Objective Improvement in Information-Geometric Optimization

Youhei Akimoto
Project TAO – INRIA Saclay
LRI, Bât. 490, Univ. Paris-Sud
91405 Orsay, France
Youhei.Akimoto@lri.fr

Yann Ollivier
CNRS & Univ. Paris-Sud
LRI, Bât. 490
91405 Orsay, France
yann.ollivier@lri.fr

ABSTRACT

Information-Geometric Optimization (IGO) is a unified framework of stochastic algorithms for optimization problems. Given a family of probability distributions, IGO turns the original optimization problem into a new maximization problem on the parameter space of the probability distributions. IGO updates the parameter of the probability distribution along the natural gradient, taken with respect to the Fisher metric on the parameter manifold, aiming at maximizing an adaptive transform of the objective function. IGO recovers several known algorithms as particular instances: for the family of Bernoulli distributions IGO recovers PBIL, for the family of Gaussian distributions the pure rank-μ CMA-ES update is recovered, and for exponential families in expectation parametrization the cross-entropy/ML method is recovered.

This article provides a theoretical justification for the IGO framework, by proving that any step size not greater than 1 guarantees monotone improvement over the course of optimization, in terms of q-quantile values of the objective function f. The range of admissible step sizes is independent of f and its domain. We extend the result to cover the case of different step sizes for blocks of the parameters in the IGO algorithm. Moreover, we prove that expected fitness improves over time when fitness-proportional selection is applied, in which case the RPP algorithm is recovered.

Categories and Subject Descriptors

G.1.6 [**Mathematics of Computing**]: Numerical Analysis—*Optimization*

General Terms

Theory

Keywords

Information-Geometric Optimization, Natural Gradient, Quantile Improvement, Step Size, Black Box Optimization

FOGA'13, January 16–20, 2013, Adelaide, Australia.

1. INTRODUCTION

Information-Geometric Optimization (IGO) [5] is a unified framework of model based stochastic search algorithms for any optimization problem. As typified by Estimation of Distribution Algorithms (EDA) [15], model based randomized search algorithms build a statistical model P_θ on the search space X to generate search points. The parameters θ of the statistical model are updated over time so that the probability distribution hopefully concentrates around the minimum of the objective function. In most model based algorithms such as EDAs and Ant Colony Optimization (ACO) algorithms [10], parameter calibration is based on the maximum likelihood principle or other intuitive ways. IGO, unlike them, performs a natural gradient ascent of θ in the parameter space Θ, having first adaptively transformed the objective function into a function on Θ. This construction offers maximal robustness guarantees with respect to changes in the representation of the problem (change of parametrization of the search space, of the parameter space, and of the fitness values).

Importantly, the IGO framework recovers several known algorithms [5, Section 4]. When IGO is instantiated using the family of Bernoulli distributions on $\{0,1\}^d$, one obtains the *population based incremental learning* (PBIL) algorithm [6]. When using the family of Gaussian distributions on $\mathbb{R}^d$, IGO instantiates as a variant of *covariance matrix adaptation evolution strategies* (CMA-ES), the so-called pure rank-μ CMA-ES update [11]. Moreover, when using an exponential family with the expectation parameters, the IGO instance is equivalent to the cross-entropy method for optimization [7]. Of course, the IGO framework not only provides information-theoretic derivations for existing algorithms but automatically offers new algorithms for possibly complicated optimization problems. For instance, the IGO update rule for the parameters of restricted Boltzmann machines has been derived [5].

Theoretical justification of the IGO framework, therefore, is important both to provide a theoretical basis for the recovered algorithms and to make the design principle for future algorithms more reliable. Here we focus on providing a measure of "progress" over the course of IGO optimization, in terms of quantile values of the objective function.

Parameter updates by gradient ascent are somewhat justified in general, at least for infinitesimally small steps, because the gradient points to the direction of steepest ascent of a function. However, this argument does not apply to the IGO algorithm: as the objective function is adaptively transformed in a time-dependent way, the function on which the

gradient is computed changes over time, so that its increase does not necessarily mean global improvement. Still, the IGO framework comes with a guarantee that an infinitesimally small IGO step along the natural gradient leads to monotone improvement of a specified quantity, for any objective function f [5, Proposition 5]: a result from [5] is that the *q-quantile value* of the objective function monotonically improves along the natural gradient. This result is limited to the exact IGO flow, i.e., an infinite number of sample points is considered and the step size of the gradient ascent is infinitesimal. Still this ensures that the randomized algorithm with large sample size stays close to the deterministic trajectory with infinite samples with high probability, provided the step size is sufficiently small. Now the question arises whether actual, non-infinitesimal step sizes still ensure monotone q-quantile improvement.

In this article, we prove that *any* step size not greater than 1 guarantees monotone q-quantile value improvement in the IGO algorithm for an exponential family with a finite step size (Theorem 6), thus extending the previous result from infinitesimal steps with continuous time to more realistic algorithmic situations. For instance, this ensures monotone q-quantile improvement in PBIL (using uniform weights, see below), or in the cross-entropy method for exponential families in expectation parameters. Interestingly, our results show that the admissible step sizes in IGO are *independent of the objective function* f, at least for large population sizes (this stems from the many invariance properties built into IGO).

We further extend the result by defining *blockwise* updates in IGO where different blocks of parameters are adjusted one after another with different step sizes. Our motivation is that in practice the pure rank-μ update CMA-ES updates the mean vector and the covariance matrix with different learning rates. We show that the blockwise update rule recovers the pure rank-μ CMA-ES update using different learning rates for the mean vector and the covariance matrix (Proposition 9). We prove that *any* distinct step sizes less than 1 guarantee monotone q-quantile improvement, which justifies the parameter setting used for the CMA-ES in practice (Theorem 10).

Other examples fitting into this framework are the Relative Payoff Procedure (also known as expectation-maximization for reinforcement learning) [9,12], or situations where fitness-proportional selection is applied using exponential families (Theorem 12). The RPP is considered as an alternative to gradient based methods that allows to use relatively large learning rates. As it turns out, the RPP can be described as a natural gradient based algorithm with step size 1, and our result is an extension of the proof of its monotone improvement to generic natural gradient algorithms.

The article is organized as follows. In Section 2, we explain the IGO framework and its implementation in practice. IGO-maximum likelihood (IGO-ML), a variant of IGO as a maximum likelihood, is presented, followed by the relation between the IGO algorithm, IGO-ML and the cross-entropy method for optimization, for exponential families of distributions. In Section 3, we prove monotone q-quantile improvement in IGO-ML. The result is extended by defining blockwise IGO-ML, and q-quantile improvement in blockwise IGO-ML is proved. We also provide a result with finite but large population sizes. Section 4 is devoted to the natural gradient algorithm with fitness-proportional selection scheme, where monotone improvement of expected fitness is proven. A short discussion in Section 5 closes the article.

2. INFORMATION-GEOMETRIC OPTIMIZATION

In this article, we consider an objective function $f : X \to \mathbb{R}$ to be minimized over any search space X. The search space X may be continuous or discrete, finite or infinite.

Let $\{P_\theta\}$ be a family of probability distributions parametrized by $\theta \in \Theta$ and let p_θ be the probability density function induced by P_θ w.r.t. an arbitrary reference measure $\mathrm{d}x$ on X, namely, $P_\theta(\mathrm{d}x) = p_\theta(x)\mathrm{d}x$. Given a family of probability distributions, IGO [5] evolves the probability distribution P_{θ^t} at each time t so that higher probabilities are assigned to better regions. To do so, IGO transforms the objective function $f(x)$ into a new one $W^f_{\theta^t}(x)$, defines a function on Θ to be maximized: $J(\theta \,|\, \theta^t) := \mathbb{E}_{P_\theta}[W^f_{\theta^t}(x)]$, and performs the steepest gradient ascent of $J(\theta \,|\, \theta^t)$ on Θ. Hopefully, after some time the distribution P_{θ^t} concentrates around minima of the objective function.

IGO is designed to exhibit as many invariance properties as possible [5, Section 2]. The first property is invariance under strictly increasing transformations of f. For any strictly increasing g, IGO minimizes $g \circ f$ as easily as f. This property is realized by a quantile based mapping of f to $W^f_{\theta^t}$ at each time. The second property is invariance under a change of coordinates in X, provided that this coordinate change globally preserves the family of probability distributions $\{P_\theta\}$. For example, the IGO algorithm for Gaussian distributions on $\mathbb{R}^d$ is invariant under any affine transformation of the coordinates whereas the IGO algorithm for isotropic Gaussian distribution is only invariant under any translation and rotation. Invariance under X-coordinate transformation is one of the key properties for the success of the CMA-ES. The last property is invariance under reparametrization of θ. At least for infinitesimal steps of the gradient ascent, IGO follows the same trajectory on the parameter space whatever the parametrization for θ is. This property is obtained by considering the intrinsic (Fisher) metric on the parameter space Θ and defining the steepest ascent on Θ w.r.t. this metric, i.e., by using a natural gradient.

The study of the intrinsic metric on the parameter space of the probability distribution, called a *statistical manifold*, is the main topic of *information geometry* [4]. The most widely used divergence between two points on the space of probability distributions is the Kullback–Leibler divergence (KL divergence)

$$D_{\mathrm{KL}}(P_\theta \,\|\, P_{\theta'}) := \int \ln \frac{p_\theta(x)}{p_{\theta'}(x)} \, P_\theta(\mathrm{d}x) \ .$$

The KL divergence is, by definition, independent of the parametrization θ. Let $\theta' = \theta + \delta\theta$. Then, the KL divergence between P_θ and $P_{\theta+\delta\theta}$ expands [13] as

$$D_{\mathrm{KL}}(P_\theta \,\|\, P_{\theta+\delta\theta}) = \frac{1}{2} \delta\theta^{\mathrm{T}} \mathcal{I}_\theta \delta\theta + O(\|\delta\theta\|^3) \ , \qquad (1)$$

where $\|\cdot\|$ is the Euclidean norm and $\mathcal{I}_\theta$ is the Fisher infor-

mation matrix at θ defined as

$$(\mathcal{I}_\theta)_{ij} := \int \frac{\partial \ln p_\theta(x)}{\partial \theta_i} \frac{\partial \ln p_\theta(x)}{\partial \theta_j} P_\theta(\mathrm{d}x) = -\int \frac{\partial^2 \ln p_\theta(x)}{\partial \theta_i \partial \theta_j} P_\theta(\mathrm{d}x) \ .$$

The expansion (1) follows from the well-known fact that the Fisher information matrix is the Hessian of KL divergence. By using the KL divergence, we have the following property of the steepest ascent direction (see [3], Theorem 1, or [5], Proposition 1).

STATEMENT 1. *Let g be a smooth function on the parameter space Θ. Let $\theta \in \Theta$ be a nonsingular point where $\nabla_\theta g(\theta) \neq 0$. Then the steepest ascent direction of g is given by the so-called natural gradient $\tilde{\nabla}_\theta g(\theta) := \mathcal{I}_\theta^{-1} \nabla_\theta g(\theta)$. More precisely,*

$$\frac{\tilde{\nabla}_\theta g(\theta)}{\|\tilde{\nabla}_\theta g(\theta)\|} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \underset{\substack{\delta\theta \text{ such that} \\ D_{\mathrm{KL}}(P_\theta \| P_{\theta+\delta\theta}) \leq \epsilon^2/2}}{\arg\max} g(\theta + \delta\theta) \ .$$

Since KL divergence does not depend on parametrization, the natural gradient is invariant under reparametrization of θ. Hence, the natural gradient step—steepest ascent step w.r.t. the Fisher metric—is invariant at least for an infinitesimal step size [5, Section 2.4].

2.1 Algorithm Description

For completeness, we include here a short description of the IGO algorithm. We refer to [5] for a more complete presentation.

First, IGO transforms the objective function into an adaptive weighted preference by a quantile based approach. This results in a rank based algorithm, invariant under increasing transformations of the objective function. Define the lower and upper P_θ-f-quantiles of $x \in X$ as

$$q_\theta^<(x) := P_\theta[y : f(y) < f(x)]$$
$$q_\theta^\leq(x) := P_\theta[y : f(y) \leq f(x)] \ .$$

The lower quantile value $q_\theta^<(x)$ is the probability of sampling strictly better points than x under the current distribution P_θ, while the upper quantile value $q_\theta^\leq(x)$ is the probability of sampling points better than or equivalent to x. Given a weight function (selection scheme) $w : [0,1] \to \mathbb{R}$ that is non-increasing, the weighted preference $W_\theta^f(x)$ is defined as

$$W_\theta^f(x) := \begin{cases} w(q_\theta^\leq(x)) & \text{if } q_\theta^<(x) = q_\theta^\leq(x), \\ \frac{1}{q_\theta^\leq(x) - q_\theta^<(x)} \int_{q_\theta^<(x)}^{q_\theta^\leq(x)} w(u)\,\mathrm{d}u & \text{otherwise.} \end{cases} \tag{2}$$

This way, the quality of a point is measured by a function of the P_θ-quantile in which it lies. A typical choice of the selection scheme w is $w(u) = \mathbb{1}_{[u \leq q]}/q$, $0 < q < 1$. We call it the *q-truncation* selection scheme. Using q-truncation amounts, in the final IGO algorithm, to giving the same positive weight to a fraction q of the best samples in a population, and weight 0 to the rest, as is often the case in practice.

Next, IGO turns the original objective function f on the search space X into a function $J(\theta \,|\, \theta^t)$ on the statistical manifold Θ by defining

$$J(\theta \,|\, \theta^t) := \mathbb{E}_{P_\theta}[W_{\theta^t}^f(x)] \ . \tag{3}$$

Note that $J(\theta \,|\, \theta^t)$ depends on the current position θ^t. Then, the gradient of $J(\theta \,|\, \theta^t)$ is computed as

$$\begin{aligned} \nabla_\theta J(\theta \,|\, \theta^t) &= \nabla_\theta \, \mathbb{E}_{P_\theta}[W_{\theta^t}^f(x)] \\ &= \nabla_\theta \int W_{\theta^t}^f(x)\, p_\theta(x)\, \mathrm{d}x \\ &= \int W_{\theta^t}^f(x)\, p_\theta(x)\, \nabla_\theta \ln p_\theta(x)\, \mathrm{d}x \\ &= \mathbb{E}_{P_\theta}[W_{\theta^t}^f(x)\, \nabla_\theta \ln p_\theta(x)] \ . \end{aligned} \tag{4}$$

Here we have used the relation $\nabla p_\theta(x) = p_\theta(x) \nabla \ln p_\theta(x)$.

Finally, IGO uses natural gradient ascent on the parameter space. The natural gradient on the statistical manifold $(\Theta, \mathcal{I})$ equipped with the Fisher metric $\mathcal{I}$ is given by the product of the inverse of the Fisher information matrix, $\mathcal{I}_\theta^{-1}$, and the vanilla gradient. That is, the natural gradient of $J(\cdot \,|\, \theta^t)$ at θ is written as $\tilde{\nabla}_\theta J(\theta \,|\, \theta^t) = \mathcal{I}_\theta^{-1} \nabla_\theta J(\theta \,|\, \theta^t)$. According to (4), we can rewrite the natural gradient as

$$\tilde{\nabla}_\theta J(\theta \,|\, \theta^t) = \mathbb{E}_{P_\theta}[W_{\theta^t}^f(x)\, \tilde{\nabla}_\theta \ln p_\theta(x)] \ . \tag{5}$$

Introducing a finite step size δt, IGO finally updates the parameter as follows

$$\theta^{t+\delta t} = \theta^t + \delta t\, \tilde{\nabla}_\theta J(\theta \,|\, \theta^t)\big|_{\theta=\theta^t} \ . \tag{6}$$

2.2 Implementation and Recovering Algorithms

When implementing IGO in practice, it is necessary to estimate the expectation in (5). The approximation is done by the Monte Carlo method using λ samples taken from P_{θ^t}. Let $x_1, \ldots, x_\lambda$ be independent samples from P_{θ^t}.

First, we need to approximate $W_{\theta^t}^f(x_i)$ for each $i = 1, \ldots, \lambda$. Define

$$\mathrm{rk}^<(x_i) := \#\{j, f(x_j) < f(x_i)\}$$
$$\mathrm{rk}^\leq(x_i) := \#\{j, f(x_j) \leq f(x_i)\} \ ,$$

let

$$\bar{w}_i := \int_{(i-1)/\lambda}^{i/\lambda} w(q)\,\mathrm{d}q, \quad \forall i \in [\![1, \lambda]\!],$$

and set

$$\widehat{w}_i = \frac{1}{\mathrm{rk}^\leq(x_i) - \mathrm{rk}^<(x_i)} \sum_{j=\mathrm{rk}^<(x_i)+1}^{\mathrm{rk}^\leq(x_i)} \bar{w}_j \ . \tag{7}$$

Then $\lambda \widehat{w}_i$ is a consistent estimator of $W_{\theta^t}^f(x_i)$, in other words, $\lim_{\lambda \to \infty} \lambda \widehat{w}_i = W_{\theta^t}^f(x_i)$ with probability one. (See the proof of Theorem 4 in [5].) If there are no ties in our sample, i.e. $f(x_i) \neq f(x_j)$ for any $i \neq j$, then $\mathrm{rk}^\leq(x_i) = \mathrm{rk}^<(x_i) + 1$ and (7) simply reads $\widehat{w}_i = \bar{w}_{\mathrm{rk}^\leq(x_i)}$, but (7) is a mathematically neater definition of rank based weights accounting for possible ties. In practice we just design the λ weight values $\bar{w}_1, \ldots, \bar{w}_\lambda$, instead of the selection scheme w.

In the rest of this article, we assume for simplicity that the selection weights $\bar{w}_i$ are non-negative and sum to 1. This is the case, for instance, if the selection scheme w is q-truncation as above.

Next, Monte Carlo sampling is applied to the expectation (5), using $\widehat{w}_i$ and x_i. Replacing the expectation with a

sample average $\frac{1}{\lambda}\sum_{i=1}^{\lambda}$ and $W_{\theta^t}^f(x_i)$ with $\lambda\widehat{w}_i$, we get

$$G^t := \sum_{i=1}^{\lambda} \widehat{w}_i \, \tilde{\nabla}_\theta \ln p_\theta(x_i)|_{\theta=\theta^t} \quad . \tag{8}$$

Again, G^t is a consistent estimator of the IGO step at θ^t, i.e., of $\tilde{\nabla}_\theta J(\theta \,|\, \theta^t)|_{\theta=\theta^t}$. See Theorem 4 in [5].

Now the practical IGO algorithm implementation can be written in the form of a black-box search algorithm as

1. Sample x_i, $i = 1, \ldots, \lambda$, independently from P_{θ^t};
2. Evaluate $f(x_i)$ and compute $\mathrm{rk}^{\leq}(x_i)$ and $\mathrm{rk}^{<}(x_i)$;
3. Evaluate $G^t = \sum_{i=1}^{\lambda} \widehat{w}_i \, \tilde{\nabla}_\theta \ln p_\theta(x_i)|_{\theta=\theta^t}$;
4. Update the parameter: $\theta^{t+\delta t} = \theta^t + \delta t \cdot G^t$.

Finally, to obtain an explicit form of the parameter update equation, we need to know the explicit form of the natural gradient of the log-likelihood, which depends on a family of probability distributions and its parametrization. Explicit forms of $\tilde{\nabla}_\theta \ln p_\theta(x)$ are known for some specific families of probability distributions with specific parametrizations, and the above algorithm sometimes coincides with several known algorithms.

Example 1. The family of Bernoulli distributions on $X = \{0,1\}^d$ is defined as $P_\theta(x) = \prod_{j=1}^d \theta_j^{x_j}(1-\theta_j)^{1-x_j}$. The natural gradient of the log-likelihood is readily computed as $\tilde{\nabla}_\theta \ln p_\theta(x) = x - \theta$ (Section 4.1 in [5]). The natural gradient update reads

$$\theta^{t+\delta t} = \theta^t + \delta t \sum_{i=1}^{\lambda} \widehat{w}_i \left(x_i - \theta^t\right) \quad .$$

This is equivalent to so-called PBIL (population based incremental learning, [6]).See Section 4.1 in [5] for details.

Example 2. The probability density function of a multivariate Gaussian distribution on $X = \mathbb{R}^d$ with mean vector m and covariance matrix C, is defined as

$$p_\theta(x) = \left(\det(2\pi C)\right)^{-1/2} \exp\left(-(x-m)^{\mathrm{T}} C^{-1} (x-m)/2\right) \quad .$$

When $\theta = (m, C)$, the explicit form of $\tilde{\nabla} \ln p_\theta(x)$ is known to be $\tilde{\nabla} \ln p_\theta(x) = \begin{bmatrix} x-m \\ (x-m)(x-m)^{\mathrm{T}} - C \end{bmatrix}$ (see [2]). Then, the natural gradient update reads

$$\theta^{t+\delta t} = \theta^t + \delta t \sum_{i=1}^{\lambda} \widehat{w}_i \begin{bmatrix} x - m^t \\ (x-m^t)(x-m^t)^{\mathrm{T}} - C^t \end{bmatrix} \quad .$$

This is equivalent to the pure rank-μ CMA-ES update [11]

$$m^{t+1} = m^t + \eta_m \textstyle\sum_{i=1}^{\lambda} \widehat{w}_i (x_i - m^t)$$
$$C^{t+1} = C^t + \eta_C \textstyle\sum_{i=1}^{\lambda} \widehat{w}_i \left((x_i - m^t)(x_i - m^t)^{\mathrm{T}} - C^t\right)$$

except that $\eta_m = \eta_C = \delta t$ in the natural gradient update.

2.3 Maximum likelihood, IGO-ML, and cross-entropy

In the sequel, we prove monotone improvement of the objective function for a variant of IGO known as IGO-maximum likelihood (IGO-ML, introduced in [5, Section 3]). The result is then transferred to IGO because the two algorithms exactly coincide in an important class of cases, namely, exponential families using mean value parametrization.

The IGO-ML algorithm [5, Section 3] updates the current parameter value θ^t by taking a weighted maximum likelihood of the current distribution and the best sampled points. Assume as above that $\sum \widehat{w}_i = 1$. Then the *IGO-ML update* is defined as

$$\theta^{t+\delta t} = \arg\max_\theta \left\{ (1-\delta t)\, \mathbb{E}_{P_{\theta^t}} \left[\ln p_\theta(x)\right] + \delta t \sum_i \widehat{w}_i \ln p_\theta(x_i) \right\} \tag{9}$$

where we note that the first part is the cross-entropy of P_{θ^t} and P_θ, and thus, taken alone, is maximized for $\theta = \theta^t$. Taking the limit $\lambda \to \infty$, we also define the *infinite-population IGO-ML update* as

$$\theta^{t+\delta t} = \arg\max_\theta \left\{ (1-\delta t)\, \mathbb{E}_{P_{\theta^t}} \left[\ln p_\theta(x)\right] + \delta t H_t(\theta) \right\} \tag{10}$$

where we set

$$H_t(\theta) := \mathbb{E}_{P_{\theta^t}} \left[W_{\theta^t}^f(x) \ln p_\theta(x) \right]$$

a "weighted cross-entropy" of θ and θ^t.

Note that the finite- and infinite-population IGO-ML updates only make sense when there is a unique maximizer θ in (9) and (10), respectively. This assumption is always satisfied, for instance, for exponential families of probability distributions, as considered below (Statement 2).

The IGO-ML update is compatible with the IGO update, in the sense that for $\delta t \to 0$ the direction and magnitude of these updates coincide [5, Section 3].

The IGO-ML method is also related to the cross-entropy (CE) or maximum-likelihood (ML) method for optimization [7], which can be written as

$$\theta^{t+1} = \arg\max_\theta \sum_{i=1}^{\lambda} \widehat{w}_i \ln p_\theta(x_i)$$

and its smoothed version which reads [7]

$$\theta^{t+\delta t} = (1-\delta t)\theta^t + \delta t \arg\max_\theta \sum_{i=1}^{\lambda} \widehat{w}_i \ln p_\theta(x_i) \quad . \tag{11}$$

Note that IGO-ML is parametrization-independent whereas for $\delta t \neq 1$ the smoothed CE/ML method is not. Consequently, in general these updates will differ.

2.4 IGO and IGO-ML for Exponential Families

An *exponential family* is a set $\{p_\theta; \theta \in \Theta\}$ of probability density functions p_θ with respect to an arbitrary measure $\mathrm{d}x$ on X defined as

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp\left(\sum_{i=1}^{n} \beta_i(\theta) T_i(x) \right) \quad , \tag{12}$$

where $\beta = (\beta_i)_{1 \leq i \leq n}$ is the so-called *natural* (i.e. canonical) *parameter*; each T_i, $1 \leq i \leq n$ is a map $T_i : X \to \mathbb{R}$ such that $\{T_1, \ldots, T_n, x \mapsto 1\}$ are linearly independent; $Z(\theta)$ is the normalization factor. This linear independence ensures that the manifold of the exponential family is nonsingular. Many probability models, including multivariate Gaussian

distributions, are expressed as exponential families. See [4, Section 2.3] for examples.

If we define

$$\eta(\theta) := \mathbb{E}_{P_\theta}[T(x)] = \int T(x)\, p_\theta(x)\, \mathrm{d}x \ , \tag{13}$$

$\eta = (\eta_i)_{1 \leq i \leq n}$ is the so-called *expectation parameter*. For example, the expectation parameter for the multivariate Gaussian distribution encodes the first moment $\mathbb{E}_{P_\theta}[x]$ and the second moment $\mathbb{E}_{P_\theta}[xx^{\mathrm{T}}]$. Other examples can be found in [4, Section 3.5].

We will repeatedly and implicitly make use of the following well-known fact for exponential families.

STATEMENT 2. *Let $x_1, \ldots, x_k$ be k points in X and let $\alpha_1, \ldots, \alpha_k$ be non-negative numbers with $\sum \alpha_i = 1$. Then the value θ of the parameter such that the associated expectation parameter satisfies $\eta(\theta) = \sum \alpha_i T(x_i)$, if it belongs to the statistical manifold, is the unique maximizer of the weighted log-likelihood: $\theta = \arg\max \sum \alpha_i \ln p_\theta(x_i)$. An analogous statement holds if the finite sum is replaced with an integral or a combination of both.*

(Uniqueness boils down to strict concavity of $\ln p_\theta(x)$ as a function of θ. The restriction placed on η to belong to the statistical manifold is necessary: for instance, for Gaussian distributions, if the number of points k is not greater than the dimension of the ambient space, a degenerate distribution θ will result.)

The following statement from [5] shows that the natural gradient of a function in the expectation parametrization is given by the vanilla gradient of the function w.r.t. the normal parameter, and vice versa.

STATEMENT 3 (PROPOSITION 22 IN [5]). *Let g be a function on the statistical manifold of an exponential family as above. Then the components of the natural gradient w.r.t. the expectation parameters are given by the vanilla gradient w.r.t. the natural parameters and vice versa, that is,*

$$\tilde{\nabla}_{\eta_i} g = \frac{\partial g}{\partial \beta_i} \quad \textit{and} \quad \tilde{\nabla}_{\beta_i} g = \frac{\partial g}{\partial \eta_i} \ .$$

According to Statement 3, each component of the natural gradient of the log likelihood $\ln p_\theta(x)$ under the exponential parametrization $\theta = \eta$ is equivalent to each component of the vanilla gradient, i.e.,

$$\tilde{\nabla}_{\eta_i} \ln p_\theta(x) = \frac{\partial \ln p_\theta(x)}{\partial \beta_i} = T_i(x) - \eta_i \ , \tag{14}$$

where the latter equality is well-known, e.g., [4, (2.33)]. The IGO update (6) under the expectation parametrization thus reads

$$\eta^{t+\delta t} = \eta^t + \delta t\, \mathbb{E}_{P_{\theta^t}}[W^f_{\theta^t}(x)(T(x) - \eta^t)] \tag{15}$$

and the natural gradient update with finite sample size reads

$$\eta^{t+\delta t} = \eta^t + \delta t \sum_{i=1}^{\lambda} \widehat{w}_i \left(T(x_i) - \eta^t\right) \ . \tag{16}$$

Suppose as above that the selection weights sum to one: $\mathbb{E}_{P_\theta}[W^f_\theta(x)] = \int_0^1 w(q)\mathrm{d}q = 1$ and thus $\sum \widehat{w}_i = 1$. Then, IGO has a close relation with the CE/ML for optimization. As is stated in Theorem 15 in [5], for an exponential family the CE/ML method (11) and the IGO instance (16), when expressed with the expectation parametrization ($\theta = \eta$), coincide with IGO-ML (9).

STATEMENT 4 (THEOREM 15 IN [5]). *For optimization using an exponential family $\{P_\theta\}$, these three algorithms coincide: IGO-ML; the IGO expressed in expectation parameters; the CE/ML expressed in expectation parameters. That is, for an exponential family with the expectation parametrization, for $0 \leq \delta t \leq 1$ we have (writing in turn IGO, CE/ML and IGO-ML)*

$$\begin{aligned}
\theta^{t+\delta t} &= \theta^t + \delta t \sum_{i=1}^{\lambda} \widehat{w}_i \left(T(x_i) - \theta^t\right) \\
&= (1 - \delta t)\theta^t + \delta t \arg\max_\theta \sum_{i=1}^{\lambda} \widehat{w}_i \ln p_\theta(x_i) \\
&= \arg\max_\theta \Big\{ (1 - \delta t)\mathbb{E}_{P_{\theta^t}}\left[\ln p_\theta(x)\right] \\
&\qquad\qquad + \delta t \sum_{i=1}^{\lambda} \widehat{w}_i \ln p_\theta(x_i) \Big\} \ .
\end{aligned} \tag{17}$$

In the limit of infinite sample size $\lambda \to \infty$ this rewrites

$$\begin{aligned}
\theta^{t+\delta t} &= \theta^t + \delta t\, \mathbb{E}_{P_{\theta^t}}\left[W^f_{\theta^t}(x)\left(T(x) - \theta^t\right)\right] \\
&= (1 - \delta t)\theta^t + \delta t \arg\max_\theta H_t(\theta) \\
&= \arg\max_\theta \Big\{ (1 - \delta t)\mathbb{E}_{P_{\theta^t}}\left[\ln p_\theta(x)\right] + \delta t H_t(\theta) \Big\}
\end{aligned} \tag{18}$$

where we recall that $H_t(\theta) = \mathbb{E}_{P_{\theta^t}}\left[W^f_{\theta^t}(x) \ln p_\theta(x)\right]$.

Remark 1. Malagò et al. [16] study information-geometric aspects of exponential families for optimization. One difference from the IGO framework is that the optimization problem is defined as the minimization of the expectation of the objective function over P_θ, namely

$$\min_\theta \mathbb{E}_{P_\theta}[f(x)],$$

which they call the stochastic relaxation of the original optimization problem. They study this for an exponential family on a discrete search space with the natural parametrization ($\theta = \beta$) and propose the natural gradient descent algorithm. Note that this requires computation of the empirical Fisher information matrix to perform natural gradient descent. However, if the algorithm is modified to use the expectation parameters instead, one can compute the natural gradient descent directly as

$$\eta^{t+\delta t} = \eta^t - \delta t \sum_{i=1}^{\lambda} f(x_i) \left(T(x_i) - \eta^t\right) \ . \tag{19}$$

We study this algorithm in Section 4.

3. QUANTILE IMPROVEMENT

One possible way to provide theoretical backing for an optimization algorithm is to show monotonic improvement at each step of the algorithm (although this is by no means necessary: e.g., for stochastic algorithms, this is not expected to hold at each step). For example, consider the sphere function $f : x \mapsto \|x\|^2$. Then, it is easy to show that the gradient steps $x^{t+\delta t} = x^t - \delta t\, \nabla_x f(x^t)$ generate a monotonically decreasing sequence $\{f(x^t)\}_{t \geq 0}$ provided $0 < \delta t \leq 1/2$. For any smooth function, infinitesimal gradient steps are guaranteed to improve the objective function values; but in general

the admissible step size strongly depends on the function and has to be adjusted by the user.

When it comes to the counterpart in IGO, however, we follow the gradient of the function $J(\theta \,|\, \theta^t)$, which depends on θ^t, so that step-by-step improvement in the objective, $J(\theta^{t+1} \,|\, \theta^t) > J(\theta^t \,|\, \theta^t)$, does not necessarily mean improvement. (It might happen that $J(\theta^t \,|\, \theta^{t+1}) > J(\theta^{t+1} \,|\, \theta^{t+1})$ and $J(\theta^{t+1} \,|\, \theta^t) > J(\theta^t \,|\, \theta^t)$ at the same time.)

A key feature of the IGO framework is its invariance under changing the objective function f by an increasing transformation (e.g. optimizing f^3 instead of f). Thus, any measure of progress that is not compatible with such transformations (e.g. the expectation $\mathbb{E}_{P_{\theta^t}} f$) is not a good candidate to always improve over the course of IGO optimization.

As a measure of improvement, Arnold et al. [5] use the notion of *q-quantile of f*. The q-quantile $Q^q_P(f)$ of f under a probability distribution P is any number m such that $P[x : f(x) \leq m] \geq q$ and $P[x : f(x) \geq m] \geq 1 - q$. For instance, $Q^q_P(f)$ is the median value of f under P if $q = 1/2$. For smooth distributions and continuous f there is only one such number m, but in general the set of such m may be a closed interval, for instance if f has "jumps". For the sake of definiteness let us use the largest such value:

$$Q^q_P(f) := \sup \big\{ m \in \mathbb{R} : P[x : f(x) \leq m] \geq q \text{ and } P[x : f(x) \geq m] \geq 1 - q \big\} \ .$$

(This is because we want to minimize the objective function f; when IGO is used for maximization instead, Theorem 6 has to be written using an infimum in the definition of $Q^q_P(f)$ instead.)

It is proven in [5] that when using the q-truncation selection scheme, the q-quantile value of f monotonically decreases along infinitesimal IGO steps.

STATEMENT 5 (PROPOSITION 5 IN [5]). *Consider the q-truncation selection scheme $w(u) = \mathbb{1}_{[u \leq q]}/q$ where $0 < q < 1$ is fixed. Then each infinitesimal IGO step* (6) *where δt is infinitesimal leads to monotonic improvement in the q-quantile of f: $Q^q_{P_{\theta^{t+\delta t}}}(f) \leq Q^q_{P_{\theta^t}}(f)$.*

3.1 Quantile Improvement in IGO-ML

In practice, explicit algorithms do not use continuous time with infinitesimal time steps: the time step δt may be quite large and its calibration may be an important issue. It is more interesting and important to see how long steps we can take along the natural gradient, i.e. how large a δt we can choose while guaranteeing q-quantile improvement.

When using IGO-ML (and thus when using IGO or CE/ML on an exponential family with the expectation parametrization), we can obtain such a conclusion; the size of the steps may even be chosen independently of the objective function.

THEOREM 6. *Let the selection scheme be $w(u) = \mathbb{1}_{[u \leq q]}/q$ where $0 < q < 1$. Assume that the* arg max *defining the IGO-ML step* (10) *is uniquely determined. Then for $0 < \delta t \leq 1$, each infinite-population IGO-ML step* (10) *leads to q-quantile improvement: $Q^q_{P_{\theta^{t+\delta t}}}(f) \leq Q^q_{P_{\theta^t}}(f)$.*

Moreover, equality can hold only if $P_{\theta^{t+\delta t}} = P_{\theta^t}$ or if $P_{\theta^{t+\delta t}}[x : f(x) = Q^q_{P_{\theta^t}}(f)] > 0$.

COROLLARY 7. *For exponential families written in expectation parameters, on any search space, the same holds for the CE/ML method and for the IGO algorithm.*

Note that the first condition for equality means the algorithm has reached a stable point.

The second condition for equality typically happens for discrete search spaces: on such spaces, the q-quantile evolves in time by discrete jumps even when θ^t moves smoothly, so we cannot expect strict quantile improvement at each step. On the other hand, with continuous distributions on continuous search spaces, the second equality condition can only occur if the objective function has a plateau (a level set with non-zero measure).

PROOF. If $P_{\theta^{t+\delta t}} = P_{\theta^t}$, obviously $Q^q_{P_{\theta^{t+\delta t}}}(f) = Q^q_{P_{\theta^t}}(f)$. Hereunder, we assume $P_{\theta^{t+\delta t}} \neq P_{\theta^t}$.

Consider the function $J(\theta \,|\, \theta^t)$ defining the expected P_{θ^t}-adjusted fitness of a random point under P_θ:

$$J(\theta \,|\, \theta^t) = \mathbb{E}_{P_\theta}\big[W^f_{\theta^t}(x)\big]$$

and remember that $J(\theta^t \,|\, \theta^t) = 1$. The idea is as follows: letting Y be the set of points with $P_{\theta^t}(Y) = q$ at which the objective function f is smallest (the sublevel set of f with P_{θ^t}-mass q), then with our choice of w, $W^f_{\theta^t}(x)$ is (up to technicalities) equal to $1/q$ on Y and 0 elsewhere, so that $J(\theta \,|\, \theta^t)$ represents $1/q$ times the P_θ-probability of falling into Y (hence $J(\theta^t \,|\, \theta^t) = 1$). Thus $J(\theta \,|\, \theta^t) > 1$ will mean that the P_θ-probability of falling into Y is larger than q, so that P_θ improves over P_{θ^t} and the q-quantile has decreased.

We are going to prove that the IGO-ML update satisfies $J(\theta^{t+\delta t} \,|\, \theta^t) > 1$ if $P_{\theta^t} \neq P_{\theta^{t+\delta t}}$. More precisely we prove that

$$J(\theta^{t+\delta t} \,|\, \theta^t) > \exp\left(\frac{1 - \delta t}{\delta t} D_{\mathrm{KL}}(P_{\theta^t} \,\|\, P_{\theta^{t+\delta t}})\right) \ .$$

This will imply quantile improvement, thanks to the following lemma, the proof of which is postponed.

LEMMA 8. *Let the selection scheme w be as above. If $J(\theta^{t+\delta t} \,|\, \theta^t) > 1$, then $Q^q_{P_{\theta^{t+\delta t}}}(f) \leq Q^q_{P_{\theta^t}}(f)$. If moreover $P_{\theta^{t+\delta t}}[x : f(x) = Q^q_{P_{\theta^t}}(f)] = 0$, then $Q^q_{P_{\theta^{t+\delta t}}}(f) < Q^q_{P_{\theta^t}}(f)$.*

The lower bound on $J(\theta^{t+\delta t} \,|\, \theta^t)$ is obtained as follows. Since $\int W^f_{\theta^t}(x) p_{\theta^t}(x) \mathrm{d}x = 1$ and $W^f_{\theta^t}(x) p_{\theta^t}(x) \geq 0$ for any x, $W^f_{\theta^t}(x) p_{\theta^t}(x)$ can be viewed as a probability density function. Since ln is concave, by Jensen's inequality we have

$$\begin{aligned} \ln J(\theta \,|\, \theta^t) &= \ln \int \frac{p_\theta(x)}{p_{\theta^t}(x)} W^f_{\theta^t}(x) p_{\theta^t}(x) \mathrm{d}x \\ &\geq \int \ln\left(\frac{p_\theta(x)}{p_{\theta^t}(x)}\right) W^f_{\theta^t}(x) p_{\theta^t}(x) \mathrm{d}x \\ &= H_t(\theta) - H_t(\theta^t) \ . \end{aligned} \tag{20}$$

Thus, if $H_t(\theta) > H_t(\theta^t)$ we have $J(\theta \,|\, \theta^t) > 1$.

Now, according to (10), $\theta^{t+\delta t}$ uniquely maximizes the quantity $(1-\delta t)\mathbb{E}_{P_{\theta^t}}\big[\ln p_\theta(x)\big] + \delta t H_t(\theta)$. Therefore, if $\theta^{t+\delta t} \neq \theta^t$, we have

$$\begin{aligned} (1 - \delta t)\mathbb{E}_{P_{\theta^t}}\big[\ln p_{\theta^{t+\delta t}}(x)\big] + \delta t H_t(\theta^{t+\delta t}) \\ > (1 - \delta t)\mathbb{E}_{P_{\theta^t}}\big[\ln p_{\theta^t}(x)\big] + \delta t H_t(\theta^t) \end{aligned}$$

and rearranging we get

$$H_t(\theta^{t+\delta t}) - H_t(\theta^t) > \frac{1-\delta t}{\delta t}\left(\mathbb{E}_{P_{\theta^t}}\left[\ln p_{\theta^t}(x)\right] - \mathbb{E}_{P_{\theta^t}}\left[\ln p_{\theta^{t+\delta t}}(x)\right]\right) = \frac{1-\delta t}{\delta t} D_{\mathrm{KL}}(P_{\theta^t} \,\|\, P_{\theta^{t+\delta t}}) \ . \quad (21)$$

The right-hand side of this inequality is non-negative for $0 < \delta t \leq 1$.

This will prove the theorem once Lemma 8 is proved, which we now proceed to do. □

Proof of Lemma 8. Hereunder, we abbreviate m for the q-quantile value $Q^q_{P_{\theta^t}}(f)$ of f under P_{θ^t}.

Let us compute the weighted preference $W^f_{\theta^t}(x)$. Since the selection scheme w satisfies $0 \leq w(u) \leq 1/q$ for all $u \in [0;1]$, we have $0 \leq W^f_{\theta^t}(x) \leq 1/q$ for any x.

We claim that $f(x) > m$ implies $W^f_{\theta^t}(x) = 0$. Indeed, suppose that x is such that $f(x) > m$. Since by definition m is the largest value such that $P_{\theta^t}[y : f(y) \geq m] \geq 1-q$, we must have $P_{\theta^t}[y : f(y) \geq f(x)] < 1-q$. Hence $P_{\theta^t}[y : f(y) < f(x)] > q$, i.e., $q^<_{\theta^t}(x) > q$. Now this implies $W^f_{\theta^t}(x) = 0$ for our choice of selection scheme w.

Thus $W^f_{\theta^t}(x)$ is at most $1/q$ and vanishes if $f(x) > m$. For any probability distribution P_θ, this implies that

$$J(\theta \,|\, \theta^t) = \mathbb{E}_{P_\theta}[W^f_{\theta^t}(x)] \leq \frac{1}{q} P_\theta[x : f(x) \leq m] \ .$$

Therefore,

$$J(\theta \,|\, \theta^t) > 1 \Longrightarrow P_\theta[x : f(x) \leq m] > q \Longrightarrow Q^q_{P_\theta}(f) \leq m \ .$$

If moreover $P_\theta[x : f(x) = m] = 0$, we have $P_\theta[x : f(x) \leq m] = P_\theta[x : f(x) < m]$ hence

$$J(\theta \,|\, \theta^t) > 1 \Longrightarrow P_\theta[x : f(x) < m] > q \Longleftrightarrow P_\theta[x : f(x) \geq m] < 1-q \Longrightarrow Q^q_{P_\theta}(f) < m \ .$$

Altogether, $J(\theta^{t+\delta t} \,|\, \theta^t) > 1$ implies quantile improvement $Q^q_{P_{\theta^{t+\delta t}}}(f) \leq Q^q_{P_{\theta^t}}(f)$. Moreover, if $P_{\theta^{t+\delta t}}[x : f(x) = m] = 0$, we have strict quantile improvement $Q^q_{P_{\theta^{t+\delta t}}}(f) < Q^q_{P_{\theta^t}}(f)$. □

This completes the proof of Theorem 6.

Example 3. Bernoulli distributions constitute an exponential family where the sufficient statistics $T_i(x)$ are x_i. The parameter θ used in PBIL (Example 1) is indeed the expectation parameter. Thus, PBIL is an instance of IGO-ML and can be viewed as a CE/ML method at the same time. Hence, by Theorem 6, each infinite-population PBIL step leads to q-quantile improvement if we employ q-truncation selection, which is not the same as the exponential weights introduced in [6].

Remark 2. The proof of the theorem is quantitative: the Kullback–Leibler divergence $D_{\mathrm{KL}}(P_{\theta^t} \,\|\, P_{\theta^{t+\delta t}})$ indicates how much progress was made. More precisely (assuming for simplicity a continuous situation with no plateaus), while the probability under P_{θ^t} to fall into the best q percent of points for P_{θ^t} is q by definition, the probability under $P_{\theta^{t+\delta t}}$ to fall into the best q percent of points for P_{θ^t} is at least $q \exp\left(\frac{1-\delta t}{\delta t} D_{\mathrm{KL}}(P_{\theta^t} \,\|\, P_{\theta^{t+\delta t}})\right)$.

3.2 Blockwise IGO-ML

The expectation parameter is not always the most obvious one. When it comes to multivariate Gaussian distributions, the expectation parameter is the mean vector and second moment, $(m,\ mm^{\mathrm{T}} + C)$. Meanwhile, the CMA-ES and the CE/ML method for continuous optimization parametrize the mean vector and covariance matrix, hence they differ from the IGO-ML algorithm. Moreover, sometimes different step sizes (learning rates) are employed for each parameter, which makes the direction of parameter update different from that of the natural gradient. Here, we justify some of these settings by guaranteeing q-quantile improvement in an extended framework.

We define an extension of IGO-ML, *blockwise IGO-ML*, that recovers the pure rank-μ CMA-ES update with different learning rates for m and C.

Definition 1. Let $\theta = (\theta_1, \ldots, \theta_k)$ be any decomposition of the parameter θ into k blocks, and let $\{\delta t_1, \ldots, \delta t_k\}$ be a step size for each block. For $1 \leq j \leq k$, define the *j-th block partial IGO-ML update* with step size δt_j as the map sending a parameter value θ to $\Phi_j(\theta)$ where

$$\Phi_j(\theta) := \underset{\substack{\theta^* \\ \theta^*_i = \theta_i \text{ for all } i \neq j}}{\arg\max} \left\{ (1-\delta t_j)\mathbb{E}_{P_\theta}\left[\ln p_{\theta^*}(x)\right] + \delta t_j \sum_i \widehat{w}_i \ln p_{\theta^*}(x_i) \right\} \ . \quad (22)$$

The *blockwise IGO-ML* updates the parameter θ as follows. Given a current parameter value θ^t, update the first block of θ^t, then the second block, etc., in that order; explicitly, set

$$\theta^{t+1} := (\Phi_k \circ \cdots \circ \Phi_2 \circ \Phi_1)(\theta^t) \ , \quad (23)$$

where we note that the same Monte Carlo sample $\{x_i\}$ from P_{θ^t} is used throughout the whole range of block updates $\Phi_1, \ldots, \Phi_k$.

The infinite-population step ($\lambda = \infty$) reads the same with

$$\Phi_j(\theta) := \underset{\substack{\theta^* \\ \theta^*_i = \theta_i \text{ for all } i \neq j}}{\arg\max} \left\{ (1-\delta t_j)\mathbb{E}_{P_\theta}\left[\ln p_{\theta^*}(x)\right] + \delta t_j\, \mathbb{E}_{P_{\theta^t}}\left[W^f_{\theta^t}(x) \ln p_{\theta^*}(x)\right] \right\} \ . \quad (24)$$

As before, the finite- and infinite-population blockwise IGO-ML updates only make sense if the arg max in (22) or (24) is uniquely determined.

Note that the blockwise IGO-ML depends on the decomposition of the parameters into blocks and their update order, while it is independent of the parametrization inside each block. Blockwise IGO-ML is not necessarily equivalent to IGO-ML even when all δt_i are equal to δt.

Proposition 9. *The pure rank-μ CMA-ES update (Example 2) is an instance of blockwise IGO-ML for Gaussian distributions, with parameter decomposition $\theta = (\theta_1, \theta_2)$ where $\theta_1 = C$, the covariance matrix, and $\theta_2 = m$, the mean vector.*

PROOF. Given $\theta^t = (C^t, m^t)$, blockwise IGO-ML first updates C as follows:

$$C^* = \arg\max_C \left\{ (1-\delta t_C)\mathbb{E}_{P_{(C^t,m^t)}}\left[\ln p_{(C,m^t)}(x)\right] + \delta t_C \sum_i \widehat{w}_i \ln p_{(C,m^t)}(x_i) \right\} . \quad (25)$$

Considering $\{P_{(C,m^t)}\}$ as an exponential family of Gaussian distributions whose mean vector is fixed to m^t, (25) can be viewed as an ordinary IGO-ML step for this restricted model. Then, since (after shifting the origin of the coordinate system to m^t) C is the expectation parameter of the restricted model, the update is given by (17) namely

$$C^* = C^t + \delta t_C \sum_{i=1}^{\lambda} \widehat{w}_i \big((x_i - m^t)(x_i - m^t)^{\mathrm{T}} - C^t\big) .$$

Next, m is updated as

$$m^* = \arg\max_m \left\{ (1-\delta t_m)\mathbb{E}_{P_{(C^*,m^t)}}\left[\ln p_{(C^*,m)}(x)\right] + \delta t_m \sum_i \widehat{w}_i \ln p_{(C^*,m)}(x_i) \right\} .$$

To derive m^*, let us differentiate the inside of arg max w.r.t. m and derive the zero point of the derivative. Seeing that $\nabla_m \ln p_{(C^*,m)}(x) = (C^*)^{-1}(x-m)$, we find the condition

$$(1-\delta t_m)(C^*)^{-1}(m^t - m^*) + \delta t_m \sum_i \widehat{w}_i (C^*)^{-1}(x_i - m^*) = 0 ,$$

which holds if and only if

$$m^* = m^t + \delta t_m \sum_{i=1}^{\lambda} \widehat{w}_i (x_i - m^t) .$$

This is equivalent to the pure rank-μ CMA-ES update. □

Quantile improvement as in Theorem 6 readily extends to this setting as follows.

THEOREM 10. *Let the selection scheme be $w(u) = \mathbb{1}_{[u\le q]}/q$ where $0 < q < 1$. Assume that the* arg max *defining each partial infinite-population IGO-ML update* (24) *is uniquely determined. Then for $0 < \delta t_j \le 1$ ($j \in [\![1;k]\!]$), each infinite-population blockwise IGO-ML step* (23) *leads to q-quantile improvement: $Q^q_{P_{\theta^{t+1}}}(f) \le Q^q_{P_{\theta^t}}(f)$.*

Moreover, equality can hold only if $P_{\theta^{t+1}} = P_{\theta^t}$ or if $P_{\theta^{t+1}}[x : f(x) = Q^q_{P_{\theta^t}}(f)] > 0$.

Consequently, each infinite-population step of the pure rank-μ CMA-ES update guarantees q-quantile improvement. Indeed, from Proposition 9 this variant of the CMA-ES is an instance of blockwise IGO-ML. Moreover, if each level set of f has zero Lebesgue measure, which often holds for continuous optimization, we have strict q-quantile improvement.

PROOF. If $P_{\theta^{t+1}} = P_{\theta^t}$, obviously $Q^q_{P_{\theta^{t+1}}}(f) = Q^q_{P_{\theta^t}}(f)$. We assume $P_{\theta^{t+1}} \ne P_{\theta^t}$ in the following.

Set $\theta^{t,0} := \theta^t$ and $\theta^{t,j} := \Phi_j(\theta^{t,j-1})$ so that $\theta^{t+1} = \theta^{t,k}$. According to Lemma 8, to prove quantile improvement it is enough to show that $J(\theta^{t+1} \,|\, \theta^t) > 1$. Moreover, this implies strict quantile improvement provided $P_{\theta^{t+1}}[x : f(x) = Q^q_{P_{\theta^t}}(f)] = 0$.

According to (20), if $H_t(\theta^{t+1}) > H_t(\theta^t)$ we have $J(\theta^{t+1} \,|\, \theta^t) > 1$. To show that $H_t(\theta^{t+1}) > H_t(\theta^t)$ we decompose $H_t(\theta^{t+1}) - H_t(\theta^t)$ into the sum of partial differences, namely,

$$H_t(\theta^{t+1}) - H_t(\theta^t) = \sum_{j=1}^{k} H_t(\theta^{t,j}) - H_t(\theta^{t,j-1}) ,$$

and we will prove that each term is non-negative. Moreover, if $P_{\theta^{t,j}} \ne P_{\theta^{t,j-1}}$ for some $j \in [\![1;k]\!]$, we will have $H_t(\theta^{t,j}) - H_t(\theta^{t,j-1}) > 0$ for this j. Since $P_{\theta^{t+1}} \ne P_{\theta^t}$ implies $P_{\theta^{t,j}} \ne P_{\theta^{t,j-1}}$ for at least one $j \in [\![1;k]\!]$, we will have that $H_t(\theta^{t+1}) - H_t(\theta^t) > 0$, resulting in $J(\theta^{t+1} \,|\, \theta^t) > 1$.

We proceed as in Theorem 6. Since $\theta^{t,j} = \Phi_j(\theta^{t,j-1})$ is the only maximizer of (24), we have

$$(1-\delta t_j)\mathbb{E}_{P_{\theta^{t,j-1}}}\left[\ln p_{\theta^{t,j}}(x)\right] + \delta t_j H_t(\theta^{t,j}) \ge (1-\delta t_j)\mathbb{E}_{P_{\theta^{t,j-1}}}\left[\ln p_{\theta^{t,j-1}}(x)\right] + \delta t_j H_t(\theta^{t,j-1})$$

with equality holding if and only if $\theta^{t,j} = \theta^{t,j-1}$. Rearranging, we get

$$H_t(\theta^{t,j}) - H_t(\theta^{t,j-1}) > \frac{1-\delta t_j}{\delta t_j} D_{\mathrm{KL}}(P_{\theta^{t,j-1}} \,\|\, P_{\theta^{t,j}})$$

if $\theta^{t,j} \ne \theta^{t,j-1}$, and $H_t(\theta^{t,j}) = H_t(\theta^{t,j-1})$ if $\theta^{t,j} = \theta^{t,j-1}$. The right-hand side of the above inequality is non-negative for $0 < \delta t \le 1$. Therefore, $H_t(\theta^{t,j}) - H_t(\theta^{t,j-1}) \ge 0$ for all $j \in [\![1;k]\!]$. Moreover, since $P_{\theta^{t+1}} \ne P_{\theta^t}$, for at least one $j \in [\![1;k]\!]$ we have $\theta^{t,j} \ne \theta^{t,j-1}$ and thus $H_t(\theta^{t,j}) - H_t(\theta^{t,j-1}) > 0$ for this j, implying that $H_t(\theta^{t+1}) - H_t(\theta^t) > 0$. This completes the proof. □

3.3 Finite Population Sizes

The results above are valid for "ideal" updates with infinite sample size. With finite sample size, the update (9) defines a stochastic sequence (depending on the random sample $\{x_i\}$) and so one cannot expect monotone q-quantile improvement at each step. Still, we can expect q-quantile improvement with high probability when the population size is sufficiently large.

We provide an analogue of Theorem 6 for finite but large population size. A similar statement holds for blockwise IGO-ML. The proof follows a standard probabilistic approximation argument.

PROPOSITION 11. *Let $w(\cdot)$ be the q-truncation selection scheme: $w(u) = \mathbb{1}_{[u\le q]}/q$ where $0 < q < 1$. Let $\{P_\theta\}$ be an exponential family of probability distributions, parametrized by its expectation parameter. Assume that the* arg max *defining the infinite-population IGO-ML step* (10) *is uniquely defined.*

Assume that for all $\theta \in \Theta$, the derivative $\partial \ln P_\theta(x)/\partial\theta$ exists for P_θ-almost all $x \in X$ and has finite second moment: $\mathbb{E}_{P_\theta}\left[\left|\partial \ln P_\theta(x)/\partial\theta\right|^2\right] < \infty$.

Let $0 < \delta t \le 1$. Let $\theta^{t+\delta t}_\lambda$ be the IGO-ML update (9) *with sample size λ, and let $\theta^{t+\delta t}_\infty$ be the infinite-population IGO-ML update* (10)*. Assume that $\theta^{t+\delta t}_\infty \ne \theta^t$.*

Then, with probability tending to 1 as $\lambda \to \infty$, the finite-population update $\theta^{t+\delta t}_\lambda$ results in q-quantile improvement:

$$Q^q_{P_{\theta^{t+\delta t}_\lambda}}(f) \le Q^q_{P_{\theta^t}}(f) .$$

Consequently, the same holds for the CE/ML method and the IGO algorithm when they are applied to an exponential family using the expectation parameters.

Note the assumption that the *ideal* dynamics has not reached equilibrium yet: $\theta_\infty^{t+\delta t} \neq \theta^t$. If $\theta_\infty^{t+\delta t} = \theta^t$, the finite-population dynamics will just randomly wander around this equilibrium value with some noise, resulting in either improvement or deterioration at each step.

Also note that the population size λ needed may depend on the current location θ^t in parameter space, as well as the objective function f. For instance, highly oscillating functions f likely require higher population sizes for a consistent estimation of the IGO-ML update.

PROOF. For exponential families, the IGO and IGO-ML updates coincide. Under the conditions of the theorem, the finite-population IGO update (8) is a consistent estimator of the infinite-population IGO update (5) [5, Proposition 18], implying that $\theta_\lambda^{t+\delta t}$ converges with probability one to $\theta_\infty^{t+\delta t}$. Under our regularity assumptions on P_θ, this implies pointwise convergence of $p_{\theta_\lambda^{t+\delta t}}$ to $p_{\theta_\infty^{t+\delta t}}$, which, since $W_{\theta^t}^f(x)$ is bounded, leads to

$$J(\theta_\lambda^{t+\delta t} \mid \theta^t) = \mathbb{E}_{P_{\theta_\lambda^{t+\delta t}}}[W_{\theta^t}^f(x)] \to \mathbb{E}_{P_{\theta_\infty^{t+\delta t}}}[W_{\theta^t}^f(x)] = J(\theta_\infty^{t+\delta t} \mid \theta^t) \quad \text{as } \lambda \to \infty.$$

Now the right-hand side is greater than 1 for $0 < \delta t \leq 1$ unless $\theta_\infty^{t+\delta t} = \theta^t$, as we have shown in the proof of Theorem 6. Thus, we have $J(\theta_\lambda^{t+\delta t} \mid \theta^t) > 1$ with high probability for sufficiently large λ. Thus Lemma 8 entails q-quantile improvement with high probability. □

4. FITNESS-PROPORTIONAL SELECTION

These results carry over to the use of a composite $g \circ f$ of a function g with the objective function f, as a selection weight instead of $W_{\theta^t}^f$ in the IGO framework. This covers, for instance, fitness-proportional selection ($g = \mathrm{Id}$). We prove that, when considering the natural gradient ascent for an exponential family (12) using the expectation parameter (13), we can guarantee monotone $\mathbb{E}_{P_\theta}[g \circ f(x)]$-value improvement for updates of step size inversely proportional to $\mathbb{E}_{P_\theta}[g \circ f(x)]$. More precisely,

THEOREM 12. *Assume $g \circ f$ is non-negative and not almost everywhere* 0. *Consider the update*

$$\theta^{t+\delta t} = \theta^t + \delta t\, \mathbb{E}_{P_{\theta^t}}\left[\frac{g \circ f(x)}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]}\left(T(x) - \theta^t\right)\right], \quad (26)$$

where $\theta = \eta$ is the expectation parameter of the exponential family $\{P_\theta\}$.

Then for $0 < \delta t \leq 1$, we have

$$\mathbb{E}_{P_{\theta^{t+\delta t}}}[g \circ f(x)] \geq \mathbb{E}_{P_{\theta^t}}[g \circ f(x)] .$$

Moreover, equality can occur only if $P_{\theta^{t+\delta t}} = P_{\theta^t}$.

Gradient based methods with fitness-proportional selection are often employed, especially in reinforcement learning, e.g. *policy gradient with parameter based exploration* (PGPE) [17]. One disadvantage of gradient based methods is that the step size has to be calibrated by the user depending on the problem at hand. Alternative methods such as *expectation-maximization* [9], including the RPP below, are sometimes employed to avoid this issue. Theorem 12, however, ensures that each natural gradient step improves the expected fitness for $0 < \delta t \leq 1$ when an exponential family is used with its expectation parameters.

Example 4. The Relative Payoff Procedure (RPP) [12] is a reinforcement learning algorithm, also known as expectation-maximization (EM) algorithm for reinforcement learning [9]. The RPP expresses a policy on the action space $X = \{0,1\}^d$ by a Bernoulli distribution $P_\theta(x)$ parametrized by the expectation parameter. The objective to be maximized is the expectation $\mathbb{E}_{P_\theta}[r(x)]$ of non-negative reward $r(x)$ after taking action $x \in X$. The RPP updates the parameters to

$$\theta^{t+1} = \frac{\mathbb{E}_{P_{\theta^t}}[xr(x)]}{\mathbb{E}_{P_{\theta^t}}[r(x)]} .$$

Remember the sufficient statistics $T(x)$ for Bernoulli distributions are $T_i(x) = x_i$. Thus the RPP is equivalent to (26) with $g \circ f(x) = r(x)$ and $\delta t = 1$ and can be viewed as a natural gradient ascent with large step.

The RPP is known from [9] to monotonically improve expected reward, thanks to its expectation-maximization interpretation. Theorem 12 can be thought of as an extension of this result, and also shows monotone improvement for the smoothed RPP, where a step size $0 < \delta t \leq 1$ is introduced.

PROOF. Most of the proof of Theorem 6 carries over. Replacing $W_{\theta^t}^f$ in (18) with $g \circ f / \mathbb{E}_{P_{\theta^t}}[g \circ f(x)]$, (18) still holds and we have

$$\begin{aligned}\theta^{t+\delta t} &= \theta^t + \delta t\, \mathbb{E}_{P_{\theta^t}}\left[\frac{g \circ f(x)}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]}\left(T(x) - \theta^t\right)\right] \\ &= \arg\max_\theta \Big\{ (1 - \delta t)\mathbb{E}_{P_{\theta^t}}[\ln p_\theta(x)] \\ &\qquad + \delta t \mathbb{E}_{P_{\theta^t}}\left[\frac{g \circ f(x)}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]} \ln p_\theta(x)\right]\Big\} \end{aligned} \quad (27)$$

Thanks to Jensen's inequality, we have the counterpart of (20) as

$$\ln \mathbb{E}_{P_\theta}[g \circ f(x)] - \ln \mathbb{E}_{P_{\theta^t}}[g \circ f(x)] \geq \frac{\mathbb{E}_{P_{\theta^t}}[g \circ f(x) \ln p_\theta(x)]}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]} - \frac{\mathbb{E}_{P_{\theta^t}}[g \circ f(x) \ln p_{\theta^t}(x)]}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]} . \quad (28)$$

Because of the second equality of (27), we have the counterpart of (21) as

$$\frac{\mathbb{E}_{P_{\theta^t}}[g \circ f(x) \ln p_{\theta^{t+\delta t}}(x)]}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]} - \frac{\mathbb{E}_{P_{\theta^t}}[g \circ f(x) \ln p_{\theta^t}(x)]}{\mathbb{E}_{P_{\theta^t}}[g \circ f(x)]} \geq \frac{1 - \delta t}{\delta t} D_{\mathrm{KL}}(P_{\theta^t} \,\|\, P_{\theta^{t+\delta t}}),$$

and moreover, since the maximizer in (27) is unique, the inequality is strict unless $\theta^t = \theta^{t+\delta t}$. Hence, since the right-hand side is non-negative, by (28) we have $\ln \mathbb{E}_{P_\theta}[g \circ f(x)] \geq \ln \mathbb{E}_{P_{\theta^t}}[g \circ f(x)]$ with equality only if $P_{\theta^t} = P_{\theta^{t+\delta t}}$. This completes the proof. □

Remark 3. As mentioned in Remark 1, Malagò et al. [16] propose the natural gradient algorithm for discrete optimization using exponential distributions. However, as they parametrize the exponential distributions by the natural parameters $\theta = \beta$, Theorem 12 does not guarantee expected

fitness improvement for their algorithms, whereas it does so for the algorithm (19) using the expectation parameters.

5. FURTHER DISCUSSION

These results contribute to bringing theory closer to practice, by waiving the need for infinitesimal step sizes in gradient ascent. Still, they cover only the "ideal" situation with infinite population size, as well as finite but very large population sizes (by a standard probabilistic approximation argument). Finite population sizes lead to stochastic behavior and so monotone objective improvement at each step occurs only with high probability.

In practice, population sizes used can be quite small, $\lambda \leq 10$, with medium to small step sizes [6,11]. It has been shown in [1, Remark 2] that when population size does not tend to infinity, the expectation of the natural gradient estimate (8) is the natural gradient (5) with a *different* selection scheme w. So using the truncation weight $w(u) = \mathbb{1}_{[u \geq q]}$ with a small population size and very small step sizes will result, by the machinery of stochastic approximation [8, 14], in simulating an infinite-population IGO step with another selection scheme, a situation outside the scope of this article. Our results, on the contrary, suggest using larger populations and larger step sizes instead.

Finally, let us stress that objective improvement is not, by itself, a sufficient guarantee that optimization performs well: in situations of premature convergence, the objective still improves at each step. Premature convergence can occur for large values of the learning rate in some instantiations of IGO and IGO-ML (see the study in [5]); our results say nothing about this phenomenon.

6. REFERENCES

[1] Y. Akimoto, A. Auger, and N. Hansen. Convergence of the continuous time trajectories of isotropic evolution strategies on monotonic $\mathcal{C}^2$-composite functions. In *Parallel Problem Solving from Nature - PPSN XII, 12th International Conference*, number 7491 in Lecture Notes in Computer Science, pages 42–51, Taormina, Italy, September 1–5 2012. Springer.

[2] Y. Akimoto, Y. Nagata, I. Ono, and S. Kobayashi. Bidirectional relation between CMA evolution strategies and natural evolution strategies. In R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, editors, *Parallel Problem Solving from Nature - PPSN XI, 11th International Conference*, volume 6238 of *Lecture Notes in Computer Science*, pages 154–163, Kraków, Poland, September 11–15 2010. Springer.

[3] S.-i. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

[4] S.-i. Amari and H. Nagaoka. *Methods of Information Geometry*. Translations of Mathematical Monographs vol. 191. American Mathematical Society, 2000.

[5] L. Arnold, A. Auger, N. Hansen, and Y. Ollivier. Information-Geometric Optimization algorithms: A unifying picture via invariance principles. *arXiv:1106.3708v1*, 2011.

[6] S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithm. In *Proceedings of the 12th International Conference on Machine Learning*, pages 38–46, 1995.

[7] P.-T. D. Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, (134):19–67, 2005.

[8] V. S. Borkar. Stochastic approximation: a dynamical systems viewpoint. Cambridge University Press, 2008.

[9] P. Dayan and G. E. Hinton. Using expectation-maximization for reinforcement learning. *Neural Computation*, 9(2):271–278, 1997.

[10] M. Dorigo, V. Maniezzo, and A. Colorni. The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics - part B*, 26(1):1–13, 1996.

[11] N. Hansen, S. D. Muller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[12] G. E. Hinton. Connectionist learning procedures. *Artificial Intelligence*, 40(1-3):185–234, 1989.

[13] S. Kullback. *Information theory and statistics*. Dover Publications Inc., Mineola, NY, 1997. Reprint of the second (1968) edition.

[14] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*. Springer Verlag, 2nd edition, 2003.

[15] P. Larrañaga and J. A. Lozano. *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publishers, 2002.

[16] L. Malagò, M. Matteucci, and G. Pistone. Towards the geometry of estimation of distribution algorithms based on the exponential family. In H.-G. Beyer and W. B. Langdon, editors, *FOGA '11: Proceedings of the 11th workshop proceedings on Foundations of genetic algorithms*, pages 230–242. ACM, 2011.

[17] F. Sehnke, C. Osendorfer, T. Rückstieß, A. Graves, J. Peters, and J. Schmidhuber. Parameter-exploring policy gradients. *Neural Networks*, 23(4):551–559, 2010.