Knowledge Discovery for Pareto based Multiobjective Optimization in Simulation

__Patrick Lange__, Rene Weller, Gabriel Zachmann

University of Bremen, Germany

cgvr.informatik.uni-bremen.de

ACM SIGIM PADS

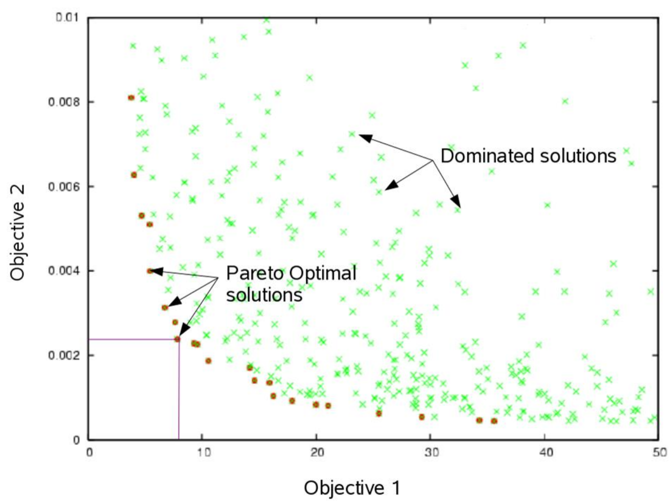
15-18 May 2016, Banff, AB, Canada

Multiobjective Optimization

- Simulation-based optimization
- Multidisciplinary design attempts to satisfy multiple, possibly conflicting, objectives at once

$$(MOP)\ \min_{x \in X} F(x) = (f_1(x), f_2(x), \ldots, f_p(x))$$

- Blackbox simulations: f_i not known
  - No partial derivatives, no constraints, no relationships…



(From: Cenaero http://www.cenaero.be/Page.asp?docid=27103&)

Motivation: Blackbox Simulations

- Engineers can not describe the relationships which are used to formulate a mathematical problem (e.g. differential equations)
- Finding a tradeoff set of input parameters which satisfy all simulation goals

Simulation goals

Parameters

$$(MOP)\ \min_{x \in X} F(x) = (f_1(x), f_2(x), \dots, f_p(x))$$

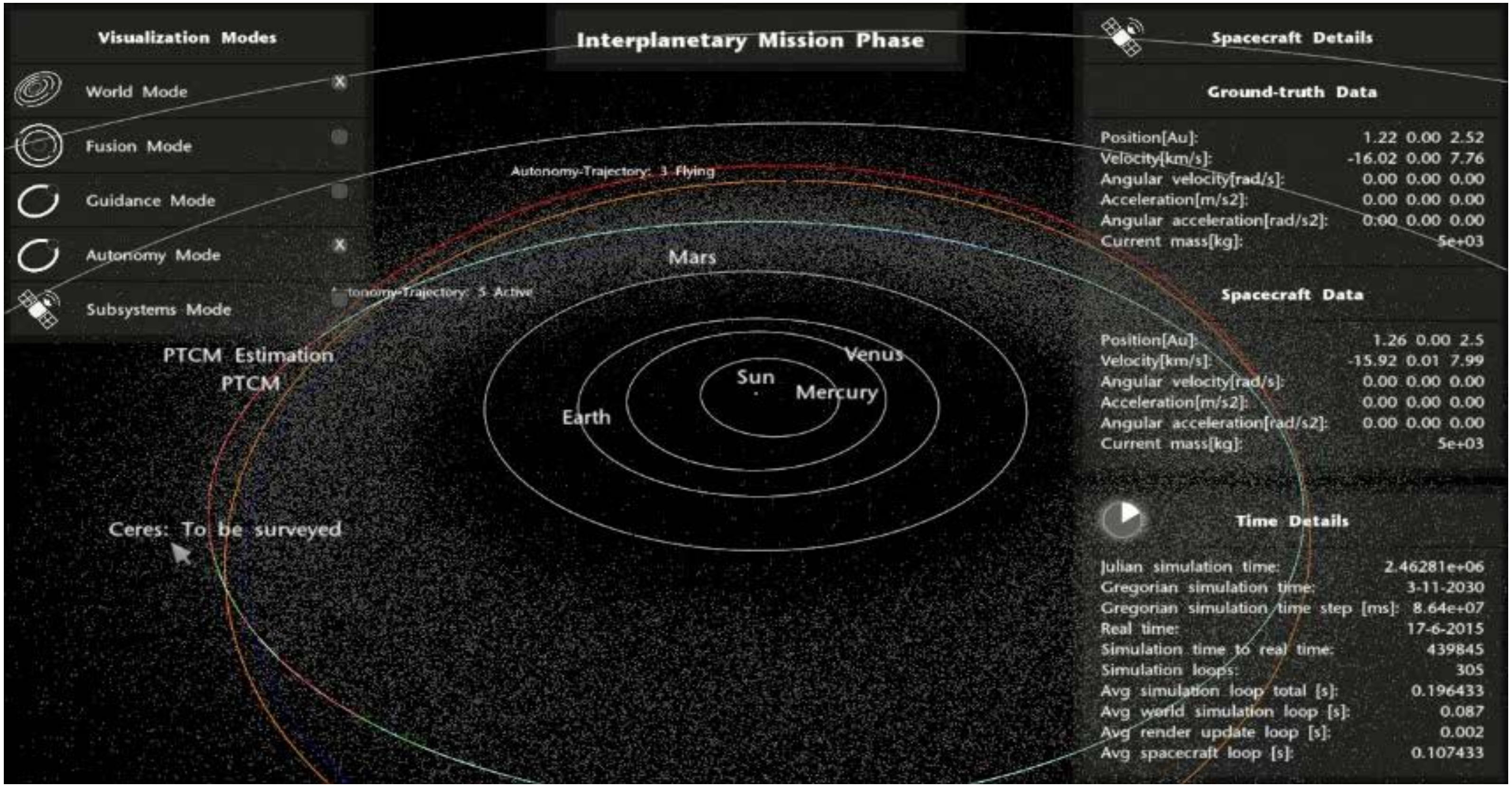
$$f_i\ not\ known$$

Satisfaction of goal states

- Application in simulation-based feasibility studies
  - Our use case scenario: Autonomous spacecraft operations for small planetary objects

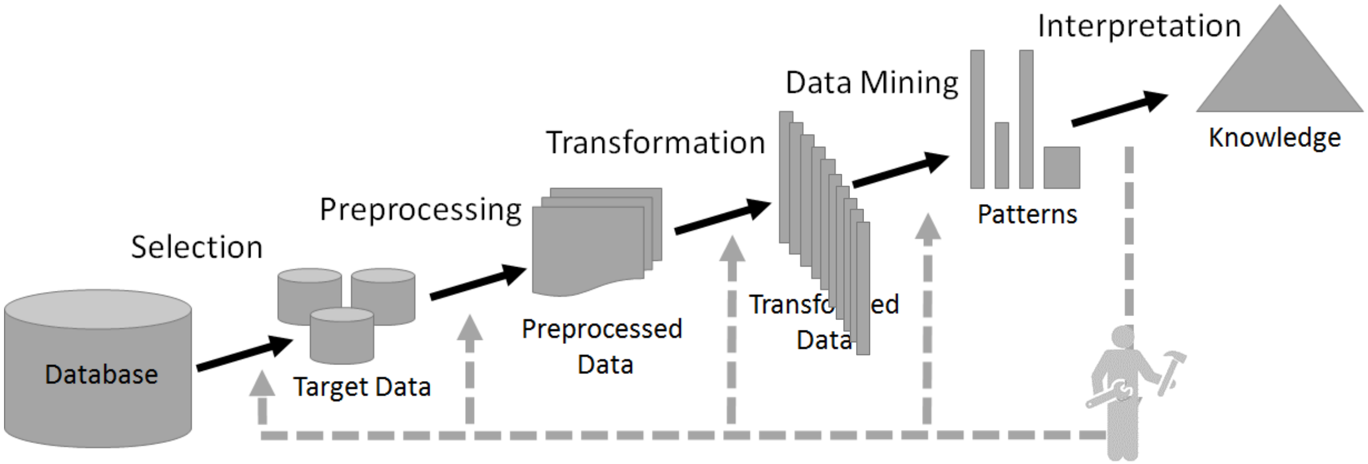
Motivation: Autonomous Spaceflight Example

- Propulsion type ⇒ Orbit transfer ⇒ Planetary visibility ⇒ Self-localization ⇒ Ground station communication ⇒ Bandwidth ⇒ Antenna diameter



The Knowledge Discovery Process

- Main idea: Use simulation itself to generate data in order to simulate, optimize or analyze the given model
- Making sense of huge data collections
- Semi-automatic five step process
- Requires several iterations of some steps
- Collection of data mining techniques

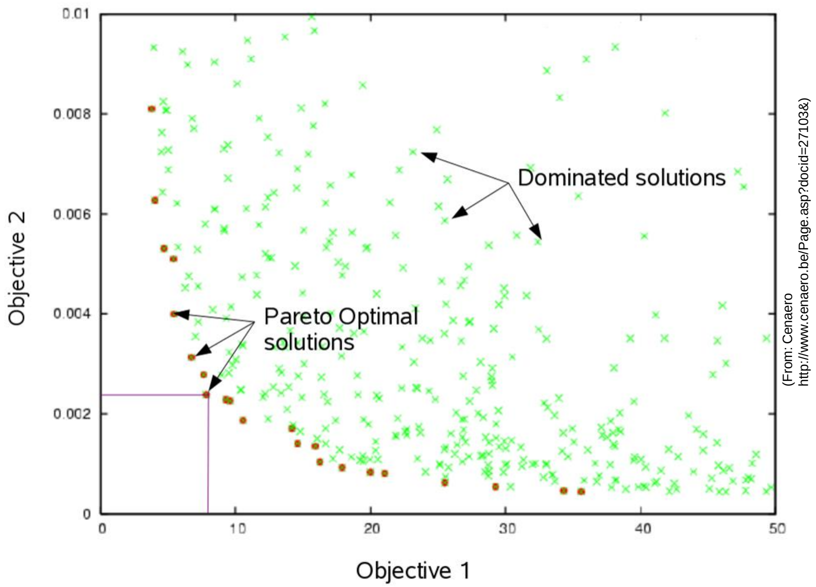


KD Processes in Simulations

- Single objective optimization
  - Landscape characterization problem exploration via support vector machines [Burl‘06]
  - Determination of adaptation strategies for linear relationships [Lattner‘11]
  - Linear regression of input parameters and classification [Painter‘06]
- Multi objective optimization
  - Analysis of existing Pareto solutions [Bandaru‘10,Sugimura‘07,Liebscher‘09,Dudas‘15]

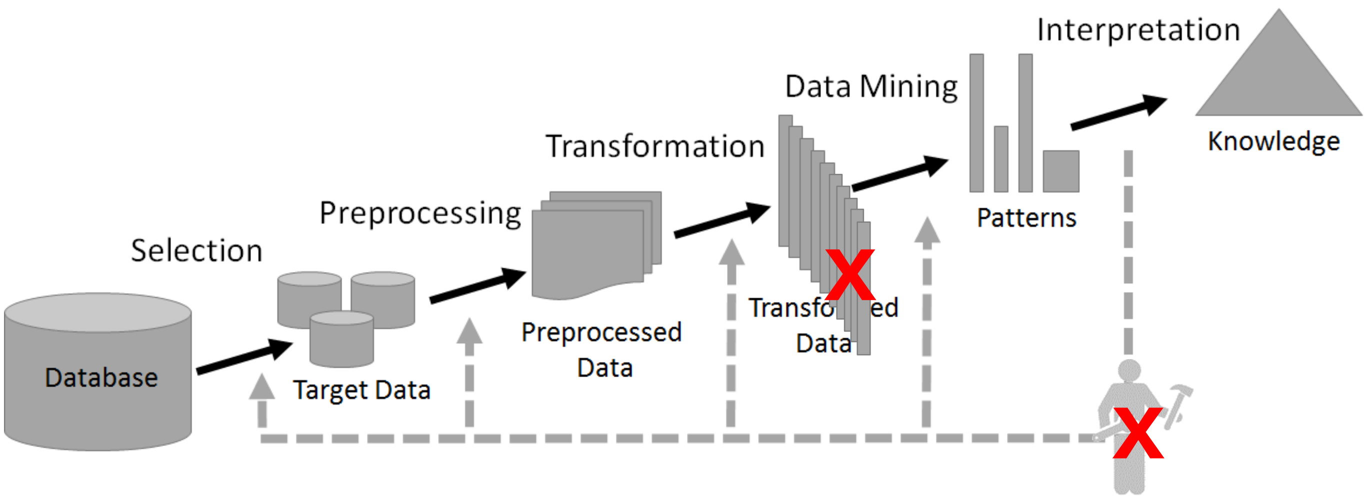
Remaining Challenges

1. Multiobjective optimization
   - Approximation of the feasible design space
2. Blackbox simulation
   - Determination of relationships between input parameters and simulation goals



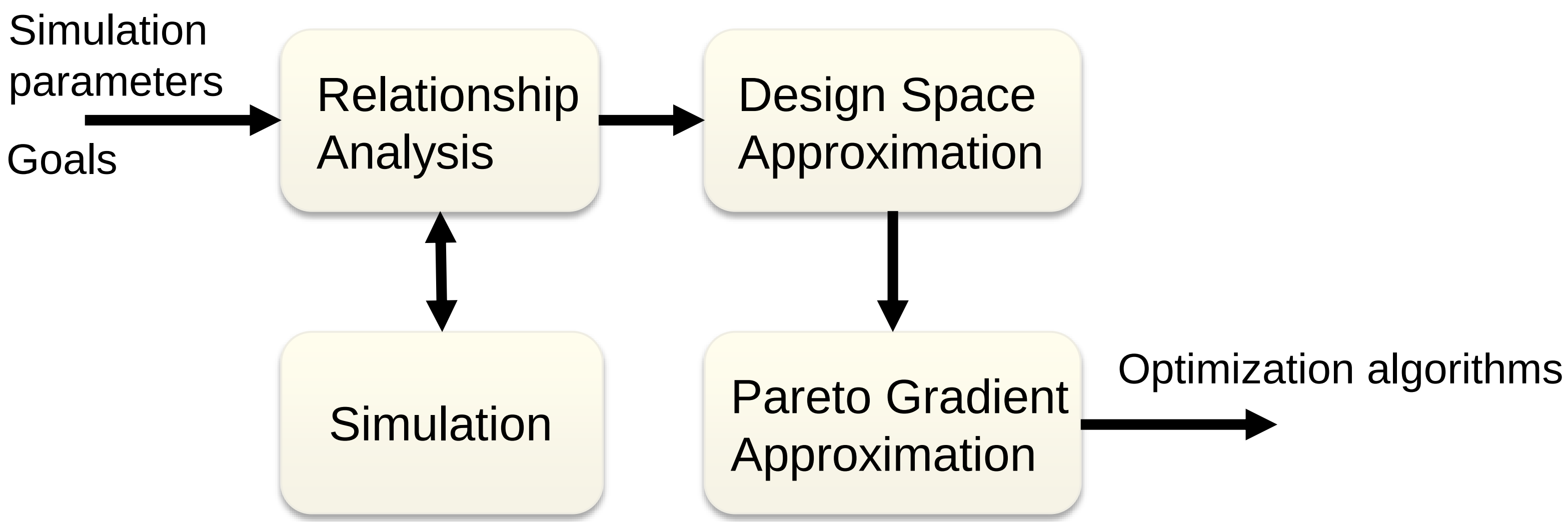
Features

1. Reduce amount of simulation data farming
2. Completely autonomous knowledge discovery process
   - Remove manual assessment of knowledge discovery results



Our Approach

- Completely autonomous knowledge discovery process
  - Uncovers hidden relationships between simulation input parameters and simulation goals with few samples from the simulation
  - Approximates feasible design space
  - Approximates Pareto gradient information for multiobjective algorithms



Goal

- Approximate objective function f and determine their input $(x_i, \ldots, x_k)$

$$f_j(x_i, \ldots, x_k) \rightarrow G_n$$

- Complexity of simulation data farming
  - Brute-force approach is too computationally expensive

$$O\big((p^2 - p) \cdot m\big) \qquad \begin{aligned} m &: \#simulation\ goals \\ p &: \#input\ parameters \end{aligned}$$

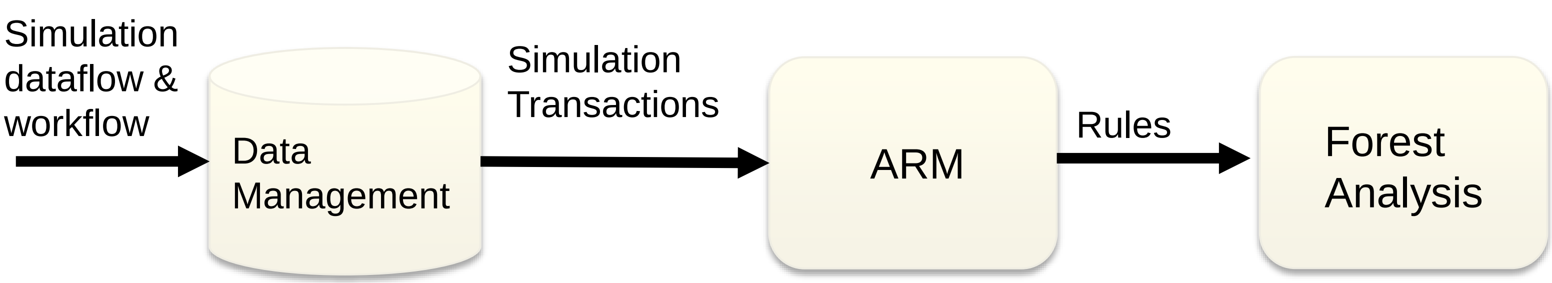
- Our two phase approach reduces the farming operations
  - Forest-based association rule analysis determines $(x_i, \ldots, x_k)$
  - Spline-based sampling approximates f_j

Association Rule Mining

- Requires centralized data management which records transactions of all software modules (e.g. GraphPool)
- Outputs list of association rules

$$\text{Module A}: \mathrm{X} \Rightarrow \ Y \qquad X \cap Y = 0 \qquad X, Y \subseteq P$$

- Association rule implies workflow from X to Y
- Example: $\text{Module Propulsion: Fuel} \Rightarrow \text{Mass}$

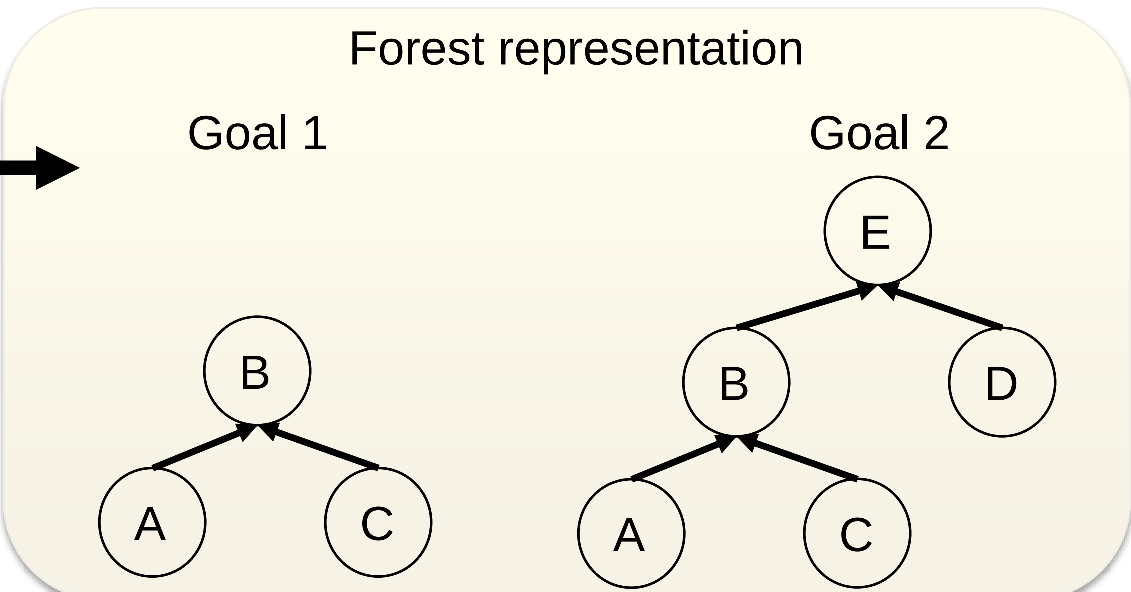


Forest-Based Association Rule Analysis

- Represent list of association rules in a tree data structures (association rule tree)
- One association rule tree for every goal

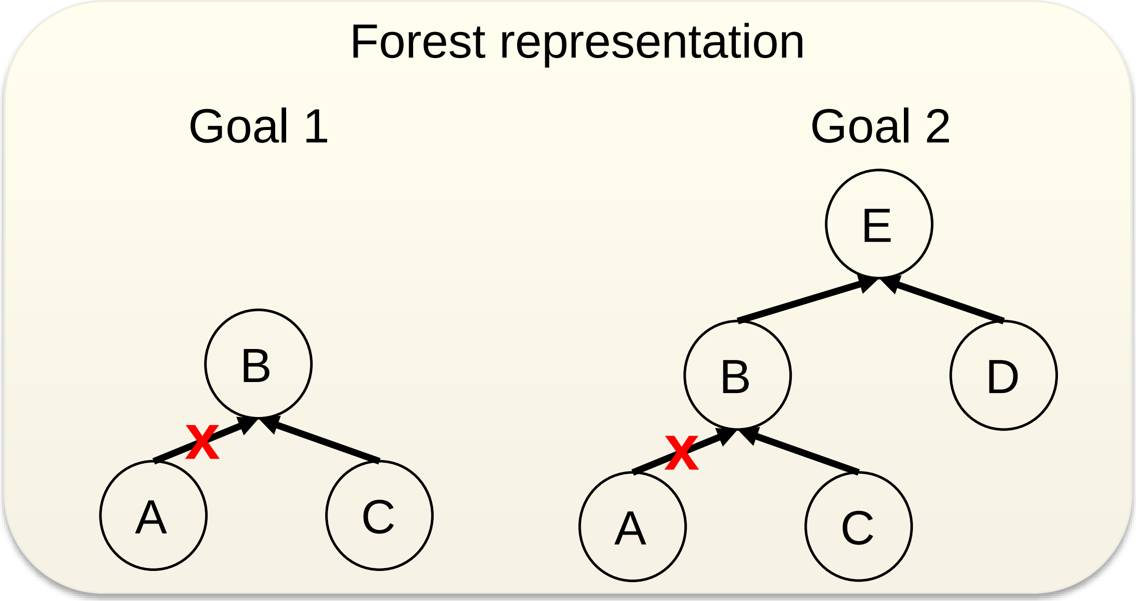
Association rules

Module 1: $A \Rightarrow B$
Module 2: $B \Rightarrow E$
Module 3: $C \Rightarrow B$
Module 4: $D \Rightarrow E$



Forest-Based Association Rule Analysis

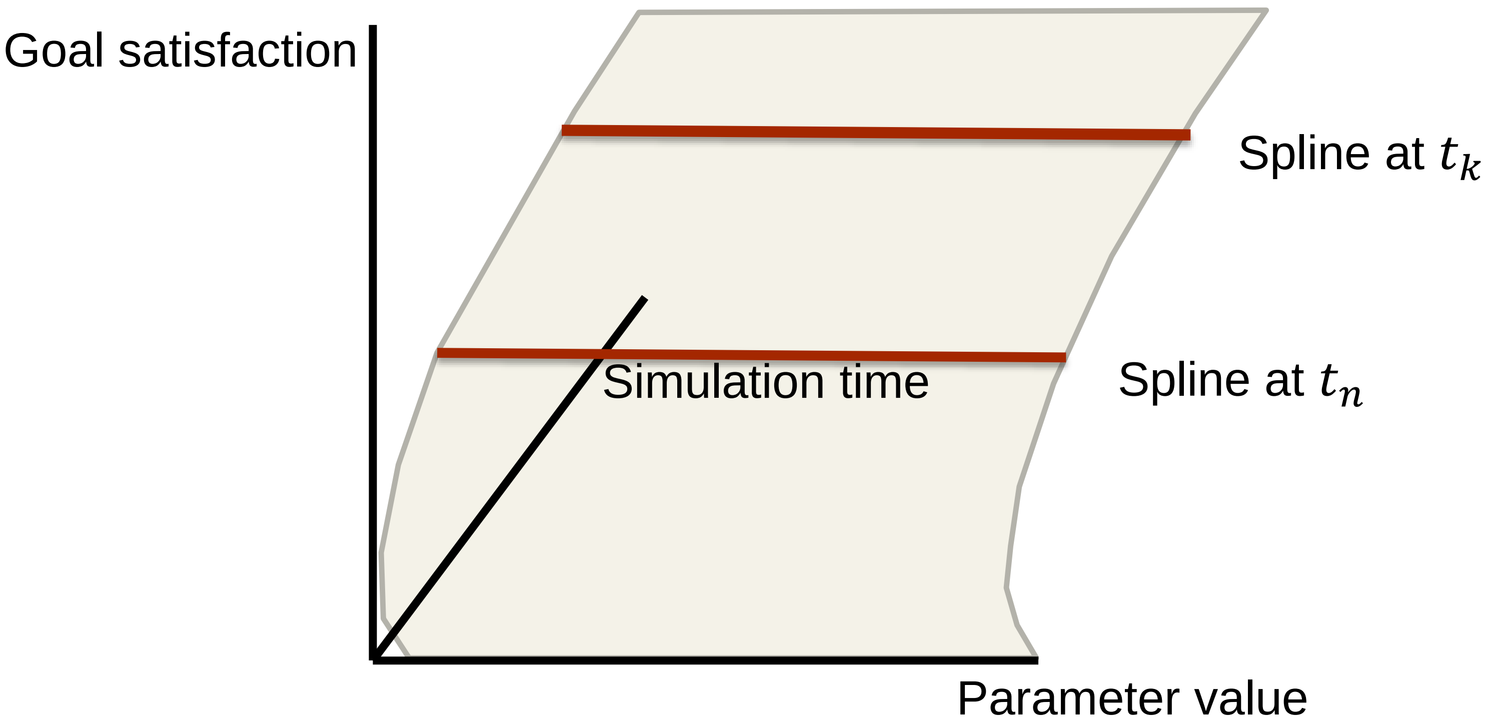
- Determination of correlation between input parameter and simulation goal
  - Prune sub-tree if no correlation can be found
- Approximate the relationship with splines



Spline-based Sampling

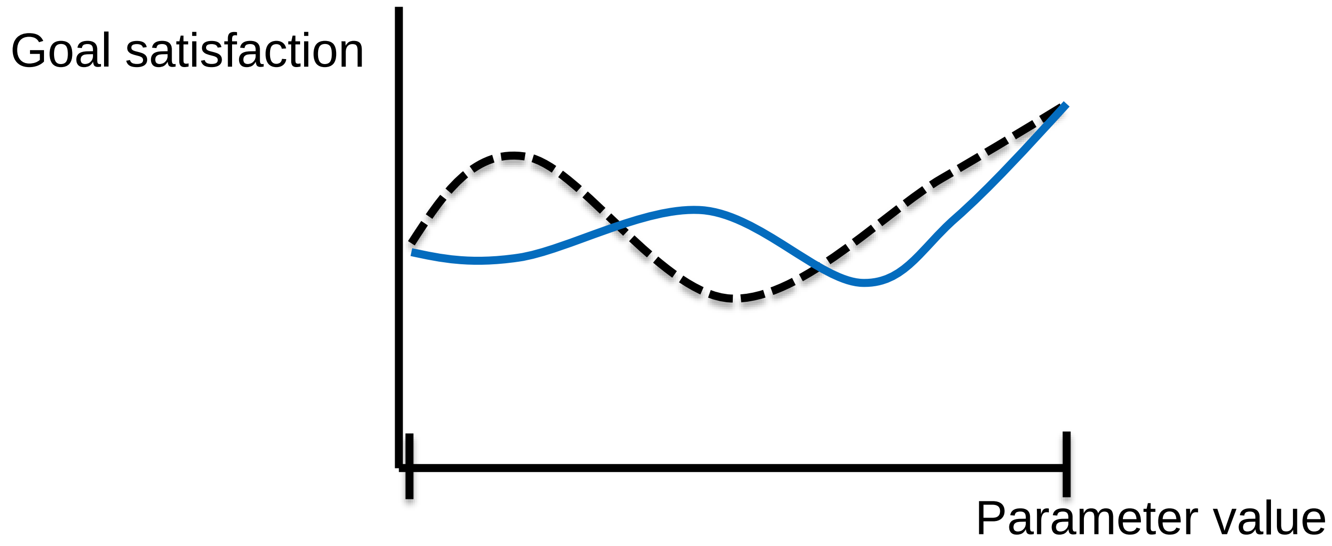
- Relationship defines three-dimensional space

1. Approximate behavior per time frame with one spline
2. Analyze spline for correlation



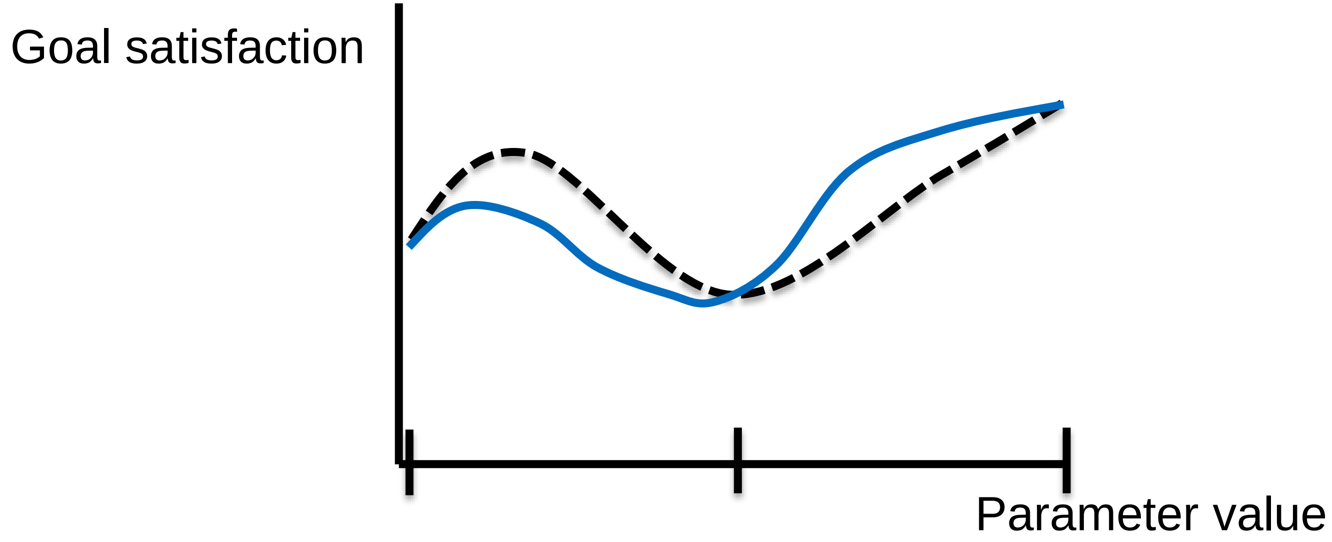
Spline-based Sampling

- Draw samples which minize euclidean distance between samples in parameter space
- Stop if spline predicts next n satisfaction states correctly



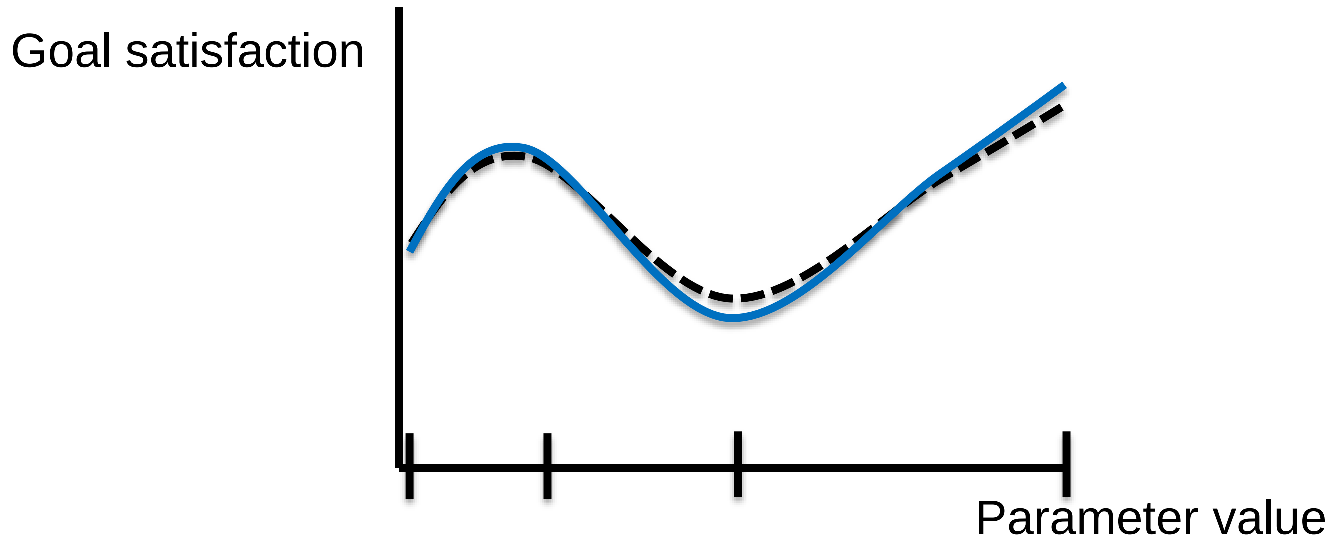
Spline-based Sampling

- Draw samples which minize euclidean distance between samples in parameter space
- Stop if spline predicts next n satisfaction states correctly



Spline-based Sampling

- Draw samples which minize euclidean distance between samples in parameter space
- Stop if spline predicts next n satisfaction states correctly

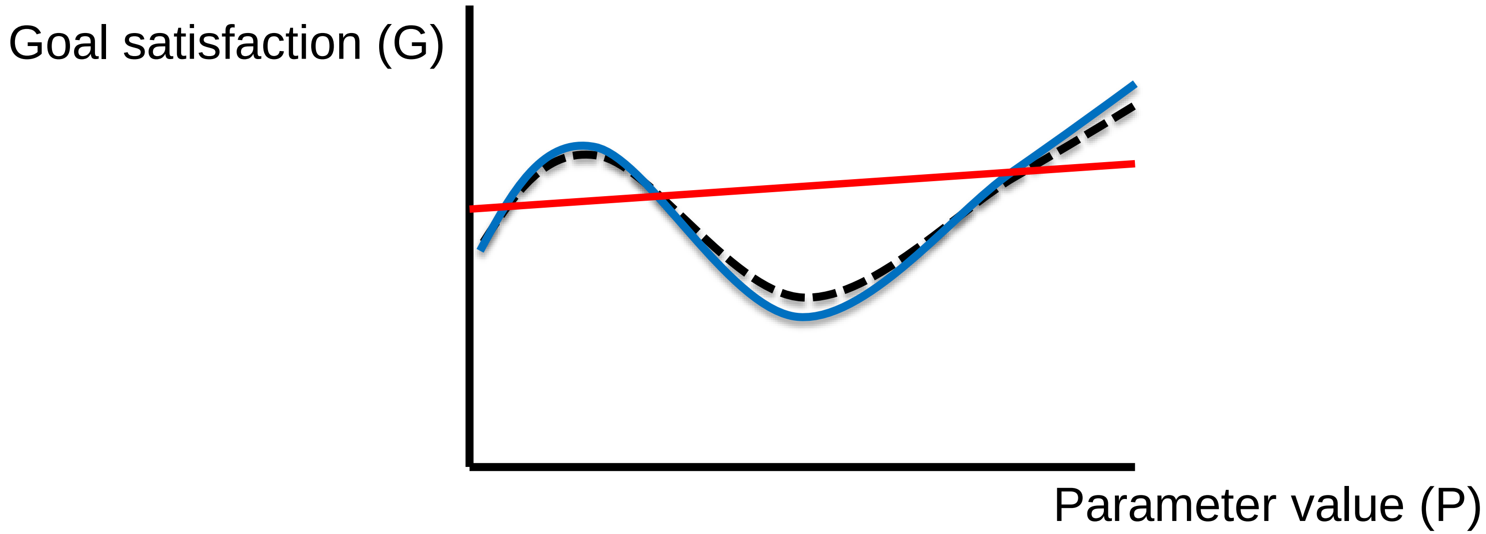


Recursive Correlation Analysis

- Compute correlation coefficient for spline

$$r = \frac{\sum(P-\overline{P})(G-\overline{G})}{\sqrt{\sum(P-\overline{P})^2}\sqrt{\sum(G-\overline{G})^2}}$$

- If coefficient does not yield correlation, split the spline and recompute the coefficient



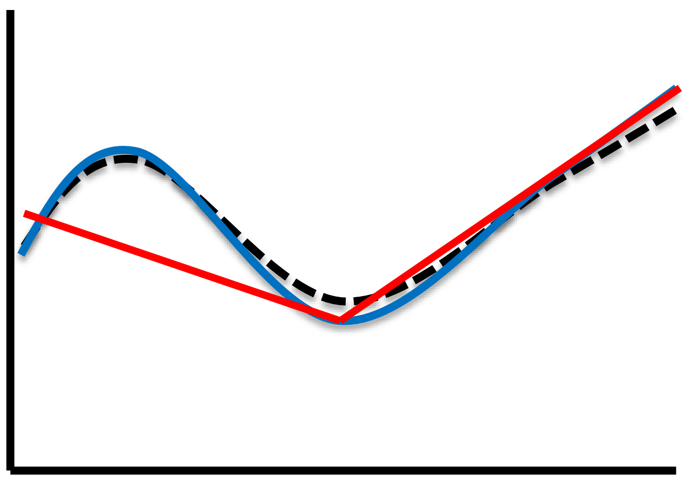
Recursive Correlation Analysis

- Compute correlation coefficient for spline

$$r = \frac{\sum(P-\overline{P})(G-\overline{G})}{\sqrt{\sum(P-\overline{P})^2}\sqrt{\sum(G-\overline{G})^2}}$$

- If coefficient does not yield correlation, split the spline and recompute the coefficient

Goal satisfaction (G)



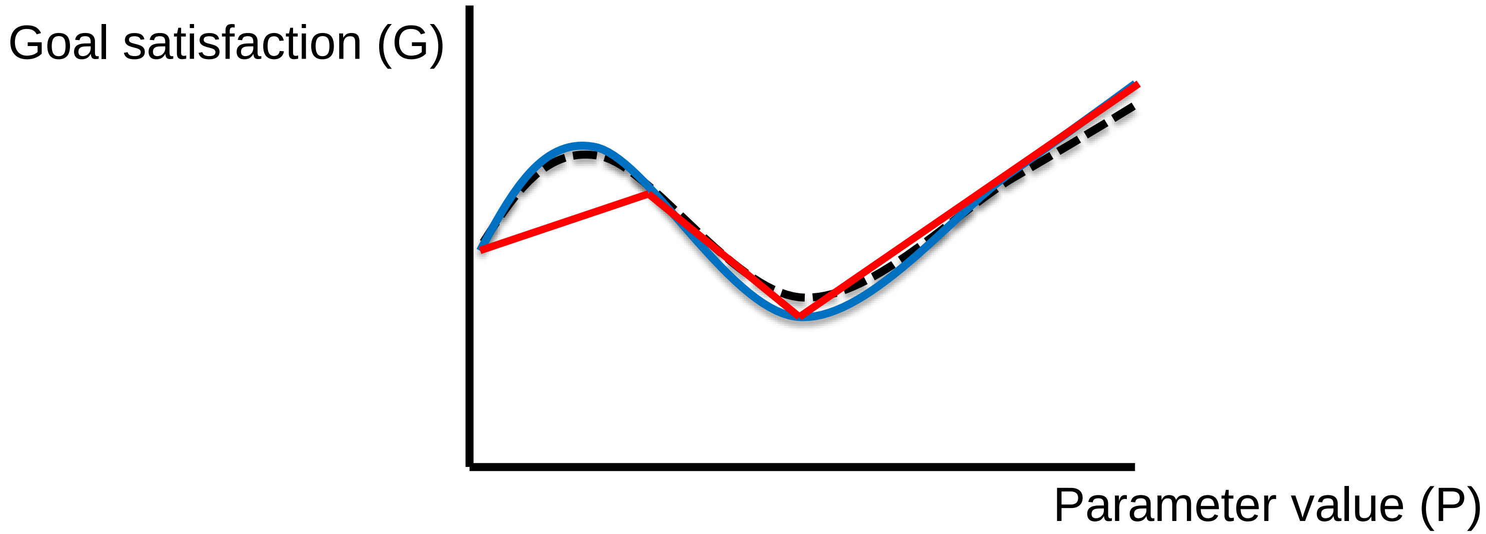
Parameter value (P)

Recursive Correlation Analysis

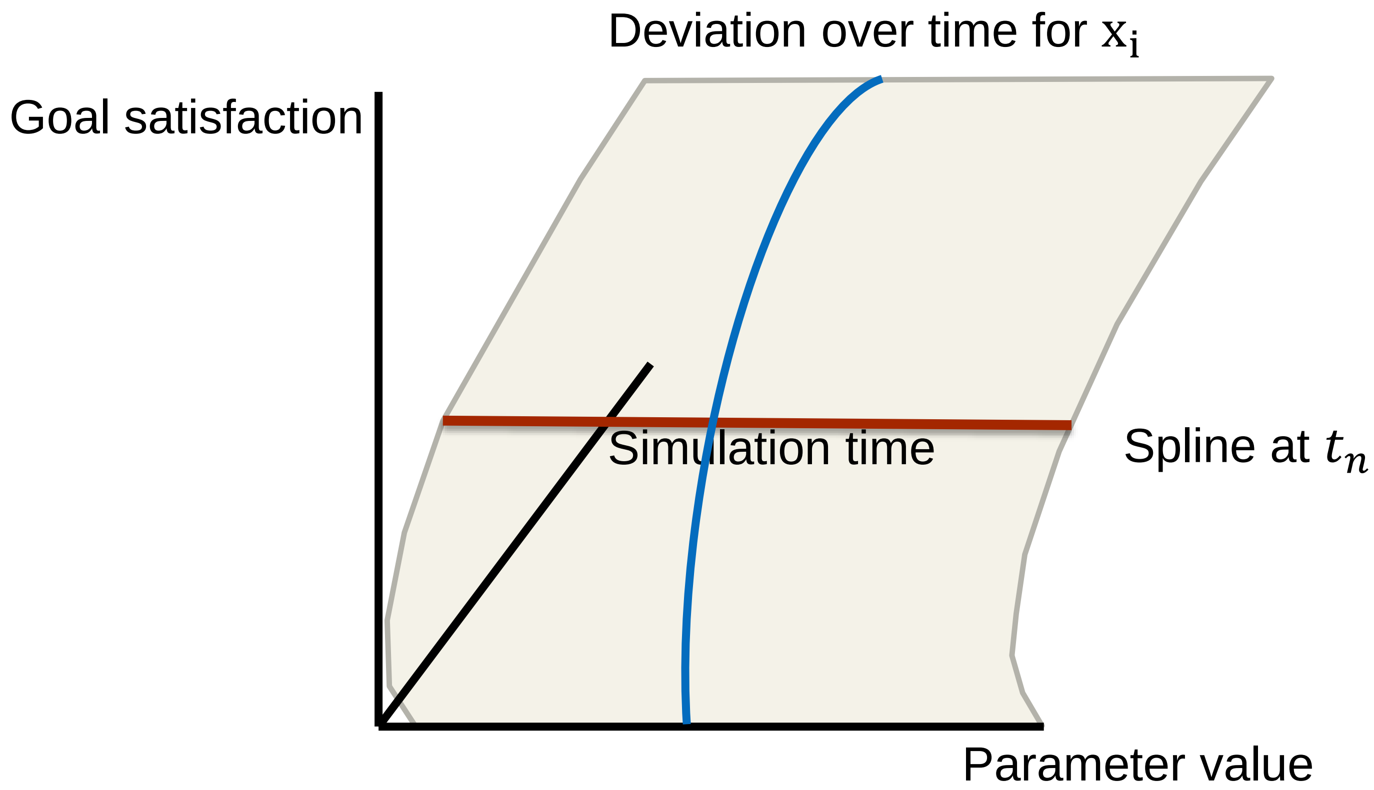
- Compute correlation coefficient for spline

$$r = \frac{\sum(P-\overline{P})(G-\overline{G})}{\sqrt{\sum\left(P-\overline{P}\right)^{2}}\sqrt{\sum\left(G-\overline{G}\right)^{2}}}$$

- If coefficient does not yield correlation, split the spline and recompute the coefficient



Feasible Design Space Approximation



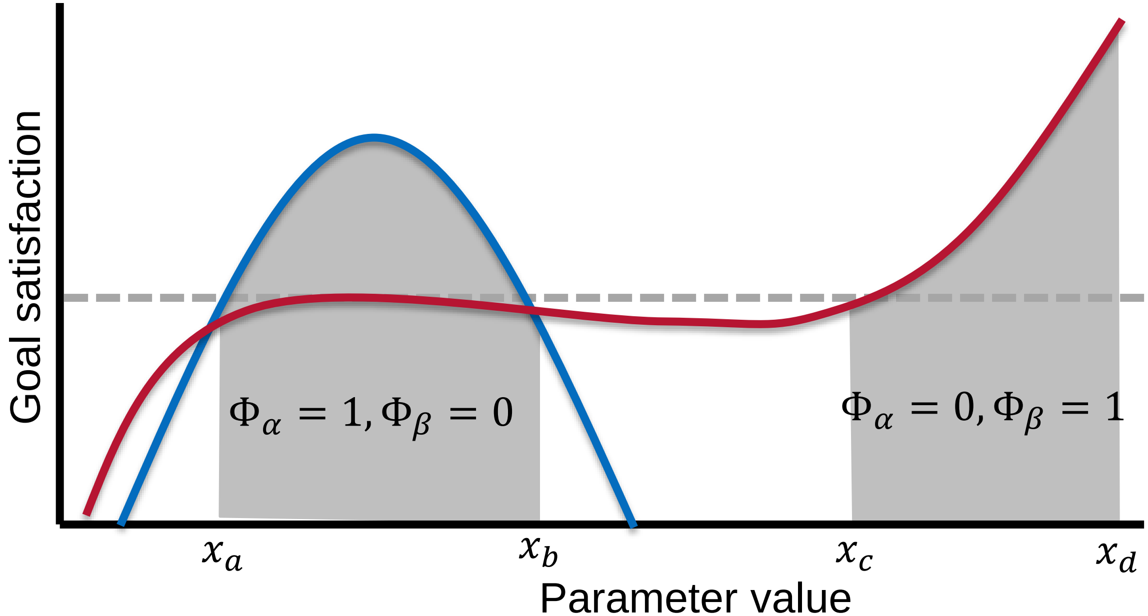
Feasible Design Space Approximation

- Weighting of spline deviation

$$\gamma(x_i, t_i) = \frac{e^{-k^2}\alpha_{t_i}(x_i) + \ \ldots + e^{-g^2}\alpha_{t_m}(x_i)}{m}$$

- Pareto space

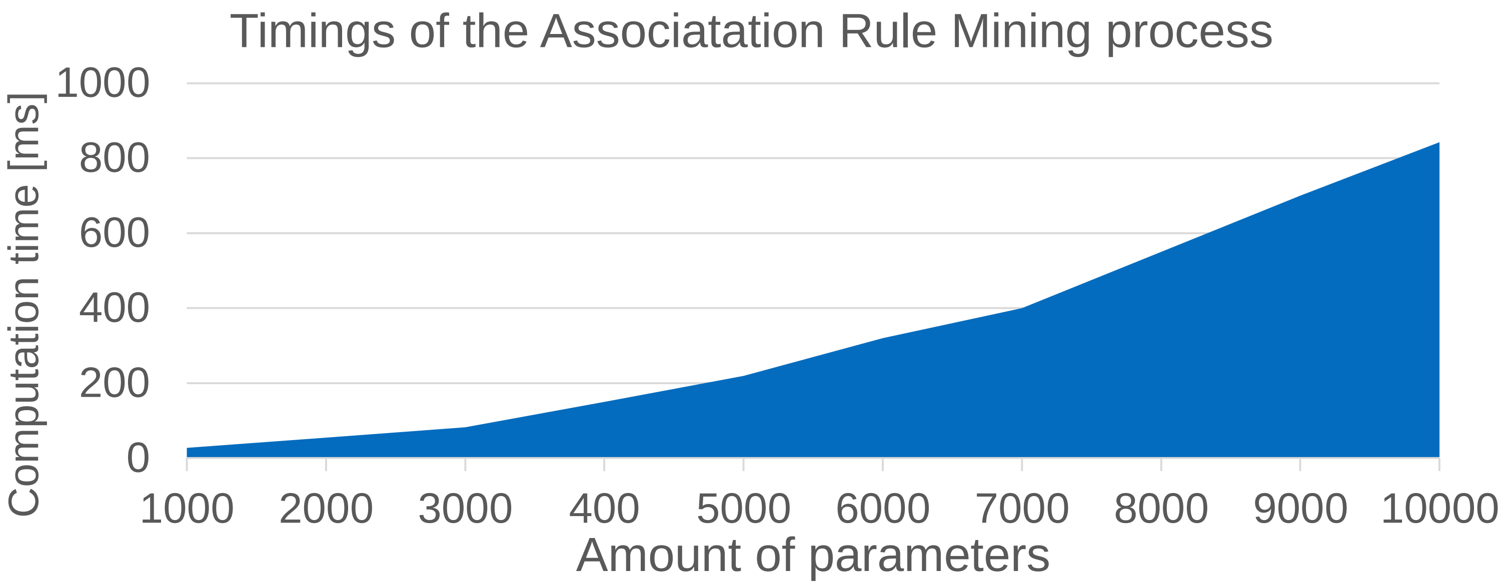
$$\omega_{pareto}\,(x_i, t_i) = \frac{\sum \Phi(|\frac{o}{n} - \frac{o}{\sum f(x_i)} \cdot \gamma(x_i, t_i)|)}{k}$$

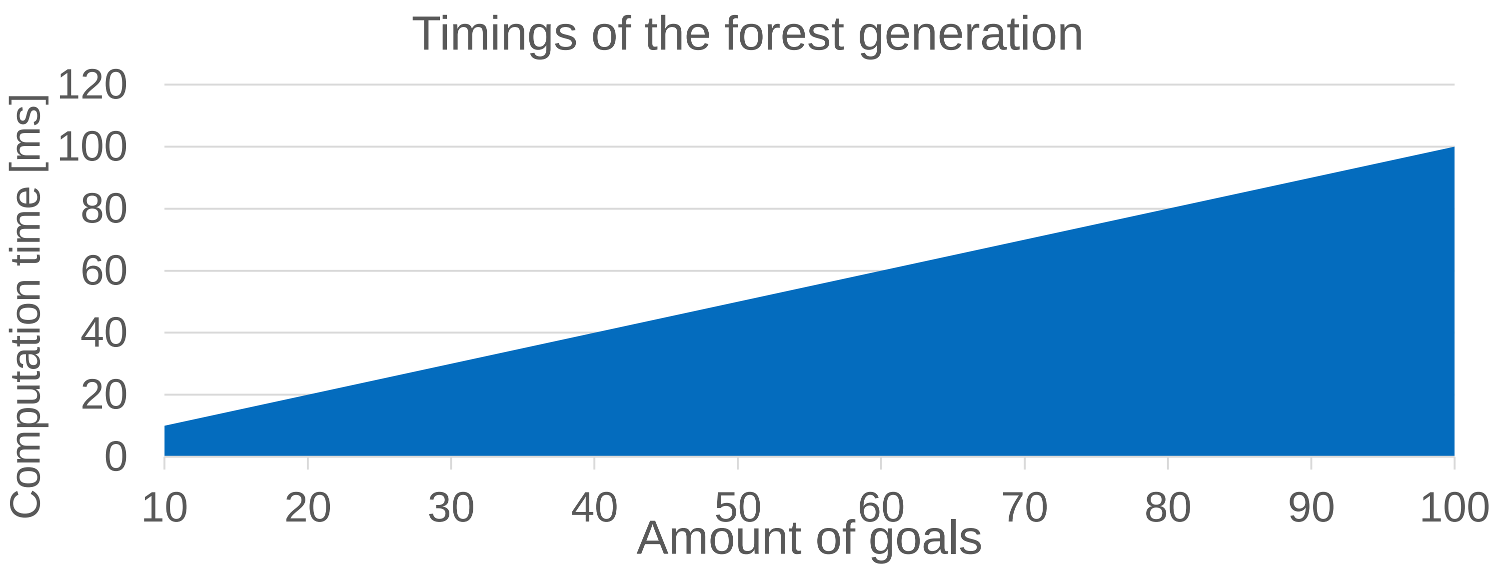


Evaluation

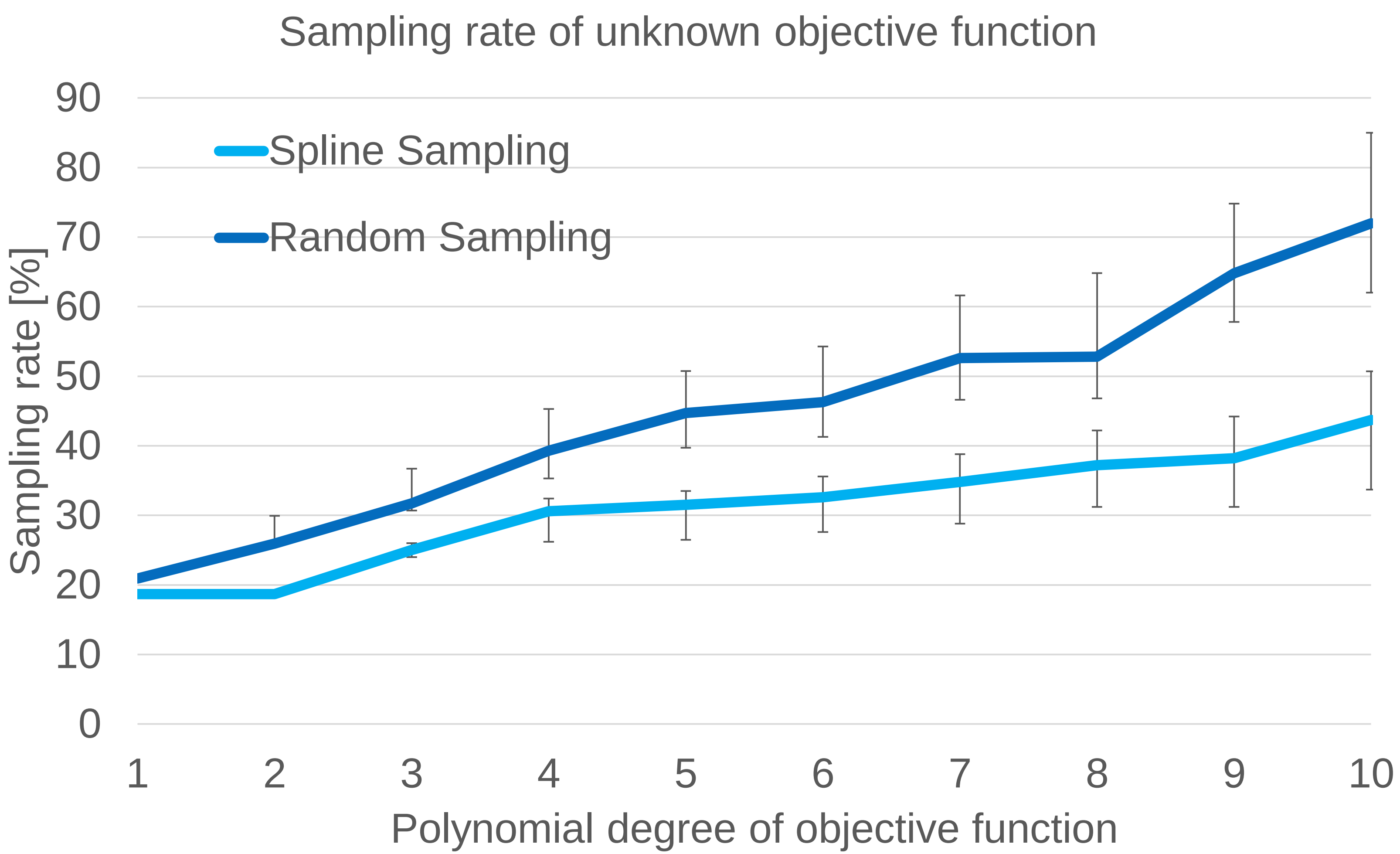
- Performance evaluation of association rule mining step, forest generation and spline-based sampling
- Two use case studies for quality performance evaluation
  - Lotka-Volterra prey predator system
  - Interplanetary cruise flight
- Synthetic optimization scenarios
  - Gradient descent, simulated annealing, evolutionary algorithm

Simulation Analysis

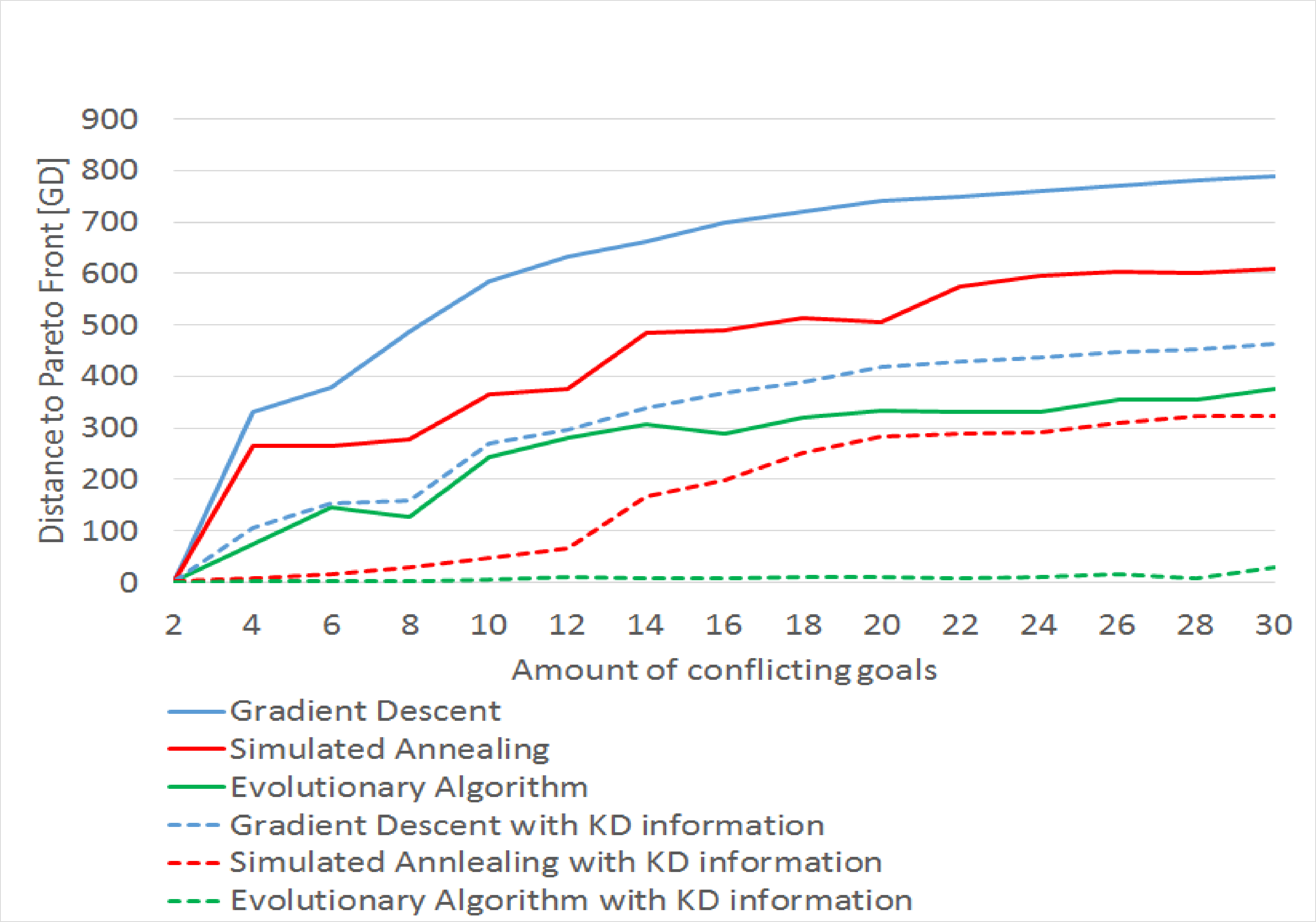




Spline-Based Sampling



Quality of Optimization Algorithms



Conclusion

- Completely autonomous knowledge discovery process
- Uncovers hidden relationships between simulation input parameters and simulation goals
  - Our technique requires up to 40 % less samples
- Approximates Pareto gradient information for multiobjective algorithms
  - Gradient descent up to a factor of 5
  - Simulated annealing up to a factor of 8
  - Evolutionary algorithm up to a factor of 12

Future Work

- Extension of spline-sampling for stochastic simulation
- Integration of gradient information into spline-based objective function sampling
- Evaluation with standard optimization problems (e.g. SimOpt library)

Thank you for your attention

Questions?

Patrick Lange, Rene Weller, Gabriel Zachmann
{lange,weller,zach}@cs.uni-bremen.de

This research is based upon the project KaNaRiA, supported by German Aerospace Center (DLR) with funds of German Federal Ministry of Economics and Technoloy (BMWi) grant *50NA1318*