\tilde{\alpha}(\mathbf{Z}) = 0$ iff $\nabla\alpha(T(\mathbf{Z})) = 0$; and since s is a bijection onto $(0,1)$, T maps $\mathbb{R}^{q\times d}$ onto the open box, so box-boundary points are approached only as $|Z^{ij}| \to \infty$. □

A.4.4 Value conversion

The following proposition converts the nuclear-norm stationarity of Theorem 6.1 into the value accuracy that Theorem 4.1 consumes, which utilizes the PŁ inequality (Karimi et al., 2016) of the form $\frac{1}{2}\|\nabla\alpha(\mathbf{X})\|_F^2 \ge \mu_{\mathrm{PL}}(\alpha_B^\star - \alpha(\mathbf{X}))$ on a basin B with maximal value $\alpha_B^\star$. This condition ensures that a small gradient implies a small value gap within that basin. It is used in Corollary A.11.

Proposition A.10 (Value conversion)**.** *Suppose the PŁ inequality $\frac{1}{2}\|\nabla\alpha(\mathbf{X})\|_F^2 \ge \mu_{\mathrm{PL}}\big(\alpha_B^\star - \alpha(\mathbf{X})\big)$ holds with $\mu_{\mathrm{PL}} > 0$ on a basin B with maximal value $\alpha_B^\star$. Then any $\mathbf{X} \in B$ with $\|\nabla\alpha(\mathbf{X})\|_* \le \epsilon$ satisfies*

$$\alpha_B^\star - \alpha(\mathbf{X}) \;\le\; \epsilon^2/(2\mu_{\mathrm{PL}}).$$

With R independent uniform restarts on Ω^q, a measurable basin of normalized volume p receives at least one start with probability $1 - (1-p)^R$.

A.4.5 Proof of Proposition A.10

For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_i \sigma_i(\mathbf{A})^2 \le \big(\sum_i \sigma_i(\mathbf{A})\big)^2 = \|\mathbf{A}\|_*^2$, so $\|\mathbf{A}\|_F \le \|\mathbf{A}\|_*$. Hence $\|\nabla\alpha(\mathbf{X})\|_* \le \epsilon$ implies $\|\nabla\alpha(\mathbf{X})\|_F \le \epsilon$, and the PŁ inequality gives $\alpha_B^\star - \alpha(\mathbf{X}) \le \|\nabla\alpha(\mathbf{X})\|_F^2/(2\mu_{\mathrm{PL}}) \le \epsilon^2/(2\mu_{\mathrm{PL}})$. For the restarts: each start is an independent uniform draw, missing B with probability $1 - p$; the probability all R miss is $(1-p)^R$. □

A.4.6 Regret bound for Muon-solved batch BO

Theorem 4.1 expresses the effect of approximate acquisition maximization through $\sum_t \epsilon_t$. We now bound this term in terms of Muon's inner iteration budget. This requires the local PŁ condition of Proposition A.10, together with assumptions ensuring that a restart enters and remains in a basin containing a global acquisition maximizer.

Let $\mathcal{H}_t$ denote the σ-algebra generated by the observed locations and values available at the start of round t, so that $\mathcal{U}_t$ is $\mathcal{H}_t$-measurable. Write $\mathcal{U}_t^\star = \sup_{\Omega^q}\mathcal{U}_t$.

Assumption 3 (Acquisition regularity and invariant PŁ basins)**.** *There are deterministic constants $L_{\mathcal{U}}, D_{\mathcal{U}}, \mu_{\mathrm{PL}} > 0$ and $p \in (0, 1]$, independent of the round, such that the following hold almost surely for every $t \le T$.*

(i) $\mathcal{U}_t$ satisfies Assumption 2 with constant $L_{\mathcal{U}}$, and

$$\sup_{\Omega^q} \mathcal{U}_t - \inf_{\Omega^q} \mathcal{U}_t \le D_{\mathcal{U}}.$$

(ii) There is an $\mathcal{H}_t$-measurable basin $B_t \subseteq \mathrm{int}(\Omega^q)$ containing a global maximizer of $\mathcal{U}_t$, with

$$\frac{\mathrm{vol}(B_t)}{\mathrm{vol}(\Omega^q)} \ge p.$$

For every $\mathbf{X} \in B_t$,

$$\frac{1}{2}\|\nabla \mathcal{U}_t(\mathbf{X})\|_F^2 \ge \mu_{\mathrm{PL}}\big(\mathcal{U}_t^\star - \mathcal{U}_t(\mathbf{X})\big).$$

(iii) For the step size in Corollary A.11, Muon iterates initialized in B_t remain in B_t without clipping throughout the T_{in} inner steps.

Part (i) supplies uniform bounds for the inner-optimization analysis. Part (ii) converts a small gradient into a small acquisition value gap and gives a lower bound on the probability that a uniform restart enters a suitable basin. Because the basin contains a global maximizer, this gap is measured against $\mathcal{U}_t^\star$. Part (iii) ensures that the PŁ inequality remains applicable along the selected trajectory and that no clipping correction is needed.

Corollary A.11 (Regret bound for Muon-solved joint batch BO)**.** *Suppose the statistical and feedback conditions of Theorem 4.1 and Assumption 3 hold. At each round, run Algorithm 1 with R conditionally i.i.d. uniform restarts on Ω^q given $\mathcal{H}_t$, T_{in} steps per restart, and*

$$\eta = \sqrt{\frac{2D_{\mathcal{U}}}{L_{\mathcal{U}} q T_{\mathrm{in}}}}.$$

Generate the restarts using fresh randomness independent of the history, f, and the observation noise. Return the iterate with the largest exactly evaluated acquisition value among all restarts, including their initial iterates. Then, with probability at least $1 - \delta - T(1-p)^R$,

$$R_N \le \sqrt{C_1 N \beta_T \gamma_N} + \frac{N L_{\mathcal{U}} D_{\mathcal{U}}}{\mu_{\mathrm{PL}} T_{\mathrm{in}}}, \qquad N = qT.$$

The first term retains the usual GP-UCB/GP-BUCB information-gain form, with β_T incorporating the batch confidence inflation. The second term makes the cost of finite acquisition optimization explicit. Choosing

$$T_{\mathrm{in}} \ge \frac{L_{\mathcal{U}} D_{\mathcal{U}} \sqrt{N}}{\mu_{\mathrm{PL}}}$$

makes the solver term at most $\sqrt{N}$. It is therefore asymptotically smaller than the first term whenever $\beta_T \gamma_N \to \infty$. This result connects inner iterations to regret under additional assumptions; it does not establish a Muon-specific improvement in the outer regret rate.

The basin and regularity requirements need not hold for a general GP acquisition. In particular, p measures volume in dimension qd, so a small basin may require many restarts. For any $\zeta \in (0, 1)$, choosing $R \ge p^{-1} \log(T/\zeta)$ suffices to ensure $T(1-p)^R \le \zeta$. The constants $L_{\mathcal{U}}, D_{\mathcal{U}}, \mu_{\mathrm{PL}}$ must also hold uniformly across rounds, although acquisition geometry changes as data accumulate. Theorem 4.1 does not require these additional conditions.

Proof. Step 1: a successful restart in every round. Let $\mathcal{G}_t$ denote the event that at least one of the R restarts initializes in B_t. Conditional on $\mathcal{H}_t$, the basin is a fixed measurable set. Writing

$$v_t = \frac{\mathrm{vol}(B_t)}{\mathrm{vol}(\Omega^q)},$$

each uniform restart misses the basin with probability $1 - v_t$. Conditional independence of the restarts and $v_t \geq p$ give

$$\Pr(\mathcal{G}_t^c \mid \mathcal{H}_t) = (1 - v_t)^R \leq (1 - p)^R.$$

Taking expectations yields

$$\Pr(\mathcal{G}_t^c) = \mathbb{E}\big[\Pr(\mathcal{G}_t^c \mid \mathcal{H}_t)\big] \leq (1 - p)^R.$$

Consequently, for $\mathcal{G} = \bigcap_{t=1}^{T} \mathcal{G}_t$, the union bound gives

$$\Pr(\mathcal{G}^c) \leq \sum_{t=1}^{T} \Pr(\mathcal{G}_t^c) \leq T(1 - p)^R.$$

Let $\mathcal{E}$ be the confidence event of Assumption 1, with $\Pr(\mathcal{E}^c) \leq \delta$. A further union bound gives

$$\Pr(\mathcal{E} \cap \mathcal{G}) \geq 1 - \delta - T(1 - p)^R.$$

No independence between $\mathcal{E}$ and $\mathcal{G}$, or between success events in different rounds, is required.

Step 2: a small gradient on a successful trajectory. Fix a realization in $\mathcal{E} \cap \mathcal{G}$ and a round t. Select a restart whose initial iterate lies in B_t, and denote its iterates by $\mathbf{X}_0, \ldots, \mathbf{X}_{T_{\text{in}}}$. Assumption 3(iii) ensures that all these iterates remain in B_t without clipping. Its initial value gap satisfies

$$\Delta_{t,0} := \mathcal{U}_t^\star - \mathcal{U}_t(\mathbf{X}_0) \leq \sup_{\Omega^q} \mathcal{U}_t - \inf_{\Omega^q} \mathcal{U}_t \leq D_{\mathcal{U}}.$$

Theorem 6.1 therefore gives

$$\frac{1}{T_{\text{in}}} \sum_{k=0}^{T_{\text{in}}-1} \|\nabla \mathcal{U}_t(\mathbf{X}_k)\|_* \leq \frac{\Delta_{t,0}}{T_{\text{in}}\eta} + \frac{L_{\mathcal{U}} q \eta}{2} \leq \frac{D_{\mathcal{U}}}{T_{\text{in}}\eta} + \frac{L_{\mathcal{U}} q \eta}{2}.$$

For the stated step size, both terms on the right equal $\sqrt{L_{\mathcal{U}} q D_{\mathcal{U}} / (2T_{\text{in}})}$. Hence

$$\min_{0 \leq k < T_{\text{in}}} \|\nabla \mathcal{U}_t(\mathbf{X}_k)\|_* \leq \frac{1}{T_{\text{in}}} \sum_{k=0}^{T_{\text{in}}-1} \|\nabla \mathcal{U}_t(\mathbf{X}_k)\|_* \leq \sqrt{\frac{2 L_{\mathcal{U}} q D_{\mathcal{U}}}{T_{\text{in}}}}.$$

Step 3: conversion to acquisition value accuracy. Let $k^\star$ attain the minimum in Step 2. Since $\mathbf{X}_{k^\star} \in B_t$, the PŁ inequality and $\|\mathbf{A}\|_F \leq \|\mathbf{A}\|_*$ imply

$$\begin{aligned} \mathcal{U}_t^\star - \mathcal{U}_t(\mathbf{X}_{k^\star}) &\leq \frac{\|\nabla \mathcal{U}_t(\mathbf{X}_{k^\star})\|_F^2}{2\mu_{\text{PL}}} \\ &\leq \frac{\|\nabla \mathcal{U}_t(\mathbf{X}_{k^\star})\|_*^2}{2\mu_{\text{PL}}} \\ &\leq \frac{L_{\mathcal{U}} q D_{\mathcal{U}}}{\mu_{\text{PL}} T_{\text{in}}}. \end{aligned}$$

The algorithm returns the best iterate across all restarts under exact acquisition evaluation. Thus its returned batch $\hat{\mathbf{X}}_t$ satisfies $\mathcal{U}_t(\hat{\mathbf{X}}_t) \geq \mathcal{U}_t(\mathbf{X}_{k^\star})$. Defining the actual optimization error by

$$\epsilon_t := \mathcal{U}_t^\star - \mathcal{U}_t(\hat{\mathbf{X}}_t),$$

we obtain

$$0 \leq \epsilon_t \leq \frac{L_{\mathcal{U}} q D_{\mathcal{U}}}{\mu_{\text{PL}} T_{\text{in}}}.$$

This argument is deterministic for each successful realization, so selecting a successful restart introduces no further probabilistic assumption.

Step 4: substitution into the regret bound. On $\mathcal{E} \cap \mathcal{G}$, the bound from Step 3 holds simultaneously for all rounds. Theorem 4.1 therefore yields

$$R_N \leq \sqrt{C_1 N \beta_T \gamma_N} + \sum_{t=1}^{T} \epsilon_t.$$

Using $N = qT$,

$$\sum_{t=1}^{T} \epsilon_t \leq T \, \frac{L_{\mathcal{U}} q D_{\mathcal{U}}}{\mu_{\text{PL}} T_{\text{in}}} = \frac{N L_{\mathcal{U}} D_{\mathcal{U}}}{\mu_{\text{PL}} T_{\text{in}}}.$$

Combining this inequality with the probability estimate in Step 1 proves the corollary. □

A.5 Stochastic Muon: Monte-Carlo gradients and a growing sample schedule

Fix the GP posterior and let α denote the underlying acquisition. When its gradient is estimated using fresh Monte Carlo samples at each step (resampled qEI, for instance), the gradient handed to Algorithm 1 is stochastic. The following theorem is the stochastic counterpart of Theorem 6.1.

Theorem A.12 (Stochastic Muon; growing samples remove the floor)**.** *Let Assumption 2 hold. Let $\mathcal{F}_k$ be the σ-algebra generated by the Monte-Carlo samples drawn at steps $0, \dots, k-1$, with $\mathbf{X}_0 \in \Omega^q$ deterministic (so that $\mathbf{X}_k$ and $\mathbf{G}_k = \nabla\alpha(\mathbf{X}_k)$ are $\mathcal{F}_k$-measurable), let the sample counts $S_k \in \{1, 2, \dots\}$ be a deterministic schedule, and let the gradient estimate $\hat{\mathbf{G}}_k$, computed from S_k fresh Monte-Carlo samples, satisfy $\mathbb{E}[\|\hat{\mathbf{G}}_k - \mathbf{G}_k\|_F^2 \,|\, \mathcal{F}_k] \le \varsigma^2/S_k$. Suppose no clipping occurs almost surely. Then Algorithm 1 with $\hat{\mathbf{Q}}_k = \mathrm{Polar}(\hat{\mathbf{G}}_k)$ and constant $\eta > 0$ satisfies*

$$\frac{1}{T_{\mathrm{in}}} \sum_{k=0}^{T_{\mathrm{in}}-1} \mathbb{E}\,\|\mathbf{G}_k\|_* \le \frac{\Delta_0}{T_{\mathrm{in}}\eta} + \frac{L_F q \eta}{2} + \frac{2\sqrt{q}\,\varsigma}{T_{\mathrm{in}}} \sum_{k=0}^{T_{\mathrm{in}}-1} \frac{1}{\sqrt{S_k}}.$$

With $S_k \equiv S$ the last term is the floor $2\sqrt{q}\varsigma/\sqrt{S}$, which no choice of step size reduces in this bound; this is the floor that momentum-free stochastic Muon analyses exhibit (Shen et al., 2025; Sato et al., 2025; Lau et al., 2025), and the reason they otherwise rely on momentum or on a batch of order ϵ^{-2} (Shen et al., 2025, Thm. 4.1). With $S_k = S_0(k+1)$ it is at most $4\sqrt{q}\varsigma/\sqrt{S_0 T_{\mathrm{in}}} \to 0$, at $O(S_0 T_{\mathrm{in}}^2)$ surrogate *samples and zero evaluations of f. If clipping is allowed, the same bound holds with $\mathbb{E}[\|\mathbf{G}_k\|_* - \hat{E}_k]$ in place of $\mathbb{E}\|\mathbf{G}_k\|_*$ on the left, where $\hat{E}_k = \sum_{(i,j)\in\mathcal{A}_k} |G_k^{ij}||\hat{Q}_k^{ij}|$ is the boundary term of Theorem 6.1 evaluated at the stochastic polar factor $\hat{\mathbf{Q}}_k = \mathrm{Polar}(\hat{\mathbf{G}}_k)$ and $\mathcal{A}_k$ is the set of coordinates clipped in $\Pi_\Omega(\mathbf{X}_k + \eta\hat{\mathbf{Q}}_k)$ (App. A.4).*

Growing the sample size to defeat a variance floor is a classical device in stochastic optimization (Friedlander & Schmidt, 2012), and Nagashima & Iiduka (2026) apply it to Muon with an exponentially growing batch. In BO, these additional samples require surrogate computation rather than new evaluations of f, making sample growth a practical alternative to momentum. The linear schedule above spends $O(S_0 T_{\mathrm{in}}^2)$ surrogate samples in total. The resampled MC-qEI experiments of Appendix B.3.2 and the momentum ablation of Appendix B.3.1 are the empirical companions of this theorem.

A.5.1 Proof of Theorem A.12

Step 1: a deterministic inequality. For any $\mathbf{G}$ and any $\hat{\mathbf{G}}$, let $\hat{\mathbf{Q}} = \mathrm{Polar}(\hat{\mathbf{G}})$. If $\hat{\mathbf{G}} = 0$ then $\hat{\mathbf{Q}} = 0$ and $\langle \mathbf{G}, \hat{\mathbf{Q}} \rangle = 0 \ge \|\mathbf{G}\|_* - 2\|\mathbf{G} - \hat{\mathbf{G}}\|_*$ holds because $\|\mathbf{G} - \hat{\mathbf{G}}\|_* = \|\mathbf{G}\|_*$. If $\hat{\mathbf{G}} \neq 0$, then

$$\begin{aligned}
\langle \mathbf{G}, \hat{\mathbf{Q}} \rangle &= \langle \hat{\mathbf{G}}, \hat{\mathbf{Q}} \rangle + \langle \mathbf{G} - \hat{\mathbf{G}}, \hat{\mathbf{Q}} \rangle \\
&\overset{(a)}{\ge} \|\hat{\mathbf{G}}\|_* - \|\mathbf{G} - \hat{\mathbf{G}}\|_* \, \|\hat{\mathbf{Q}}\|_2 \\
&\overset{(b)}{=} \|\hat{\mathbf{G}}\|_* - \|\mathbf{G} - \hat{\mathbf{G}}\|_* \\
&\overset{(c)}{\ge} \|\mathbf{G}\|_* - 2\|\mathbf{G} - \hat{\mathbf{G}}\|_*,
\end{aligned}$$

where (a) uses Lemma A.1(i) for the first term and the nuclear–spectral Hölder pairing for the second, (b) uses $\|\hat{\mathbf{Q}}\|_2 = 1$ (Lemma A.1(iii)), and (c) is the reverse triangle inequality $\|\hat{\mathbf{G}}\|_* \ge \|\mathbf{G}\|_* - \|\mathbf{G} - \hat{\mathbf{G}}\|_*$.

Step 2: the error in expectation. $\mathbf{G}_k - \hat{\mathbf{G}}_k$ is a $q \times d$ matrix, so $\mathrm{rank}(\mathbf{G}_k - \hat{\mathbf{G}}_k) \le q$ and $\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_* \le \sqrt{q}\,\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_F$ (Cauchy–Schwarz on the singular values). Conditional Jensen then gives

$$\mathbb{E}\big[\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_* \,\big|\, \mathcal{F}_k\big] \le \sqrt{q}\,\mathbb{E}\big[\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_F \,\big|\, \mathcal{F}_k\big] \le \sqrt{q}\,\sqrt{\mathbb{E}\big[\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_F^2 \,\big|\, \mathcal{F}_k\big]} \le \frac{\sqrt{q}\,\varsigma}{\sqrt{S_k}}.$$

Step 3: conditional ascent. $\mathbf{X}_k$, hence $\mathbf{G}_k$, is $\mathcal{F}_k$-measurable. With no clipping, $\mathbf{X}_{k+1} - \mathbf{X}_k = \eta\hat{\mathbf{Q}}_k$ and $\|\eta\hat{\mathbf{Q}}_k\|_F^2 = \eta^2 \mathrm{rank}(\hat{\mathbf{G}}_k) \le \eta^2 q$, so Lemma A.2 plus Steps 1–2 give

$$\mathbb{E}\big[\alpha(\mathbf{X}_{k+1}) \,\big|\, \mathcal{F}_k\big] \;\ge\; \alpha(\mathbf{X}_k) + \eta\Big(\|\mathbf{G}_k\|_* - \frac{2\sqrt{q}\,\varsigma}{\sqrt{S_k}}\Big) - \frac{L_F q}{2}\eta^2.$$

Taking total expectations, summing over k, and bounding the telescoped sum by Δ_0 yields, after division by $T_{\text{in}}\eta$, the display. (The conditional expectation is well defined: α is bounded on the compact Ω^q, and $\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_*$ is integrable by the second-moment hypothesis; in Step 3′ below, $0 \le \hat{E}_k \le \|\mathbf{G}_k\|_F \|\hat{\mathbf{Q}}_k\|_F \le \sqrt{q}\,\|\mathbf{G}_k\|_F$ is bounded.)

Step 3′: with clipping. If the projection is active, write $\mathbf{D}_k = \mathbf{X}_{k+1} - \mathbf{X}_k$ and $\mathbf{S}_k = \eta\hat{\mathbf{Q}}_k - \mathbf{D}_k$. Step 1 of the proof of Theorem 6.1 applies verbatim with $\hat{\mathbf{Q}}_k$ in place of $\mathbf{Q}_k$ (it uses only the coordinate-wise form of the projection and $\eta > 0$), so $\mathbf{S}_k$ is supported on $\mathcal{A}_k$ with $S_k^{ij}\hat{Q}_k^{ij} \ge 0$ and $|S_k^{ij}| \le \eta|\hat{Q}_k^{ij}|$, and $|D_k^{ij}| \le \eta|\hat{Q}_k^{ij}|$ for every coordinate. Hence, by Step 1 above,

$$\langle \mathbf{G}_k,\, \mathbf{D}_k\rangle = \eta\langle \mathbf{G}_k,\, \hat{\mathbf{Q}}_k\rangle - \langle \mathbf{G}_k,\, \mathbf{S}_k\rangle \;\ge\; \eta\big(\|\mathbf{G}_k\|_* - 2\|\mathbf{G}_k - \hat{\mathbf{G}}_k\|_* - \hat{E}_k\big),\; \hat{E}_k = \sum_{(i,j)\in\mathcal{A}_k} |G_k^{ij}||\hat{Q}_k^{ij}|,$$

and $\|\mathbf{D}_k\|_F^2 \le \eta^2\|\hat{\mathbf{Q}}_k\|_F^2 \le \eta^2 q$. Repeating Steps 2–3 with these two bounds gives the display of the theorem with $\mathbb{E}[\|\mathbf{G}_k\|_* - \hat{E}_k]$ on the left.

Step 4: the schedule. For $S_k = S_0(k+1)$,

$$\sum_{k=0}^{T_{\text{in}}-1} \frac{1}{\sqrt{S_k}} = \frac{1}{\sqrt{S_0}} \sum_{m=1}^{T_{\text{in}}} \frac{1}{\sqrt{m}} \le \frac{1}{\sqrt{S_0}}\Big(1 + \int_1^{T_{\text{in}}} t^{-1/2}dt\Big) \le \frac{2\sqrt{T_{\text{in}}}}{\sqrt{S_0}},$$

so the third term is at most $(2\sqrt{q}\varsigma/T_{\text{in}}) \cdot 2\sqrt{T_{\text{in}}/S_0} = 4\sqrt{q}\varsigma/\sqrt{S_0 T_{\text{in}}}$, and the total sample count is $\sum_k S_k = S_0 T_{\text{in}}(T_{\text{in}}+1)/2 = O(S_0 T_{\text{in}}^2)$. □

B Additional experimental results

This appendix is organized by the main-text section each result belongs to: Appendix B.1 complements Figure 1 of Section 4, Appendix B.3 complements the solver results of Section 6 and their stochastic extension in Appendix A.5, and Appendix B.4 complements the benchmarks of Section 7, one subsection per benchmark in the order of the main text. Throughout, numbers are mean ± standard deviation over seeds and the curve bands are ± one standard error.

B.1 Section 4: GP-draw objectives

Table 1 reports the final cumulative and simple regret behind Figure 1. The joint solvers start each round from the same $R = 8$ uniform batches, compute-matched greedy receives their gradient budget split q ways, and unlimited greedy runs 6 restarts of 100 L-BFGS-B iterations per row. Muon, Adam, L-BFGS-B and matched greedy finish within 4% of one another in R_N at every d, differences of about one standard deviation, while projected GD finishes 4 to 12% higher and unlimited greedy 4 to 7% lower. Theorem 4.1 does not attribute these observed differences to $\sum_t \epsilon_t$. Notice here we use $\beta_t \equiv 4$ as a practical exploration setting similar to the modified experimental schedule of Desautels et al. (2014), and this choice is not covered by the theoretical confidence guarantee of Assumption 1 (also see Appendix C.1).

Table 1: GP-draw objectives, chained batch-UCB, $q = 8$ (Figure 1). Final cumulative regret R_N and final simple regret, mean ± SD over 50 draws (25 at $d = 50$). All joint-solver methods share the same $R = 8$ uniformly sampled starting batches.

Method	R_N, d=10	R_N, d=20	R_N, d=50	simple, d=10	simple, d=20	simple, d=50
Muon	1060 ± 61	2420 ± 106	5000 ± 140	1.74 ± 0.57	2.58 ± 0.85	7.26 ± 0.9
Proj. GD	1190 ± 55	2640 ± 92	5220 ± 140	2.47 ± 0.59	4.16 ± 0.71	9.00 ± 0.75
Adam	1100 ± 64	2490 ± 92	5070 ± 145	1.94 ± 0.68	2.81 ± 0.85	7.62 ± 1.05
L-BFGS-B	1070 ± 61	2450 ± 92	5040 ± 140	1.65 ± 0.57	2.77 ± 0.85	7.46 ± 0.65
Greedy (matched)	1060 ± 67	2440 ± 85	5040 ± 135	1.70 ± 0.71	2.87 ± 0.78	7.75 ± 0.95
Greedy (unlimited)	992 ± 71	2260 ± 106	4800 ± 140	1.52 ± 0.47	2.15 ± 0.71	5.63 ± 0.80

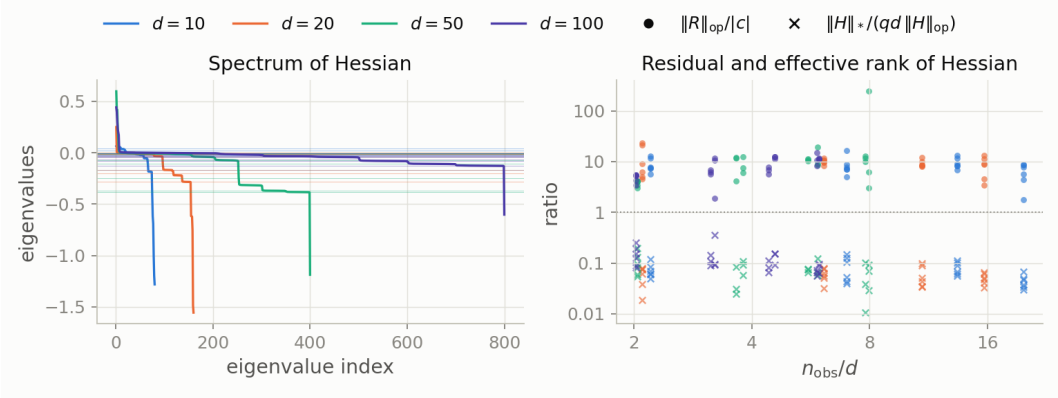


Figure 5: Curvature structure along Muon-driven BO runs on GP-draw objectives with $q \in \{8, 16\}$, $d \in \{10, 20, 50, 100\}$, and $n_{\text{init}} = 2d + 2$. Each configuration uses three runs and four snapshots per run, giving 96 snapshots in total. Left panel shows eigenvalues of the scaled batch Hessian at the first snapshot of one run per dimension with $q = 8$ ($n_{\text{obs}} = 22, 42, 102, 202$), with horizontal lines marking the empirically estimated bulk coefficients of its diagonal blocks (Corollary 5.2). Right panel presents relative residual magnitude $\|\mathbf{R}\|_2/|c|$ for diagonal blocks (dots) and normalized nuclear effective rank $\|\mathbf{H}\|_*/(qd\,\|\mathbf{H}\|_2)$ of the batch Hessian (crosses), plotted against n_{obs}/d across all snapshots. The dashed horizontal line marks 1.

B.2 Section 5: Curvature Structure

Figure 5 summarizes 96 Hessian snapshots from Muon-driven BO runs on GP draws at $d \in \{10, 20, 50, 100\}$. The left panel compares representative scaled Hessian spectra with the empirically estimated block coefficients of Corollary 5.2. The right panel tracks the relative residual magnitude $\|\mathbf{R}\|_2/|c|$ and the normalized nuclear effective rank $\|\mathbf{H}\|_*/(qd\,\|\mathbf{H}\|_2)$ across snapshots. The spectra show a largely flat region together with outlying eigenvalues. The residual ratio is typically between 3 and 30, with one snapshot at $d = 50$ near 250, indicating that the residual often dominates the scaled-metric component in operator norm. Meanwhile, the effective-rank fraction ranges from approximately 0.01 to 0.3. These spectra suggest that curvature is concentrated in relatively few directions, although they do not establish the directional-curvature condition of Proposition 5.3. A polar-step advantage additionally requires favorable gradient–curvature alignment. This concentration is also empirical: at the sample sizes considered, the theoretical residual-rank bound equals the block dimension d and imposes no nontrivial rank restriction.

B.3 Section 6: The Inner Solver

B.3.1 Momentum

The main text uses and analyzes momentum-free Muon. Here, we evaluate a momentum variant that replaces $\text{Polar}(\mathbf{G}_k)$ with $\text{Polar}(\mathbf{M}_k)$, where

$$\mathbf{M}_k = \beta_{\text{mom}}\mathbf{M}_{k-1} + (1 - \beta_{\text{mom}})\mathbf{G}_k.$$

The polar factor is invariant to positive rescaling, so accumulating momentum does not directly amplify the update through the magnitude of $\mathbf{M}_k$. Momentum instead changes its direction and potentially its rank. We tune the step constant separately for each variant on frozen acquisition surfaces. We compare $\beta_{\text{mom}} \in \{0, 0.9, 0.95\}$ on Ackley at $d = 20$ using batch-UCB with $q = 8$ and $N = 160$. Each variant uses eight shared restarts, 200 inner iterations per restart, cosine step-size decay, and 20 seeds. The larger inner budget allows more iterations for momentum to influence the trajectory. Table 2 shows similar final simple regret across the three variants: 8.30 ± 0.84, 8.44 ± 1.22, and 8.49 ± 0.94, respectively. Cumulative regret and the average returned acquisition value are also similar. These results show no clear benefit from momentum in this configuration, supporting our use of the simpler momentum-free update.

Table 2: Effect of momentum on Ackley at $d = 20$ using batch-UCB with 8 restarts and 200 inner iterations per restart. Columns report final simple regret, cumulative regret, and the average acquisition value of the selected batches across rounds, each as mean $\pm$ standard deviation over 20 seeds.

Arm	final simple regret	cumulative regret R_N	mean $\mathcal{U}_t$ per round
Muon (β_{mom}=0)	8.30 ± 0.84	2178 ± 61	-77.9 ± 4
Muon (β_{mom}=0.9)	8.44 ± 1.2	2179 ± 66	-77.6 ± 4.1
Muon (β_{mom}=0.95)	8.49 ± 0.94	2199 ± 52	-79.0 ± 4.6

B.3.2 Resampled Monte-Carlo qEI

We evaluate MC-qEI with 64 fresh Monte Carlo samples per gradient evaluation, introducing sampling noise that is redrawn at every inner step. This setting motivates the stochastic analysis in Theorem A.12, although the fixed sample size used here differs from the growing-sample schedule of the theorem. The analysis does not predict which optimizer performs best. Resampling also makes the objective and gradient estimates vary between evaluations, which can complicate standard L-BFGS-B line searches and curvature updates. Table 3 reports final simple regret on GP-draw objectives at $d \in \{10, 20, 50\}$, with evaluation budgets $N = 200, 320, 400$, respectively, and 20 seeds per configuration. We compare Muon, Adam, projected GD, and L-BFGS-B using uniform restarts, with tuned step sizes for the fixed-step methods. Muon and Adam attain the same mean regret at $d = 10$ (2.44 ± 0.43 and 2.44 ± 0.39). Muon achieves the lowest mean at $d = 20$ and $d = 50$, although its margins over Adam are small relative to the reported standard deviations. Projected GD and L-BFGS-B attain higher mean regret across all three dimensions, as Balandat et al. (2020) predicts for a quasi-Newton method on resampled gradients. These results show that Muon remains competitive with Adam under resampled acquisition gradients, without establishing a general advantage in the stochastic setting.

Table 3: Final simple regret on GP-draw objectives using resampled MC-qEI with $S = 64$ Monte Carlo samples and $q = 8$, reported as mean $\pm$ standard deviation over 20 seeds. Bold indicates the lowest mean in each column, including ties at the displayed precision.

Method	d=10	d=20	d=50
Muon	$\mathbf{2.44} \pm 0.43$	$\mathbf{4.62} \pm 0.44$	$\mathbf{9.49} \pm 0.33$
Adam	$\mathbf{2.44} \pm 0.39$	4.91 ± 0.39	9.76 ± 0.55
Proj. GD	3.50 ± 0.4	5.83 ± 0.55	10.2 ± 0.55
L-BFGS-B	3.55 ± 0.33	5.77 ± 0.47	10.3 ± 0.53

B.4 Section 7: benchmarks

B.4.1 Setup

We use batch size $q = 8$ and evaluation budgets $(N, n_{\text{init}}) = (100, 10)$, $(160, 20)$, $(400, 50)$, and $(600, 80)$, for $d \in \{10, 20, 50, 100\}$, respectively. The surrogate is a Matérn-$5/2$ GP with ARD length scales, standardized targets, and observation noise variance 10^{-6}. Hyperparameters are refitted every 1, 1, 3, and 8 rounds, respectively. We compare Muon with projected gradient ascent (projected GD), Adam (Kingma & Ba, 2015), multistart L-BFGS-B with a BoTorch `optimize_acqf` recipe (Balandat et al., 2020), and compute-matched greedy GP-BUCB. For fair comparison, all methods share the same surrogate specification, hyperparameter-fitting procedure, evaluation budget, and scrambled-Sobol initial design for each seed. Muon, projected GD, and Adam each use $R = 8$ uniform restarts with $T_{\text{in}} = 60$ steps and cosine step-size decay. Compute-matched greedy GP-BUCB divides the corresponding RT_{in} gradient-evaluation budget across the q candidate selections.

We report additional results for chained batch-UCB and qLogEI. For chained batch-UCB (2), we include cumulative-regret curves, corresponding to the quantity bounded by Theorem 4.1, and unlimited greedy GP-BUCB as an additional comparator. We also evaluate qLogEI (Ament et al., 2023), a numerically stabilized log formulation of batch expected improvement, using a sample-

average approximation with 128 fixed base samples. These experiments assess whether Muon's observed benefits extend to another acquisition. The local comparisons in Propositions 5.3–5.4 apply wherever their differentiability and curvature assumptions hold, but neither the single-point EI Hessian decomposition (Theorem 5.1) nor our chained batch-UCB regret guarantee (Theorem 4.1 and Corollary A.11) automatically extends to qLogEI. We retain the benchmark setup used for chained batch-UCB and compare joint solvers with sequential-greedy construction following BoTorch's `sequential=True` recipe under compute-matched and unlimited optimization budgets. Appendix B.4.6 reports final simple regret for all methods. Muon's advantages are less consistent under qLogEI, indicating that the empirical benefit depends on the acquisition and benchmark.

B.4.2 LEVY

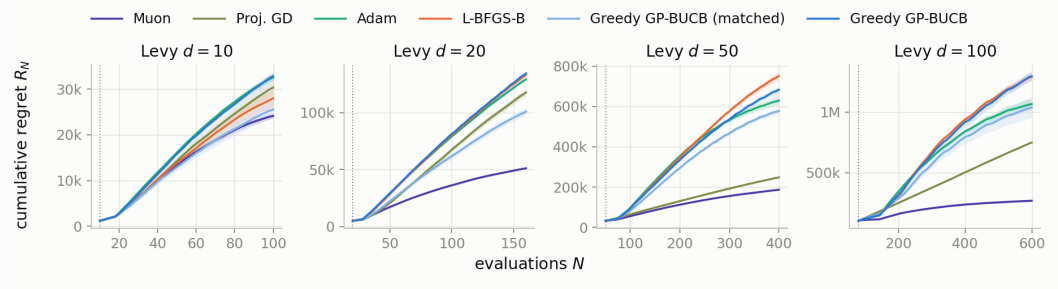


Figure 6: Cumulative regret R_N on Levy using batch-UCB with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

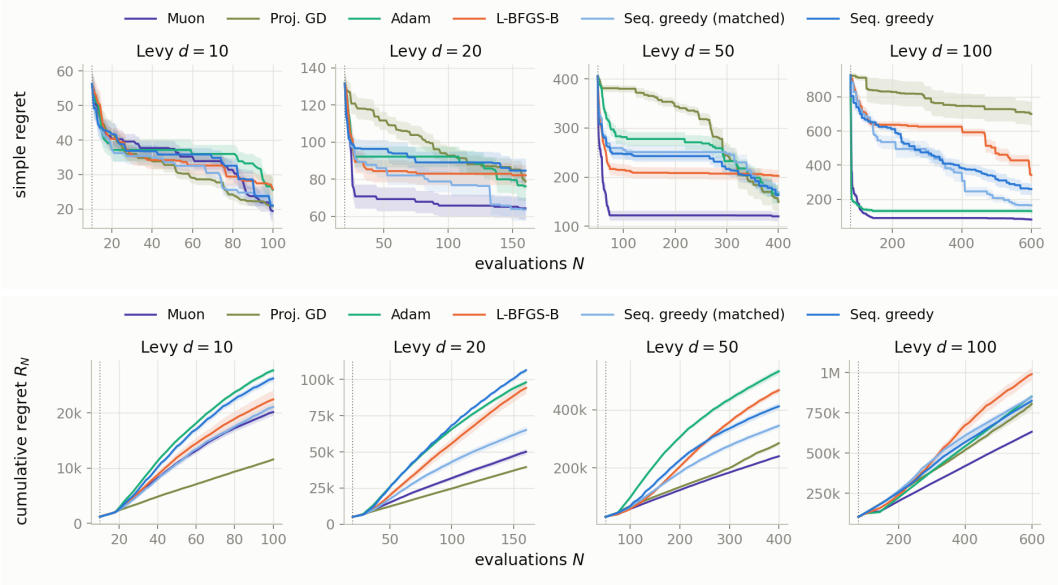


Figure 7: Simple regret (top) and cumulative regret (bottom) on Levy using qLogEI with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

Figure 6 reports cumulative regret for the Levy experiments in Section 7.1, and Figure 7 shows simple and cumulative regret under qLogEI. Under chained batch-UCB, Muon achieves the lowest mean cumulative regret across all dimensions. Under qLogEI, Muon achieves the lowest mean final simple regret at $d \in \{10, 50, 100\}$. At $d = 20$, compute-matched sequential greedy is slightly lower (63.6 ± 27 versus 64.2 ± 26 for Muon). At $d = 100$, Muon achieves 81.8 ± 26, compared to 342 ± 123 for L-BFGS-B and 130 ± 38 for Adam over 10 seeds.

B.4.3 RASTRIGIN

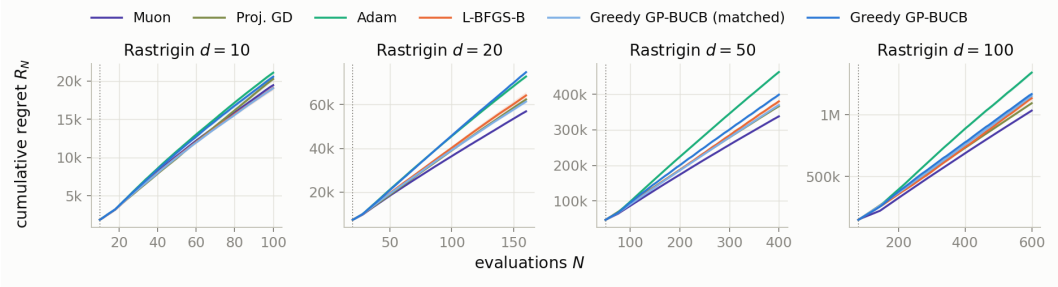


Figure 8: Cumulative regret R_N on Rastrigin using batch-UCB with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

Figure 8 shows cumulative regret for the Rastrigin experiments in Section 7.1, and Figure 9 shows simple and cumulative regret under qLogEI. Under chained batch-UCB, Muon achieves the lowest mean cumulative regret across all dimensions except for $d = 10$. L-BFGS-B achieves the lowest mean final simple regret under qLogEI at $d = 10$ and $d = 20$, with 86.3 ± 14 and 220 ± 20, compared to 102 ± 17 and 222 ± 19 for Muon. Muon performs best at $d = 50$ (549 ± 41), while Adam achieves the lowest mean at $d = 100$ (927 ± 51).

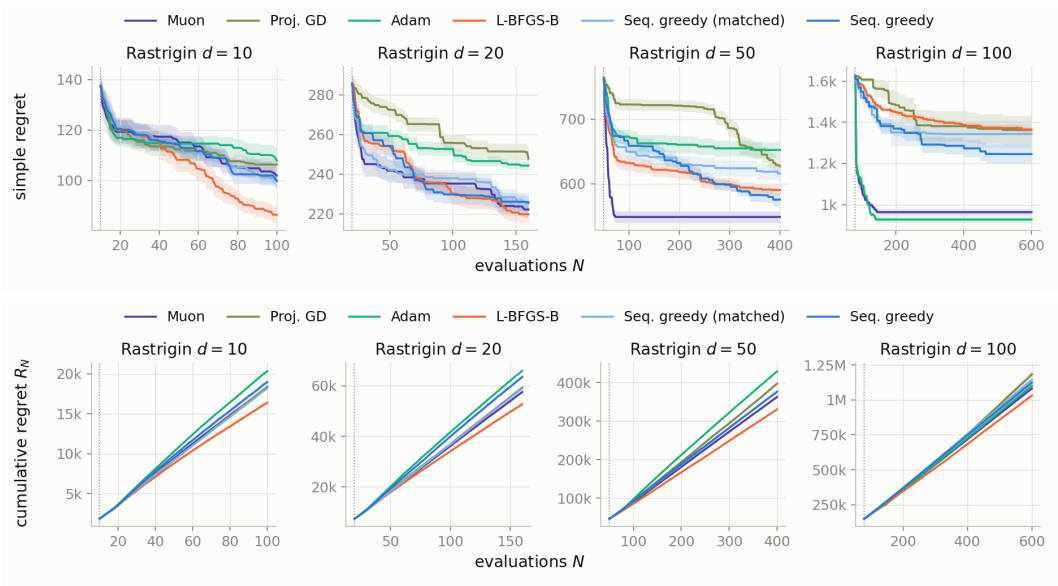


Figure 9: Simple regret (top) and cumulative regret (bottom) on Rastrigin using qLogEI with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

B.4.4 ACKLEY

Figure 10 reports cumulative regret for the Ackley experiments in Section 7.1, and Figure 11 shows simple and cumulative regret under qLogEI. Similarly, under the batch-UCB, Muon achieves the lowest mean cumulative regret across all dimensions except for $d = 10$. Muon achieves the lowest mean final simple regret under qLogEI at $d \in \{10, 20, 100\}$, reaching 3.78 ± 0.45, 5.90 ± 0.72, and 8.48 ± 0.51, respectively. At $d = 50$, L-BFGS-B is slightly lower (6.40 ± 1.2 versus 6.46 ± 0.8 for Muon).

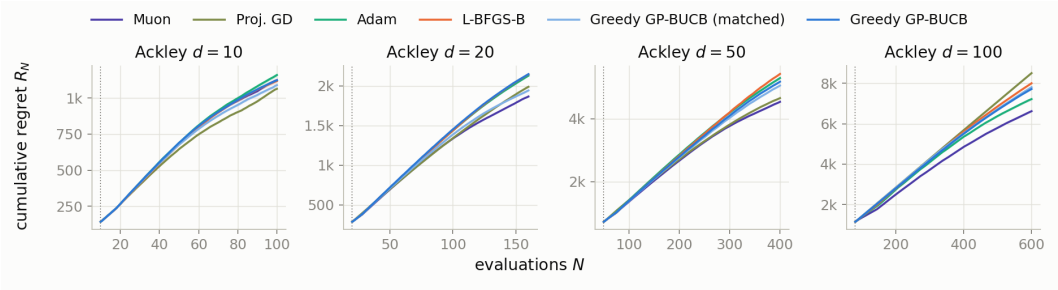


Figure 10: Cumulative regret R_N on Ackley using batch-UCB with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

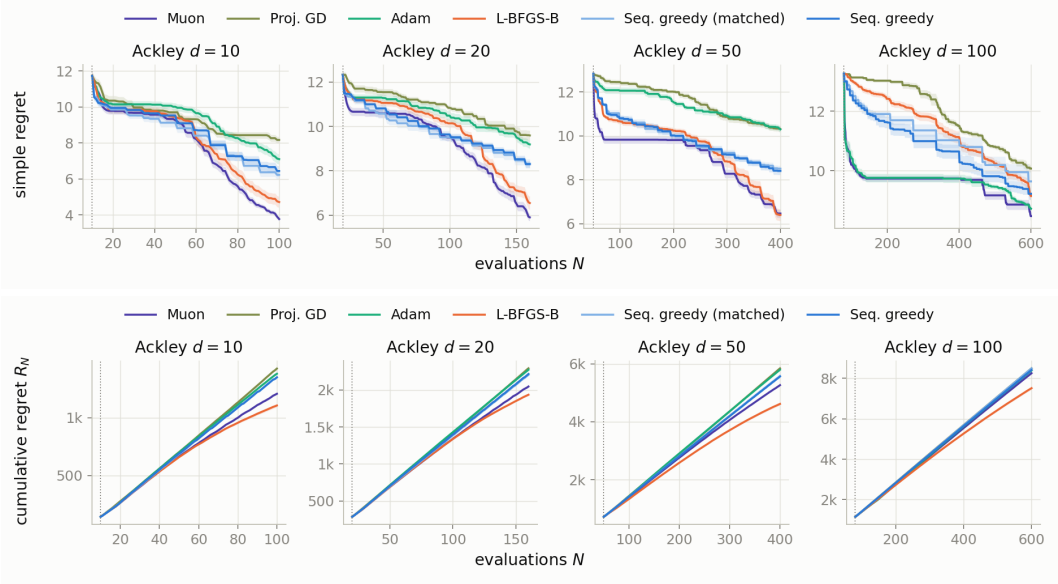


Figure 11: Simple regret (top) and cumulative regret (bottom) on Ackley using qLogEI with $q = 8$, plotted against the number of evaluations. Curves show means over 20 seeds (10 at $d = 100$), with bands indicating $\pm$ one standard error.

B.4.5 ROVER-60

We use $N = 600$ evaluations, $n_{\text{init}} = 50$, $q = 8$, hyperparameter refits every third round, and 20 seeds. This budget is substantially smaller than the 20,000 evaluations used by Eriksson et al. (2019), so our results are not directly comparable. For this task, the fixed-step solvers use 10 restarts and 200 steps per restart, matching the restart count and iteration cap of the BoTorch L-BFGS-B configuration. All gradient-based joint-solvers use the same BoTorch-style starting batches. The cumulative regret is computed as the cumulative shortfall to the reward upper bound 5, $R_N = \sum_{j=1}^{N}(5 - r(\mathbf{x}_j))$, where $r(\mathbf{x}_j)$ is the trajectory reward. This quantity upper-bounds the actual cumulative regret relative to the unknown attainable maximum reward. Table 4 below lists every arm of Figure 4, with projected GD achieving the highest final best reward and Muon attaining the lowest cumulative regret.

B.4.6 FINAL SIMPLE REGRET OF EVERY ARM

Table 5 reports final simple regret for all methods under batch-UCB, including unlimited greedy GP-BUCB and random screening, which selects the highest-acquisition-value batch from 1024 uniformly sampled batches. Table 6 reports the corresponding results for qLogEI. Under batch-UCB, Muon achieves the lowest mean final simple regret in ten of the twelve configurations, and the gaps are small relative to the reported standard deviations in the remaining two cases. Under qLogEI, Muon

Table 4: Results on Rover-60 using batch-UCB with $q = 8$ and $N = 600$. Columns report average gradient evaluations per round, final best reward (higher is better), and cumulative regret R_N (lower is better). Performance values are mean $\pm$ standard deviation over 20 seeds; bold indicates the best mean in each performance column. All gradient-based joint solvers use the same BoTorch-style starting batches.

Method	gradient calls / round	final best reward	cumulative regret R_N
Muon	2,007	1.83 ± 0.87	$\mathbf{7128} \pm 396$
Proj. GD	2,007	$\mathbf{1.93} \pm 1$	7516 ± 400
Adam	2,007	0.20 ± 1.2	9693 ± 730
L-BFGS-B	2,073	1.62 ± 0.69	7688 ± 485
Greedy GP-BUCB	35,499	0.33 ± 0.92	9255 ± 371
Greedy GP-BUCB (matched)	4,485	0.27 ± 1.2	9701 ± 493
Random screen	0	-3.85 ± 1.4	12316 ± 99

achieves the lowest mean final simple regret in seven of the twelve configurations, L-BFGS-B in three, compute-matched sequential greedy in one, and Adam in one. Similarly, in the five qLogEI configurations where Muon is not the best, the gap between Muon and the lowest-mean method is small relative to the reported standard deviations, except for Rastrigin at $d = 10$.

Table 5: Final simple regret under chained batch-UCB with $q = 8$, reported as mean $\pm$ standard deviation over the stated number of seeds. Bold indicates the lowest mean in each row. "Greedy" denotes GP-BUCB with either a compute-matched budget or an unlimited budget of six L-BFGS-B restarts with up to 100 iterations per candidate.

Cell	seeds	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)	Random screen
Levy-10	20	$\mathbf{20.4} \pm 16$	35.1 ± 18	31.7 ± 16	27.1 ± 16	28.6 ± 13	31.9 ± 15	28.5 ± 9.8
Levy-20	20	$\mathbf{50.5} \pm 21$	86.1 ± 30	97.6 ± 30	92.9 ± 21	84.1 ± 25	89.5 ± 31	88.7 ± 17
Levy-50	20	$\mathbf{131} \pm 16$	290 ± 42	182 ± 72	193 ± 42	207 ± 67	248 ± 63	331 ± 28
Levy-100	10	$\mathbf{51.5} \pm 20$	747 ± 60	325 ± 89	439 ± 82	323 ± 149	426 ± 79	791 ± 51
Rastrigin-10	20	106 ± 18	112 ± 13	119 ± 15	107 ± 17	105 ± 13	111 ± 15	$\mathbf{103} \pm 15$
Rastrigin-20	20	$\mathbf{221} \pm 20$	248 ± 20	256 ± 20	247 ± 24	234 ± 30	246 ± 14	250 ± 13
Rastrigin-50	20	$\mathbf{552} \pm 34$	668 ± 54	651 ± 54	612 ± 36	577 ± 80	592 ± 58	718 ± 14
Rastrigin-100	10	$\mathbf{900} \pm 20$	1520 ± 38	1340 ± 82	1370 ± 117	1280 ± 199	1160 ± 73	1510 ± 30
Ackley-10	20	3.54 ± 0.49	$\mathbf{3.14} \pm 0.37$	5.26 ± 2	4.11 ± 0.89	3.86 ± 1.4	4.34 ± 1.1	8.42 ± 0.8
Ackley-20	20	$\mathbf{4.38} \pm 0.41$	6.84 ± 1	8.76 ± 1.2	8.48 ± 1.3	6.18 ± 1.5	8.57 ± 0.94	10.6 ± 0.4
Ackley-50	20	$\mathbf{4.83} \pm 0.63$	5.15 ± 0.31	9.17 ± 0.85	9.38 ± 0.98	7.60 ± 0.63	8.20 ± 1.1	11.8 ± 0.4
Ackley-100	10	$\mathbf{5.99} \pm 0.47$	12.7 ± 0.24	6.66 ± 0.38	10.4 ± 0.6	9.78 ± 0.54	9.09 ± 0.35	12.7 ± 0.14

Table 6: Final simple regret under qLogEI with $q = 8$, reported as mean $\pm$ standard deviation over the stated number of seeds. Bold indicates the lowest mean in each row. "Greedy" denotes BoTorch's sequential-greedy qLogEI construction with either a compute-matched or unlimited budget.

Cell	seeds	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)	Random screen
Levy-10	20	$\mathbf{19.4} \pm 13$	20.8 ± 5.8	25.6 ± 16	25.8 ± 13	21.1 ± 8.5	21.0 ± 12	25.2 ± 8.5
Levy-20	20	64.2 ± 26	78.8 ± 20	76.1 ± 36	82.0 ± 25	$\mathbf{63.6} \pm 27$	84.6 ± 29	85.0 ± 19
Levy-50	20	$\mathbf{120} \pm 41$	150 ± 121	167 ± 85	202 ± 58	169 ± 63	163 ± 67	339 ± 26
Levy-100	10	$\mathbf{81.8} \pm 26$	700 ± 225	130 ± 38	342 ± 123	165 ± 76	260 ± 108	781 ± 41
Rastrigin-10	20	102 ± 17	106 ± 12	108 ± 17	$\mathbf{86.3} \pm 14$	100 ± 12	99.7 ± 14	106 ± 9.4
Rastrigin-20	20	222 ± 19	248 ± 19	244 ± 9.8	$\mathbf{220} \pm 20$	225 ± 16	226 ± 21	241 ± 21
Rastrigin-50	20	$\mathbf{549} \pm 41$	627 ± 49	652 ± 54	590 ± 37	616 ± 49	575 ± 45	709 ± 29
Rastrigin-100	10	963 ± 51	1360 ± 218	$\mathbf{927} \pm 51$	1360 ± 47	1340 ± 206	1250 ± 152	1510 ± 41
Ackley-10	20	$\mathbf{3.78} \pm 0.45$	8.17 ± 0.8	7.10 ± 1.3	4.72 ± 1.3	6.23 ± 1.2	6.44 ± 1.3	7.64 ± 0.94
Ackley-20	20	$\mathbf{5.90} \pm 0.72$	9.60 ± 0.98	9.19 ± 0.8	6.55 ± 1.3	8.32 ± 0.94	8.30 ± 0.67	9.71 ± 0.54
Ackley-50	20	6.46 ± 0.8	10.3 ± 0.58	10.3 ± 0.67	$\mathbf{6.40} \pm 1.2$	8.50 ± 0.63	8.40 ± 0.8	11.4 ± 0.28
Ackley-100	10	$\mathbf{8.48} \pm 0.51$	10.1 ± 0.38	8.72 ± 0.27	9.15 ± 0.54	9.64 ± 1.1	9.23 ± 0.6	12.3 ± 0.17

B.4.7 THE BATCH-SIZE SWEEP AND WALL-CLOCK

Table 7 reports the effect of batch size $q \in \{2, 4, 8, 16, 32\}$ on final simple regret. We evaluate batch-UCB on Levy, Rastrigin, and Ackley at $d \in \{20, 50\}$, using the fixed total evaluation budgets of Section B.4.1 and 15 seeds per configuration. Muon achieves the strictly lowest mean regret in 24 of the 30 configurations. Unlimited greedy GP-BUCB is lowest on Levy-20 at $q = 2$, and Adam is lowest on Levy-50 at $q = 2$ and $q = 4$. Compute-matched greedy GP-BUCB is lowest on Levy-50 and Rastrigin-50 at $q = 32$, and projected GD essentially ties with Muon on Ackley-50 at $q = 2$.

The effect of batch size also varies across benchmarks. Muon's regret decreases monotonically with batch size on Levy-50 and Rastrigin-50 and decreases overall, though not monotonically, on Rastrigin-20. Its regret increases with batch size on Ackley-20 and Ackley-50. Several other methods also worsen overall on Ackley, but the trend is not universal or monotonic. These results indicate that the benefit of increasing q depends on the problem. Note that here the configurations with $q = 32$ and $d = 20$ additionally explore $q > d$, beyond the $q \leq d$ regime assumed in our theoretical results. The polar update still remains well-defined in this regime, so these configurations provide additional empirical comparisons under the same evaluation protocol, while our theoretical claims retain their stated scope.

Table 8 reports inner optimization time per round. Compute-matched greedy GP-BUCB takes approximately 0.50–0.95 times Muon's wall-clock time: matching nominal iteration budgets does not match runtime, because single-point acquisition gradients are cheaper to evaluate than joint batch gradients. Unlimited greedy GP-BUCB, using six L-BFGS-B restarts with up to 100 iterations per candidate, costs approximately 4–50 times as much as the joint solve, with the relative cost increasing with q. Appendix C.3 reports wall-clock results for the main experiments.

Table 7: Final simple regret across batch sizes under batch-UCB with a fixed total evaluation budget N, reported as mean $\pm$ standard deviation over 15 seeds. Bold indicates the lowest mean in each row.

Cell	q	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)
Levy-20	2	64.2 ± 26	57.0 ± 39	53.7 ± 36	52.6 ± 19	53.6 ± 28	$\mathbf{44.7} \pm 18$
	4	$\mathbf{60.9} \pm 24$	87.2 ± 33	63.0 ± 36	79.4 ± 35	72.1 ± 38	76.1 ± 36
	8	$\mathbf{48.6} \pm 23$	88.7 ± 32	102 ± 33	94.1 ± 23	87.2 ± 26	95.8 ± 31
	16	$\mathbf{55.1} \pm 21$	79.8 ± 22	89.8 ± 27	92.0 ± 29	67.4 ± 33	95.2 ± 20
	32	$\mathbf{48.1} \pm 17$	73.6 ± 17	78.1 ± 20	82.7 ± 21	66.0 ± 14	84.8 ± 25
Levy-50	2	191 ± 31	252 ± 38	$\mathbf{112} \pm 38$	153 ± 54	140 ± 31	131 ± 27
	4	158 ± 25	255 ± 50	$\mathbf{146} \pm 38$	182 ± 66	166 ± 70	198 ± 74
	8	$\mathbf{131} \pm 18$	294 ± 43	181 ± 77	201 ± 43	198 ± 74	244 ± 66
	16	$\mathbf{125} \pm 17$	289 ± 37	242 ± 70	148 ± 54	157 ± 46	233 ± 66
	32	114 ± 21	317 ± 43	196 ± 54	142 ± 54	$\mathbf{81.1} \pm 21$	209 ± 58
Rastrigin-20	2	$\mathbf{233} \pm 20$	246 ± 24	249 ± 19	261 ± 21	259 ± 16	249 ± 29
	4	$\mathbf{229} \pm 16$	252 ± 16	254 ± 22	257 ± 20	244 ± 20	250 ± 23
	8	$\mathbf{220} \pm 18$	247 ± 22	255 ± 19	252 ± 24	229 ± 31	244 ± 13
	16	$\mathbf{224} \pm 19$	244 ± 17	245 ± 23	236 ± 14	242 ± 16	241 ± 18
	32	$\mathbf{212} \pm 15$	239 ± 14	237 ± 21	232 ± 17	229 ± 19	236 ± 17
Rastrigin-50	2	$\mathbf{613} \pm 24$	672 ± 36	690 ± 62	692 ± 46	653 ± 35	662 ± 58
	4	$\mathbf{573} \pm 34$	674 ± 35	649 ± 62	656 ± 46	619 ± 54	622 ± 39
	8	$\mathbf{555} \pm 35$	666 ± 54	649 ± 58	614 ± 34	584 ± 85	596 ± 58
	16	$\mathbf{540} \pm 19$	617 ± 50	612 ± 43	574 ± 37	563 ± 54	574 ± 66
	32	517 ± 26	536 ± 46	599 ± 58	576 ± 39	$\mathbf{483} \pm 43$	611 ± 34
Ackley-20	2	$\mathbf{3.56} \pm 0.36$	5.33 ± 1	7.83 ± 1	7.45 ± 1.3	6.34 ± 1.2	6.79 ± 1.2
	4	$\mathbf{3.79} \pm 0.29$	6.16 ± 0.89	8.51 ± 1	7.93 ± 0.93	6.98 ± 1.3	7.25 ± 1.1
	8	$\mathbf{4.39} \pm 0.46$	7.02 ± 1.1	8.69 ± 0.93	8.78 ± 1.3	6.18 ± 1.5	8.45 ± 1
	16	$\mathbf{5.38} \pm 0.38$	7.49 ± 1	8.71 ± 1.2	9.00 ± 1.5	5.75 ± 1.1	8.54 ± 1.4
	32	$\mathbf{5.69} \pm 0.58$	8.13 ± 0.81	8.21 ± 1.1	8.09 ± 1.4	6.01 ± 1.1	8.23 ± 1.3
Ackley-50	2	4.16 ± 0.54	$\mathbf{4.13} \pm 0.26$	8.42 ± 0.97	7.10 ± 1.1	5.90 ± 1.2	6.20 ± 0.7
	4	$\mathbf{4.53} \pm 0.5$	4.57 ± 0.43	9.13 ± 0.81	7.81 ± 0.85	6.80 ± 0.93	6.98 ± 1.3
	8	$\mathbf{4.83} \pm 0.58$	5.17 ± 0.33	9.11 ± 0.93	9.39 ± 0.93	7.59 ± 0.62	8.25 ± 1.1
	16	$\mathbf{4.85} \pm 0.46$	5.66 ± 0.66	9.26 ± 1.2	9.07 ± 0.81	6.73 ± 0.62	8.58 ± 1
	32	$\mathbf{5.07} \pm 0.28$	7.80 ± 0.5	9.29 ± 1.1	8.91 ± 0.7	7.69 ± 1.1	8.99 ± 0.74

Table 8: Inner optimization time per round (seconds) across batch sizes under batch-UCB with a fixed total evaluation budget N. Values are averaged over 15 seeds and all rounds for the joint solvers and greedy GP-BUCB with compute-matched and unlimited budgets.

d	q	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)
20	2	0.13	0.11	0.11	0.22	0.06	0.55
20	4	0.19	0.16	0.17	0.41	0.15	1.62
20	8	0.33	0.30	0.30	0.78	0.29	4.85
20	16	0.70	0.65	0.65	1.88	0.63	17.4
20	32	1.61	1.55	1.55	4.50	1.17	59.2
50	2	0.28	0.26	0.26	0.54	0.17	1.54
50	4	0.51	0.47	0.48	1.18	0.39	5.15
50	8	1.03	0.98	0.98	2.76	0.84	19.2
50	16	2.73	2.65	2.64	7.52	2.07	84.5
50	32	8.12	7.96	7.98	22.2	5.51	419

C Experimental setup

C.1 Protocol

Implementation. The implementation is NumPy/SciPy with analytic gradients throughout, finite-difference-checked in the test suite (GP posterior mean, standard deviation and joint covariance under the RBF and Matérn-$5/2$ kernels; MC-qEI and qLogEI via forward-mode Cholesky sensitivity; the chained batch-UCB). The chained variances of (2) are computed as Schur complements $v_i = \Sigma_{ii} - \mathbf{\Sigma}_{i,<i}(\mathbf{\Sigma}_{<i,<i} + \sigma_n^2\mathbf{I})^{-1}\mathbf{\Sigma}_{<i,i}$ of the joint posterior covariance $\mathbf{\Sigma}_t(\mathbf{X})$, with the following derivatives. Let $c_t(\mathbf{x}, \mathbf{y})$ be the latent posterior covariance, $\mathbf{w}_i = (\mathbf{\Sigma}_{<i,<i} + \sigma_n^2\mathbf{I})^{-1}\mathbf{\Sigma}_{<i,i}$, and

$$\mathbf{s}_{i,m} = \nabla_1 c_t(\mathbf{x}_m, \mathbf{x}_i) - \sum_{j<i} w_{i,j}\nabla_1 c_t(\mathbf{x}_m, \mathbf{x}_j), \qquad m \le i.$$

Then

$$\nabla_{\mathbf{x}_m} v_i = \begin{cases} 2\mathbf{s}_{i,i}, & m = i, \\ -2w_{i,m}\mathbf{s}_{i,m}, & m < i, \\ 0, & m > i. \end{cases}$$

Here ∇_1 differentiates the first covariance argument, the prefix is empty for $i = 1$, and the row index m in $\mathbf{s}_{i,m}$ is essential. Differentiating $v_i = \Sigma_{ii} - \mathbf{b}_i^\mathsf{T}\mathbf{A}_i^{-1}\mathbf{b}_i$ (with $\mathbf{A}_i = \mathbf{\Sigma}_{<i,<i} + \sigma_n^2\mathbf{I}$ and $\mathbf{b}_i = \mathbf{\Sigma}_{<i,i}$) gives $dv_i = d\Sigma_{ii} - 2\mathbf{w}_i^\mathsf{T} d\mathbf{b}_i + \mathbf{w}_i^\mathsf{T}(d\mathbf{A}_i)\mathbf{w}_i$, and the changes to the row and column of $\mathbf{A}_i$ yield the prefix formula above. The hallucination noise equals the GP noise, floored at 10^{-8}. qLogEI uses the log-space smoothing of Ament et al. (2023) with temperatures 10^{-3} (per-sample improvement) and 10^{-2} (maximum over rows) and 128 fixed base samples. MC-qEI resamples 64 base samples per gradient. The polar update is computed from the eigendecomposition of the Gram matrix $\mathbf{G}\mathbf{G}^\mathsf{T}$, with diagonal regularization $10^{-10}\mathbf{I}$. The SVD-based reference implementation retains singular modes satisfying $\sigma_i > 10^{-12}\sigma_{\max}$. Newton–Schulz iterations were used only in a check, not in any reported run.

Surrogate. We use a Matérn-$5/2$ GP with ARD length scales and a zero prior mean on standardized targets. On the benchmark objectives, which are evaluated without noise, the surrogate noise variance is fixed at 10^{-6} and is not fitted. Hyperparameters are fitted by maximizing the marginal likelihood with two restarts. Each length scale is initialized to half the observed data range in its coordinate, with a minimum value of 0.1. We refit hyperparameters every round at $d \in \{10, 20\}$, every third round at $d = 50$ and on Rover, and every eighth round at $d = 100$, retaining the previous hyperparameters between refits. For the GP-draw experiments in Figures 1, 5, and 2, the kernel amplitude, length scales, and noise variance remain fixed at their true values.

Objectives, budgets, seeds. Each objective used in the paper is optimized on its own standard box constraint. $[-10, 10]^d$ for Levy, $[-5.12, 5.12]^d$ for Rastrigin, $[-5, 10]^d$ for Ackley, and $[0, 1]^d$ for the GP draws and Rover. The budgets, written as (total evaluations, initial design, refit period),

are $(100, 10, 1)$ at $d = 10$, $(160, 20, 1)$ at $d = 20$, $(400, 50, 3)$ at $d = 50$, $(600, 80, 8)$ at $d = 100$ and $(600, 50, 3)$ on Rover-60; GP draws use $N = 200/320/400/600$ at $d = 10/20/50/100$ with $n_{\text{init}} = 2d + 2$. The initial design is a scrambled Sobol sequence shared across methods for each seed (enforced by a test). The seed counts are 20 at $d \leq 50$ and 10 at $d = 100$ for the main cells, 50 draws for the GP-draw regret curves (25 at $d = 50$), 30 draws per d for the T_{in} study, 15 seeds for the q sweep, and 20 for Rover and the momentum study; the curvature snapshots come from 3 runs per cell and the frozen-surface rate study from 6 surfaces per d. GP draws are sampled from the Matérn-$5/2$ prior with $\ell = 0.2\sqrt{d/10}$ (so that $\ell/\sqrt{d}$ is constant across d) through 2^{14} random Fourier features, with observation noise variance 10^{-3}. The default exploration setting is $\beta_t \equiv 4$ throughout. A sensitivity check at $\beta_t \in \{1, 2\}$ and with the scaled schedule of Desautels et al. (2014) on the $d = 10$ GP-draw cell changes the final R_N of Muon and of matched greedy by at most 3%.

Inner solvers. Muon, projected GD and Adam use $R = 8$ restarts drawn uniformly from the box (identical across methods for a given seed) and $T_{\text{in}} = 60$ steps with cosine decay of the step. Adam uses $\beta_1 = 0.9$, $\beta_2 = 0.999$ with bias correction. Projected GD is $\mathbf{X} \leftarrow \Pi_\Omega(\mathbf{X} + \eta\mathbf{G})$. L-BFGS-B follows the BoTorch `optimize_acqf` recipe: 512 raw candidate batches, half of them drawn around the incumbents (`sample_around_best`), 10 restarts selected by a Boltzmann rule ($\eta = 2$), and at most 200 iterations per restart with the SciPy L-BFGS-B line search. Greedy GP-BUCB solves the q hallucinated single-point problems by multistart L-BFGS-B, either unlimited (6 restarts of 100 iterations per row) or compute-matched, in which the joint solvers' $R \cdot T_{\text{in}}$ gradient evaluations are split q ways as an iteration cap per row. Note that the matched variant still makes around 2 times as many gradient calls as the fixed-step solvers due to the line search. Sequential-greedy qLogEI uses BoTorch's `sequential=True` construction with the same L-BFGS-B budgets. The random screen scores 1024 uniformly random batches. On Rover the fixed-step solvers use 10×200 and start from BoTorch's start batches (Section 7.2). Every solver returns the iterate with the best exactly evaluated acquisition value.

C.2 Step sizes

Each fixed-step method uses an initial step size $\eta = s\,\bar{w}/\sqrt{d}$ with cosine decay, where $\bar{w}$ is the average box width and s is a tuned constant. For the synthetic benchmarks, we select s using the acquisition value attained on four frozen acquisition surfaces per configuration. These surfaces are constructed by refitting GP posteriors to data from completed runs. All methods use the same starting batches and the inner optimization budget of the corresponding experiment, and s is selected from a grid search. For Rover, the constants are tuned using the shared starting batches and an inner budget of 10 restarts with 200 iterations each. Table 9 lists the constants used in the reported experiments.

Table 9: Tuned step constants s for the reported experiments, using $\eta = s\,\bar{w}/\sqrt{d}$ with cosine decay. The left and right column groups correspond to batch-UCB and qLogEI, respectively. Rover experiments use batch-UCB only.

		Batch-UCB			qLogEI		
Benchmark	d	Muon	Proj. GD	Adam	Muon	Proj. GD	Adam
Levy	10	0.5	1	0.25	0.5	1	0.25
Levy	20	0.2	0.5	0.25	0.5	0.5	0.25
Levy	50	0.2	0.02	0.5	1	4	0.5
Levy	100	1	0.03	0.2	1	0.2	0.2
Rastrigin	10	0.5	1	0.25	0.5	1	0.25
Rastrigin	20	0.2	0.5	0.25	0.5	0.5	0.25
Rastrigin	50	0.2	0.2	0.5	1	4	0.5
Rastrigin	100	1	0.02	0.5	1	0.5	0.2
Ackley	10	0.5	1	0.25	0.5	1	0.25
Ackley	20	0.2	0.5	0.25	0.5	0.5	0.25
Ackley	50	1	4	0.5	1	4	0.5
Ackley	100	1	0.2	0.2	1	0.2	0.2
Rover	60	0.5	1	4	–	–	–

C.3 COMPUTE

All experiments were executed on Linux virtual machines with 8 CPU cores. Tables 10 and 11 report the mean time spent on inner optimization per run, in seconds, for the main experiments. At $d \geq 50$, GP refitting accounts for most of the remaining runtime. The joint fixed-step solvers have similar computational costs, differing by at most 5%. Under batch-UCB at $d \in \{50, 100\}$, L-BFGS-B with BoTorch's configuration takes approximately 2.6–2.8 times as long as the joint fixed-step solvers (relative to Muon), while unlimited greedy GP-BUCB takes approximately 15–18 times as long (relative to Muon).

Table 10: Mean inner optimization time per run (seconds) under batch-UCB with $q = 8$, averaged across the three benchmarks and all seeds.

d	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)	Random screen
10	3	2	2	6	2	26	2
20	5	5	5	14	5	83	5
50	42	40	40	117	34	763	41
100	195	188	187	517	120	2947	192

Table 11: Mean inner optimization time per run (seconds) under qLogEI with $q = 8$, averaged across the three benchmarks and all seeds. "Greedy" denotes sequential-greedy qLogEI construction.

d	Muon	Proj. GD	Adam	L-BFGS-B	Greedy (matched)	Greedy (unlimited)	Random screen
10	8	8	8	30	5	51	3
20	23	22	22	87	10	154	6
50	149	146	146	559	30	430	43
100	870	862	864	3172	158	1932	201