Derivative Free Optimization

Optimization and AMS Masters - University Paris Saclay

Information Geometric Optimization

Anne Auger
anne.auger@inria.fr
https://www.lri.fr/~auger/teaching.html

1 Information Geometric Optimization for the family of Bernoulli measure

We consider the problem of optimization on the discrete space $\Omega = \{0,1\}^d$. We will derive the IGO algorithm for optimizing $f : \Omega \to \mathbb{R}$ associated to the family of Bernoulli measure $P_\theta(x) = p_{\theta_1}(x_1) \times \ldots \times p_{\theta_d}(x_d)$ on Ω with $p_{\theta_i}(x_i) = \theta_i^{x_i}(1-\theta_i)^{1-x_i}$ and $\theta_i \in [0,1]$, that is x_i is equal 1 with probability θ_i and x_i equals 0 with probability $1-\theta_i$.

1. Compute the Fisher information matrix associated to P_θ.
2. Compute the inverse of the Fisher information matrix.
3. Compute $\tilde{\nabla} \ln P_\theta$
4. Show that the IGO algorithm for the family of Bernoulli measures write for all $i = 1, \ldots d$

$$\theta_i^{t+\delta t} = (1 - \delta t \bar{w})\theta_i^t + \delta t \sum_{j=1}^{N} \hat{w}_j [X_j]_i$$

   where $X_j \sim P_\theta$, $[x]_i$ denotes the coordinate i of the vector x and $\hat{w_j} = \frac{1}{N} w\left(\frac{\mathrm{rk}(X_i)+1/2}{N}\right)$.

5. Show that the previous equation can be rewritten as

$$\theta_i^{t+\delta t} = (1-c)\theta_i^t + c \sum_{j=1}^{N} \tilde{w_j} [X_j]_i$$

   where $\tilde{w_j}$ are weights summing to 1 and c is a so-called learning rate.

This algorithm corresponds to the so-called PBIL algorithm (see slides during the class).