A simple embedding method for the evaluation of fitness in a noisy environment

JUSTE RAIMBAULT[1,2]
[1] UPS CNRS 3611 ISC-PIF
[2] UMR CNRS 8504 Géographie-cités

Abstract

Stochastic outputs of fitness functions challenge evolutionary meta-heuristics for multi-objective optimization. Several techniques to deal with noise have consequently been recently introduced. We describe here a novel strategy based on embedding into a larger objective space with additional criteria to be optimized that take into account the stochasticity.

1   Context

The problem we aim at tackling is related to stochastic models of simulation: optimization problems in a random landscape. Following [Rakshit et al., 2017], the noisy optimization corresponding to a deterministic function contaminated with noise is analog to (i) finding robust (stable, or not sensitive) solutions in a chaotic deterministic landscape; (ii) optimizing an approximated function; (iii) optimizing in a dynamic landscape (all being strictly equivalent under specific assumptions). We add an additional analog problem, of greater interest in the case of simulation models, that is (iv) the optimization of a random variable landscape.

Particular formulations

1. Fitness with additive noise :
$$\vec{f}(\vec{x}) = \vec{f_d}(\vec{x}) + \vec{b}(\vec{x})$$

2. Fitness with multiplicative noise : equivalent to additive noise if $\vec{f} > 0$ (take the log)

3. Non-stationary noise : $\vec{b}(\alpha(\vec{x}))$; for example parameters of the distribution varying in space (compatible with other cases)

4. Self-amplifying noise (or linked to the value of the deterministic part)

5. Random variable with conditional expectancy locally "well-defined" : e.g., local variance small compared to global variance. Equivalent to an additive noise by writing
$$\vec{f}(\vec{x}) = \mathbb{E}\left[\vec{f}\right](\vec{x}) + \left(\vec{f}(\vec{x}) - \mathbb{E}\left[\vec{f}\right](\vec{x})\right)$$

6. "Chaotic" random variable (overall variance of the same magnitude as local variance): we won't consider models where the sensitivity to parameters is negligible compared to the randomness ?

Literature   A comprehensive survey is done by [Rakshit et al., 2017] with more than 200 papers reviewed. The dominant formulation of the problem is the noisy optimization. Four main strategies to deal with noise in EA are unveiled: (i) explicit averaging of evaluated points, with numerous sampling strategies; (ii) estimation of the fitness distribution; (iii) implicit averaging with dynamical population sizing, that extends into more elaborated evolutionary search strategies; (iv) selection strategies.

Positioning The dominant narrative seems to be the noisy $f + b$, which is a bit different from considering the output of a simulation model. We could in theory deal with multimodal distributions in some parts of the parameter space, in which case we would be closer to the strategies (ii) which in a sense follow an objective similar to Approximate Bayesian Computation [Sunnåker et al., 2013] (estimating the statistical distribution over the parameter space). The method should indirectly take this aspect into account.

The strategy we introduce does not act directly on the *structure* of the meta-heuristic, but on its application problem. The broad idea is to solve a more general problem which additional objectives should allow to tackle the stochasticity in the initial space. Our proposal indirectly acts as a (iv) strategy, as we will change the objective space. Such an embedding approach has to the best of our knowledge never been proposed in the literature.

The idea is analog to state augmentation in Bayesian signal processing to deal with non-stationarity [Simon, 2006]. It also bears similarities with determination of the embedding dimension when forecasting chaotic time series, since it diminishes apparent noise by extending the phase space ([Szpiro, 1997] proposes to forecast chaotic time series with a GA).

2   Proposed method

Base idea Embedding approach taking n_{evals} as a supplementary objective in the optimization.

$\rightarrow$ currently $\frac{1}{n_{evals}}$ is taken as objective, as they need to be minimized.

Generalization Embed into a larger space with additional objective function of number of repetitions/convergence levels/other criteria ?

$\rightarrow$ flexibility on criteria should allow to be flexible on either distrib estimation, or median/average with given CI, or . . .

Ideas

- Adaptive in sigma ? size of the confidence interval proportional to $\sigma/\sqrt{n}$

- Use bootstrap to get local confidence intervals ? $\rightarrow$ cf Welsh Confidence Interval based samplings

2.1   Method formalization

The idea of the method is described by [Pumain and Reuillon, 2017]. The base of the method is the standard NSGA2 with a fixed resampling rate r. Each individual has a maximal number of evaluations, but new samples discard the oldest ones to always allow for possible new diversity.

Let $\vec{f}$ the random variable to be optimized. $\vec{f}(\vec{x}_i)^{(k,t)}$ denote a realization obtained at generation t, for the parameter point $\vec{x}_i$ and correspond to the kth evaluation at this point. A realization of the genetic algorithm can be interpreted as sequences i_t and k_t such that $\vec{x}_{i_t}$) are the successive populations and $\vec{f}(\vec{x}_{i_t})^{(k_t,t)}$ the corresponding realizations.

We consider $\vec{g}$ a set of additional objective functions, which at generation t will be a function of the previous realizations $\vec{f}(\vec{x}_{i_{t-1}})^{(k_{t-1},t-1)}$, and we update the population by running a step of the algorithm on the extended objective $(\vec{f}, \vec{g})$.

A simple example is to add the inverse of the number of evaluation : $g(\vec{x}) = \frac{1}{|k_t(\vec{x})|}$. Other alternatives can be adaptive and consider an estimated standard deviation to take as objective a rough estimation of a confidence interval around the mean (in the case of a Gaussian distribution) with $\frac{\hat{\sigma}}{\sqrt{n}}$.

3 Experiments

3.1 Questions to be answered

1. Does our approach "improves" the optimization in some cases ? (possible indicators below : expected run time, relative hypervolume, etc)

2. On which type of problems ? (benchmark functions : to be compared with other state-of-the-art algo ; real-life model)

3. In which particular settings : form of the additional objectives (see above), constant/adaptive re-sampling rates, objective estimation method, . . .

4. Overall advantages of the approach : gain in computation time / tradeoff of convergence ?

3.2 Experience plan

3.2.1 Benchmark problems

Optimization

- Noisy bbob from COCO [Hansen et al., 2016] : 30 single objective noisy functions, with varying dimensions (*Q : varying noise distrib ?*)

- ad-hoc noisy multi-obj ?

- ad-hoc non-stationary noise ?

- Vary the distribution of noise ?

Simulation model Proposition (simple but applied model) : [Raimbault, 2017] ; distribution of output indicators seem not stationary, strange shapes in some places for Moran.
$\rightarrow$ check first behavior of MSE to real points if it has similar behavior.

3.2.2 Benchmark algorithms

$\rightarrow$ find algos to compare to.

- static sampling (assumed very bad)

- state-of-the-art dynamic sampling ?

- fitness estimation

- (selection strategies) $\rightarrow$ *may be more difficult to implement depending on MGO capabilities ?*

3.3 Performance measures

$\rightarrow$ find relevant performance measures

- Expected Run Time

- *convergence* : generational distance, hypervolume, etc.

References

[Hansen et al., 2016] Hansen, N., Tusar, T., Mersmann, O., Auger, A., and Brockhoff, D. (2016). COCO: The Experimental Procedure. *ArXiv e-prints*.

[Pumain and Reuillon, 2017] Pumain, D. and Reuillon, R. (2017). Evaluation of the simpoplocal model. In *Urban Dynamics and Simulation Models*, pages 37–56. Springer.

[Raimbault, 2017] Raimbault, J. (2017). Calibration of a Density-based Model of Urban Morphogenesis. *ArXiv e-prints*.

[Rakshit et al., 2017] Rakshit, P., Konar, A., and Das, S. (2017). Noisy evolutionary optimization algorithms–a comprehensive survey. *Swarm and Evolutionary Computation*, 33:18–45.

[Simon, 2006] Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.

[Sunnåker et al., 2013] Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate bayesian computation. *PLOS Computational Biology*, 9(1):1–10.

[Szpiro, 1997] Szpiro, G. G. (1997). Forecasting chaotic time series with genetic algorithms. *Physical Review E*, 55(3):2557.