DERIVATIVE FREE OPTIMIZATION

CLASS 7

2025/2026

Effect of population size λ.

Test on frastrigin – multimodal, exponential number (in the dim.) of local optima

$opt = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ $f(opt) = 0$

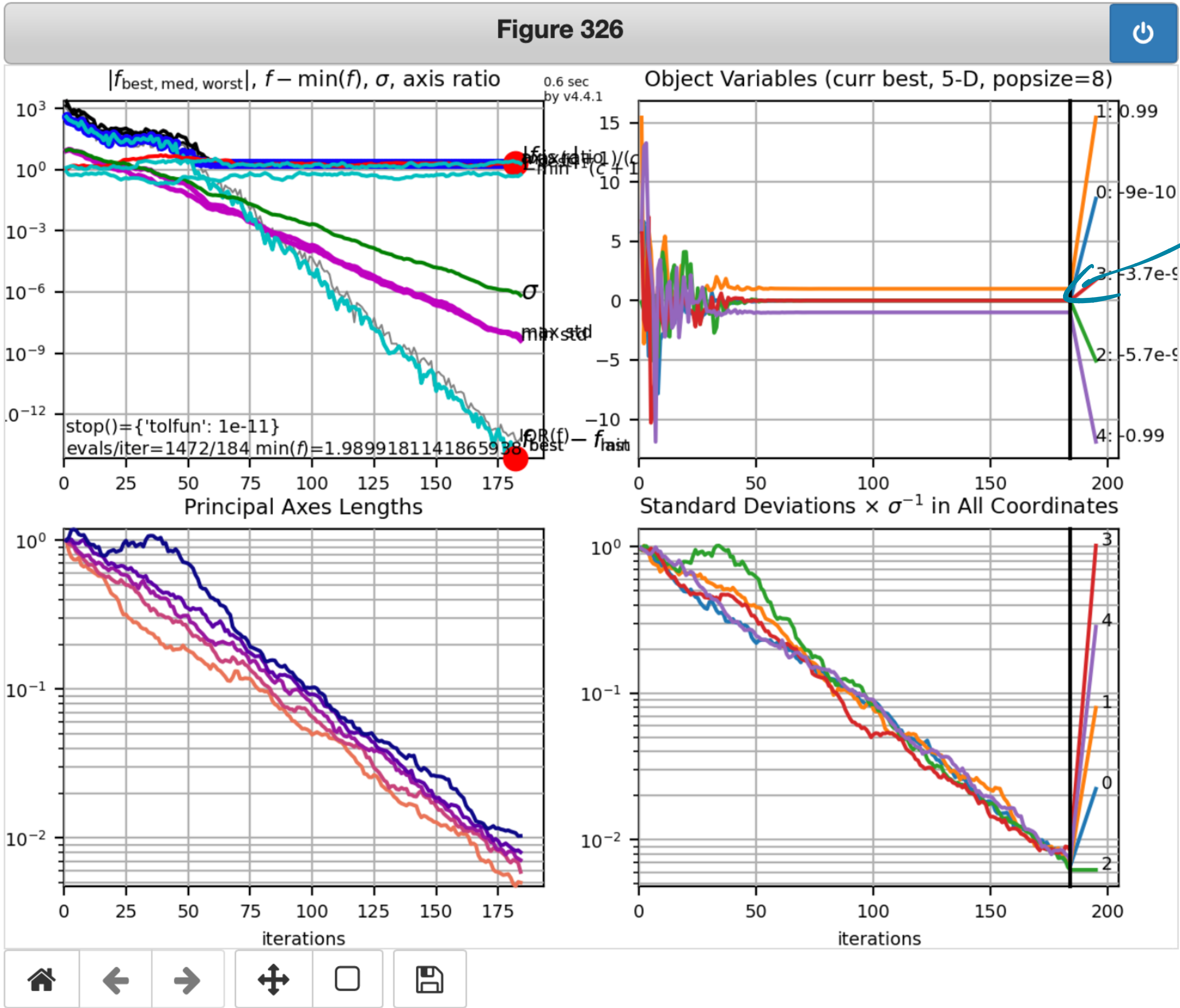
The default $\lambda = 4 + \lfloor 3 \ln(n) \rfloor$ dim of problem

In [6]:

```
n=5
default_popsize = 4+int(3*np.log(n))
res = cma.fmin(cma.ff.rastrigin, 10*np.ones(n), 10, {'popsize':1*default_popsize})
cma.plot()
```

```
(4_w,8)-aCMA-ES (mu_w=2.6,w_1=52%) in dimension 5 (seed=229793, Fri Jan 23 10:20:00 2026)
Iterat #Fevals   function value  axis ratio  sigma  min&max std  t[m:s]
    1      8 3.968707954379850e+02 1.0e+00 8.94e+00  9e+00  9e+00 0:00.0
    2     16 2.836001566652973e+02 1.2e+00 9.94e+00  9e+00  1e+01 0:00.0
    3     24 2.482790101213030e+02 1.5e+00 9.47e+00  9e+00  1e+01 0:00.0
  100    800 1.989926126312014e+00 2.1e+00 2.17e-03  1e-04  2e-04 0:00.1
  184   1472 1.989918114186594e+00 2.1e+00 6.82e-07  4e-09  6e-09 0:00.2
termination on {'tolfun': 1e-11} (Fri Jan 23 10:20:00 2026)
final/bestever f-value = 1.989918e+00 1.989918e+00 after 1473/1473 evaluations
incumbent solution: [ 7.87090245e-11, 9.94958637e-01, -2.48539450e-09, -3.91206338e-10, -9.94958638e-01]
std deviations: [4.51191323e-09, 4.73469851e-09, 4.25171675e-09, 5.73708814e-09, 4.90756493e-09]
```

std deviations: [4.31119152e-09, 4.75409851e-09, 4.23171075e-09, 3.75768814e-09, 4.90756495e-0



CONVERGENCE TO LOCAL MINIMUM

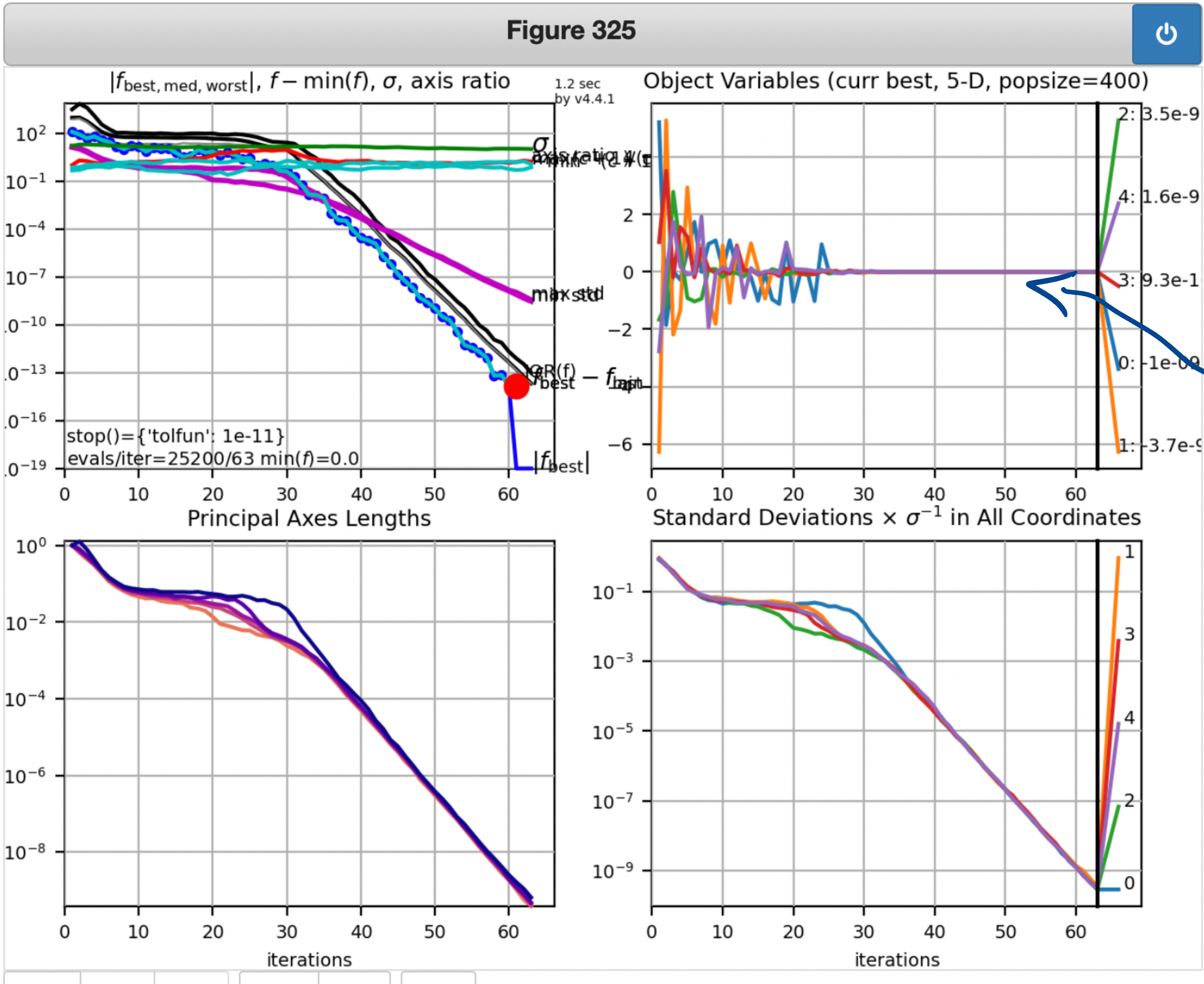
↳ The coordinates are not all in 0.

On several trials with default pop-size, we observed convergence to a local minimum.

We increased λ and with.

$$\lambda \leftarrow 50 \times \text{default pop-size}$$

we observe.



Here we converge to the global minimum

all coordinates are in 0.

CONCLUSION: When the function is multimodal, it can be usefull to increase the population size.

On Invariances

Invariance

The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms.
— Albert Einstein

- Empirical performance results
  - from benchmark functions
  - from solved real world problems

  are only useful if they do generalize to other problems

- Invariance is a strong non-empirical statement about generalization
  generalizing (identical) performance from a single function to a whole class of functions

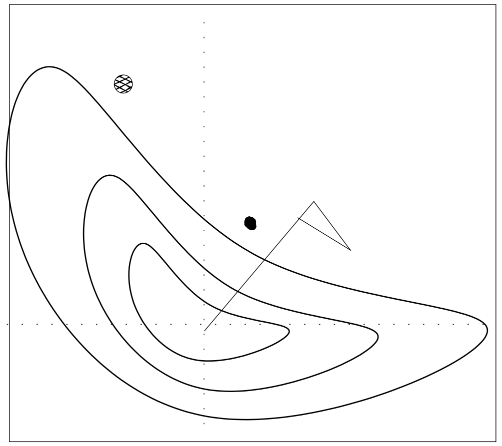
consequently, invariance is important for the evaluation of search algorithms

Rotational Invariance in Search Space

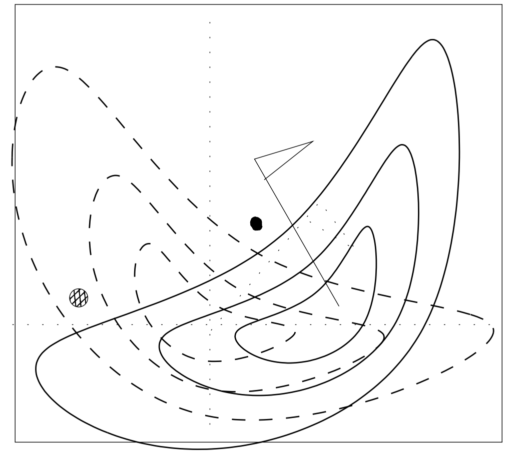
- invariance to orthogonal (rigid) transformations $\mathbf{R}$, where $\mathbf{R}\mathbf{R}^{\mathrm{T}} = \mathbf{I}$

e.g. true for simple evolution strategies

recombination operators might jeopardize rotational invariance



$$f(\boldsymbol{x}) \leftrightarrow f(\mathbf{R}\boldsymbol{x})$$



Identical behavior on f and $f_{\mathbf{R}}$

$$\begin{aligned} f: &\quad \boldsymbol{x} \mapsto f(\boldsymbol{x}), &\quad \boldsymbol{x}^{(t=0)} &= \boldsymbol{x}_0 \\ f_{\mathbf{R}}: &\quad \boldsymbol{x} \mapsto f(\mathbf{R}\boldsymbol{x}), &\quad \boldsymbol{x}^{(t=0)} &= \mathbf{R}^{-1}(\boldsymbol{x}_0) \end{aligned}$$

45

No difference can be observed w.r.t. the argument of f

[4] Salomon 1996. "Reevaluating Genetic Algorithm Performance under Coordinate Rotation of Benchmark Functions; A survey of some theoretical and practical aspects of genetic algorithms." BioSystems, 39(3):263-278

[5] Hansen 2000. Invariance, Self-Adaptation and Correlated Mutations in Evolution Strategies. *Parallel Problem Solving from Nature PPSN VI*

Main Invariances in Optimization

Invariance to strictly increasing transformations of f: identical behavior when optimizing

$x \mapsto f(x)$

$x \mapsto g(f(x))$ **where** $g : \mathrm{Im}(f) \to \mathbb{R}$ **is strictly increasing**

Translation invariance: identical behavior when optimizing

$x \mapsto f(x)$

$x \mapsto f(x - a)$ **for all** $a \in \mathbb{R}^n$

Scale invariance: identical behavior when optimizing

$x \mapsto f(x)$

$x \mapsto f(\alpha x)$ **for all** $\alpha \in \mathbb{R}_>$

Rotational invariance: identical behavior when optimizing

$x \mapsto f(x)$

$x \mapsto f(Rx)$ **for all** R **is an orthogonal matrix**

Affine invariance: identical behavior when optimizing

$x \mapsto f(x)$

$x \mapsto f(Ax + b)$ **for all** $A \in \mathbb{R}^{n \times n}$ **an invertible matrix and** $b \in \mathbb{R}^n$

Main Invariances in Optimization

Invariance to strictly increasing transformations of f: identical behavior when optimizing

$$x \mapsto f(x)$$

$x \mapsto g(f(x))$ **where** $g : \mathrm{Im}(f) \to \mathbb{R}$ **is strictly increasing**

Translation invariance: identical behavior when optimizing

$$x \mapsto f(x)$$

$x \mapsto f(x-a)$ **for all** $a \in \mathbb{R}^n$

Scale invariance: identical behavior when optimizing

$$x \mapsto f(x)$$

$x \mapsto f(\alpha x)$ **for all** $\alpha \in \mathbb{R}_>$

Rotational invariance: identical behavior when optimizing

$$x \mapsto f(x)$$

$x \mapsto f(Rx)$ **for all** R **an orthogonal matrix**

Affine invariance: identical behavior when optimizing

$$x \mapsto f(x)$$

$x \mapsto f(Ax+b)$ **for all** $A \in \mathbb{R}^{n\times n}$ **an invertible matrix and** $b \in \mathbb{R}^n$

provided initial state is change accordingly

Hierarchy of Invariance

Affine invariance

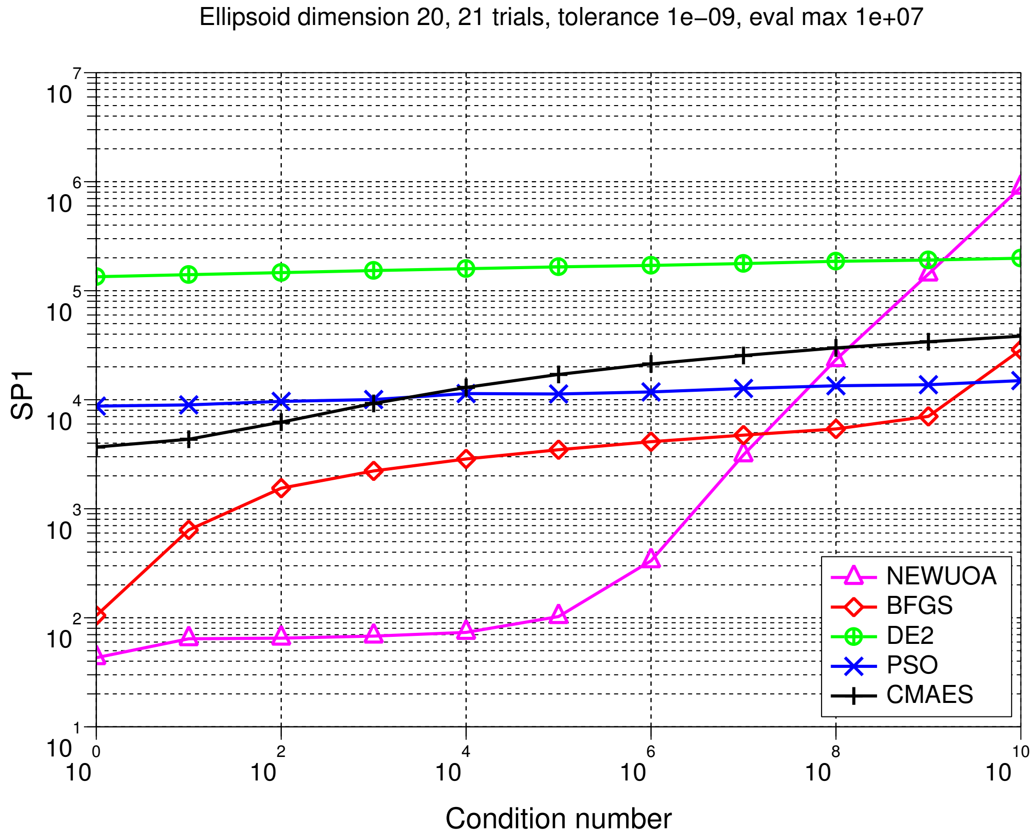
$\swarrow\!\!\swarrow$ $\Downarrow$ $\searrow\!\!\searrow$

Rotational Invariance **Scale-invariance** **translation invariance**

Testing for invariances

Comparison to BFGS, NEWUOA, PSO and DE

f convex quadratic, separable with varying condition number α



BFGS (Broyden et al 1970)
NEWUAO (Powell 2004)
DE (Storn & Price 1996)
PSO (Kennedy & Eberhart 1995)
CMA-ES (Hansen & Ostermeier 2001)

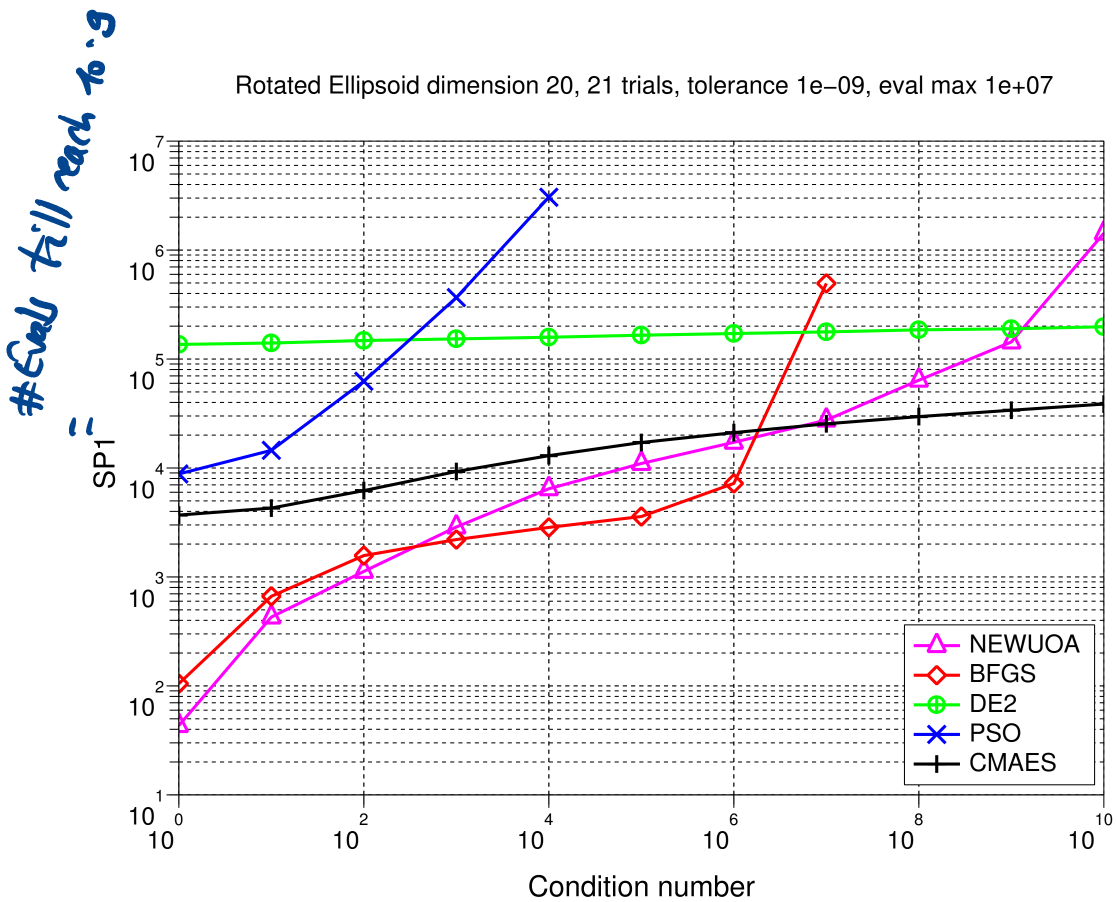
$f(\boldsymbol{x}) = g(\boldsymbol{x}^{\mathrm{T}}\mathbf{H}\boldsymbol{x})$ with
$\boldsymbol{H}$ diagonal
g identity (for BFGS and NEWUOA)
g any order-preserving = strictly increasing function (for all other)

SP1 = average number of objective function evaluations[5] to reach the target function value of $g^{-1}(10^{-9})$

[5] Auger et.al. (2009): Experimental comparisons of derivative free optimization algorithms, SEA

Comparison to BFGS, NEWUOA, PSO and DE

f convex quadratic, non-separable (rotated) with varying condition number α



BFGS (Broyden et al 1970)
NEWUAO (Powell 2004)
DE (Storn & Price 1996)
PSO (Kennedy & Eberhart 1995)
CMA-ES (Hansen & Ostermeier 2001)

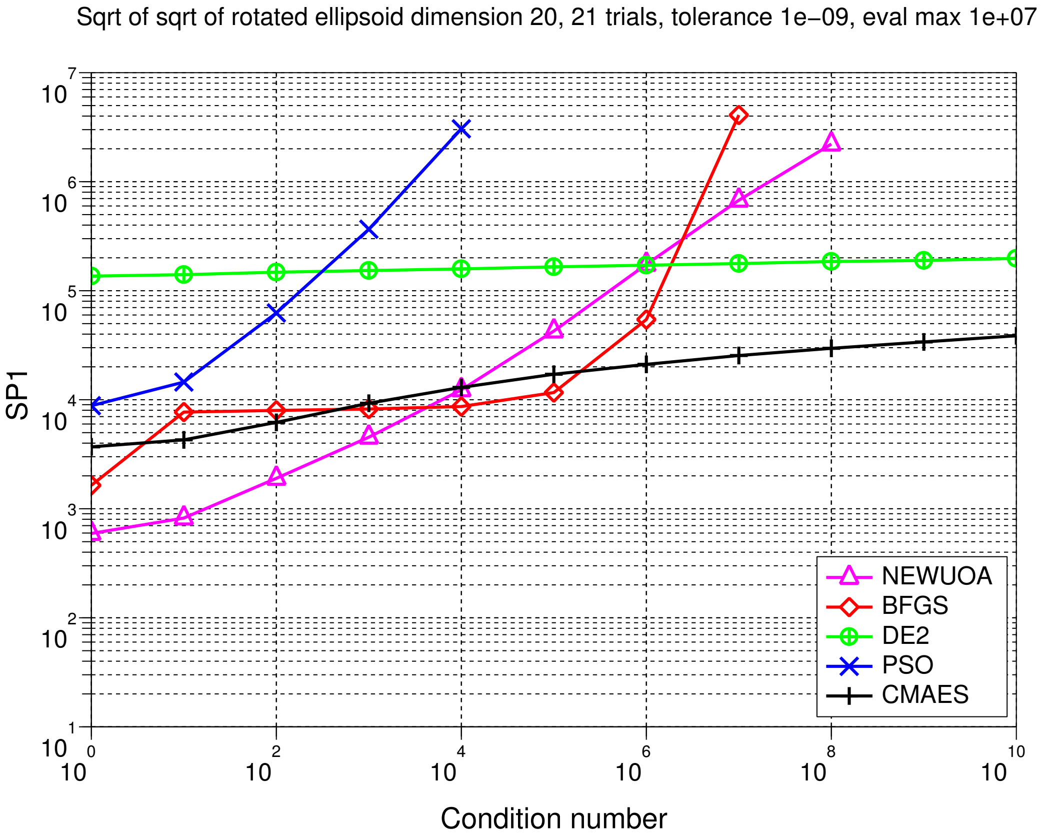
$f(\boldsymbol{x}) = g(\boldsymbol{x}^{\mathrm{T}}\mathbf{H}\boldsymbol{x})$ with
$\boldsymbol{H}$ full
g identity (for BFGS and NEWUOA)
g any order-preserving = strictly increasing function (for all other)

SP1 = average number of objective function evaluations[6] to reach the target function value of $g^{-1}(10^{-9})$

[6] Auger et.al. (2009): Experimental comparisons of derivative free optimization algorithms, SEA

Comparison to BFGS, NEWUOA, PSO and DE

f non-convex, non-separable (rotated) with varying condition number α



BFGS (Broyden et al 1970)
NEWUAO (Powell 2004)
DE (Storn & Price 1996)
PSO (Kennedy & Eberhart 1995)
CMA-ES (Hansen & Ostermeier 2001)

$f(\boldsymbol{x}) = g(\boldsymbol{x}^{\mathrm{T}}\mathbf{H}\boldsymbol{x})$ with
$\boldsymbol{H}$ full
$g : x \mapsto x^{1/4}$ (for BFGS and NEWUOA)
g any order-preserving = strictly increasing function (for all other)

SP1 = average number of objective function evaluations[7] to reach the target function value of $g^{-1}(10^{-9})$

[7] Auger et.al. (2009): Experimental comparisons of derivative free optimization algorithms, SEA