Foundations of Real-Valued Evolutionary Algorithms

Dirk V. Arnold

Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia
Canada

dirk@cs.dal.ca
Optimisation Problems

• optimisation problems are abundant in science and engineering

• mathematically, we want to minimise (or maximise) some function $f : D \rightarrow \mathbb{R}$
Evolutionary Algorithms

- evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution
- they model the interplay of variation and selection in a population of individuals
- fitness is determined by the objective function
Evolutionary Algorithms

- evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution
- they model the interplay of variation and selection in a population of individuals
- fitness is determined by the objective function
Evolutionary Algorithms

- evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution
- they model the interplay of variation and selection in a population of individuals
- fitness is determined by the objective function
Evolutionary Algorithms

• evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution

• they model the interplay of variation and selection in a population of individuals

• fitness is determined by the objective function
Evolutionary Algorithms

- evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution
- they model the interplay of variation and selection in a population of individuals
- fitness is determined by the objective function
Evolutionary Algorithms

- Evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution.
- They model the interplay of variation and selection in a population of individuals.
- Fitness is determined by the objective function.
Evolutionary Algorithms

- evolutionary algorithms (EAs) are optimisation strategies which derive inspiration from Darwinian evolution
- they model the interplay of variation and selection in a population of individuals
- fitness is determined by the objective function
Two-phase Jet Nozzle

- optimisation of a single-component two-phase nozzle
- objective: maximisation of efficiency
- 330 compatible segments
  - $10^{60}$ different configurations
- the efficiency increased from 55% to 80%

Optimisation of Coffee Blends (1)

- most coffees are blends of up to ten different kinds of single-origin coffee
- the quality and availability of the single-origin coffees varies from year to year
- brand name coffees have distinct “target tastes” that need to be matched
- experts judge based on criteria such as aroma, brightness, acidity, and body, and achieve the target taste using their experience
Optimisation of Coffee Blends  (2)

• interactive evolution of coffee blends:
  – replace the experts’ heuristics with random steps
  – experts pick the best among a population of five blends

• after 11 generations, the taste of the blend was indistinguishable from the target

• the composition of the blend was very different from what the experts would have chosen

• different blends can have identical flavour; the expert solution is not necessarily the cheapest one

Real-Valued Optimisation

• in this tutorial, we consider optimisation problems of the form $f : \mathbb{R}^N \rightarrow \mathbb{R}$

• unconstrained numerical optimisation algorithms:
  – quasi-Newton and conjugate gradient methods
  – implicit filtering
  – pattern search
  – stochastic approximation
  – response surface methods
  – evolutionary algorithms
  – …
Real-Valued Evolutionary Algorithms

• evolutionary algorithms
  – are easy to understand and implement
  – do not rely on derivative information and make no assumptions
    with regard to the function being optimised
  – typically employ populations
  – involve some element of randomness
  – proceed based on incomplete information
  – strive to be “adaptive”

• types of real-valued EAs include
  – evolution strategies
  – evolutionary programming
  – differential evolution
Outline

• evolution strategies
  – the $(\mu/\rho \pm \lambda)$-ES
  – performance for the line model
  – performance for the sphere model
  – performance for other problems

• covariance matrix adaptation
  – CMA-ES
  – other approaches
Benefits of “Keeping It Simple”

- derive scaling laws
- compare with optimal behaviour
- highlight differences between strategy variants; reveal strengths and weaknesses
- recommend parameter settings
- develop intuition with regard to working principles of operators
- develop and improve adaptation strategies
Evolution Strategies: The $(\mu/\rho \uparrow \lambda)$-ES (1)

- population size: $|\mathcal{P}| = \mu$

- number of offspring: $|Q| = \lambda$

**selection**: the $\mu$ best candidate solutions in

- $\mathcal{P} \cup Q$ for plus-selection
- $Q$ for comma-selection

survive (“truncation selection”)

**variation**: for every offspring to be generated, randomly choose $\rho$ parents; recombine and mutate
Evolution Strategies: The $(\mu/\rho + \lambda)$-ES (2)

recombination: with $i_1, i_2, \ldots, i_\rho$ the indices of the (randomly chosen) parents of a candidate solution to be generated, let

$$x = \frac{1}{\rho} \sum_{j=1}^{\rho} x_{i_j}$$

mutation: add a normally distributed random vector with mean zero

$$y = x + \sigma z$$
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E}\left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz
\]

\[
= \frac{\sigma}{\sqrt{2\pi}}
\]

\[
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The $(1 + 1)$-ES

- consider maximisation of objective function $f(x) = x$

- $(1 + 1)$-ES:

$$x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}$$

- progress rate:

$$\varphi_{1+1} = E \left[ x^{(t+1)} - x^{(t)} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}$$
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = E\left[x^{(t+1)} - x^{(t)}\right]
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The $(1 + 1)$-ES

- consider maximisation of objective function $f(x) = x$

- $(1 + 1)$-ES:

$$x^{(t+1)} = \begin{cases} 
x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
x^{(t)} & \text{otherwise} 
\end{cases}$$

- progress rate:

$$\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}$$
**Line Model: The $(1 + 1)$-ES**

- consider maximisation of objective function $f(x) = x$

- $(1 + 1)$-ES:

  \[ x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise} 
  \end{cases} \]

- progress rate:

  \[
  \varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_0^\infty z e^{-\frac{1}{2}z^2} \, dz
  \]

  \[
  = \frac{\sigma}{\sqrt{2\pi}}
  \]

  \[
  = \frac{\sigma}{\sqrt{2\pi}}
  \]
Line Model: The $(1 + 1)$-ES

- consider maximisation of objective function $f(x) = x$

- $(1 + 1)$-ES:

$$x^{(t+1)} = \begin{cases} x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\ x^{(t)} & \text{otherwise} \end{cases}$$

- progress rate:

$$\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}$$
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{z^2}{2}} \, dz
\]

\[
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E}\left[ x^{(t+1)} - x^{(t)} \right] \\
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z \, e^{-\frac{1}{2}z^2} \, dz \\
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The (1 + 1)-ES

- consider maximisation of objective function $f(x) = x$

- (1 + 1)-ES:

$$x^{(t+1)} = \begin{cases} x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\ x^{(t)} & \text{otherwise} \end{cases}$$

- progress rate:

$$\varphi_{1+1} = E \left[ x^{(t+1)} - x^{(t)} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}$$
Line Model: The \((1 + 1)-ES\)

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)-ES:\)

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = E \left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz = \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right] \\
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz \\
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The $(1 + 1)$-ES

- consider maximisation of objective function $f(x) = x$

- $(1 + 1)$-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz
\]

\[
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz
\]

\[
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E}\left[ x^{(t+1)} - x^{(t)} \right] \\
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} ze^{-\frac{1}{2}z^2} \, dz \\
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = E \left[ x^{(t+1)} - x^{(t)} \right] = \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz = \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

• consider maximisation of objective function \(f(x) = x\)

• \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
x^{(t)} & \text{otherwise}
\end{cases}
\]

• progress rate:

\[
\varphi_{1+1} = E \left[ x^{(t+1)} - x^{(t)} \right] \\
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-\frac{1}{2}z^2} \, dz \\
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The \((1 + 1)\)-ES

- consider maximisation of objective function \(f(x) = x\)

- \((1 + 1)\)-ES:

\[
x^{(t+1)} = \begin{cases} 
  x^{(t)} + \sigma z & \text{if } f(x^{(t)} + \sigma z) > f(x^{(t)}) \\
  x^{(t)} & \text{otherwise}
\end{cases}
\]

- progress rate:

\[
\varphi_{1+1} = \mathbb{E} \left[ x^{(t+1)} - x^{(t)} \right] \\
= \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} \frac{1}{2}z^2 \, dz \\
= \frac{\sigma}{\sqrt{2\pi}}
\]
Line Model: The $(1, \lambda)$-ES

- $(1, \lambda)$-ES:

\[ x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda} \]

where $k; \lambda$ denotes the index of the $k$th best offspring

- progress rate:

\[ \varphi_{1,\lambda} = \sigma E[z_{1;\lambda}] \]

\[ \equiv \sigma c_{1,\lambda} \]

- $c_{1,\lambda}$ is referred to as the $(1, \lambda)$-progress coefficient
Line Model: The \((1, \lambda)\)-ES

- \((1, \lambda)\)-ES:

\[ x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda} \]

where \(k; \lambda\) denotes the index of the \(k\)th best offspring

- progress rate:

\[ \varphi_{1,\lambda} = \sigma \mathbb{E}[z_{1;\lambda}] \equiv \sigma c_{1,\lambda} \]

- \(c_{1,\lambda}\) is referred to as the \((1, \lambda)\)-progress coefficient
Line Model: The \((1, \lambda)\)-ES

- \((1, \lambda)\)-ES:

\[
x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda}
\]

where \(k; \lambda\) denotes the index of the \(k\)th best offspring

- progress rate:

\[
\varphi_{1,\lambda} = \sigma \mathbb{E}[z_{1;\lambda}]
\]

\[
\equiv \sigma c_{1,\lambda}
\]

- \(c_{1,\lambda}\) is referred to as the \((1, \lambda)\)-progress coefficient
Line Model: The $(1, \lambda)$-ES

- $(1, \lambda)$-ES:
  \[
x(t+1) = x(t) + \sigma z_{1;\lambda}
  \]
  where $k; \lambda$ denotes the index of the $k$th best offspring

- progress rate:
  \[
  \varphi_{1,\lambda} = \sigma \mathbb{E}[z_{1;\lambda}]
  \equiv \sigma c_{1,\lambda}
  \]

- $c_{1,\lambda}$ is referred to as the $(1, \lambda)$-progress coefficient
Line Model: The \((1, \lambda)\)-ES

- \((1, \lambda)\)-ES:

\[ x^{(t+1)} = x^{(t)} + \sigma z_{1:}\lambda \]

where \(k; \lambda\) denotes the index of the \(k\)th best offspring

- progress rate:

\[ \varphi_{1,\lambda} = \sigma \mathbb{E}[z_{1;\lambda}] \equiv \sigma c_{1,\lambda} \]

- \(c_{1,\lambda}\) is referred to as the \((1, \lambda)\)-progress coefficient
Line Model: The $(1, \lambda)$-ES

- $(1, \lambda)$-ES:

\[ x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda} \]

where $k; \lambda$ denotes the index of the $k$th best offspring

- progress rate:

\[ \varphi_{1,\lambda} = \sigma E[z_{1;\lambda}] \]
\[ \equiv \sigma c_{1,\lambda} \]

- $c_{1,\lambda}$ is referred to as the $(1, \lambda)$-progress coefficient
Line Model: The \((1, \lambda)\)-ES

- \((1, \lambda)\)-ES:

\[
x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda}
\]

where \(k; \lambda\) denotes the index of the \(k\)th best offspring

- Progress rate:

\[
\varphi_{1,\lambda} = \sigma E[z_{1;\lambda}] \\
\equiv \sigma c_{1,\lambda}
\]

- \(c_{1,\lambda}\) is referred to as the \((1, \lambda)\)-progress coefficient
Line Model: The \((1, \lambda)\)-ES

- \(z_{k;\lambda}\) is the \((\lambda + 1 - k)\)th order statistic of a sample of \(\lambda\) independent, standard normally distributed random variables

\[
E[z_{k;\lambda}] = \frac{\lambda}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} \Phi(z)^{\lambda-k} [1 - \Phi(z)]^{k-1} \, dz
\]

- for large \(\lambda\),

\[
c_{1,\lambda} \propto \sqrt{2 \log \lambda}
\]

\(\Rightarrow\) the growth of the progress rate with \(\lambda\) is very slow

Noisy Line Model: The \((1, \lambda)\)-ES

- sources of noise in real-world problems:
  - inaccurate measurements
  - Monte Carlo methods
  - subjective selection
  - ...
- noisy fitness: \(f_\epsilon(x) = f(x) + \sigma_\epsilon z_\epsilon\)
- progress rate:
  \[
  \varphi_{1,\lambda} = \frac{\sigma c_{1,\lambda}}{\sqrt{1 + \vartheta^2}} \quad \vartheta = \frac{\sigma_\epsilon}{\sigma}: \text{noise-to-signal ratio}
  \]

\(\Rightarrow\) larger steps reduce the noise-to-signal ratio by amplifying the signal
Line Model: The $(\mu, \lambda)$-ES

- $(\mu, \lambda)$-ES: short for $(\mu/1, \lambda)$-ES; no recombination
- problem: the distribution of the population in search space needs to be modelled
- progress rate:

$$\varphi_{\mu,\lambda} = \sigma c_{\mu,\lambda}$$

- it can be shown that $c_{\mu,\lambda} \leq c_{1,\lambda}$ for any $\mu$

  $\Rightarrow$ keeping any but the best offspring deteriorates performance

Line Model: The $(\mu, \lambda)$-ES

- $(\mu, \lambda)$-ES: short for $(\mu/1, \lambda)$-ES; no recombination

- problem: the distribution of the population in search space needs to be modelled

- progress rate:

  $$\varphi_{\mu,\lambda} = \sigma c_{\mu,\lambda}$$

- it can be shown that $c_{\mu,\lambda} \leq c_{1,\lambda}$ for any $\mu$

  $\Rightarrow$ keeping any but the best offspring deteriorates performance

Line Model: The \((\mu, \lambda)\)-ES

- \((\mu, \lambda)\)-ES: short for \((\mu/1, \lambda)\)-ES; no recombination

- problem: the distribution of the population in search space needs to be modelled

- progress rate:

\[
\varphi_{\mu, \lambda} = \sigma c_{\mu, \lambda}
\]

- it can be shown that \(c_{\mu, \lambda} \leq c_{1, \lambda}\) for any \(\mu\)

\[\Rightarrow\] keeping any but the best offspring deteriorates performance

Line Model: The \((\mu, \lambda)\)-ES

- \((\mu, \lambda)\)-ES: short for \((\mu/1, \lambda)\)-ES; no recombination

- problem: the distribution of the population in search space needs to be modelled

- progress rate:

  \[ \varphi_{\mu, \lambda} = \sigma c_{\mu, \lambda} \]

- it can be shown that \(c_{\mu, \lambda} \leq c_{1, \lambda}\) for any \(\mu\)

  \(\Rightarrow\) keeping any but the best offspring deteriorates performance

Line Model: The \((\mu, \lambda)\)-ES

- \((\mu, \lambda)\)-ES: short for \((\mu/1, \lambda)\)-ES; no recombination

- problem: the distribution of the population in search space needs to be modelled

- progress rate:

  \[
  \varphi_{\mu,\lambda} = \sigma c_{\mu,\lambda}
  \]

- it can be shown that \(c_{\mu,\lambda} \leq c_{1,\lambda}\) for any \(\mu\)

  \(\Rightarrow\) keeping any but the best offspring deteriorates performance

Line Model: The \((\mu, \lambda)\)-ES

- \((\mu, \lambda)\)-ES: short for \((\mu/1, \lambda)\)-ES; no recombination

- problem: the distribution of the population in search space needs to be modelled

- progress rate:

\[
\varphi_{\mu, \lambda} = \sigma c_{\mu, \lambda}
\]

- it can be shown that \(c_{\mu, \lambda} \leq c_{1, \lambda}\) for any \(\mu\)

\[\Rightarrow\] keeping any but the best offspring deteriorates performance

Line Model: The $((\mu, \lambda))$-ES

- $((\mu, \lambda))$-ES: short for $((\mu/1, \lambda))$-ES; no recombination
- problem: the distribution of the population in search space needs to be modelled
- progress rate:
  \[
  \varphi_{\mu,\lambda} = \sigma c_{\mu,\lambda}
  \]
- it can be shown that $c_{\mu,\lambda} \leq c_{1,\lambda}$ for any $\mu$
  \[\Rightarrow\] keeping any but the best offspring deteriorates performance

Noisy Line Model: The \((\mu, \lambda)-\text{ES}\)

- the signal strength is \(\sqrt{\sigma^2 + D^2}\),
  where \(D^2\) is the variance of the population

- \(D\) is proportional to \(\sigma\) and increases with increasing \(\mu\)

\(\Rightarrow\) increasing the size of the population decreases the noise-to-signal ratio


Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_k; \lambda
\]

- progress rate:

\[
\varphi_{\mu/\mu, \lambda} = \sigma c_{\mu/\mu, \lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu, \lambda} = \frac{\sigma c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
    x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda}
\]

- progress rate:

\[
    \varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda}
\]

- in the presence of noise:

\[
    \varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_k; \lambda
\]

- progress rate:

\[
\varphi_{\mu/\mu, \lambda} = \sigma_c_{\mu/\mu, \lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu, \lambda} = \frac{\sigma_c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The $(\mu/\mu, \lambda)$-ES

- $(\mu/\mu, \lambda)$-ES: recombination of $\rho = \mu$ parents contracts the population to a point
  
  $$x(t+1) = x(t) + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_k; \lambda$$

- progress rate:
  
  $$\varphi_{\mu/\mu, \lambda} = \sigma c_{\mu/\mu, \lambda}$$

- in the presence of noise:
  
  $$\varphi_{\mu/\mu, \lambda} = \frac{\sigma c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}}$$
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_k; \lambda
\]

- progress rate:

\[
\varphi_{\mu/\mu, \lambda} = \sigma c_{\mu/\mu, \lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu, \lambda} = \frac{\sigma c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \sigma \sum_{k=1}^{\mu} z_{k;\lambda}
\]

- progress rate:

\[
\varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The $\left( \mu/\mu, \lambda \right)$-ES

- $\left( \mu/\mu, \lambda \right)$-ES: recombination of $\rho = \mu$ parents contracts the population to a point

\[ x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda} \]

- progress rate:

\[ \varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda} \]

- in the presence of noise:

\[ \varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}} \]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda}
\]

- progress rate:

\[
\varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[ x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda} \]

- progress rate:

\[ \varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda} \]

- in the presence of noise:

\[ \varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}} \]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_k; \lambda
\]

- progress rate:

\[
\varphi_{\mu/\mu, \lambda} = \sigma c_{\mu/\mu, \lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu, \lambda} = \frac{\sigma c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

- \((\mu/\mu, \lambda)\)-ES: recombination of \(\rho = \mu\) parents contracts the population to a point

\[
x^{(t+1)} = x^{(t)} + \frac{\sigma}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda}
\]

- progress rate:

\[
\varphi_{\mu/\mu,\lambda} = \sigma c_{\mu/\mu,\lambda}
\]

- in the presence of noise:

\[
\varphi_{\mu/\mu,\lambda} = \frac{\sigma c_{\mu/\mu,\lambda}}{\sqrt{1 + \vartheta^2}}
\]
Line Model: The \((\mu/\mu, \lambda)\)-ES

• \((\mu/\mu, \lambda)\)-progress coefficient:

\[
c_{\mu/\mu, \lambda} = \frac{\lambda - \mu}{2\pi} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} z e^{-z^2} [\Phi(z)]^{\lambda-\mu-1} [1 - \Phi(z)]^{\mu-1} \, dz
\]

• in general, \(c_{\mu/\mu, \lambda} \lesssim c_{\mu, \lambda} \leq c_{1, \lambda}\)

The Sphere Model (1)

- sphere model: minimise

\[ f(x) = \sum_{i=1}^{N} x_i^2 \]

- assume that \( N \) is large


The Sphere Model  (2)

• consider candidate solution

\[ y = x + \sigma z \]

• fitness:

\[
\begin{align*}
    f(y) &= r^2 \\
    &= (R - \sigma z_A)^2 + \sigma^2 \|z_B\|^2 \\
    &= R^2 - 2R\sigma z_A + \sigma^2 \|z\|^2 \\
    &= f(x) - 2R\sigma z_A + \sigma^2 \|z\|^2
\end{align*}
\]

• \( \sigma^2 \|z\|^2 \) deteriorates fitness, limiting useful mutation strengths
The Sphere Model (3)

- if \( z \) is a mutation vector, then
  1. \( z_A \) is standard normally distributed
  2. \( \|z\|^2 \) is \( \chi^2_N \)-distributed

- the \( \chi^2_N \)-distribution has mean \( N \) and standard deviation \( \sqrt{2N} \)

- the coefficient of variation of the \( \chi^2_N \)-distribution tends to zero as \( N \) increases

\[ \Rightarrow \text{for the range of mutation strengths of interest and large enough } N, \|z\|^2 \text{ can be replaced with } N \]
The Sphere Model (4)

- offspring fitness:
  \[ f(y) = f(x) - 2R\sigma z_A + N\sigma^2 \]

- with normalised mutation strength \( \sigma^* = \sigma N/R \):
  \[ f(y) = f(x) \left[ 1 - \frac{2}{N} \left( \sigma^* z_A - \frac{\sigma^*}{2} \right) \right] \]

- evolution strategies converge linearly on the sphere model provided that the mutation strength is adapted properly

- the rate of convergence is inversely proportional to \( N \)
Sphere Model: The $(1 + 1)$-ES

- **progress rate** (Rechenberg, 1973):

  \[ \varphi_{1+1}^* = \frac{\sigma^*}{\sqrt{2\pi}} e^{-\frac{1}{8}\sigma^*^2} - \frac{\sigma^*^2}{2} \left[ 1 - \Phi \left( \frac{\sigma^*}{2} \right) \right] \]

- **success probability**:

  \[ P_{\text{succ}} = 1 - \Phi \left( \frac{\sigma^*}{2} \right) \]

- **maximal progress rate**:

  \[ \varphi_{1+1}^* = 0.202 \]

  at mutation strength \( \sigma^* = 1.224 \)
Sphere Model: The \((1 + 1)\)-ES

- 1/5th success rule (Rechenberg, 1973):
  Decrease the mutation strength if the percentage of successful mutations is below one fifth; increase it if it is above.

- simple implementation:

  \[
  \sigma^{(t+1)} = \sigma^{(t)} \cdot \begin{cases} 
  2^{1/N} & \text{on success} \\
  2^{-0.25/N} & \text{otherwise}
  \end{cases}
  \]

Sphere Model: The $\text{(1, } \lambda\text{)}$-ES

- progress rate (Rechenberg, 1984):
  \[ \varphi^*_{1,\lambda} = \sigma^* c_{1,\lambda} - \frac{\sigma^*}{2} \]

- optimal progress rate:
  \[ \varphi^*_{1,\lambda} = \frac{c_{1,\lambda}^2}{2} \propto \log \lambda \]
  at mutation strength $\sigma^* = c_{1,\lambda}$

- the $(1, \lambda)$-ES is less efficient than the simple $(1+1)$-ES unless offspring can be evaluated in parallel
Noisy Sphere Model: The \((1, \lambda)\)-ES

- assume noise of strength \(\sigma_{\epsilon}(x)\), proportional to \(R^2\)

- progress rate (Beyer, 1993):

\[
\varphi_{1,\lambda} = \frac{\sigma^* c_{1,\lambda}}{\sqrt{1 + \vartheta^2}} - \frac{\sigma^*}{2}
\]

where \(\vartheta = \sigma_{\epsilon}/\sigma^*\) is the noise-to-signal ratio

- averaging over multiple fitness evaluations reduces the noise strength, but is expensive
Noisy Sphere Model: Rescaled Mutations

- \((1, \lambda)\)-ES with rescaled mutations:
  - generate and evaluate offspring
    \[ y_i = x^{(t)} + \kappa \sigma z_i \]
  - then let
    \[ x^{(t+1)} = x^{(t)} + \sigma z_{1;\lambda} \]

- progress rate:
  \[
  \varphi^* = \frac{\sigma^* c_{1,\lambda}}{\sqrt{1 + \left(\frac{\sigma^* \epsilon}{(\kappa \sigma^*)}\right)^2}} - \frac{\sigma^*}{2}
  \]


Mutative Self-Adaptation

- every candidate solution carries its own set of strategy parameters
  - mutation:
    \[
    \sigma_i = \sigma^{(t)} \exp(\tau N(0, 1)) \\
    y_i = x^{(t)} + \sigma_i z_i
    \]
  - selection:
    \[
    \sigma^{(t+1)} = \sigma_{1;\lambda} \\
    x^{(t+1)} = y_{1;\lambda}
    \]

- have a competition of strategic ideas; a “good” set of strategy parameters increases the chance of generating a good set of object parameters and is thus likely to prevail under selection

Hierarchically Organised Evolution Strategies

- problems with mutative self-adaptation:
  - selection of strategy parameters is indirect and noisy
  - rewarding short-term success may be shortsighted

- hierarchically organised ES “try out” strategy parameter settings for longer periods of time
  \[\Rightarrow\] evolve several populations in isolation from each other; compare their relative success after a number of time steps


Sphere Model: The \((\mu/\mu, \lambda)\)-ES

- progress rate (Rechenberg, 1994):

\[
\varphi_{\mu/\mu,\lambda}^* = \sigma^* c_{\mu/\mu,\lambda} - \frac{\sigma^*}{2\mu}
\]

- maximal progress rate:

\[
\varphi_{\mu/\mu,\lambda}^* = \frac{\mu c_{\mu/\mu,\lambda}^2}{2} \propto \mu
\]

at mutation strength \(\sigma^* = \mu c_{\mu/\mu,\lambda}\)

- the maximal (serial!) efficiency is asymptotically equal to that of the \((1 + 1)\)-ES
Sphere Model: The \((\mu/\mu, \lambda)\)-ES

- the components toward the optimum of the selected mutation vectors are correlated, the other components are not

\[
\begin{align*}
\text{consider vector} & \\
\mathbf{z}^{(\text{avg})} &= \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbf{z}_k; \lambda
\end{align*}
\]

- the length of the “harmful” components of the mutation vectors is reduced

- the purpose of recombination is similarity extraction;

\[\Rightarrow \text{“genetic repair principle”} \quad (\text{Beyer, 1995})\]
Noisy Sphere Model: The $(\mu/\mu, \lambda)$-ES

- progress rate in the presence of noise:
  \[
  \varphi_{\mu/\mu, \lambda}^* = \frac{\sigma^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \vartheta^2}} - \frac{\sigma^*^2}{2\mu}
  \]
  where $\vartheta = \sigma^*_\epsilon / \sigma^*$ is the noise-to-signal ratio

- the larger mutation strengths (compared to the $(1, \lambda)$-ES) reduce $\vartheta$

- the $(\mu/\mu, \lambda)$-ES implicitly rescales mutation vectors

Sphere Model: Optimally Weighted Recombination

- weighted recombination: replace

\[ x = \frac{1}{\mu} \sum_{k=1}^{\mu} x_k; \lambda \]

with

\[ x = \sum_{k=1}^{\lambda} w_k x_k; \lambda \]

- for maximal progress, choose

\[ w_k \propto E[z_k; \lambda] \]

- the proportionality constant determines the amount of implicit re-scaling; the speed-up is 2.5-fold

Cumulative Step Length Adaptation (1)

- postulate: consecutive steps of the strategy should be uncorrelated

- if consecutive steps are positively correlated, then the step length should be increased

- if consecutive steps are negatively correlated, then the step length should be decreased

Cumulative Step Length Adaptation  (2)

• in order to detect correlations, information from a number of steps needs to be accumulated

⇒ (for the \((\mu/\mu, \lambda)\)-ES) define the search path

\[
s^{(t+1)} = (1 - c)s^{(t)} + \sqrt{\mu c(2 - c)}z^{(avg)}
\]

• under random selection, the expected squared length of the search path is \(N\)

• the step length is updated according to

\[
\sigma^{(t+1)} = \sigma^{(t)} \exp \left( \frac{\|s^{(t+1)}\|^2 - N}{2DN} \right)
\]
Cumulative Step Length Adaptation (3)

- on the noisy sphere, cumulative step length adaptation generates

\[ \sigma^* = \mu C_{\mu/\mu,\lambda} \sqrt{2 - \left( \frac{\sigma_{\epsilon}^*}{\mu C_{\mu/\mu,\lambda}} \right)^2} \]

and achieves progress rate

\[ \varphi^* = \frac{\sqrt{2 - 1}}{2} \mu C_{\mu/\mu,\lambda} \left( 2 - \left( \frac{\sigma_{\epsilon}^*}{\mu C_{\mu/\mu,\lambda}} \right)^2 \right) \]

Comparison of Strategies (1)

Direct Pattern Search: (Hooke and Jeeves, 1961) precursor of many direct search strategies

Multi-Directional Search: (Torczon, 1989) successor of Nelder and Mead’s simplex method, the most popular strategy for noisy optimisation

Implicit Filtering: (Gilmore and Kelley, 1995) gradient strategy using finite differencing and Armijo line searches; designed for noisy optimisation

Evolution Strategy: ($\mu_0/\mu, \lambda$)-ES with cumulative step length adaptation and various population sizes

Comparison of Strategies  (2)

- incomplete graphs result from failure to achieve linear convergence
- larger populations buy robustness at the price of efficiency
- strengths of ES: genetic repair and relatively robust step length adaptation
Ridge Model

- ridge model:

\[ f(x) = x_1 - d \left( \sum_{i=2}^{N} x_i^2 \right)^{\alpha/2} \]


Convex Quadratic Functions (1)

- convex quadratic functions:

\[ f(x) = \sum_{i=1}^{N} (a_i x_i)^2 \]

- Hessian matrix:

\[
H = \begin{pmatrix}
2a_1^2 & 0 & \cdots & 0 \\
0 & 2a_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 2a_N^2
\end{pmatrix}
\]

- condition number: \( \frac{a_{\text{max}}}{a_{\text{min}}} \)

- for strategies that are not rotationally invariant, the coordinate system should be rotated
Convex Quadratic Functions (2)

• examples of convex quadratic functions:

\[ f_{\text{cigar}}(\mathbf{x}) = x_1^2 + \sum_{i=2}^{N} (1000x_i)^2 \]

\[ f_{\text{discus}}(\mathbf{x}) = (1000x_1)^2 + \sum_{i=2}^{N} x_i^2 \]

\[ f_{\text{ellipsoid}}(\mathbf{x}) = \sum_{i=1}^{N} \left(1000(i-1)/(N-1)x_i\right)^2 \]

\[ f_{\text{twoaxes}}(\mathbf{x}) = \sum_{i=1}^{\lfloor N/2 \rfloor} x_i^2 + \sum_{\lfloor N/2 \rfloor + 1}^{N} (1000x_i)^2 \]
Convex Quadratic Functions (3)

- performance of the $(1+1)$-ES using isotropically distributed mutations (Jägersküpper, 2006):
  - if the condition number is $O(1)$, the number of steps needed to reduce the approximation error to a $2^{-b}$-fraction is $\Theta(bN)$
  - for $f_{\text{twoaxes}}$ with condition number $\xi$ polynomially bounded in $N$ such that $1/\xi \to 0$ as $N \to \infty$, the number of steps is $\Theta(b\xi N)$

- similar results hold for the $(\mu/\mu, \lambda)$-ES with cumulative step length adaptation (Arnold, 2007)


Nonisotropically Distributed Mutations

- nonisotropic mutation distributions can be vastly more efficient than isotropic ones
- ideally, the mutation covariance matrix should be proportional to the inverse of the local Hessian of the objective


Covariance Matrix Adaptation (1)

- the CMA-ES adapts the mutation covariance matrix based on information gathered in past steps (Hansen and Ostermeier, 2001)
  - variances in directions that have previously been successful are increased
  - other variances decay over time

- the strategy is rotationally invariant


Covariance Matrix Adaptation Evolution Strategy (1)

- state variables of the \((\mu/\mu, \lambda)\)-CMA-ES: \(x, \sigma, C, s_\sigma, s_C\)

- in every iteration:
  1. Compute an eigen decomposition \(C = BD(BD)^T\).
  2. Generate \(\lambda\) offspring \(y_i = x + \sigma BDz_i\).
  3. Compute the mean of the \(\mu\) best offspring
     \[
     z^{(\text{avg})} = \frac{1}{\mu} \sum_{k=1}^{\mu} z_k; \lambda.
     \]
  4. Update the search point
     \[
     x \leftarrow x + \sigma BDz^{(\text{avg})}.
     \]
  5. Update \(s_\sigma\) and \(s_C\).
  6. Update \(C\) and \(\sigma\).
Covariance Matrix Adaptation Evolution Strategy (2)

• update of the search paths:

\[ s_\sigma \leftarrow (1 - c_\sigma) s_\sigma + \sqrt{\mu c_\sigma(2 - c_\sigma)} B z^{(avg)} \]
\[ s_C \leftarrow (1 - c_C) s_C + \sqrt{\mu c_C(2 - c_C)} B D z^{(avg)} \]

• update of the step length:

\[ \sigma \leftarrow \sigma \exp \left( \frac{\|s_\sigma\|^2 - N}{2DN} \right) \]

• update of the covariance matrix:

\[ C \leftarrow (1 - c_{cov}) C + c_{cov} s_C s_C^T \]

• \( c_\sigma, c_C, \) and \( c_{cov} \) determine how quickly old information fades
Covariance Matrix Adaptation Evolution Strategy (3)

- better use can be made of large populations by using covariance matrix update

\[ C \leftarrow (1 - c_{\text{cov}})C + c_{\text{cov}} \left( \alpha_{\text{cov}} s_{C} s_{C}^T + (1 - \alpha_{\text{cov}})Z \right) \]

where

\[ Z = BD \left( \frac{1}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda} z_{k;\lambda}^T \right) (BD)^T \]

- \( c_{\text{cov}} \) can be chosen larger (roughly by a factor of \( \mu \)) than for the previous update rule

- \( \alpha_{\text{cov}} \) weights the path-based and population-based contributions

Covariance Matrix Adaptation Evolution Strategy (4)

- performance on $f_{\text{ellipsoid}}$ with $N = 10$

small population ($\mu = 2$, $\lambda = 8$): large population ($\mu = 10$, $\lambda = 40$):
Active Covariance Matrix Adaptation  (1)

- idea: use information not only from successful, but also from unsuccessful offspring
  \[ \Rightarrow \text{actively decrease variances in directions that repeatedly yield bad offspring} \]

- new update rule:
  \[ C \leftarrow (1 - c_{\text{cov}})C + c_{\text{cov}} s_C s_C^T + \beta Z \]

where
\[
Z = BD \left( \frac{1}{\mu} \sum_{k=1}^{\mu} z_{k;\lambda} z_{k;\lambda}^T - \frac{1}{\mu} \sum_{k=\lambda-\mu+1}^{\lambda} z_{k;\lambda} z_{k;\lambda}^T \right) (BD)^T
\]

Active Covariance Matrix Adaptation (2)

- performance across a set of test functions ($N = 10$)

small population ($\mu = 2, \lambda = 8$): large population ($\mu = 10, \lambda = 40$):
Active Covariance Matrix Adaptation (3)

• dependence on $N$

• test function: $f_{\text{ellipsoid}}$

small population ($\mu = 2$, $\lambda = 8$): large population ($\mu = N$, $\lambda = 4N$):
(1 + 1)-CMA-ES

- the eigen decomposition is computationally expensive for large $N$

- for the (1 + 1)-CMA-ES,
  - the mutation strength can be controlled using the $1/5$th rule
  - matrix $A = BD$ can be updated directly, with no need to decompose the covariance matrix

- for multimodal (and presumably for noisy) functions, the longer steps of the $(\mu/\mu, \lambda)$-CMA-ES are advantageous

Differential Evolution

• mutation (DE/rand/1): randomly pick \( i \neq j \neq k \) and let

\[
y = x_i + k(x_j - x_k)
\]

⇒ the step length is determined by the diversity of the population

• the rate at which diversity decreases is influenced by the population size, the replacement mechanism, and the factor \( k \)

• adaptive variants exist

PCX — Parent Centric Recombination

- assume wlog that \( x_1 \) has been picked as parent; compute

\[
y = x_1 + \sigma_1 z_1 d_1 + \sigma_2 D \sum_{i=2}^{\mu} z_i e_i
\]

where \( x \) is the population centroid, \( d_i = x_i - x \), \( e_i \) is the normalised vector consisting of those components of \( d_i \) that are perpendicular to \( x_1 \), and \( D \) is the average distance of the \( x_i \) from the line through \( x \) and \( x_1 \).

Summary and Further Topics

topics covered:

• evolution strategies:
  – benefits of populations and recombination
  – coping with noise

• step length adaptation: success rate based; mutative; cumulative

• covariance matrix adaptation: passive and active; alternative approaches

further topics:

• multimodal problems: restart strategies; niching methods; spatially organised populations

• constraint handling techniques