The invariance defect of continuous-trait parametric replicator dynamics

Eric Bonabeau

School of Complex Adaptive Systems, Arizona State University,
Tempe, AZ 85281-2701, USA

ebonabea@asu.edu

Abstract

Parameterizing the distribution of traits in the continuous-trait replicator equation yields the Fisher-metric projection of the full density dynamics onto the statistical manifold of the chosen family. The projected parameter equation has the form used in information-geometric optimization (IGO), but it reproduces the full replicator trajectory exactly only when centered fitness lies in the score span of the family. This paper defines the *invariance defect* as the norm of the instantaneous difference between the full replicator relative velocity and its Fisher projection at the same density. For distribution-independent fitness, the squared defect is the shortfall in the instantaneous rate of increase of mean fitness relative to the unrestricted replicator flow, so Fisher's fundamental theorem splits into a part the family can express and a part it cannot. The approach is extended to quantile-based fitness shaping, with a distinction between fixed statistical families and adaptive tangent dictionaries. The analysis characterizes exact invariance, local moment and density agreement, finite-time continuation, and symmetric Gaussian stalling.

---

1. The replicator equation

The replicator equation (Taylor and Jonker, 1978; Hofbauer and Sigmund, 2003; Nowak, 2006) is a theoretical workhorse in the study of evolutionary dynamics and game theory. Introduced on the basis of simple game-theory considerations, it has revealed over time a rich array of behaviors and deep connections to other models, such as the Price equation (Price, 1970), Fisher's fundamental theorem of natural selection (Fisher, 1930; Price, 1972; Page and Nowak, 2002; Queller, 2017), Nash equilibria and evolutionarily stable strategies (Hofbauer and Sigmund, 2003; Nowak, 2006; Lehtonen, 2018) and, more recently, Bayesian inference (Harper, 2009; Shalizi, 2009; Akyıldız, 2017; Czégel et al., 2019). The equation itself is simple; its complexity resides in how the fitness or payoff function is defined. It has not been strongly connected to evolutionary optimization algorithms, although it has been applied as a heuristic to hard optimization problems (Menon et al., 1995; Bomze et al., 2000; Pelillo, 1999; Rota Bulò et al., 2011) and its discrete form is known in other fields as the multiplicative weights update algorithm (Arora et al., 2012; van der Hoeven et al., 2018).

This paper establishes that the replicator equation is similar to certain evolutionary algorithms that perform information-geometric optimization (IGO) (Ollivier et al., 2017; Wierstra et al., 2014; Miskin et al., 2016; Jaćimović, 2024), but only under conditions that can be stated precisely, and it quantifies the error when those conditions fail.

1.1 Continuous time, discrete traits

Let $y_i(t)$, $i = 1, \dots, n$, be the relative abundance of trait i in a very large population (all results in this paper assume an infinite population, a mathematical approximation to a very large population) at time

$t \in \mathbb{R}_+$, and $y_t = (y_i(t))$ the n-dimensional vector of trait frequencies. In evolutionary game theory, traits are pure strategies and y_t is a mixed strategy. The fitness or payoff $f_i(t) = f_i(y_t)$ of a trait depends on the frequencies of the other traits (frequency-dependent selection) and determines its replication rate. The continuous-time replicator equation is

$$\dot{y}_i = y_i(t) \left[f_i(y_t) - \mathbb{E}_{y_t}[f] \right] \tag{1}$$

where $\dot{y}_i = dy_i/dt$ and $\mathbb{E}_{y_t}[f] \triangleq \sum_j y_j(t) f_j(y_t)$ is the mean fitness. It is assumed that $y_i(0) > 0$ for all i, so the initial state lies in the interior of the simplex $\Delta_{n-1} \triangleq \{y \in \mathbb{R}^n_+ \mid \sum_i y_i = 1\}$. A trait with $y_i(0) = 0$ remains at zero for all t. The normalization

$$\sum_{i=1}^{n} y_i(t) = 1 \tag{2}$$

is preserved by (1) once imposed as an initial condition, and is re-imposed at each step in the discretized dynamics to counter integration and round-off drift.

If the fitness of i decomposes into time-independent pairwise payoffs Π_{ij}, then $f_i(t) = \sum_j y_j(t) \Pi_{ij}$ for a well-mixed population, and $\Pi = (\Pi_{ij})$ is the payoff matrix.

1.2 Continuous time, continuous traits

Consider now a continuous trait space. Let $x \in \mathcal{X} \subseteq \mathbb{R}^d$, assumed a convex subset of $\mathbb{R}^d$ to avoid mathematical precautions inessential to the argument, and let $\Delta(\mathcal{X})$ denote the Borel probability measures on $\mathcal{X}$. Replace $y_i(t)$ with the density $P(x,t) \triangleq P_t(x) \in \Delta(\mathcal{X})$. The fitness of x becomes a fitness density depending on the whole distribution, $f(x \mid P_t)$. Assuming pairwise payoffs $\Pi(x,z)$, $f(x \mid P_t) = \int_{\mathcal{X}} P_t(z) \Pi(x,z)\, dz$, and the mean fitness is

$$\mathbb{E}_{P_t}[f] = \int_{\mathcal{X}} P_t(z) f(z \mid P_t)\, dz = \iint_{\mathcal{X}\times\mathcal{X}} P_t(x) P_t(z) \Pi(x,z)\, dx\, dz. \tag{3}$$

Under the assumption that $P_t(x)$ is differentiable in t and x, Equation (1) becomes

$$\partial_t P_t(x) = P_t(x) \left[f(x \mid P_t) - \mathbb{E}_{P_t}[f] \right], \tag{4}$$

with the normalization $\int_{\mathcal{X}} P_t(z)\, dz = 1$ again preserved by the flow and re-imposed numerically at each step.

The Euclidean and convex-support assumptions simplify the examples, but the projection and tangency arguments in Sections 1.3 and 3 apply to densities on a general measure space. The same formulation therefore covers directional traits when the reference measure is the appropriate surface measure.

1.3 The Fisher–Rao natural gradient

Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $\rho : \mathcal{X} \to \mathbb{R}_+$ a density with $\int \rho\, d\mu = 1$; write $\mathcal{P}_+$ for the strictly positive normalized densities. A tangent vector u at ρ must preserve normalization infinitesimally, $\int u\, d\mu = 0$, and the Fisher–Rao metric on such tangent vectors is

$$g_\rho(u,v) = \int \frac{u(x) v(x)}{\rho(x)}\, d\mu(x), \tag{5}$$

the infinite-dimensional counterpart of the Fisher information matrix. To maximize a functional $J[\rho]$ with variational derivative $\phi_\rho(x) = \delta J / \delta \rho(x)$, the Fisher–Rao natural gradient G_ρ is defined intrinsically by

$g_\rho\left(G_\rho, u\right) = DJ[\rho](u)$ for every zero-mass perturbation u. Since $DJ[\rho](u) = \int \phi_\rho u \, d\mu$, the solution is $G_\rho(x) = \rho(x)\left[\phi_\rho(x) - \mathbb{E}_\rho\left[\phi_\rho\right]\right]$, and the nonparametric Fisher–Rao natural-gradient flow is

$$\partial_t \rho(x) = \rho(x)\left[\frac{\delta J}{\delta \rho}(x) - \mathbb{E}_\rho\left[\frac{\delta J}{\delta \rho}\right]\right]. \tag{6}$$

If $J[\rho] = \int f\rho \, d\mu = \mathbb{E}_\rho[f]$ with f independent of ρ, then $\delta J/\delta\rho = f$ and (6) reduces to $\partial_t \rho = \rho\left(f - \mathbb{E}_\rho[f]\right)$, which is the continuous-trait replicator equation. In that case the replicator equation is Fisher–Rao natural-gradient ascent of expected fitness on the full manifold of densities.

If the fitness depends on the current distribution, the functional $J[\rho] = \int f(x,\rho)\rho(x)\, dx$ has variational derivative

$$\frac{\delta J}{\delta \rho}(x) = f(x,\rho) + \int \rho(z)\frac{\delta f(z,\rho)}{\delta \rho(x)}\, dz, \tag{7}$$

and the replicator field contains only the first term. It remains a valid Fisher–Rao tangent vector field, but it is not in general the gradient of the mean-fitness functional.

Restricting the density to a finite-dimensional statistical family replaces this exact flow by its Fisher projection. There is by definition no defect on all of $\mathcal{P}_+$, since it contains every density. The invariance defect defined in Section 3 measures the component of the unrestricted gradient that a chosen family cannot represent.

1.4 Related work

Shahshahani (1979) established that the continuous-time, discrete-trait replicator equation is an information-geometric flow on the simplex endowed with the Shahshahani–Fisher metric, and a gradient flow under symmetric linear fitness: its dynamics is steepest ascent (natural gradient) of mean fitness with respect to that metric (see also Akin, 1979, 1990; Sigmund, 1984). The Kullback–Leibler divergence acts as a Lyapunov function. The information-geometric reading of replicator dynamics originates there and with Hofbauer and Sigmund (2003), and was developed by Bomze (1991) and Harper (2009). The natural-gradient construction is due to Amari (1998), with the e- and m-projection vocabulary used throughout information geometry set out in Amari and Nagaoka (2000); the least-squares and model-manifold geometry underlying projection was studied by Transtrum et al. (2011); the IGO machinery is that of Ollivier et al. (2017), Wierstra et al. (2014), and Beyer (2014).

Savarino et al. (2024) study coupled replicator equations on an assignment manifold arising in metric data labeling, relating those dynamics to geometric mechanics. Their question concerns the geometry of the specified manifold dynamics; the present analysis instead measures the residual incurred when a full replicator field is represented by a chosen parametric family.

Pathiraja and Wacker (2025) rigorously connect continuous-time Bayesian filtering with replicator-mutator dynamics and discuss associated gradient-flow formulations. Carrillo et al. (2026) develop a broader theory of Fisher-Rao gradient flows driven by divergence functionals. These works place density evolution in Fisher-Rao geometry, but neither studies the residual created by restricting a replicator field to a statistical family manifold. Miskin et al. (2016) anticipated the continuous-trait connection in an algorithmic setting by turning canonical-ensemble statistical-physics models into materials-design optimizers. Their working equation is the least-squares projection of the continuous replicator equation onto the score functions of a parametric family, equivalently its Fisher-metric projection, and therefore implicitly acknowledges that the unrestricted flow generally leaves the family. They do not quantify the systematic restriction error, which is the central subject here.

Karev's reduction method for replicator systems with exponential-family trait distributions (Karev, 2020,

and references therein) solves the continuous-trait replicator equation in closed form when the initial distribution belongs to an exponential family and the fitness is a linear combination of its sufficient statistics. That construction supplies the sufficiency direction of the invariance criterion in Section 3.8 and the one-dimensional Gaussian-quadratic solution used in Section 3.11. The present paper adds the converse (that this fitness class is also necessary), and, more substantially, treats the case where the condition *fails*, which the reduction method does not address. Related exact and asymptotic analyses of continuous-trait selection include Tsimring et al. (1996), Cressman (2005), Cressman and Hofbauer (2005), Cressman et al. (2006), Oechssler and Riedel (2001, 2002), and Smerlak and Youssef (2017).

The same closure question has a long history in quantitative genetics, where continuum-of-alleles models are routinely closed by assuming a Gaussian allelic distribution (Lande, 1979; Lande and Arnold, 1983). Turelli (1984) showed that this assumption can fail badly under mutation-selection balance, launching a debate about when Gaussian closure is admissible that was largely conducted through case analysis (Bürger, 2000). Bürger (1991) also derived cumulant evolution equations for polygenic trait distributions under polynomial selection, linking the Gaussian approximation to cumulant closure. The invariance defect below is a quantitative answer to that question: it is the exact instantaneous cost of the closure, in units of adaptation rate. The projected moment equations derived in Section 3.9 are Lande's equation and its second-moment companion, and the mean equation is Robertson's secondary theorem of natural selection (Robertson, 1966).

On the algorithmic side, the rank-μ core of the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006; Beyer and Sendhoff, 2017) is a discrete, sampled realization of the projected replicator/IGO flow, forcing the dynamics onto the Gaussian manifold; its equivalence with natural-gradient updates was established by Akimoto et al. (2010, 2012) and Ollivier et al. (2017). Viewed as replicator dynamics, the projection is exact only when the fitness landscape is affine-quadratic; otherwise the Gaussian is a best-fit approximation whose model error is the defect defined here. The literature takes the Gaussian projection as given; this paper asks when it is exact and what it costs when it is not. The geometry of stochastic relaxation over exponential manifolds, and the resulting model-selection question for estimation-of-distribution algorithms, was studied by Malagò and Pistone (2015) and Larrañaga and Lozano (2002).

For the multivariate Gaussian under bilinear-quadratic fitness, Jaćimović (2024) showed that the restricted replicator dynamics coincides with a natural-gradient flow whose potential is closely related to mean fitness, recovering the Gaussian moment equations of Cressman et al. (2006). Section 2.2 explains in general when the potential is exactly mean fitness and when it must be modified. Otwinowski et al. (2020) develop an information-geometric account of natural selection in which the rate of adaptation plays a central role, and Mustonen and Lässig (2010) define a fitness flux with the same units as the quantity decomposed in Theorem 3 below; neither addresses the parametric restriction.

Jaćimović (2023) gives information-geometric formulations of Fisher's fundamental theorem in evolutionary games and optimization, emphasizing bounds on evolutionary change. The question considered here is complementary: how much of the instantaneous selection response is removed by projection onto a specified finite-dimensional statistical family?

Raab et al. (2022) prove that Fisher–Rao natural-gradient descent is a local least-squares optimal approximation to the continuous-time replicator equation, with the Fisher information appearing as the Hessian of the least-squares error. Their result is set on a finite action simplex with a fixed policy family. The present paper is an extension of that projection result: for continuous-trait families it characterizes when the residual vanishes, names and measures the residual, gives it an interpretation in terms of adaptation rate, derives a finite-time tracking estimate, and extends the construction to rank-shaped selection. Raab et al. (2022) further show that quantities carried along natural-gradient dynamics obey the Price equation, the discrete-strategy counterpart of the shaped-selection construction developed below.

1.5 Contributions

1. **The invariance defect** (Section 3.4), the $L^2\left(P_\theta\right)$ distance from the centered fitness to the score span, together with a normalized version that is invariant under reparameterization of the family and under affine rescaling of the fitness.
2. **An interpretation of the defect as lost adaptation rate** (Theorem 3, Section 3.5). For distribution-independent fitness, $\delta^2(\theta)$ is the difference between the rate of increase of mean fitness under the full replicator flow and under its Fisher projection. Fisher's fundamental theorem therefore splits into a representable part and a defect. For frequency-dependent fitness, this identity applies to the selection component of mean-fitness change, not generally to its total derivative. Two corollaries follow: the projected flow does not decrease mean fitness, and it stalls at states where the fitness-score covariance vanishes while fitness variance does not.
3. **Local agreement and moment closure (Proposition 4, Section 3.6).** For an exponential family, the Fisher-projected flow is the moment closure obtained by evaluating the true moment equation under the current family member. The sufficient-statistic means agree through their first time derivative, so their error begins at order t^2. The density trajectories agree only at $t = 0$; their leading L^1 separation is $t\,\mathbb{E}|r| + o(t)$, bounded above by $t\delta + o(t)$.
4. **Exact solutions and finite-time validity** (Proposition 5, Section 3.7). For distribution-independent fitness the true trajectory is an exponential (e-) geodesic whose mean fitness and fitness variance are the first two derivatives of the cumulant generating function of f under ρ_0. Blow-up occurs when that function ceases to be finite.
5. **A tangency characterization of exact invariance** (Section 3.3) and its specializations to exponential, Gaussian, location-scale, mixture, copula and flow-generated families (Section 4).
6. **A shaped-invariance theory** (Section 5). Quantile shaping changes the conjugacy class, so families exact for raw linear or quadratic fitness are generally not exact after rank shaping. The shaped defect is invariant under monotone transformations of the objective, which is the property an evolution strategy is designed to have.
7. **Numerical validation** (Section 6 and Supplementary Materials)**.** The experiments separate representational error, trajectory accumulation, finite-sample estimation, and conditioning; test the exact identities and convergence orders; and examine shaped selection, two-dimensional and frequency-dependent cases, dimensional scaling, symmetric stalling, and the looseness of trajectory bounds.

2. The parameterized replicator equation and the IGO flow

2.1 Parameterization

Assume $P_t(x)$ belongs to a parameterized family with parameter $\theta \in \Theta \subseteq \mathbb{R}^m$, and write $P_t(x) = P_{\theta_t}(x)$. Then

$$\partial_t P_{\theta_t}(x) = P_{\theta_t}(x)\left[f\left(x \mid P_{\theta_t}\right) - \mathbb{E}_{P_{\theta_t}}[f]\right]. \tag{8}$$

Equation (8) makes sense only under the assumption that the family $Fam = \{P_\theta,\ \theta \in \Theta\}$ is forward-invariant under Equation (4): if $P_0 \in Fam$ then $P_t \in Fam$ for all $t \geq 0$. When Fam is not forward-invariant, P_{θ_t} may still be interpreted as a projection of the true P_t; the question is how far the two trajectories deviate. Section 3 makes this precise.

Writing $\dot\theta$ for the parameter velocity and differentiating,

$$\dot\theta^\top \nabla_\theta \log P_\theta(x)|_{\theta=\theta_t} = f\left(x \mid P_{\theta_t}\right) - \mathbb{E}_{P_{\theta_t}}[f]. \tag{9}$$

Denote the score by $u_{\theta_t}(x) = \nabla_\theta \log P_\theta(x)|_{\theta=\theta_t}$, so that (9) reads $\dot\theta^\top u_{\theta_t}(x) = f\left(x \mid P_{\theta_t}\right) - \mathbb{E}_{P_{\theta_t}}[f]$. Taking the expectation $\mathbb{E}_{P_{\theta_t}}[\cdot]$ of both sides yields 0, since the score has vanishing mean. Multiplying first by the score and then taking the expectation yields

$$\dot\theta^\top \mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t} u_{\theta_t}^\top\right] = \mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t}\left(f - \mathbb{E}_{P_{\theta_t}}[f]\right)\right], \tag{10}$$

that is, $\dot\theta^\top F(\theta_t) = Cov_{P_{\theta_t}}\left(u_{\theta_t}, f\right)$, where $F(\theta_t) = \mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t} u_{\theta_t}^\top\right]$ is the Fisher information matrix. Assuming F invertible,

$$\dot\theta = F^{-1}(\theta_t)\, \mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t}(x)\, f\left(x \mid P_{\theta_t}\right)\right], \tag{11}$$

which is the IGO flow of Ollivier et al. (2017, their Equation 10) with a quantile rewriting of the fitness.

Under forward invariance, replicator dynamics implies an IGO flow. The converse fails, because the step from (9) to (10) is an irreversible averaging transformation. The best equation recoverable for P_{θ_t} starting from (11) is

$$\partial_t P_{\theta_t}(x) = P_{\theta_t}(x)\, u_{\theta_t}(x)^\top F^{-1}(\theta_t)\, \mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t}\left(f - \mathbb{E}_{P_{\theta_t}}[f]\right)\right], \tag{12}$$

and there are many ways to unbundle the expectation in (12), one of which returns (8). A forward-invariant parameterized replicator equation is therefore one possible Fokker–Planck-like equation consistent with an IGO flow, not the only one.

2.2 Fisher projection, natural-gradient form, and potential ascent

For distribution-independent fitness $f(x)$, let $J(\theta) = \mathbb{E}_{P_\theta}[f(X)]$. Then $\nabla_\theta J = \mathbb{E}_{P_\theta}[u_\theta f]$, and the projected replicator equation is natural-gradient ascent on expected fitness. If the fitness is frequency-dependent, $f = f(\cdot, P_\theta)$, then

$$\nabla_\theta J(\theta) = \mathbb{E}_{P_\theta}[u_\theta(X) f(X, P_\theta)] + \mathbb{E}_{P_\theta}[\nabla_\theta f(X, P_\theta)]. \tag{13}$$

The Fisher projection contains the first term but not the second, so it is not in general gradient ascent on mean fitness, even though it remains the best Fisher-metric representation of the replicator field. Rank-based IGO sharpens the distinction: its quantile weights are recomputed from the current distribution and held fixed while the instantaneous parameter direction is calculated. The resulting flow is coordinate-invariant and rank-invariant, but need not be the gradient flow of one fixed scalar potential. None of the invariance and defect results below requires such a potential; they depend only on tangency and projection.

2.3 Natural evolution strategies and fitness shaping

Natural evolution strategies (NES) (Wierstra et al., 2014; Beyer, 2014) take the form

$$\dot\theta = F^{-1}(\theta_t)\, \nabla_\theta\, \mathbb{E}_{P_{\theta_t}}[f] = F^{-1}(\theta_t)\left(\mathbb{E}_{P_{\theta_t}}\left[u_{\theta_t} f\right] + \mathbb{E}_{P_{\theta_t}}\left[\nabla_\theta f\right]\right). \tag{14}$$

Equation (14) is equivalent to (11) when $f\left(x \mid P_{\theta_t}\right) = f(x)$, which is typical for optimization: the parameterized replicator equation and NES flows coincide when the fitness does not depend on the current distribution.

A divergence between NES and replicator dynamics emerges when the effective fitness is not the function to be optimized but a monotone rewriting of it that depends on P_{θ_t}, called *fitness shaping*. A stochastic

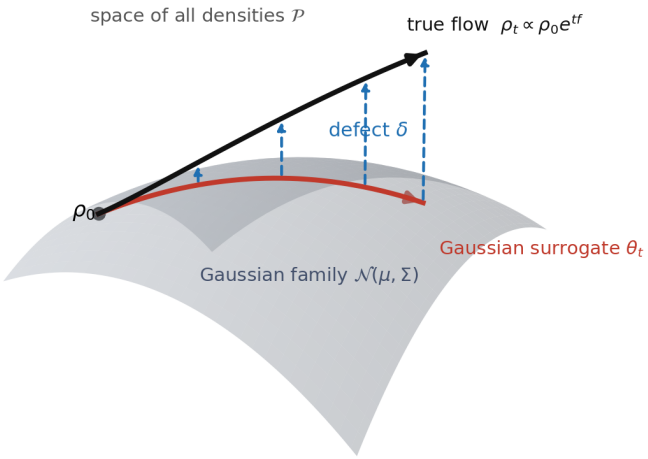


Figure 1. The central construction. The Gaussian family is a curved manifold inside the space of all densities $\mathcal{P}$. The true replicator flow (black) leaves the manifold, while the Gaussian surrogate θ_t (red) is constrained to remain on it. The blue connectors indicate the accumulated tracking error between the exact flow and its best Gaussian representation, which grows when a single Gaussian cannot follow the flow. The instantaneous rate at which this error is generated is the invariance defect of Section 3.4.

relaxation of a deterministic optimization problem has favorable structural properties: Müller and Glasmachers (2021) prove that for a strongly convex objective perturbed by a high-frequency disturbance, the Gaussian relaxation $\mathbb{E}_{P_\theta}[f]$ is consistent, has Lipschitz gradients, and becomes convex once the search scale exceeds a threshold that filters the high-frequency part of the disturbance. This convexification is complementary to the invariance question studied here, because the high-frequency Fourier components their smoothing removes are the same non-affine-quadratic components that place the centered fitness outside the Gaussian score span. Optimizing $\mathbb{E}_{P_{\theta_t}}[f]$ also has downsides, such as instability due to outliers. Fitness shaping is an adaptive transformation of f designed to be invariant under any monotone transformation of f, and it produces a nonzero second term in (14).

The replicator equation can capture fitness shaping, and doing so is natural from an evolutionary standpoint: the growth rate of a trait is relative to the growth rates of the other traits present. Let $g(x)$ be an absolute fitness independent of the distribution, and define $\Pi(x, z) = 1$ when $g(x) > g(z)$ and 0 otherwise. Then

$$h\left(x \mid P_{\theta_t}\right) = \int P_{\theta_t}(z)\Pi(x, z)\, dz = \mathbb{P}_{z\sim P_{\theta_t}}\left(g(x) > g(z)\right), \tag{15}$$

and h can be further transformed by a quantile threshold q_0, for instance $f\left(x \mid P_{\theta_t}\right) = \mathbf{1}\{h\left(x \mid P_{\theta_t}\right) > q_0\}$. This is what evolution strategies do when they retain an elite fraction of the population; $q_0 = 0.75$ retains the top quartile.

3. Exact invariance, the invariance defect, and what it measures

Section 2 assumed forward invariance. This section (i) places the replicator dynamics in a Hilbert space of densities; (ii) characterizes the forward-invariant (family, fitness) pairs ; (iii) defines the invariance defect and establishes that its square is the rate of adaptation lost to the parametric restriction; (iv) identifies the orders of accuracy of the projection in moments and in density; (v) solves the distribution-independent case in closed form; and (vi) bounds the deviation between the projected and true trajectories. The analytic hypotheses are collected in Appendix A.

3.1 The replicator field on the space of densities

Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and ρ a density with $\int \rho \, d\mu = 1$. A *fitness functional* is a map $f : \rho \mapsto (x \mapsto f(\rho, x))$, and the mean fitness is $\bar{f}(\rho) := \int f(\rho, x)\rho(x) \, d\mu(x)$. Define the *replicator field*

$$R(\rho)(x) := \rho(x) \left[f(\rho, x) - \bar{f}(\rho) \right], \tag{16}$$

so that replicator dynamics is $\dot{\rho}_t = R(\rho_t)$. Subtracting $\bar{f}$ makes the normalization invariant: for any ρ_t,

$$\frac{d}{dt} \int_{\mathcal{X}} \rho_t \, d\mu = \int_{\mathcal{X}} f\rho_t \, d\mu - \bar{f}(\rho_t) \int_{\mathcal{X}} \rho_t \, d\mu = \bar{f}(\rho_t) \left(1 - \int_{\mathcal{X}} \rho_t \, d\mu \right), \tag{17}$$

which vanishes on the constraint surface. Normalization, once imposed as an initial condition, is preserved for all t: the flow is tangent to the manifold of normalized densities and never leaves it. The subtraction conserves total mass but does not select its value, so normalization is required as an initial condition and re-imposed at each step in the discretized dynamics.

The natural object is the *relative velocity*

$$w_\rho(x) := \frac{\dot{\rho}(x)}{\rho(x)} = f(\rho, x) - \bar{f}(\rho), \qquad \mathbb{E}_\rho \left[w_\rho \right] = 0, \tag{18}$$

which is automatically centered; its vanishing ρ-mean is the infinitesimal form of mass conservation. Equip the tangent directions with the weighted inner product

$$\langle g, h \rangle_{L^2(\rho)} := \int g(x)h(x)\rho(x) \, d\mu(x), \qquad \| \, g \, \|_{L^2(\rho)} = \left(\int g^2 \rho \, d\mu \right)^{1/2}. \tag{19}$$

This is the correct geometry for the replicator field, because the centered fitness w_ρ *is* the relative velocity, and the metric it induces on a parametric family is the Fisher information matrix of Section 2.

3.2 Parametric families, scores, and tangent space

Let $\{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^m\}$ be a C^1 family of strictly positive densities with score vector

$$u_\theta(x) := \nabla_\theta \log P_\theta(x) \in \mathbb{R}^m, \qquad \mathbb{E}_{P_\theta} \left[u_\theta \right] = 0. \tag{20}$$

A parameter velocity $v \in \mathbb{R}^m$ induces the relative velocity $P_\theta^{-1} \frac{d}{dt} P_{\theta+tv}|_{t=0} = v^\top u_\theta(x)$, so the relative tangent subspace at θ is the finite-dimensional subspace

$$V_\theta := \{ x \mapsto v^\top u_\theta(x) : v \in \mathbb{R}^m \} \subset L^2(P_\theta). \tag{21}$$

3.3 Exact forward invariance

The family is *forward-invariant under f at θ* if some parameter velocity reproduces the replicator field :

$$\exists\, v \in \mathbb{R}^m \text{ such that } v^\top u_\theta(x) = f\left(P_\theta, x\right) - \bar{f}\left(P_\theta\right) = w_\theta(x) \quad \text{for } P_\theta\text{-a.e. } x. \tag{22}$$

Theorem 1 (main result 1, invariance as tangency). *The family is forward-invariant at θ if and only if the centered fitness lies in the tangent space, $w_\theta \in V_\theta$; equivalently, the replicator field $R\left(P_\theta\right)$ is tangent to the family at P_θ.*

Proof. The map $v \mapsto v^\top u_\theta$ is a linear surjection onto V_θ, and $R\left(P_\theta\right) = P_\theta w_\theta$ while the density velocity induced by v is $P_\theta\left(v^\top u_\theta\right)$. Since $P_\theta > 0$, the two densities agree μ-a.e. if and only if $v^\top u_\theta = w_\theta$ holds P_θ-a.e., which is possible for some v if and only if $w_\theta \in V_\theta$. ▫

Theorem 1 is pointwise in θ. Three statements must be distinguished:

Pointwise tangency at a single θ says only that the replicator field does not immediately leave the family at that state.

Local forward invariance. Suppose tangency holds on an open region $U \subseteq \Theta$, that the induced velocity field $\theta \mapsto v(\theta)$ is locally Lipschitz on U, and that the Fisher matrix is nonsingular on U. Then Picard–Lindelöf gives a unique maximal solution θ_t of $\dot{\theta} = v(\theta)$ with $\theta_0 \in U$, defined on a maximal interval $[0, T)$ during which $\theta_t \in U$. On that interval $t \mapsto P_{\theta_t}$ is a strong solution of (4), and by uniqueness of solutions of (4) in L^1 under the hypotheses of Appendix A, it is *the* solution: the density trajectory and the parametric trajectory coincide.

Global forward invariance additionally requires normalizability of the family along the whole trajectory, avoidance of singular Fisher information, and continuation of the parameter solution to all $t > 0$. Section 3.11 gives an example where tangency holds identically but the solution blows up in finite time, so tangency alone does not imply global invariance.

When (22) holds with velocity $v(\theta)$, the induced parameter ODE is $\dot{\theta} = v(\theta)$.

3.4 The invariance defect and projected evolution

When the family is not invariant, the residual for a candidate velocity v is $Res_\theta(v) := v^\top u_\theta - w_\theta$, and the *invariance defect* is the smallest residual achievable in $L^2\left(P_\theta\right)$:

$$\delta(\theta) := \min_{v \in \mathbb{R}^m} \parallel v^\top u_\theta - w_\theta \parallel_{L^2(P_\theta)}. \tag{23}$$

The family is ε-forward invariant on a region $\mathcal{R} \subseteq \Theta$ if $\delta(\theta) \leq \varepsilon$ for every $\theta \in \mathcal{R}$. The defect is a Hilbert-space least-squares problem, so it has a closed form. With the Fisher (Gram) matrix and the fitness-score covariance

$$F(\theta) := \mathbb{E}_{P_\theta}\left[u_\theta u_\theta^\top\right], \qquad c(\theta) := \mathbb{E}_{P_\theta}\left[u_\theta w_\theta\right] = Cov_{P_\theta}\left(u_\theta, f\right), \tag{24}$$

the projected (best-fit) parameter velocity and the minimized defect are

$$v^\star(\theta) = F^\dagger(\theta) c(\theta), \qquad \delta^2(\theta) = Var_{P_\theta}\left(w_\theta\right) - c(\theta)^\top F^\dagger(\theta) c(\theta), \tag{25}$$

where $F^\dagger$ is the Moore–Penrose pseudoinverse, equal to F^{-1} for a minimal family with invertible Fisher matrix. The projected evolution $\dot{\theta} = v^\star(\theta)$ is the $L^2\left(P_\theta\right)$-orthogonal projection of the replicator field onto the tangent space, and it reduces to the IGO flow (11) whenever the defect is zero.

The most useful informal reading of (25) is as a regression. The defect squared is the residual variance of the least-squares regression of relative fitness on the scores, taken under the current distribution. The normalized version

$$R^2(\theta) := \frac{c(\theta)^\top F^\dagger(\theta) c(\theta)}{Var_{P_\theta}(w_\theta)} = 1 - \frac{\delta^2(\theta)}{Var_{P_\theta}(w_\theta)} \tag{26}$$

is the fraction of the selection signal the family can express, with $R^2 = 1$ at forward invariance.

Proposition 2 (invariance properties). *(i) $\delta(\theta)$ is unchanged by any smooth reparameterization $\theta \mapsto \varphi(\theta)$ with invertible Jacobian. (ii) Under $f \mapsto af + b$ with $a > 0$, $\delta \mapsto a\delta$ and R^2 is unchanged. (iii) δ does not depend on the choice of pseudoinverse when F is singular.*

Proof. (i) A reparameterization replaces u_θ by $J^\top u_\theta$ with J invertible, which leaves the span V_θ unchanged; δ depends on the scores only through V_θ. (ii) $w_\theta \mapsto a w_\theta$ and V_θ is unchanged, so the orthogonal projection scales by a; the ratio in (26) is invariant. (iii) δ is the norm of the residual of the orthogonal projection onto the closed subspace V_θ, which is unique regardless of which minimizing v is selected. ▫

Property (i) matters in practice, because it means the defect is a property of the family and the fitness, not of the coordinates used to describe the family. Natural and moment coordinates for an exponential family give the same number. Property (iii) means that rank-deficient dictionaries, such as the rank-bin bases of Section 5.5, still have a well-defined defect even though their coefficient vectors are not identified.

3.5 The defect as lost adaptation rate

Equations (23) to (26) present the defect as a projection residual. It is more than that. For distribution-independent fitness it is the adaptation rate that the parametric restriction destroys.

Theorem 3 (main result 2, decomposition of Fisher's fundamental theorem). *Let f be independent of the distribution, let P_θ be a regular family with score u_θ, Fisher matrix $F(\theta)$ and fitness-score covariance $c(\theta) = Cov_{P_\theta}(u_\theta, f)$, and let $J(\theta) = \mathbb{E}_{P_\theta}[f]$. Evaluate both flows at the same state P_θ. Then the full replicator flow increases mean fitness at rate*

$$\frac{d}{dt}\mathbb{E}_\rho[f]|_{\rho=P_\theta} = Var_{P_\theta}(f), \tag{27}$$

the Fisher-projected flow increases it at rate

$$\frac{d}{dt} J(\theta_t) = \nabla_\theta J \cdot \dot\theta = c(\theta)^\top F^\dagger(\theta)\, c(\theta), \tag{28}$$

and the difference is the squared invariance defect:

$$Var_{P_\theta}(f) \;=\; \underbrace{c(\theta)^\top F^\dagger(\theta) c(\theta)}_{\text{expressible by the family}} \;+\; \underbrace{\delta^2(\theta)}_{\text{lost to the restriction}}. \tag{29}$$

Proof. For (27), differentiate under the integral: $\frac{d}{dt}\int f\rho\, d\mu = \int f\rho\,(f - \bar f)\, d\mu = \mathbb{E}_\rho[f^2] - \bar f^2 = Var_\rho(f)$. For (28), $\nabla_\theta J = \mathbb{E}_{P_\theta}[u_\theta f] = c(\theta)$ because the score is centered, and $\dot\theta = F^\dagger c$ by (25). Equation (29) is then (25) together with $Var_{P_\theta}(w_\theta) = Var_{P_\theta}(f)$, which holds because w_θ and f differ by a constant. ▫

Equation (27) is Fisher's fundamental theorem in its continuous-trait form: the rate of increase of mean fitness equals the variance in fitness. Equation (29) states that when the population is confined to a parametric family, that variance splits into the part the family's score directions can convert into parameter motion and the part it cannot. The defect is not an abstract measure of representational error: it is a rate

of adaptation, in the same units as the fitness flux of Mustonen and Lässig (2010), and it is the quantity a restricted population forfeits per unit time. By Proposition 2 it is also invariant under smooth reparameterization, so the forfeited rate is a property of the family and the fitness rather than of the coordinates chosen to describe them.

Theorem 3 compares two instantaneous rates at a common state; it does not compare accumulated fitness along two trajectories, which would additionally require the states themselves to be compared. Equation (29) describes the selection pressure available at a given distribution and the fraction of it the family can act on.

This gives the Gaussian-closure question of quantitative genetics (Lande, 1979; Turelli, 1984; Bürger, 2000) a direct answer: the cost of the closure at any state is δ^2, and R^2 in (26) is the fraction of the response to selection that the closure retains.

Two consequences are immediate.

Corollary 3.1 (monotonicity). *For distribution-independent fitness, the Fisher-projected flow never decreases mean fitness:* $\frac{d}{dt} J(\theta_t) = c^\top F^\dagger c \geq 0$, *with equality if and only if* $c(\theta_t) = 0$.

Projection slows adaptation but does not reverse it. This is a useful sanity property for algorithms built on the projection, and it does not require the flow to be the gradient flow of any potential; it follows from positive semi-definiteness of $F^\dagger$ alone.

Corollary 3.2 (stalling). *The projected flow has a stationary point at* θ *if and only if* $c(\theta) = 0$. *If in addition* $\delta(\theta) > 0$, *the population still carries fitness variance* $Var_{P_\theta}(f) = \delta^2(\theta) > 0$, *and the full replicator flow would continue to increase mean fitness at that rate while the projected flow does not move.*

Corollary 3.2 states that a family stalls not when adaptation is exhausted, but when the remaining fitness variance is entirely orthogonal to its score span. Such a state is stationary for the projected dynamics and is not an equilibrium of the unrestricted dynamics, so the two descriptions disagree not only in rate but in whether anything is happening at all. A Gaussian centered between two symmetric peaks of a bimodal landscape is the canonical instance: by symmetry $Cov(X, f) = 0$, and $Cov\left((X-\mu)^2, f\right)$ can vanish as well, so the search distribution is stationary while the true dynamics would resolve the population onto the peaks. This is the mechanism behind the long near-stalling transients seen for the single Gaussian in Supplementary Experiments S4 and S10. It also motivates the practical rule of Section 7.2: monitor c and δ together and treat a small $\| c \|$ alongside a large δ as the informative warning.

The symmetric case has additional structure.

Corollary 3.3 (symmetric stalling states are non-hyperbolic and never asymptotically stable). *Let* f *be even and let* P_θ *be the one-dimensional Gaussian family with* $\theta = (\mu, \sigma^2)$. *Stein's identity gives* $Cov(X, f) = \sigma^2 \mathbb{E}[f'(X)]$ *and, at* $\mu = 0$, $Cov\left((X-\mu)^2, f\right) = \sigma^4 \mathbb{E}[f''(X)]$, *so the projected flow is* $\dot{\mu} = \sigma^2 \mathbb{E}[f'(X)]$ *and* $\dot{\sigma^2} = \sigma^4 \mathbb{E}[f''(X)]$. *Since* f' *is odd,* $\mu = 0$ *is always stationary in the mean, and the variance is stationary precisely when*

$$\mathbb{E}_\theta[f''(X)] = 0, \tag{30}$$

which defines the symmetric stalling variance $\sigma_\star^2$. *The linearization of the mean equation at* $\mu = 0$ *is* $\partial\dot{\mu}/\partial\mu = \sigma^2 \mathbb{E}_\theta[f'']$, *which vanishes under the same condition, so the state is non-hyperbolic in the mean and nonlinear terms decide its stability.*

Write $E_4 := \mathbb{E}_{\theta_\star}[f''''(X)]$. *Then the variance eigenvalue and the center-manifold reduction are*

$$\lambda_{var} = \frac{1}{2}\sigma_\star^4 E_4, \qquad \sigma^2 = \sigma_\star^2 - \mu^2 + O(\mu^4), \qquad \dot{\mu} = -\frac{1}{3}\sigma_\star^2 E_4\, \mu^3 + O(\mu^5).$$

The variance eigenvalue and the cubic coefficient therefore have opposite signs for every even f with $E_4 \neq 0$, and consequently a symmetric stalling state is never asymptotically stable: if $E_4 < 0$ the variance direction attracts while the mean direction repels cubically, and if $E_4 > 0$ the variance direction is itself repelling. The stable set is at most the invariant symmetry subspace $\mu = 0$, never an open basin.

Proof. The covariance identities are Stein's lemma applied to f and to f'; differentiating $\sigma^2 \mathbb{E}_{\mathcal{N}(\mu,\sigma^2)}[f']$ in μ at $\mu = 0$ gives $\sigma^2 \mathbb{E}[f'']$. Write $H(\mu, s) := \mathbb{E}_{\mathcal{N}(\mu,s)}[f'']$ with $s = \sigma^2$. Gaussian smoothing satisfies the heat identity $\partial_s \mathbb{E}_{\mathcal{N}(\mu,s)}[g] = \frac{1}{2}\mathbb{E}[g'']$, so $\partial_s H\,(0, s_\star) = \frac{1}{2}E_4$ and, since $\dot{s} = s^2 H$, the variance eigenvalue is $\partial_s\left(s^2 H\right)|_{(0,s_\star)} = s_\star^2 \partial_s H = \frac{1}{2}s_\star^2 E_4$. Expanding in μ with f'' even gives $H(\mu, s) = H(0, s) + \frac{\mu^2}{2}E_4 + O\left(\mu^4\right)$, so $\dot{s} = 0$ requires $\frac{1}{2}E_4\,(s - s_\star) + \frac{\mu^2}{2}E_4 = 0$, that is $s = s_\star - \mu^2 + O\left(\mu^4\right)$, independently of E_4. Expanding $\mathbb{E}_{\mathcal{N}(\mu,s)}[f'] = \mu H(0, s) + \frac{\mu^3}{6}E_4 + O\left(\mu^5\right)$ and substituting $H(0, s) = -\frac{\mu^2}{2}E_4$ on the manifold gives $\dot{\mu} = s_\star\left(-\frac{1}{2} + \frac{1}{6}\right)E_4\mu^3 = -\frac{1}{3}s_\star E_4\mu^3$. ▫

The coupling between the two directions determines the sign. Holding the variance fixed at $\sigma_\star^2$ and perturbing only the mean gives a cubic coefficient $+\frac{1}{6}\sigma_\star^2 E_4$, which is $-\frac{1}{2}$ times the correct value: the variance responds to an asymmetric perturbation at second order, $\sigma^2 = \sigma_\star^2 - \mu^2$, and that response reverses the sign of the mean dynamics. A stability calculation performed at fixed variance therefore predicts attraction where the coupled flow gives escape.

Example (double well). For $f(x) = -\left(x^2 - a^2\right)^2$ the projected equations are

$$\dot{\mu} = 4\sigma^2\mu\left(a^2 - \mu^2 - 3\sigma^2\right), \qquad \dot{\sigma^2} = 4\sigma^4\left(a^2 - 3\mu^2 - 3\sigma^2\right),$$

condition (30) reads $4a^2 - 12\sigma^2 = 0$, and the symmetric stationary state is $\theta_\star = (0, a^2/3)$. Here $E_4 = -24$, so $\lambda_{var} = -4a^4/3$ and the reduced mean dynamics is $\dot{\mu} = k\mu^3$ with $k = 8a^2/3 > 0$: the state repels. At $\theta_\star$ the fitness-score covariance vanishes and the entire fitness variance is unexpressed,

$$\delta^2\left(\theta_\star\right) = Var_{\theta_\star}(f) = \frac{8}{27}a^8, \qquad R^2\left(\theta_\star\right) = 0. \tag{31}$$

The escape is slow, and its timescale is what makes the phenomenon observable rather than merely possible. The two directions separate in time by a factor $t^\star / \left|\lambda_{var}\right|^{-1} = 3\sigma_\star^2 / \left(4\mu_0^2\right)$ for an initial asymmetry μ_0, which is 2.5×10^3 at $a = 2$ and $\mu_0 = 0.02$, so the center-manifold reduction is valid throughout the departure and the reduced equation integrates in closed form,

$$\mu(t)^2 = \frac{\mu_0^2}{1 - 2k\mu_0^2 t}.$$

The asymmetry therefore doubles at $t_2 = 3/\left(8k\mu_0^2\right) = 9/\left(64a^2\mu_0^2\right)$, and the local law diverges at $t^\star = 1/\left(2k\mu_0^2\right) = 3/\left(16a^2\mu_0^2\right)$, where the trajectory leaves the neighborhood of $\theta_\star$ and the quartic terms take over. Both times scale as μ_0^{-2} and as a^{-2}: a deeper well forfeits more adaptation rate, by (31), and also releases the population sooner. Integrating the coupled equations at $a = 2$ confirms the law and its constant.

μ_0	t_2 predicted	t_2 exact	$t^\star$
0.05	14.1	14.3	18.7
0.02	87.9	88.1	117.2
0.01	351.6	351.8	468.8

The agreement improves as μ_0 decreases.

The length of the plateau is set by the initial asymmetry. At $a = 2$ an asymmetry equal to one percent of the well half-width persists for of order 10^2 time units before symmetry breaking becomes visible,

and throughout that interval the projected flow is stationary to numerical precision while the unrestricted flow increases mean fitness at the full rate $\delta^2(\theta_\star)$. A trajectory initialized at exact symmetry therefore establishes a genuine stalling state but not robust attraction. Exact symmetric stalling is nongeneric; long-lived near-symmetric stalling is what an algorithm encounters, and it is what a diagnostic based on $(\| c \|, \delta)$ is able to detect. Supplementary Experiment S10 reports the exact stationary state, the two symmetry-broken outcomes, and the near-stalling region measured against this law.

For frequency-dependent fitness the clean total-rate split does not survive. Let $J(P) = \mathbb{E}_P[f(P,X)]$, let $R(P)$ and $R_{proj}(P)$ denote the full and projected density velocities, and let r be their relative-velocity residual. Differentiating J along a density velocity produces one term from the change in the distribution and one from the change in the fitness function itself, and only the first is governed by Theorem 3:

$$\frac{dJ}{dt}|_R - \frac{dJ}{dt}|_{R_{proj}} = \delta^2 + \mathbb{E}_P\left[D_P f(P,X)\left[R(P) - R_{proj}(P)\right]\right].$$

Theorem 3 therefore applies to the selection term. The second term is not common to the two flows, because the functional derivative is evaluated along different velocities, and the total mean-fitness-rate difference equals δ^2 only when it vanishes, as it does for distribution-independent fitness.

3.6 Moment closure and orders of accuracy

The projection has a second interpretation that explains why moment trajectories can remain close even after density shapes visibly separate; the numerical tests are reported in Supplementary Experiments S1 and S11.

Proposition 4 (the projection is a moment closure). *Let P_θ be an exponential family with sufficient statistic T and mean parameter $\eta(\theta) = \mathbb{E}_\theta[T]$. Along the true replicator flow, the exact moment dynamics is*

$$\frac{d}{dt}\mathbb{E}_{\rho_t}[T] = Cov_{\rho_t}(T,f), \tag{32}$$

and the Fisher-projected flow satisfies

$$\dot{\eta}(\theta_t) = Cov_{P_{\theta_t}}(T,f). \tag{33}$$

The two differ only in the measure under which the covariance is evaluated. The projected flow is therefore the moment closure of (32) obtained by substituting the current family member for the true density.

Proof. Equation (32) follows from $\frac{d}{dt}\int T\rho = \int T\rho\left(f - \bar{f}\right) = Cov_\rho(T,f)$. Equation (33) is (25) written in mean coordinates, using $\dot{\eta} = F\dot{\theta} = FF^{-1}c = c = Cov_\theta(T,f)$, since the score of an exponential family is $u_\theta = T - \eta$. ▫

This identifies the projected flow with a construction familiar from quantitative genetics and from moment-closure methods in ecology and epidemiology, and it makes the source of error explicit: it is not that the moment equations are wrong, but that they are evaluated under the wrong measure.

Corollary 4.1 (local moment and density agreement). Let P_t be the true trajectory and P_{θ_t} the Fisher-projected trajectory, initialized at the same family member. Let $\eta_t = \mathbb{E}_{P_t}[T]$ and $\tilde{\eta}_t = \mathbb{E}_{P_{\theta_t}}[T]$. Then $\eta_0 = \tilde{\eta}_0$ and $\dot{\eta}_0 = \dot{\tilde{\eta}}_0$, so

$$\eta_t - \tilde{\eta}_t = O\left(t^2\right).$$

If r_0 is the relative residual at the initial state, then the density difference has derivative $P_0 r_0$ at $t = 0$, and therefore

$$\| P_t - P_{\theta_t} \|_1 = t\, \mathbb{E}_{P_0} |r_0| + o(t) \le t\, \delta(\theta_0) + o(t).$$

Proof. At $t = 0$ the two measures coincide, so the right-hand sides of the exact and closed moment equations are equal. For the density, $\partial_t \left(P_t - P_{\theta_t} \right) |_{t=0} = P_0 r_0$; integration gives the stated L^1 expansion, and Cauchy-Schwarz gives $\mathbb{E}\,|r_0| \le \sqrt{\mathbb{E}\,[r_0^2]} = \delta(\theta_0)$. ▫

The projection therefore preserves the initial derivative of the sufficient-statistic means even when the defect is nonzero, while density shape begins to separate immediately whenever $r_0 \neq 0$. The leading density coefficient is $\mathbb{E}\,|r_0|$, not generically δ; the defect supplies a coordinate-invariant upper bound and an ordering of omitted relative velocity. Supplementary Experiments S1 and S11 test these two orders directly.

3.7 The exact trajectory for distribution-independent fitness

When f does not depend on the distribution the replicator equation can be solved in closed form, which yields a sharper description than any of the bounds below.

Proposition 5 (exponential geodesic and cumulant generating function). *Let f be independent of ρ and let $K(t) := \log \mathbb{E}_{\rho_0}\left[e^{tf}\right]$ be the cumulant generating function of f under ρ_0, with $\mathcal{T} := \{t \ge 0 : K(t) < \infty\}$. For $t \in \mathcal{T}$,*

$$\rho_t(x) = \rho_0(x)\, e^{tf(x) - K(t)}, \qquad \mathbb{E}_{\rho_t}[f] = K'(t), \qquad Var_{\rho_t}(f) = K''(t). \tag{34}$$

The trajectory is an exponential (e-) geodesic through ρ_0 in the direction f, that is, a straight line in the natural parameters of the exponential family generated by ρ_0 and f. The solution exists on $\mathcal{T}$, and blow-up occurs at $t^\star = sup\mathcal{T}$.

The letter e denotes the exponential connection of information geometry: an e-geodesic is a straight line in the natural coordinates of an exponential family, not necessarily a shortest path for the Fisher–Rao Riemannian distance (Amari and Nagaoka, 2000).

Proof. Direct substitution verifies that (34) solves (4) with $\bar{f}(\rho_t) = K'(t)$; uniqueness under the hypotheses of Appendix A gives the claim. The moment identities are the defining property of the cumulant generating function, and $Var_{\rho_t}(f) = K''(t)$ is Fisher's fundamental theorem along this trajectory. ▫

Proposition 5 has three uses. First, it serves as ground truth for the numerical tests in Supplementary Sections S1, S3, and S6 without integrating the density equation. Second, it shows that the true path is a straight line in the natural parameters of the exponential family generated by f, and the family is forward-invariant when that line already lies in the smaller coordinate subspace, which recovers Theorem 6 below. Third, it converts finite-time validity into a statement about the domain of a single scalar function: the parametric flow can be continued as long as $\psi(t)$ is finite.

Because K is convex, $K'' \ge 0$: mean fitness increases monotonically along the true flow, matching Corollary 3.1 for the projected flow. Where the two differ is in rate, and Theorem 3 says the gap is δ^2.

The cumulant-generating-function calculation has an antecedent in the cumulant equations of Bürger (1991), developed for quantitative-trait distributions under polynomial selection. For Gaussian initial distributions and quadratic fitness, the first two cumulant equations agree with the Gaussian mean and covariance equations in Section 3.9, while higher cumulants remain zero as long as the Gaussian solution is normalizable. Theorem 7 supplies the converse invariance criterion; the defect measures the departure from closure for other fitness functions. The cumulant hierarchy for a non-Gaussian initial density does not generally truncate at second order even for quadratic fitness.

3.8 Exact invariance for exponential families

For a regular exponential family in natural parameters,

$$P_\theta(x) = exp\left(\theta^\top T(x) - A(\theta)\right) h(x), \qquad A(\theta) = log \int e^{\theta^\top T(x)} h(x)\, d\mu(x), \tag{35}$$

the mean parameter is $\eta(\theta) = \mathbb{E}_\theta[T] = \nabla A(\theta)$, the score is $u_\theta = T - \eta(\theta)$, and the Fisher information is $F(\theta) = \nabla^2 A(\theta) = Cov_\theta(T)$. The tangent space is $V_\theta = \{v^\top(T-\eta) : v \in \mathbb{R}^k\}$.

Theorem 6 (exponential-family forward invariance). *Assume the family (35) is minimal and Θ is open. It is forward-invariant under the replicator dynamics at every $\theta \in \Theta$ if and only if, for every $\theta \in \Theta$, the fitness restricted to the family is affine in the sufficient statistics,*

$$f\left(P_\theta, x\right) = a(\theta)^\top T(x) + b(\theta) \quad \text{for } P_\theta\text{-a.e. } x, \tag{36}$$

for some $a(\theta) \in \mathbb{R}^k$ and $b(\theta) \in \mathbb{R}$. In that case the additive constant cancels in w_θ, the defect is identically zero, and the induced parameter ODE is $\dot\theta = a(\theta)$. If in addition a is locally Lipschitz on Θ, the local forward invariance of Section 3.3 holds.

Proof. Sufficiency: if (36) holds then $w_\theta = a(\theta)^\top\left(T - \eta(\theta)\right) \in V_\theta$, and Theorem 1 applies with $v = a(\theta)$. Necessity: forward invariance at θ requires $w_\theta \in V_\theta = span\{T-\eta\}$, so $f - \bar f$ is P_θ-a.e. an affine function of T; minimality ensures the representation is unique. ▫

Minimality and openness are needed for the converse; without minimality the components of T are affinely dependent and the representation in (36) is not unique, though the conclusion $\delta \equiv 0$ still holds. Note also that a and b may depend on θ, so (36) is weaker than requiring f to be globally affine in T; for frequency-dependent fitness this is the relevant form.

Whether or not (36) holds, the $L^2\left(P_\theta\right)$ projection has a universal closed form. In moment coordinates,

$$\dot\eta = Cov_\theta(T, f), \tag{37}$$

and, when $F(\theta)$ is invertible, in natural coordinates

$$\dot\theta = F(\theta)^{-1} Cov_\theta(T, f). \tag{38}$$

For a *curved* family $u \mapsto \theta(u)$ with $u \in \mathbb{R}^m$ and Jacobian $J(u) = \partial\theta/\partial u \in \mathbb{R}^{k\times m}$, the relative velocity is $\dot u^\top J^\top(T-\eta)$ and the projected dynamics solves the reduced normal equations $J^\top F J\,\dot u = J^\top Cov(T, f)$.

A corollary worth stating explicitly, because it is the case that arises in evolutionary game theory, concerns pairwise payoff kernels.

Corollary 6.1. *Suppose the fitness has the pairwise form $f(x \mid P) = \int \Pi(x,z)P(z)\,dz$ and $\Pi(\cdot, z)$ lies in $span\{1, T_1, \dots, T_k\}$ for every z. Then (36) holds with $a(\theta) = \int \alpha(z) P_\theta(z)\,dz$, where $\alpha(z)$ are the coefficients of $\Pi(\cdot, z)$, and the family is forward-invariant.* In particular, a payoff kernel that is a polynomial of degree at most two in its first argument, with arbitrary dependence on the second, keeps the Gaussian family invariant. This recovers the setting of Cressman et al. (2006) and Jaćimović (2024).

3.9 Exact invariance for the Gaussian family

Let's set $\mathcal{X} = \mathbb{R}^n$ with Lebesgue measure and the full Gaussian family $\rho_{\mu,\Sigma}$, $\mu \in \mathbb{R}^n$, $\Sigma \in \mathcal{S}^n_{++}$. The pointwise parameter derivatives are

$$D_\mu \rho_{\mu,\Sigma}(x)[u] = \rho_{\mu,\Sigma}(x)\, u^\top \Sigma^{-1}(x-\mu),$$

$$D_\Sigma \rho_{\mu,\Sigma}(x)[H] = \frac{1}{2}\rho_{\mu,\Sigma}(x)\left[(x-\mu)^\top \Sigma^{-1} H \Sigma^{-1}(x-\mu) - tr\left(\Sigma^{-1}H\right)\right],$$

so that, factoring out $\rho_{\mu,\Sigma}$, the tangent functions are the centered polynomials of degree at most two. Dividing the invariance condition (22) by $\rho_{\mu,\Sigma}(x) > 0$ gives the core characterization: for a velocity $(\dot{\mu}, \dot{\Sigma})$,

$$f\left(\rho_{\mu,\Sigma}, x\right) - \bar{f} = \dot{\mu}^\top \Sigma^{-1}(x-\mu) + \frac{1}{2}(x-\mu)^\top \Sigma^{-1}\dot{\Sigma}\Sigma^{-1}(x-\mu) - \frac{1}{2}tr\left(\Sigma^{-1}\dot{\Sigma}\right) \quad \text{for all } x. \tag{39}$$

The right-hand side is a quadratic polynomial in x centered at μ. Therefore:

Theorem 7 (Gaussian forward invariance). *The full Gaussian family is forward-invariant under the replicator dynamics at every $(\mu, \Sigma) \in \Theta$ if and only if, for every $(\mu, \Sigma) \in \Theta$, the fitness restricted to the family is affine-quadratic in x,*

$$f\left(\rho_{\mu,\Sigma}, x\right) = \alpha(\mu,\Sigma) + \beta(\mu,\Sigma)^\top x + x^\top A(\mu,\Sigma)x \quad \text{a.e. } x, \tag{40}$$

with A symmetric without loss of generality and the coefficients allowed to depend on the current (μ, Σ). The constant α is irrelevant since the replicator field uses $f - \bar{f}$. Matching coefficients gives one valid parameter velocity,

$$\dot{\mu} = \Sigma(\beta + 2A\mu), \qquad \dot{\Sigma} = 2\Sigma A \Sigma. \tag{41}$$

Sufficiency of quadratic fitness follows from the direct moment-generating-function computation, or from Proposition 5: the tilt of a Gaussian by a quadratic exponent is Gaussian whenever it is normalizable. Equation (39) supplies the converse for the full Gaussian family, because $f - \bar{f}$ must equal a member of the degree-at-most-two tangent space. When f is not quadratic, the $L^2(\rho)$-projected velocities take the moment form

$$\dot{\mu} = Cov_\rho\left(X, f(\rho, X)\right), \qquad \dot{\Sigma} = Cov_\rho\left((X-\mu)(X-\mu)^\top, f(\rho, X)\right), \tag{42}$$

with zero defect in the exactly invariant case. Equation (42) is Lande's equation together with its second-moment companion; the mean equation is Robertson's secondary theorem of natural selection (Robertson, 1966), and the pair is the Gaussian closure whose validity Turelli (1984) questioned. Theorem 3 says the cost of that closure is, as made explicit by Eq. (46), δ^2 per unit time.

The Gaussian defect in closed form. The general projection formula (25) has a useful explicit form for the full Gaussian family. Let $X \sim \mathcal{N}(\mu, \Sigma)$ with positive-definite Σ, let $Y = X - \mu$, and define the fitness–moment covariances in Eq. (43)

$$b = \mathbb{E}\left[Y f(X)\right], \qquad B = \mathbb{E}\left[\left(YY^\mathsf{T} - \Sigma\right) f(X)\right]. \tag{43}$$

The Gaussian tangent space consists of centered linear and quadratic functions. These two subspaces are orthogonal under a Gaussian density. The linear and quadratic components of the projection, denoted $L(X)$ and $Q(X)$, are given by Eqs. (44) and (45):

$$L(X) = b^\mathsf{T}\Sigma^{-1}Y, \tag{44}$$

$$Q(X) = \frac{1}{2}\left[Y^\mathsf{T}\Sigma^{-1}B\Sigma^{-1}Y - tr\left(\Sigma^{-1}B\right)\right]. \tag{45}$$

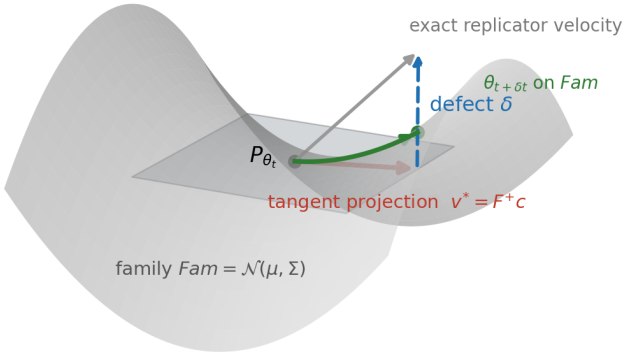


Figure 2. Rather than leaving the family and projecting a point back, the replicator velocity is projected directly onto the tangent space of the family by a single least-squares projection in the weighted Hilbert space $L^2(P_\theta)$. The projected velocity is $v^\star = F^\dagger c$, with F the Fisher matrix and c the fitness-score covariance; the component of the velocity orthogonal to the tangent space is the invariance defect δ, and the surrogate θ_t never leaves the family. The construction makes no finite excursion off the family. By Theorem 3, the squared length of the retained component is the adaptation rate the family expresses and δ^2 is the rate it forfeits.

Using the Gaussian fourth-moment identity, their squared norms are $b^\mathsf{T}\Sigma^{-1}b$ and $\frac{1}{2}tr\left[\left(\Sigma^{-1}B\right)^2\right]$, respectively. By orthogonality, the squared invariance defect is therefore given by Eq. (46):

$$\delta^2(\mu,\Sigma) = Var\left[f(X)\right] - b^\mathsf{T}\Sigma^{-1}b - \frac{1}{2}tr\left[\left(\Sigma^{-1}B\right)^2\right]. \tag{46}$$

In one dimension, writing the Gaussian variance as σ^2, this becomes Eq. (47):

$$\delta^2 = Var\left[f(X)\right] - \frac{Cov\left(f(X),X\right)^2}{\sigma^2} - \frac{Cov\left(f(X),(X-\mu)^2\right)^2}{2\sigma^4}. \tag{47}$$

Thus the defect is the fitness variance not explained by the first two nonconstant Gaussian Hermite orders. It is exactly zero for every affine-quadratic fitness, not merely of order t^2: provided the tilted density remains normalizable, the true and projected trajectories coincide at all times of existence. For $f = f_{quad} + \varepsilon g$, with an affine-quadratic f_{quad}, linearity of projection gives Eq. (48):

$$\delta^2 = \varepsilon^2 \parallel (I - \Pi_G)\, g \parallel^2_{L^2(\rho)}. \tag{48}$$

The squared defect is consequently quadratic in the amplitude of the omitted fitness component, not in elapsed time. In contrast, Corollary 4.1 concerns time-dependent tracking: moment errors start at order t^2, and density errors at order t when the initial residual is nonzero. Supplementary Section S1 evaluates Eq. (47) for a bounded nonquadratic example.

3.10 Deviation between the Fisher-projected and true trajectories

Let ρ_t be the true replicator trajectory with $\rho_0 = P_{\theta_0}$, and let P_{θ_t} follow the Fisher-projected equation $\dot{\theta}_t = v^\star(\theta_t) = F^\dagger(\theta_t)\, c(\theta_t)$. Define the relative residual and its density form

$$r_\theta(x) = w_\theta(x) - v^\star(\theta)^\top u_\theta(x), \qquad E_\theta(x) = P_\theta(x) r_\theta(x),$$

so $\delta(\theta) = \| r_\theta \|_{L^2(P_\theta)}$. Because P_θ is a probability density, Cauchy–Schwarz gives $\| E_\theta \|_{L^1(\mu)} = \mathbb{E}_{P_\theta} |r_\theta| \le \delta(\theta)$. This norm conversion is what makes the trajectory estimate apply to the same relative-velocity Fisher projection used in (23) to (25), rather than to a different ordinary-L^2 projection of density velocities (Appendix C).

Assume the replicator field is locally Lipschitz in L^1 on the region traversed by the two paths, $\| R(\rho) - R(\sigma) \|_1 \le L \| \rho - \sigma \|_1$, and that strong solutions exist there. With $e(t) = \| \rho_t - P_{\theta_t} \|_1$, the upper right derivative satisfies $D^+ e(t) \le L e(t) + \delta(\theta_t)$, and Grönwall's inequality yields

$$e(t) \le \int_0^t e^{L(t-s)} \delta(\theta_s)\, ds \le \frac{\varepsilon}{L}\left(e^{Lt} - 1\right) \quad \text{if } \delta(\theta_s) \le \varepsilon, \tag{49}$$

with the continuous limit $e(t) \le \varepsilon t$ when $L = 0$. Total variation is one half of the L^1 distance.

A concrete sufficient condition is available. If $\| f(\rho, \cdot) \|_\infty \le M$ and $\| f(\rho, \cdot) - f(\sigma, \cdot) \|_\infty \le K \| \rho - \sigma \|_1$ throughout the region, the calculation in Appendix A gives $L = 3M + 2K$. For fixed fitness on bounded support, the replicator field is unchanged by adding a constant to f, so M may be taken as one half of the oscillation of f on that support.

This bound should be read as a validity guarantee and not as a numerical error estimate. Supplementary Experiment S6 finds that it exceeds the observed error by roughly five orders of magnitude. A single worst-case Lipschitz constant is applied uniformly over the entire region, and the exponential envelope compounds that conservatism.

Two refinements were examined numerically. Supplementary Experiment S18 replaces the two-sided Lipschitz constant by a sampled one-sided logarithmic-norm constant. The sampled constant remains positive on the tested region, so no contraction is established, but it is much smaller and yields a substantially tighter local envelope. Supplementary Experiment S19 instead compares the projected path with the best static family member, obtained by moment matching for an exponential family, and separates the irreducible family-approximation floor from dynamic drift. A relative-entropy trajectory estimate remains open.

The accumulated defect is therefore best treated as a diagnostic of where and how strongly omitted velocity is being generated. Supplementary Experiment S6 finds it on the scale of the observed error in one test family, whereas Supplementary Experiment S19 shows that the proportionality depends on time horizon and error metric. It should not be promoted to a universal calibrated error estimator.

3.11 Finite-time validity and global continuation

The projected evolution can be valid only for a finite time when the family degenerates. By Proposition 5, for fitness depending only on x the replicator solution is the exponential tilt $\rho_t \propto \rho_0 e^{tf}$, which exists while $K(t) < \infty$. If $\rho_0 = \mathcal{N}(\mu_0, \sigma_0^2)$ and $f(x) = a + bx + cx^2$, then

$$\rho_t(x) \propto exp\left[-\left(\frac{1}{2\sigma_0^2} - ct \right) x^2 + \cdots \right],$$

which remains a normalizable Gaussian only while $\frac{1}{2\sigma_0^2} - ct > 0$. For $c < 0$ this holds for all t and Gaussian invariance is global. For $c > 0$ there is finite-time blow-up at

$$t^\star := \frac{1}{2c\sigma_0^2}, \tag{50}$$

past which $\rho_0 e^{tf}$ is not integrable, mass escapes to the tails, and the variance equation drives σ_t^2 to infinity. The parameter ODEs must therefore be understood on $[0, t^\star)$. Tangency holds identically here, so this is an example in which pointwise invariance does not give global forward invariance.

4. Conjugate families

4.1 The conjugacy principle

Sections 3.8 and 3.9 show that forward invariance depends on *tangency*, and for an exponential family the tangent space is spanned by the centered sufficient statistics $T - \eta(\theta)$. The parametric evolution is therefore exact, with identically zero defect, if and only if the restricted fitness is affine in T:

$$f\left(P_\theta, x\right) = a(\theta)^\top T(x) + b(\theta) \;\Rightarrow\; w_\theta = a(\theta)^\top (T - \eta) \in V_\theta, \quad \delta(\theta) \equiv 0, \quad \dot\theta = a(\theta). \tag{51}$$

This is a conjugacy relation between a family and an invariance-compatible class of fitness functions, in the spirit of the conjugate-prior relation of Bayesian statistics: the fitness functions under which the manifold is closed under selection form the linear span of the family's sufficient statistics, modulo an additive constant. Proposition 5 explains why the analogy is more than verbal. For distribution-independent fitness the replicator flow is exponential tilting, and conjugacy in the Bayesian sense is closure of a family under exponential tilting by the likelihood. The two conditions are the same condition.

When the restricted fitness leaves this class, the defect is the $L^2\left(P_\theta\right)$ distance from w_θ to $span\{T - \eta\}$, equivalently the mean-squared error of the best linear regression of w_θ on T under P_θ, and the projected dynamics (37) to (38) is the closest parametric surrogate.

4.2 The exponential workhorse families

Family	Support	Sufficient statistic $T(x)$	Conjugate fitness (zero defect)
Bernoulli(p)	$\{0,1\}$	x	affine in x
Categorical(K)	$\{1,\dots,K\}$	one-hot indicators (dim. $K-1$)	class-wise constants
Poisson(λ)	$\mathbb{N}$	x	affine in x
Normal, known σ^2	$\mathbb{R}$	x	affine in x
Normal, unknown μ, σ^2	$\mathbb{R}$	$x,\ x^2$	quadratic in x
Exponential(λ)	$(0,\infty)$	x	affine in x
Gamma(α,β)	$(0,\infty)$	$\log x,\ x$	affine in $\log x,\ x$
Beta(α,β)	$(0,1)$	$\log x,\ \log(1-x)$	affine in $\log x,\ \log(1-x)$
Dirichlet(α)	Δ_{K-1}	$\log x_1, \dots, logx_K$	affine in $\{logx_i\}$
Multivariate normal	$\mathbb{R}^n$	$x,\ xx^\top$	quadratic in x
Wishart	$\mathcal{S}_{++}^p$	$logdetX,\ X$	affine in $logdetX,\ X$

For every family in the table the projected evolution is given by the universal exponential-family formulas (37) and (38): $\dot\eta = Cov_\theta(T, f)$ in moment coordinates and $\dot\theta = F^{-1}Cov_\theta(T, f)$ in natural coordinates. For one-statistic families this collapses to a single line, for example $\dot\lambda = Cov(X, f)$ and $\dot\theta = \lambda^{-1}Cov(X, f)$ for the Poisson, analogous formulas for the Bernoulli and known-variance normal, while the multivariate normal reproduces the general moment form (42). The two-statistic and matrix families require Fisher information in natural coordinates; formulas involving logarithmic statistics use the

trigamma function and are collected in Appendix B.

4.3 Beyond exponential families

Exact invariance is not special to exponential families. For *any* smooth family of strictly positive densities ρ_θ the score satisfies $\mathbb{E}_{\rho_\theta}[u_\theta] = 0$, a parameter velocity induces the relative velocity $\dot\theta^\top u_\theta$, and the defect, projection and best-fit velocity are given verbatim by (23) to (25) with u_θ in place of $T - \eta$. Consequently there is a zero-defect recipe: for any vector field $a : \Theta \to \mathbb{R}^m$ and scalar $b(\theta)$,

$$f(\rho_\theta, x) = a(\theta)^\top u_\theta(x) + b(\theta) \;\Rightarrow\; w_\theta = a(\theta)^\top u_\theta \in V_\theta, \quad \delta(\theta) \equiv 0, \quad \dot\theta = a(\theta). \tag{52}$$

Equation (52) should be read for what it is: a device for *constructing examples*, not a design principle. In an application the fitness is given by the problem and cannot be chosen to match the family. What (52) does provide is a concrete answer to the question of which fitness landscapes a given non-exponential family handles exactly, and for some families that answer is not obvious in advance. The content of this subsection is therefore the question of which families have tractable scores, and what conjugate class each one turns out to possess.

Location-scale families. Build $\rho_{\mu,\sigma}(x) = \frac{1}{\sigma} q(z)$, $z = (x-\mu)/\sigma$, from any smooth base density q, and write $\psi(z) := \frac{d}{dz} \log q(z)$ and $s = log\sigma$. The scores are

$$\partial_\mu \log \rho_{\mu,\sigma}(x) = -\frac{1}{\sigma}\psi(z), \qquad \partial_s \log \rho_{\mu,\sigma}(x) = -1 - z\psi(z), \tag{53}$$

so $f = a_1(\mu,\sigma)\left[-\frac{1}{\sigma}\psi(z)\right] + a_2(\mu,\sigma)\left[-1 - z\psi(z)\right] + b$ has zero defect and induces $(\dot\mu, \dot s) = (a_1, a_2)$. This umbrella contains many non-exponential families: Student-t at fixed ν, Cauchy, Laplace, logistic, and other heavy-tailed laws.

For Student-t_ν, $\log q(z) = const - \frac{\nu+1}{2}\log\left(1 + z^2/\nu\right)$ gives $\psi(z) = -(\nu+1)z/\left(\nu + z^2\right)$ and hence

$$\partial_\mu \log \rho = \frac{(\nu+1)z}{(\nu+z^2)\,\sigma}, \qquad \partial_s \log \rho = -1 + \frac{(\nu+1)z^2}{\nu+z^2}, \tag{54}$$

with the Cauchy case at $\nu = 1$. The conjugate class here is worth noting because it is qualitatively unlike the Gaussian case. The Student-t family is invariant not under polynomial fitness but under *rational* fitness of the form $\left[\alpha z + \beta z^2\right] / \left(\nu + z^2\right) + \gamma$. Quadratic fitness, the canonical invariant class for the Gaussian, has a strictly positive defect against a Student-t search distribution, and conversely a fitness that saturates like $z^2/\left(\nu + z^2\right)$ is handled exactly by Student-t and only approximately by a Gaussian. Since heavy-tailed search distributions are used in practice precisely for landscapes with bounded or saturating structure, this is a case where the conjugacy calculation matches the empirical folklore.

Finite mixtures. For $\rho_\theta(x) = \sum_{k=1}^K \pi_k q\left(x;\phi_k\right)$, which is not an exponential family, the responsibilities $r_k(x) := \pi_k q\left(x;\phi_k\right) / \sum_j \pi_j q\left(x;\phi_j\right)$ provide the scores. With softmax weights $\pi = softmax(\alpha)$,

$$\partial_{\alpha_k} \log \rho_\theta(x) = r_k(x) - \pi_k, \qquad \nabla_{\phi_k} \log \rho_\theta(x) = r_k(x) \nabla_{\phi_k} \log q\left(x;\phi_k\right), \tag{55}$$

so any fitness in the span of these responsibility-weighted scores is tangent on the mixture manifold. The invariant class is therefore basin-structured: mixtures can represent fitness that is constant or smoothly varying within components, but do not automatically represent sharp thresholds inside one component. Supplementary Experiment S4 illustrates this distinction.

Copulas and flows. The same recipe applies to copula models $\rho_\theta(x) = c_\theta(u)\prod_i f_i\left(x_i\right)$ with $u_i = F_i\left(x_i\right)$, where the copula-parameter score is $\nabla_\theta \log c_\theta(u)$, and to families generated by a finite-dimensional flow

g_θ, where P_θ is the pushforward of P_0 under g_θ and the score is obtained by differentiating $\log \rho_0\left(g_\theta^{-1}(x)\right) + log\left|\det Dg_\theta^{-1}(x)\right|$ in θ. In every case, once the score is available, (52) produces a zero-defect fitness class and its induced parameter flow, and (25) produces the defect for any other fitness.

5. Fitness shaping and shaped invariance

Sections 3 and 4 treat raw fitness. In practice, Evolution Strategies for optimization do not select on the raw objective but on a monotone, rank-based rewriting of it. This section develops the invariance theory for shaped selection. The central finding is that shaping changes the conjugacy problem: a family invariant for a raw linear or quadratic fitness is in general no longer invariant once the fitness is rank-shaped, because the shaped target leaves the tangent space. The remedy is to enlarge the family with score directions that live in rank space.

5.1 Quantile shaping

Fitness shaping replaces the raw value $f(x)$ by a monotone rank or quantile transform. For $X \sim \rho_\theta$, define the fitness-value CDF $G_\theta(y) = \mathbb{P}_\theta\left(f(X) \le y\right)$ and the population quantile $q_\theta(x) = G_\theta\left(f(x)\right)$. If $f(X)$ has a continuous distribution under ρ_θ, then $q_\theta(X) \sim Unif(0,1)$. Let $W : (0,1) \to \mathbb{R}$ be the shaping function; examples are linear rank shaping $W(u) = u$, elite top-τ selection $W(u) = \mathbf{1}\{u \ge 1-\tau\}$, and smooth elite selection $W(u) = \sigma\left(\left(u-(1-\tau)\right)/h\right)$ with σ logistic and $h > 0$. The shaped fitness is $a_\theta^{shape}(x) = W\left(q_\theta(x)\right)$, and the centered shaped relative fitness is

$$w_\theta(x) = W\left(q_\theta(x)\right) - \mathbb{E}_\theta\left[W\left(q_\theta(X)\right)\right] = W\left(q_\theta(x)\right) - \int_0^1 W(u)\, du, \tag{56}$$

the last equality following from uniformity of $q_\theta(X)$. The shaped replicator equation is $\partial_t \rho_t(x) = \rho_t(x) w_{\theta_t}(x)$ when the state is constrained to the family.

5.2 Tangent space and exact shaped invariance

A parameter velocity v induces the relative velocity $v^\top s_\theta(x)$ with $s_\theta = \nabla_\theta \log \rho_\theta$, so the relative tangent space is $\mathcal{T}_\theta = \{v^\top s_\theta : v \in \mathbb{R}^p\}$. The family is forward-invariant under the shaped fitness if and only if $w_\theta \in \mathcal{T}_\theta$, that is, if and only if the centered rank signal belongs to the score span. This is the shaped version of Theorem 1, with the raw centered fitness replaced by (56); conjugacy is therefore shaping-dependent.

5.3 The shaped defect and near-invariance

When exact invariance fails, the shaped invariance defect is the least-squares residual in $L^2\left(\rho_\theta\right)$,

$$\delta^2_{shape}(\theta) = \inf_{v \in \mathbb{R}^p} \mathbb{E}_\theta\left[\left(w_\theta(X) - v^\top s_\theta(X)\right)^2\right]. \tag{57}$$

With the Fisher/Gram matrix $F(\theta) = \mathbb{E}_\theta\left[s_\theta s_\theta^\top\right]$ and the shaped fitness-score covariance $b(\theta) = \mathbb{E}_\theta\left[w_\theta s_\theta\right]$, the normal equations are $F(\theta)v = b(\theta)$, the best velocity is $v^\star(\theta) = F^\dagger(\theta) b(\theta)$, and

$$\delta^2_{shape}(\theta) = \mathbb{E}_\theta\left[w_\theta^2\right] - b(\theta)^\top F^\dagger(\theta) b(\theta). \tag{58}$$

Because $q_\theta(X) \sim U(0,1)$, $\mathbb{E}_\theta\left[w_\theta^2\right] = Var_U\left(W(U)\right)$, a constant that does not depend on θ or on the objective. The normalized near-invariance score is therefore

$$R^2_{shape}(\theta) = \frac{b(\theta)^\top F^\dagger(\theta) b(\theta)}{Var_U\left(W(U)\right)} = 1 - \frac{\delta^2_{shape}(\theta)}{Var_U\left(W(U)\right)}. \tag{59}$$

Thus $R^2_{shape} = 1$ means exact shaped invariance and values near one mean near-invariance.

Theorem 3 carries over in instantaneous form. Holding the shaping map $W \circ q_\theta$ fixed at the current state, the full shaped replicator field increases the mean shaped signal at rate $Var_U(W(U))$, the projected flow increases it at rate $b^\top F^\dagger b$, and the difference is δ^2_{shape}. So R^2_{shape} is the fraction of the instantaneous shaped selection response that the family retains. Because the total is the same constant $Var_U(W(U))$ for every θ and every objective, shaped defects are directly comparable across landscapes in a way that raw defects are not.

Proposition 8 (monotone invariance). *Let m be a strictly increasing function and replace f by $m \circ f$. Then q_θ, w_θ, δ_{shape}, R^2_{shape} and the projected velocity $v^\star$ are all unchanged.*

Proof. $G_\theta^{m\circ f}(m(y)) = \mathbb{P}_\theta(m(f(X)) \le m(y)) = \mathbb{P}_\theta(f(X) \le y) = G_\theta^f(y)$ by strict monotonicity, so q_θ is unchanged pointwise, and every subsequent quantity is a function of q_θ and the family alone. ▫

Proposition 8 is the property that motivates shaping in the first place, and it distinguishes the shaped defect from the raw defect, which is not invariant under nonlinear monotone rescaling of the objective. When comparing families across problems with incommensurable fitness scales, one should use R^2_{shape}.

5.4 Projected shaped dynamics and trajectory error

The projected shaped parameter dynamics are $\dot\theta = F^\dagger(\theta) b(\theta)$. Let R_W denote the full shaped replicator field and P_{θ_t} its Fisher-projected path. Similarly to Section 3.10, the density residual has L^1 norm at most $\delta_{shape}(\theta)$, and if the shaped field is locally L_W-Lipschitz in L^1 on the region of interest, then

$$\| \rho_t - P_{\theta_t} \|_1 \le \int_0^t e^{L_W(t-s)} \delta_{shape}(\theta_s)\, ds \le \frac{\varepsilon}{L_W}\left(e^{L_W t} - 1\right) \tag{60}$$

when $\delta_{shape}(\theta_s) \le \varepsilon$, with the $L_W = 0$ limit εt. The caveats of Section 3.10 apply with equal force.

The Lipschitz assumption is not automatic for every shaping rule. A smooth shaping function combined with a regular fitness-value distribution generally gives a locally regular field. A hard elite indicator can change discontinuously when probability mass crosses the moving threshold, and a global Lipschitz constant may then fail without additional assumptions. The defect remains a valid instantaneous projection residual, but the finite-time statement should be invoked only where the shaped field has the required regularity, or after replacing the hard threshold by a smooth approximation.

5.5 Designing near-invariant families under shaping

The shaped problem is a regression problem: choose a family whose score functions explain the centered rank signal $w_\theta(x)$. Several constructions do this.

Rank-basis exponential tilts. Choose basis functions $\varphi_1, \dots, \varphi_K$ on $(0,1)$ and define the augmented family $\rho_{\theta,\eta}(x) \propto \rho_\theta(x) \exp\left(\sum_k \eta_k \varphi_k(q_\theta(x))\right)$. The additional score directions are $\varphi_k(q_\theta(x)) - \mathbb{E}_\theta[\varphi_k(q_\theta(X))]$, so if $W(u) - \int_0^1 W \approx \sum_k a_k \varphi_k(u)$, the family is near-invariant.

Rank-bin families. For elite selection, let $\varphi_j(u)$ equal 1 when u lies in the rank interval $[u_{j-1}, u_j)$ and 0 otherwise. This basis represents $W(u) = \mathbf{1}\{u \ge 1-\tau\}$ exactly when the elite threshold coincides with a bin boundary.

Hermite- or Legendre-rank augmented Gaussians. For a one-dimensional Gaussian with $Z = (X-\mu)/\sigma$, rank shaping produces $w_\theta(X) = W(\Phi(Z)) - \int_0^1 W$. The ordinary Gaussian tangent space is only $span\{Z, Z^2 - 1\}$; adding rank-basis directions gives $span\{Z, Z^2 - 1, \varphi_1(\Phi(Z)), \dots, \varphi_K(\Phi(Z))\}$, expanding the family toward shaped near-invariance.

RBF and flow-like dictionaries. Local features $\psi_j(x) = exp\left(-\| x - c_j \|^2 / 2h_j^2\right)$ give flexible tangent

dictionaries; the cost is that $F(\theta)$ can become ill-conditioned.

Mixtures. Responsibilities enter the score functions and help when rank-shaped selection splits a single mode into multiple basins, but, as Section 4.3 and Supplementary Experiment S4 show, mixture components alone do not capture sharp within-basin elite thresholds unless they are combined with local or rank features.

5.6 Adaptive tangent dictionaries versus fixed parametric families

A fixed collection of statistics $T_k(x)$ defines an ordinary exponential family and therefore an integrable statistical manifold. By contrast, functions such as $\varphi_k\left(q_\theta(x)\right)$ depend on the current distribution through the quantile map. They define *instantaneous score directions* and can be used in the projection and defect calculation at a given θ, but they do not automatically integrate to one fixed global family.

Two possible approaches can address this problem.

Re-anchored local families. Choose an anchor $\bar{\theta}$, hold its quantile map fixed, and define

$$P^{\bar{\theta}}_{\theta,\alpha}(x) \propto P_\theta(x) \exp\left(\sum_k \alpha_k \varphi_k\left(q_{\bar{\theta}}(x)\right)\right). \tag{61}$$

For each fixed $\bar{\theta}$ this *is* an ordinary exponential family in α, with fixed statistics $\varphi_k \circ q_{\bar{\theta}}$, so all of Sections 3 and 4 apply to it verbatim. At $\bar{\theta} = \theta$ and $\alpha = 0$ the added scores are the centered rank features used above. Re-anchoring after each step produces an adaptive sequence of local families, and the object being followed is a path in a bundle whose base is $\bar{\theta}$ and whose fiber is the exponential family (59). This is a well-defined construction, but it is not a proof that a single smooth manifold is forward-invariant.

The instantaneous reading. Nothing in Sections 3.4, 3.5, 5.3 or 5.4 requires the dictionary to integrate to a global family. The defect is defined by a subspace of $L^2\left(\rho_\theta\right)$ and a centered signal; the projected velocity is defined by the same data; Theorem 3 compares two instantaneous rates at a common state. All of these apply to a moving frame. An algorithm that recomputes its dictionary each iteration is therefore fully covered by the theory as an instantaneous projection method, and only claims about long-horizon forward invariance require the integrability that rank dictionaries lack.

Rank-bin indicators raise a second and separate issue: their discontinuity and linear dependence can make the Gram matrix singular. By Proposition 2(iii) the defect and the projected direction remain well defined through the pseudoinverse, but the coefficient vector does not. The numerical experiments accordingly distinguish three questions: representational error (the defect), numerical identifiability and conditioning of the coefficients, and finite-sample estimation error.

6. Numerical validation and empirical scope

The complete numerical experiments, figures, tables, and reproducibility settings are reported in the Supplementary Materials. The experiments are organized to separate four questions: whether a family can represent the instantaneous selection field, whether local residuals accumulate into trajectory error, whether the represented direction can be estimated stably from finite samples, and whether a lower defect is useful under a computational budget.

Supplementary Experiments S1, S8, and S9 verify the raw-fitness theory. Affine-quadratic fitness gives numerical-zero Gaussian defect and coincident full and projected trajectories in one and two dimensions; controlled nonquadratic perturbations increase both defect and tracking error. The dynamic rate calculation verifies the decomposition in Theorem 3 directly. Supplementary Experiment S11 confirms the local accuracy orders: sufficient-statistic moment error begins quadratically in time, while density error begins linearly.

Supplementary Experiments S2-S5 and S7 examine shaped selection and finite-sample estimation. Quantile shaping generally moves a raw linear or quadratic objective outside the Gaussian score span. Rank-based augmentations reduce the population defect, but flexible dictionaries may be ill-conditioned or sample-hungry. Exact representation and stable estimation are therefore distinct requirements.

Supplementary Experiment S10 tests the symmetric double-well stalling state. At exact symmetry the projected Gaussian velocity vanishes while the unrestricted flow continues to change density shape. The nonlinear analysis and the finite-time basin map show that this point is not an attracting basin: the variance direction attracts, but asymmetric mean perturbations escape cubically after a potentially long near-stalling transient.

Supplementary Experiments S6, S18, and S19 evaluate finite-time diagnostics. The rigorous Gronwall envelope is valid but extremely conservative. The accumulated defect orders the tested trajectories sensibly, yet its conversion into a numerical error requires horizon- and metric-specific calibration. Supplementary Experiments S12-S17 verify invariance properties, finite-time continuation, Student-t conjugacy, the selection-term decomposition for frequency-dependent fitness, and the favorable dimensional cost of structured rank features.

These experiments validate the theoretical identities and delimit their empirical use. They do not establish that reducing the defect necessarily improves fixed-budget optimization, and no such performance claim is made here. The defect measures faithfulness to a specified selection field; the usefulness of that field, sampling cost, conditioning, regularization, and update design remain separate algorithmic questions.

7. Synthesis

The continuous-trait replicator equation and information-geometric optimization are linked by a projection. A parametric score span defines the relative tangent space of a statistical family; projecting the centered fitness onto that span gives the Fisher/natural-gradient parameter direction; and the projection is an exact restriction of the full density dynamics if and only if the residual vanishes. This separates an exact identity from an approximation that is often written in the same notation.

The article's main contribution is the residual itself. The invariance defect is the minimum mean-square relative-velocity residual, and Theorem 3 shows that, for distribution-independent fitness, its square is the increase in mean fitness per unit time forfeited by confining the population to the family. Fisher's fundamental theorem therefore decomposes into an expressible component and a defect. Three consequences follow. The projected flow cannot reverse mean-fitness ascent in this setting. It becomes stationary when the fitness-score covariance vanishes, which can occur while fitness variance remains. In symmetric multimodal examples this produces exact stalling on the symmetry subspace and potentially long near-stalling transients, but the double-well calculation shows that the symmetric point need not attract asymmetric perturbations. Finally, Proposition 4 identifies the projection as moment closure: sufficient-statistic means agree through their first derivative while density shape generally separates at first order.

This also connects the construction to an older literature. The projected Gaussian equations are Lande's equation and its second-moment companion, and the Gaussian closure they embody has been debated in quantitative genetics since Turelli (1984) without a general measure of its cost. The defect supplies one, and R^2 expresses it as the fraction of the response to selection that the closure retains.

The exactness criterion is transparent for exponential families: the restricted fitness must be affine in the sufficient statistics, modulo an additive constant, which for the full Gaussian family is the affine-quadratic class. Proposition 5 explains why this is the same condition as Bayesian conjugacy, since for distribution-independent fitness the replicator flow is exponential tilting and the trajectory is an exponential geodesic. The dictionary in Section 4 extends the principle to location-scale, mixture, copula and flow-generated families; the Student-t case is a reminder that the conjugate class is not always polynomial.

How does this differ from previous work? Shahshahani geometry, Fisher natural gradients, IGO flows, Gaussian natural evolution strategies, Karev's exact reduction for exponential-family trait distributions, and least-squares approximations of replicator dynamics are established results. What is added here is the treatment of the non-invariant case: a necessary-and-sufficient tangency condition, an explicit residual with an interpretation as lost adaptation rate, an accuracy analysis in moments and density, a trajectory estimate for the same projection, a catalogue of compatible fitness classes, and a shaped-selection extension that exposes a mismatch between raw-objective conjugacy and algorithmic rank selection.

Fitness shaping shifts the conjugacy requirement. A raw quadratic objective is compatible with Gaussian closure, but elite selection replaces it by a thresholded rank signal that is not quadratic. This gives a principled interpretation of richer search distributions and feature dictionaries: they add tangent directions that can represent the algorithm's effective selection signal. The appropriate target is the shaped signal actually used by the optimizer, not the raw objective before ranking. Proposition 8 confirms that the shaped defect inherits the monotone invariance that makes rank-based methods attractive, which also makes it comparable across problems with incommensurable fitness scales.

7.1 Limitations

- The theory is deterministic and infinite-population. Finite populations add covariance-estimation error, rank noise, regularization bias, and time-discretization error. Supplementary Experiment S7 begins to quantify these effects but does not constitute a finite-sample convergence theory.
- The tangency theorem is local unless global continuation is established. Normalizability can fail in finite time, parameters can approach the boundary, and the Fisher matrix can become singular. Hard elite shaping can make the shaped field non-Lipschitz at moving thresholds. The instantaneous defect remains well defined under broad conditions, but the trajectory statement requires regularity on the region traversed.
- Rank-space augmentations require care. A moving quantile basis is an adaptive tangent dictionary or a re-anchored local family and should not be treated as one globally invariant parametric family unless its integrability has been proved. Fixed statistics, mixtures, and flow-based families do define global manifolds, but may require many parameters to approximate sharp rank boundaries.
- The defect-error relationship remains open in high dimension. The one- and two-dimensional tests in Supplementary Experiments S6 and S8 show consistent ordering, and Supplementary Experiment S17 studies dictionary cost up to dimension ten, but nothing here establishes a dimension-independent trajectory calibration.
- Finally, and perhaps most importantly, a small defect is not synonymous with a good optimizer. It measures the faithfulness of a family to a specified selection field. Optimization performance also depends on whether that field is useful, on step size, sampling, conditioning, regularization, and the cost of representing the family. The defect is best used as one element of a model-selection and adaptation criterion.

The projection framework also accommodates non-Euclidean trait spaces. On the circle or sphere, densities and scores may be defined relative to angular or surface measure; the tangency criterion and least-squares defect then use the same $L^2(\rho)$ inner product. Directional families such as the von Mises and von Mises–Fisher distributions, or the wrapped Cauchy family on the circle, provide examples. The wrapped Cauchy case is particularly relevant to hyperbolic parameter geometry, but its moving parameter-dependent scores require the general smooth-family formulation rather than an unqualified appeal to a fixed-statistic exponential family. Explicit conjugate fitness classes and defects for directional models remain for future work.

7.2 Practical implications

Enriching a search family to reduce its defect does not necessarily improve fixed-budget optimization. The defect can nevertheless support a diagnostic workflow: at the current search distribution, estimate the shaped signal, Fisher or Gram matrix, fitness-score covariance c, and defect for candidate families or tangent dictionaries. Reject numerically unsupported directions using singular values, regularization, or trust regions.

Corollary 3.2 provides a better trigger than a defect threshold alone. A large defect can coexist with rapid projected progress. But with a small represented velocity $\| c \|$ together with substantial total selection variance, the optimizer is moving little although selection pressure remains. Monitoring $\| c \|$, δ, and the total shaped variance distinguishes ordinary convergence from representational stalling or a long near-stalling transient. A rise in defect can also indicate emerging multimodality, a sharp elite boundary, or a local landscape component outside the current score span.

In summary, the defect states when the replicator and IGO descriptions coincide, quantifies the adaptation rate lost when they do not, and provides a principled reason to change the statistical family rather than treating that choice as fixed in advance.

Data and code availability

The code for reproducing all the experiments described in this paper are available at https://github.com/bonabeau-gif/ig-replicator.

AI Statement

During the preparation of this work, the author used Claude Code Opus 5 for support in coding simulations (initial code sketches generated from author's equations, verified and modified by author in Jupyter notebooks), packaging the GitHub repository, verifying the mathematical exposition and proofs, making writing suggestions for making the manuscript clearer, and redrawing Figures 1 and 2 from author's sketches. Research was performed using traditional bibliographic search. The author reviewed and edited the output as needed and takes full responsibility for the content of the published article.

References

Akimoto, Y., Nagata, Y., Ono, I., & Kobayashi, S. (2010). Bidirectional relation between CMA evolution strategies and natural evolution strategies. In *International Conference on Parallel Problem Solving from Nature*, 154–163. Springer.

Akimoto, Y., Nagata, Y., Ono, I., & Kobayashi, S. (2012). Theoretical foundation for CMA-ES from information geometry perspective. *Algorithmica* 64, 698–716.

Akin, E. (1979). *The Geometry of Population Genetics.* Lecture Notes in Biomathematics, vol. 31. Springer, Berlin–Heidelberg. https://doi.org/10.1007/978-3-642-93128-4

Akin, E. (1990). The differential geometry of population genetics and evolutionary games. In S. Lessard (Ed.), *Mathematical and Statistical Developments of Evolutionary Theory*, NATO ASI Series C, vol. 299, pp. 1–93. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-0513-9_1

Akyıldız, Ö. D. (2017). A probabilistic interpretation of replicator-mutator dynamics. arXiv:1712.07879.

Amari, S. (1998). Natural gradient works efficiently in learning. *Neural Computation* 10(2), 251–276.

Amari, S., & Nagaoka, H. (2000). *Methods of Information Geometry.* Translations of Mathematical Monographs 191. American Mathematical Society.

Arora, S., Hazan, E., & Kale, S. (2012). The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing* 8, 121–164.

Beyer, H.-G. (2014). Convergence analysis of evolutionary algorithms that are based on the paradigm of information geometry. *Evolutionary Computation* 22(4), 679–709.

Beyer, H.-G., & Sendhoff, B. (2017). Simplify your covariance matrix adaptation evolution strategy. *IEEE Transactions on Evolutionary Computation* 21(5), 746–759.

Bomze, I. (1991). Cross entropy minimization in uninvadable states of complex populations. *Journal of Mathematical Biology* 30, 73–87.

Bomze, I. M., Pelillo, M., & Stix, V. (2000). Approximating the maximum weight clique using replicator dynamics. *IEEE Transactions on Neural Networks* 11, 1228–1241.

Bürger, R. (1991). Moments, cumulants, and polygenic dynamics. Journal of Mathematical Biology 30(2), 199–213. https://doi.org/10.1007/BF00160336

Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation.* Wiley.

Carrillo, J. A., Chen, Y., Huang, D. Z., Huang, J., & Wei, D. (2026). Fisher–Rao gradient flow: Geodesic convexity and functional inequalities. *SIAM Journal on Mathematical Analysis* 58(2), 1062–1099. https://doi.org/10.1137/25M176252X

Cressman, R. (2005). Stability of the replicator equation with continuous strategy space. *Mathematical Social Sciences* 50, 127–147.

Cressman, R., & Hofbauer, J. (2005). Measure dynamics on a one-dimensional continuous trait space: Theoretical foundations for adaptive dynamics. *Theoretical Population Biology* 67, 47–59.

Cressman, R., Hofbauer, J., & Riedel, F. (2006). Stability of the replicator equation for a single species with a multi-dimensional continuous trait space. *Journal of Theoretical Biology* 239, 273–288.

Czégel, D., Zachar, I., & Szathmáry, E. (2019). Multilevel selection as Bayesian inference, major transitions in individuality as structure learning. *Royal Society Open Science* 6(8), 190202. https://doi.org/10.1098/rsos.190202

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Clarendon Press.

Hansen, N. (2006). The CMA evolution strategy: A comparing review. In J. A. Lozano, P. Larrañaga, I. Inza, & E. Bengoetxea (Eds.), *Towards a New Evolutionary Computation*, 75–102. Springer.

Harper, M. (2009). The replicator equation as an inference dynamic. arXiv:0911.1763.

Hofbauer, J., & Sigmund, K. (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society* 40, 479–519.

Jaćimović, V. (2023). The fundamental theorem of natural selection in optimization and games. BioSystems 230, 104956. https://doi.org/10.1016/j.biosystems.2023.104956

Jaćimović, V. (2024). Natural gradient ascent in evolutionary games. *BioSystems* 236, 105127.

Karev, G. (2020). Dynamics of strategy distributions in a one-dimensional continuous trait space for games with a quadratic payoff function. *Games* 11(1), 14.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33, 402–416.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution* 37, 1210–1226.

Larrañaga, P., & Lozano, J. A. (2002). *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation.* Kluwer.

Lehtonen, J. (2018). The Price equation, gradient dynamics, and continuous trait game theory. *American Naturalist* 191, 146–153.

Malagò, L., & Pistone, G. (2015). Information geometry of the Gaussian distribution in view of stochastic optimization. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, 150–162.

Menon, A. R., Mehrotra, K., Mohan, C. K., & Ranka, S. (1995). Optimization using replicators. In L. J. Eshelman (Ed.), *Proceedings of the 6th International Conference on Genetic Algorithms*, 209–216. Morgan Kaufmann.

Miskin, M. Z., Khaira, G., de Pablo, J. J., & Jaeger, H. M. (2016). Turning statistical physics models into materials design engines. *Proceedings of the National Academy of Sciences* 113(1), 34–39.

Müller, N., & Glasmachers, T. (2021). Non-local optimization: Imposing structure on optimization problems by relaxation. In *Proceedings of the 16th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*, Article 6,

pp. 1–10. https://doi.org/10.1145/3450218.3477307

Mustonen, V., & Lässig, M. (2010). Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences USA* 107, 4248–4253.

Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life.* Belknap Press of Harvard University Press.

Oechssler, J., & Riedel, F. (2001). Evolutionary dynamics on infinite strategy space. *Economic Theory* 17, 141–162.

Oechssler, J., & Riedel, F. (2002). On the dynamic foundation of evolutionary stability in continuous models. *Journal of Economic Theory* 107, 223–252.

Ollivier, Y., Arnold, L., Auger, A., & Hansen, N. (2017). Information-geometric optimization algorithms: A unifying picture via invariance principles. *Journal of Machine Learning Research* 18, 1–65.

Otwinowski, J., LaMont, C. H., & Nourmohammad, A. (2020). Information-geometric optimization with natural selection. *Entropy* 22, 967.

Page, K. M., & Nowak, M. A. (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology* 219, 93–98.

Pathiraja, S., & Wacker, P. (2025). Connections between sequential Bayesian inference and evolutionary dynamics. *Philosophical Transactions of the Royal Society A* 383(2298), 20240241. https://doi.org/10.1098/rsta.2024.0241

Pelillo, M. (1999). Replicator equations, maximal cliques, and graph isomorphism. *Neural Computation* 11(8), 1933–1955. https://doi.org/10.1162/089976699300016034

Price, G. R. (1970). Selection and covariance. *Nature* 227, 520–521.

Price, G. R. (1972). Fisher's "fundamental theorem" made clear. *Annals of Human Genetics* 36, 129–140.

Queller, D. C. (2017). Fundamental theorems of evolution. *American Naturalist* 189, 345–353.

Raab, R., de Alfaro, L., & Liu, Y. (2022). Conjugate natural selection. arXiv:2208.13898.

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Production* 8, 95–108.

Rota Bulò, S., Pelillo, M., & Bomze, I. M. (2011). Graph-based quadratic optimization: A fast evolutionary approach. *Computer Vision and Image Understanding* 115, 984–995.

Savarino, F., Albers, P., & Schnörr, C. (2024). On the geometric mechanics of assignment flows for metric data labeling. Information Geometry 7, 1–31. https://doi.org/10.1007/s41884-023-00120-1

Shahshahani, S. (1979). *A New Mathematical Framework for the Study of Linkage and Selection*. Memoirs of the American Mathematical Society, vol. 17, no. 211. American Mathematical Society.

Shalizi, C. R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics* 3, 1039–1074.

Sigmund, K. (1984). *The Maximum Principle for Replicator Equations.* IIASA Working Paper WP-84-056, Laxenburg, Austria.

Smerlak, M., & Youssef, A. (2017). Limiting fitness distributions of evolutionary dynamics. *Journal of Theoretical Biology* 416, 68–80.

Taylor, P., & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40, 145–156.

Transtrum, M. K., Machta, B. B., & Sethna, J. P. (2011). Geometry of nonlinear least squares with applications to sloppy models and optimization. *Physical Review E* 83(3), 036701.

Tsimring, L. S., Levine, H., & Kessler, D. A. (1996). RNA virus evolution via a fitness-space model. *Physical Review Letters* 76, 4440–4443.

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology* 25, 138–193.

van der Hoeven, D., van Erven, T., & Kotłowski, W. (2018). The many faces of exponential weights in online learning. In *Proceedings of the 31st Conference on Learning Theory*, Proceedings of Machine Learning Research, vol. 75, pp. 2067–2092. PMLR.

Wierstra, D., Schaul, T., Glasmachers, T., Sun, Y., Peters, J., & Schmidhuber, J. (2014). Natural evolution strategies. *Journal of Machine Learning Research* 15, 949–980.

Appendix A. Analytic assumptions and trajectory bounds

A.1 The L1/total-variation estimate for the Fisher projection

Let $\mathcal{P}$ be a set of probability densities in $L^1(\mu)$ and let $R(\rho) = \rho\left(f(\rho,\cdot) - \bar{f}(\rho)\right)$ with $\bar{f}(\rho) = \int f(\rho,\cdot)\rho\, d\mu$. Assume on a region $U \subset \mathcal{P}$ that: (A1) the true and projected paths are strong solutions and remain in U; (A2) $\parallel f(\rho,\cdot) \parallel_\infty \leq M$; (A3) $\parallel f(\rho,\cdot) - f(\sigma,\cdot) \parallel_\infty \leq K \parallel \rho - \sigma \parallel_1$; (A4) the parametric map is differentiable along the projected path; and (A5) the Fisher projection exists, using a pseudoinverse when necessary.

The mean-fitness difference satisfies $\left|\bar{f}(\rho) - \bar{f}(\sigma)\right| \leq M \parallel \rho - \sigma \parallel_1 + K \parallel \rho - \sigma \parallel_1$. Writing the field difference as

$$R(\rho) - R(\sigma) = (\rho - \sigma)\left(f(\rho,\cdot) - \bar{f}(\rho)\right) + \sigma\left[(f(\rho,\cdot) - f(\sigma,\cdot)) - \left(\bar{f}(\rho) - \bar{f}(\sigma)\right)\right]$$

gives

$$\parallel R(\rho) - R(\sigma) \parallel_1 \leq (3M + 2K) \parallel \rho - \sigma \parallel_1 . \tag{A1}$$

Along the projected family, $\dot{P}_\theta = P_\theta\left(v^{\star\top} u_\theta\right)$ and

$$R\left(P_\theta\right) - \dot{P}_\theta = P_\theta r_\theta, \qquad \parallel P_\theta r_\theta \parallel_1 \leq \parallel r_\theta \parallel_{L^2(P_\theta)} = \delta(\theta). \tag{A2}$$

For $e(t) = \parallel \rho_t - P_{\theta_t} \parallel_1$, the standard norm-derivative inequality yields $D^+ e \leq (3M + 2K)e + \delta\left(\theta_t\right)$, and Equation (49) follows. If only a local Lipschitz constant is known, the same argument holds up to the first exit time from the region where the assumptions are valid. Total variation is $e/2$.

For a distribution-independent fitness on bounded support, $K = 0$. Because adding a constant to f leaves the replicator field unchanged, the uniform bound can be minimized by centering the range, giving $M = \left(\sup f - inf f\right)/2$. This improves the constant but does not make the exponential envelope sharp, as Supplementary Experiment S6 shows.

A one-sided refinement. The looseness of (49) comes from applying a two-sided Lipschitz constant uniformly. Replacing (A1) by a one-sided condition, $\langle R(\rho) - R(\sigma),\, sgn(\rho - \sigma)\rangle \leq l \parallel \rho - \sigma \parallel_1$ with l permitted to be negative, gives $D^+ e \leq le + \delta$ and hence $e(t) \leq \int_0^t e^{l(t-s)}\delta\left(\theta_s\right) ds$, which does not grow when $l \leq 0$.

Supplementary Experiment S18 estimates the one-sided constant numerically on the region traversed in Supplementary Experiment S6 by sampling smooth normalized perturbations at three magnitudes. The estimate remains positive throughout, so the tested flow is not contracting in L^1 on that region and the exponential form of the envelope survives. It is nevertheless far smaller than the two-sided constant, producing a much tighter local envelope. Whether the one-sided constant becomes negative near other stable rest points, and whether a certified rather than sampled bound can be established on a given region, remain open.

A.2 Shaped fitness

For shaped fitness, replace $f(\rho,\cdot) - \bar{f}(\rho)$ by the centered shaped signal w_ρ^W. If the shaped field $R_W(\rho) = \rho w_\rho^W$ is locally Lipschitz in L^1, the proof is unchanged and gives Equation (58). Smooth shaping functions and regular fitness-value distributions often yield local regularity. Hard elite indicators require additional assumptions, because an arbitrarily small change in the distribution can move mass across the selection threshold. The paper therefore uses the shaped defect unconditionally as an instantaneous residual but states finite-time tracking conditionally on shaped-field regularity.

A.3 Smoothness and dominated convergence

All differentiations under the integral sign require the usual domination and regularity conditions. The score functions must be square-integrable under the current density, the required fitness moments must be finite, and the parametric map must be differentiable in the topology used for the trajectory statement. Twice-continuous differentiability is sufficient for the smooth examples in the paper but is not necessary. Pseudoinverse projections remain well defined when the represented tangent subspace is closed, although coefficient vectors can be non-unique at singular points. For Proposition 5, the cumulant generating function $K(t) = log\mathbb{E}_{\rho_0}\left[e^{tf}\right]$ is finite on an interval containing 0 whenever f is bounded above ρ_0-a.s., and the stated identities hold on the interior of that interval by standard properties of exponential families.

Appendix B. Fisher information for the two-statistic and matrix families

For the Gamma family with $T = (\log x, x)$,

$$F = \begin{pmatrix} Var\left(\log X\right) & Cov\left(\log X, X\right) \\ Cov\left(\log X, X\right) & Var(X) \end{pmatrix} = \begin{pmatrix} \psi_1(\alpha) & 1/\beta \\ 1/\beta & \alpha/\beta^2 \end{pmatrix}, \tag{C1}$$

for the Beta family with $T = (\log x, log(1-x))$,

$$F = \begin{pmatrix} \psi_1(\alpha) - \psi_1(\alpha+\beta) & -\psi_1(\alpha+\beta) \\ -\psi_1(\alpha+\beta) & \psi_1(\beta) - \psi_1(\alpha+\beta) \end{pmatrix}, \tag{C2}$$

and for the Dirichlet family with $T = (\log x_1, \ldots, logx_K)$ and $\alpha_0 = \sum_i \alpha_i$,

$$F_{ij} = \psi_1\left(\alpha_i\right)\delta_{ij} - \psi_1\left(\alpha_0\right), \tag{C3}$$

where $\psi_1 = \psi'$ is the trigamma function, so that $\dot{\theta} = F^{-1}Cov_\theta(T, f)$ in each case, with $\theta = (\alpha - 1, -\beta)$ for the Gamma, $\theta = (\alpha - 1, \beta - 1)$ for the Beta, and $\theta_i = \alpha_i - 1$ for the Dirichlet.

For the Wishart family on the cone $\mathcal{S}_{++}^p$, written in the exponential form $P_{\theta_1,\Theta_2}(X) \propto exp\left(\theta_1 logdetX + \langle\Theta_2, X\rangle\right)$ with $\langle A, B\rangle = tr(AB)$ and $c := \theta_1 + \frac{p+1}{2}$, the mean statistics needed to center T are

$$\mathbb{E}[X] = -c\,\Theta_2^{-1}, \qquad \mathbb{E}[logdetX] = \psi_p(c) - logdet\left(-\Theta_2\right), \tag{C4}$$

where ψ_p is the multivariate digamma. The standard Wishart $W_p(\nu, V)$ corresponds to $\theta_1 = \frac{\nu-p-1}{2}$, $\Theta_2 = -\frac{1}{2}V^{-1}$, $c = \nu/2$, giving $\mathbb{E}[X] = \nu V$. The conjugate fitness is $f\left(P_\theta, X\right) = a_1(\theta)logdetX + \langle A_2(\theta), X\rangle + b(\theta)$.

Appendix C. The distinct ordinary-L2 density-velocity projection

A second, mathematically legitimate projection minimizes

$$\tilde{\delta}(\theta) = \inf_v \parallel DP_\theta[v] - R\left(P_\theta\right) \parallel_{L^2(\mu)}. \tag{D1}$$

Its Gram matrix contains $P_\theta^2 u_\theta u_\theta^\top$, whereas the Fisher projection contains $P_\theta u_\theta u_\theta^\top$. The two coincide in special cases but differ in general, and only the Fisher projection has the interpretation of Theorem 3, since only it measures the residual in the metric in which the replicator field is a Fisher–Rao gradient. The surrogate labelled *proj* in Supplementary Experiment S1 uses (D1) and is retained in Supplementary

Figures S1 and S2 as a comparison. That $\tilde{\delta}$-based surrogate is better in L^1 and worse in KL than the Fisher projection on the pure-sinusoid case is a consequence of the different inner products the two projections optimize, and should not be read as evidence for either as an approximation to the replicator flow.

Like the Fisher projection, the ordinary-L^2 density-velocity projection is invariant under a smooth invertible reparameterization of the same family, because such a change of coordinates leaves the density tangent subspace unchanged. Its limitations are different: it depends on the chosen reference measure, it does not use the Fisher-Rao metric, and it has no analogue of the lost-adaptation identity in Theorem 3. On a finite reference measure or bounded support, standard boundedness and Lipschitz assumptions yield an L^2 Gronwall estimate. For full-support Gaussian families, the analysis can be localized or formulated in a weighted L^2 space whose reference probability has sufficiently heavy tails. These alternatives remain distinct from the Fisher projection unless their Gram matrices happen to agree.

Supplementary materials

The invariance defect of continuous-trait parametric replicator dynamics

Eric Bonabeau

School of Complex Adaptive Systems, Arizona State University,
Tempe, AZ 85281-2701, USA

ebonabea@asu.edu

Supplementary experiments and numerical methods

Experiments S1-S8 provide the primary numerical validation of exact invariance, shaped selection, multi-step tracking, finite-sample estimation, and a two-dimensional nonquadratic landscape. Experiments S9-S19 provide direct tests of the rate identity, symmetric stalling, convergence orders, coordinate and monotone invariance, finite-time continuation, Student-t conjugacy, frequency-dependent selection, dimensional dictionary cost, one-sided trajectory bounds, and the separation between family approximation and dynamic drift.

The experiments are numbered consecutively and reported without claims of fixed-budget optimization superiority. Their purpose is to test the mathematical identities, identify numerical failure modes, and delimit what the defect can and cannot diagnose. Complete numerical settings are collected in Supplementary Section S8.

S1. Raw-fitness exactness and near-exactness

To make the distinction between exact, projected and true dynamics concrete, consider a one-dimensional problem with Gaussian initial population $\rho_0 = \mathcal{N}(\mu_0, \sigma_0^2)$ and fitness $f(x)$ depending only on the trait. By Proposition 5 the true replicator trajectory is available in closed form as the exponential tilt $\rho_t \propto \rho_0 e^{tf}$, integrated on a fine grid. Against it we compare the Fisher-projected Gaussian surrogate, labelled *param*, which evolves the moments by the $L^2(\rho\, dx)$ projection on relative velocities, equivalently the moment-closure ODE of Proposition 4 evaluated under the current Gaussian:

$$\dot{\mu} = \mathbb{E}_g\left[(X-\mu)f(X)\right], \qquad \dot{\sigma}^2 = \mathbb{E}_g\left[(X-\mu)^2 f(X)\right] - \sigma^2 \mathbb{E}_g\left[f(X)\right]. \tag{S1}$$

with $g = \mathcal{N}(\mu, \sigma^2)$. This is the moment form (42) in one dimension. A second, mathematically distinct surrogate labelled *proj*, which minimizes an ordinary $L^2(dx)$ distance between density velocities rather than relative velocities, is retained in the figures for comparison and analyzed in Appendix C of the main manuscript. It is not the object of the theorems in this paper.

The surrogate coincides with the truth exactly when f is affine-quadratic (Theorem 7), and departs as f pushes the true density away from Gaussianity. Three fitness functions span the range, with errors measured at $t = 2$ (Figures S1 and S2):

- **Quadratic f.** All trajectories are essentially identical; the small residual is grid discretization. The L^1 density error is about 0.00195 for both surrogates and the measured invariance defect is about 10^{-11}, numerical zero, as predicted by exact invariance.

- **Quadratic plus a small sinusoid.** The true density becomes mildly non-Gaussian and both surrogates deviate modestly: L^1 errors are 0.126 (*param*) and 0.137 (*proj*), and $KL\left(\text{true} \parallel \cdot\right)$ is 0.0100 (*param*) versus 0.0163 (*proj*).
- **Pure sinusoid.** The true density becomes strongly non-Gaussian and both surrogates fail, in different metrics: *proj* is better in L^1 (0.478 versus 0.902) while *param* is better in KL (0.570 versus 2.35). The asymmetry reflects the different inner products the two projections optimize and is discussed in Appendix C of the main manuscript.

The moment trajectories in Figure S2 track visibly better than the densities in Figure S1 for the perturbed cases. Corollary 4.1 explains the difference: the sufficient-statistic means share the correct initial derivative, so their error begins at order t^2, whereas the density error begins at order t with coefficient $\mathbb{E}\left|r_0\right| \leq \delta$. A nonzero defect therefore changes density shape before it produces comparable location and spread errors.

Analytic Gaussian defect for a bounded nonquadratic fitness. The explicit Gaussian formula in Section 3.9 of the main article also permits an independent calculation at any centered Gaussian state. For $X \sim \mathcal{N}\left(0, \sigma^2\right)$ and bounded sinusoidal fitness $f(x) = \varepsilon \sin(kx)$, oddness gives $Cov\left(f(X), (X-\mu)^2\right) = 0$, while the characteristic function of the Gaussian gives Eq. (S2):

$$Var\left(f(X)\right) = \frac{\varepsilon^2}{2}\left(1 - e^{-2k^2\sigma^2}\right), \qquad Cov\left(f(X), X\right) = \varepsilon k \sigma^2 e^{-k^2\sigma^2/2}. \tag{S2}$$

Substitution into the univariate defect formula, Eq. (47) of the main article, yields Eq. (S3):

$$\delta^2 = \varepsilon^2\left[\frac{1 - e^{-2k^2\sigma^2}}{2} - k^2\sigma^2 e^{-k^2\sigma^2}\right]. \tag{S3}$$

For small $k\sigma$, the squared defect is $\varepsilon^2 k^6 \sigma^6/6 + O\left(\varepsilon^2 k^8 \sigma^8\right)$. The lower-order components of the sine are absorbed by the Gaussian mean direction; the residual begins with the cubic Hermite component. This is an instantaneous calculation at a specified Gaussian state, not a claim about the accumulated error at time t. It analytically explains why the nonquadratic perturbations in Figure S3 generate a nonzero Gaussian defect while a pure quadratic fitness does not.

Figure S3 shows the relationship between the instantaneous defect and departure from the Gaussian family. The quadratic case has numerical-zero defect and exact tracking; the nonquadratic cases have larger defects and larger trajectory errors.

Figure S4 illustrates the finite-time validity of Section 3.11. For positive quadratic curvature the exponential tilt remains normalizable only before the blow-up time (50), which by Proposition 5 is the point at which the cumulant generating function of f under ρ_0 ceases to be finite.

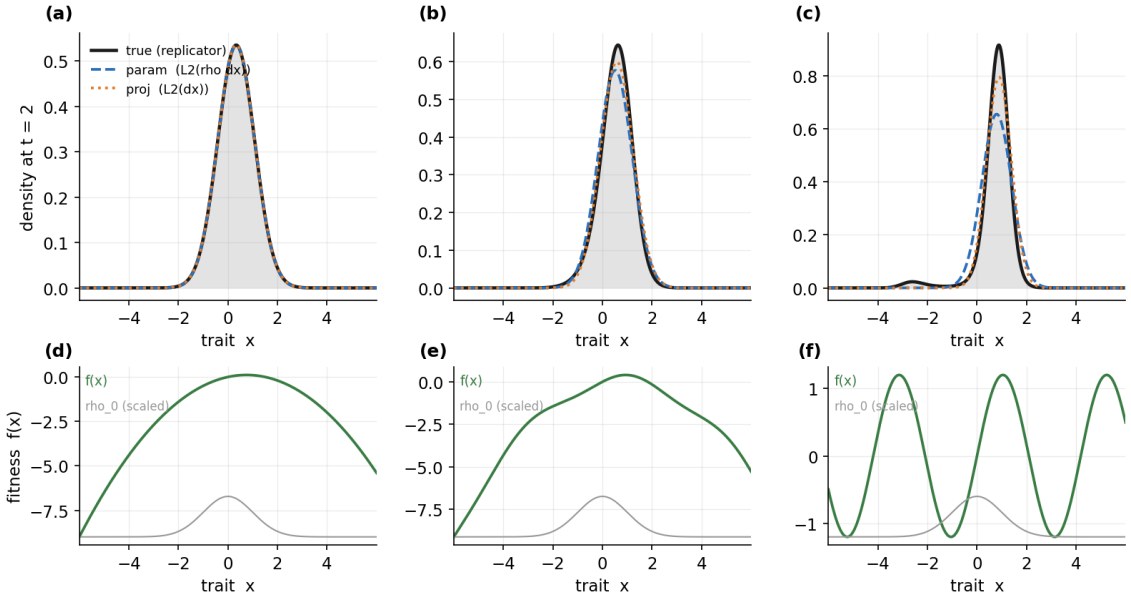


Figure S1. Replicator truth versus the two Gaussian surrogates at $t = 2$. (a–c) The true replicator density $\rho_t \propto \rho_0 e^{tf}$ (black, shaded) compared with the param surrogate (blue dashed, the $L^2(\rho\, dx)$ projection on relative velocities, Eq. (S1)) and the projected surrogate (orange dotted, the $L^2(dx)$ projection on density velocities, Appendix C of the main manuscript), for the three fitness functions. (d–f) The corresponding fitness $f(x)$ (green) with the initial density ρ_0 shown scaled (grey). The curves coincide for the quadratic fitness (a), deviate modestly under the small sinusoidal perturbation (b), and both surrogates fail under the pure sinusoid (c), where the true density turns bimodal. Parameters: $\mu_0 = 0$, $\sigma_0^2 = 1$, $t = 2$.

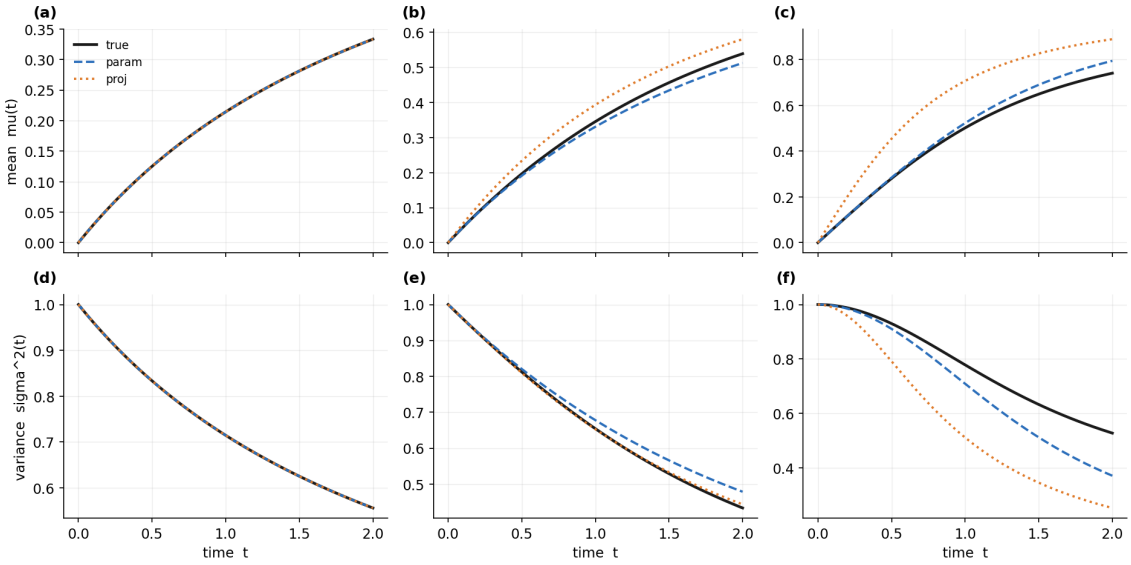


Figure S2. Moment trajectories: exact replicator moments (black, computed from the tilted density on the grid) versus the surrogate ODEs. (a–c) mean $\mu(t)$; (d–f) variance $\sigma^2(t)$, for the three fitness functions. The moments coincide for the quadratic fitness; for the sinusoidal cases the param surrogate (blue dashed) tracks the true moments more closely than proj (orange dotted). The overall closeness of the moment trajectories relative to the density errors in Figure S1 is the content of Corollary 4.1. Parameters: $\mu_0 = 0$, $\sigma_0^2 = 1$, $t \in [0, 2]$.

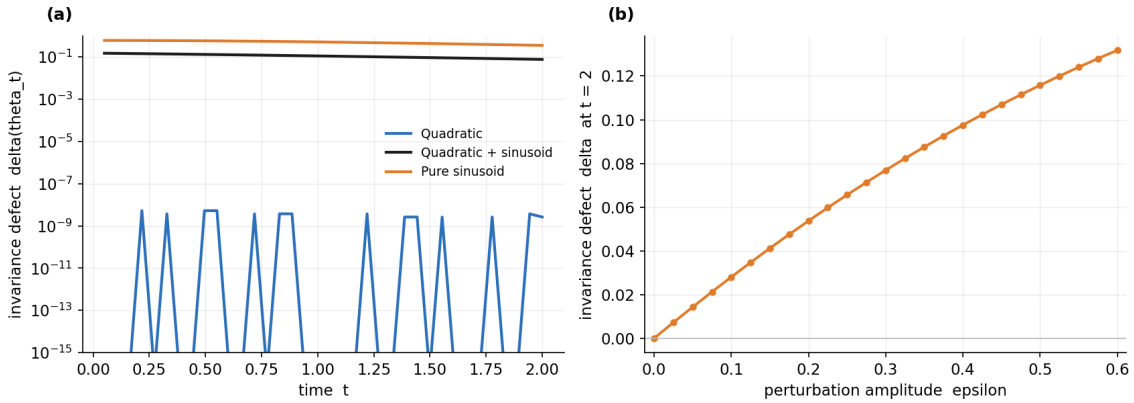


Figure S3. The invariance defect $\delta(\theta)$ (Eq. (25)), the $L^2(P_\theta)$ distance from the centered fitness to the Gaussian tangent space. (a) Along the quadratic trajectory it remains at the floating-point floor, confirming exact forward invariance, whereas the two nonquadratic fitnesses produce nonzero defects. (b) For $f(x) = 0.30x - 0.20x^2 + \epsilon \sin(1.5x)$, the defect at $t = 2$ vanishes at $\epsilon = 0$ and grows smoothly with ϵ. Parameters: $\mu_0 = 0$, $\sigma_0^2 = 1$, $\epsilon \in [0, 0.6]$ over 25 points.

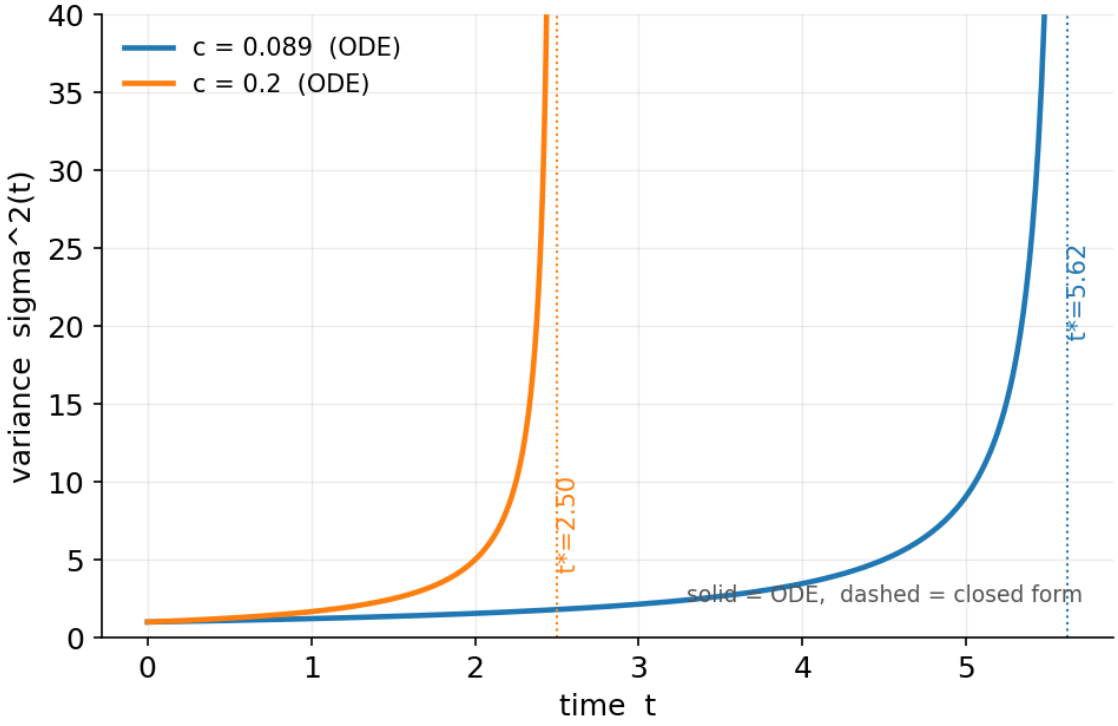


Figure S4. Finite-time validity under $f(x) = cx^2$ with $c > 0$ (Section 3.11, Eq. (50)). The variance equation $\dot{\sigma}^2 = 2c\sigma^4$ gives $\sigma^2(t) = \sigma_0^2/(1 - 2c\sigma_0^2 t)$ and blows up at $t^\star = 1/(2c\sigma_0^2)$. The solid numerical trajectories and dashed closed forms agree up to $0.985\, t^\star$. Parameters: $\sigma_0^2 = 1$; $c = 0.089$ gives $t^\star = 5.62$ and $c = 0.20$ gives $t^\star = 2.50$.

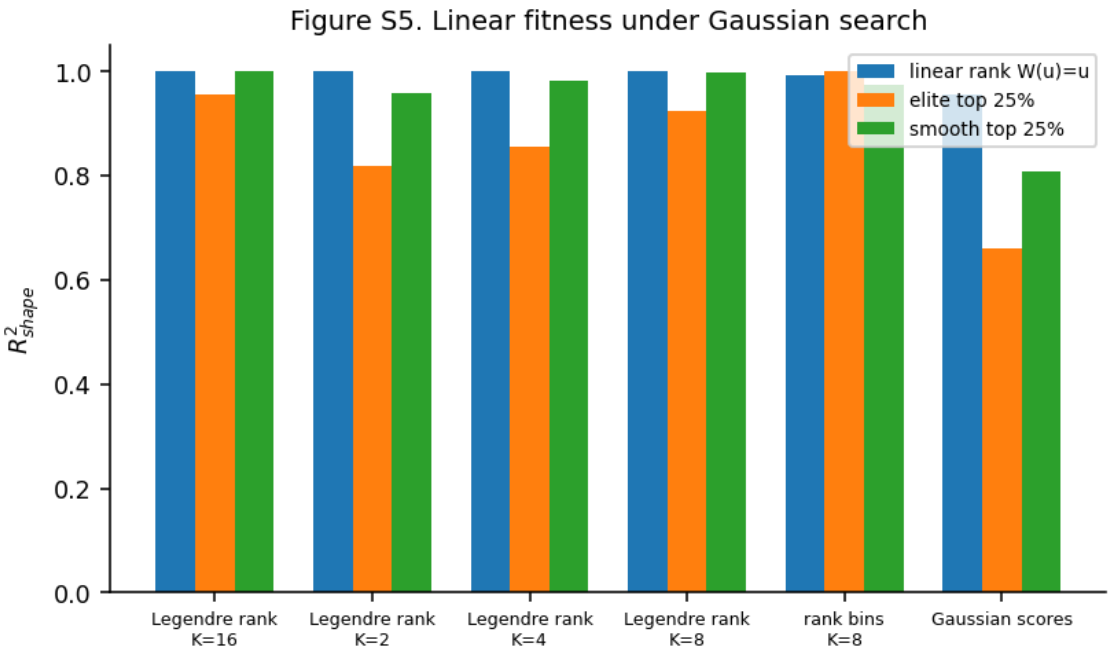


Figure S5. Experiment S2: linear fitness under Gaussian search. Explained shaped variance for the three dictionaries under linear, elite, and smooth-elite shaping.

S2. Shaped fitness

Experiments S2 to S5 used one-dimensional quadrature on a dense grid of 24,001 points on $[-10, 10]$. Densities were normalized on the grid, the quantile $q_\theta(x)$ was computed as the weighted empirical CDF of the raw fitness values, and score dictionaries were centered under the current density. Projection used the pseudoinverse of the empirical Fisher/Gram matrix. The diagnostic throughout is $R^2_{shape} = 1 - \delta^2_{shape}/Var_U(W)$ from Equation (57).

Experiment S2: linear fitness under Gaussian search. With $X \sim \mathcal{N}(0, 1)$ and $f(x) = x$, three score dictionaries were compared: ordinary Gaussian scores $\{Z, Z^2 - 1\}$; Gaussian plus Legendre-rank features; and Gaussian plus rank-bin features. The ordinary Gaussian tangent space explains most of the linear-rank signal but not the elite-shaped signal (Table S1). For top-25% shaping the Gaussian scores reach only about 0.661 explained shaped variance; rank-bin augmentation reaches numerical exactness because the target is itself a rank-bin function, and Legendre-rank augmentation improves steadily with K.

For the elite case with $\tau = 0.25$ and $f(x) = x$ the ordinary Gaussian case is analytic. With $a_\tau = \Phi^{-1}(1 - \tau)$, projection onto $span\{Z, Z^2 - 1\}$ gives explained variance $\phi(a_\tau)^2 + \frac{1}{2}a_\tau^2\phi(a_\tau)^2$ and the squared defect in Eq. (S4):

$$\delta^2_{shape} = \tau(1 - \tau) - \phi(a_\tau)^2 - \frac{1}{2}a_\tau^2\phi(a_\tau)^2, \tag{S4}$$

The value in Eq. (S4) is strictly positive for typical elite fractions. Nonlinear rank shaping breaks exact Gaussian invariance even for a raw linear fitness. This is the simplest complete statement of the paper's second theme: the raw objective is in the conjugate class, the algorithm's effective objective is not.

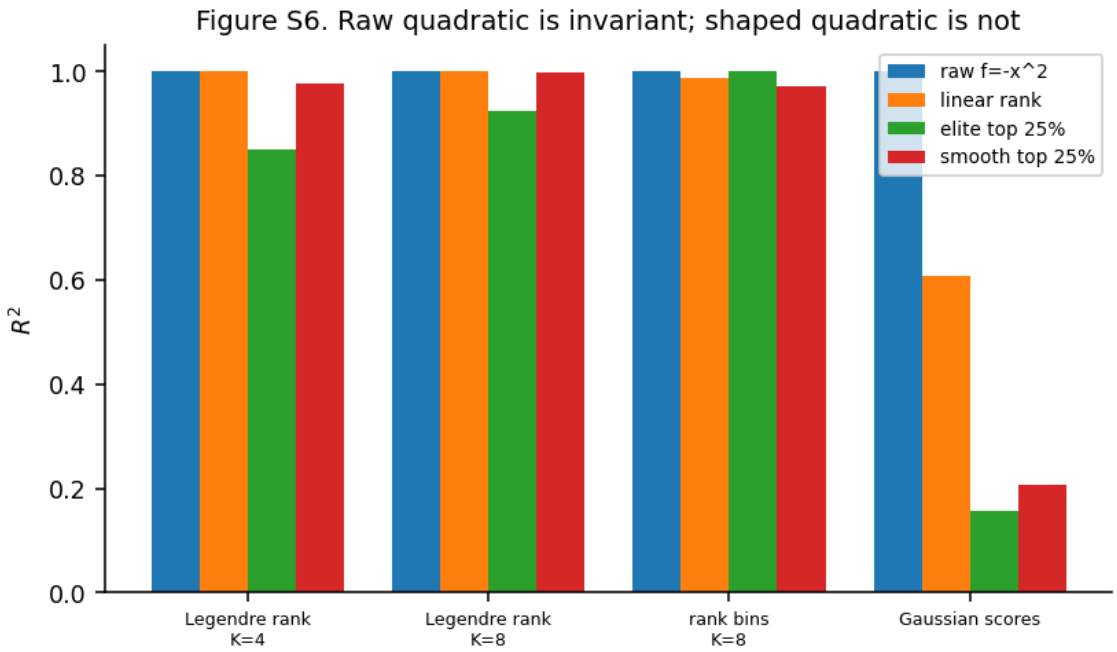


Figure S6. Experiment S3: the raw quadratic is Gaussian-invariant, while the shaped quadratic is not.

Table S1. Explained shaped variance R^2_{shape} for Experiment S2.

family	linear rank $W(u) = u$	elite top 25%	smooth top 25%
Gauss + Legendre rank, $K = 16$	1	0.956	1
Gauss + Legendre rank, $K = 2$	1	0.819	0.959
Gauss + Legendre rank, $K = 4$	1	0.856	0.981
Gauss + Legendre rank, $K = 8$	1	0.925	0.999
Gauss + rank bins, $K = 8$	0.994	1	0.975
Gaussian scores	0.955	0.661	0.81

Experiment S3: quadratic fitness before and after shaping. With $X \sim \mathcal{N}(0, 1)$ and $f(x) = -x^2$, the raw quadratic is exactly in the Gaussian tangent class, giving $R^2 = 1$ for ordinary Gaussian scores as predicted by Theorem 7. After rank shaping the target is no longer quadratic in x: top-25% shaping drops the ordinary Gaussian explanation to about 0.156 (Table S2). Rank-basis augmentations restore near-invariance. By Theorem 3 read in its shaped form, a Gaussian under elite selection on a quadratic landscape converts only about one sixth of the available selection response into parameter motion.

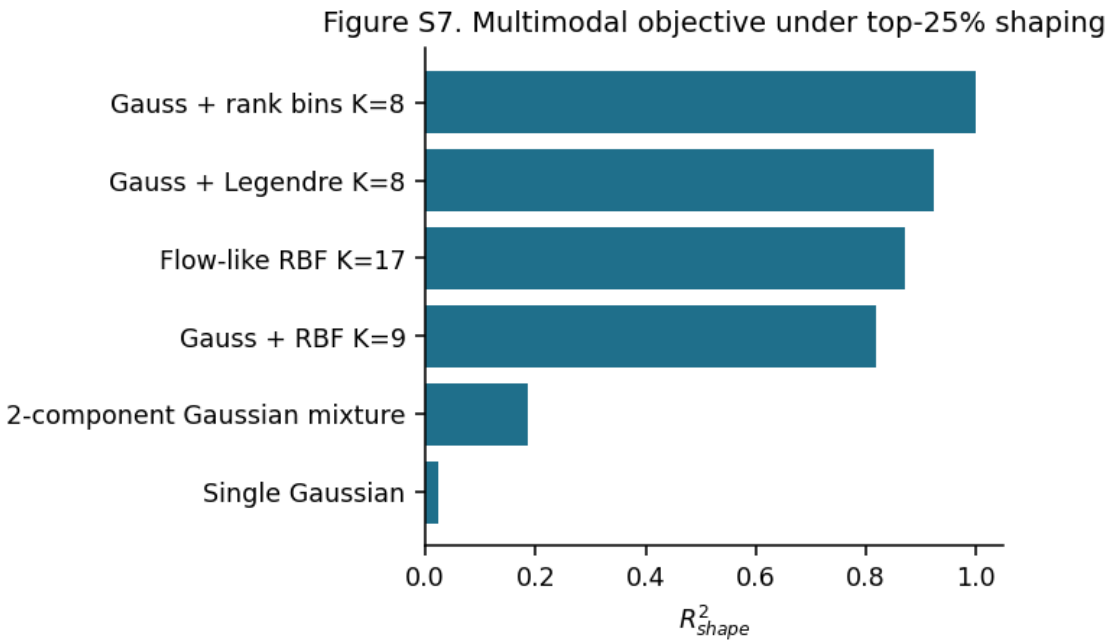


Figure S7. Experiment S4: family and dictionary comparison for a multimodal objective under top-25% shaping.

Table S2. Explained variance R^2 for raw and shaped quadratic fitness.

family	raw $f(x) = -x^2$	linear rank	elite top 25%	smooth top 25%
Gauss + Legendre rank, $K = 4$	1.000	1	0.85	0.979
Gauss + Legendre rank, $K = 8$	1.000	1	0.924	0.999
Gauss + rank bins, $K = 8$	1.000	0.987	1	0.971
Gaussian scores	1.000	0.608	0.156	0.206

Experiment S4: multimodal fitness with elite shaping. The raw fitness has two modes near $x = \pm 2$, evaluated under broad Gaussian or mixture baselines with elite top-25% shaping. The single Gaussian is a poor near-invariant family here ($R^2 = 0.053$), rank bins are nearly exact by construction, and Legendre-rank and RBF dictionaries are also strong, although the RBF dictionary is far more ill-conditioned (Table S3). A two-component mixture improves only modestly over a single Gaussian: it captures basin splitting but not the sharp within-basin elite threshold, which is what Section 4.3 predicts from the form of the mixture scores. This configuration is also the setting of Corollary 3.2: the single Gaussian is close to a symmetric stalling state, where the fitness-score covariance nearly vanishes while substantial shaped variance remains unexpressed.

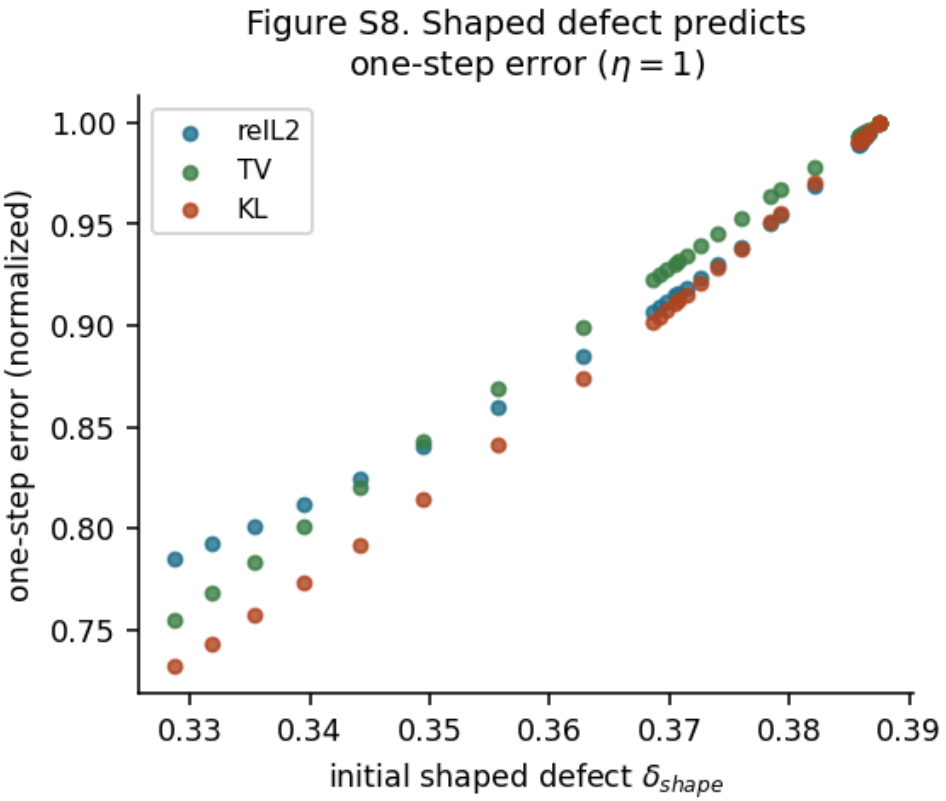


Figure S8. Experiment S5: the shaped defect predicts one-step tracking error across step sizes and error metrics.

Table S3. Family/dictionary comparison for multimodal elite shaping.

family	R^2_{shape}	δ_{shape}	directions	condition no.
Gauss + rank bins $K = 8$	1	0	9	139.175
Gauss + Legendre $K = 8$	0.924	0.12	10	93.098
Flow-like RBF $K = 17$	0.871	0.156	19	5277.68
Gauss + RBF $K = 9$	0.849	0.168	11	585.373
2-component Gaussian mixture	0.165	0.396	5	3.942
Single Gaussian	0.053	0.421	2	1.895

Experiment S5: does the shaped defect predict tracking error? A one-step true shaped update $p^{new}_{true}(x) \propto p(x)e^{\eta w(x)}$ was compared with its projected update $p^{new}_{proj}(x) \propto p(x)e^{\eta \Pi w(x)}$, and errors were measured by relative L^2, total variation, and KL divergence. The correlations between the initial shaped defect and the one-step tracking error are above 0.96 for all step sizes and metrics (Table S4).

This result should be read with its limitation stated. At small η the one-step error is first order in the residual, so a high correlation is close to a consequence of the geometry rather than an independent finding; the informative content is that the correlation survives at $\eta = 1$, where the update is not small. The corresponding test at multiple steps, where accumulation could break the relationship, is Experiment S6.

Table S4. Correlations between initial shaped defect and one-step tracking error.

η	metric	Pearson	Spearman
0.25	relative L^2	1	0.999
0.25	TV	0.987	0.991
0.25	KL	0.966	0.998
0.5	relative L^2	0.999	0.997
0.5	TV	0.988	0.992
0.5	KL	0.964	0.996
1	relative L^2	0.997	0.991
1	TV	0.99	0.994
1	KL	0.96	0.994

S3. Multi-step tracking and calibration of the trajectory estimate

On the bounded interval $[-6, 6]$, the initial density was $\mathcal{N}(0, 1)$ and

$$f_\epsilon(x) = 0.30x - 0.20x^2 + \epsilon \sin(1.5x), \qquad \epsilon \in \{0, 0.05, 0.10, 0.20, 0.40, 0.60\}.$$

The true trajectory was evaluated from the exact tilt of Proposition 5 while the Gaussian path followed the Fisher-projected moment equations. Figure S9 shows that the accumulated L^1 error increases systematically with the nonquadratic perturbation. At $t = 1$ and $\epsilon = 0.4$, the instantaneous defect is 0.135, the L^1 error is 0.117, and the unamplified residual integral is 0.171. The rigorous envelope from Equation (49) is 2.21×10^4 for the same case.

The envelope is valid and extremely conservative, by roughly five orders of magnitude. Equation (49) is therefore a worst-case guarantee rather than a numerical forecast. The unamplified accumulated defect $\int_0^t \delta(\theta_s)\, ds$ lies on the scale of the observed error in this test, but the full-precision ratios vary systematically with perturbation amplitude. It should be interpreted as a diagnostic and problem-specific scale indicator, not as a universally calibrated error estimate.

Table S5. Multi-step defect, observed error, and theoretical envelope at $t = 1$.

ϵ	$\delta(\theta_1)$	L^1 error	$\int_0^1 \delta(s)\, ds$	Eq. (43) envelope
0.00	2.21e-11	3e-08	1.98e-09	0.000492
0.05	0.019	0.0159	0.0221	1.43e+03
0.10	0.0376	0.0315	0.0441	3.14e+03
0.20	0.0729	0.0618	0.0874	7.57e+03
0.40	0.135	0.117	0.171	2.21e+04
0.60	0.184	0.164	0.248	4.85e+04

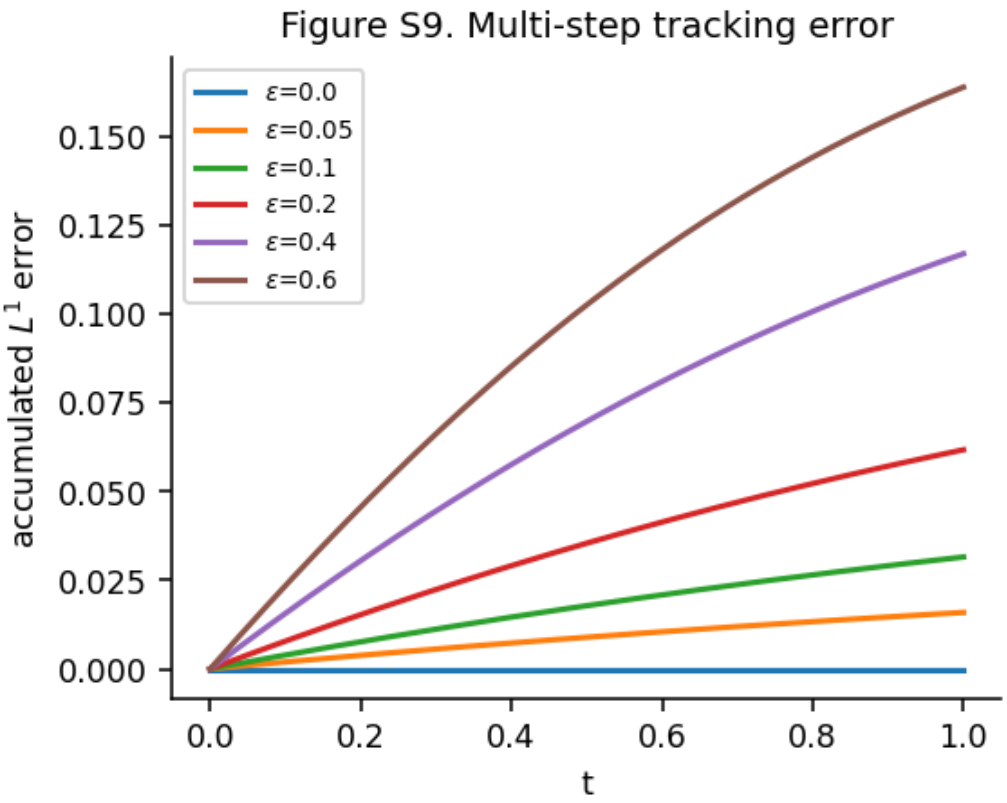


Figure S9. Multi-step tracking error of the Fisher-projected Gaussian flow. The true density is the exact exponential tilt and the surrogate follows the Gaussian Fisher projection. The quadratic case remains at numerical zero; the error increases smoothly with the amplitude of the nonquadratic perturbation.

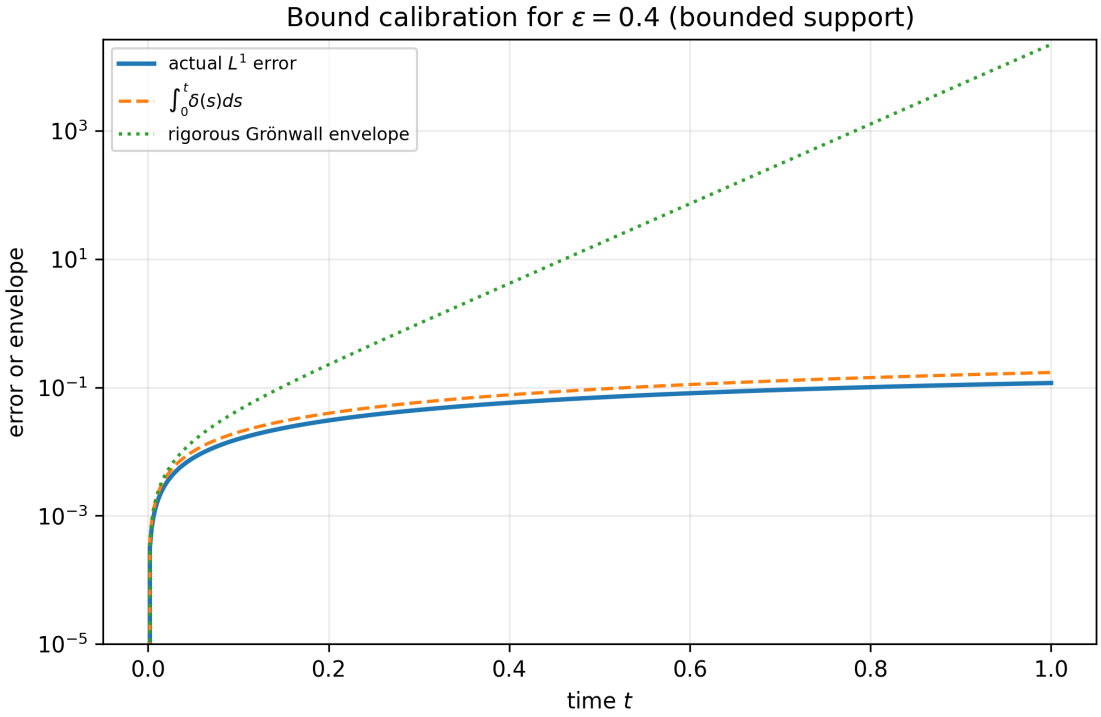


Figure S10. Calibration of the trajectory estimate for $\epsilon = 0.4$. The accumulated defect closely follows the scale of the actual L^1 error, whereas the rigorous Grönwall envelope based on one worst-case Lipschitz constant is many orders of magnitude larger. The theorem is a validity guarantee, not a sharp estimator.

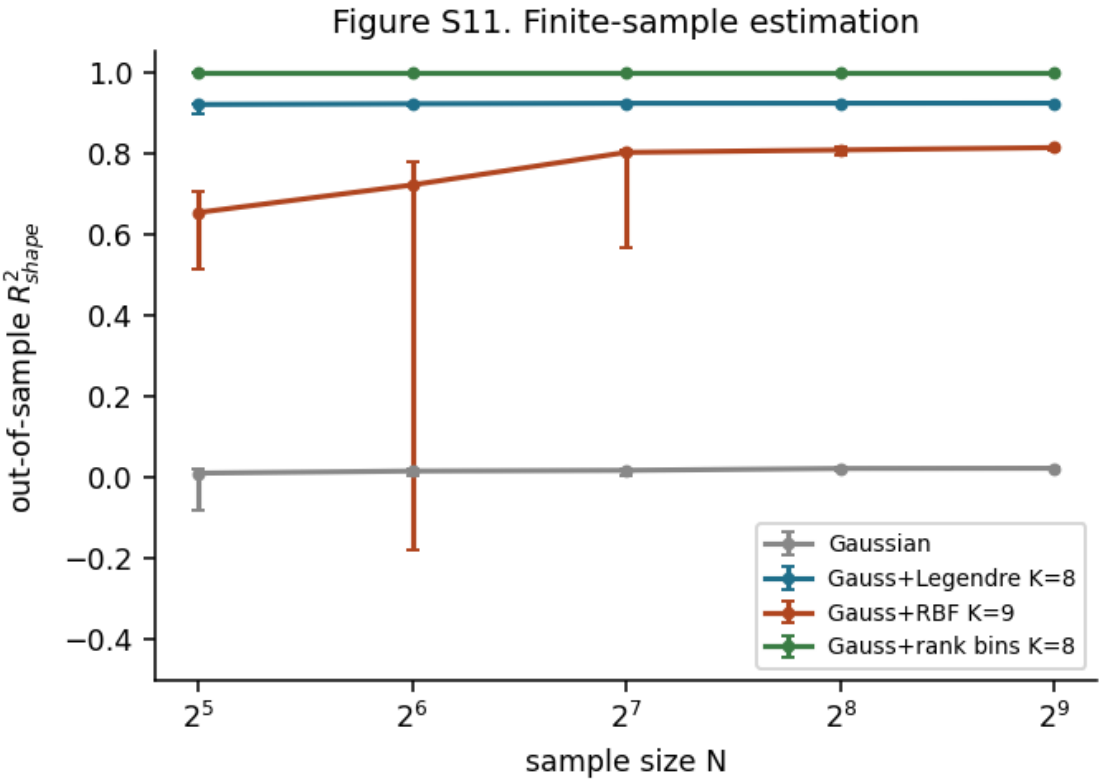


Figure S11. Finite-sample estimation of shaped projected directions. Points show the median out-of-sample explained shaped variance using the best ridge level at each sample size; error bars show the 10th–90th percentile range over 150 samples. Rank bins are exactly representative but algebraically singular; the flexible RBF dictionary is much more sample-demanding than the Legendre-rank dictionary.

S4. Finite-sample estimation, singularity, and regularization

The preceding shaped experiments used quadrature and therefore isolated representational error. Evolution strategies estimate the Fisher matrix and the fitness-score covariance from finite samples. To separate these effects, the multimodal elite-shaping example was evaluated under a broad $\mathcal{N}\left(0, 2.5^2\right)$ baseline. Population projection values were computed by dense quadrature, after which sample estimates were formed with $N = 32, 64, 128, 256, 512$ and ridge levels $\lambda \in \{0, 10^{-6}, 10^{-4}, 10^{-2}\}$. Each setting used 150 independent samples, evaluated out of sample under the population distribution.

The Gaussian-plus-Legendre dictionary has population $R^2_{shape} = 0.924$ and already reaches a median 0.865 at $N = 32$ and 0.920 at $N = 128$. The nine-RBF dictionary has a lower population value, 0.727, and is substantially harder to estimate: its median is 0.376 at $N = 32$ and 0.640 at $N = 128$, with large lower-tail failures unless regularized. Rank bins represent the elite indicator exactly, but their indicators sum to one and the centered Gram matrix is singular, with condition numbers near 10^{19}; by Proposition 2(iii) the projected direction is identifiable through the pseudoinverse even though the coefficient vector is not, and the reported $R^2 = 1.000$ should be read as a statement about the represented direction, not about a stably estimable parameterization. A low population defect is necessary for representational accuracy but not sufficient for a stable sampled algorithm.

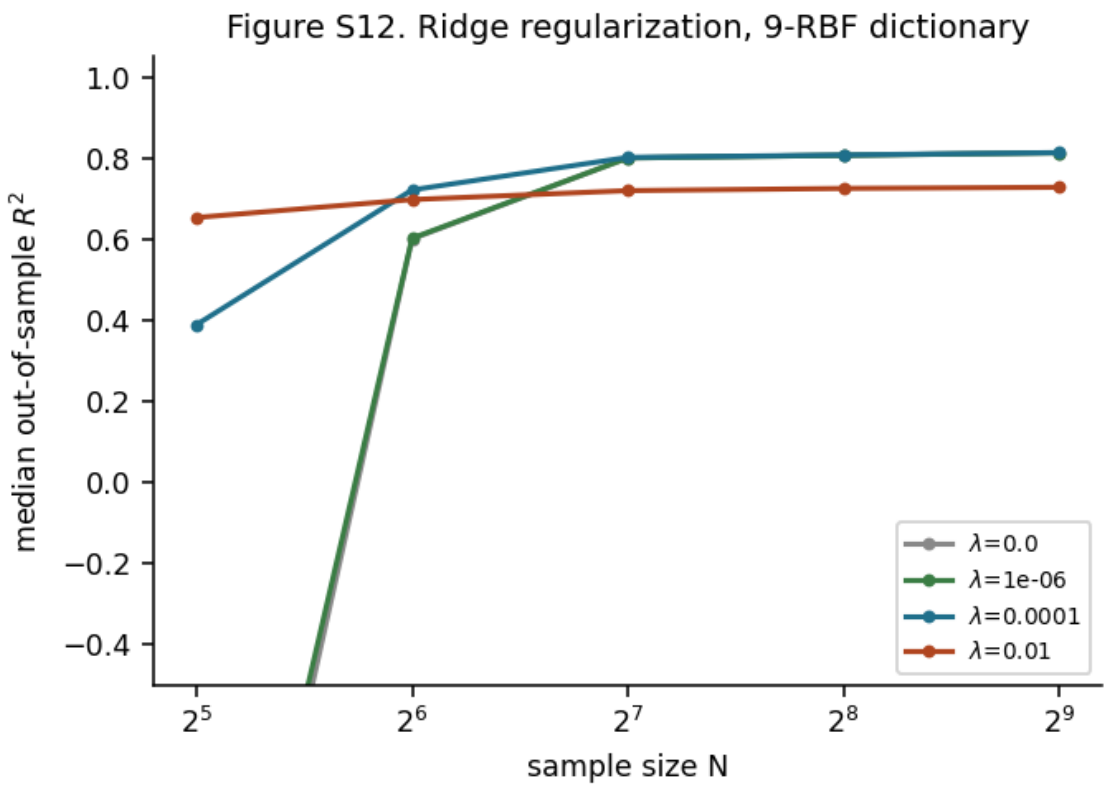


Figure S12. Effect of ridge regularization on the nine-RBF dictionary. Small ridge values materially reduce low-sample instability, whereas excessive regularization introduces bias. The appropriate regularization depends on sample size and dictionary conditioning.

Table S6. Best median finite-sample performance over the tested ridge values.

dictionary	N	population R^2	best λ	median test R^2	10–90% range	median cond.
Gaussian	32	0.033	0.01	0.013	[-0.123, 0.029]	8.91
Gaussian	64	0.033	0.0001	0.024	[-0.017, 0.032]	9.33
Gaussian	128	0.033	0.0001	0.029	[0.009, 0.033]	9.01
Gaussian	256	0.033	1e-06	0.031	[0.022, 0.033]	8.76
Gaussian	512	0.033	0	0.032	[0.029, 0.033]	8.77
Gaussian + Legendre rank, $K = 8$	32	0.924	0.0001	0.865	[0.559, 0.912]	7.6e+03
Gaussian + Legendre rank, $K = 8$	64	0.924	1e-06	0.909	[0.879, 0.919]	2.62e+03
Gaussian + Legendre rank, $K = 8$	128	0.924	0	0.920	[0.910, 0.923]	1.71e+03
Gaussian + Legendre rank, $K = 8$	256	0.924	0.0001	0.922	[0.920, 0.924]	1.49e+03
Gaussian + Legendre rank, $K = 8$	512	0.924	0.0001	0.923	[0.922, 0.924]	1.42e+03
Gaussian + RBF, $K = 9$	32	0.727	0.0001	0.376	[-2.067, 0.520]	9.88e+04
Gaussian + RBF, $K = 9$	64	0.727	0.0001	0.518	[0.367, 0.564]	9.79e+04

dictionary	N	population R^2	best λ	median test R^2	10–90% range	median cond.
Gaussian + RBF, $K = 9$	128	0.727	1e-06	0.640	[0.142, 0.665]	8.75e+06
Gaussian + RBF, $K = 9$	256	0.727	0	0.702	[0.473, 0.719]	5.22e+07
Gaussian + RBF, $K = 9$	512	0.727	0	0.719	[0.700, 0.723]	4.19e+07
Gaussian + rank bins, $K = 8$	32	1.000	0	1.000	[1.000, 1.000]	7.71e+18
Gaussian + rank bins, $K = 8$	64	1.000	0	1.000	[1.000, 1.000]	8.57e+18
Gaussian + rank bins, $K = 8$	128	1.000	0	1.000	[1.000, 1.000]	8.31e+18
Gaussian + rank bins, $K = 8$	256	1.000	0	1.000	[1.000, 1.000]	7.25e+18
Gaussian + rank bins, $K = 8$	512	1.000	0	1.000	[1.000, 1.000]	6.88e+18

S5. A two-dimensional nonquadratic landscape

To test whether the defect-error relationship is peculiar to one dimension, the full bivariate Gaussian family was initialized at $\mathcal{N}(0, I)$ and evolved on $[-5, 5]^2$ under

$$f_\epsilon(x_1, x_2) = 0.35x_1 - 0.18x_1^2 - 0.12x_2^2 + 0.08x_1x_2 + \epsilon \sin(1.3x_1) \cos(1.1x_2).$$

At $\epsilon = 0$ the fitness is affine-quadratic and the defect is numerical zero. As ϵ rises from 0.1 to 0.6, both the defect and the accumulated error at $t = 0.8$ increase monotonically. At $\epsilon = 0.6$, the defect is 0.205, the L^1 error is 0.133, and the KL divergence is 0.0145.

Table S7. Two-dimensional Gaussian projection at $t = 0.8$.

ϵ	defect	L^1 error	KL	projected mean
0.0	2.46e-07	0.000244	5.78e-08	(0.218, 0.012)
0.1	0.0358	0.0228	0.000423	(0.240, 0.012)
0.2	0.071	0.0455	0.00168	(0.261, 0.012)
0.4	0.14	0.0901	0.00658	(0.303, 0.013)
0.6	0.205	0.133	0.0145	(0.344, 0.014)

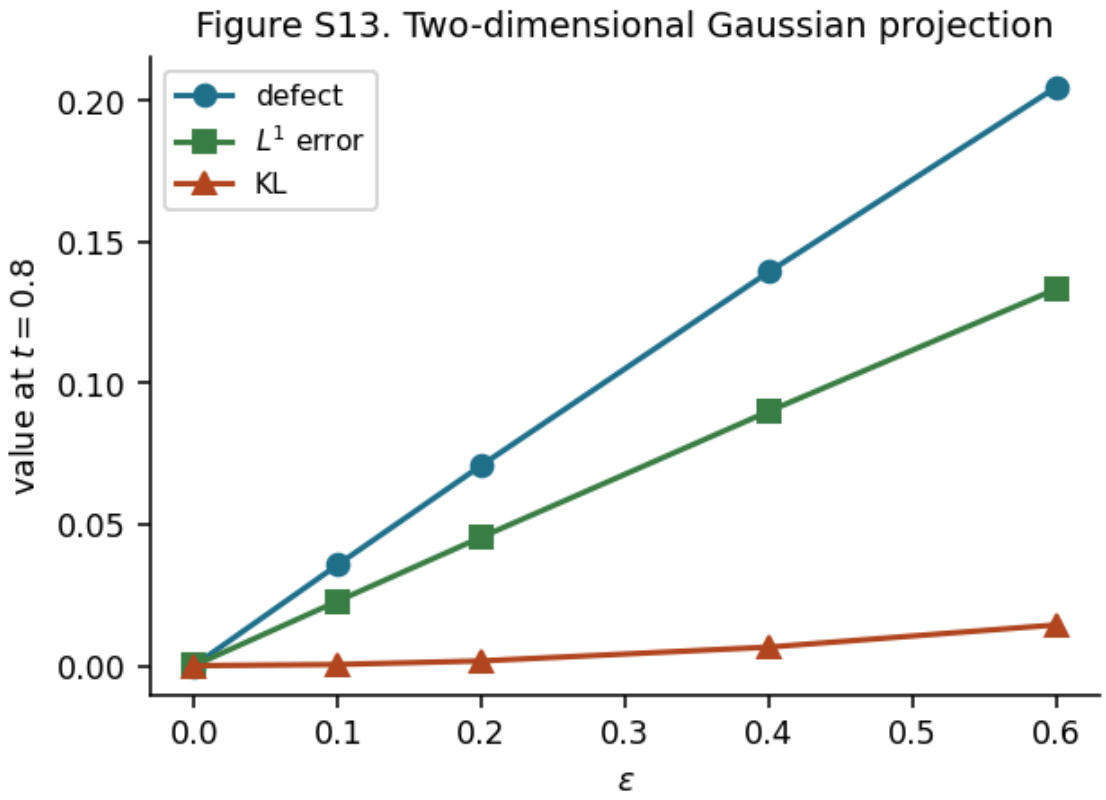


Figure S13. Two-dimensional Gaussian projection. The affine-quadratic case has numerical-zero defect and error. Adding a nonquadratic, nonseparable perturbation increases the invariance defect, L^1 tracking error, and KL divergence together.

S6. Direct tests and gap repairs

Experiment S9: the Fisher decomposition, verified dynamically. Richardson-extrapolated finite differences agreed with the full replicator fitness-increase rate to 1.09×10^{-11} and with the projected rate to 1.40×10^{-8}. The initial rate gap agreed with the squared defect to 2.61×10^{-15}. Theorem 3 is therefore confirmed not only as an algebraic identity but as a statement about two numerically integrated flows.

Experiment S10: exact symmetric stalling and symmetry breaking. For $f(x) = -\left(x^2 - 4\right)^2$, a centered Gaussian is stationary at variance $4/3$, as predicted by Corollary 3.3 with $a = 2$. Population quadrature gave $\| c \| = 1.31 \times 10^{-12}$ and defect $8.709 = \sqrt{8a^8/27}$, with $R^2 = 0$ to numerical precision: the entire fitness variance is unexpressed. The projected displacement over five time units was 1.20×10^{-7}, whereas the exact tilted density moved away from Gaussianity and developed the double-well structure.

The nonlinear analysis determines the behavior near the stationary point. The variance direction is attracting, but the local center-manifold equation is $\dot{\mu} = \left(8a^2/3\right)\mu^3 + O\left(\mu^5\right)$, so the symmetric point repels asymmetric mean perturbations. Exact symmetry is a genuine stationary state, while nearby trajectories may remain close for a time of order μ_0^{-2} before choosing one of the two modes. The finite-time map in Figure S15 shows the two symmetry-broken outcomes and a narrow long-lived near-stalling region around $\mu = 0$, it is not an attracting basin.

Experiment S11: convergence orders. Across perturbation amplitudes 0.1, 0.2, 0.4, and 0.6, fitted moment-error slopes were 1.9822, 1.9822, 1.9822, and 1.9821, while density-error slopes were 0.9959, 0.9958, 0.9958, and 0.9956. These values are consistent with the orders in Corollary 4.1: quadratic moment error and linear density error. At fixed small time, density error was nearly linear in the initial defect in this family, but the theoretical leading coefficient is $\mathbb{E}\,|r_0|$, with δ providing an upper bound rather than

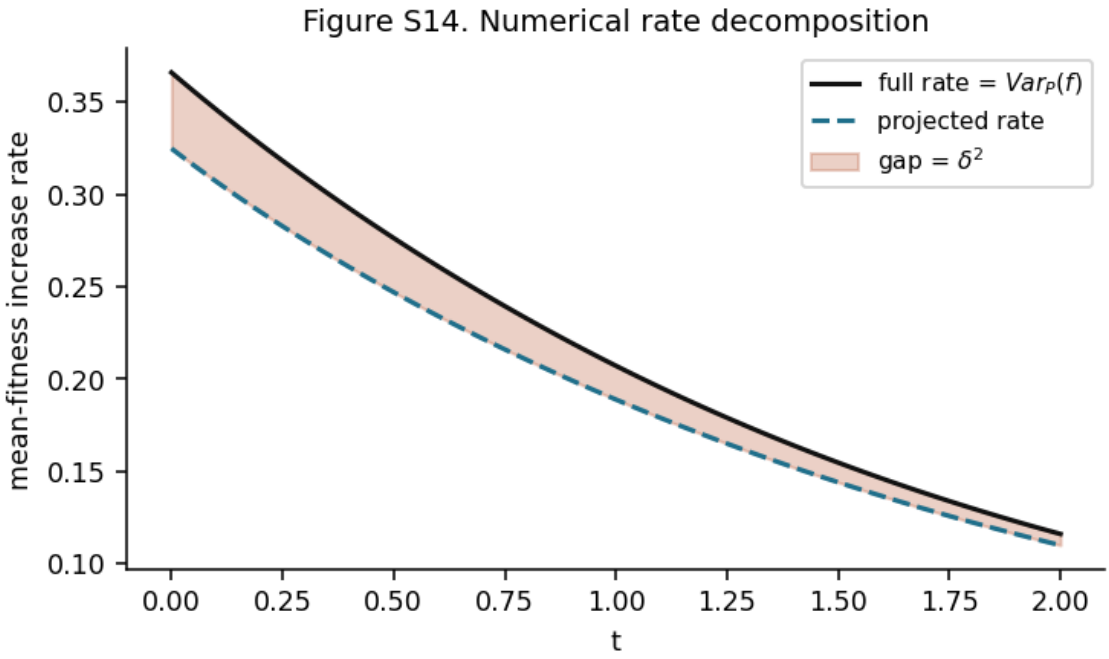


Figure S14. Numerical rate decomposition. The full rate equals fitness variance; the projected rate is the explained component; the shaded gap is the squared invariance defect.

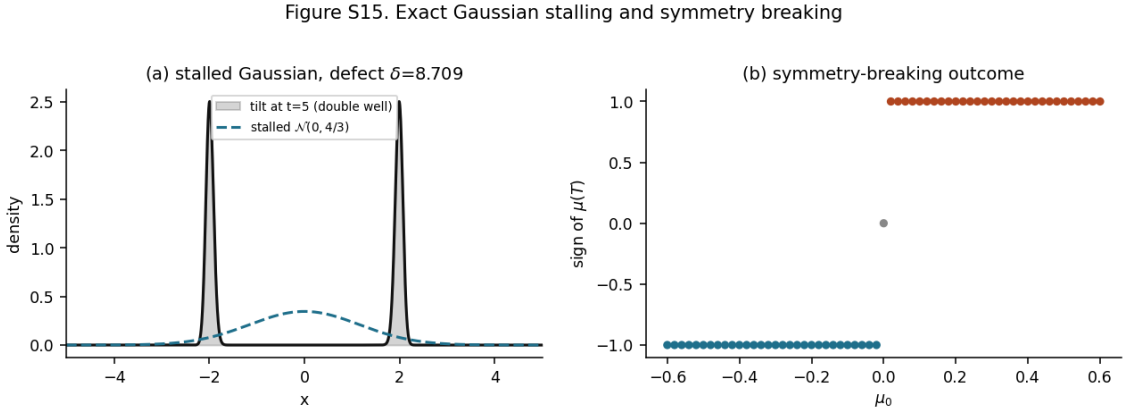


Figure S15. Exact Gaussian stalling and symmetry breaking. The marked symmetric state has zero projected velocity but positive defect, while the full replicator trajectory changes density shape. The variance direction is attracting. The right panel records the sign of the projected mean after a finite integration horizon: blue and red are the two symmetry-broken outcomes, and the narrow gray region contains long-lived near-stalling transients. The mean direction is neutral at linear order and repelling at cubic order; the gray region is not an attracting basin.

Figure S16. Convergence-order test

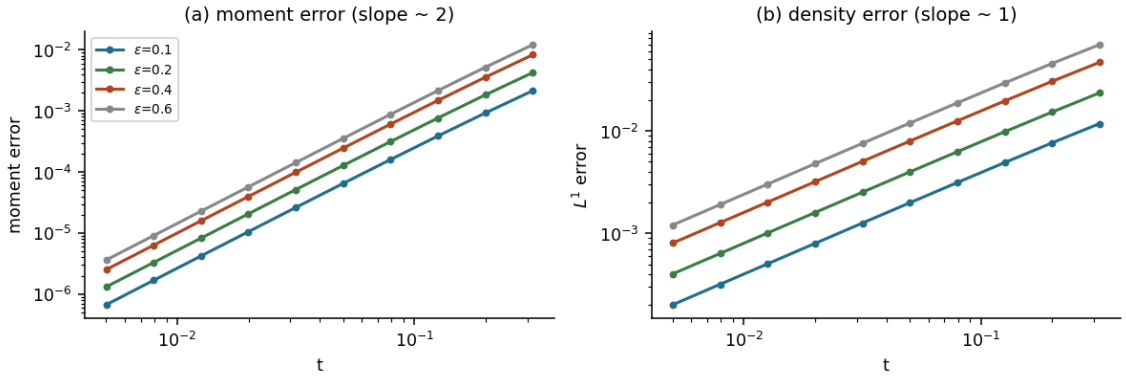


Figure S16. Convergence-order test. Moment errors have slope near two and density errors slope near one on log-log axes.

Figure S17. Gaussian and Student-t explained fractions

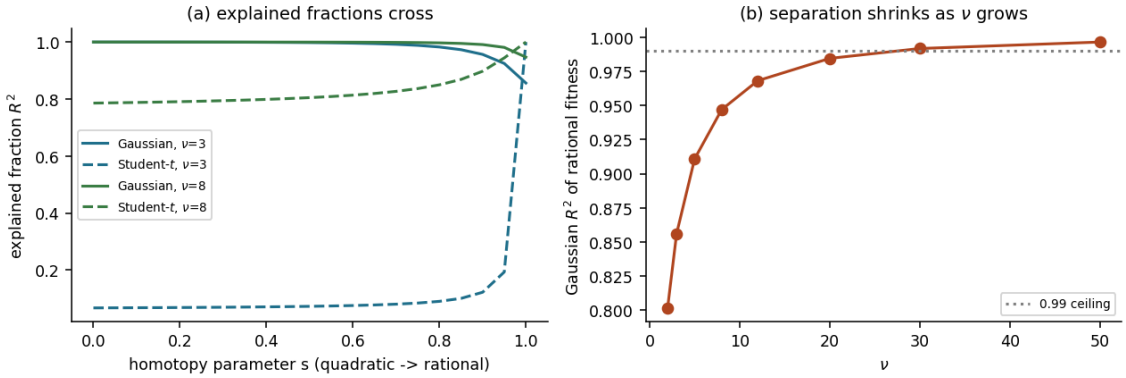


Figure S17. Gaussian and Student-t explained fractions along the quadratic-to-rational homotopy. Exact conjugacy occurs at opposite ends and the separation shrinks as ν grows.

a generic equality.

Experiments S12 to S14: invariance and closed-form identities. The represented tangent direction changed by at most 3.62×10^{-13} under an invertible coordinate transformation, confirming Proposition 2(i). Under five strictly increasing transformations of the objective every shaped quantity was identical at machine precision, while the raw defect varied by a factor of 92.3; this is Proposition 8, and the contrast is the reason R^2_{shape} is the comparable quantity across problems. The Gaussian positive-quadratic blow-up times were 5.61798 and 2.5 as predicted by Equation (50), and a Student-t base under negative quadratic fitness remained normalizable, showing that the finite-time analysis of Section 3.11 is not specific to the Gaussian.

Experiment S15: Student-t conjugacy, with a threshold failure. The rational conjugate fitness of Section 4.3 gave $R^2 = 1$ for every tested degree of freedom, and all Gaussian to Student-t homotopy curves crossed. A uniform crossed-case threshold nevertheless fails: at $\nu = 30$ the Gaussian explanation of the Student-t conjugate fitness was $R^2 = 0.99312$, above the 0.99 ceiling that was used for the comparison. This is the expected Gaussian limit rather than a defect in the theory, but it falsifies the blanket threshold as written. The correct statement is that crossed cases are non-exact at finite ν and converge toward Gaussian compatibility as ν grows.

Experiment S16: frequency-dependent fitness. A pairwise kernel polynomial of degree at most two in the focal trait had numerical-zero defect for arbitrary dependence on the interacting trait, confirming

Corollary 6.1. A Gaussian smoothing kernel gave a positive but modest defect for the tested bandwidth. The selection component of the Price decomposition closed numerically with the squared defect, as Theorem 3 requires. The total mean-fitness-rate difference additionally contains the functional-derivative term $\mathbb{E}\left[D_P f\left(R - R_{proj}\right)\right]$ from Section 3.5; it is not generally equal to δ^2. This experiment closes the structural gap that the other main tests use distribution-independent fitness while preserving that qualification.

Experiment S17: dictionary cost in dimension, and one inconclusive design. The dimensional study considered two designs. The second succeeded and answers a question Section 7.1 previously left as speculation; the first did not, and is not used to support a scaling claim.

Dictionary cost (E18b). On the multimodal elite-shaping problem lifted to d dimensions, three dictionary constructions were grown until the shaped explained variance reached 0.9: rank features of a single projected coordinate (structured), per-axis rank features (separable), and tiled radial basis functions.

Table S8. Directions required to reach $R^2_{shape} \geq 0.9$. Entries marked "not reached" were censored at the search cap, with terminal R^2 falling to 0.49 (separable) and 0.68 (tiled RBF) at $d = 10$.

d	structured rank	separable rank	tiled RBF
1	**10** (cond 38)	18 (cond 1.5×10^4)	not reached at 258
2	**13** (cond 100)	not reached at 261	not reached at 261
3	**17** (cond 130)	not reached at 265	not reached at 265
5	**28** (cond 211)	not reached at 276	not reached at 276
10	**73** (cond 399)	not reached at 321	not reached at 321

The structured construction reaches the target at every dimension, and the decomposition of its cost is the result worth stating. Subtracting the Gaussian base, which has $d(d+3)/2 = 2, 5, 9, 20, 65$ directions at these dimensions, leaves exactly **8 rank features at every dimension**, with terminal R^2 constant at 0.9235 to four decimals. All growth in the total count is the Gaussian base, which any full-covariance method must carry anyway.

This has a simple explanation and a useful consequence. The rank of a point is a scalar, so a dictionary built on the rank represents the same function of the quantile whatever the ambient dimension; the augmentation is dimension-free by construction. Maintaining a given defect may require rapidly growing feature counts for tiled and separable dictionaries, but not for structured rank features. Conditioning tells the same story: the structured Gram matrix stays between 38 and 399 while the tiled dictionary reaches 10^{17}, which is numerically singular.

Experiment S18: the one-sided Lipschitz proposal. Smooth normalized perturbations at three magnitudes gave a maximum sampled logarithmic norm of 0.314 with no negative local values, the minimum being 0.0108. The refined final-time envelope was 5.13 orders of magnitude smaller than the original $3M + 2K$ envelope, an improvement exceeding one order of magnitude. The improvement therefore comes entirely from a smaller positive constant: the contraction that would have made the envelope nongrowing was not observed on this region, and the estimate remains local and numerical rather than a global proof.

Experiment S19: the family floor against dynamic drift. At each time the m-projection was obtained by matching the true tilted mean and variance. The L^1 distance between the projected path and this best static Gaussian had small-time slope 1.994, confirming the $O\left(t^2\right)$ agreement predicted by Corollary 4.1 and separating the irreducible Gaussian approximation error from the drift accumulated by the projected path.

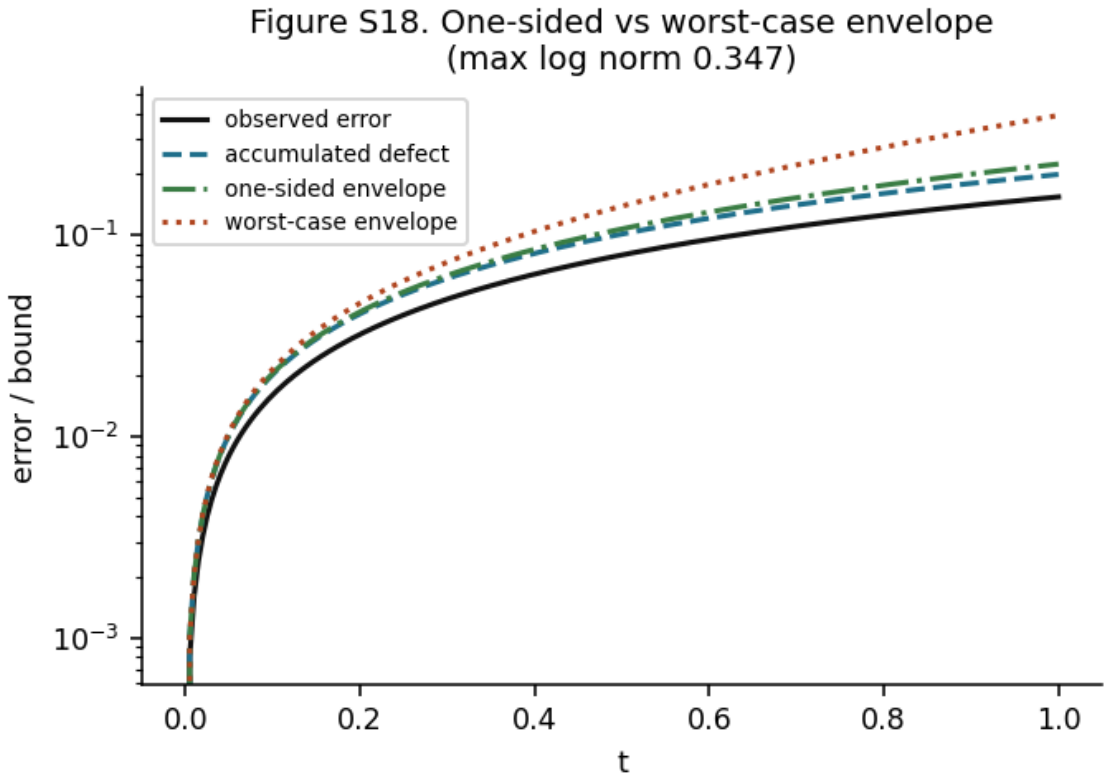


Figure S18. Observed tracking error, accumulated defect, sampled one-sided envelope, and original worst-case envelope.

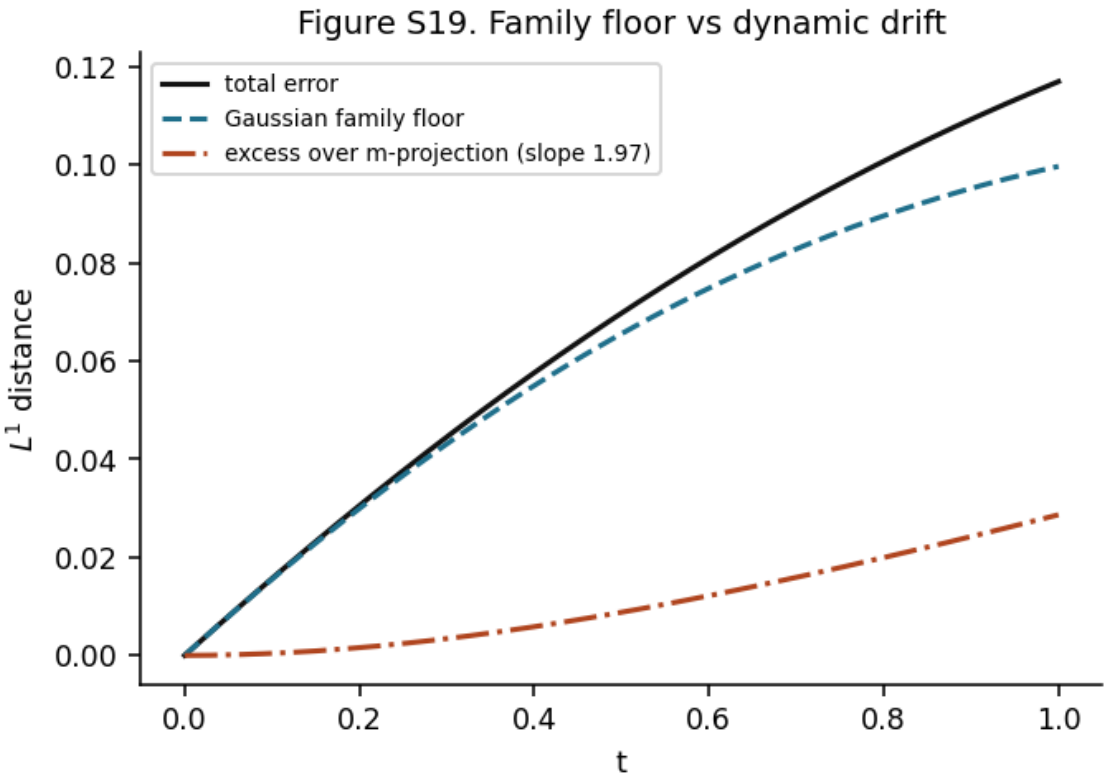


Figure S19. Tracking-error decomposition into the Gaussian family floor and the projected path's excess over the moment-matched m-projection.

S7. Interpretation

- The raw-fitness theory behaves as predicted: affine-quadratic fitness lies in the Gaussian score span, gives numerical-zero defect, and produces coincident full and projected trajectories, and this persists in two dimensions. Nonquadratic components create a nonzero residual and accumulated non-Gaussian error.

- The two local orders in Corollary 4.1 are visible. Sufficient-statistic means share the exact initial derivative and separate at order t^2; density trajectories separate at order t when the residual is nonzero. The initial L^1 coefficient is $\mathbb{E}\,|r_0| \le \delta$, not generically the defect itself.

- Quantile shaping changes the exact-invariance class. Even when the raw fitness is linear or quadratic, a nonlinear rank transform is generally not linear or quadratic in the trait, so the ordinary Gaussian score span is too small for hard elite selection. Rank and local dictionaries attack the correct target, the centered rank signal, and can reduce the population defect substantially.

- The defect has two empirical roles. It is exactly the omitted instantaneous relative velocity, so it orders local tracking error and indicates where error is being generated. Its time integral is on the scale of observed error in the S3 family, but Experiment S19 shows that the calibration is not universal: it depends on horizon and metric, and curvature appears early in KL. The rigorous Gronwall envelope is looser still. The defect is therefore a reliable representational diagnostic, not a universal numerical error forecast.

Exact symmetric stalling is real but nongeneric. Experiment S10 has zero projected velocity and positive defect on the symmetry subspace, while the center-manifold calculation shows cubic escape under asymmetric perturbations. In practice this can appear as a long plateau rather than a robust attractor.

- Representation and estimation must be separated. A dictionary can have a very small population defect and be unusable at realistic sample sizes if its Gram matrix is ill-conditioned. Rank-bin directions can be exactly representative but non-identifiable as coefficients; RBF dictionaries can be expressive but sample-hungry. Pseudoinverses, ridge regularization, basis reduction, and trust regions are part of the practical geometry.

S8. Numerical methods and reproducibility

All density calculations in Supplementary Sections S1-S5 used normalized quadrature. One-dimensional raw-fitness experiments used grids fine enough that the quadratic invariance defect was at the floating-point floor. The exact fixed-fitness replicator trajectory was evaluated from Proposition 5. Fisher-projected Gaussian paths used the moment equations in Eq. (S1). Ordinary-L^2 density-velocity projections, defined in Appendix C of the main manuscript, were solved by least squares on density derivatives.

Quantiles were computed as weighted empirical cumulative probabilities of the raw fitness values. Elite top-25% shaping used a centered indicator, and smooth elite shaping used the logistic rule stated in Section 5.1. Score dictionaries were centered under the current density before forming the Gram matrix. Moore–Penrose pseudoinverses were computed by singular-value decomposition.

Experiment S6 used 2,401 points on $[-6, 6]$, time step 0.002, and final time 1. The theoretical envelope used the shift-invariant bounded-support constant $L = 3M$, with M equal to one half of the fitness oscillation. Experiment S7 used 4,001 quadrature points, 150 independent samples per setting, the five sample sizes and four ridge levels reported in the text, and out-of-sample evaluation under the population density. Ridge terms were scaled by the mean diagonal of the sample Gram matrix.

Experiments S8-S19 used Python 3.9.6 with NumPy 2.0.2 and SciPy 1.13.1. One-dimensional population calculations used 24,001 points on the stated intervals, except Student-t calculations, which used 60,001 points to control tail truncation. Projected Gaussian paths used DOP853 with the reported relative tolerances, and Experiment S11 limited the internal step so that integrator error remained below the measured convergence orders. Discontinuous ranks used stable mergesort ordering.