# Multi-objective negotiation of power profiles for datacenter powered with renewable energies

**Léo Grange**

University of Toulouse
Institut de Recherche en Informatique de Toulouse (IRIT)

July 2018

Institut de Recherche
en Informatique de Toulouse

as green as possible

# Context and overview

# Datacenter consumption and renewable sources

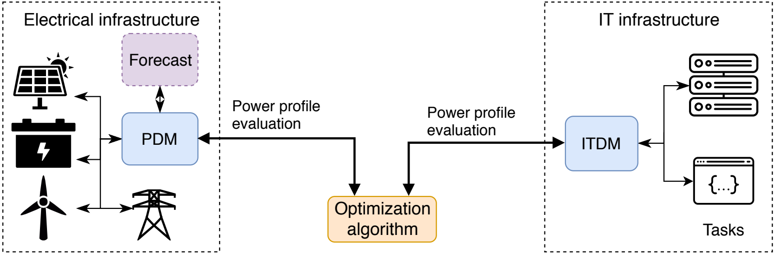
- Worldwide: 270 TWh in 2012
  - $\approx$ Italy electricity consumption
  - High economical and environmental costs

## Possible mitigations

- Improve energy efficiency, software and hardware
- Use renewable energy sources power
  - Solar, wind: intermittent and little predictability
  - New challenges to make efficient use in datacenters

- ANR Datazero: on-site renewable sources
- IT and electrical cooperation

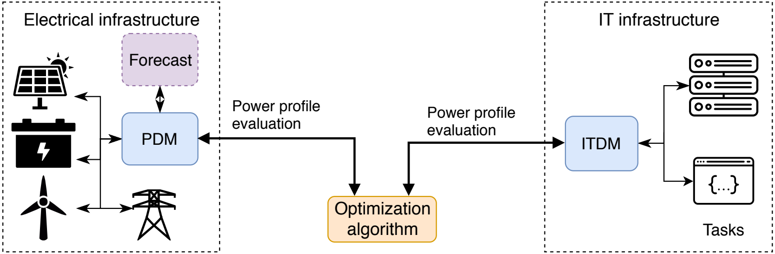
## Context and overview of the problem




### Separated IT and electrical optimizations

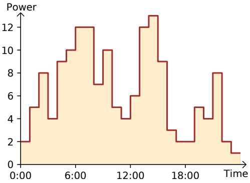
- Ability to evaluate power plan impact
- Internal objective (utility)
- Black box functions $\mathbb{R}^T \to \mathbb{R}$
- Computationally expensive

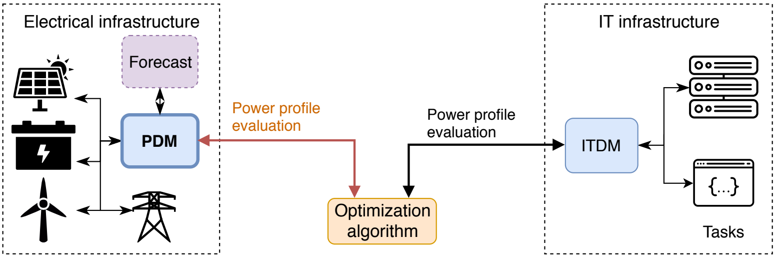
## Context and overview of the problem




### Separated IT and electrical optimizations

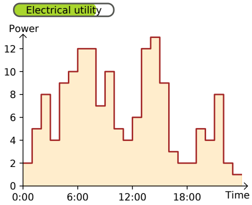
- Ability to evaluate power plan impact
- Internal objective (utility)
- Black box functions $\mathbb{R}^T \to \mathbb{R}$
- Computationally expensive

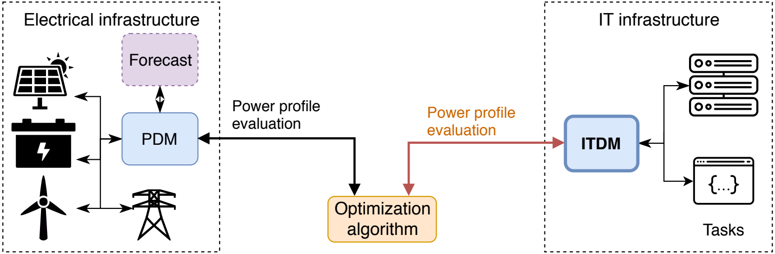





## Separated IT and electrical optimizations

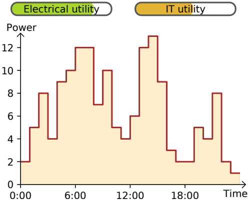
- Ability to evaluate power plan impact
- Internal objective (utility)
- Black box functions $\mathbb{R}^T \to \mathbb{R}$
- Computationally expensive






## Separated IT and electrical optimizations

- Ability to evaluate power plan impact
- Internal objective (utility)
- Black box functions $\mathbb{R}^T \rightarrow \mathbb{R}$
- Computationally expensive



# Multi-objective aspect

- Each DM has one or more objectives to satisfy
- Objectives may differ between DM
  - QoS related for ITDM, environmental impact for PDM

## Managing different objectives

- Avoiding the problem: find common objective (money?)
- Scalarization (e.g. weighted sum)
- Finding a set of good solutions (set of possible trade-offs)

# Multi-objective aspect

- Each DM has one or more objectives to satisfy
- Objectives may differ between DM
  - QoS related for ITDM, environmental impact for PDM

## Managing different objectives

- Avoiding the problem: find common objective (money?)
- Scalarization (e.g. weighted sum)
- Finding a set of good solutions (set of possible trade-offs)

# Multi-objective aspect

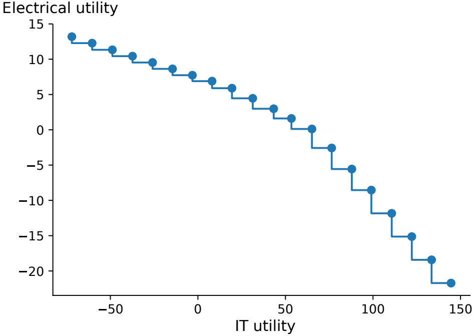
- Each DM has one or more objectives to satisfy
- Objectives may differ between DM
  - QoS related for ITDM, environmental impact for PDM

## Managing different objectives

- Avoiding the problem: find common objective (money?)
- Scalarization (e.g. weighted sum)
- Finding a set of good solutions (set of possible trade-offs)

# Multi-objective optimization and heuristics

- Find Pareto front (best trade-offs)

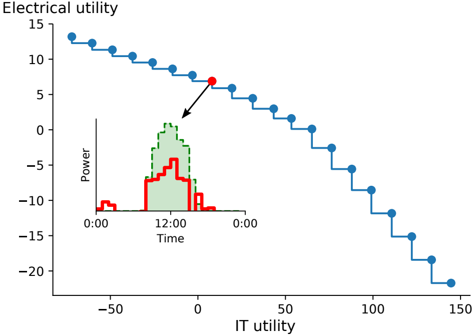



## Multi-Objective Evolutionary Algorithms

- Well studied area, various approaches
- Focused on SPEA2 (genetic algorithm)

# Multi-objective optimization and heuristics

- Find Pareto front (best trade-offs)

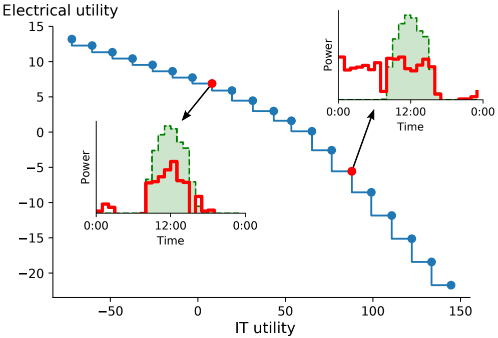



## Multi-Objective Evolutionary Algorithms

- Well studied area, various approaches
- Focused on SPEA2 (genetic algorithm)

# Multi-objective optimization and heuristics

- Find Pareto front (best trade-offs)

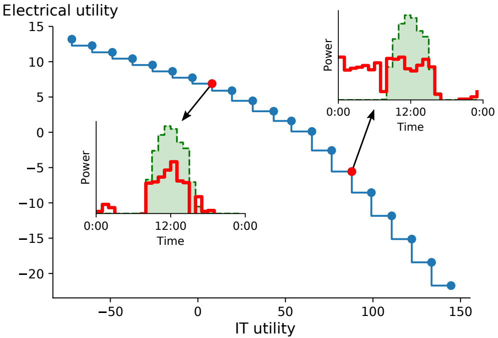



## Multi-Objective Evolutionary Algorithms

- Well studied area, various approaches
- Focused on SPEA2 (genetic algorithm)

# Multi-objective optimization and heuristics

- Find Pareto front (best trade-offs)




## Multi-Objective Evolutionary Algorithms

- Well studied area, various approaches
- Focused on SPEA2 (genetic algorithm)

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

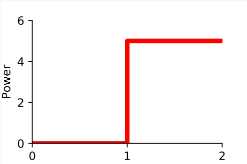
## Example: utility function, 2 time steps

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

## Example: utility function, 2 time steps



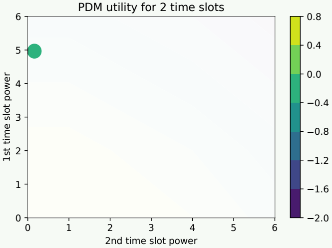



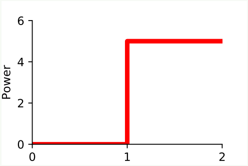

$u = -0.29$

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

## Example: utility function, 2 time steps



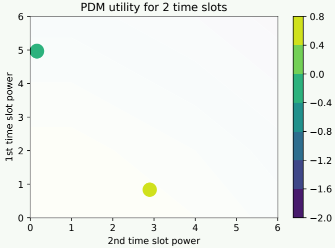



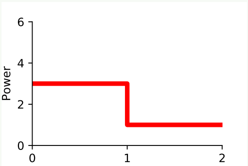

$u = -0.29$

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

## Example: utility function, 2 time steps



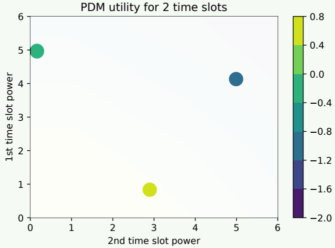



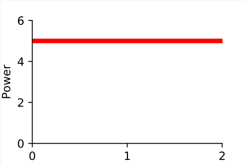

$u = 0.41$

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

## Example: utility function, 2 time steps



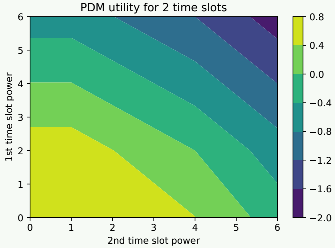




$$u = -1.2$$

# Approximation of power profile utility

- Evaluation of power profile is costly
  - Genetic algorithms require many evaluations
- Workaround: Utility approximation
  - Fast approximation based on known solutions
  - Evaluate only potentially good ones

## Example: utility function, 2 time steps




Only 2 dimensions, easy regression. What about 80 dimensions?

# Constraints for approximation methods

Goal: find a function $\mathbb{R}^T \to \mathbb{R}$ (profile to utility).

- Online learning with few training data
  - Utility function changes between negotiations

Curse of dimensionality...

- $\mathbb{R}^T$ is huge ($T > 100$ in many scenarios)
- Most regression method become impractical

# Constraints for approximation methods

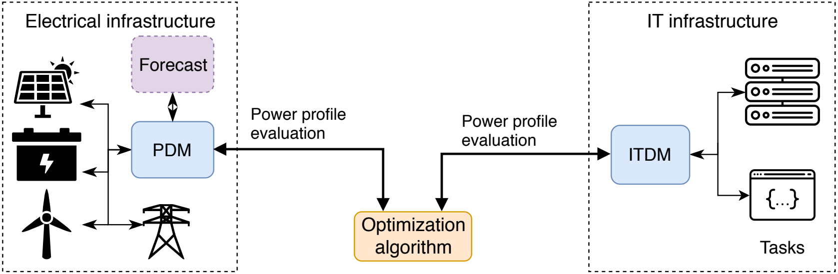
Goal: find a function $\mathbb{R}^T \to \mathbb{R}$ (profile to utility).

- Online learning with few training data
  - Utility function changes between negotiations

**Curse of dimensionality...**

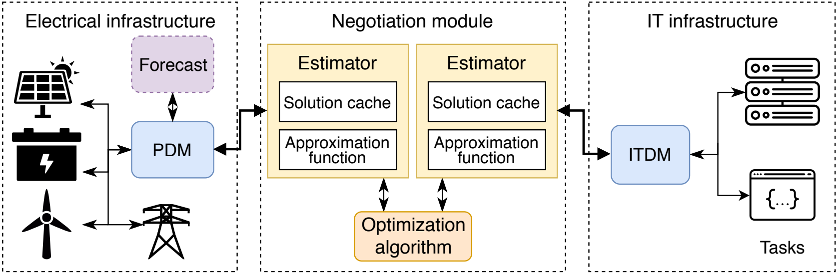
- $\mathbb{R}^T$ is huge ($T > 100$ in many scenarios)
- Most regression method become impractical

# Approximation in the overall infrastructure




- Improving negotiation for utility approximation
  - Estimator between negotiation algorithm and DM
  - Acts like a smart cache

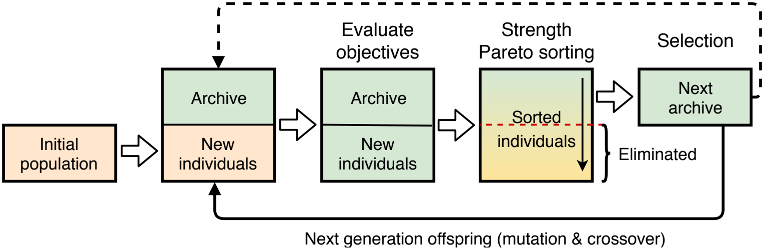
# Approximation in the overall infrastructure




- Improving negotiation for utility approximation
  - Estimator between negotiation algorithm and DM
  - Acts like a smart cache

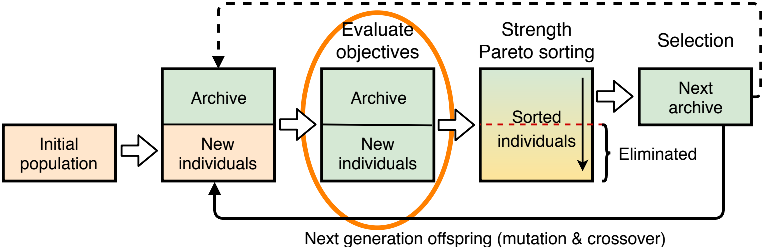
# Approach

# Integration of objective approximation




- Asynchronous approximation integration
  - Evaluation may be replaced by approximation
  - Mix of evaluated and approximated individuals

# Integration of objective approximation



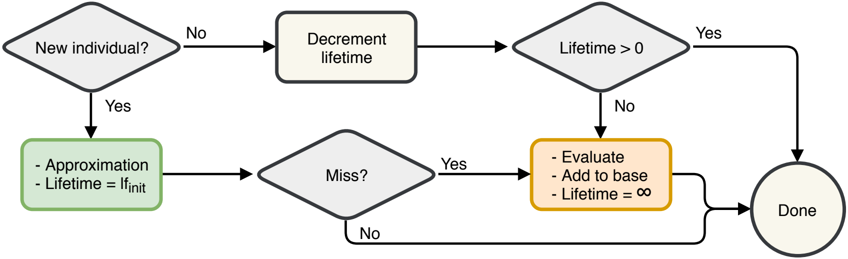

- Asynchronous approximation integration
  - Evaluation may be replaced by approximation
  - Mix of evaluated and approximated individuals

# Integration of objective approximation (2)

## Attribution of objective values

- Lifetime associated to individuals
- Evaluated if conserved until lifetime is zero (archive)
  - Added to knowledge base

For each individual:

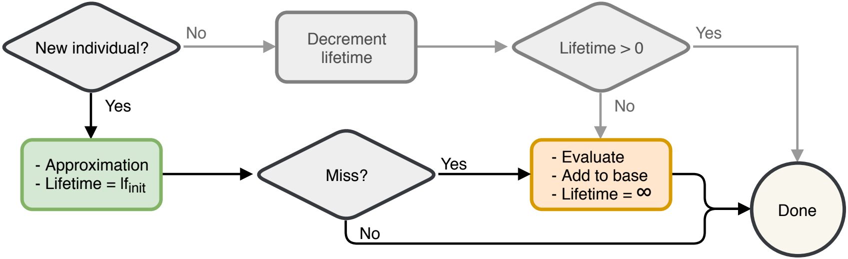


# Integration of objective approximation (2)

## Attribution of objective values

- Lifetime associated to individuals
- Evaluated if conserved until lifetime is zero (archive)
  - Added to knowledge base

For each individual:

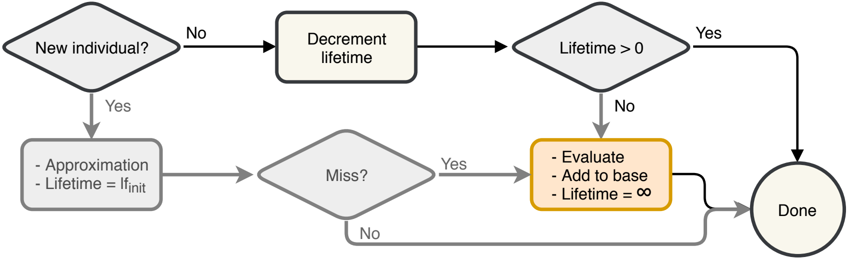


# Integration of objective approximation (2)

## Attribution of objective values

- Lifetime associated to individuals
- Evaluated if conserved until lifetime is zero (archive)
  - Added to knowledge base

For each individual:

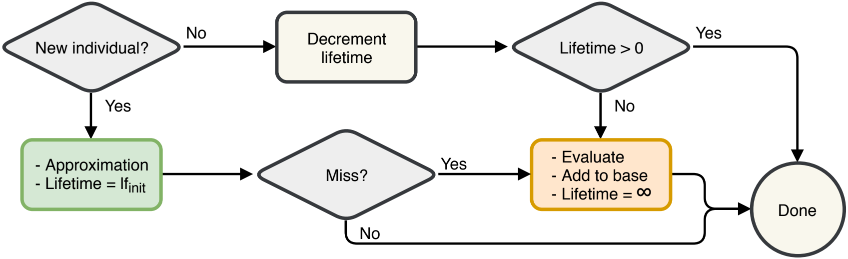


# Integration of objective approximation (2)

## Attribution of objective values

- Lifetime associated to individuals
- Evaluated if conserved until lifetime is zero (archive)
  - Added to knowledge base

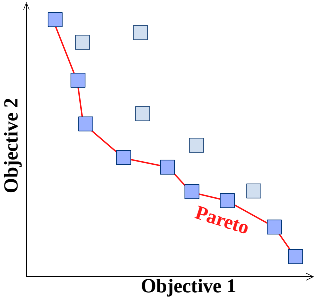
For each individual:



# Limitations of SPEA2

## SPEA2 not well adapted to asynchronous approximation

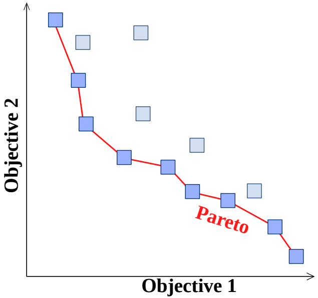
- Limited archive of individuals
- Bad (optimistic) approximations → good solutions lost
- Still approximated solutions after ending condition reached



# Limitations of SPEA2

## SPEA2 not well adapted to asynchronous approximation

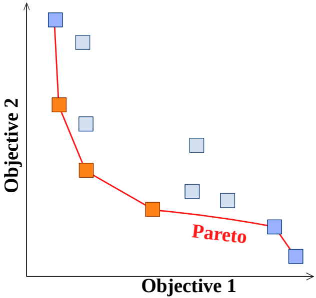
- Limited archive of individuals
- Bad (optimistic) approximations $\rightarrow$ good solutions lost
- Still approximated solutions after ending condition reached



# Limitations of SPEA2

## SPEA2 not well adapted to asynchronous approximation

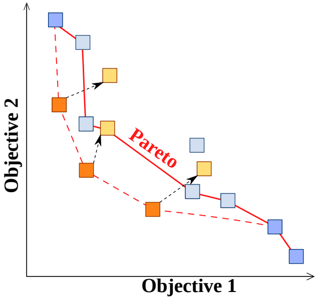
- Limited archive of individuals
- Bad (optimistic) approximations → good solutions lost
- Still approximated solutions after ending condition reached



# Limitations of SPEA2

## SPEA2 not well adapted to asynchronous approximation

- Limited archive of individuals
- Bad (optimistic) approximations → good solutions lost
- Still approximated solutions after ending condition reached



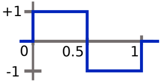
# Uncertain-SPEA2 (USPEA2)

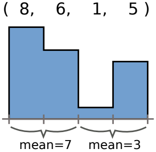
Modify SPEA2 to manage uncertain solutions (approximations)

- Add an archive of evaluated solutions (certain archive)
- Avoiding duplication of individuals

Stopping USPEA2 at any time results in a set of valid solutions.

# Overview



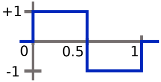



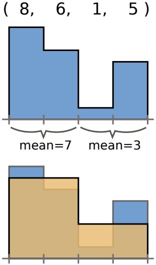

## Haar wavelet transform

Extract frequency and temporal features of a signal.

- Fast to compute
- Works well with discrete series
- $\approx$ successive mean between time steps
- Conserve euclidean distances

# Overview



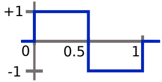



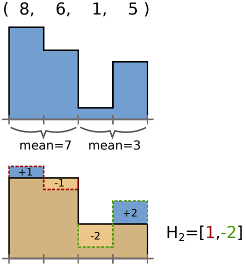

## Haar wavelet transform

Extract frequency and temporal features of a signal.

- Fast to compute
- Works well with discrete series
- ≈ successive mean between time steps
- Conserve euclidean distances

# Overview



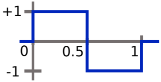



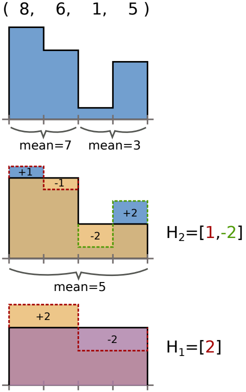

## Haar wavelet transform

Extract frequency and temporal features of a signal.

- Fast to compute
- Works well with discrete series
- $\approx$ successive mean between time steps
- Conserve euclidean distances

# Overview



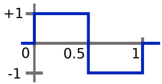




## Haar wavelet transform

Extract frequency and temporal features of a signal.

- Fast to compute
- Works well with discrete series
- ≈ successive mean between time steps
- Conserve euclidean distances

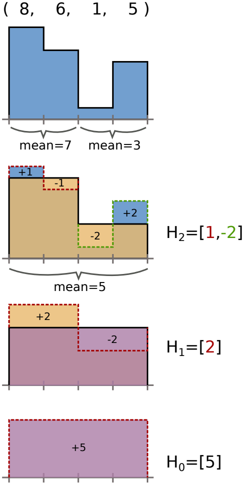
# Overview



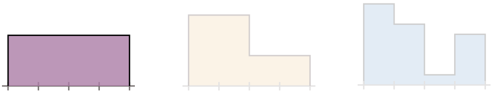

## Haar wavelet transform

Extract frequency and temporal features of a signal.

- Fast to compute
- Works well with discrete series
- ≈ successive mean between time steps
- Conserve euclidean distances

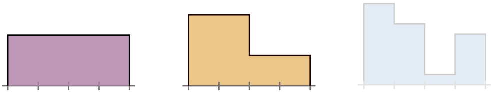


# Multi-resolution Haar approximation



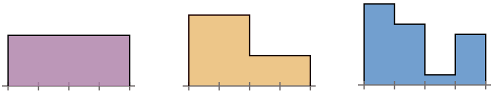
- Distance between partial Haar representations from known solutions
  - Lowest frequencies features first
- Select known solutions closer than a threshold
- If enough solutions: repeat with higher frequency
- Result: weighted average of close utilities
- Complexity: $O(n \log(n))$ ($n$ solutions in base)

# Multi-resolution Haar approximation



- Distance between partial Haar representations from known solutions
  - Lowest frequencies features first
- Select known solutions closer than a threshold
- If enough solutions: repeat with higher frequency
- Result: weighted average of close utilities
- Complexity: $O(n\log(n))$ ($n$ solutions in base)

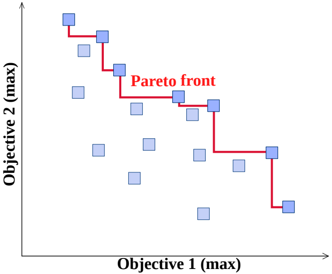
# Multi-resolution Haar approximation



- Distance between partial Haar representations from known solutions
  - Lowest frequencies features first
- Select known solutions closer than a threshold
- If enough solutions: repeat with higher frequency
- Result: weighted average of close utilities
- Complexity: $O(n \log(n))$ ($n$ solutions in base)

# Methodology and evaluation

# Quality indicators




## Hypervolume indicator

Area covered between Pareto front of solution set and any reference point.

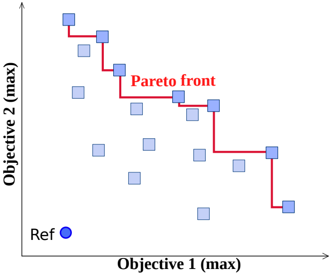
- $\geq$ if solutions are better (dominate)
- $\geq$ if solution set is more spread

## Generational distance

Average distance between approximation front and best known Pareto front

- Requires a good comparison set

# Quality indicators




## Hypervolume indicator

Area covered between Pareto front of solution set and any reference point.

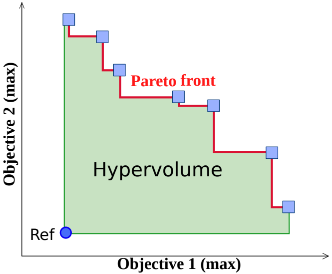
- $\geq$ if solutions are better (dominate)
- $\geq$ if solution set is more spread

## Generational distance

Average distance between approximation front and best known Pareto front

- Requires a good comparison set

# Quality indicators




## Hypervolume indicator

Area covered between Pareto front of solution set and any reference point.

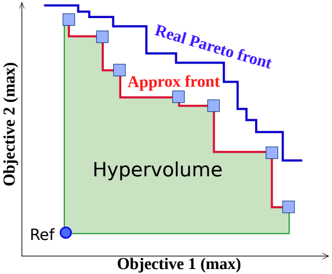
- $\geq$ if solutions are better (dominate)
- $\geq$ if solution set is more spread

## Generational distance

Average distance between approximation front and best known Pareto front

- Requires a good comparison set

# Quality indicators




## Hypervolume indicator

Area covered between Pareto front of solution set and any reference point.

- $\geq$ if solutions are better (dominate)
- $\geq$ if solution set is more spread

## Generational distance

Average distance between approximation front and best known Pareto front

- Requires a good comparison set

# Infrastructure and decision modules

Simplified models, keep optimum computable

## IT decision module

- «Fluid» workload: total amount of CPU time
- Utility: revenue
  - Reward for each unit scheduled
  - Incentive to execute unit early

## Electrical decision module

- Solar panels, batteries, electrical grid in/out
- Utility: equivalent CO2 emission
  - Zero for renewable
  - Grid electricity average emission
  - Battery aging, based on manufacturing cost

# Infrastructure and decision modules

Simplified models, keep optimum computable

## IT decision module

- «Fluid» workload: total amount of CPU time
- Utility: revenue
  - Reward for each unit scheduled
  - Incentive to execute unit early

## Electrical decision module

- Solar panels, batteries, electrical grid in/out
- Utility: equivalent CO2 emission
  - Zero for renewable
  - Grid electricity average emission
  - Battery aging, based on manufacturing cost

# Evaluation

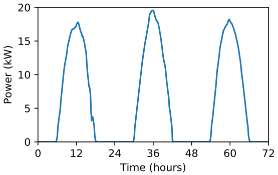
- 3 days scenarios
- Workload: 75% of maximum data center capacity
  - ExcessRenew: sunny days, initial battery 50%
  - Normal: less sunny days
  - FewRenew: almost no sun, initial battery at 25%




- Optimal formulation → comparison Pareto front
- (U)SPEA2 ending condition: budget of utility evaluations

# Evaluation

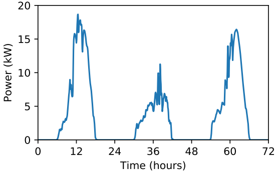
- 3 days scenarios
- Workload: 75% of maximum data center capacity
  - ExcessRenew: sunny days, initial battery 50%
  - Normal: less sunny days
  - FewRenew: almost no sun, initial battery at 25%




- Optimal formulation → comparison Pareto front
- (U)SPEA2 ending condition: budget of utility evaluations

# Evaluation

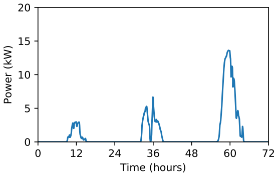
- 3 days scenarios
- Workload: 75% of maximum data center capacity
  - ExcessRenew: sunny days, initial battery 50%
  - Normal: less sunny days
  - FewRenew: almost no sun, initial battery at 25%




- Optimal formulation → comparison Pareto front
- (U)SPEA2 ending condition: budget of utility evaluations

# Evaluation

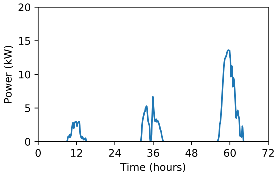
- 3 days scenarios
- Workload: 75% of maximum data center capacity
  - ExcessRenew: sunny days, initial battery 50%
  - Normal: less sunny days
  - FewRenew: almost no sun, initial battery at 25%




- Optimal formulation → comparison Pareto front
- (U)SPEA2 ending condition: budget of utility evaluations

# Evaluation

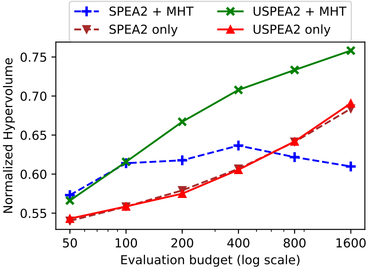
- 3 days scenarios
- Workload: 75% of maximum data center capacity
  - ExcessRenew: sunny days, initial battery 50%
  - Normal: less sunny days
  - FewRenew: almost no sun, initial battery at 25%

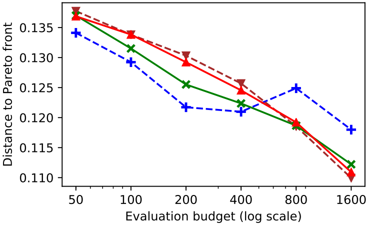



- Optimal formulation → comparison Pareto front
- (U)SPEA2 ending condition: budget of utility evaluations

# Budget of evaluations

Scenario *Normal*, 80 time steps

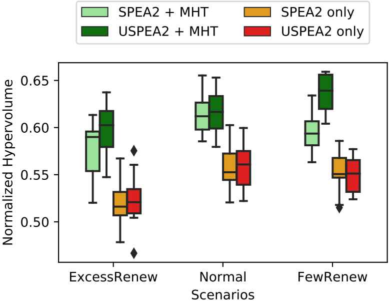


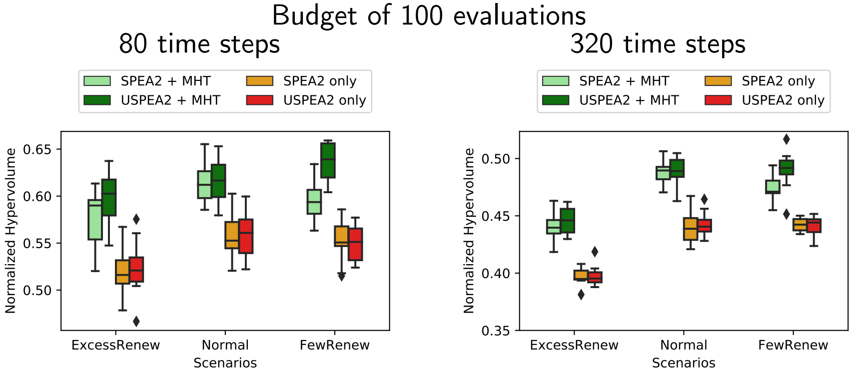
# Scenarios and number of time steps

Budget of 100 evaluations

80 time steps

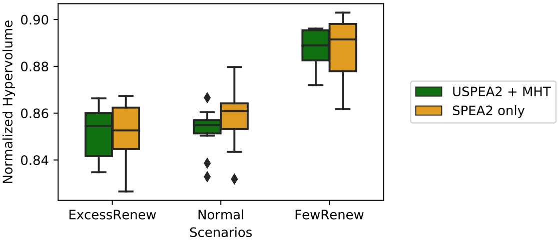


# Scenarios and number of time steps



# Some unexpected results

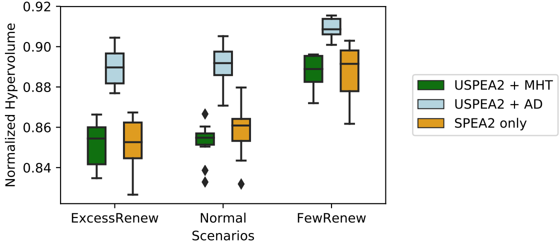
Initial profiles: Best profiles from each DM




- Approximation method AD: naive and usually ≈ baseline

# Some unexpected results

Initial profiles: Best profiles from each DM




- Approximation method AD: naive and usually ≈ baseline

# Conclusion

# Conclusion

- Find set of trade-offs power plans
- Approximation of power profile utility valuable
  - More objective space covered
  - Similar hypervolume for $1/3^{rd}$ to $1/5^{th}$ evaluations
  - Difficult to predict performances in advance...
- USPEA2 ensure stable results with approximation
  - Similar to SPEA2 without approximation

## Future works

- Repeated planning policies: choosing a solution
- Better understanding of MOEA/approximation relationship

# Conclusion

- Find set of trade-offs power plans
- Approximation of power profile utility valuable
  - More objective space covered
  - Similar hypervolume for $1/3^{rd}$ to $1/5^{th}$ evaluations
  - Difficult to predict performances in advance...
- USPEA2 ensure stable results with approximation
  - Similar to SPEA2 without approximation

## Future works

- Repeated planning policies: choosing a solution
- Better understanding of MOEA/approximation relationship

# Conclusion

- Find set of trade-offs power plans
- Approximation of power profile utility valuable
  - More objective space covered
  - Similar hypervolume for $1/3^{rd}$ to $1/5^{th}$ evaluations
  - Difficult to predict performances in advance...
- USPEA2 ensure stable results with approximation
  - Similar to SPEA2 without approximation

## Future works

- Repeated planning policies: choosing a solution
- Better understanding of MOEA/approximation relationship

# Questions?