Statistical Learning for Power Systems Optimization: An Active Set Approach

Line A. Roald (UW Madison)

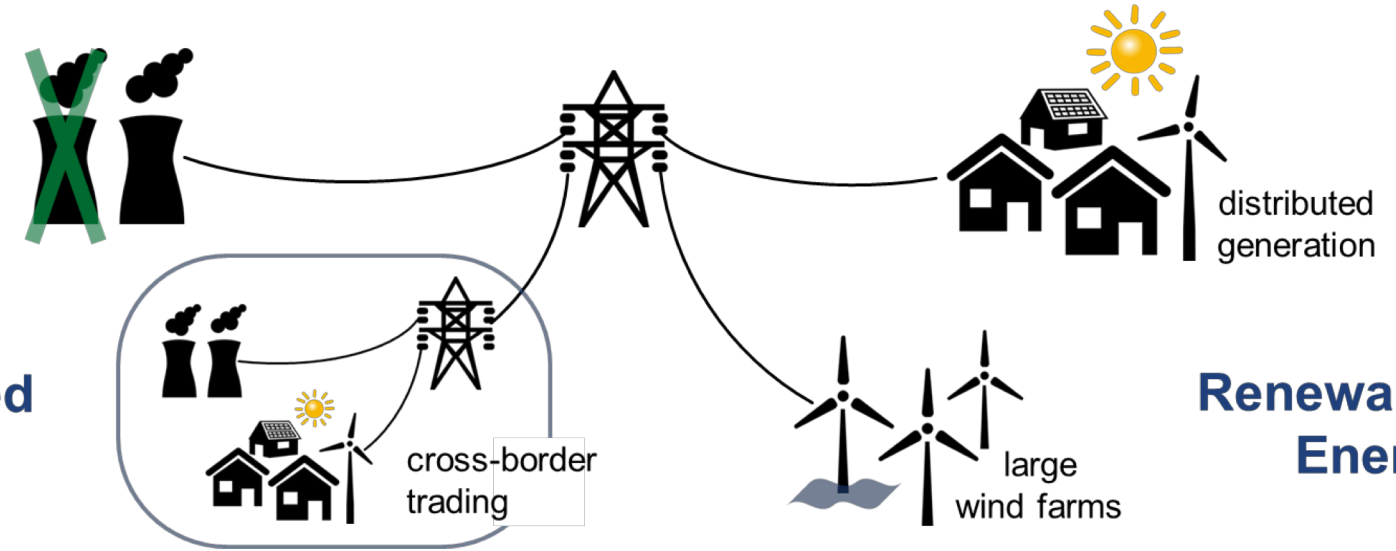
with **Sidhant Misra** (LANL)

and **Yee Sian Ng** (MIT)

Yee Sian Ng, Sidhant Misra, Line Roald and Scott Backhaus, «Statistical Learning for DC Optimal Power Flow», Power System Computation Conference (PSCC), 2018

Sidhant Misra, Line Roald and Yee Sian Ng, «Learning for Constrained Optimization», submitted, available online: https://arxiv.org/abs/1802.09639

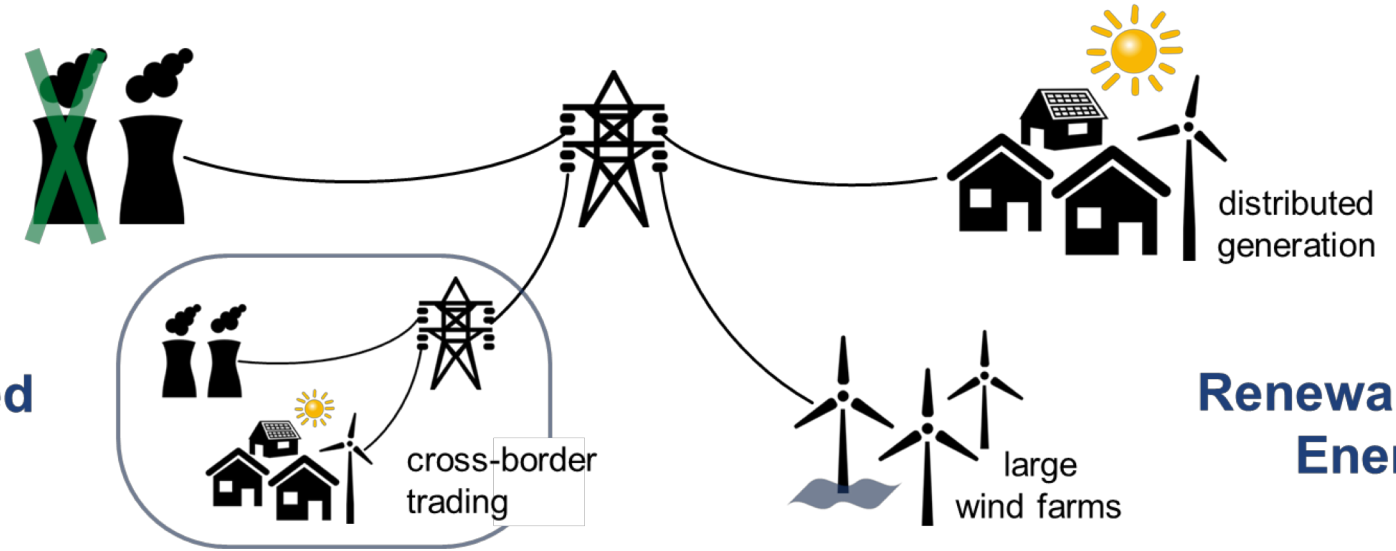
Transmission System Operation



Liberalized Markets

Renewable Energy

Transmission System Operation

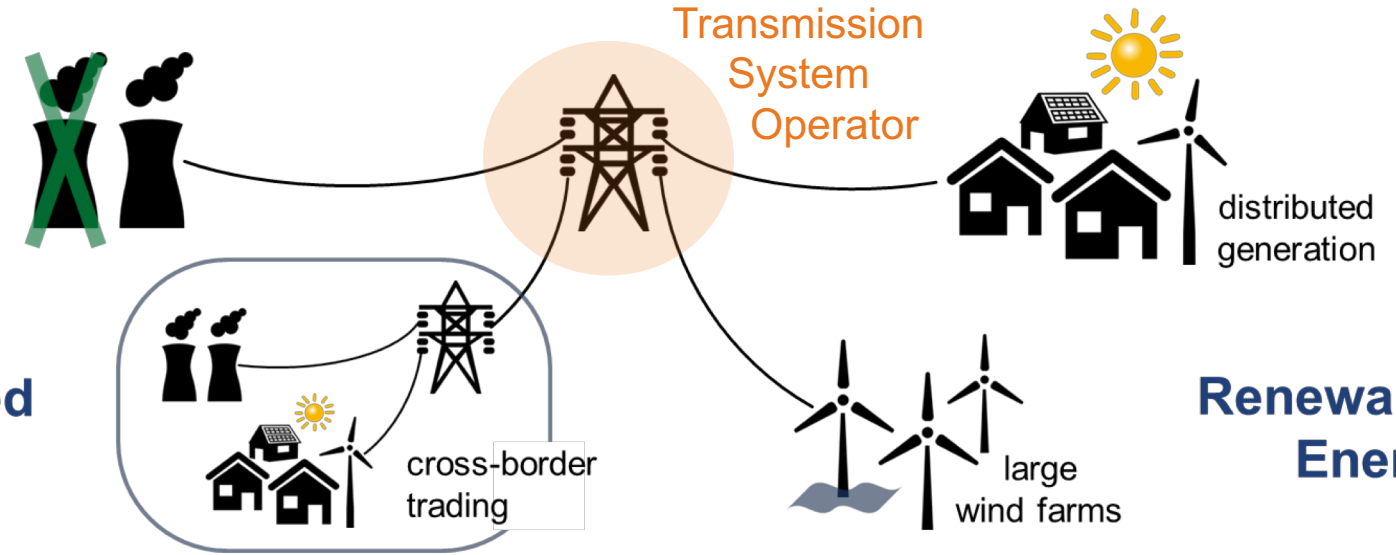


Liberalized Markets

Renewable Energy

- Higher and frequently changing power flows
- Increased uncertainty

Transmission System Operation



Liberalized Markets

Renewable Energy

- Higher and frequently changing power flows
- Increased uncertainty

The Optimal Power Flow Problem

DC Optimal Power Flow

Goal: Low cost operation, while enforcing technical limits

$$\min_{P_G} \quad C_G^T P_G$$

minimize generation cost

$$\text{s.t.} \quad \sum_{i=1}^{N_B} \left(P_{G(i)} + P_{W(i)} - P_{D(i)} \right) = 0$$

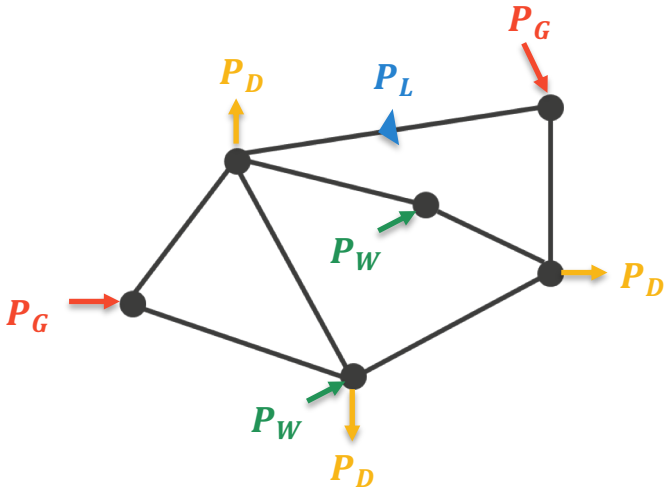
balanced operation

$$P_{G(g)} \leq P_{G(g)}^{max},$$
$$P_{G(g)} \geq P_{G(g)}^{min},$$
$$\forall\, g = 1, \ldots, N_G$$

generator limits

$$A_{(l,\cdot)}(P_G + P_W - P_D) \leq P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G + P_W - P_D) \geq -P_{L(l)}^{max},$$
$$\forall\, l = 1, \ldots, N_L$$

transmission line limits



DC Optimal Power Flow

Renewable energy/load is changing!

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

minimize generation cost

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

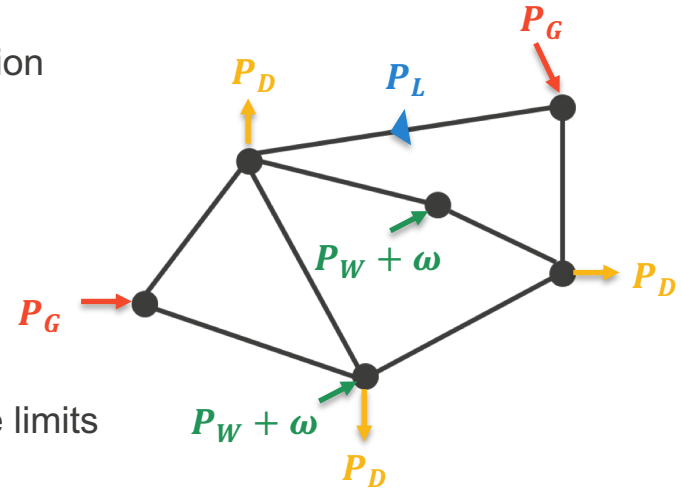
balanced operation

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

generator limits

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

transmission line limits



DC Optimal Power Flow

Renewable energy/load is changing! → Resolve problem for each new $\boldsymbol{\omega}$ to obtain $P_G^*(\boldsymbol{\omega})$ **every 5-15 min**

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

minimize generation cost

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

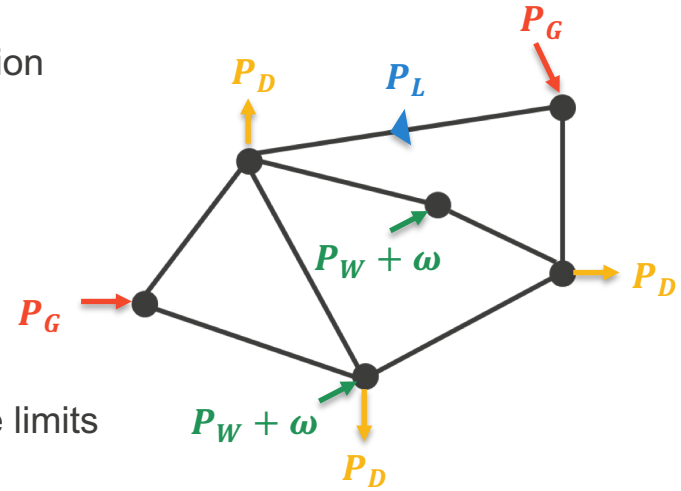
balanced operation

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

generator limits

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

transmission line limits



Repeated Solution Process

OPF at $T_1, \boldsymbol{\omega_1}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\ g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\ l = 1, \dots, N_L$$

Repeated Solution Process

OPF at T_1, $\boldsymbol{\omega_1}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at T_2, $\boldsymbol{\omega_2}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

Repeated Solution Process

OPF at $T_1, \boldsymbol{\omega_1}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_2, \boldsymbol{\omega_2}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_3, \boldsymbol{\omega_3}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

Repeated Solution Process

OPF at $T_1, \boldsymbol{\omega_1}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_2, \boldsymbol{\omega_2}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_3, \boldsymbol{\omega_3}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

s.t. $\sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

.....

Repeated Solution Process

OPF at $T_1, \boldsymbol{\omega_1}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_2, \boldsymbol{\omega_2}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

OPF at $T_3, \boldsymbol{\omega_3}$

$$\min_{P_G(\boldsymbol{\omega})} \quad C_G^T P_G(\boldsymbol{\omega})$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B}\left(P_{G(i)}(\boldsymbol{\omega}) + P_{W(i)} + \boldsymbol{\omega}_{(i)} - P_{D(i)}\right) = 0$$

$$P_{G(g)}(\boldsymbol{\omega}) \le P_{G(g)}^{max},$$
$$P_{G(g)}(\boldsymbol{\omega}) \ge P_{G(g)}^{min},$$
$$\forall\, g = 1, \dots, N_G$$

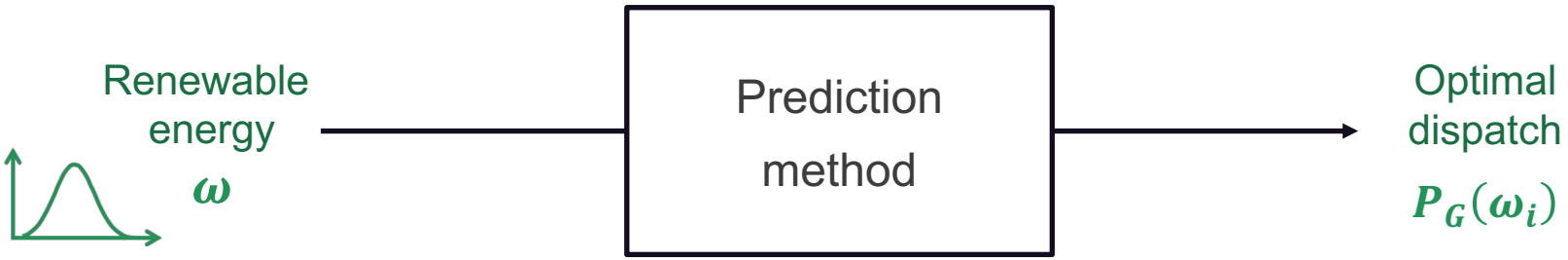
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \le P_{L(l)}^{max},$$
$$A_{(l,\cdot)}(P_G(\boldsymbol{\omega}) + P_W + \boldsymbol{\omega} - P_D) \ge -P_{L(l)}^{max},$$
$$\forall\, l = 1, \dots, N_L$$

.....

Can we use learning to speed up the solution process

by using information from past solutions $\left(\boldsymbol{\omega_i}, \boldsymbol{P(\omega_i)}\right)$?

Repeated Solution Process



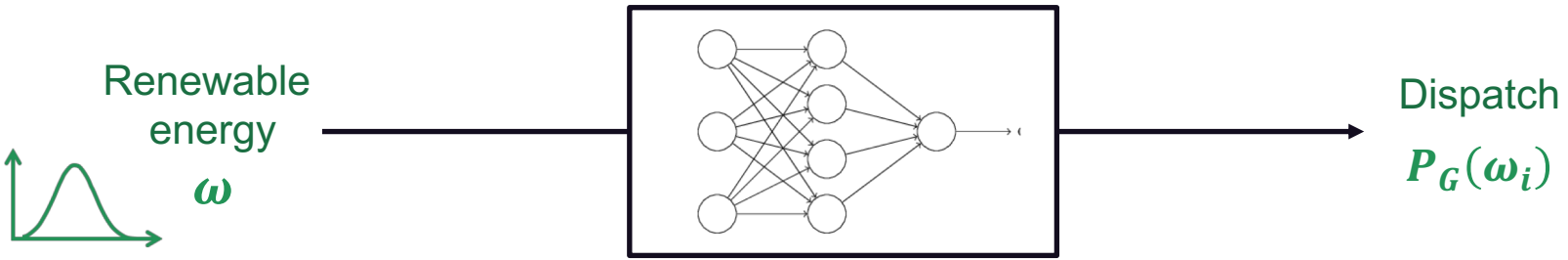
Can we use learning to speed up the solution process

by using information from past solutions $(\omega_i, P(\omega_i))$?

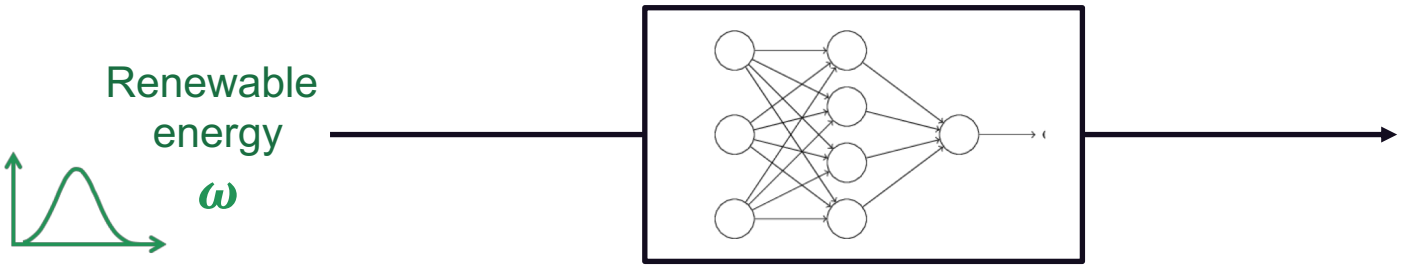
First attempt:

Train a neural net

First Attempt – Train a Neural Net



First Attempt – Train a Neural Net

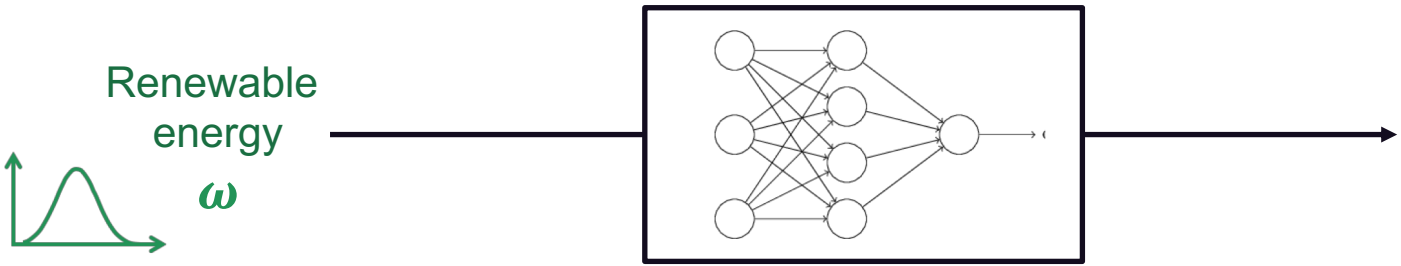




- **This didn't work well...**

(though I will admit that we gave up quite fast)

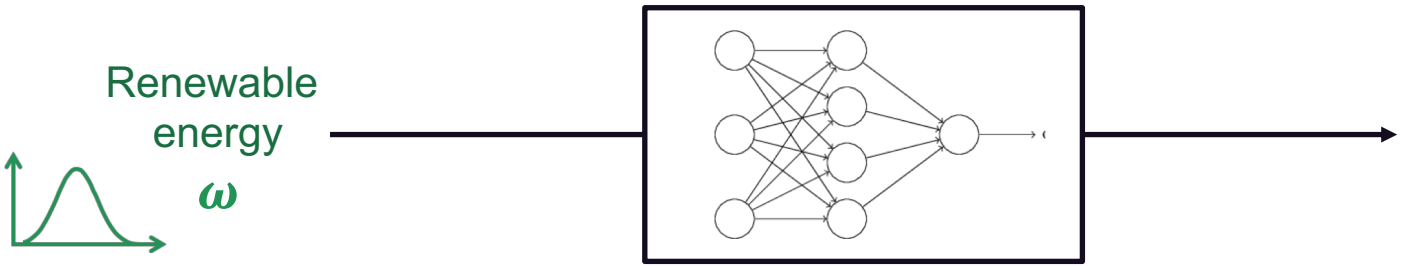
First Attempt – Train a Neural Net





- **This didn't work well...**
  - Hard to satisfy safety constraints!

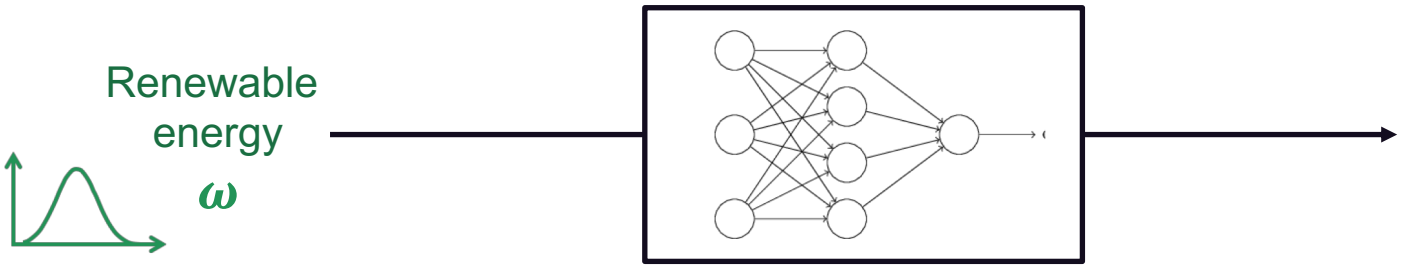
First Attempt – Train a Neural Net





- **This didn't work well...**
  - Hard to satisfy safety constraints!
  - Projection back onto feasible space cause suboptimality…

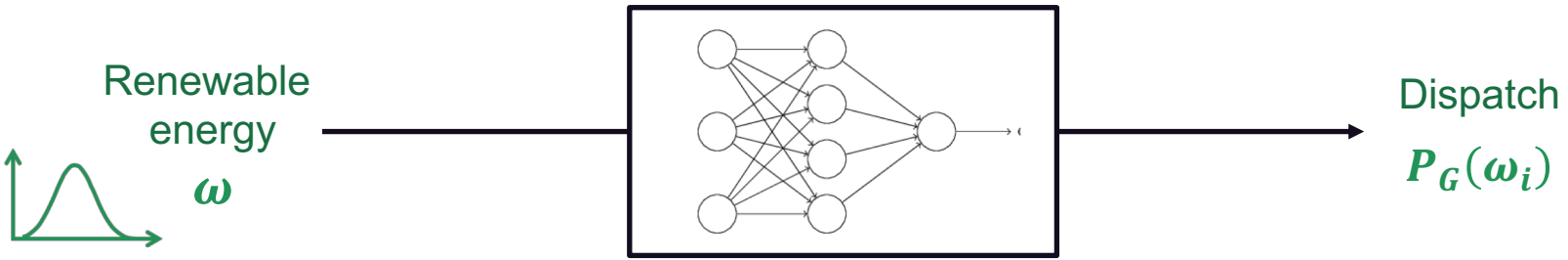
First Attempt – Train a Neural Net





- **This didn't work well...**
  - Hard to satisfy safety constraints!
  - Projection back onto feasible space cause suboptimality…
  - Challenging: High-dimensional input → High dimensional output

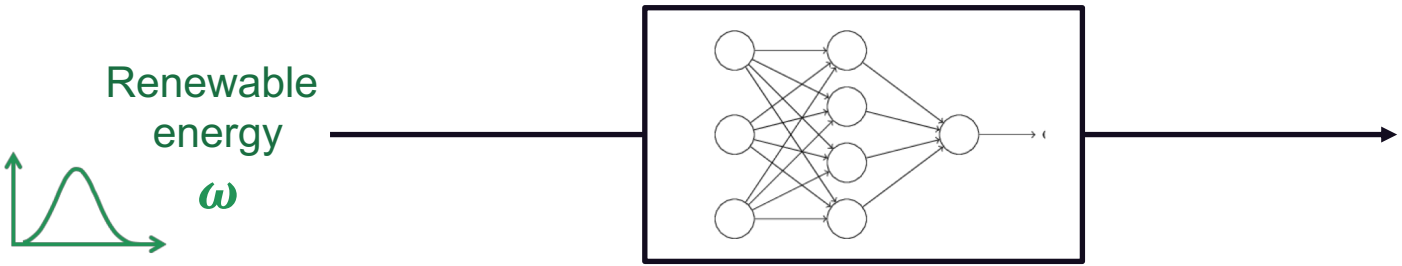
First Attempt – Train a Neural Net





- **This didn't work well...**
  - Hard to satisfy safety constraints!
  - Projection back onto feasible space cause suboptimality…
  - Challenging: High-dimensional input → High dimensional output
- **This can work well under some circumstances** Wide enough and deep enough [Karg and Lucia, 2018]

First Attempt – Train a Neural Net





We have a **mathematical optimization problem**

– can we use **more information** about the **problem structure**?

Think again:

How can we leverage pre-existing knowledge about the solution?

Think again:

How can we leverage pre-existing knowledge about the solution?

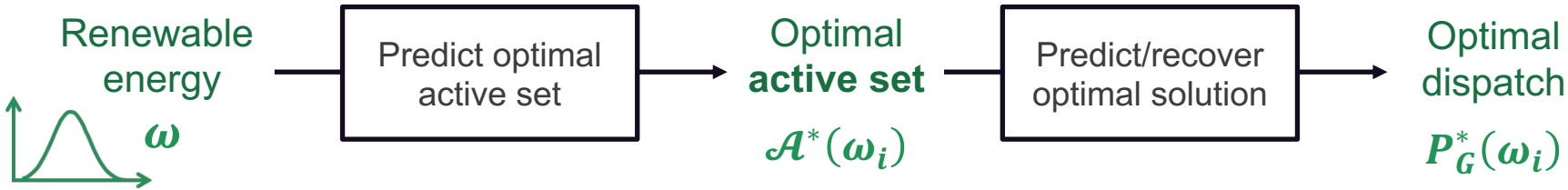
New idea:
Learn the optimal set of active constraints!

Optimal set of active constraints

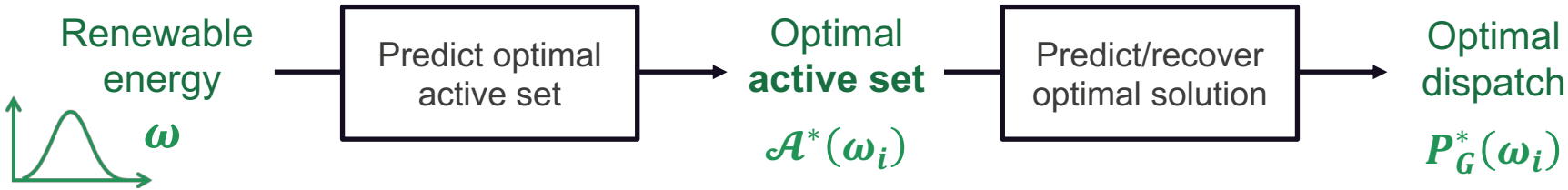
Set of constraints that are active at optimum!

- Power flow (equality) constraints are **always** active
- Only **very few** of the inequality constraints are active
  - Generator upper and lower bounds = “production pattern”
  - Transmission line constraints = “congestion pattern”
- **Together represent an operational pattern**

Learn optimal set of active constraints

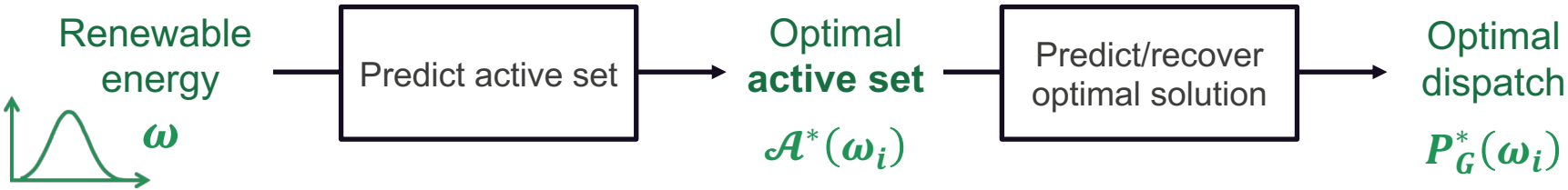


Learn optimal set of active constraints



- **Why?**
  - Optimal active set is the “minimal” information we need to recover optimal solution
  - Inherently encodes information about physical constraints and technical limits
  - Finite, low dimensional object
  - Nice physical interpretation (power system operational pattern)

Learn optimal set of active constraints



- **Related applications**
  - LMP prediction – predict when LMPs change = optimal active set changes
  - Explicit MPC – each optimal active set corresponds to an optimal affine control policy
  - …

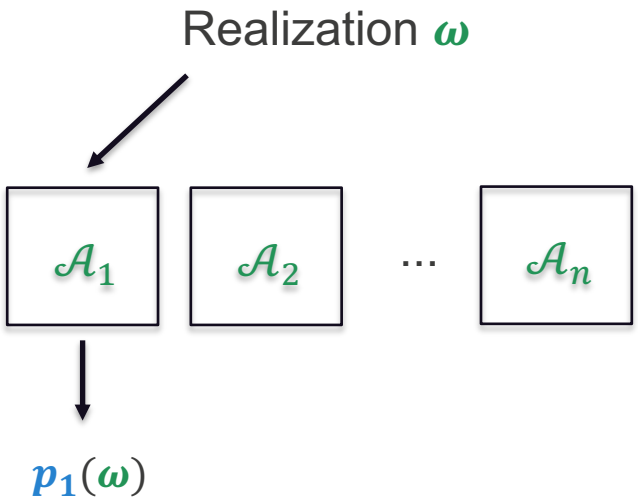
An Example - Ensemble policy

Realization ω



Candidates for optimal active set

Ensemble Policy



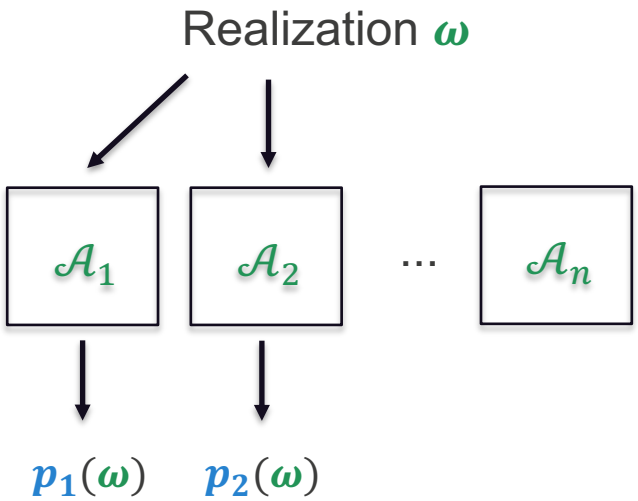
Candidates for
optimal active set

Solve problem given the active set

→ Solve a reduced problem with fewer constraints!

→ Solve a set of linear equations (linear problem)!

Ensemble Policy



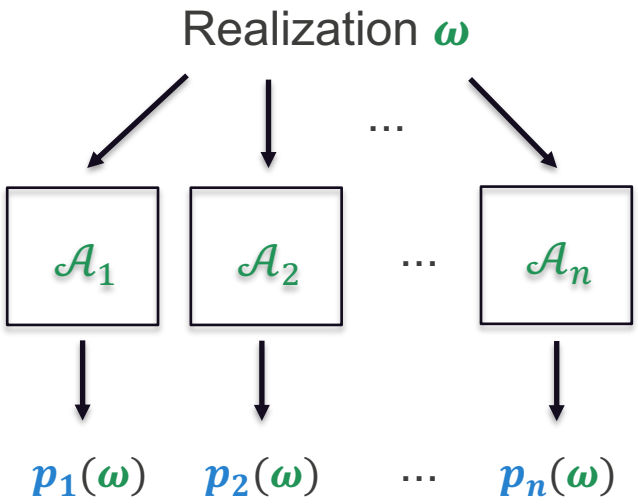
Candidates for
optimal active set

Solve problem given the active set

→ Solve a reduced problem with fewer constraints!

→ Solve a set of linear equations (linear problem)!

Ensemble Policy



Candidates for optimal active set

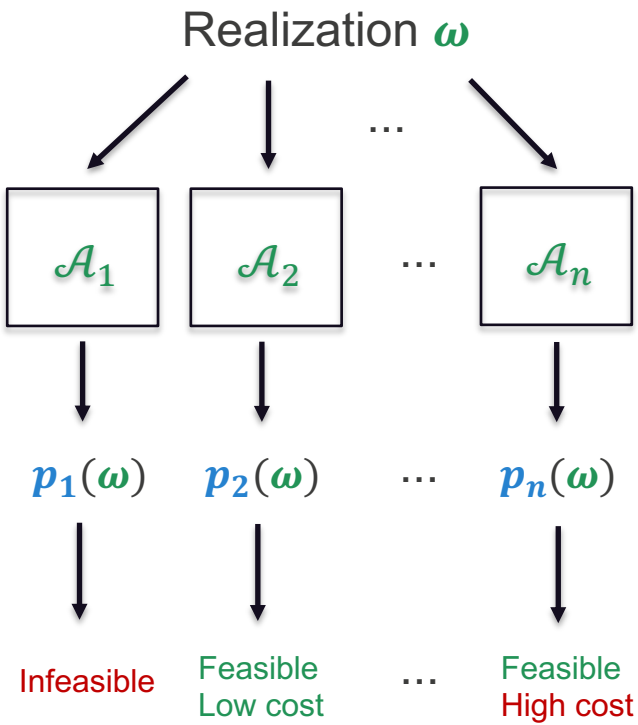
Solve problem given the active set

→ Solve a reduced problem with fewer constraints!

→ Solve a set of linear equations (linear problem)!

Easier than solving the full optimal power flow problem

Ensemble Policy

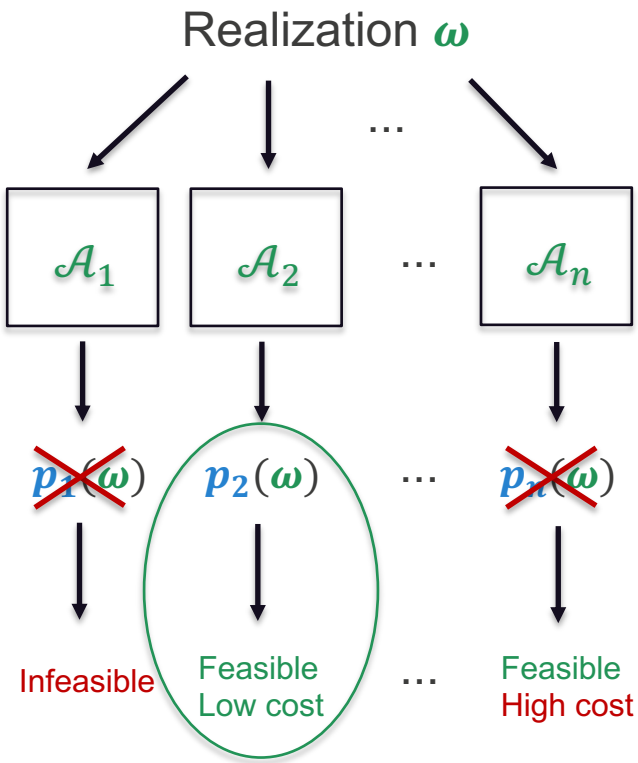


Candidates for optimal active set

Solve problem given the optimal active set

Evaluate cost and feasibility

Ensemble Policy



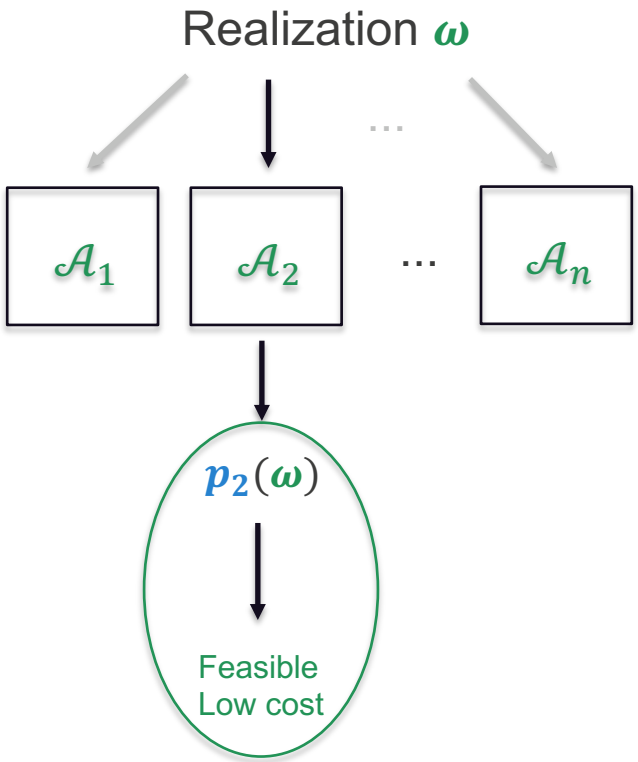
Candidates for optimal active set

Solve problem given the optimal active set

Evaluate cost and feasibility

Pick **best (optimal?)** solution

Limits of the Approach

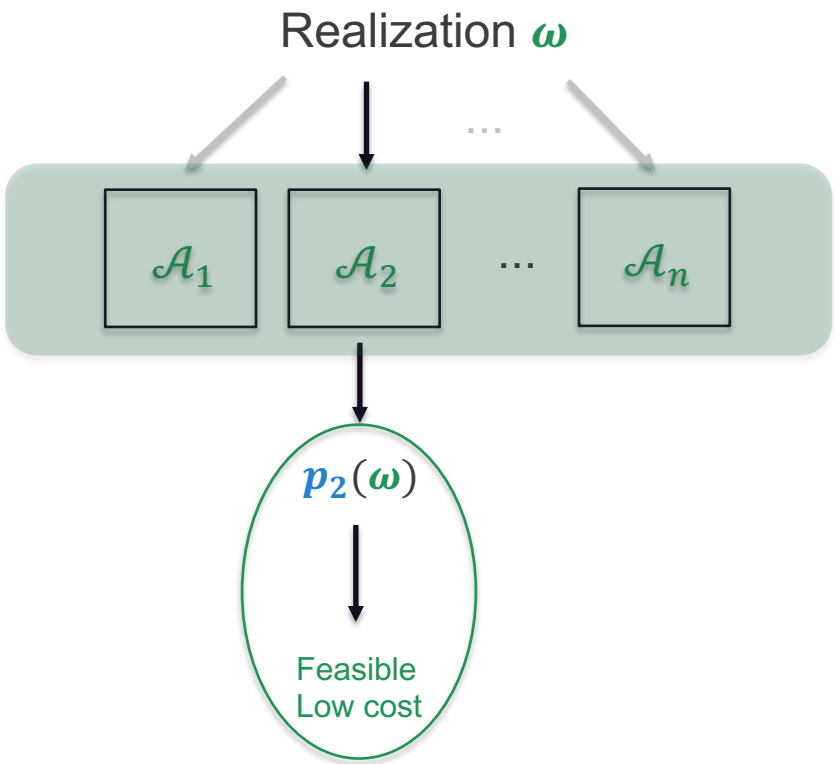


Works well if the number of active sets $\mathcal{A}_n$ is **small**!

Total number of possible active sets is exponential ☹

Maybe only a few are practically relevant? ☺

Limits of the Approach



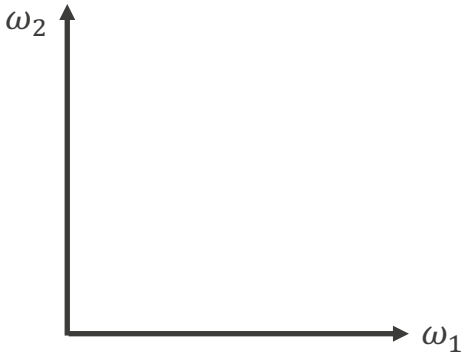
How to identify the collection of **relevant** active sets?

High probability that one of the active sets is optimal for a new realization ω

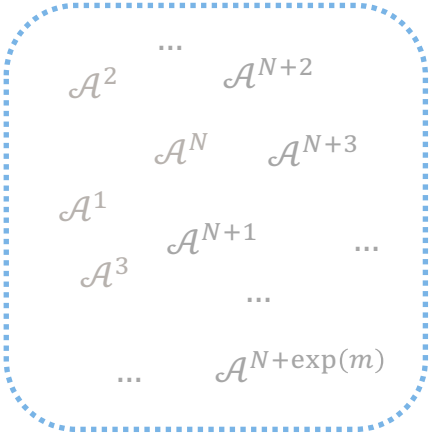
Do NOT search entire parameter space!

Using Sampling to Learn Important Active Sets

Learning Collection of Optimal Active Sets



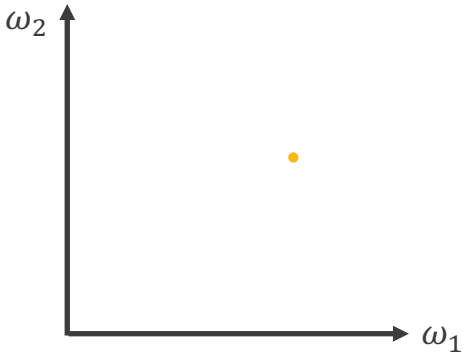
- samples of input parameters



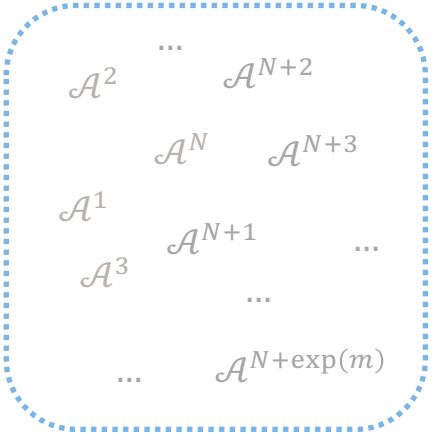
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Learning Collection of Optimal Active Sets



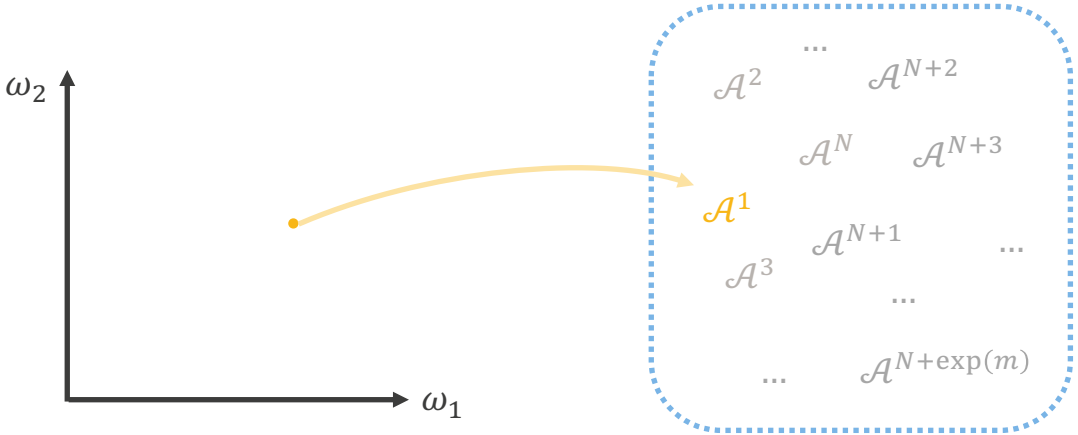
- samples of input parameters



all possible active sets

color: discovered active sets
grey: undiscovered active sets

Learning Collection of Optimal Active Sets

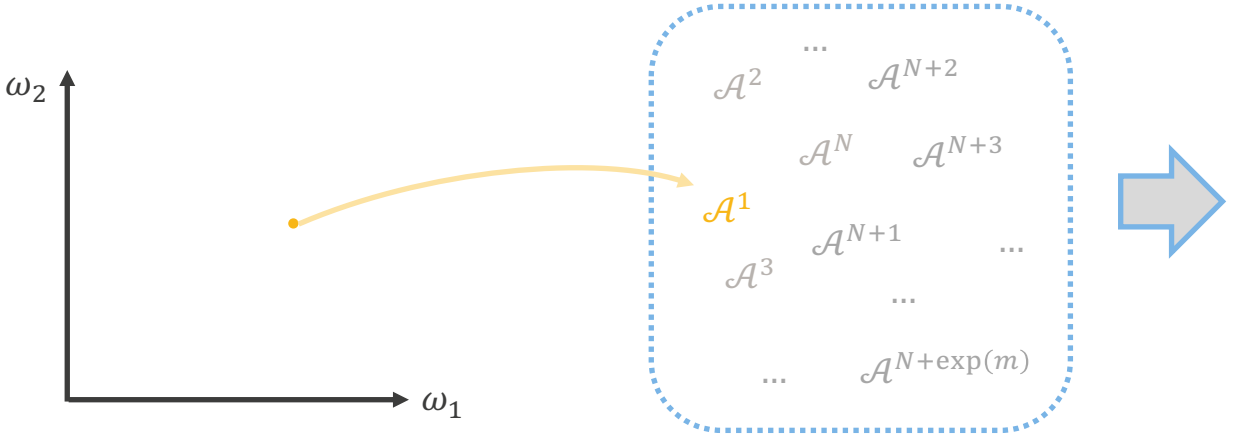


- samples of input parameters

all possible active sets

color: discovered active sets
grey: undiscovered active sets

Learning Collection of Optimal Active Sets



- samples of input parameters

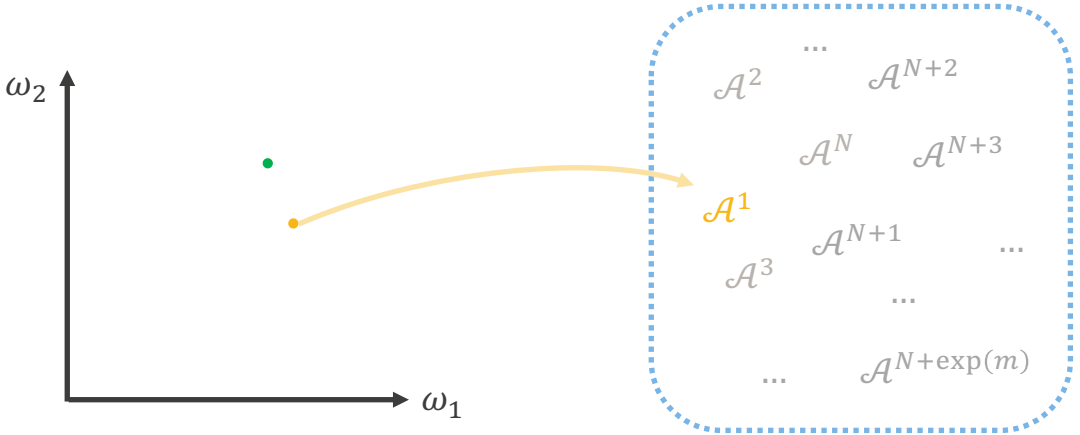
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

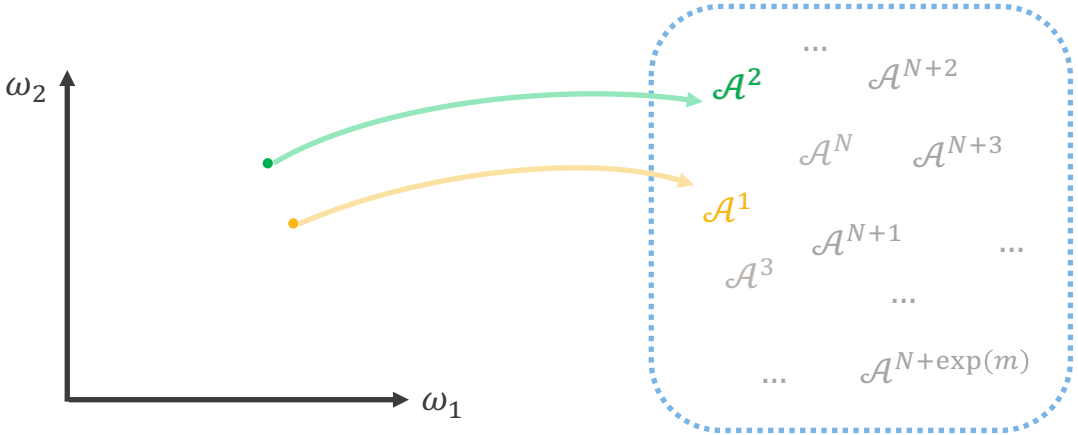
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

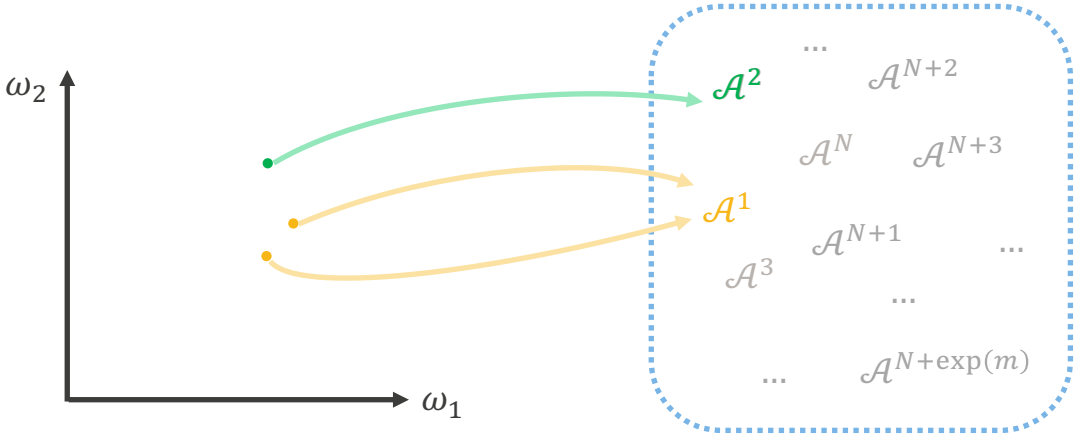
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

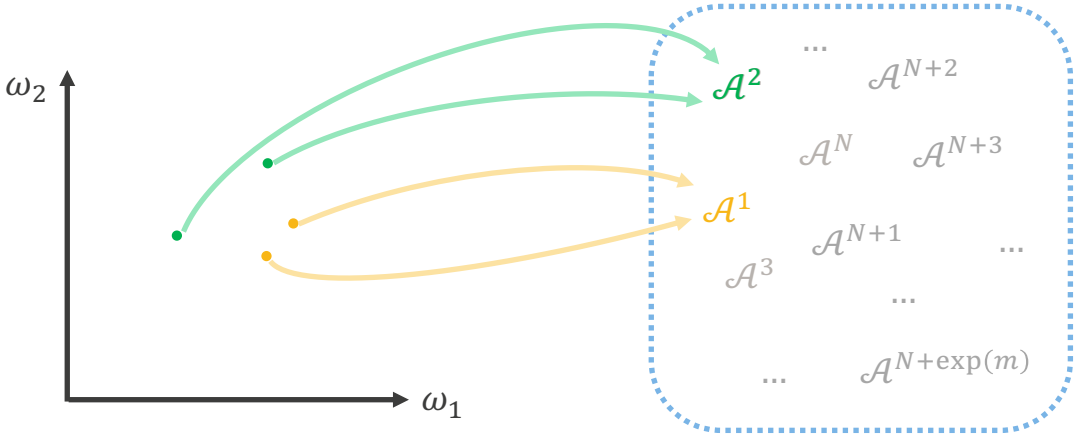
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

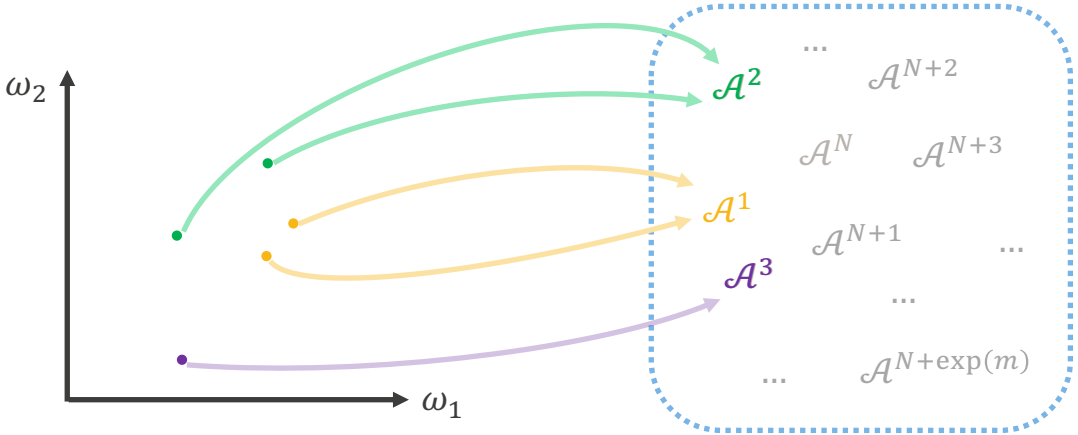
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

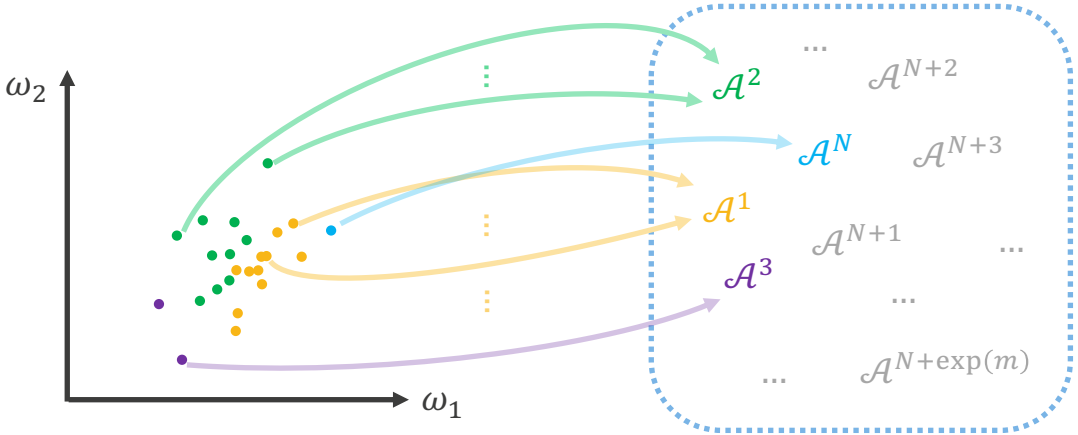
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2, \mathcal{A}^N\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

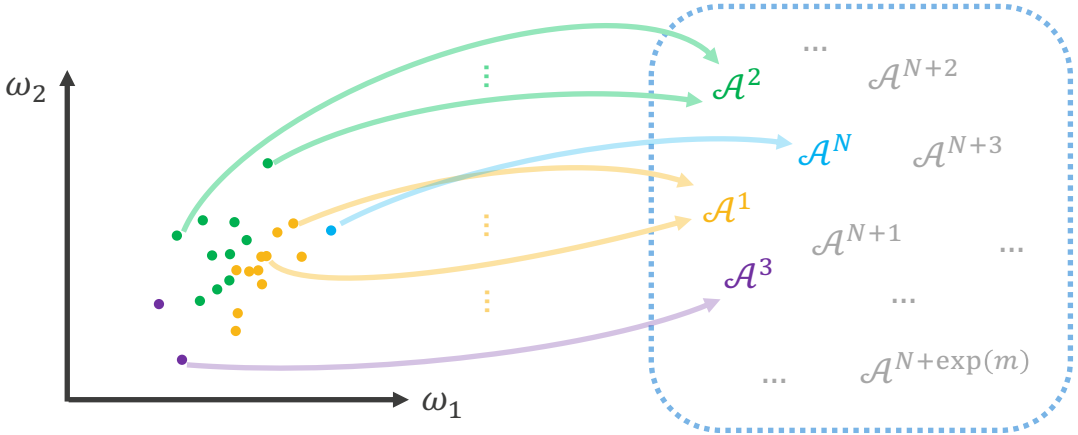
all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2, \ldots, \mathcal{A}^N\}$$

Learning Collection of Optimal Active Sets



- samples of input parameters

all possible active sets

color: discovered active sets
grey: undiscovered active sets

Collection of observed active sets

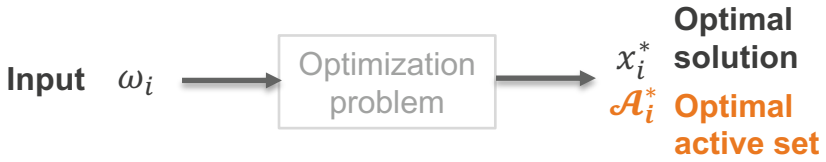
$$\mathcal{O}_M = \{\mathcal{A}^1, \mathcal{A}^2, \ldots, \mathcal{A}^N\}$$

When do I stop???

Streaming Algorithm to Learn Collection of Optimal Active Sets

Learning Collection of Optimal Active Sets

Training data:
Renewable energy realizations and corresponding active set $(\boldsymbol{\omega_i}, \boldsymbol{\mathcal{A}_i^*})$



Goal: Find a active sets that together have a high probability of being optimal!

Learning Collection of Optimal Active Sets

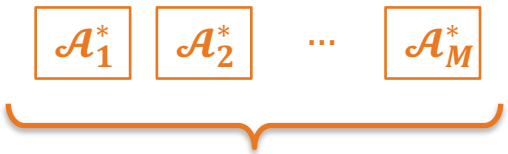
1. Observe optimal active sets for M samples



Collection of **observed** active sets

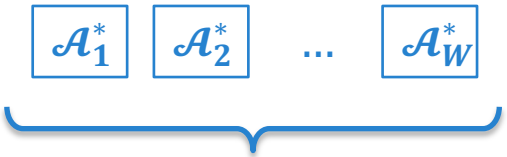
Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples

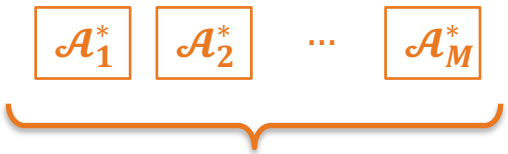


How frequently do we observe sets we have **not seen** before?

Rate of discovery: $\boldsymbol{R_{M,W}} = \frac{N_{unobserved}}{N}$

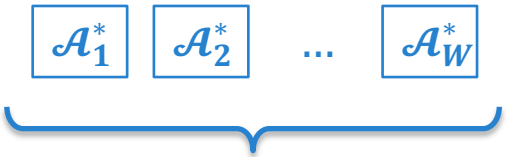
Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



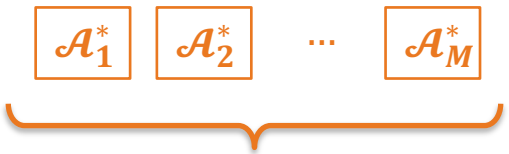
How frequently do we observe sets we have **not seen** before?

Rate of discovery: $\boldsymbol{R_{M,W}} = \frac{N_{unobserved}}{N}$

- If the rate of discovery is below a certain threshold, stop.

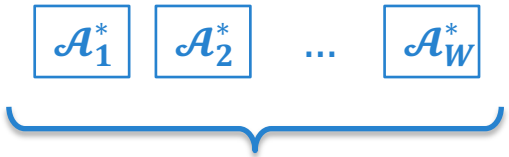
Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



How frequently do we observe sets we have **not seen** before?

Rate of discovery: $\boldsymbol{R_{M,W}} = \frac{N_{unobserved}}{N}$

- If the rate of discovery is below a certain threshold, stop.
- If the rate of discovery is too high, add more samples.

Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



How frequently do we observe sets we have **not seen** before?

Rate of discovery: $\boldsymbol{R_{M,W}} = \frac{N_{unobserved}}{N}$

- If the rate of discovery is below a certain threshold, stop.
- If the rate of discovery is too high, add more samples.

Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



How frequently do we observe sets we have **not seen** before?

- If the rate of discovery is below a certain threshold, stop.
- If the rate of discovery is too high, add more samples.

Correct threshold = performance guarantees

[Misra, Roald, Ng, 2018]

Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



How frequently do we observe sets we have **not seen** before?

- If the rate of discovery is below a certain threshold, stop.
- If the rate of discovery is too high, add more samples.

Correct threshold = performance guarantees

[Misra, Roald, Ng, 2018]

Guaranteed to converge!

Learning Collection of Optimal Active Sets

1. Observe optimal active sets for M samples



Collection of **observed** active sets

2. Check "rate of discovery" for W samples



How frequently do we observe sets we have **not seen** before?

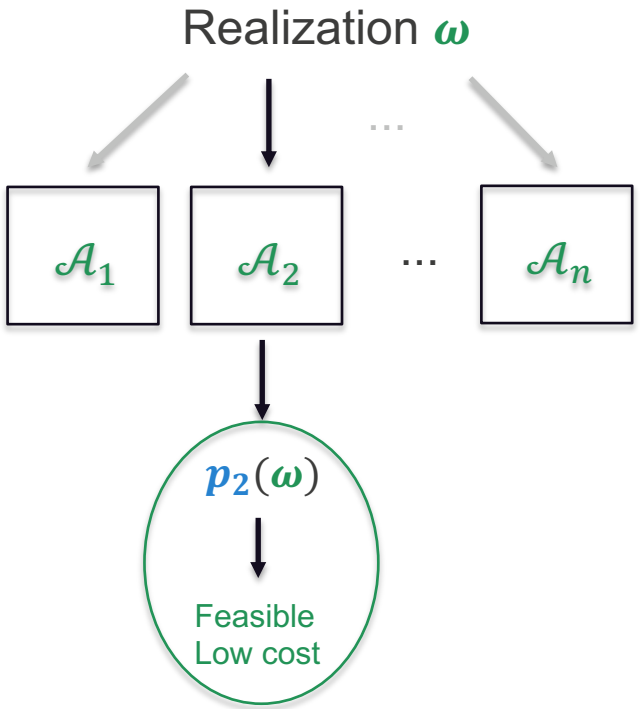
Correct threshold = performance guarantees

[Misra, Roald, Ng, 2018]

- If the rate of discovery is below a certain threshold, stop.
- If the rate of discovery is too high, add more samples.

Guaranteed to converge fast for systems with low number of optimal active sets!

Practicability of the approach



Simultaneously establishes

- Collection of optimal active sets
- Practicability of the approach

Guaranteed to converge fast for systems with low number of optimal active sets!

Results for the

Optimal Power Flow Problem

Streaming Algorithm Results – PGLib-OPF v 17.08

Probabilistic guarantee: $\mathbb{P}_{\omega}\left(\pi(\mathcal{U}_M)\right) < \alpha = 0.05,$

Max. difference: $\epsilon = 0.04$

Confidence level: $\delta = 0.01$

Hyperparameter: $\gamma = 2$

Termination: $R_{M,W} \leq 0.01$

$W = \frac{2\gamma}{\epsilon^2}\max\{\log M, \log \underline{M}\}$ with $\underline{M} = 1 + \left(\frac{\gamma}{\delta(\gamma-1)}\right)^{\frac{1}{\gamma-1}}$

Initial W: $W = 13'259$

(constant until $M = 201$)

Streaming Algorithm Results – PGLib-OPF v 17.08

Probabilistic guarantee: $\mathbb{P}_{\omega}\left(\pi(\mathcal{U}_M)\right) < \alpha = 0.05,$

Max. difference: $\epsilon = 0.04$

Confidence level: $\delta = 0.01$

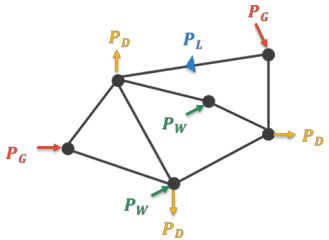
Hyperparameter: $\gamma = 2$

Termination: $R_{M,W} \leq 0.01$

$W = \frac{2\gamma}{\epsilon^2}\max\{\log M, \log \underline{M}\}$ with $\underline{M} = 1 + \left(\frac{\gamma}{\delta(\gamma-1)}\right)^{\frac{1}{\gamma-1}}$

Initial W: $W = 13'259$

(constant until $M = 201$)

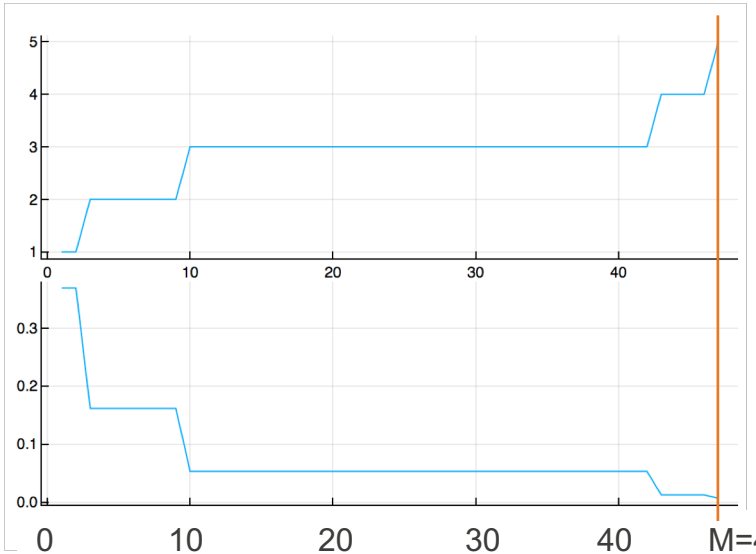


Uncertain loads:

Normal distribution
$\omega \sim \mathcal{N}(0, \sigma = 0.03d)$

Uniform distribution
with support
$\omega \in [-0.09d, 0.09d]$

Example – RTE 1951 bus test case

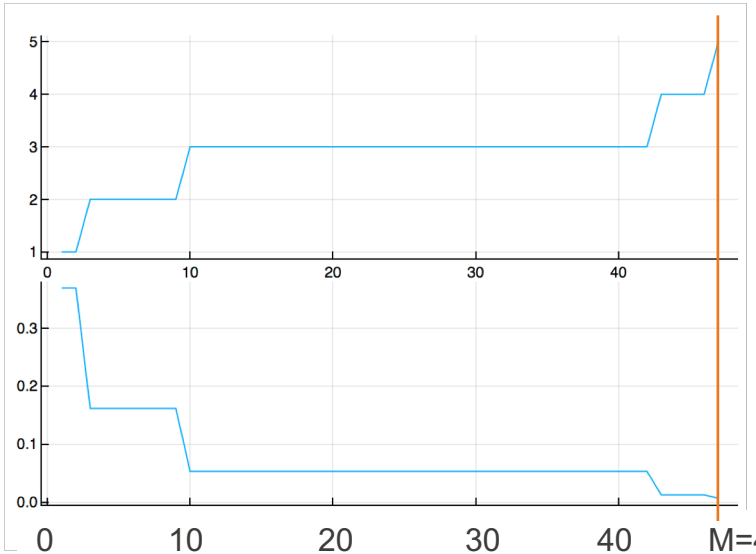


Number of active sets
5

Rate of discovery
0.0069

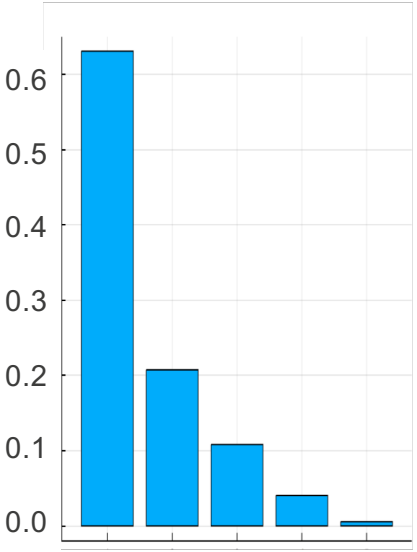
When there are few relevant active sets, the algorithm terminates fast!

Example – PSERC 200 bus test case



Number of active sets
5

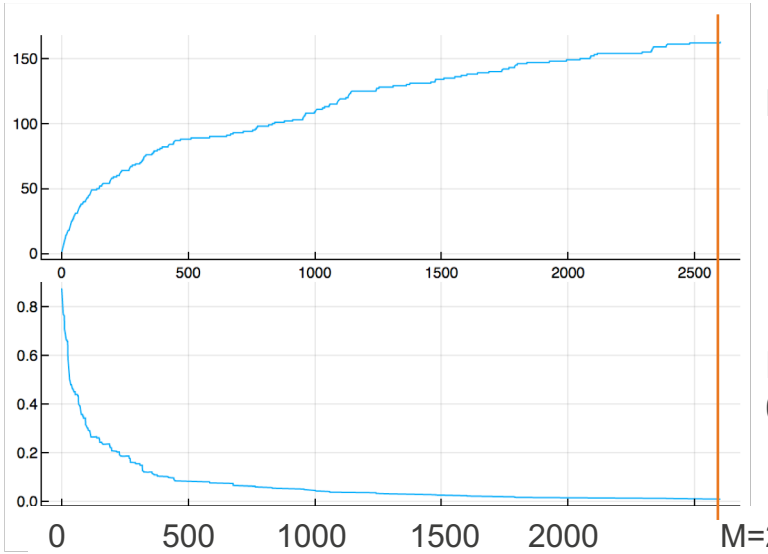
Rate of discovery
0.0069



Probability of observed sets

When there are few relevant active sets, the algorithm terminates fast!

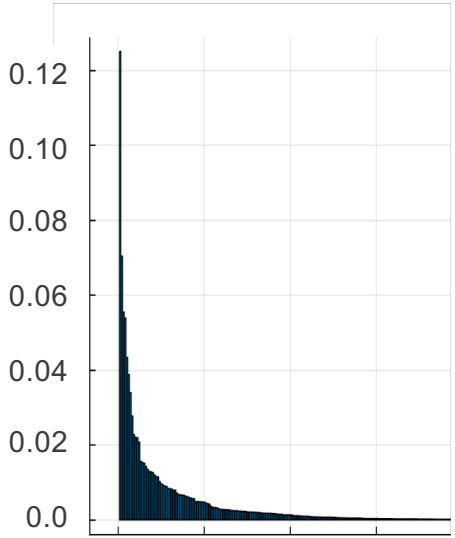
Example – PSERC 200 bus test case



Number of active sets
163

Rate of discovery
0.0099

M=2602
W=19’661



Probability of observed sets

When there are many relevant active sets, the algorithm terminates slowly!

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd										
case5_pjm										
case14_ieee										
case30_ieee										
case39_epri										
case118_ieee										
case57_ieee										
case1888_rte										
case1951_rte										
case162_ieee_dtc										
case24_ieee_rts										

System sizes ranging from 3 to 1951 nodes

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd										
case5_pjm										
case14_ieee										
case30_ieee										
case39_epri										
case118_ieee										
case57_ieee										
case1888_rte										
case1951_rte										
case162_ieee_dtc										
case24_ieee_rts										

Normal distribution
$\omega \sim \mathcal{N}(0, \sigma = 0.03d)$

Uniform distribution
with support
$\omega \in [-0.09d, 0.09d]$

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941

Few active sets!

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941

Few active sets!

Terminates fast.

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941

Few active sets!

Terminates fast.

High probability of optimal solutions!

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941
High-Complexity										
case73_ieee_rts	19	1258	17'844	0.0087	0.9931	130	22'000	24'977	0.0136	-
case300_ieee	24	1257	17'842	0.0073	0.9919	293	9095	22'789	0.0099	0.9897
case200_pserc	174	4649	21'112	0.0099	0.9909	236	6741	22'040	0.0099	0.9901
case240_pserc	2993	22'000	24'997	0.0795	-	2993	22'000	24'997	0.0795	-

Few active sets!

Terminates fast.

High probability of optimal solutions!

Not always:

Large number of active sets

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941
High-Complexity										
case73_ieee_rts	19	1258	17'844	0.0087	0.9931	130	22'000	24'977	0.0136	-
case300_ieee	24	1257	17'842	0.0073	0.9919	293	9095	22'789	0.0099	0.9897
case200_pserc	174	4649	21'112	0.0099	0.9909	236	6741	22'040	0.0099	0.9901
case240_pserc	2993	22'000	24'997	0.0795	-	2993	22'000	24'997	0.0795	-

Few active sets!

Terminates fast.

High probability of optimal solutions!

Not always:

Large number of active sets

Slow to terminate.

Streaming Algorithm Results – PGLib-OPF v 17.08

Max. undiscovered: $\alpha = 0.05$, Max. difference: $\epsilon = 0.04$ ⇨ Termination: $R_{M,W} \leq 0.01$

	Normal distribution					Uniform distribution				
	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$	K_M	M	W_M	$R_{M,W}$	$\mathbb{P}(p^*)$
Low-Complexity										
case3_lmbd	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case5_pjm	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case14_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case30_ieee	1	1	13'259	0.0	1.0	1	1	13'259	0.0	1.0
case39_epri	2	2	13'259	0.0	1.0	2	2	13'259	0.0008	0.9998
case118_ieee	2	33	13'259	0.0	1.0	2	4	13'259	0.0019	0.9984
case57_ieee	2	2	13'259	0.0003	0.9997	3	46	13'259	0.0	1.0
case1888_rte	3	6	13'259	0.0	1.0	3	10	13'259	0.0	1.0
case1951_rte	5	47	13'259	0.0069	0.9943	11	63	13'259	0.0084	0.9901
case162_ieee_dtc	7	91	13'259	0.0054	0.9925	17	192	13'259	0.0085	0.9926
case24_ieee_rts	10	1456	18'209	1.0	1.0	11	64	13'259	0.0047	0.9941
High-Complexity										
case73_ieee_rts	19	1258	17'844	0.0087	0.9931	130	22'000	24'977	0.0136	-
case300_ieee	24	1257	17'842	0.0073	0.9919	293	9095	22'789	0.0099	0.9897
case200_pserc	174	4649	21'112	0.0099	0.9909	236	6741	22'040	0.0099	0.9901
case240_pserc	2993	22'000	24'997	0.0795	-	2993	22'000	24'997	0.0795	-

Few active sets!

Terminates fast.

High probability of optimal solutions!

Not always:

Large number of active sets
Slow to terminate.
Lower probability of optimal solutions!

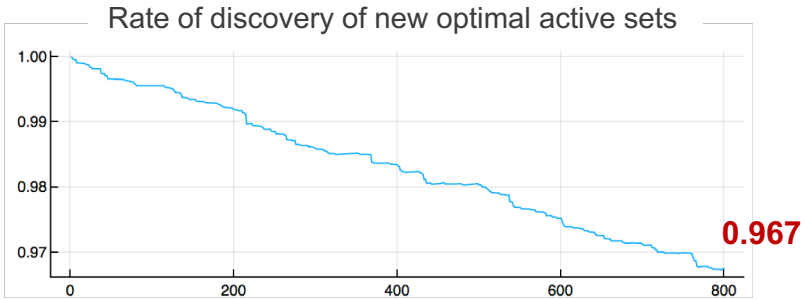
What about the non-linear, non-convex AC OPF?

(just very preliminary results…)

AC Optimal Power Flow

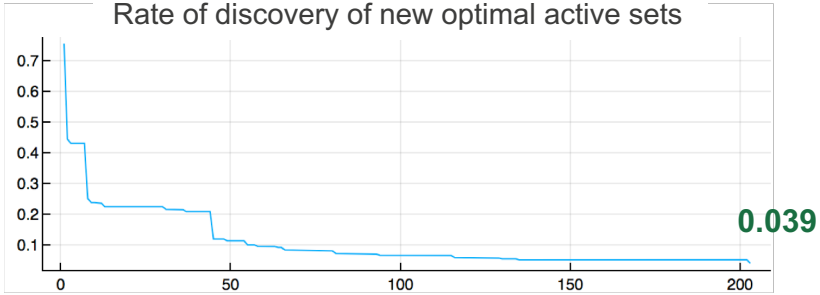
Max. undiscovered: $\alpha = 0.1$, Max. difference: $\epsilon = 0.05$ ⇨ Termination: $R_{M,W} \leq 0.05$

RTE 1951 bus test case



(did not terminate)

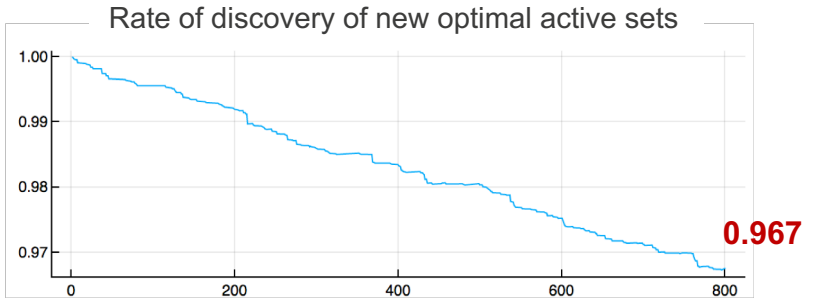
PSERC 200 bus test case

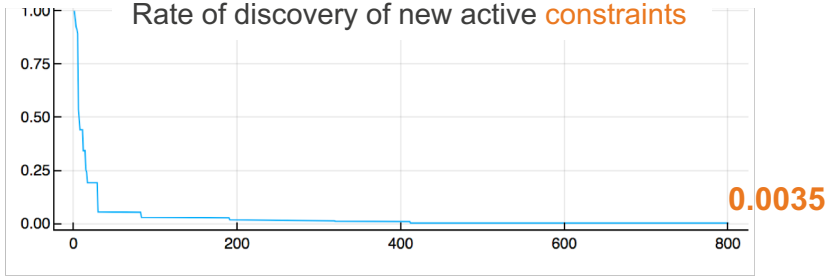


AC Optimal Power Flow

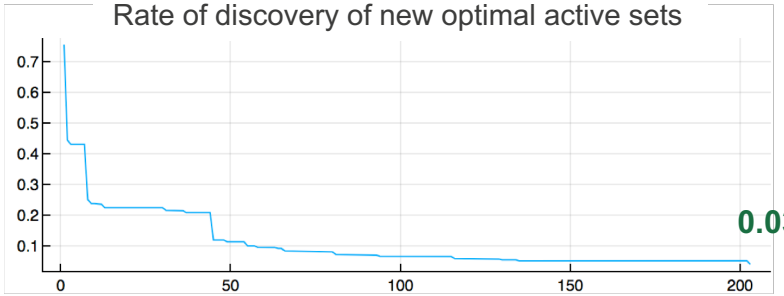
Max. undiscovered: $\alpha = 0.1$, Max. difference: $\epsilon = 0.05$ ⇨ Termination: $R_{M,W} \leq 0.05$

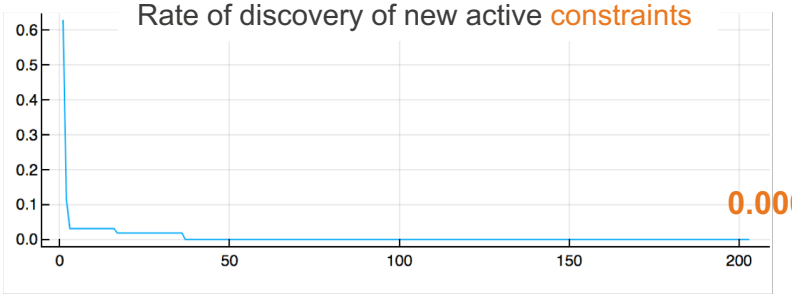
RTE 1951 bus test case





PSERC 200 bus test case



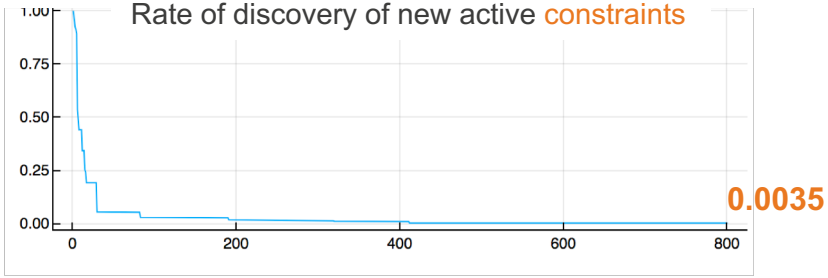


AC Optimal Power Flow

Max. undiscovered: $\alpha = 0.1$, Max. difference: $\epsilon = 0.05$ ⇨ Termination: $R_{M,W} \leq 0.05$

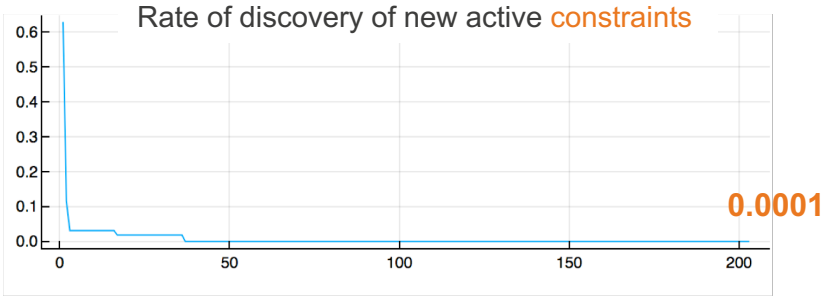
RTE 1951 bus test case

Only 164 of 5192 transmission line constraints ever active



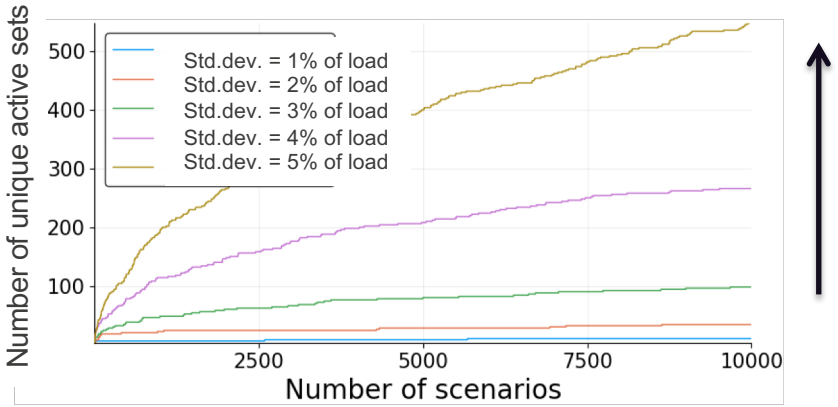
PSERC 200 bus test case

Only 28 of 490 transmission line constraints ever active



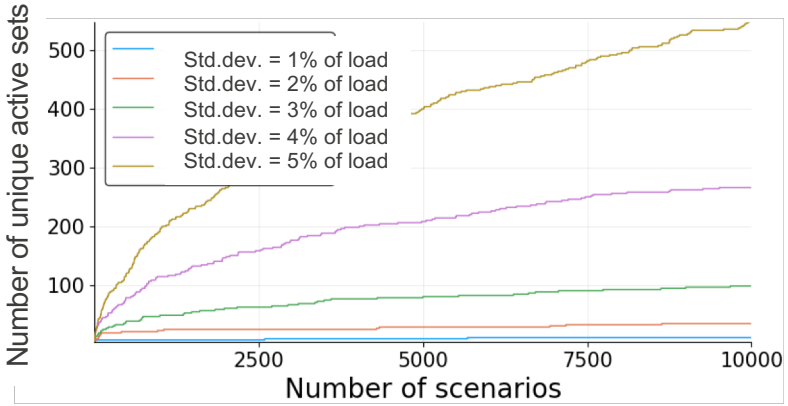
Practical Implications for Power Systems Operation

IEEE 300 bus system with normally distributed load



Increasing parameter uncertainty =
Increasing number of optimal active sets

IEEE 300 bus system with normally distributed load

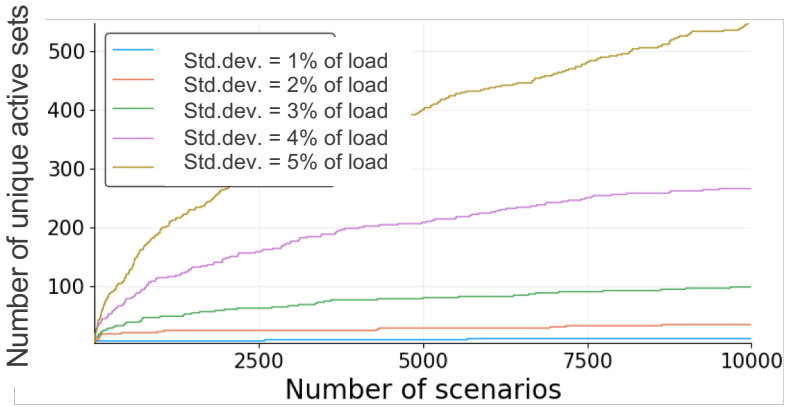


Increasing parameter uncertainty =
Increasing number of optimal active sets

«Power systems operation becomes more unpredictable and complex with increasing uncertainty»

General perception among system operators

IEEE 300 bus system with normally distributed load



Increasing parameter uncertainty =
Increasing number of optimal active sets

«Power systems operation becomes more unpredictable and complex with increasing uncertainty»

General perception among system operators

What does this imply for system risk? Price stability?

Summary

 # 1 – Leverage pre-existing knowledge (mathematical model) improves learning outcomes

 # 2 – Using active sets as an intermediate step is useful

- encodes all information about optimal solution
- finite (and typically low?) number of active sets

 # 3 – Streaming algorithm establishes practicability of the task

- Probabilistic performance guarantees
- Guaranteed to terminate
- Guaranteed to terminate fast for nice problems

 # 4 – This works well for some practical power system problems!

Summary

 # 1 – Leverage pre-existing knowledge (mathematical model) improves learning outcomes

 # 2 – Using active sets as an intermediate step is useful

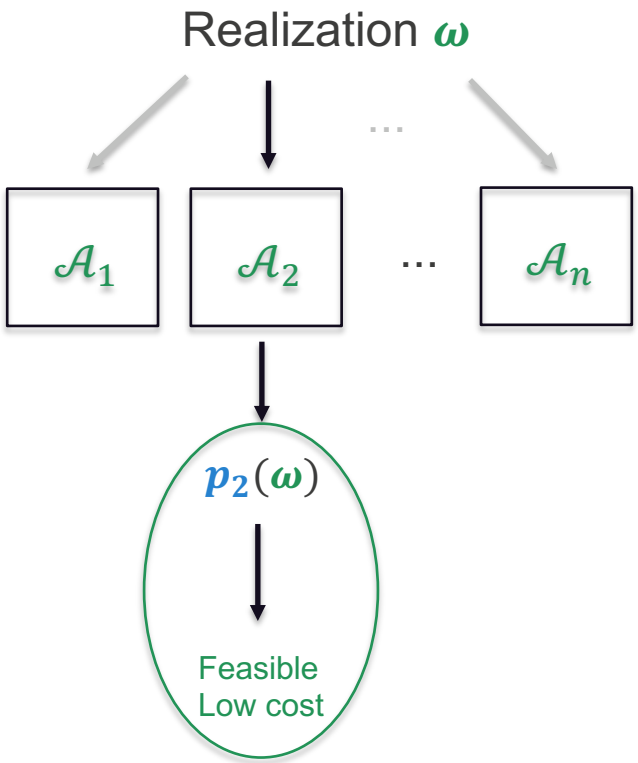
- encodes all information about optimal solution
- finite (and typically low?) number of active sets

 # 3 – Streaming algorithm establishes practicability of the task

- Probabilistic performance guarantees
- Guaranteed to terminate
- Guaranteed to terminate fast for nice problems

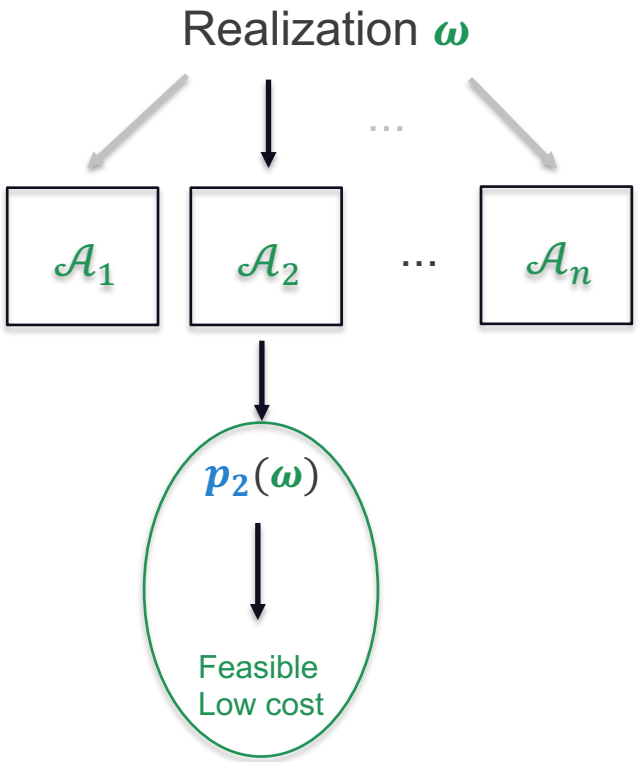
- **Quite general strategy**
  - Streaming algorithm can work for very general problems: Non-convex AC OPF, mixed integer problems …
  - Disclaimer: Application must be such that the number of active sets is small.
  - Alternative strategy: Learn possible *active constraints* instead of active sets

Outlook



Classification?

Outlook



Classification?

Efficient solution?

THANK YOU!

Line Roald, roald@lanl.gov

Yee Sian Ng, Sidhant Misra, Line Roald and Scott Backhaus, «Statistical Learning for DC Optimal Power Flow», Power System Computation Conference (PSCC), 2018

Sidhant Misra, Line Roald and Yee Sian Ng, «Learning for Constrained Optimization», submitted, available online: https://arxiv.org/abs/1802.09639

Summary of theoretical results

1. **Guaranteed to terminate, no need to decide on the number of M samples apriori**
   Definition of the window size W and termination criterion

Theorem 1 and 2 [Misra, Roald, Ng, 2018]**:**

If the window size $W(M)$ is defined as

$$W = \frac{2\gamma}{\epsilon^2}\max\{\log M, \log \underline{M}\} \quad \text{with } \underline{M} = 1 + \left(\frac{\gamma}{\delta(\gamma-1)}\right)^{\frac{1}{\gamma-1}}$$

Then $\mathbb{P}\left(\pi(U_M) - R_{M,W} \leq \epsilon \quad \forall\, M > 1\right) \leq 1 - \delta$

- ϵ difference between true and empirical probability of unobserved active sets
- δ confidence level
- γ hyperparameter > 1

2. **Guaranteed to terminate fast for *benign* systems**

If a (small) number of relevant active sets K_0 that contains more than $1 - \alpha_0$ probability mass, then with probability at least $1 - \delta - \delta_0$ the algorithm terminates in less iterations than

$$M = \frac{1}{\alpha - \alpha_0}\left(K_0 \log 2 + \log\frac{1}{\delta_0}\right)$$