Steady-State Selection and Efficient Covariance Matrix Update in the Multi-objective CMA-ES

Christian Igel[1], Thorsten Suttorp[1], and Nikolaus Hansen[2]

[1] Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany
{christian.igel, thorsten.suttorp}@neuroinformatik.rub.de
[2] Institute of Computational Science, ETH Zurich, 8092 Zurich, Switzerland
nikolaus.hansen@inf.ethz.ch

Abstract. The multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES) combines a mutation operator that adapts its search distribution to the underlying optimization problem with multi-criteria selection. Here, a generational and two steady-state selection schemes for the MO-CMA-ES are compared. Further, a recently proposed method for computationally efficient adaptation of the search distribution is evaluated in the context of the MO-CMA-ES.

1 Introduction

Evolution strategies (ES) for real-valued optimization rely on Gaussian random variations. Appropriately adapting the covariance matrices of these mutations during optimization allows for learning a variable metric for the search distribution. It is well known that such an automatic adaptation of the mutation distribution drastically improves the search performance on non-separable and/or badly scaled single-objective functions [1,2,3,4].

In [5], we incorporated the step size and covariance matrix adaptation from the covariance matrix adaptation ES (CMA-ES, [3]) into a multi-objective framework. The resulting MO-CMA-ES used generational selection based on [6] combined with the sorting criterion proposed in [7,8]. We chose generational selection in order to make our performance comparisons with alternative methods easier to interpret. However, in [7,8] steady-state selection is used with good results and the question arises whether the MO-CMA-ES would profit from this selection scheme. In [9], we presented a new, computationally efficient update scheme for covariance matrices. The complexity reduction from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ per update of the mutation distribution, where n is the dimensionality of the search space, comes at the cost of slower adaptation rates. However, as in the MO-CMA-ES many mutation distributions need to be traced, this approach seems to be particularly promising for the MO-CMA-ES.

In this work, we first investigate the computationally efficient update proposed in [9] within the framework of the MO-CMA-ES. Second, we compare variants of the MO-CMA-ES with different steady-state selection schemes and generational selection, respectively.

S. Obayashi et al. (Eds.): EMO 2007, LNCS 4403, pp. 171–185, 2007.

2 Covariance Matrix Adaptation

Let us consider an additive mutation $\boldsymbol{v}_i^{(g)} \in \mathbb{R}^n$ of individual i in generation g. The mutation $\boldsymbol{v}_i^{(g)}$ is a realization of an n-dimensional random vector distributed according to a zero-mean Gaussian distribution with covariance matrix $\boldsymbol{C}_i^{(g)}$, that is, $\boldsymbol{v}_i^{(g)} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{C}_i^{(g)}\right)$. To sample this mutation distribution, n independent standard normally distributed random numbers are drawn to generate a realization of an n-dimensional normally distributed random vector $\boldsymbol{z}_i^{(g)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ with unit covariance matrix and zero mean. Then this random vector is rotated and scaled by a linear transformation $\boldsymbol{A}_i^{(g)} \in \mathbb{R}^{n\times n}$ such that

$$\boldsymbol{A}_i^{(g)} \boldsymbol{z}_i^{(g)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_i^{(g)}) \text{ for } \boldsymbol{z}_i^{(g)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \ .$$

Thus, for sampling the mutation distribution the covariance matrix $\boldsymbol{C}_i^{(g)}$ has to be decomposed into Cholesky factors $\boldsymbol{C}_i^{(g)} = \boldsymbol{A}_i^{(g)} {\boldsymbol{A}_i^{(g)}}^T$. One of the decisive features of ES is that the covariance matrices are subject to adaptation. The general policy is to alter the covariance matrices such that steps promising larger fitness gain are sampled more often. Here we consider matrix updates of the form $\boldsymbol{C}^{(g+1)} = \alpha \boldsymbol{C}^{(g)} + \beta \boldsymbol{V}^{(g)}$, where $\boldsymbol{V}^{(g)} \in \mathbb{R}^{n\times n}$ is positive definite and $\alpha, \beta \in \mathbb{R}^+$ are weighting factors (e.g., see [3,10]). Let $\boldsymbol{v}^{(g)} \in \mathbb{R}$ be a step in the search space promising large fitness gain. To increase the probability that $\boldsymbol{v}^{(g)}$ is sampled in the next iteration, the rank-one update

$$\boldsymbol{C}^{(g+1)} = \alpha \boldsymbol{C}^{(g)} + \beta \boldsymbol{v}^{(g)} {\boldsymbol{v}^{(g)}}^T \tag{1}$$

can be used. This update rule shifts the mutation distribution towards the line distribution $\mathcal{N}\big(\mathbf{0}, \boldsymbol{v}^{(g)} {\boldsymbol{v}^{(g)}}^T\big)$, which is the distribution with the highest probability to generate $\boldsymbol{v}^{(g)}$ among all normal distributions with zero mean [3].

In general, each factorizing of a covariance matrix requires $O(n^3)$ operations. Thus, in an ES with additive covariance matrix update the Cholesky factorization of the covariance matrix is the computationally dominating factor apart from the fitness function evaluations. In [9] we therefore proposed not to factorize the covariance matrix, but to use an incremental rank-one update rule for the Cholesky factorization. This reduces the computational complexity to $O(n^2)$. The idea is not to compute the covariance matrix explicitly, but to operate on Cholesky factors only. Setting $\boldsymbol{v}^{(g)} = \boldsymbol{A}^{(g)} \boldsymbol{z}^{(g)}$ with $\boldsymbol{z}^{(g)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ we can rewrite the rank-one update of the covariance matrix equation (1) as

$$\boldsymbol{C}^{(g+1)} = \alpha \boldsymbol{C}^{(g)} + \beta \boldsymbol{A}^{(g)} \boldsymbol{z}^{(g)} \left[\boldsymbol{A}^{(g)} \boldsymbol{z}^{(g)}\right]^T \ . \tag{2}$$

Using the following theorem, we turn this update for $\boldsymbol{C}^{(g)}$ into an update for $\boldsymbol{A}^{(g)}$.

Theorem 1 ([9]). *Let $\boldsymbol{C}_t \in \mathbb{R}^{n\times n}$ be a symmetric nonnegative definite matrix with Cholesky factorization $\boldsymbol{C}_t = \boldsymbol{A}_t {\boldsymbol{A}_t}^T$. Assuming that $\boldsymbol{C}_t$ is updated using*

$$\boldsymbol{C}_{t+1} = \alpha \boldsymbol{C}_t + \beta \boldsymbol{v}_t \boldsymbol{v}_t^T ,$$