

---

# Benchmarking Cross Entropy Exploiting partial Decomposability:

Analysis of a Batch Sampling Method for Large-Scale Structured Optimization

---

**Siamak (Mohsen) Ravanbakhsh, Russell Greiner**

Department of Computing Science, University of Alberta

{ mravanba, rgreiner } @cs.ualberta.ca

While many of the Monte Carlo methods for optimization are sequential sampling methods (*e.g.*, simulated annealing), some important ones are batch sampling techniques, including Cross Entropy (*a.k.a.* CE) [1] and Probability Collectives [2]. These methods each start by defining a distribution over the optimization domain, which is iteratively refined, to improve the chance that an instance from this distribution is likely to be (near) an optimum. CE performs iterative improvements by sampling the current distribution, identifying the "elite" instances (the ones producing the best scores), then producing the maximum likelihood distribution over these instances. At convergence, CE returns the mode of final distribution as an approximation for a global optimum.

These batch sampling methods are highly parallelizable, and also provide sensitivity information about the optimization landscape: Finding a distribution is peaked wrt a variable means a small change to that variable's value (in that region) will have a large impact on the value of the optimization function. This observation inspired our "Cross Entropy Method for Exploiting partial Decomposability", CEED [3,4], which can deal with an optimization function that is the sum of loss sub-functions  $L(x) = \sum_i L_i(\vec{x}^i)$ , each involving a subset of variables  $\vec{X}^i \subseteq \vec{X}$ . (Note a variable may appear in several loss sub-functions.) CEED first uses the CE approach for each  $L_i$  individually: given a current distribution  $D_{t-1}^i$  over its  $\vec{X}^i$  variables, CEED produces a new distribution  $D_t^i$  based on the elite subset of instances drawn from the  $D_{t-1}^i$ . CEED then forms a single distribution  $D_t$  over *all* of the  $\vec{X}$  variables by combining the distributions  $\{D_t^i\}$ : this involves first computing a single distribution over each variable  $X_j$ , by combining the marginals wrt each  $D_t^i$  that involves  $X_j$ . To be effective, this combination process must use the sensitivity of each  $D_t^i$  wrt  $X_j$ ; here we considered Fisher Information and the variance of the estimated max-likelihood parameters for the marginals of  $D_t^i$  over each single variable [3]. While variance of this ML estimate can be readily computed from the estimated parameters for discrete variables, it is not readily available for continuous variables (when dealing with a member of the exponential family) We therefore use an adaptive binning approach to approximate a continuous variable as a categorical one.

This report explores the effect of different characteristics of the problem as well as that of CEED's parameters on the convergence rate, for a difficult high-dimensional multi-modal optimization problem. We evaluate how entanglement of the loss subfunctions  $\{L_i\}$  (corresponding to the variables in common across subfunctions) affects the performance of CEED, and illuminate the situations where CEED is most advantageous. The results demonstrate that incorporating information from the different  $L_i$ 's, rather than only looking at their sum, can increase the convergence rate by many folds. We compare our CEED method against its "father" CE, as well as other methods, including simulated annealing method. These analyses will also help us to further extend CEED, towards developing extensions capable of solving some major challenges, such as protein folding, protein-protein and protein-ligand docking and in general any energy minimization task that involves many losses (energy terms), when most of the terms depend on only a small subset of variables.

## References

- [1] R. Rubinstein and D. Kroese, *The Cross-Entropy Method: a Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning.*, Springer, Jan 2004. [2] David H. Wolpert, *Information Theory The Bridge Connecting Bounded Rational Game Theory and Statistical Physics In Complex Engineering Systems*, edited by Yaneer Bar-Yam and Dan Braha, Perseus books, in print. [3] S. Ravanbakhsh (2009) *A Stochastic Optimization Method for Partially Decomposable Problems, with Application to Analysis of NMR Spectra*. M.Sc. Thesis, U of Alberta, Edmonton, Canada. [4] S. Ravanbakhsh, B. Pezos and R. Greiner (2010), *A Cross-Entropy Method that Optimizes Partially Decomposable Problems*. AAAI, pp 1280-1286