Surrogate Model Selection for Evolutionary Optimization Using Landscape Analysis

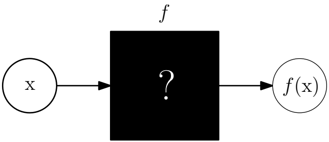
Zbyněk Pitra[1,2], Lukáš Bajer[1,3], Jakub Repický[1,4] Martin Holeňa[1]

[1]Institute of Computer Science, Czech Academy of Sciences
[2]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University
[3]Cisco Systems
[4]Faculty of Mathematics and Physics, Charles University in Prague

Prague, Czech Republic

2019

Continuous Black-box Optimization

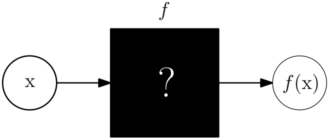


- objective function evaluated empirically or through simulations
- optimization (minimization) is finding such $\mathbf{x}^\star \in \mathbb{R}^n$ that

$$f(\mathbf{x}^\star) = \min_{\forall \mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

- expensive scenario – limited number of evaluations

Continuous Black-box Optimization

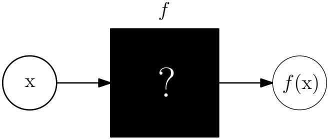


- objective function evaluated empirically or through simulations
- optimization (minimization) is finding such $\mathbf{x}^\star \in \mathbb{R}^n$ that

$$f(\mathbf{x}^\star) = \min_{\forall \mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

- expensive scenario – limited number of evaluations

Continuous Black-box Optimization



- objective function evaluated empirically or through simulations
- optimization (minimization) is finding such $\mathbf{x}^\star \in \mathbb{R}^n$ that

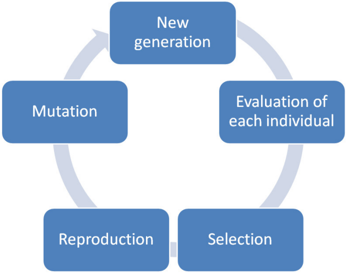
$$f(\mathbf{x}^\star) = \min_{\forall \mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

- expensive scenario – limited number of evaluations

Evolutionary Algorithms and Surrogate Modeling

Evolutionary Algorithms

- escape from local optima
- require many function evaluations

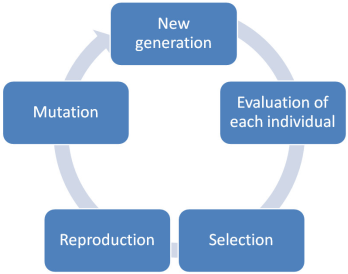


sicara.ai

EVOLUTIONARY ALGORITHMS AND SURROGATE MODELING

Evolutionary Algorithms

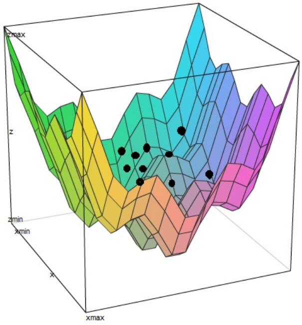
- escape from local optima
- require many function evaluations



sicara.ai

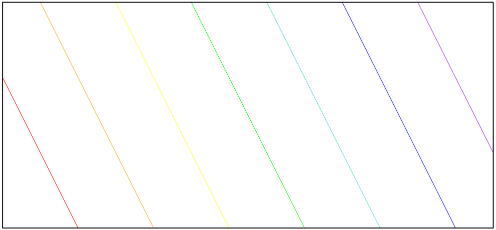
Surrogate Modeling

- approximating regression model
- not expensive
- less accurate



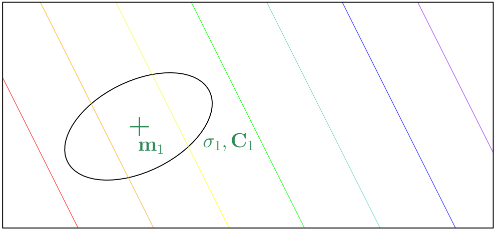
CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \dots w_\lambda$ appropriately



CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

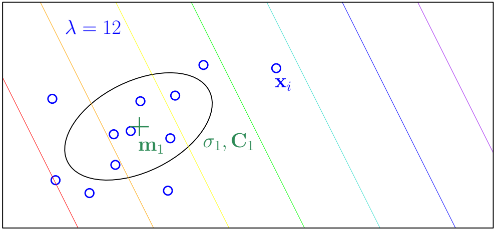


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}),$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

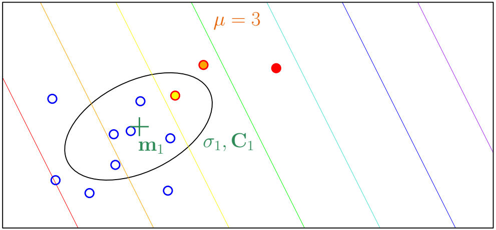


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

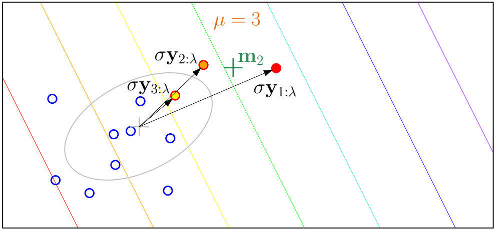


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w \quad$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

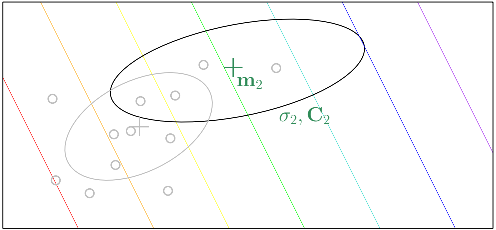


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ — sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w \quad$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ — update mean
3. update $\mathbf{C}$
4. update step-size σ

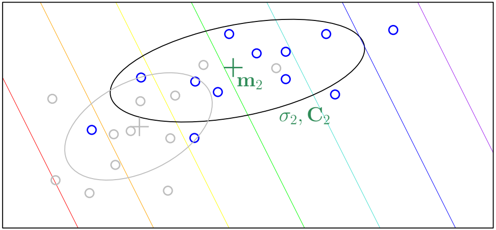


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w \quad$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

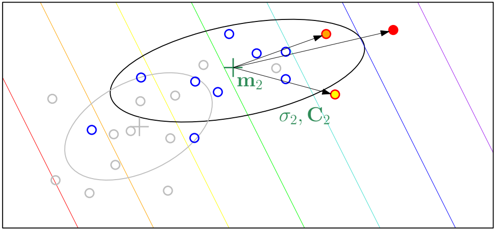


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w \quad$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

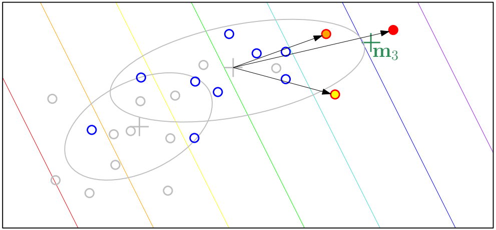


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w \quad \text{where } \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ

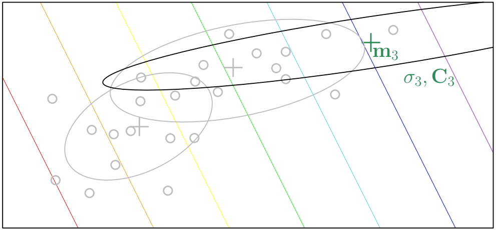


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

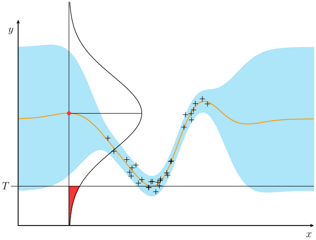
1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ sampling
2. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w$ where $\mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ update mean
3. update $\mathbf{C}$
4. update step-size σ



Gaussian Processes

A collection of random variables, any finite subset of which have a joint Gaussian distribution.

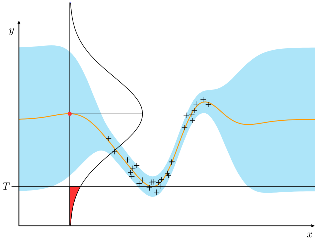
- specified by a **mean function** and a **covariance function**
- prediction in a point given as a **univariate** Gaussian



Gaussian Processes

A collection of random variables, any finite subset of which have a joint Gaussian distribution.

- specified by a **mean function** and a **covariance function**
- prediction in a point given as a **univariate** Gaussian

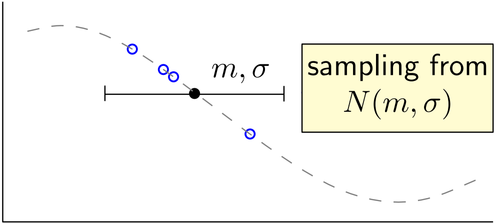


CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the original fitness
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

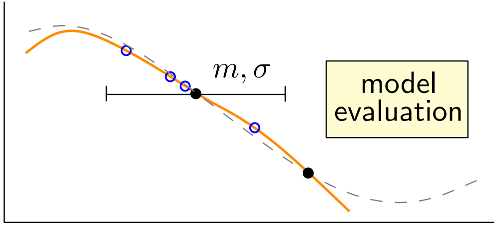


Evolution Control in the CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ {sampling}
2. $\mathbf{y}_i \leftarrow \text{evolutionControl}(\mathbf{x}_i)$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

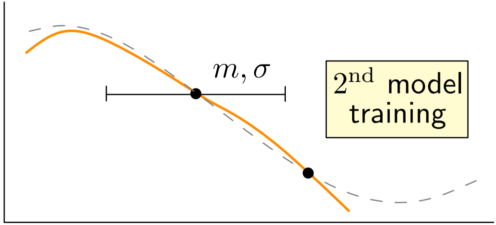


Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the original fitness f & build a model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

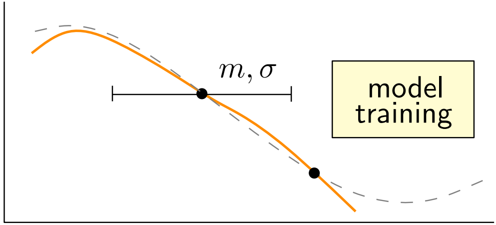


Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

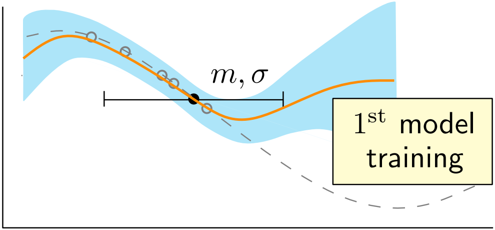


Doubly Trained Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ {sampling}
2. build a model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

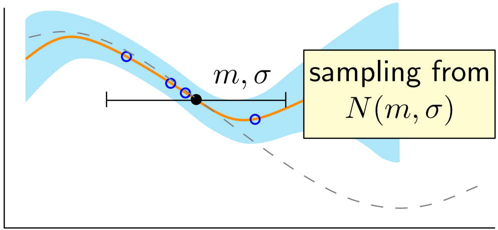


Doubly Trained Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ {sampling}
2. build a model $f_\mathcal{M}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

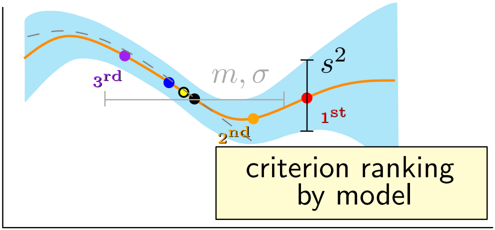


Doubly Trained Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

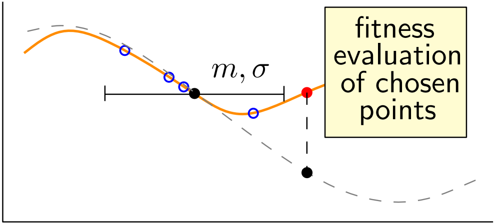


Doubly Trained Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \dots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \dots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the original fitness f & build a model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

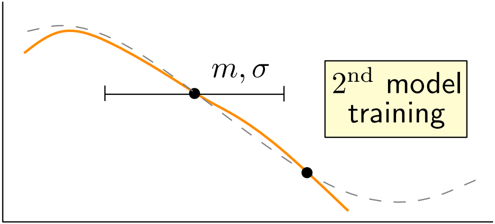


Doubly Trained Surrogate CMA-ES

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad$ for $i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the original fitness f & build a model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

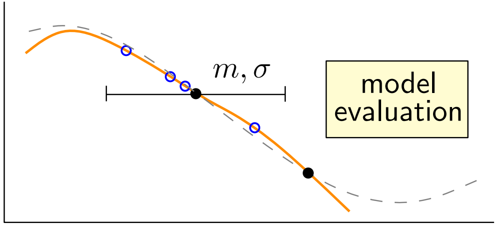


Doubly Trained Surrogate CMA-ES

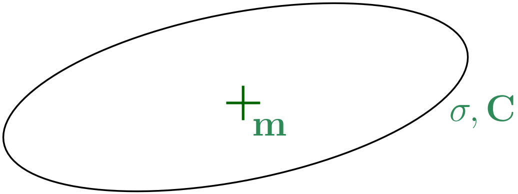
Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+, \lambda \in \mathbb{N}$
Initialize: $\mathbf{C} = \mathbf{I}$ (and several other parameters)
Set the weights $w_1, \ldots w_\lambda$ appropriately

while not terminate

1. $\mathbf{x}_i = \mathbf{m} + \sigma \mathbf{y}_i, \quad \mathbf{y}_i \sim N(\mathbf{0}, \mathbf{C}), \quad \text{for } i = 1, \ldots, \lambda$ {sampling}
2. evaluate $\mathbf{x}_i$ with the model $f_{\mathcal{M}}$
3. $\mathbf{m} \leftarrow \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda} = \mathbf{m} + \sigma \mathbf{y}_w, \quad \mathbf{y}_w = \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda}$ {update mean}
4. update step-size σ
5. update $\mathbf{C}$

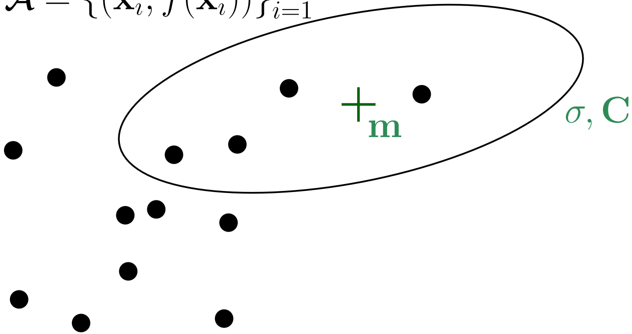


SURROGATE MODEL SELECTION PROBLEM

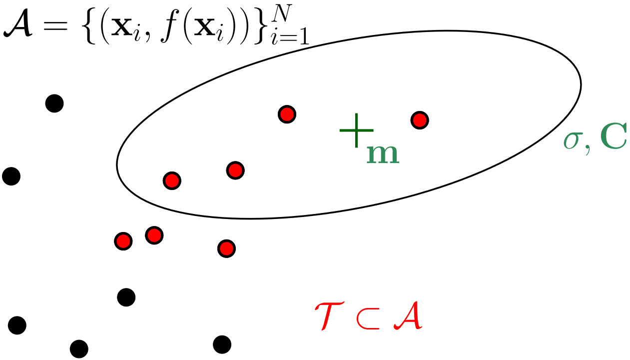


Surrogate Model Selection problem

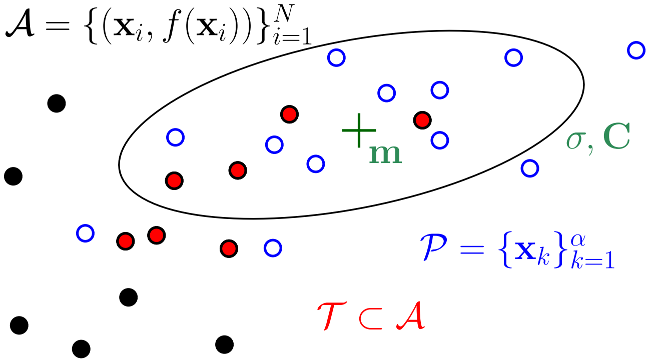
$\mathcal{A} = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^{N}$



SURROGATE MODEL SELECTION PROBLEM



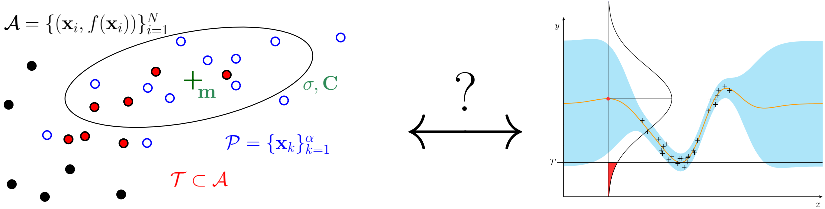
Surrogate Model Selection problem



RESEARCH TASK

Question

What relationships are between the suitability of GPs with different covariances and the properties of training data sampled from the optimized function?



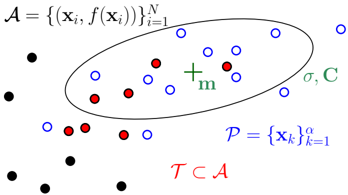
Research task

Metalearning of Optimization Algorithms

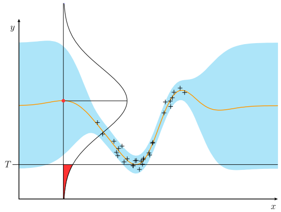
Metalearning of Classification

↓

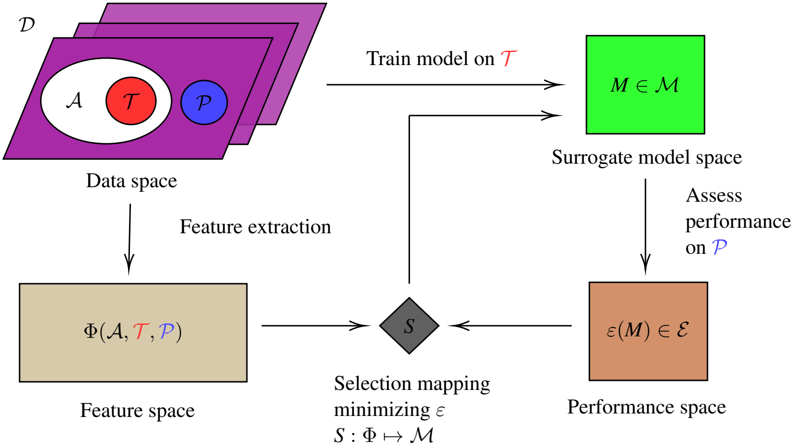
Metalearning of Regression Models



$\overset{?}{\longleftrightarrow}$



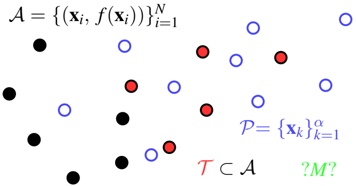
SURROGATE MODEL SELECTION SYSTEM

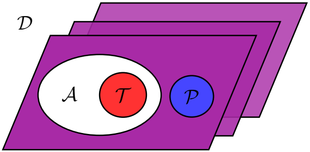


EXPERIMENTAL SETTINGS

DATASET

- Snapshots from independent runs of the DTS-CMA-ES
  - 24 noiseless benchmark functions
  - 5 dimensions
  - 5 instances
  - 8 covariance functions
  - 25 generations
- 120 000 data





Experimental settings

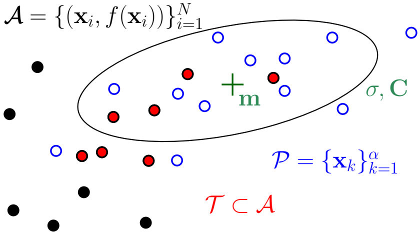
Dataset - Sample Sets

$$\mathcal{S} = \left\{ (\mathbf{x}_i, y_i) \in \mathbb{R}^D \times \mathbb{R} \,\middle|\, i = 1, \ldots, N \right\}$$

Archive $\mathcal{A}$

Training set $\mathcal{T}$

Training + Population set $\mathcal{T}_{\mathcal{P}}$



Experimental settings

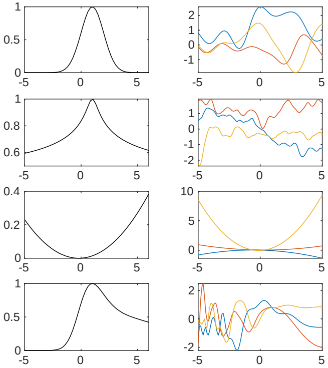
Model space ∼ Covariance functions

squared-exponential (SE)

rational quadratic (RQ)

quadratic (Q)

SE with variable length-scale (Gibbs)



Experimental settings

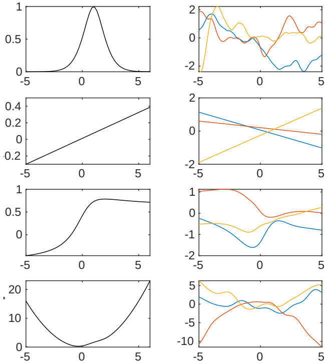
Model space ∼ Covariance functions

Matérn class (Mat)

linear (LIN)

neural network (NN)

squared exponential + quadratic (SE+Q)



Experimental settings

Feature space ∼ Exploratory landscape features

$$\varphi : \bigcup_{N \in \mathbb{N}} \mathbb{R}^{N,D} \times \mathbb{R}^{N,1} \mapsto \mathbb{R}$$

- Distribution
- Levelset
- Meta-Model
- Nearest better clustering (NBC)
- Dispersion
- Information content
- *Dimension*
- *Number of observations*

New CMA-ES features

Experimental settings

Feature space ∼ Exploratory landscape features

$$\varphi : \bigcup_{N \in \mathbb{N}} \mathbb{R}^{N,D} \times \mathbb{R}^{N,1} \mapsto \mathbb{R}$$

- Distribution
- Levelset
- Meta-Model
- Nearest better clustering (NBC)
- Dispersion
- Information content
- *Dimension*
- *Number of observations*

New CMA-ES features

Experimental settings

CMA-ES features

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

Experimental settings

CMA-ES features

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_r$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

EXPERIMENTAL SETTINGS

CMA-ES FEATURES

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

Experimental settings

CMA-ES features

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

Experimental settings

CMA-ES features

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

EXPERIMENTAL SETTINGS

CMA-ES FEATURES

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- $\mathbf{C}$ evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathrm{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
$\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

EXPERIMENTAL SETTINGS

CMA-ES FEATURES

- Generation number $\varphi_g = g$
- Step-size $\varphi_\sigma = \sigma$
- Number of restarts $\varphi_{\text{restart}} = n_{\text{r}}$
- CMA mean distance $\varphi_{d(\mathbf{m})} = \sqrt{(\mathbf{m} - \mu_{\mathbf{X}})^\top \mathbf{C_X}^{-1} (\mathbf{m} - \mu_{\mathbf{X}})}$
- **C** evolution path length square $\varphi_{\mathbf{p}_c} = \|\mathbf{p}_c\|^2$
- σ evolution path ratio $\varphi_{\mathbf{p}_\sigma} = \frac{\|\mathbf{p}_\sigma\|}{E\|\mathsf{N}(\mathbf{0},\mathbf{I})\|}$
- CMA similarity likelihood
  $\varphi_{\mathcal{L}} = -\frac{N}{2}\left(D \log 2\pi\sigma^2 + \log\det \mathbf{C}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)^\top \mathbf{C}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}}{\sigma}\right)$

Experimental settings

Performance space ∼ Ranking Difference Error

$$RDE_{\mu}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{\sum_{i:\rho(i)\leq\mu} |\hat{\rho}(i) - \rho(i)|}{\max_{\pi \in \text{Permutations of } (1,\ldots,\lambda)} \sum_{i:\pi(i)\leq\mu} |\, i - \pi(i)|}$$

λ – population size
$\mathbf{y}, \hat{\mathbf{y}} \in \mathbb{R}^{\lambda}$, $\mu = \lceil \frac{\lambda}{2} \rceil$
$\rho(i)$ – ranks of the i-th element in vector $\mathbf{y}$
$\hat{\rho}(i)$ – ranks of the i-th element in vector $\hat{\mathbf{y}}$

Statistical testing

- RDE_{μ} data diversity
  - Friedman test and Tukey's post-hoc test
  - Significant differences among all pairs of covariances (except one)
- Univariate features descriptivity
  - Kolmogorov-Smirnov test
  - Significant differences between features on sample sets with particular best covariance and all data
- Multivariate features descriptivity
  - Classification tree

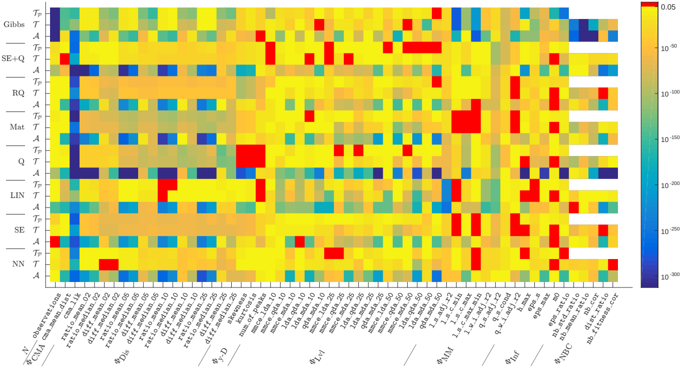
Statistical testing

- RDE_{μ} data diversity
  - Friedman test and Tukey's post-hoc test
  - Significant differences among all pairs of covariances (except one)
- Univariate features descriptivity
  - Kolmogorov-Smirnov test
  - Significant differences between features on sample sets with particular best covariance and all data
- Multivariate features descriptivity
  - Classification tree

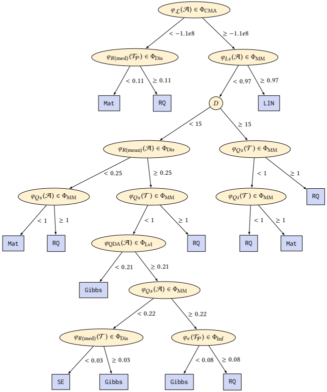
Statistical testing

- RDE_{μ} data diversity
  - Friedman test and Tukey's post-hoc test
  - Significant differences among all pairs of covariances (except one)
- Univariate features descriptivity
  - Kolmogorov-Smirnov test
  - Significant differences between features on sample sets with particular best covariance and all data
- Multivariate features descriptivity
  - Classification tree

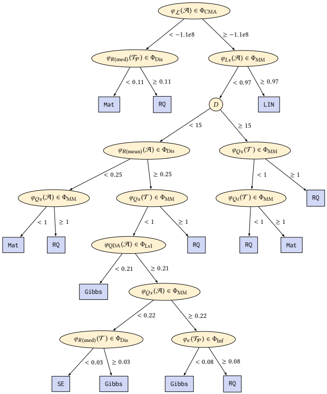
KOLMOGOROV-SMIRNOV TEST

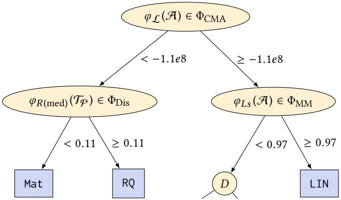


Selection mapping ∼ Decision tree

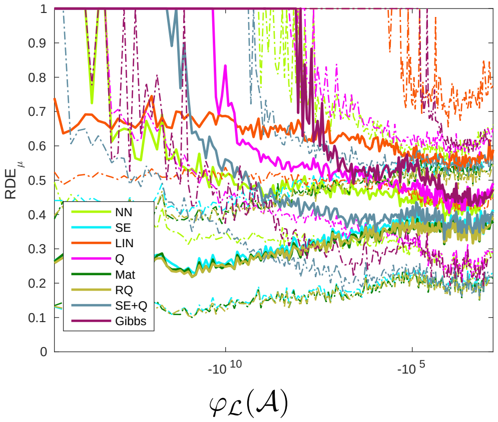


SELECTION MAPPING ~ DECISION TREE



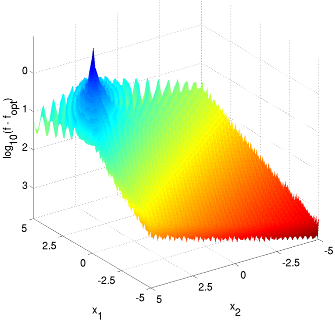


Feature vs. Error

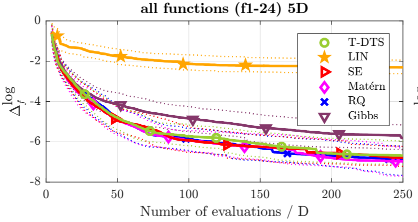


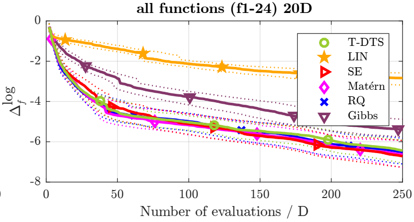
Decision tree validation

- COCO framework
  - 24 noiseless benchmarks
  - 5 dimensions
  - 15 instances
- Algorithms
  - T-DTS – DTS-CMA-ES adaptively changing kernel according to features using decision tree
  - 5 DTS-CMA-ES versions using fixed kernels

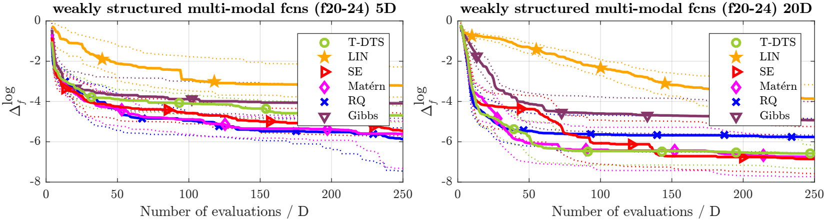


Experimental results





Experimental results



Summary of results

- Statistical testing
  - Significant differences in covariance performance ordering
  - Significant differences in feature distribution
  - CMA-ES based features are useful
- Decision tree with DTS
  - Surrogate model selection methodology can be utilized for GP kernel selection
  - Selection of GP kernel using classification tree in DTS-CMA-ES provided a performance equivalent to versions with successful fixed kernels
- Future research:
  - Feature reliability
    - Number of observations, dimension $\rightarrow$ density
    - Data distribution
  - Covariance selection methods

QUESTIONS?

z.pitra@gmail.com

lubajer@cisco.com

j.repicky@gmail.com

martin@cs.cas.cz