Advanced Optimization

Master AIC - Paris Saclay University

Exercices - Stochastic Continuous Optimization

Anne Auger and Dimo Brockhoff
anne.auger@inria.fr

I Adaptation of the Covariance Matrix: Rank-one Update

In this first exercice we want to understand the so-called rank-one update mechanism to update the covariance matrix in the CMA-ES algorithm. We consider thus the following algorithm implementing solely the rank-one update (while the full CMA-ES algorithm combines other updates for the covariance matrix and step-size adaptation)

[Objective: minimize $f : \mathbb{R}^n \to \mathbb{R}$]
1. `Initialize` $\mathbf{C}_0 = I_d$, $\mathbf{m}_0 \in \mathbb{R}^n$, $t = 0$
2. `set` $w_1 \geq w_2 \geq \ldots w_\mu \geq 0$ with $\sum w_i = 1$; $\mu_{\text{eff}} = 1/\sum w_i^2$, $0 < c_{\text{cov}} < 1$ (typically $c_{\text{cov}} \approx 2/n^2$)
3. `while not terminate`
4. Sample λ independent candidate solutions:
5. $\mathbf{X}_{t+1}^i = \mathbf{m}_t + \mathbf{y}_{t+1}^i$ for $i = 1 \ldots \lambda$
6. with $(\mathbf{y}_{t+1}^i)_{1 \leq i \leq \lambda}$ i.i.d. following $\mathcal{N}(\mathbf{0}, \mathbf{C}_t)$
7. Evaluate and rank solutions:
8. $f(\mathbf{X}_{t+1}^{1:\lambda}) \leq \ldots \leq f(\mathbf{X}_{t+1}^{\lambda:\lambda})$
9. Update the mean vector:
10. $$\mathbf{m}_{t+1} = \mathbf{m}_t + \underbrace{\sum_{i=1}^{\mu} w_i \mathbf{y}_{t+1}^{i:\lambda}}_{\mathbf{y}_{t+1}^w}$$
11. Update the covariance matrix using the rank-one update:
12. $\mathbf{C}_{t+1} = (1 - c_{\text{cov}})\mathbf{C}_t + c_{\text{cov}} \mu_{\text{eff}} \mathbf{y}_{t+1}^w (\mathbf{y}_{t+1}^w)^T$
13. t=t+1

1. Why is the update in line 12 called the rank-one update?
2. Plot the lines of equal density of the initial sampling distribution $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$ (with $\mathbf{C}_0$ being equal to the identity)

In order to understand geometrically the effect of adding the matrix $c_{\text{cov}} \mu_{\text{eff}} \mathbf{y}_{t+1}^w (\mathbf{y}_{t+1}^w)^T$ to the matrix $(1 - c_{\text{cov}})\mathbf{C}_t$ (line 12 of the algorithm), we consider $t = 0$ and want to plot the lines of equal density associated to the multivariate normal distribution with mean vector $\mathbf{m}_1$ and covariance matrix $\mathbf{C}_1$. In order to simplify we assume that $\mu_{\text{eff}} = 1$.

3. Compute the eigenvalues of the matrix $A = c_{\text{cov}} \mathbf{y}_1^w (\mathbf{y}_1^w)^T$. **Hint:** you can in particular show that the matrix has a rank of 1, deduce how many non-zero eigenvalues the matrix has. You can also show that $\mathbf{y}_1^w$ is an eigenvector of the matrix and compute its associated eigenvalue.

4. We remind that for a symmetric matrix A of $\mathbb{R}^n$ we have $\mathbb{R}^n = \mathrm{Ker}(A) \overset{\perp}{\bigoplus} \mathrm{Im}(A)$. Show that there exists an orthogonal basis of normalized eigenvectors of A of the form $(\mathbf{y}_1^w/\|\mathbf{y}_1^w\|, u_2, \dots, u_n)$.

4. Show that the basis $(\mathbf{y}_1^w/\|\mathbf{y}_1^w\|, u_2, \dots, u_n)$ is also a basis composed of eigenvectors of the matrix $\mathbf{C}_1 = (1 - c_{\mathrm{cov}})I_d + c_{\mathrm{cov}}\mathbf{y}_1^w(\mathbf{y}_1^w)^T$. Compute the associated eigenvalues.

5. Assume $n = 2$, using the previous question plot the lines of equal density of $\mathcal{N}(\mathbf{m}_1, \mathbf{C}_1)$.

6. Deduce that the rank-one update increases the probability of successful steps[1] to appear again.

[1]The terminology "step" refers to what is added to the mean to create a new solution. For instance in Line 5. of the algorithm, the first sampled solution equals $\mathbf{X}_{t+1}^1 = \mathbf{m}_t + \mathbf{y}_{t+1}^1$. We call $\mathbf{y}_{t+1}^1$ the step that created the solution $\mathbf{X}_{t+1}^1$.