Evolution Strategies

Prof. Dr. Thomas Bäck
Managing Director, NuTech Solutions GmbH
Chief Scientist, NuTech Solutions, Inc.

Agenda

- Background
- EA Principles
- Evolution Strategies
- Applications
- Special Topics

Modeling - Simulation - Optimization

- Modeling / Data Mining

  ! ! ! → ??? → ! ! !

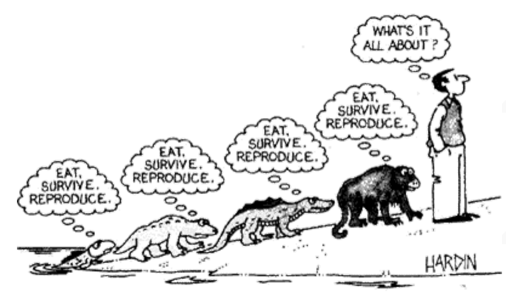
- Simulation

  ! ! ! → ! ! ! → ???

- Optimization

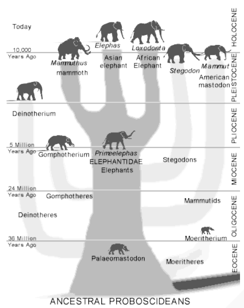
  ??? → ! ! ! → ! ! !

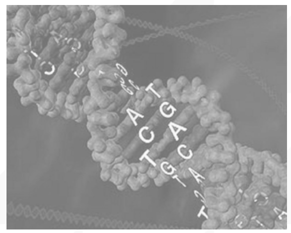
Background I



Daniel Dannett: Biology = Engineering

Background II

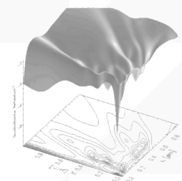




Phenotypic evolution ... and genotypic

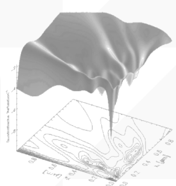
Optimization Problem

- $f: M \to \Re, f(\vec{x}) \to \min$
- f: Objective function, can be
  - Multimodal, with many local optima
  - Discontinuous
  - Stochastically perturbed
  - High-dimensional
  - Varying over time.
- $M \subseteq M_1 \times M_2 \times ... \times M_n$ can be heterogenous.
- Constraints can be defined over $M, f(\vec{x})$



Optimization Algorithms

- Direct optimization algorithm: Evolutionary Algorithms $f(\vec{x})$
- First order optimization algorithm: e.g, gradient method $f(\vec{x}), \nabla f(\vec{x})$
- Second order optimization algorithm: e.g., Newton method $f(\vec{x}), \nabla f(\vec{x}), \nabla^2 f(\vec{x})$



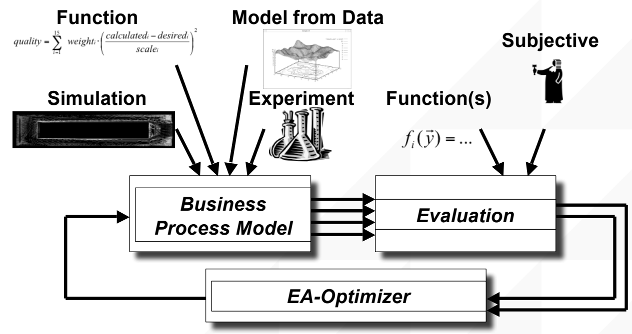
Business Issues

- Supply Chain Optimization
- Scheduling & Timetabling
- Product Development, R&D
- Management Decision Making, e.g., project portfolio optimization
- Optimization of Marketing Strategies; Channel allocation
- Multicriteria Optimization (cost / quality)
- ... And many others

Evolutionary Algorithm Applications

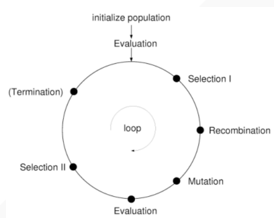
Principles

General Aspects

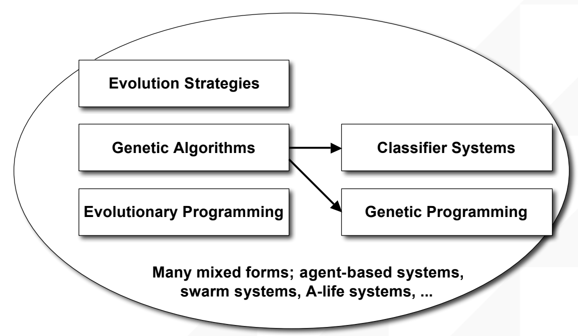


Unifying Evolutionary Algorithm

t := 0;
initialize(P(t));
evaluate(P(t));
while not *terminate* **do**
  P'(t) := *mating_selection*(P(t));
  P''(t) := *variation*(P'(t));
  evaluate(P''(t));
  P(t+1) := *environmental_selection*(P''(t) u Q);
  t := t+1;
od



Evolutionary Algorithm Taxonomy

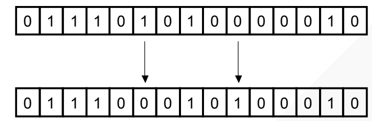


Genetic Algorithms vs. Evolution Strategies

- Genetic Algorithm
  - Binary representation
  - Fixed mutation rate p_m (= 1/n)
  - Fixed crossover rate p_c
  - Probabilistic selection
  - Identical population size
  - No self-adaptation
- Evolution Strategies
  - Real-valued representation
  - Normally distributed mutations
  - Fixed recombination rate (= 1)
  - Deterministic selection
  - Creation of offspring surplus
  - Self-adaptation of strategy parameters: Variance(s), Covariances

NuTech Solutions science for business

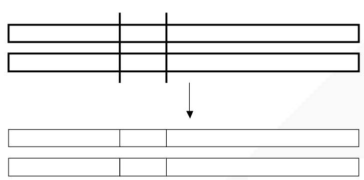
Genetic Algorithms: Mutation



- Mutation by bit inversion with probability pm.
- pm identical for all bits.
- pm small (e.g., pm = 1/n).

NuTech Solutions science for business

Genetic Algorithms: Crossover



- Crossover applied with probability pc.
- pc identical for all individuals.
- k-point crossover: k points chosen randomly.
- Example: 2-point crossover.

NuTech Solutions science for business

Genetic Algorithms: Selection

- Fitness proportional:
  - f fitness
  - λ population size

$$p_i = \frac{f(\vec{a}_i)}{\sum_{j=1}^{\lambda} f(\vec{a}_j)}$$

- Tournament selection:
  - Randomly select $q << \lambda$ individuals.
  - Copy best of these q into next generation.
  - Repeat λ times.
  - q is the tournament size (often: $q = 2$).

NuTech Solutions science for business

Evolution Strategies

An instance of evolutionary algorithms

Advantages of Evolution Strategies

- Self-Adaptation of strategy parameters.
- Direct, global optimizers !
- Extremely good in solution quality.
- Very small number of function evaluations.
- Dynamical optimization problems.
- Design optimization problems.
- Discrete or mixed-integer problems.
- Experimental design optimisation.
- Combination with Meta-Modeling techniques.

Evolution Strategies

- Real-valued / discrete / mixed-integer search spaces.
- Emphasis on mutation: n-dimensional, normally distributed, expectation zero.
- Different recombination operators.
- Deterministic selection: (μ,λ) , $(\mu+\lambda)$
- Self-adaptation of strategy parameters.
- Creation of offspring surplus, i.e., $\lambda >> \mu$.

Mutation

- **Creation of a new solution:**
  $x_i' = x_i + \sigma_i' \cdot N_i(0,1)$
- σ-adaptation by means of
  - 1/5-success rule.
  - **Self-adaptation.**
- More complex / powerful strategies:
  - Individual step sizes σ_i.
  - Covariances.
- Convergence speed:
  $\Rightarrow$ Ca. 10 · n down to 5 · n is possible.

Self-Adaptation

- Motivation: General search algorithm

$$\vec{x}_{t+1} = \vec{x}_t + s_t \cdot \vec{v}_t$$

Step size

Direction

- Geometric convergence: Arbitrarily slow, if s wrongly controlled !
- No deterministic / adaptive scheme for arbitrary functions exists.
- Self-adaptation: On-line evolution of strategy parameters.
- Various schemes:
  - Schwefel one σ, n σ, covariances; Rechenberg MSA.
  - Ostermeier, Hansen: Derandomized, Covariance Matrix Adaptation.
  - EP variants (meta EP, Rmeta EP).
  - Bäck: Application to p in GAs.

Self-Adaptation

- Learning while searching: Intelligent Method.
- Different algorithmic approaches, e.g:
  - Pure self-adaptation:

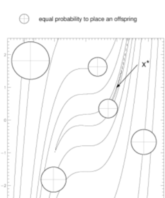
$$\sigma_i' = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1))$$
$$x_i' = x_i + \sigma_i' \cdot N_i(0,1)$$

  - Mutational step size control MSC:

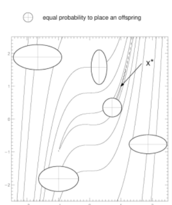
$$\sigma' = \begin{cases} \sigma \cdot \alpha\ , & \text{if } u \approx U(0,1) \leq 1/2 \\ \sigma / \alpha\ , & \text{if } u \approx U(0,1) > 1/2 \end{cases}$$
$$x_i' = x_i + \sigma_i' \cdot N_i(0,1)$$

  - Derandomized step size adaptation
  - Covariance adaptation

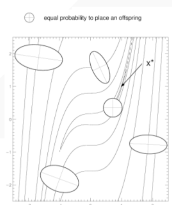
Self-Adaptive Mutation



n = 2, n_σ = 1, n_α = 0

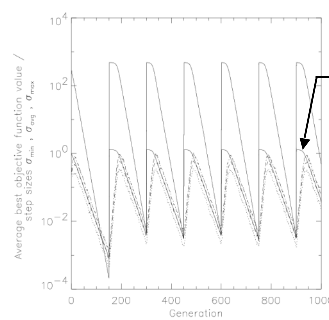


n = 2, n_σ = 2, n_α = 0



n = 2, n_σ = 2, n_α = 1

Self-Adaptation: Dynamic Sphere



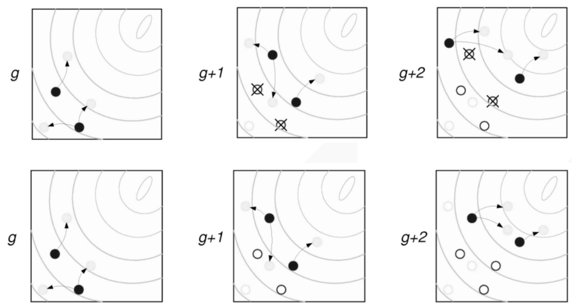
- Optimum σ:

$$\sigma_{opt} = c_{\mu,\lambda} \frac{R}{n}$$

- Transition time proportionate to n.
- Optimum σ learned by self-adaptation.

Selection

(μ,λ)

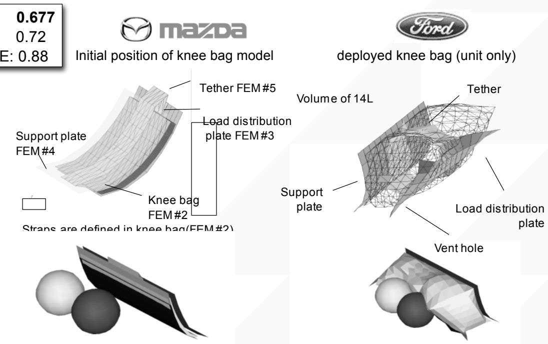


Possible Selection Operators

- (1+1)-strategy: one parent, one offspring.
- (1,λ)-strategies: one parent, λ offspring.
  - Example: (1,10)-strategy.
  - Derandomized / self-adaptive / mutative step size control.
- (μ,λ)-strategies: μ>1 parents, λ>μ offspring
  - Example: (2,15)-strategy.
  - Includes recombination.
  - Can overcome local optima.
- ($\mu+\lambda$)-strategies: elitist strategies.

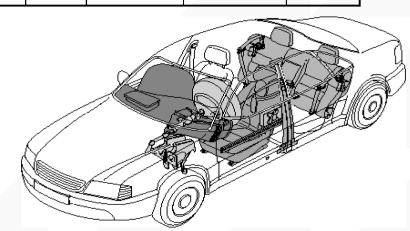
Examples I: Inflatable Knee Bolster Optimization

Low Cost ES:	**0.677**
GA (Ford):	0.72
Hooke Jeeves DoE:	0.88



IKB: Previous Designs

# Variables	Characteristics	HIC	CG	Left foot load	Right foot load	$P_{Combined}$
4	Unconstrained	576,324	44,880	4935	3504	12,393
5	Unconstrained	384,389	41,460	4707	4704	8,758
9	Unconstrained	292,354	38,298	5573	5498	6,951
10	Constrained	305,900	39,042	6815	6850	7,289

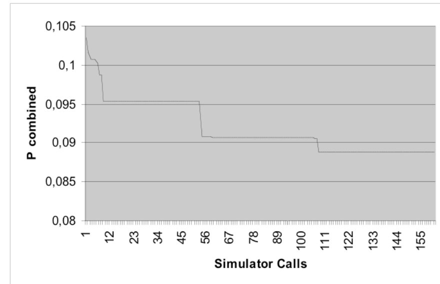


IKB: Problem Statement

- Objective: Min Ptotal   Subject to: Left   Femur load <= 7000
  Right   Femur load <= 7000

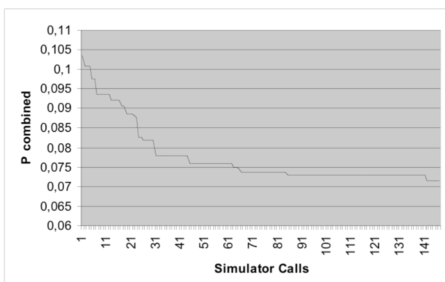
Design Variable	Description	Base Design 1	Base Design 2	GA (Yan Fu)
dx	IKB center offset x	0	0	0,01
dz	IKB center offset y	0	0	-0,01
rcdex	KB venting area ratio	1	1	2
massrat	KB mass inflow ratio	1	1	1,5
rcdexd	DB venting area ratio	1	1	2,5
Dmassratf	DB high output mass inflow ratio	1	1	1,1
Dmassratl	DB low output mass inflow ratio	1	1	1
dbfire	DB firing time	0	0	-0,003
dstraprat	DB strap length ratio	1	1	1,5
emr	Load of load limiter (N)	3000	3000	2000
Performance Response	**Description**			
NCAP_HIC_50	HIC	590	555.711	305,9
NCAP_CG_50	CG	47	47.133	39,04
NCAP_FMLL_50	Left foot load	760	6079	6815
NCAP_FMRL_50	Right foot load	900	5766	6850
P combined (Quality)		13.693	13.276	7,289

IKB Results I: Hooke-Jeeves



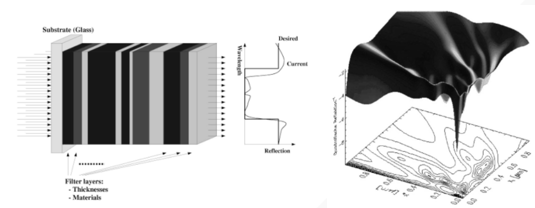
Quality: 8.888   Simulations: 160

IKB Results II: (1+1)-ES



Quality: 7.142   Simulations: 122

Optical Coatings: Design Optimization



- Nonlinear mixed-integer problem, variable dimensionality.
- Minimize deviation from desired reflection behaviour.
- Excellent synthesis method; robust and reliable results.

Dielectric Filter Design Problem

Client:

Corning, Inc.,
Corning, NY

- Dielectric filter design.
- n=40 layers assumed.
- Layer thicknesses xi in [0.01, 10.0].
- Quality function: Sum of quadratic penalty terms

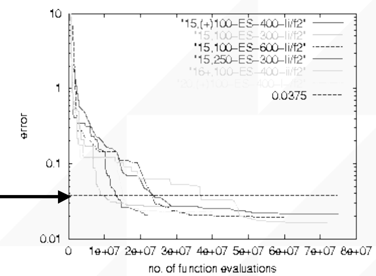
$$quality = \sum_{i=1}^{15} weight_i \cdot \left(\frac{calculated_i - desired_i}{scale_i} \right)^2 \rightarrow \min$$

- Penalty terms = 0 iff constraints satisfied.

33

Results: Overview of Runs

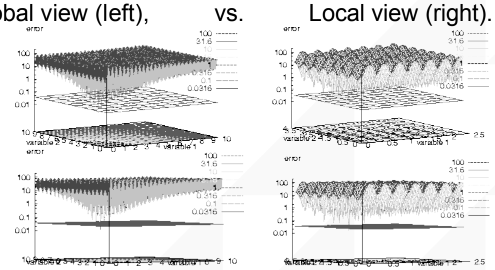
- Factor 2 in quality.
- Factor 10 in effort.
- Reliable, repeatable results.



34

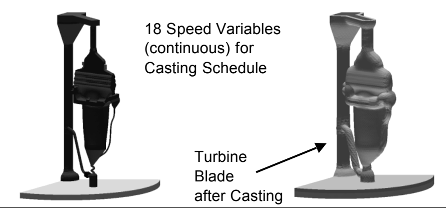
Problem Topology Analysis: An Attempt

- Grid evaluation for 2 variables.
- Close to the optimum (from vector of quality 0.0199).
- Global view (left), vs. Local view (right).



35

Examples II: Bridgman Casting Process

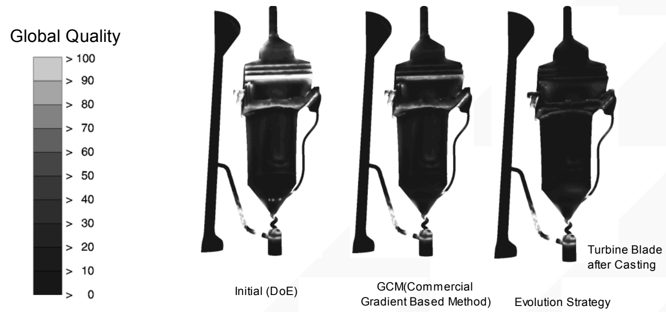


- FE mesh of 1/3 geometry: 98.610 nodes, 357.300 tetrahedrons, 92.830 radiation surfaces

  ⇒ large problem:
  - run time varies: 16 h 30 min to 32 h (SGI, Origin, R12000, 400 MHz)
  - at each run: 38,3 GB of view factors (49 positions) are treated!

36

Examples II:
Bridgman Casting Process



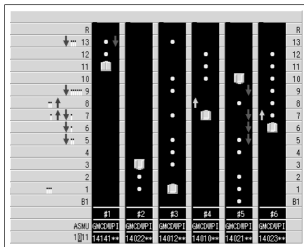
Quality Comparison of the Initial and Optimized Configurations

Examples IV:
Traffic Light Control

- Generates green times for next switching schedule.
- Minimization of total delay / number of stops.
- Better results (3 – 5%) / higher flexibility than with traditional controllers.
- Dynamic optimization, depending on actual traffic (measured by control loops).

- **Client:**
  Dutch Ministry of Traffic
  Rotterdam, NL

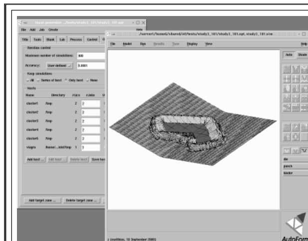
Examples V:
Elevator Control



- Minimization of passenger waiting times.
- Better results (3 – 5%) / higher flexibility than with traditional controllers.
- Dynamic optimization, depending on actual traffic.

- **Client:**
  Fujitec Co. Ltd., Osaka, Japan

Examples VI:
Metal Stamping Process

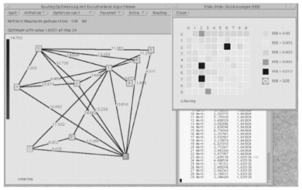


- Minimization of defects in the produced parts.
- Optimization on geometric parameters and forces.
- Fast algorithm; finds very good results.

- **Client:**
  AutoForm Engineering GmbH,
  Dortmund

Examples VII: Network Routing

SIEMENS

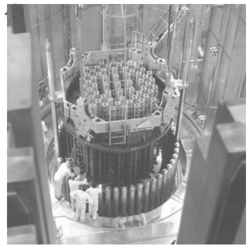


- Minimization of end-to-end-blockings under service constraints.
- Optimization of routing tables for existing, hard-wired networks.
- 10%-1000% improvement.

- **Client:**
  SIEMENS AG, München

Examples VIII: Nuclear Reactor Refueling

SIEMENS



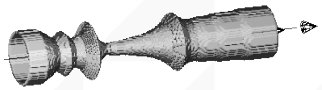
- Minimization of total costs.
- Creates new fuel assembly reload patterns.
- Clear improvements (1%-5%) of existing expert solutions.
- Huge cost saving.

- **Client:**
  SIEMENS AG, München

Two-Phase Nozzle Design (Experimental)

- Experimental design optimisation: Optimise efficieny.

- ...evolves...



- Final design: 32% improvement in efficiency.

Advantages of Evolution Strategies

- Self-Adaptation of strategy parameters.
- Direct, global optimizers !
- Extremely good in solution quality.
- Very small number of function evaluations.
- Dynamical optimization problems.
- Design optimization problems.
- Discrete or mixed-integer problems.
- Experimental design optimisation.
- Combination with Meta-Modeling techniques.

Multi Criteria Optimization (1)

- Most Problems: More than one aspect to optimise.
- Conflicting Criteria !
- Classical optimization techniques map multiple criteria to one single value, e.g. by weighted sum:

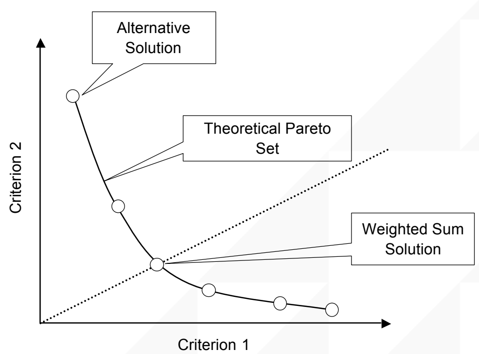
$$f(x) = \sum_i w_i f_i(x)$$

- But: How can optimal weights be determined?
- Evolution Strategies can directly use the concept of Pareto Dominance

Multi Criteria Optimization (2)

- Multi Criteria Optimization does not mean:
  - Decide on „What is a good compromise“ before optimization (e.g. by choosing weighting factors).
  - Find one single optimal solution.
- Multi Criteria Optimization means:
  - Decide on a compromise after optimization.
  - Find a set of multiple compromise solutions.
- Evolutionary Multi Criteria Optimization means:
  - Use the population structure to represent the set of multiple compromise solutions.
  - Use the concept of Pareto Dominance

Multi Criteria Optimization (3)



Pareto Dominance

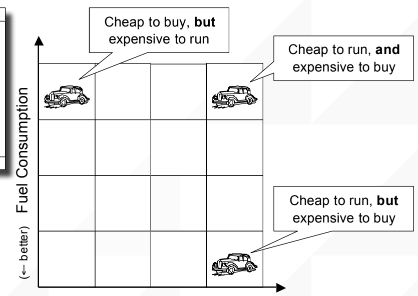
- **Assume two design solutions a and b with**
  $F(a) = (f_1(a),...,f_k(a))$ and $F(b) = (f_1(b),...,f_k(b))$
- If all $f_i(a)$ are better than $f_i(b)$, then a dominates b.
- If all $f_i(b)$ are better than $f_i(a)$, then b dominates a.
- If there are i and j, such that
  - $f_i(a)$ is better than $f_i(b)$, but
  - $f_j(b)$ is better than $f_j(a)$, then
- a and b do not dominate each other („are equal“, „are incomparable“)

Pareto Dominance (2)

Two Criteria Example:
The economic car

1. **Minimize initial costs**
2. **Minimize long term costs**

⇒ Never choose the red car!

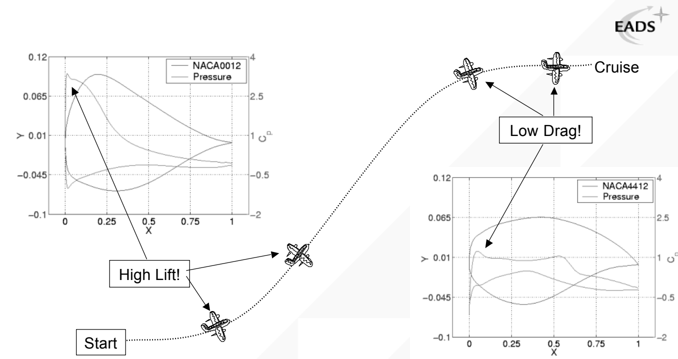


NuTech Solutions science for business

49

Multipoint Airfoil Optimization (1)

EADS



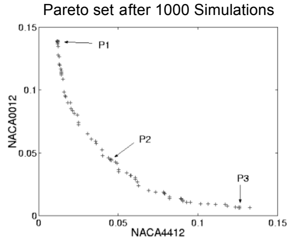
NuTech Solutions science for business

50

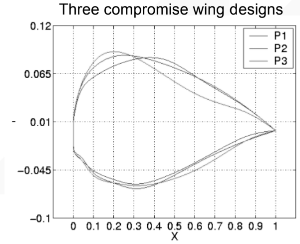
Multipoint Airfoil Optimization (2)

Find **pressure** profiles that are a **compromise** between two given **target pressure** distributions under two given flow conditions!

Pareto set after 1000 Simulations



Three compromise wing designs



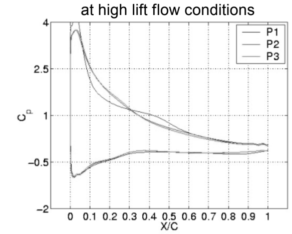
NuTech Solutions science for business

51

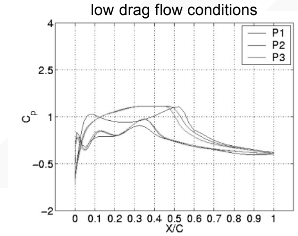
Multipoint Airfoil Optimization (3)

Find **pressure** profiles that are a **compromise** between two given **target pressure** distributions under two given flow conditions!

Pressure profile at high lift flow conditions



Pressure profile at low drag flow conditions



NuTech Solutions science for business

52

Noisy Fitness Functions: Thresholding

- Fitness evaluation is disturbed by noise, e.g.: stochastic distribution of passengers within an elevator system.
- Traffic control problems in general.
- Probability of generating a real improvement is very small.
- Introduce explicit barrier into the (1+1)-ES to distinguish real improvements from overvalued individuals:

> *Only accept offspring if it outperforms the parent by at least a value of τ (threshold).*

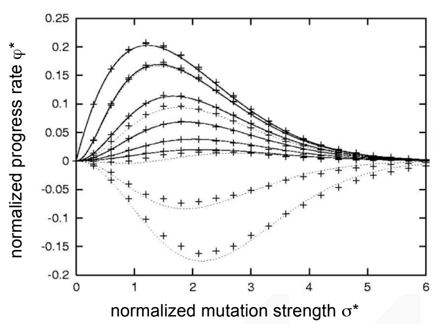
Finding the Optimal Threshold

- For Gaussian noise $\varepsilon \approx N(0, \sigma_\varepsilon^2)$
  - General optimal threshold:

    $\hat{\tau}^* = \sigma_\varepsilon^{*2}$
  - For the sphere model $Q(R) = Q_0 + cR^\alpha$ (where R is the distance to the optimum):

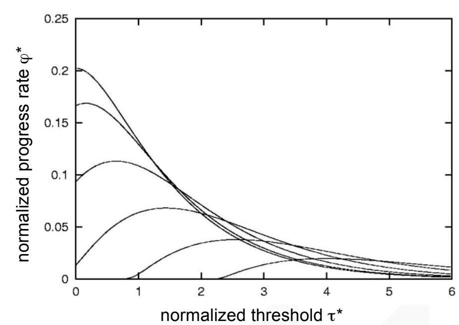
$$\hat{\tau} = \frac{\sigma_\varepsilon^2 N}{\alpha(Q - Q_0)}$$

Influence of Thresholding (I)



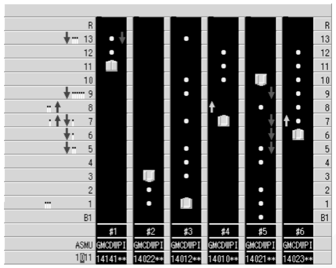
solid lines:
opt. threshold
$\hat{\tau}^* = \sigma_\varepsilon^{*2}$
dashed lines:
$\tau^* = 0$
crosses:
data measured in ES-runs (sphere)
noise strength (from top to bottom)
$\sigma_\varepsilon^* = 0.0, 0.4, 0.8, 1.2, 1.6, 2.0$

Influence of Thresholding (II)



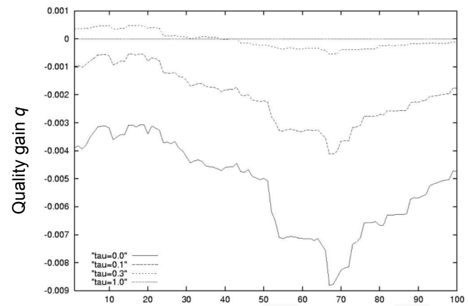
noise strength (from top to bottom)
$\sigma_\varepsilon^* = 0.0, 0.4, 0.8, 1.2, 1.6, 2.0$

Applications: elevator control



- Simulation of an elevator group controller takes a long time
- Instead use artificial problem tightly related to the real-world problem: S-Ring

Application in Elevator Controller: S-Ring



distance to optimum *R*: 0 = greedy, 100 = optimum

- Only thresholding leads to a positive quality gain ($\tau = 0.3$)
- A too large threshold value does not permit any progress ($\tau = 1.0$)