5. 差分进化

堵威

华东理工大学 自动化系

2021.4.8

本章内容

1. 背景知识

2. 基本差分进化

3. 差分进化变种

4. 离散优化

5. 差分进化与遗传算法

本章内容

1. 背景知识
2. 基本差分进化
3. 差分进化变种
4. 离散优化
5. 差分进化与遗传算法

背景知识

- 差分进化的概述
  - Differential Evolution：Rainer Storn和Kenneth V. Price于1995年针对连续优化问题提出的基于种群差异的元启发式随机搜索算法
  - 由实际问题推动：切比雪夫多项式系数的解和数字滤波器系数的优化
  - 参数少，易于编写、实施

背景知识

- 差分进化的概述
  - 性能有竞争力

1996	the First International Contest on Evolutionary Optimization (1st ICEO), 第三名
1997	the Second International Contest on Evolutionary Optimization (2nd ICEO), 最好的之一
2005	2005 CEC competition on real parameter optimization (10D), 实数优化竞赛，经典DE第一名，SaDE，第三名
2006	CEC 2006 competition on constrained real parameter optimization, 受限优化竞赛，第一名
2007	CEC 2007 competition on multiobjective optimization, 多目标优化竞赛，第二名
2008	CEC 2008 competition on large scale global optimization, 大规模优化竞赛，第三名
2009	CEC 2009 competition on multiobjective optimization 多目标优化竞赛，(第一名, DE-based MOEA/D) CEC 2009 competition on evolutionary computation in dynamic and uncertain environments, 动态优化竞赛，第一名

本章内容

1. 背景知识

2. **基本差分进化**

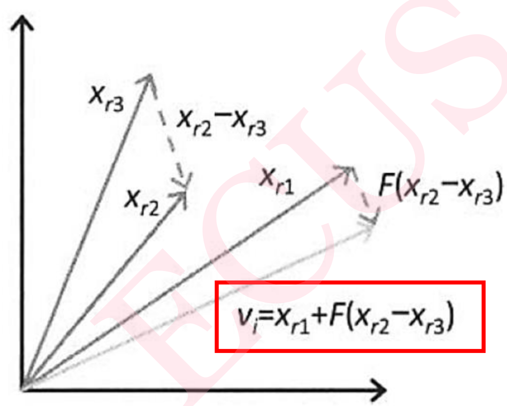
3. 差分进化变种

4. 离散优化

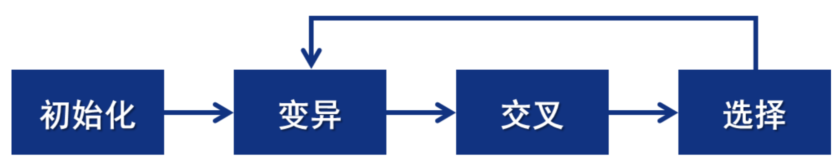
5. 差分进化与遗传算法

基本差分进化

- 基本框架
  - 差分进化是为了优化n维连续域中的函数而设计的基于种群的算法，种群中每一个个体都是用来表示候选解的一个n维向量
  - 基本思路：取两个个体之间的差分向量，将这个差分向量的一个伸缩版加到第三个个体上从而产生一个新的候选解（变异）



- 随机选出$\mathbf{x}_{r2}$和$\mathbf{x}_{r3}$这两个个体，$r_2 \neq r_3$
- 这两个个体之间的差的伸缩版加到第三个随机选出的个体$\mathbf{x}_{r1}$，这里$r_1 \notin \{r_2, r_3\}$，得到变异$\mathbf{v}_i$



基本差分进化

- 基本框架
  - 在建立变异向量v_i之后，它与一个差分进化个体x_i结合（交叉），这里$i \notin \{r_1, r_2, r_3\}$，产生一个试验向量$u_i$，按如下方式实施交叉：对于$j \in [1,n]$，

$$u_{ij} = \begin{cases} v_{ij} & \text{if } (r_{cj} < c) \text{ or } (j = \mathcal{J}_r) \\ x_{ij} & \text{otherwise} \end{cases}$$

其中，n是问题的维数，也是x_i，v_i和u_i的维数；u_{ij}是u_i的第j个分量；v_{ij}是v_i的第j个分量；x_{ij}是x_i的第j个分量；r_{cj}是在[0,1]均匀分布的随机数；$c \in [0,1]$是给定的交叉率；J_r是在[1,n]上均匀分布的一个随机整数。

基本差分进化

- 基本框架
  - 选择：在生成试验向量$\mathbf{u}_i$之后，比较向量$\mathbf{u}_i$和$\mathbf{x}_i$，适应性更强的向量留下来作为**差分进化**的下一代

经典差分进化（DE/rand/1/bin）伪代码

F=尺度参数，F∈[0.4,0.9]
c=交叉率，c∈[0.1,1]
初始化候选解种群{$\mathbf{x}_i$}, i∈[1,N]
While not (终止准则)
    For 每一个个体$\mathbf{x}_i$, i∈[1,N]
        $r_1 \leftarrow$ 随机整数∈[1,N]：$r_1 \neq i$
        $r_2 \leftarrow$ 随机整数∈[1,N]：$r_2 \notin \{i, r_1\}$
        $r_3 \leftarrow$ 随机整数∈[1,N]：$r_3 \notin \{i, r_1, r_2\}$
        $\mathbf{v}_i \leftarrow \mathbf{x}_{r1}+F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$ (变异向量)
        $J_r \leftarrow$ 随机整数， $J_r \in [1,n]$
        For 每一维j∈[1,n]
            $r_{cj} \leftarrow$ 随机数，$r_{cj} \in [0,1]$
            If ($r_{cj} < c$) or ($j = J_r$) then
                $u_{ij} \leftarrow v_{ij}$
            else
                $u_{ij} \leftarrow x_{ij}$
            End if
        下一维
    下一个个体
    For 每一个种群中的个体 i∈[1, N]
        If $f(\mathbf{u}_i) < f(\mathbf{x}_i)$ then $\mathbf{x}_i \leftarrow \mathbf{u}_i$ Endif
    下一个个体
下一代

基本差分进化

- 基本框架
  - 调试参数：种群N，尺度参数F，交叉率c
  - 参数设定的经验值：Storn和Price建议N的取值为[5D，10D]，50,100；F的取值[0.4,0.9]；c的取值[0.1,1]
  - 为什么经典差分进化算法(DE/rand/1/bin)会有效？

1）随着种群逐渐缩小到问题的最优解，形式为($\mathbf{x}_{r2}$-$\mathbf{x}_{r3}$)的扰动会减小

2）不同维度上的扰动大小也不同，依赖于问题的规模（$\mathbf{x}_{r2}$-$\mathbf{x}_{r3}$的第p个分量大小与种群在第p维与问题的解的接近程度成比例）

3）在维度之间的扰动步长是相关的

本章内容

1. 背景知识

2. 基本差分进化

3. 差分进化变种

4. 离散优化

5. 差分进化与遗传算法

差分进化变种

- 差分进化的命名
  - DE/rand/1/bin：DE/x/y/z
  - DE：differential evolution
  - x：被扰动的基向量
  - y：扰动x的差分向量个数
  - z：交叉的方式

DE/rand/1/bin

random

1个差分向量

binomial：二项分布的

差分进化变种

- 变异向量
  - 经典差分进化算法中使用的变异向量为DE/rand/1：

$$\mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$$

  - 被扰动的基向量为随机挑选的$\mathbf{x}_{r1}$
  - 变种1：被扰动的基向量为种群中最好的个体$\mathbf{x}_{best}$

$$\text{DE/best/1:}\quad \mathbf{v}_i \leftarrow \mathbf{x}_{best} + F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$$

将种群中最好的个体作为基向量，可以起到增加开发减少探索的效果

差分进化变种

- 变异向量
  - 变种2：差分向量的个数设为2

DE/rand/2：　$\mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_{r2}-\mathbf{x}_{r3}) + F(\mathbf{x}_{r4}-\mathbf{x}_{r5})$

DE/best/2：　$\mathbf{v}_i \leftarrow \mathbf{x}_{best} + F(\mathbf{x}_{r2}-\mathbf{x}_{r3}) + F(\mathbf{x}_{r4}-\mathbf{x}_{r5})$

其中，r_4为随机整数，$r_4 \in [1,N]$：$r_4 \notin \{i, r_1, r_2, r_3\}$；$r_5$为随机整数，$r_5 \in [1,N]$：$r_5 \notin \{i, r_1, r_2, r_3, r_4\}$

- 增加了差分向量对变异向量的影响，增加了算法探索的能力，总的差分向量不限于一对向量的差的方向，有了更多的自由度。

差分进化变种

- 变异向量
  - 变种3：用当前的$\mathbf{x}_i$作为基向量

DE/target/1： $\mathbf{v}_i \leftarrow \mathbf{x}_i + F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$

DE/target/2： $\mathbf{v}_i \leftarrow \mathbf{x}_i + F(\mathbf{x}_{r2}-\mathbf{x}_{r3}) + F(\mathbf{x}_{r4}-\mathbf{x}_{r5})$

  - 变种4：在生成差分向量时利用种群中最好的个体$\mathbf{x}_b$，让变异向量全部都移向$\mathbf{x}_b$，从$\mathbf{x}_b$减掉的向量可以是随机个体或目标向量

$\mathbf{v}_i \leftarrow \mathbf{x}_i + F(\mathbf{x}_b-\mathbf{x}_i)$

$\mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_b-\mathbf{x}_{r3})$

$\mathbf{v}_i \leftarrow \mathbf{x}_b + F(\mathbf{x}_{r2}-\mathbf{x}_{r3})+F(\mathbf{x}_b-\mathbf{x}_{r5})$

$\mathbf{v}_i \leftarrow \mathbf{x}_i + F(\mathbf{x}_b-\mathbf{x}_i)+F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$（DE/target-to-best/1
或DE/current-to-best/1）

差分进化变种

- 变异向量
  - 通过随机决定变异向量的方式，可以将不同的方案组合起来，如：

p_f = 变异概率，$p_f \in [0,1]$

$a \leftarrow$ 随机数，$a \in [0,1]$

If $a < p_f$ then

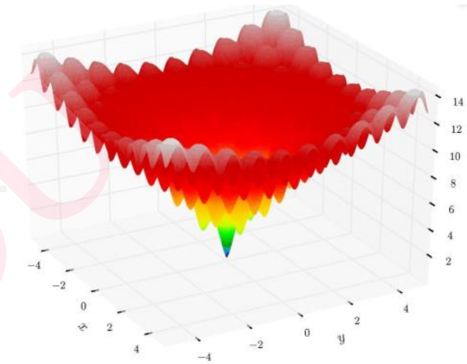
$\mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_{r2} - \mathbf{x}_{r3})$

else

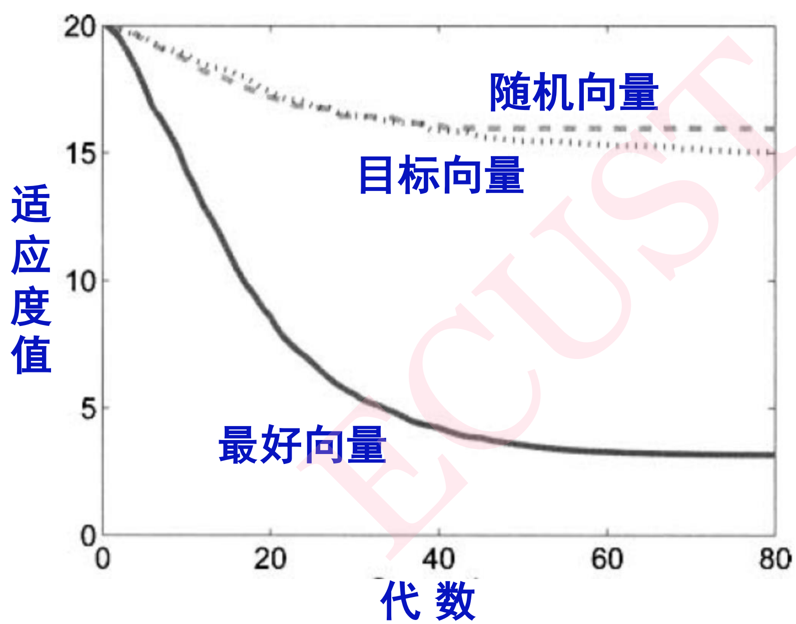
$\mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_{r2} - \mathbf{x}_{r3}) + F(\mathbf{x}_{r4} - \mathbf{x}_{r5})$

End if

差分进化变种

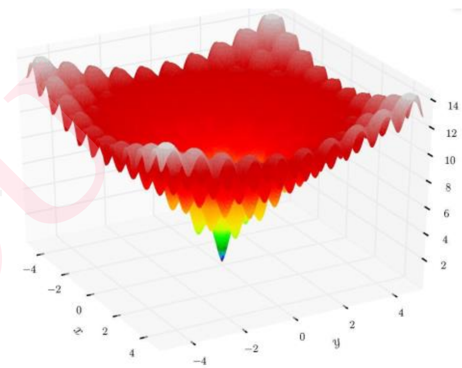


- 变异向量
  - 例：将差分进化应用于20维Ackley函数，分别采用随机向量作为基向量、最好向量作为基向量和目标向量作为基向量

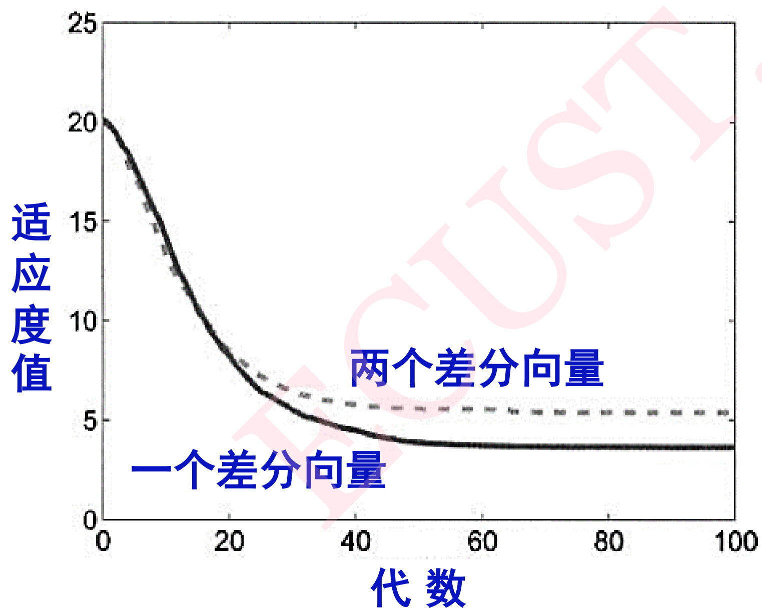


- 随机向量、目标向量方案效果相似
- 最好向量作为基向量能大大提高性能

差分进化变种



- 变异向量
  - 例：将差分进化应用于20维Ackley函数，采用最好向量作为基向量，差分向量个数分别取1个和2个。



- 在该问题上，使用一个差分向量比使用两个差分向量略好一些

差分进化变种

- 试验向量
  - 经典差分进化算法中使用的生成试验向量的方案（交叉）：

$$u_{ij} = \begin{cases} v_{ij} & \text{if } (r_{cj} < c) \text{ or } (j = \mathcal{J}_r) \\ x_{ij} & \text{otherwise} \end{cases}$$

上述方案中将v_{ij}复制给u_{ij}的概率与是否将v_{ij-1}复制给u_{ij-1}的概率相同，也就是说没有把由$\mathbf{v}_i$或$\mathbf{x}_i$得到的解的特征一起保留下来。

  - 有许多问题的适应度依赖于解的特征组合而不是解的单个特征，因此需要将解的特征组合保留下来

差分进化变种

- 试验向量
  - DE/rand/1/L：生成一个随机数L∈[1,n]，将$\mathbf{v}_i$的L个连续的特征复制给$\mathbf{u}_i$，然后从$\mathbf{x}_i$把余下的性质复制给$\mathbf{u}_i$

假设一个n=7维的问题，DE/rand/1/L首先生成一个随机整数L∈[1,n]，假设L=3。生成一个随机的起点s∈[1,n]，假设s=6：

$u_{i1} \leftarrow v_{i1}$，$u_{i2} \leftarrow x_{i2}$，$u_{i3} \leftarrow x_{i3}$，$u_{i4} \leftarrow x_{i4}$，

$u_{i5} \leftarrow x_{i5}$(终点)，$u_{i6} \leftarrow v_{i6}$ (起点s)，$u_{i7} \leftarrow v_{i7}$.

差分进化变种

- 试验向量
  - DE/rand/1/L循环伪代码：

$L \leftarrow$ 随机整数，$L \in [1,n]$
$s \leftarrow$ 随机整数，$s \in [1,n]$
$J \leftarrow \{s, \min\{n, s+L-1\}\} \cup \{1, s+L-n-1\}$
For 每一维$j \in [1,n]$
    If $j \in J$ then
        $u_{ij} \leftarrow v_{ij}$
    else
        $u_{ij} \leftarrow x_{ij}$
    End if
下一维

For 每一维$j \in [1,n]$
    $r_{cj} \leftarrow$ 随机数，$r_{cj} \in [0,1]$
    If $(r_{cj} < c)$ or $(j = J_r)$ then
        $u_{ij} \leftarrow v_{ij}$
    else
        $u_{ij} \leftarrow x_{ij}$
    End if
下一维

差分进化变种

- 试验向量
  - 思考：对于给定的指标i，变异向量的元素v_{ij}平均有多少个被复制到试验向量的特征u_{ij}?

对于DE/rand/1/bin，在“for 每一维”循环有n次迭代；

这些迭代中有1次会以100%的概率将v_{ij}复制给u_{ij}，

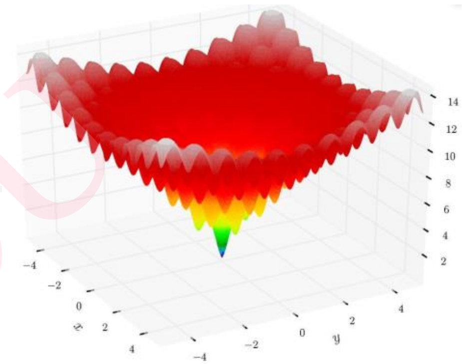
其余n-1次迭代以c的概率将v_{ij}复制给u_{ij}；

复制给试验向量的元素v_{ij}的个数的期望为：$E=1+c(n-1)$

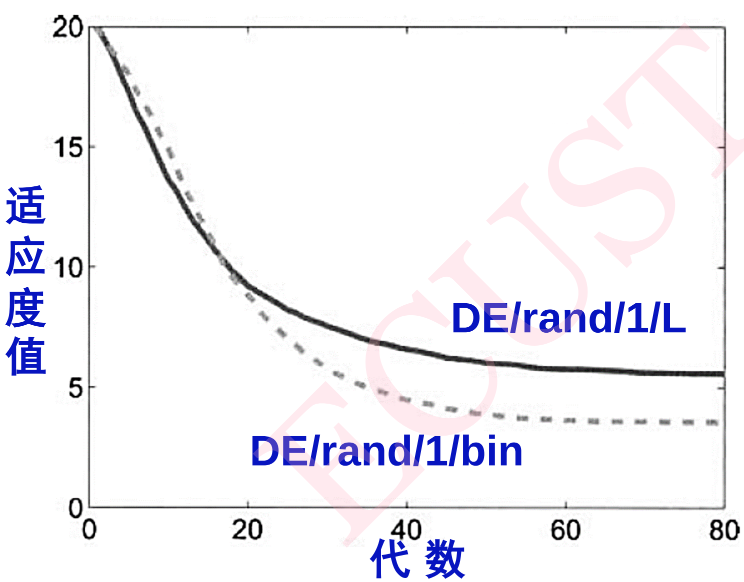
对于DE/rand/1/L，L在[1,n]上均匀分布；

复制给试验向量的元素v_{ij}的个数的期望为：$E=n/2$

差分进化变种



- 试验向量
  - 例：将DE/rand/1/bin和DE/rand/1/L应用于20维Ackley函数，种群规模为50，F=0.4，c=0.49



- DE/rand/1/bin效果优于DE/rand/1/L
- Ackley函数解的特征间没有耦合，可分问题

差分进化变种

- 尺度参数的调整
  - 差分进化的尺度参数F决定了差分向量对变异向量的影响
  - F为常数、F为随机变量
  - 改变尺度参数F的几种方式：

颤振(dither)：F仍为标量，在“for 每一个个体”的循环中每次都随机改变，

$$F \leftarrow U[F_{min}, F_{max}], \ \mathbf{v}_i \leftarrow \mathbf{x}_{r1} + F(\mathbf{x}_{r2}-\mathbf{x}_{r3})$$

尺度参数是一个均匀分布在F_{min}和F_{max}之间的随机标量，也可以服从高斯分布

差分进化变种

- 尺度参数的调整
  - 改变尺度参数F的几种方式：

抖动(jitter)： F为n元向量，在“for 每一个个体”的循环中随机改变它的每一个元素，

对每一维$j \in [1, n]$

$F_j \leftarrow U[F_{min}, F_{max}]$

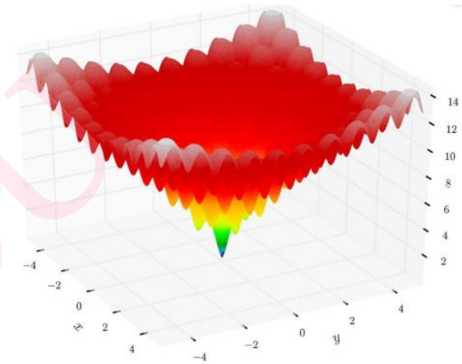
$\mathbf{v}_{ij} \leftarrow \mathbf{x}_{r1,j} + F_j(\mathbf{x}_{r2,j} - \mathbf{x}_{r3,j})$

下一维

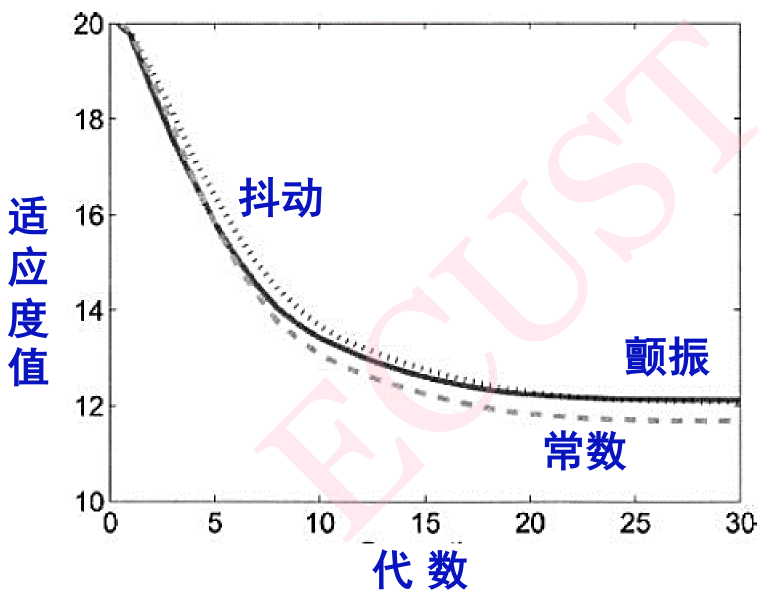
在生成变异向量时差分向量的每一个元素会以不同的量伸缩。

F取常数对简单问题比较管用；对于多峰问题，随机化F比较管用；对可分问题，抖动有效；对不可分问题，颤振有效。

差分进化变种

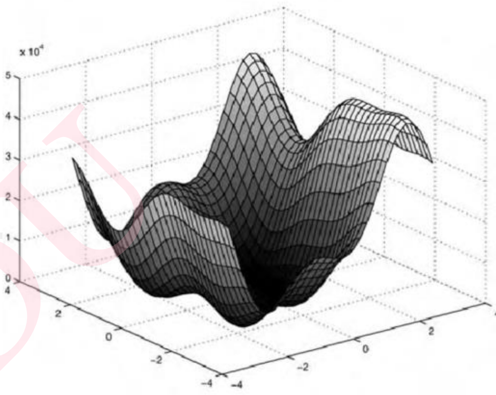


- 尺度参数的调整
  - 例：探索颤振和抖动的作用。应用于20维Ackley函数，种群规模为50，交叉率为0.9，F_{min}=0.2, F_{max}=0.6

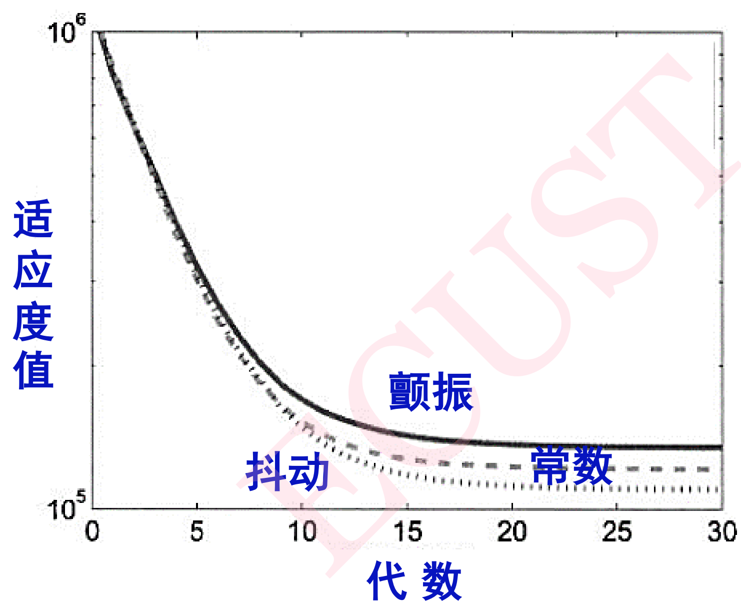


- 颤振和抖动的效果差不多，F为常数性能最好
- 随机化的F会让性能衰退

差分进化变种



- 尺度参数的调整
  - 例：探索颤振和抖动的作用。应用于20维Fletcher函数，种群规模为50，交叉率为0.9，F_{min}=0.2, F_{max}=0.6



- 抖动F表现稍好于常数F
- 常数F比颤振F稍好

颤振和抖动的性能取决于具体问题