Operations Research
and Complex Systems
Department of Electrical Engineering
Universidade Federal de Minas Gerais

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Towards Statistical Convergence Criteria for Mutation-Based Evolutionary Algorithms

Prof. Felipe Campelo, Ph.D.

Dept. Electrical Engineering, Universidade Federal de Minas Gerais

Curitiba, Brazil - October 2015

Motivation

A simple question

When should my algorithm stop?

Motivation

Evolutionary Algorithms

Evolutionary Algorithms in a nutshell:

initial population + iterate(variation + selection);

Choice of variation operators can often be justified (widely studied);

Stop conditions still chosen in an essentially arbitrary fashion;

Algorithm 1: Basic $(1+1)$-ES

begin
    $k \leftarrow 0;$
    $\mathbf{x}_k \leftarrow$ determine initial point;
    $\psi_k \leftarrow \psi(\mathbf{x}_k);$
    while *stop_criterion() == FALSE* **do**
        $\tilde{\mathbf{x}} \leftarrow \mathbf{x}_k + \mathcal{N}(0, \sigma \mathbf{I}_n);$
        $\tilde{\psi} \leftarrow \psi(\tilde{\mathbf{x}});$
        if $\tilde{\psi} > \psi_k$ **then**
            $\mathbf{x}_{k+1} \leftarrow \tilde{\mathbf{x}};$
            $\psi_{k+1} \leftarrow \tilde{\psi};$
        else
            $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k;$
            $\psi_{k+1} \leftarrow \psi_k;$
        $k \leftarrow k + 1;$

Motivation

Usual stop criteria

- Maximum number of iterations;
- Maximum number of function evaluations;
- Stabilization of best function value;
- Lack of variance in function values;
- Loss of diversity;

Motivation

Usual stop criteria

- Maximum number of iterations;
- Maximum number of function evaluations;
- Stabilization of best function value;
- Lack of variance in function values;
- Loss of diversity;

> *"But if you don't know where you're going,*
> *any road will take you there"*
> – George Harrison

Image: `http://www.magixl.com/cliparts/rock.php`

Motivation

Some notable exceptions

- Trautmann *et al.*[1] and Guerrero *et al.*[2]: online statistical convergence criterion for MOEAs;
  - Statistical tests on the variance of quality indicators and on the regression coefficient of the convergence curve;
- Roche *et al.*[3]: statistical inference for designing termination conditions for DE;
  - Maximum distance to current best;
- Wessing *et al.*[4]: statistical summarization of objective values and mutation strength;
  - Preliminary results for multimodal optimization

[1] Trautmann *et al.*, Evol. Comp. 17(4):493-509, 2009.

[2] Guerrero *et al.*, Proc. IEEE CEC 2010, pp. 4314-4321.

[3] Roche *et al.*, Proc. ECAL 2011, pp. 680-687.

[4] Wessing *et al.*, LNCS 8672, pp. 141-150, 2014.

Even most of the best stop criteria currently available still lack a clear definition of what exactly they are trying to achieve;

To derive a good convergence criterion, we need to answer a few questions. A good starting point would be:

How good is good enough?

Motivation

A scientific approach

- How good is good enough?
- What can we safely assume about the behavior of the algorithm when a *good enough* state has been reached?
- How can we detect this behavior?
- Bonus: can we use this insight to *accelerate* convergence, instead of simply detecting it?

Proposed method

Basic summary

Assume that we can determine the probability of a given EA to generate offspring that are better than their parents, as a function of the parent's distance from the optimum;

If the above assumption is verified, then it should be possible to use the *observed* rate of successful offspring as a way to indirectly measure the current distance from the optimum;

Statistical tests could then be used to ascertain, with a quantifiable degree of confidence, whether or not the algorithm has converged to within the desired distance from the optimum.

Proposed method

Statistical stop criteria

Let $\epsilon^* = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^*\|_2 \leq R^*\}$ be a neighborhood of the (locally-) optimal point $\mathbf{x}^*$, considered to be close enough to represent convergence to that point for practical reasons;

If we can determine the maximum probability of a successful offspring for any point contained within ϵ^* (call it p_s^*), then we could use the observed ratio of sucess to derive an upper bound on p_s^* with a predefined level of confidence.

Proposed method

Statistical stop criteria: $(1+1)$-ES

Let M be the random variable that corresponds to the number of points generated before a successful mutation occurs.

Given that the true (unknown) probability of success p_s is kept constant as long as $\mathbf{x}_k$ does not change, we have that M follows a Geometric distribution.

The number of iterations without improvement required to conclude that $p_s \leq p_s^*$ with confidence level $(1-\alpha)$ can be derived from the CI formula as:

$$m^* = \left\lceil \frac{\ln(\alpha/2)}{\ln(1-p_s^*)} \right\rceil$$

Algorithm 1: Basic $(1+1)$-ES

```
begin
    k ← 0;
    x_k ← determine initial point;
    ψ_k ← ψ(x_k);
    while stop_criterion() == FALSE do
        x̃ ← x_k + N(0, σI_n);
        ψ̃ ← ψ(x̃);
        if ψ̃ > ψ_k then
            x_{k+1} ← x̃;
            ψ_{k+1} ← ψ̃;
        else
            x_{k+1} ← x_k;
            ψ_{k+1} ← ψ_k;
        k ← k + 1;
```

Proposed method

Statistical stop criteria: $(1+\lambda)$-ES

For the $(1+\lambda)$ case, let M be the random variable that corresponds to the number of successful points generated in a batch of λ candidates.

In this case, M follows a binomial distribution, and the formulas for calculating the confidence interval on p_s yield a suggested offspring population size:

$$\lambda^* = \left\lceil \max roots \left\{ a\lambda^3 + b\lambda^2 + c\lambda + d \right\} \right\rceil$$

$$a = (2u/z_\alpha)^2$$

$$b = 4u(3u-1)$$

$$c = -z_\alpha^2(8u+1)$$

$$d = 4uz_\alpha^2\left[3u + z_\alpha^2(u+1)\right]$$

Algorithm 2: Basic $(1+\lambda)$-ES

begin
- $k \leftarrow 0$;
- $\mathbf{x}_k \leftarrow$ determine initial point;
- $\psi_k \leftarrow \psi(\mathbf{x}_k)$;
- **while** *stop criterion* $=$ *false* **do**
  - $i \leftarrow 1$;
  - **while** $i \le \lambda$ **do**
    - $\tilde{\mathbf{x}}_i \leftarrow \mathbf{x}_k + \mathcal{N}(0, \sigma\mathbf{I}_n)$;
    - $\tilde{\psi}_i \leftarrow \psi(\tilde{\mathbf{x}}_i)$;
    - $i \leftarrow i+1$;
  - $c \leftarrow \left\{ j \mid \tilde{\psi}_j \ge \tilde{\psi}_i \forall i \ne j \right\}$;
  - **if** $\tilde{\psi}_c > \psi_k$ **then**
    - $\mathbf{x}_{k+1} \leftarrow \tilde{\mathbf{x}}_c$;
    - $\psi_{k+1} \leftarrow \tilde{\psi}_c$;
  - **else**
    - $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k$;
    - $\psi_{k+1} \leftarrow \psi_k$;
  - $k \leftarrow k+1$;

Proposed method

Calculating the probability of success

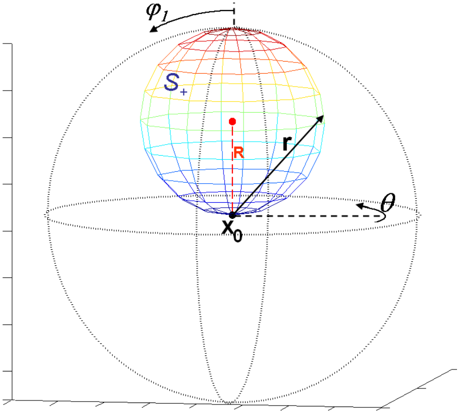
The criteria presented depend on the ability to calculate p_s^*, *i.e.*, the critical value for the probability of success;

Proposed method

Calculating the probability of success

The criteria presented depend on the ability to calculate p_s^*, *i.e.*, the critical value for the probability of success;

In this study, only the most trivial case (spherical function, spherical Gaussian mutation) was derived.



Proposed method

Calculating the probability of success

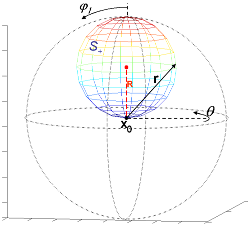
The criteria presented depend on the ability to calculate p_s^*, *i.e.*, the critical value for the probability of success;

In this study, only the most trivial case (spherical function, spherical Gaussian mutation) was derived:

$$p_s(n, \sigma, R) = \Phi(n) \int_{\varphi_1=0}^{\pi/2} \sin^{n-2}(\varphi_1)\, \gamma(n/2, z)\, d\varphi_1$$

with $z = 2\left[\cos(\varphi_1)(R/\sigma)\right]^2$, and:

$$\Phi(n) = \begin{cases} 1/\pi & \text{, if } n = 2 \\ \left[\dfrac{\left(\sqrt{\pi}\right)^{-n}}{2^{(n/2)-1}}\right] \displaystyle\prod_{j=0}^{n-3} \int_0^{\pi} \sin^j(t)\, dt & \text{, if } n \geq 3 \end{cases}$$



Proposed method

Some (sobering) considerations

The formulas obtained for this case are mathematically accurate, but can yield scarily large numbers;

For instance, suppose a $(1+1)$-ES (Gaussian mutation, intensity σ) set to solve a 10-dimensional sphere function. If I want to guarantee convergence to within one σ of the optimum, how many iterations should I use?

`http://drwho.cpdee.ufmg.br:3838/StatStop/`

Proposed method

Some (sobering) considerations

The formulas obtained for this case are mathematically accurate, but can yield scarily large numbers;

For instance, suppose a $(1+1)$-ES (Gaussian mutation, intensity σ) set to solve a 10-dimensional sphere function. If I want to guarantee convergence to within one σ of the optimum, how many iterations should I use?

`http://drwho.cpdee.ufmg.br:3838/StatStop/`

The answer is almost **34,000** iterations *without improvement*!

Proposed method

Some (sobering) considerations

The formulas obtained for this case are mathematically accurate, but can yield scarily large numbers;

For instance, suppose a $(1+1)$-ES (Gaussian mutation, intensity σ) set to solve a 10-dimensional sphere function. If I want to guarantee convergence to within one σ of the optimum, how many iterations should I use?

`http://drwho.cpdee.ufmg.br:3838/StatStop/`

The answer is almost **34,000** iterations *without improvement*!

So what is this approach good for?

Applications

Some tantalizing possibilities: adaptation mechanism

Notice that the probability of success depends only on problem dimension and the ratio R/σ;

Useful as a basis for an evidence-based adaptation mechanism for σ.

For instance:

1. Start with a relatively large σ and use the test to determine convergence to within, say, $R = 3\sigma$ of the optimum (low m^*);
2. Whenever this "convergence" has been reached, make $\sigma \leftarrow \sigma/2$;
3. Repeat until $\hat{R} = 3\sigma \leq R^*$

Applications

Some tantalizing possibilities: generalization

Objective functions are usually not spherical - but most can be modeled as quadratic in the neighborhood of a (local) optimum;

Also, fixed isotropic mutations are not exactly state-of-the-art. Adaptive covariance matrices power some of the best methods around;

While the present work presents formulas only for the trivial case, the derivation of p^* for arbitrary multivariate Gaussians on arbitrary quadratic functions is essentially an implementation problem - mathematical formulas were derived in the 1960s.

Combined with the possibility of an adaptation mechanism, this may provide insights and improvements on relevant techniques such as the CMA-ES.

Conclusions

Final considerations

Statistical convergence criteria were presented for mutation-based EAs (relatively general);

A formula for calculating the probability of a successful offspring were derived for spherical functions and isotropic Gaussian mutation (quite specific and of very little practical importance, although theoretically interesting)

These results show some promising possibilities for the near future: generalization and self-adaptation mechanisms;

About this material

Conditions of use and referencing

The slides and source code are available online. You can use and modify them freely, but please cite the originals:

Felipe Campelo, *Towards Statistical Convergence Criteria for Mutation-Based Evolutionary Algorithms*. Online: `http://git.io/vCgNN`, Latin American School on Computational Intelligence, October 2015, Curitiba, Brazil; Creative Commons BY-NC-SA 4.0.

Or, if you learn something useful, please cite the paper:

Felipe Campelo, *Towards Statistical Convergence Criteria for Mutation-Based Evolutionary Algorithms*, Proc. 2nd Latin American School on Computational Intelligence (LA-CCI), Curitiba, Brazil, 2015.