

ONLINE VARIANTS OF THE CROSS-ENTROPY METHOD

ISTVÁN SZITA AND ANDRÁS LŐRINCZ

ABSTRACT. The cross-entropy method [2] is a simple but efficient method for global optimization. In this paper we provide two online variants of the basic CEM, together with a proof of convergence.

1. INTRODUCTION

It is well known that the cross entropy method (CEM) has [2] similarities to many other selection based methods, such as genetic algorithms, estimation-f-distribution algorithms, ant colony optimization, and maximum likelihood parameter estimation. In this paper we provide two online variants of the basic CEM. The online variants reveal similarities to several other optimization methods like stochastic gradient or simulated annealing. However, it is not our aim to analyze the similarities and differences between these methods, nor to argue that one method is superior to the other. Here we provide asymptotic convergence results for the new CE variants, which are online.

2. THE ALGORITHMS

2.1. The basic CE method. The cross-entropy method is shown in Figure 1. For an explanation of the algorithm and its derivation, see e.g. [2]. Extensions of the method allow various generalizations, e.g., decreasing α , varying population size, added noise etc. In this paper we restrict our attention to the basic algorithm.

2.2. CEM for the combinatorial optimization task. Consider the following problem:

The combinatorial optimization task. Let $n \in \mathbb{N}$, $D = \{0, 1\}^n$ and $f : D \rightarrow \mathbb{R}$. Find a vector $\mathbf{x}^* \in D$ such that $\mathbf{x}^* = \arg \min_{\mathbf{x} \in D} f(\mathbf{x})$.

To apply the CE method to this problem, let the distribution g be the product of n independent Bernoulli distributions with parameter vector $\mathbf{p}_t \in [0, 1]^n$ and set the initial parameter vector to $\mathbf{p}_0 = (1/2, \dots, 1/2)$. For Bernoulli distributions, the parameter update is done by the following simple procedure:

2.3. Online CEM. The algorithm performs batch updates, the sampling distribution is updated once after drawing and evaluating N samples. We shall transform this algorithm into an online one. Batch processing is used in two steps of the algorithm:

- in the update of the distribution g_t , and

Department of Information Systems
 Faculty of Infomatics
 Eötvös Loránd University
 Pázmány Péter sétány 1/C
 Budapest, Hungary, H-1117
 Emails: szityu@gmail.com, andras.lorincz@elte.hu.

```

% inputs:
% population size N
% selection ratio  $\rho$ 
% smoothing factor  $\alpha$ 
% number of iterations T



$\mathbf{p}_0$  := initial distribution parameters


for t from 0 to T - 1,
  % draw N samples and evaluate them
  for i from 1 to N,
    draw  $\mathbf{x}^{(i)}$  from distribution  $g(\mathbf{p}_t)$ 
     $f_i := f(\mathbf{x}^{(i)})$ 
  sort  $\{(\mathbf{x}^{(i)}, f_i)\}$  in descending order w.r.t.  $f_i$ 
  % compute new elite threshold level
   $\gamma_{t+1} := f_{[\rho \cdot N]}$ 
  % get elite samples
   $E_{t+1} := \{\mathbf{x}^{(i)} \mid f_i \geq \gamma_{t+1}\}$ 
   $\mathbf{p}' := \text{CEBatchUpdate}(E_{t+1}, \mathbf{p}_t, \alpha)$ 
end loop

```

FIGURE 1. The basic cross-entropy method.

```

procedure  $\mathbf{p}_{t+1} := \text{CEBatchUpdate}(E, \mathbf{p}_t, \alpha)$ 
% E: set of elite samples
%  $\mathbf{p}_t$ : current parameter vector
%  $\alpha$ : smoothing factor

 $N_b = \lceil \rho \cdot N \rceil$ 
 $\mathbf{p}' := (\sum_{\mathbf{x} \in E} \mathbf{x}) / N_b$ 
 $\mathbf{p}_{t+1} := (1 - \alpha) \cdot \mathbf{p}_t + \alpha \cdot \mathbf{p}'$ 

```

FIGURE 2. The batch cross-entropy update for Bernoulli parameters.

- when the elite threshold is computed (which includes the sorting of the N samples of the last episode).

As a first step, note that the contribution of a single sample in the distribution update is $\alpha_1 := \alpha / \lceil \rho \cdot N \rceil$, if the sample is contained in the elite set and zero otherwise. We can perform this update immediately after generating the sample, provided that we know whether it is an elite sample or not. To decide this, we have to wait until the end of the episode. However, with a small modification we can get an answer immediately: we can check whether the new sample is among the best ρ -percentile of the *last N samples*. This corresponds to a sliding window of length N . Algorithmically, we can implement this as a queue Q with at most N elements. The algorithm is summarized in Figure 3.

```

% inputs:
% window size N
% selection ratio  $\rho$ 
% smoothing factor  $\alpha$ 
% number of samples K



$\mathbf{p}_0$  := initial distribution parameters


```

```

Q := {}
for t from 0 to K - 1,
  % draw one samples and evaluate it
  draw  $\mathbf{x}^{(t)}$  from distribution  $g(p_t)$ 
   $f_t := f(\mathbf{x}^{(t)})$ 
  % add sample to queue
  Q := Q  $\cup \{(t, \mathbf{x}^{(t)}, f_t)\}$ 
  if LengthOf(Q) > N, % no updates until we have collected N samples
    delete oldest element of Q
    % compute new elite threshold level
     $\{f'_t\} := \text{sort } f\text{-values in } Q \text{ in descending order}$ 
     $\gamma_{t+1} := f'_{[\rho \cdot N]}$ 
    if  $f(\mathbf{x}^{(t)}) \geq \gamma_{t+1}$  then
      %  $\mathbf{x}^{(t)}$  is an elite sample
       $\mathbf{p}_{t+1} := \text{CEOnlineUpdate}(\mathbf{x}^{(t)}, \mathbf{p}_t, \alpha / [\rho \cdot N])$ 
    endif
  endif
end loop

```

FIGURE 3. Online cross-entropy method, first variant.

For Bernoulli distributions, the parameter update is done by the simple procedure shown in Fig 4.

```

procedure  $\mathbf{p}_{t+1} := \text{CEOnlineUpdate}(\mathbf{x}, \mathbf{p}_t, \alpha_1)$ 
%  $\mathbf{x}$ : elite sample
%  $\mathbf{p}_t$ : current parameter vector
%  $\alpha_1$ : stepsize

 $\mathbf{p}_{t+1} := (1 - \alpha_1) \cdot \mathbf{p}_t + \alpha_1 \cdot \mathbf{x}$ 

```

FIGURE 4. The online cross-entropy update for Bernoulli parameters.

Note that the behavior of this modified algorithm is slightly different from the batch version, as the following example highlights: suppose that the population size is $N = 100$, and we have just drawn the 114th sample. In the batch version, we will check whether this sample belongs to the elite of the set $\{\mathbf{x}^{(101)}, \dots, \mathbf{x}^{(200)}\}$ (after all of these samples

are known), while in the online version, it is checked against the set $\{\mathbf{x}^{(14)}, \dots, \mathbf{x}^{(114)}\}$ (which is known immediately).

2.4. Online CEM, memoryless version. The sliding window online CEM algorithm (Fig. 3) is fully incremental in the sense that each sample is processed immediately, and the per-sample processing time does not increase with increasing t . However, processing time (and required memory) does depend on the size of the sliding window N : in order to determine the elite threshold level γ_t , we have to store the last N samples and sort them.¹ In some applications (for example, when a connectionist implementation is sought for), this requirement is not desirable. We shall simplify the algorithm further, so that both memory requirement and processing time is constant. This simplification will come at a cost: the performance of the new variant will depend on the range and distribution of the sample values.

Consider now the sample at position $N_e = \lceil \rho \cdot N \rceil$, the value of which determines the threshold. The key observation is that its position cannot change arbitrarily in a single step. First of all, there is a small chance that it will be removed from the queue as the oldest sample. Neglecting this small-probability event, the position of the threshold sample can either jump up or down one place or remain unchanged. More precisely, there are four possible cases, depending on (1) whether the new sample belongs to the elite and (2) whether the sample that just drops out of the queue belonged to the elite

- (A) both the new sample and the dropout sample are elite. The threshold position remains unchanged. So does the threshold level except with a small probability when the new or the dropout sample were exactly at the boundary. We will ignore this small-probability event.
- (B) the new sample is elite but the dropout sample is not. The threshold level increases to $\gamma_{t+1} := \gamma_t + f_{N_{e+1}} - f_{N_e}$ (ignoring a low-probability event)
- (C) neither the new sample nor the dropout sample are elite. The threshold remains unchanged (with high probability).
- (D) the new sample is not elite but the dropout sample is. The threshold level decreases to $\gamma_{t+1} := \gamma_t + f_{N_{e-1}} - f_{N_e}$.

Let \mathcal{F}_t denote the σ -algebra generated by knowing all random outcomes up to time step t . Assuming that the positions of the new sample and the dropout sample are distributed uniformly, we get that

$$\begin{aligned} & E(\gamma_{t+1} \mid \mathcal{F}_t, \text{new sample is elite}) \\ &= \gamma_t + \Pr(\text{case A}) \cdot E(f_{N_{e+1}} - f_{N_e} \mid \mathcal{F}_t) + \Pr(\text{case B}) \cdot 0 \\ &\approx \gamma_t + (1 - \rho) \cdot E(f_{N_{e+1}} - f_{N_e} \mid \mathcal{F}_t) \\ &= \gamma_t + (1 - \rho) \cdot \Delta_t, \end{aligned}$$

¹Processing time can be reduced to $O(\log N)$ if insertion sort is used: in each step, there is only one new element to be inserted into the sorted queue.

where we introduced the notation $\Delta_t = E(f_{N_{e+1}} - f_{N_e} \mid \mathcal{F}_t)$. Similarly,

$$\begin{aligned} & E(\gamma_{t+1} \mid \mathcal{F}_t, \text{new sample is not elite}) \\ &= \gamma_t + \Pr(\text{case C}) \cdot 0 + \Pr(\text{case D}) \cdot E(f_{N_{e-1}} - f_{N_e} \mid \mathcal{F}_t) \\ &\approx \gamma_t + \rho \cdot E(f_{N_{e-1}} - f_{N_e} \mid \mathcal{F}_t) \\ &\approx \gamma_t - \rho \cdot \Delta_t, \end{aligned}$$

using the approximation that $E(f_{N_{e-1}} - f_{N_e} \mid \mathcal{F}_t) \approx -E(f_{N_{e+1}} - f_{N_e} \mid \mathcal{F}_t) = \Delta_t$.

Δ_t can drift as t grows, and its exact value cannot be computed without storing the f -values. Therefore, we have to use some approximation. We present three possibilities:

- (1) use a constant stepsize Δ . Clearly, this approximation works best if the distribution of f -value differences does not change much during the optimization process.
- (2) assume that function values are distributed uniformly over an interval $[a, b]$. In this case, $\Delta_t = (b - a)/(N + 1)$. On the other hand, let $D_t = E(|f(\mathbf{x}^{(t)}) - f(\mathbf{x}^{(t+1)})|)$. $f(\mathbf{x}^{(t)})$ and $f(\mathbf{x}^{(t+1)})$ are independent, uniformly distributed samples, so we obtain $D_t = (b - a)/3$, i.e., $\Delta_t = \Delta_0^{\text{uniform}} D_t$ with $\Delta_0^{\text{uniform}} = \frac{3}{N+1}$. From this, we can obtain an online approximation scheme

$$\Delta_{t+1} := (1 - \beta)\Delta_t + \beta \cdot \Delta_0^{\text{uniform}} |f(\mathbf{x}^{(t)}) - f(\mathbf{x}^{(t+1)})|,$$

where β is an exponential forgetting parameter.

- (3) assume that function values have a normal distribution $\sim N(\mu, \sigma^2)$. In this case, $\Delta_t = \sigma(\Phi^{-1}(1 - \rho + \frac{1}{N}) - \Phi^{-1}(1 - \rho))$, where Φ is the Gaussian error function. On the other hand, let $D_t = E(|f(\mathbf{x}^{(t)}) - f(\mathbf{x}^{(t+1)})|)$. $f(\mathbf{x}^{(t)})$ and $f(\mathbf{x}^{(t+1)})$ are independent, normally distributed samples, so we obtain $D_t = \frac{\sigma}{2\sqrt{\pi}}$, i.e., $\Delta_t = \Delta_0^{\text{Gauss}} D_t$ with $\Delta_0^{\text{Gauss}} = 2\sqrt{\pi}(\Phi^{-1}(1 - \rho + \frac{1}{N}) - \Phi^{-1}(1 - \rho))$. From this, we can obtain an online approximation scheme

$$\Delta_{t+1} := (1 - \beta)\Delta_t + \beta \cdot \Delta_0^{\text{Gauss}} |f(\mathbf{x}^{(t)}) - f(\mathbf{x}^{(t+1)})|,$$

where β is an exponential forgetting parameter.

- (4) we can obtain a similar approximation for many other distributions f , but the constant Δ_0^f does not necessarily have an easy-to-compute form.

The resulting algorithm using option (1) is summarized in Fig. 5.

3. CONVERGENCE ANALYSIS

In this section we show that despite the various approximations used, the three variants of the CE method possess the same asymptotical convergence properties. Naturally, the actual performance of these algorithms may differ from each other.

3.1. The classical CE method. Firstly, we review the results of Costa et al. [1] on the convergence of the classical CE method.

Theorem 3.1. *If the basic CE method is used for combinatorial optimization with smoothing factor α , $\rho > 0$ and $p_{0,i} \in (0, 1)$ for each $i \in \{1, \dots, m\}$, then \mathbf{p}_t converges to a 0/1 vector with probability 1. The probability that the optimal probability is generated during the process can be made arbitrarily close to 1 if α is sufficiently small.*

```

% inputs:
% window size N
% selection ratio  $\rho$ 
% smoothing factor  $\alpha$ 
% number of samples K



$\mathbf{p}_0$  := initial distribution parameters

 $\gamma_0$  := arbitrary
for t from 0 to K - 1,
    % draw one samples and evaluate it
    draw  $\mathbf{x}^{(t)}$  from distribution  $g(\mathbf{p}_i)$ 
    if  $f(\mathbf{x}^{(t)}) \geq \gamma_t$  then
        %  $x^{(t)}$  is an elite sample
        % compute new elite threshold level
         $\gamma_{t+1} := \gamma_t + (1 - \rho) \cdot \Delta$ 
         $\mathbf{p}_{t+1} := \text{CEOnlineUpdate}(\mathbf{x}^{(t)}, \mathbf{p}_t, \alpha/(\rho \cdot N))$ 
    else
        % compute new elite threshold level
         $\gamma_{t+1} := \gamma_t - \rho \cdot \Delta$ 
    endif
    % optional step: update  $\Delta$ 
    %  $\Delta := (1 - \beta)\Delta + \beta \cdot \Delta_0 |f(\mathbf{x}^{(t)}) - f(\mathbf{x}^{(t-1)})|$ 
end loop

```

FIGURE 5. Online cross-entropy method, memoryless variant.

The statements of the theorem are rather weak, and are not specific to the particular form of the algorithm: basically they state that (1) the algorithm is a “trapped random walk”: the probabilities may change up and down, but eventually they converge to either one of the two absorbing values, 0 or 1; and (2) if the random walk can last for a sufficiently long time, then the optimal solution is sampled with high probability. We shall transfer the proof to the other two algorithms below.

3.2. The online CE methods.

Theorem 3.2. *If either variant of the online CE method is used for combinatorial optimization with smoothing factor α , $\rho > 0$ and $p_{0,i} \in (0, 1)$ for each $i \in \{1, \dots, n\}$, then \mathbf{p}_t converges to a 0/1 vector with probability 1. The probability that the optimal probability is generated during the process can be made arbitrarily close to 1 if α is sufficiently small.*

Proof. The proof follows closely the proof of Theorems 1-3 in [1]. We begin with introducing several notations. Let \mathbf{x}^* denote the optimum solution, let \mathcal{F}_t denote the σ -algebra generated by knowing all random outcomes up to time step t . Let $\phi_t := \Pr(\mathbf{x} = \mathbf{x}^* \mid \mathcal{F}_{t-1})$ the probability that the optimal solution is generated at time t and $\phi_{t,i} := \Pr(x_i = x_i^* \mid \mathcal{F}_{t-1})$ the probability that component i is identical to

that of the optimal solution. Clearly, $\phi_{t,i} = p_{t-1,i} \mathbf{1}\{x_i^* = 1\} + (1 - p_{t-1,i}) \mathbf{1}\{x_i^* = 0\}$ and $\phi_t = \prod_{i=1}^n \phi_{t,i}$.

Let $p_{t,i}^{\min}$ and $p_{t,i}^{\max}$ denote the minimum and maximum possible value of $p_{t,i}$, respectively. In each step of the algorithms, $p_{t,i}$ is either left unchanged or modified with stepsize $\alpha_1 := \alpha/N_e$. Consequently,

$$p_{t,i}^{\min} = p_{0,i}(1 - \alpha_1)^t$$

and

$$\begin{aligned} p_{t,i}^{\max} &= p_{0,i}(1 - \alpha_1)^t + \sum_{j=1}^t \alpha_1(1 - \alpha_1)^{t-j} \\ &= p_{0,i}(1 - \alpha_1)^t + \sum_{j=1}^t (1 - (1 - \alpha_1))(1 - \alpha_1)^{t-j} \\ &= p_{0,i}(1 - \alpha_1)^t + 1 - (1 - \alpha_1)^t. \end{aligned}$$

Using these quantities,

$$\begin{aligned} \phi_{t,i}^{\min} &= p_{t-1,i}^{\min} \mathbf{1}\{x_i^* = 1\} + (1 - p_{t-1,i}^{\max}) \mathbf{1}\{x_i^* = 0\} \\ &= (1 - \alpha_1)^t (p_{0,i} \mathbf{1}\{x_i^* = 1\} + (1 - p_{0,i}) \mathbf{1}\{x_i^* = 0\}) \\ &= \phi_{1,i}(1 - \alpha_1)^t, \\ \phi_t^{\min} &= \prod_{i=1}^n \phi_{t,i}^{\min} = \phi_1(1 - \alpha_1)^{nt}. \end{aligned}$$

Let $E_t = \cap_{m=1}^t \{\mathbf{x}^{(m)} \neq \mathbf{x}^*\}$ denote the event that the optimal solution was not generated up to time t . Let \mathcal{R}_t denote the set of possible values of ϕ_t . Clearly, for all $r \in \mathcal{R}_t$, $r \geq \phi_t^{\min}$. Note also that $\Pr(\mathbf{x}^{(t)} = \mathbf{x}^* \mid \phi_t, E_{t-1}) = r$ by the construction of the random sampling procedure of CE. Then

$$\begin{aligned} \Pr(\mathbf{x}^{(t)} = \mathbf{x}^* \mid E_{t-1}) &= \sum_{r \in \mathcal{R}_t} \Pr(\mathbf{x}^{(t)} = \mathbf{x}^* \mid \phi_t, E_{t-1}) \Pr(\phi_t = r \mid E_{t-1}) \\ &= \sum_{r \in \mathcal{R}_t} r \Pr(\phi_t = r \mid E_{t-1}) \\ &\geq \phi_t^{\min} = \phi_1(1 - \alpha_1)^{nt}. \end{aligned}$$

Using this, we can estimate the probability that the optimum solution has not been generated up to time step T :

$$\begin{aligned} \Pr(E_T) &= \Pr(E_1) \prod_{t=2}^T \Pr(E_t \mid E_{t-1}) \\ &= \Pr(E_1) \prod_{t=2}^T (1 - \Pr(\mathbf{x}^{(t)} = \mathbf{x}^* \mid E_{t-1})) \\ &\leq \Pr(E_1) \prod_{t=2}^T (1 - \phi_1(1 - \alpha_1)^{nt}). \end{aligned}$$

Using the fact that $(1 - u) \leq e^{-u}$, we obtain

$$\begin{aligned} \Pr(E_T) &\leq \Pr(E_1) \prod_{t=2}^T \exp(-\phi_1(1 - \alpha_1)^{nt}) \\ &= \Pr(E_1) \exp\left(-\phi_1 \sum_{t=1}^T (1 - \alpha_1)^{nt}\right). \end{aligned}$$

Let

$$h(\alpha_1) := \sum_{t=1}^{\infty} (1 - \alpha_1)^{nt} = \frac{1}{1 - (1 - \alpha_1)^n} - 1.$$

With this notation,

$$\lim_{T \rightarrow \infty} \Pr(E_T) \leq \Pr(E_1) \exp(-\phi_1 h(\alpha_1)).$$

However, $h(\alpha_1) \rightarrow 0$ as $\alpha_1 \rightarrow 0$, so $\lim_{T \rightarrow \infty} \Pr(E_T)$ can be made arbitrarily close to zero, if α_1 is sufficiently small.

To prove the second part of the theorem, define $Z_{t,i} = p_{t,i} - p_{t-1,i}$. For the sake of notational convenience, we fix a component i and omit it from the indices. Note that $Z_t \neq 0$ if and only if $\mathbf{x}^{(t)}$ is considered an elite sample. Clearly, if $\mathbf{x}^{(t)}$ is not elite, then no probability update is made. On the other hand, an update modifies p_t towards either 0 or 1. Since $0 < p_t < 1$ with no equality allowed, this update will change the probabilities indeed. Consider the subset of time indices when probabilities are updated, $I = \{t : Z_t \neq 0\}$. We need to show that $|I| = \infty$. This is the only part of the proof where there is a slight extra work compared to the proof of the batch variant.

We will show that each unbroken sequence of zeros in $\{Z_t\}$ is finite with probability 1. Consider such a 0-sequence that starts at time t_1 , and suppose that it is infinite. Then, the sampling distribution p_t is unchanged for $t \geq t_1$, and so is the distribution F of the f -values. Let us examine the first online variant of the CEM. Divide the interval $[t_1, \infty)$ to $N + 1$ -step long epochs. The contents of the queue at time step $t_1, t_1 + (N + 1), t_1 + 2(N + 1), \dots$ are independent and identically distributed, because (a) the samples are generated independently from each other and (b) the different queues have no common elements. For a given queue Q_t (with all elements sampled from distribution F) and a new sample $\mathbf{x}^{(t)}$ (also from distribution F), the probability that $\mathbf{x}^{(t)}$ is not elite is exactly $1 - \rho$. Therefore the probability that no sample is considered elite for $t \geq t_1$ is at most $\lim_{k \rightarrow \infty} (1 - \rho)^k = 0$.

The situation is even simpler for the memoryless variant of the online CEM: suppose again that no sample is considered elite for $t \geq t_1$, and all samples are drawn from the distribution F . F is a distribution over a finite domain, so it has a finite minimum f^{\min} . As all samples are considered non-elite, the elite threshold is decreased by a constant amount $\rho\Delta$ in each step, eventually becoming smaller than f^{\min} , which results in a contradiction.

So, for both online methods we can consider the (infinitely long) subsequences $\{Z_t\}_{t \in I}, \{\mathbf{x}^{(t)}\}_{t \in I}, \{p_t\}_{t \in I}$ etc. For the sake of notational simplicity, we shall index these subsequences with $t = 1, 2, 3, \dots$

From now on, the proof continues identically to the original. We will show that Z_t changes signs for a finite number of times with probability 1. To this end, let τ_k be the random iteration number when Z_t changes sign for the k th time. For all k ,

- (1) $\tau_k = \infty \Rightarrow \tau_{k+1} = \infty$,
- (2) $Z_{\tau_k} < 0 \Rightarrow p_{\tau_k} = (1 - \alpha_1)p_{\tau_{k-1}} + \alpha_1 \cdot 0 \leq (1 - \alpha_1) < 1$,
- (3) $Z_{\tau_k} > 0 \Rightarrow p_{\tau_k} = (1 - \alpha_1)p_{\tau_{k-1}} + \alpha_1 \cdot 1 \geq \alpha_1 > 0$.

From this point on, the proof of Theorem 3 in [1] can be applied without change, showing that the number of sign changes is finite with probability 1, then proving that this implies convergence to either 0 or 1.

□

REFERENCES

- [1] Andre Costa, Owen D. Jones, and Dirk P. Kroese. Convergence properties of the cross-entropy method for discrete optimization. *Operations Research Letters*, 2007. To appear.
- [2] Reuven Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 1:127–190, 1999.