Supplementary Information
Supplementary Note: Theoretical Analysis of Domain-Adaptive Cuckoo Search Estimators

Elvis Han Cui, Heather Xihe Yu, and Weng Kee Wong

Related manuscript: Domain-Adaptive Cuckoo Search for Structured Statistical Optimization: Applications to Principal Curves, Single-Cell Trajectory Inference, and Exact Experimental Design.

Summary. This Supplementary Note gives conditional theoretical guarantees for the domain-adaptive Cuckoo Search (CS) estimators considered in the main manuscript. The results cover global convergence in objective value, a finite-iteration probability bound, consistency under vanishing optimization error, and asymptotic normality for smooth continuous-parameter submodels. The arguments combine standard random-search hitting-probability ideas [7], the CS and Lévy-flight proposal mechanisms [10, 4], and classical M-estimation theory [5, 8].

Domain-Adaptive Cuckoo Search as a Random Search Estimator

The three algorithms in the main manuscript can be viewed as instances of a common domain-adaptive CS framework. Let Θ be the feasible parameter space and let $M_n : \Theta \to \mathbb{R}$ be the empirical objective to be minimized. For objective functions originally written as maximization criteria, such as D-optimal exact design, one may set $M_n = -\Phi_n$ without changing the analysis.

At iteration t, the algorithm stores a population of nests

$$\mathcal{X}_t = \{\theta_{1,t}, \ldots, \theta_{N_p,t}\} \subset \Theta$$

and an incumbent best solution

$$\widehat{\theta}_{n,t} \in \underset{\theta \in \mathcal{X}_0 \cup \cdots \cup \mathcal{X}_t}{\arg\min} M_n(\theta).$$

Candidate moves are generated from a mixture of two mechanisms: a Lévy-flight proposal and an abandonment or replacement proposal, following the basic CS construction of Yang and Deb [10]. In implementation, Lévy-stable proposals can be generated by Mantegna's algorithm [4]. The domain-adaptive variants in the main text modify the proposal by projection or discrete-coordinate updates, for example integrality projection in exact design or integer jumps for the single-cell Generalized Trend Model (scGTM) dispersion parameter. We write the resulting proposal kernel as

$$K_t(\theta, A), \qquad \theta \in \Theta, \quad A \subseteq \Theta.$$

The subscript t allows adaptive tuning of step sizes, discovery probabilities, projections, or mixed-domain updates.

Let $\mathcal{F}_t$ denote the sigma-field generated by all random variables used by the algorithm up to the end of iteration t. The incumbent $\widehat{\theta}_{n,t}$ is therefore $\mathcal{F}_t$-measurable. This notation is useful because the proposal distribution may depend on the past through adaptive step sizes, projection maps, or abandonment probabilities. The convergence statements below only require lower bounds on conditional hitting probabilities given $\mathcal{F}_t$; they do not require the sequence of proposals to be independent or identically distributed.

When an empirical objective M_n is being optimized, there are two sources of error:

$$M(\widehat{\theta}_{n,T}) - M(\theta_0) = \{M(\widehat{\theta}_{n,T}) - M_n(\widehat{\theta}_{n,T})\} + \{M_n(\widehat{\theta}_{n,T}) - \inf_{\theta\in\Theta} M_n(\theta)\} + \{\inf_{\theta\in\Theta} M_n(\theta) - M(\theta_0)\}.$$

The middle term is the algorithmic optimization error, while the first and third terms are controlled by empirical-process approximation. This decomposition explains why the statistical guarantees below require both uniform convergence of M_n to M and a sufficiently small optimization error, paralleling the usual separation between statistical approximation and numerical optimization in extremum-estimator theory [5, 8].

The theory below deliberately separates two questions. First, does the stochastic search find arbitrarily good objective values as the number of iterations grows? Second, if the empirical objective is a statistical criterion, when does the computed minimizer inherit standard statistical properties such as consistency or asymptotic normality? The first question is algorithmic; the second is an M-estimation question with an additional optimization-error term.

Global Convergence in Objective Value

Supplementary Assumption 1 (Feasible space and objective)**.** *The feasible set Θ is a compact metric space and $M : \Theta \to \mathbb{R}$ is continuous. The set of global minimizers $\Theta^\star = \arg\min_{\theta\in\Theta} M(\theta)$ is nonempty.*

Supplementary Assumption 2 (Uniform reachability of near-optimal regions)**.** *For every $\epsilon > 0$, define the ϵ-optimal region*

$$A_\epsilon = \{\theta \in \Theta : M(\theta) \le M^\star + \epsilon\}, \qquad M^\star = \min_{\theta\in\Theta} M(\theta).$$

There exists $p_\epsilon > 0$ such that, at every iteration and for every current population state, the probability that at least one generated candidate lies in A_ϵ is at least p_ϵ.

Assumption 2 is a minorization condition on the adaptive proposal mechanism. It is the condition that turns a stochastic heuristic into a probabilistic global-search argument, and is closely related to classical random-search convergence conditions [7]. It holds, for example, if the combined Lévy-flight and abandonment proposal has a density that is bounded below on every open subset of Θ after projection onto the feasible set. In mixed continuous-discrete problems, it is sufficient that the continuous proposal has this local support property within each discrete stratum and that the integer-update mechanism can reach every admissible discrete value with positive probability in finitely many steps.

Supplementary Assumption 3 (Elitist retention)**.** *The incumbent best objective value is never discarded:*

$$M(\widehat{\theta}_t) \le \min_{0\le s<t} M(\widehat{\theta}_s)$$

for all $t \ge 1$.

Supplementary Theorem 1 (Almost-sure global convergence in objective value)**.** *Under Assumptions 1–3,*

$$M(\widehat{\theta}_t) \longrightarrow M^\star \qquad \textit{almost surely as } t \to \infty.$$

Proof. Fix $\epsilon > 0$ and let E_t^ϵ be the event that at least one candidate generated at iteration t lies in A_ϵ. By Assumption 2,

$$\mathbb{P}(E_t^\epsilon \mid \mathcal{F}_{t-1}) \ge p_\epsilon$$

for every possible past history. Hence, conditional on not having reached A_ϵ before iteration t, the probability of missing it again is at most $1 - p_\epsilon$. Iterating this conditional bound gives

$$\mathbb{P}\left(\bigcap_{t=1}^{T} (E_t^\epsilon)^c\right) \le (1 - p_\epsilon)^T.$$

Let $\tau_\epsilon = \inf\{t : E_t^\epsilon \text{ occurs}\}$. The preceding display implies

$$\mathbb{P}(\tau_\epsilon > T) \le (1 - p_\epsilon)^T \to 0,$$

so $\tau_\epsilon < \infty$ almost surely. Once a point in A_ϵ has been generated, elitist retention implies

$$M(\widehat{\theta}_t) \le M^\star + \epsilon \qquad \text{for all sufficiently large } t.$$

Because $M(\widehat{\theta}_t) \ge M^\star$ always, we obtain

$$0 \le \limsup_{t\to\infty}\{M(\widehat{\theta}_t) - M^\star\} \le \epsilon$$

almost surely. Taking ϵ along a countable sequence such as $\epsilon = 1/k$ proves $M(\widehat{\theta}_t) \to M^\star$ almost surely. □

Supplementary Corollary 1 (Finite-iteration probability bound)**.** *Under the same assumptions, for any $\epsilon > 0$ and integer $T \ge 1$,*

$$\mathbb{P}\{M(\widehat{\theta}_T) - M^\star > \epsilon\} \le (1 - p_\epsilon)^T \le \exp(-p_\epsilon T).$$

This bound is intentionally problem-dependent: the constant p_ϵ encodes the size and accessibility of the ϵ-optimal region. It does not claim a universal rate for all nonconvex objectives. If, locally, $M(\theta) - M^\star$ behaves quadratically in d effective continuous dimensions and the proposal density is bounded below around $\Theta^\star$, then the volume of A_ϵ is typically of order $\epsilon^{d/2}$. In that case $p_\epsilon \gtrsim C\epsilon^{d/2}$ and Corollary 1 implies the stochastic order

$$M(\widehat{\theta}_T) - M^\star = O_p(T^{-2/d})$$

up to constants and logarithmic factors. This is a local volume-search rate, not a gradient-descent rate.

Supplementary Corollary 2 (Sufficient iteration schedule for vanishing optimization error)**.** *For an empirical criterion M_n, define*

$$M_n^\star = \inf_{\theta\in\Theta} M_n(\theta), \qquad A_{n,\epsilon} = \{\theta \in \Theta : M_n(\theta) \le M_n^\star + \epsilon\}.$$

Suppose the reachability condition holds for M_n with deterministic lower bounds $p_{n,\epsilon} > 0$, so that each iteration hits $A_{n,\epsilon}$ with conditional probability at least $p_{n,\epsilon}$. If $T_n p_{n,\epsilon} \to \infty$ for every fixed $\epsilon > 0$, then

$$M_n(\widehat{\theta}_{n,T_n}) - M_n^\star = o_p(1).$$

More sharply, if $\epsilon_n \downarrow 0$ and $T_n p_{n,\epsilon_n} \to \infty$, then

$$M_n(\widehat{\theta}_{n,T_n}) - M_n^\star = O_p(\epsilon_n).$$

Proof. Applying Corollary 1 to the empirical objective gives

$$\mathbb{P}\{M_n(\widehat{\theta}_{n,T_n}) - M_n^\star > \epsilon\} \le \exp(-T_n p_{n,\epsilon}).$$

For fixed $\epsilon > 0$, the right-hand side tends to zero under $T_n p_{n,\epsilon} \to \infty$, which proves the $o_p(1)$ statement. The same argument with $\epsilon = \epsilon_n$ gives the sharper bound. □

Corollary 2 provides one concrete route to Assumption 6: the number of CS iterations must grow fast enough relative to the accessibility of near-optimal regions. In practice, $p_{n,\epsilon}$ is rarely known, so the result should be viewed as a theoretical calibration rather than a tuning rule.

Consistency of CS-Based Estimators

We now treat the algorithm output as an approximate minimizer of an empirical statistical criterion. This is the standard extremum-estimator or M-estimator viewpoint, with an additional numerical optimization error [5, 8]. Let $M_n : \Theta \to \mathbb{R}$ be an empirical objective and let $M : \Theta \to \mathbb{R}$ be its population target. Let θ_0 be the target parameter, defined as the unique minimizer of M.

Supplementary Assumption 4 (Uniform law of large numbers)**.**

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{p} 0.$$

Supplementary Assumption 5 (Identification)**.** *For every $\delta > 0$,*

$$\inf_{\theta : d(\theta, \theta_0) \geq \delta} \{M(\theta) - M(\theta_0)\} > 0.$$

Supplementary Assumption 6 (Vanishing optimization error)**.** *The CS output after $T = T_n$ iterations satisfies*

$$M_n(\widehat{\theta}_{n,T}) \leq \inf_{\theta \in \Theta} M_n(\theta) + r_{n,T}, \qquad r_{n,T} = o_p(1).$$

Supplementary Theorem 2 (Consistency under vanishing optimization error)**.** *Under Assumptions 1, 4, 5, and 6,*

$$\widehat{\theta}_{n,T} \xrightarrow{p} \theta_0.$$

Proof. Let

$$\Delta_n = \sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)|.$$

For any $\theta \in \Theta$,

$$M(\widehat{\theta}_{n,T}) \leq M_n(\widehat{\theta}_{n,T}) + \Delta_n.$$

By the approximate optimality condition,

$$M_n(\widehat{\theta}_{n,T}) \leq M_n(\theta_0) + r_{n,T} \leq M(\theta_0) + r_{n,T} + \Delta_n.$$

Thus

$$M(\widehat{\theta}_{n,T}) - M(\theta_0) \leq 2\Delta_n + r_{n,T} = o_p(1).$$

For any fixed $\delta > 0$, Assumption 5 gives a separation constant

$$\eta_\delta = \inf_{\theta : d(\theta, \theta_0) \geq \delta} \{M(\theta) - M(\theta_0)\} > 0.$$

Therefore

$$\mathbb{P}\{d(\widehat{\theta}_{n,T}, \theta_0) \geq \delta\} \leq \mathbb{P}\{M(\widehat{\theta}_{n,T}) - M(\theta_0) \geq \eta_\delta\} \to 0,$$

which proves $\widehat{\theta}_{n,T} \to_p \theta_0$. □

Supplementary Corollary 3 (Set-valued consistency)**.** *If the unique minimizer θ_0 is replaced by a nonempty compact minimizer set*

$$\Theta_0 = \arg\min_{\theta \in \Theta} M(\theta),$$

and if for every $\delta > 0$,

$$\inf_{\theta : d(\theta, \Theta_0) \geq \delta} \{M(\theta) - \inf_{\vartheta \in \Theta} M(\vartheta)\} > 0,$$

then the same conditions imply

$$d(\widehat{\theta}_{n,T}, \Theta_0) \xrightarrow{p} 0.$$

Proof. The proof is identical to Theorem 2, with the separation condition applied outside the δ-neighborhood of Θ_0. This formulation is useful when the statistical target is identifiable only up to curve orientation, monotone reparameterization, label conventions, or nonunique exact-design optima. □

The theorem applies directly to continuous-parameter submodels. For mixed continuous-discrete models such as scGTM with an integer dispersion parameter, consistency can be stated on the product space $\Theta_c \times \Theta_d$ when the admissible integer set Θ_d is finite or effectively truncated. For exact designs with integer allocations and a fixed sample size N, the target is a finite-dimensional discrete-continuous optimum; consistency should be interpreted as convergence to the set of population-optimal exact designs rather than convergence to a regular Euclidean parameter.

Smooth Submodels

Asymptotic normality is a local property and should not be attributed to the Cuckoo Search mechanism itself. It can be inherited by CS estimators when the empirical objective is smooth, the population optimum is a unique interior point, and the optimization error is negligible relative to the statistical error, as in standard differentiable M-estimation theory [5, 8].

Supplementary Assumption 7 (Local smoothness and nonsingularity)**.** *The parameter θ_0 is an interior point of an open subset of $\mathbb{R}^q$. In a neighborhood of θ_0, M_n is twice continuously differentiable with gradient ∇M_n and Hessian $\nabla^2 M_n$. Moreover,*

$$H = \nabla^2 M(\theta_0)$$

exists and is positive definite.

Supplementary Assumption 8 (Score central limit theorem and Hessian convergence)**.**

$$\sqrt{n}\nabla M_n(\theta_0) \Rightarrow N(0, \Sigma), \qquad \sup_{\|\theta-\theta_0\| \le \delta_n} \|\nabla^2 M_n(\theta) - H\| \xrightarrow{p} 0$$

for every deterministic sequence $\delta_n \downarrow 0$.

Supplementary Assumption 9 (First-order optimization accuracy)**.** *The CS output is first-order accurate at the $n^{-1/2}$ scale:*

$$\|\nabla M_n(\widehat{\theta}_{n,T})\| = o_p(n^{-1/2}).$$

Supplementary Remark 1 (Value accuracy versus first-order accuracy)**.** *The value-level requirement $M_n(\widehat{\theta}_{n,T}) - \inf_\Theta M_n = o_p(1)$ is sufficient for consistency, but it is not sufficient for asymptotic normality. Normality is a local expansion result at the $n^{-1/2}$ scale and therefore requires a stronger optimization condition. Assumption 9 can be replaced by the more classical condition that the algorithmic error relative to the exact empirical minimizer is $o_p(n^{-1/2})$. Under local strong convexity, a sufficient value-level condition is*

$$M_n(\widehat{\theta}_{n,T}) - \inf_{\theta\in\Theta} M_n(\theta) = o_p(n^{-1}),$$

which implies $n^{-1/2}$-scale proximity to the exact minimizer. This stronger accuracy is typically relevant only when the final CS solution is polished locally or the search budget increases with n.

Supplementary Theorem 3 (Asymptotic normality for smooth submodels)**.** *Assume Theorem 2 and Assumptions 7–9. Then*

$$\sqrt{n}(\widehat{\theta}_{n,T} - \theta_0) \Rightarrow N\left(0, H^{-1}\Sigma H^{-1}\right).$$

Proof. By Taylor expansion around θ_0,

$$\nabla M_n(\widehat{\theta}_{n,T}) = \nabla M_n(\theta_0) + \nabla^2 M_n(\widetilde{\theta}_n)(\widehat{\theta}_{n,T} - \theta_0),$$

where $\widetilde{\theta}_n$ lies between $\widehat{\theta}_{n,T}$ and θ_0. Consistency and Hessian convergence imply $\nabla^2 M_n(\widetilde{\theta}_n) = H + o_p(1)$. Rearranging and using Assumption 9,

$$\sqrt{n}(\widehat{\theta}_{n,T} - \theta_0) = -H^{-1}\sqrt{n}\nabla M_n(\theta_0) + o_p(1).$$

The result follows from the score central limit theorem. □

Supplementary Remark 2 (Scope of the normality result)**.** *Theorem 3 is appropriate for smooth continuous-parameter components, such as a differentiable likelihood submodel or a smooth local approximation to a regularized projection objective. It does not apply without modification to integer design allocations, nonsmooth projection maps, boundary optima, nonunique curves, Hausdorff-loss objectives, or discrete dispersion parameters. In such cases, consistency or set convergence is the safer asymptotic statement.*

Implications for the Three Applications

Cuckoo Search for principal-curve estimation (Cuckoo-PC). For principal-curve estimation, the relevant parameter includes projection indices and smoother coefficients or fitted functions. Principal curves originate from the self-consistency formulation of Hastie and Stuetzle [2], while the saddle-point pathology of the classical distance functional is described by Duchamp and Stuetzle [1]. When the smoother is restricted to a finite-dimensional sieve or a compact smoothness class and the metric discrepancy is continuous, the consistency theorem applies to the regularized population target, provided that this target is identifiable up to orientation or monotone reparameterization. If several parameterizations represent the same geometric curve, Corollary 3 is the natural formulation. The asymptotic normality result can apply only to a locally regular finite-dimensional approximation. For the fully nonparametric curve, a function-space consistency statement, for example under an integrated squared-error or Hausdorff-type metric after alignment, is more appropriate than ordinary Euclidean normality.

Mixed-domain scGTM optimization. The scGTM objective is a likelihood-type criterion with continuous kinetic parameters and a positive-integer dispersion parameter, motivated by single-cell trajectory modeling for the RNA-seq data analyzed in Wang et al. [9]. If the integer parameter is restricted to a finite admissible set and the likelihood identifies a unique mixed-domain parameter, the CS output is consistent under the vanishing optimization-error condition. Conditional on selecting the correct integer value with probability tending to one, the continuous parameters can satisfy the asymptotic normality theorem under the usual local likelihood regularity conditions.

Exact experimental design. For exact design, the optimization target is a finite-sample design object containing support points and integer allocations. This differs from the approximate-design framework, where equivalence theorems and convex design theory provide powerful optimality checks [3, 6]. The CS convergence theorem gives convergence in the design criterion under reachability and elitist retention. When the total sample size N is fixed, classical asymptotic normality of an estimator is not the natural target; the appropriate statement is convergence to the set of criterion-optimal exact designs. The set-valued formulation is important because two designs may be equivalent under permutation of support points or may attain the same determinant value. If N grows and one embeds exact designs in approximate-design space, further asymptotic analysis would require a separate increasing-sample design framework that controls both allocation discretization and support-point approximation.

Practical Interpretation

The theoretical guarantees should be read as conditional results. Adaptive Cuckoo Search is a stochastic global search procedure; its convergence speed depends on the geometry of the objective landscape and the proposal probability of reaching near-optimal regions. The strongest general statement is almost-sure convergence in objective value under compactness, reachability, and elitist retention. Statistical consistency follows when empirical objectives converge uniformly and the algorithmic error vanishes. Asymptotic normality is available only for smooth continuous-parameter submodels and only when the optimization error is smaller than the $n^{-1/2}$ statistical scale.

References

[1] Duchamp, T. and Stuetzle, W. (1996). Extremal properties of principal curves in the plane. *The Annals of Statistics*, 24(4), 1511–1520.

[2] Hastie, T. and Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406), 502–516.

[3] Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B*, 21(2), 272–319.

[4] Mantegna, R. N. (1994). Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. *Physical Review E*, 49(5), 4677–4683.

[5] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Elsevier.

[6] Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM.

[7] Solis, F. J. and Wets, R. J.-B. (1981). Minimization by random search techniques. *Mathematics of Operations Research*, 6(1), 19–30.

[8] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[9] Wang, W., Vilella, F., Alama, P., Moreno, I., Mignardi, M., Isakova, A., Pan, W., Simon, C., and Quake, S. R. (2020). Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. *Nature Medicine*, 26(10), 1644–1653.

[10] Yang, X.-S. and Deb, S. (2009). Cuckoo search via Lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing*, pp. 210–214. IEEE.