Algoritmo CMA-σ-SA-ES
Giovanni Gamaliel López Padilla

1. ¿Como se ajusta el tamaño del paso?

El tamaño de cada paso z se ajusta de la siguiente manera:

1. Mediante el uso de un sigma promedio ($\langle\sigma\rangle$) se perturba mediante una variable aleatoria de tal manera que el desplazamiento de cada hijo esta definido en la ecuación 1.
$$\sigma_h = \langle\sigma\rangle e^{\tau N(0,1)} \tag{1}$$
donde τ esta definido para cada dimensión. El uso de la distribución normal es debido a que nos proporcionara valores isotropicos, lo cual nos favorecera en los puntos de busqueda.[1]

2. A partir de la factorización de Cholesky de la matriz de Covarianza se obtiene la matriz triangular superior que efectuara una transformación a un vector aleatorio (ecuación 2). Cada componente del vector aleatorio esta basado en una distribución normal.
$$S_h = \text{Cholesky}(C)N(0,1) \tag{2}$$
El término S_h contiene la dirección acerca de donde se tiene que ir dirigiendo la solución. La razón de esto se encuentra en la matriz de covarianza, ya que mientras más cercano esten los hijos al mínimo global el espacio de exploración será menor.

3. Al realizar el producto de σ_h y S_h (ecuación 3), obtendremos el desplazamiento y dirección del hijo con respecto el centro de exploración (elipsoide).
$$Z_h = \sigma_h S_h \tag{3}$$

4. Al tener λ número de hijos escogeremos los μ hijos tales que la diferencia con el mínimo global sea menor. El centro del espacio de exploración será desplazado siguiendo la ecuación 4:
$$Y = Y + \langle z\rangle \tag{4}$$
donde $\langle z\rangle$ es el promedio de z_μ de los μ hijos.

5. Los cuales proporcionaran sus S_h para modificar la matriz de covariancia de modo que en cada iteracción se aproximen a la solución (ecuacion 5).
$$C = \left(1 - \frac{1}{\tau_e}\right) C + \frac{1}{\tau_e}\left\langle ss^T\right\rangle \tag{5}$$
donde
$$\left\langle ss^T\right\rangle = \frac{1}{\mu}\sum_{i=1}^{\mu} s_i s_i^T$$

$s_i s_i^T$ es una matriz de tamaño nxn donde n es la dimensión de la función donde se encontrará el mínimo local. Analizando más a la suma de matrices $s_i s_i^T$ se encuentra que son simetricas ya que cada una se ellas lo es y su suma preserva esta caracteristica. Ejemplo:
Sea:

$$s = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

entonces:

$$\begin{aligned} ss^T &= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \end{aligned}$$

y como la matriz de covariancia también es simetrica, entonces la combinación lineal de C y $\langle ss^T \rangle$ también es simetrica.

2. ¿Cómo se orienta la matriz de covarinza?

El comportamiento de la elipse esta descrito en la ecuación 4, esto es debido a que la podemos reescribir como la ecuacion 6. [2]

$$Y = m + \sigma N(0,1) \tag{6}$$

donde

$$m = \sum_i^{\mu} w_i y_i$$

donde w_i es conocido como peso, la cual es positiva. Para este caso, w_i es el mismo valor para los μ mejores hijos.

$$w_i = \frac{1}{\mu}$$

Como σ de la ecuación 6 es calculada con la ecuación 2 donde $S_h = \sigma$, entonces se trata de un problema con $(n^2+n)/2$ grados de libertad donde las componentes estas relacionadas.[3]

Por lo que la orientación de la matriz de covariancia es manipulado por la ecuación 5, donde la matriz $\langle ss^T \rangle$ indica como cambiar su orientación ya que copntiene los pesos de cada hijo seleccionado.

3. ¿Como la elipse se alarga/acorta en cada dirección

La definición de la matriz de covarianza es:

$$C = \begin{pmatrix} Var(x_1) & Cov(x_1, x_2) & \dots & Cov(x_1, x_n) \\ Cov(x_2, x_1) & Var(x_2, x_2) & \dots & Cov(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(x_n, x_1) & Cov(x_n, x_2) & \cdots & Var(x_n) \end{pmatrix}$$

donde

$$Cov(x, y) = Cov(y, x)$$

por lo tanto, la matriz de covarianza es simetrica. Para este caso se tiene que los valores $x_1 = X$ y $x_2 = Y$, por lo que la matriz tendria la siguiente forma:

$$C = \begin{pmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{pmatrix}$$

4. ¿Cómo da $\mathrm{Z_h}$ en cierta dirección?

Debido a que $\mathrm{Z_h}$ es calculado en la ecuación 3, anteriormente se comento que las ecuaciones 1 y 2 guardan información acerca del desplazamiento y dirección de la posición de busqueda para los nuevos hijos (ecuación 3). Esto es porque $\langle\sigma\rangle$ se encuentra pesada por los μ hijos y la matriz de covariancia C contiene información acerca de la dirección preferencial para encontrar el mínimo local.

Referencias

[1] A. Auger and N. Hansen. Tutorial cma-es — evolution strategies andcovariance matrix adaptation, 7 2013. pp 18. http://www.cmap.polytechnique.fr/~nikolaus.hansen/gecco2013-CMA-ES-tutorial.pdf.

[2] A. Auger and N. Hansen. Tutorial cma-es — evolution strategies andcovariance matrix adaptation, 7 2013. pp 25. http://www.cmap.polytechnique.fr/~nikolaus.hansen/gecco2013-CMA-ES-tutorial.pdf.

[3] A. Auger and N. Hansen. Tutorial cma-es — evolution strategies andcovariance matrix adaptation, 7 2013. pp 21. http://www.cmap.polytechnique.fr/~nikolaus.hansen/gecco2013-CMA-ES-tutorial.pdf.

[4] N. Hansen. The cma evolution strategy: A tutorial, 2016. https://arxiv.org/abs/1604.00772.