Efficient Creation of Datasets for Data-Driven Power System Applications

Andreas Venzke[1], Daniel K. Molzahn[2], Spyros Chatzivasileiadis[1]

[1]Department of Electrical Engineering, Technical University of Denmark (DTU)

[2]School of Electrical and Computer Engineering,Georgia Institute of Technology

Paper ID: EPSR-D-19-02014

Outline

1. Motivation and Scope of Our Work
2. Infeasibility Certificates
3. Efficient Algorithm to Create Datasets
4. Balanced Datasets for PGLib-OPF cases
5. Conclusions

Outline

1. **Motivation and Scope of Our Work**
2. Infeasibility Certificates
3. Efficient Algorithm to Create Datasets
4. Balanced Datasets for PGLib-OPF cases
5. Conclusions

Motivation

- Machine learning including neural networks have been applied to a range of power system applications since decades (Wehenkel, Springer, 2012)
  $\Rightarrow$ Recent advances in deep learning have sparked renewed interest

Motivation

- Machine learning including neural networks have been applied to a range of power system applications since decades (Wehenkel, Springer, 2012)
  $\Rightarrow$ Recent advances in deep learning have sparked renewed interest

- Machine learning for power system security assessment
  - **static** security criteria such as N-1 security criterion and operational constraints
  - **dynamic** stability criteria such as small-signal or transient stability

- Application example: Convolutional neural networks to predict
  - N-1 security (Du et al., TPWRS, 2019)
  - N-1 security + small signal stability (Arteaga et al., PowerTech, 2019)

  $\Rightarrow$ **Two orders** of magnitude faster than conventional approaches

Scope of Our Work

Performance of machine learning **highly** depends on dataset quality

- Historical data: limited information of **abnormal** situations
- High problem **dimensionality** (number of control variables)

Scope of Our Work

Performance of machine learning **highly** depends on dataset quality

- Historical data: limited information of **abnormal** situations
- High problem **dimensionality** (number of control variables)

Requires **balanced** datasets of secure and not secure operating points with **detailed** security boundary description

- **Computationally hard** due to large **input dimensionality**
- E.g. creating dataset for 162-bus system for 1 MW discretization and one loading level requires to evaluate $\approx 10^{29}$ operating points

Scope of Our Work

1. **Importance sampling**: Bias sampling towards region of interest (Krishnan et al., TPWRS, 2011)
   ⇒ Input space in power systems is largely unbalanced

2. **Composite approaches** or **vine copulas** to enrich historical data (Sun et al., PSCC, 2016), (Konstantelos et al., TPWRS, 2019)
   ⇒ Highly dependent on data quality

Scope of Our Work

1. **Importance sampling**: Bias sampling towards region of interest (Krishnan et al., TPWRS, 2011)
   ⇒ Input space in power systems is largely unbalanced

2. **Composite approaches** or **vine copulas** to enrich historical data (Sun et al., PSCC, 2016), (Konstantelos et al., TPWRS, 2019)
   ⇒ Highly dependent on data quality

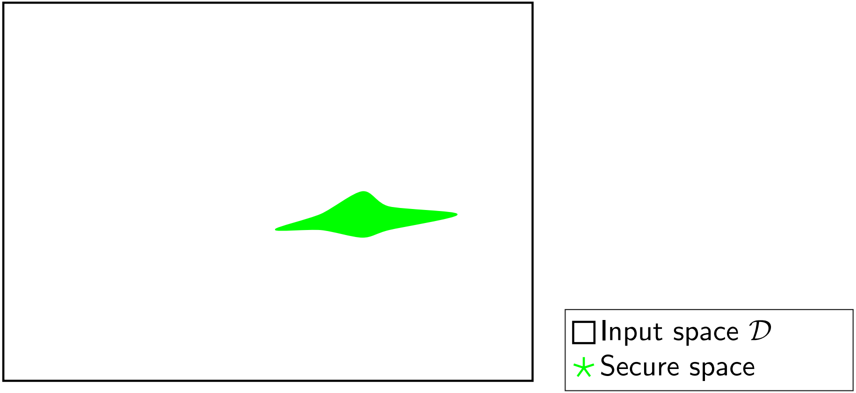
Our approach is **modular** and **parallelizable**:

- Uses **infeasibility certificates** to classify large parts of input space **a-priori** as **not secure**
- Obtains a detailed **boundary** description
- Creates **balanced** datasets and can leverage historical data

Outline

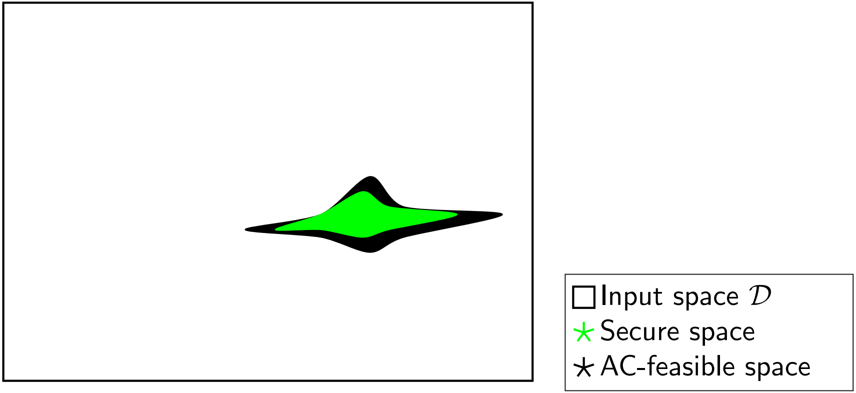
1. Motivation and Scope of Our Work
2. **Infeasibility Certificates**
3. Efficient Algorithm to Create Datasets
4. Balanced Datasets for PGLib-OPF cases
5. Conclusions

Infeasiblity Certificates



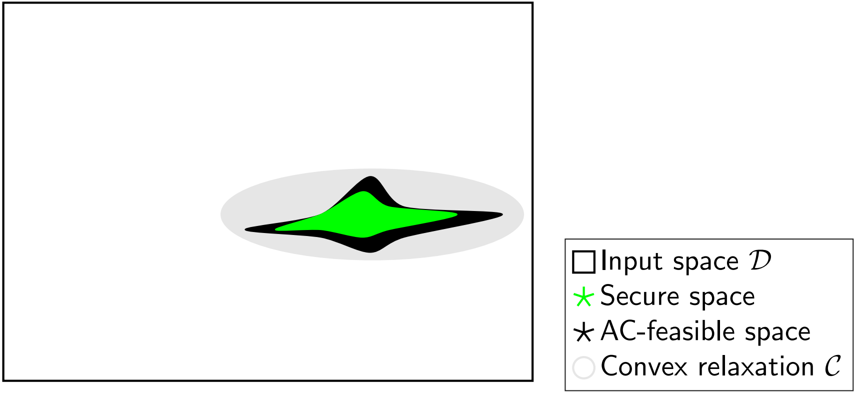
- Goal: For a defined input space $\mathcal{D}$ of dimension d obtain a balanced dataset characterizing the boundary of the secure space
- E.g. for operating point $\mathbf{x} = \begin{bmatrix} \mathbf{P_G} \\ \mathbf{V_G} \end{bmatrix} \in \mathcal{R}^d$, $\mathcal{D} : \begin{bmatrix} \mathbf{P_G^{\text{min}}} \\ \mathbf{V_G^{\text{min}}} \end{bmatrix} \leq \mathbf{x} \leq \begin{bmatrix} \mathbf{P_G^{\text{max}}} \\ \mathbf{V_G^{\text{max}}} \end{bmatrix}$

Infeasiblity Certificates



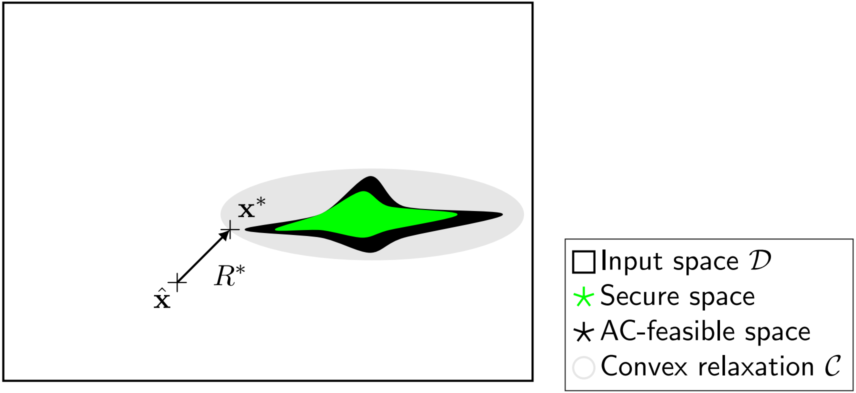
- Pre-requisite for assessment of **dynamic** stability criteria is feasibility to **static** security criteria such as N-1 security and operational constraints
- Static security criteria are described in the AC optimal power flow (**AC-OPF**) problem

Infeasiblity Certificates



- **Convex relaxations** of AC-OPF can provide **infeasibility** certificate

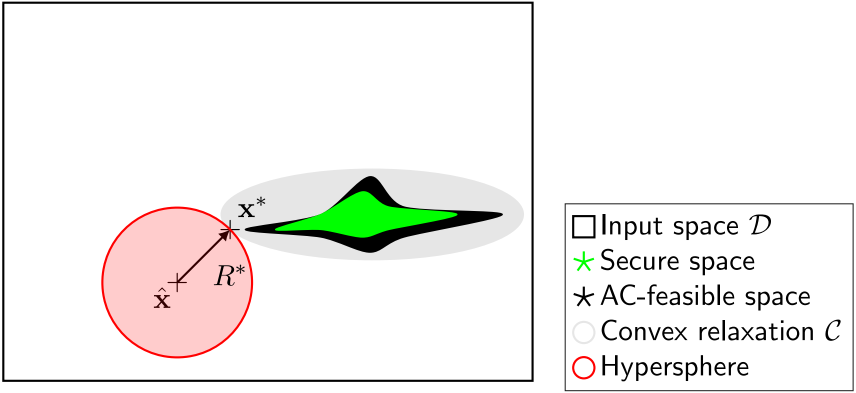
Infeasiblity Certificates



- For **infeasible** operating point $\hat{\mathbf{x}}$ compute closest dispatch $\mathbf{x}^*$ **feasible** to convex relaxation of AC-OPF:

$$\min_{\mathbf{x}^*} \quad |\mathbf{x}^* - \hat{\mathbf{x}}|_2$$
$$\text{s.t.} \quad \mathbf{x}^* \text{feasible to convex relaxation of AC-OPF}$$

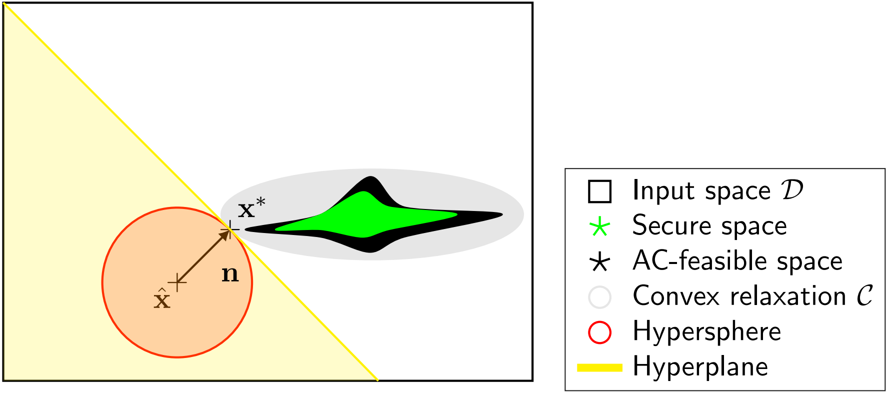
Infeasiblity Certificates



- (Molzahn, TPWRS, 2018) proposed **hyperspheres** with radius R^* as infeasibility certificates:

$$\forall \mathbf{x} : |\hat{\mathbf{x}} - \mathbf{x}|_2 < R^* \Rightarrow \mathbf{x} \notin \mathcal{C}$$

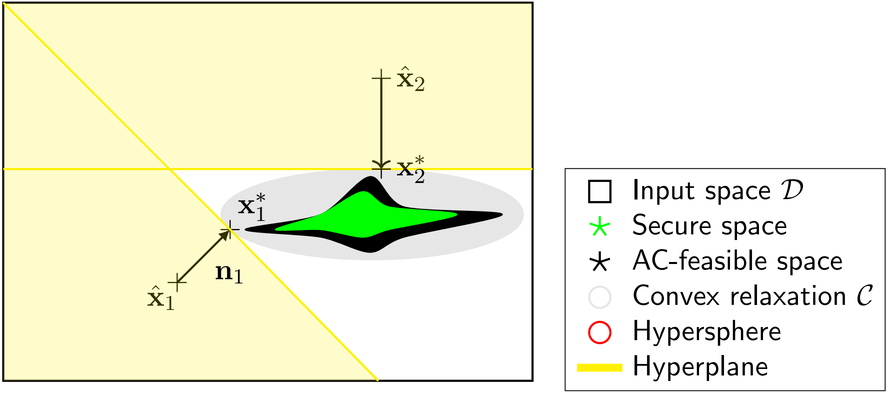
Infeasiblity Certificates



- We propose to use separating **hyperplanes** as infeasiblity certificates:

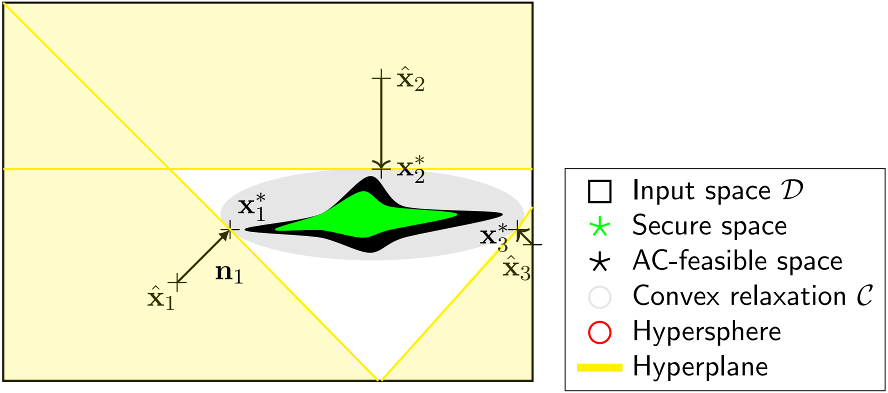
$$\forall \mathbf{x} : \mathbf{n}^T(\mathbf{x} - \mathbf{x}^*) < 0 \Rightarrow \mathbf{x} \notin \mathcal{C}$$

Infeasiblity Certificates



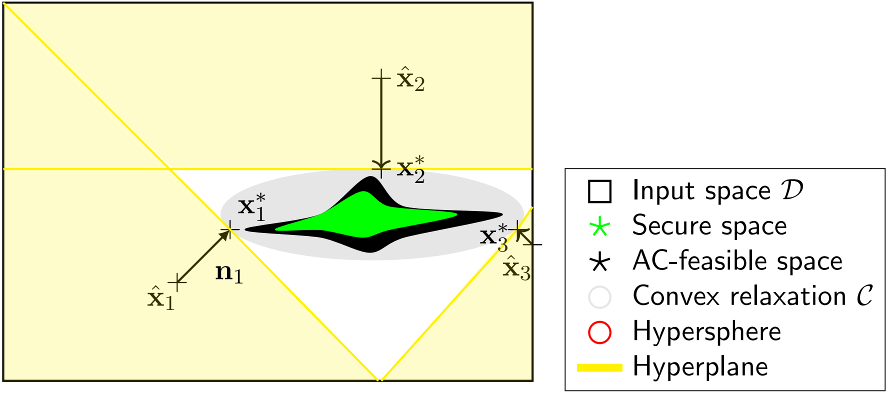
- Observation: Multiple hyperplanes form **convex polyhedron**

Infeasiblity Certificates



- Observation: Multiple hyperplanes form **convex polyhedron**

Infeasiblity Certificates



- Benefits of hyperplanes:
  1. Larger parts of input space classified as infeasible
  2. Efficient methods available to sample from within a convex polyhedron

Algorithm to Minimize Unclassified Input Space

Goal: minimize volume of **convex polyhedron** using N_1 iterations

Algorithm to Minimize Unclassified Input Space

Goal: minimize volume of **convex polyhedron** using N_1 iterations

1. Set iteration count $k \leftarrow 0$. Initialize unclassified input space as $\mathbf{A}^{(0)}\mathbf{x} \leq \mathbf{b}^{(0)}$ with:
   $\mathbf{A}^{(0)} := [\mathbf{I}^{|x|\times|x|} \;\; -\mathbf{I}^{|x|\times|x|}]^T$
   $\mathbf{b}^{(0)} := [(\mathbf{x}^{\text{max}})^T \; (\mathbf{x}^{\text{min}})^T]^T$

Algorithm to Minimize Unclassified Input Space

Goal: minimize volume of **convex polyhedron** using N_1 iterations

1. Set iteration count $k \leftarrow 0$. Initialize unclassified input space as $\mathbf{A}^{(0)}\mathbf{x} \leq \mathbf{b}^{(0)}$ with:
   $$\mathbf{A}^{(0)} := [\mathbf{I}^{|x|\times|x|} \ \ -\mathbf{I}^{|x|\times|x|}]^T$$
   $$\mathbf{b}^{(0)} := [(\mathbf{x}^{\text{max}})^T \ (\mathbf{x}^{\text{min}})^T]^T$$

2. If $k \leq N_1$ continue, otherwise terminate. Increase $k \leftarrow k+1$. Draw random $\mathbf{x}^{(k)}$ from inside $\mathbf{A}^{(k)}\mathbf{x} \leq \mathbf{b}^{(k)}$. Compute closest feasible sample with $\hat{\mathbf{x}} := \mathbf{x}^{(k)}$ and obtain $\mathbf{x}^*$ and R^*.

Algorithm to Minimize Unclassified Input Space

Goal: minimize volume of **convex polyhedron** using N_1 iterations

1. Set iteration count $k \leftarrow 0$. Initialize unclassified input space as $\mathbf{A}^{(0)}\mathbf{x} \leq \mathbf{b}^{(0)}$ with:
   $\mathbf{A}^{(0)} := [\mathbf{I}^{|x|\times|x|} \;\; -\mathbf{I}^{|x|\times|x|}]^T$
   $\mathbf{b}^{(0)} := [(\mathbf{x}^{\text{max}})^T \;\; (\mathbf{x}^{\text{min}})^T]^T$

2. If $k \leq N_1$ continue, otherwise terminate. Increase $k \leftarrow k+1$. Draw random $\mathbf{x}^{(k)}$ from inside $\mathbf{A}^{(k)}\mathbf{x} \leq \mathbf{b}^{(k)}$. Compute closest feasible sample with $\hat{\mathbf{x}} := \mathbf{x}^{(k)}$ and obtain $\mathbf{x}^*$ and R^*.

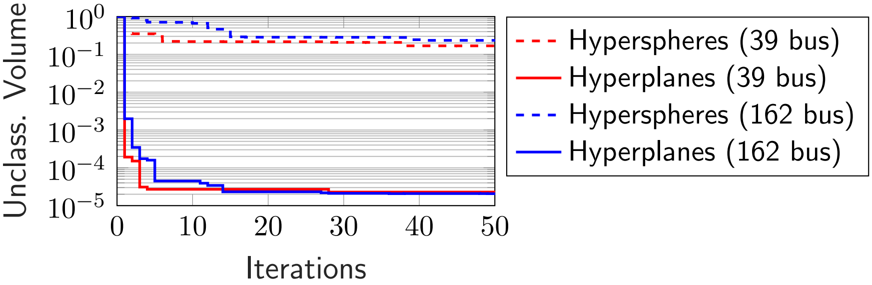
3. If $R^* > 0$ reduce unclassified region by adding hyperplane.
   $\mathbf{A}^{(k+1)} = [(\mathbf{A}^{(k)})^T \;\; \mathbf{n}^T]^T$
   $\mathbf{b}^{(k+1)} = [(\mathbf{b}^{(k)})^T \;\; \mathbf{n}^T\mathbf{x}^*]^T$
   Return to step 2.

Comparison of Infeasibility Certificates

- Reference normalized volume is 1
- Use *volesti* (Emiris et al., ACM TOMS, 2018) to approximate the volume of a convex polyhedron
- Rejection sampling for the volume of hyperspheres

Comparison of Infeasibility Certificates

- Reference normalized volume is 1
- Use *volesti* (Emiris et al., ACM TOMS, 2018) to approximate the volume of a convex polyhedron
- Rejection sampling for the volume of hyperspheres

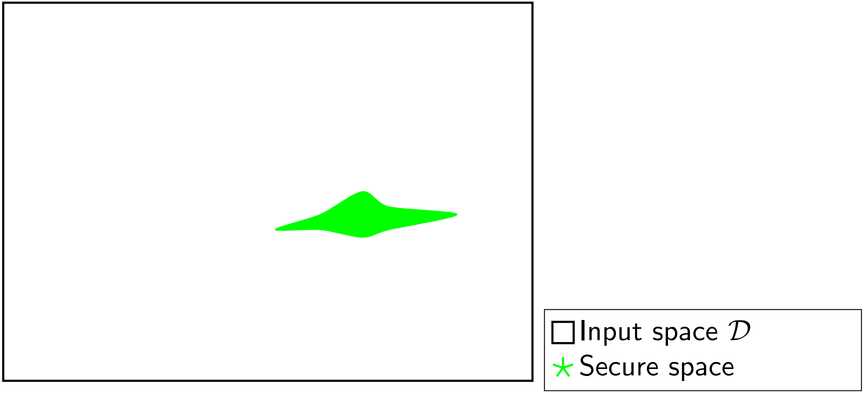


- Volume reduction from 1 to 10^{-5} implies that in expectation at least 10^5 random samples are required to identify 1 potential feasible sample.

Outline

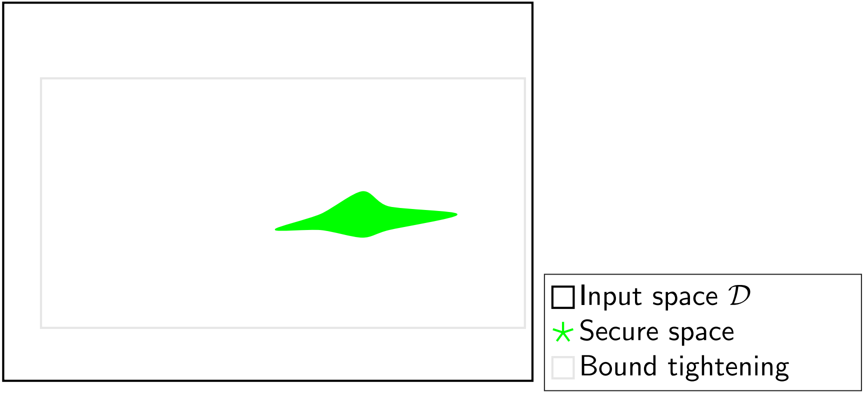
1. Motivation and Scope of Our Work
2. Infeasibility Certificates
3. **Efficient Algorithm to Create Datasets**
4. Balanced Datasets for PGLib-OPF cases
5. Conclusions

Efficient Algorithm to Create Datasets



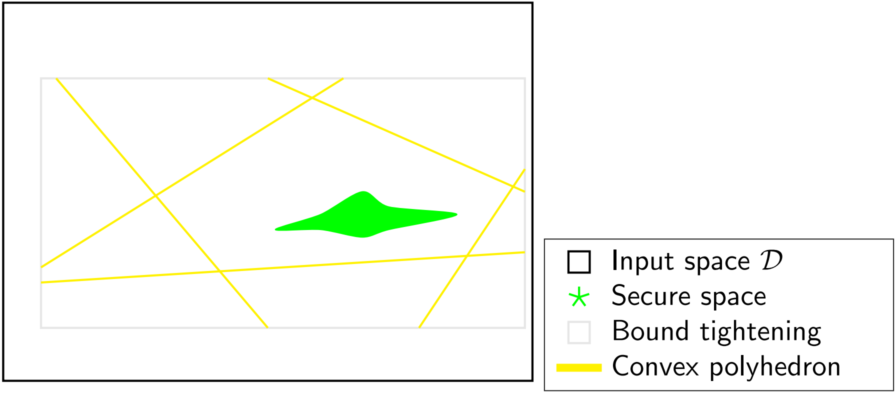
0 Goal: Computationally efficient creation of balanced datasets of secure and insecure operating points with a detailed boundary description

Efficient Algorithm to Create Datasets



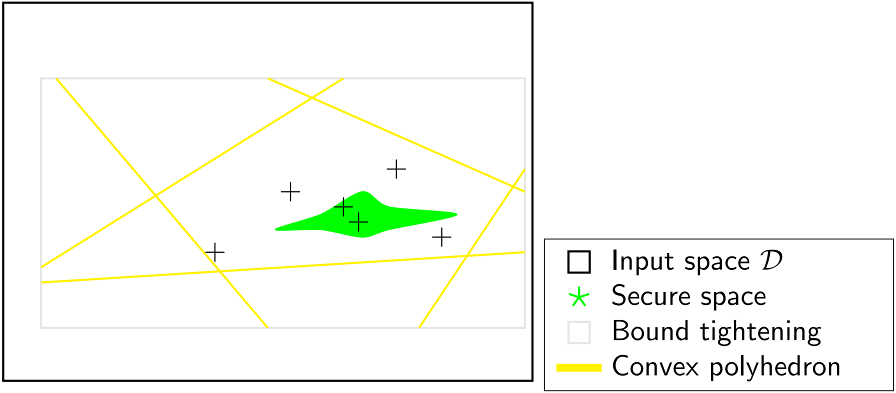
1. **Bound tightening** for AC-OPF problem
2. Direct tightening of input bounds by maximizing and minimizing $\mathbf{x}$ subject to constraints of convex relaxations of AC-OPF
   $\Rightarrow$ reduce unclassified volume to V^{BT}

Efficient Algorithm to Create Datasets



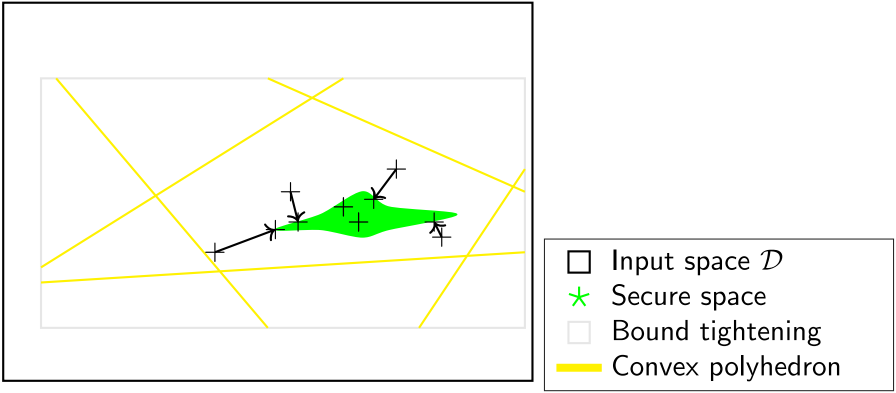
❸ Run algorithm to minimize unclassified input space for N_1 steps
$\Rightarrow$ convex polyhedron $\mathbf{Ax} \leq \mathbf{b}$ with unclassified volume V^{HP}

Efficient Algorithm to Create Datasets



❹ Draw and classify N_2 samples from within convex polyhedron

Efficient Algorithm to Create Datasets

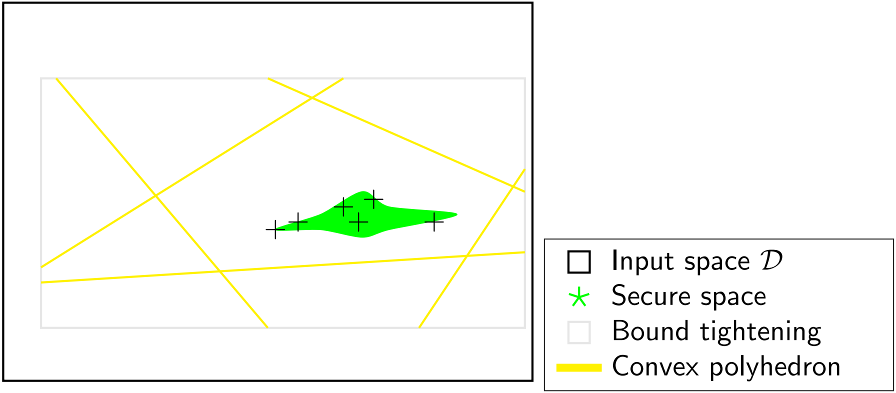


5. Obtain detailed boundary identification. For each infeasible sample $\mathbf{x}$ identify **closest** secure operation point.

- Static Security: Solve OPF minimizing distance to feasible solution
- Dynamic Security: Directed walks using sensitivity measures

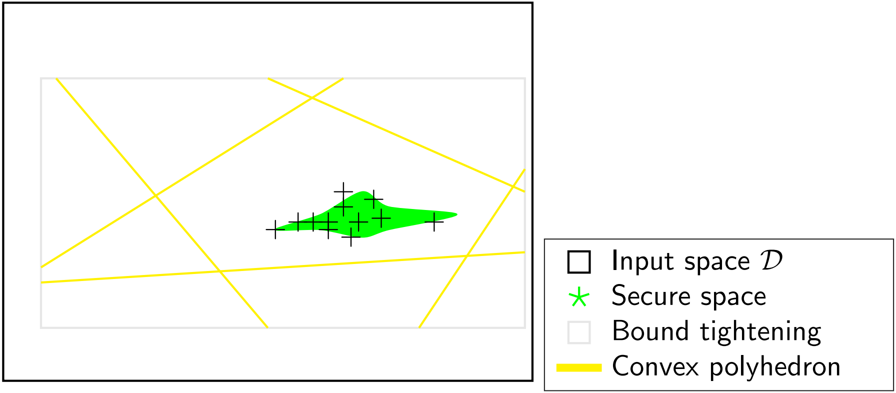
For dynamic security more details in our paper: Thams, F., Venzke, A., Eriksson, R. and Chatzivasileiadis, S., 2019. Efficient database generation for data-driven security assessment of power systems. IEEE Transactions on Power Systems.

Efficient Algorithm to Create Datasets



❻ Fit multi-variate Gaussian distribution to obtained secure samples (and if available historical data). Bias sampling towards inside the secure space by multiplying covariance matrix with $s_{\text{red}} \leq 1$.

Efficient Algorithm to Create Datasets



❼ Draw and classify N_3 samples from fitted distribution.

$\Rightarrow$ Algorithm **parallelizable** and **modular**
$\Rightarrow$ Can include e.g. importance sampling or vine-copulas as well

Outline

1. Motivation and Scope of Our Work
2. Infeasibility Certificates
3. Efficient Algorithm to Create Datasets
4. **Balanced Datasets for PGLib-OPF cases**
5. Conclusions

Simulation Setup

- We use the **QC relaxation** and existing bound tightening techniques:
  - Analytical bound tightening (Shchetinin et al., TPWRS, 2019)
  - Optimization-based bound tightening (Sundar et al., Arxiv, 2018)

Simulation Setup

- We use the **QC relaxation** and existing bound tightening techniques:
  - Analytical bound tightening (Shchetinin et al., TPWRS, 2019)
  - Optimization-based bound tightening (Sundar et al., Arxiv, 2018)
- Security criteria:

1. Operational constraints (AC-OPF)
2. N-1 security + uncertainty in generation and loading (N-1 SC-AC-OPF)

- We evaluate 1 on 13 PGLib-OPF test cases up to 500 buses and input dimensionality $|\mathbf{x}| = d$ of 125
- We evaluate 2 on 39- and 162 bus test cases considering 5 line outages, 3 uncertain loads, and 3 uncertain renewable generators

Simulation Setup

- We use the **QC relaxation** and existing bound tightening techniques:
  - Analytical bound tightening (Shchetinin et al., TPWRS, 2019)
  - Optimization-based bound tightening (Sundar et al., Arxiv, 2018)
- Security criteria:

1. Operational constraints (AC-OPF)
2. N-1 security + uncertainty in generation and loading (N-1 SC-AC-OPF)

- We evaluate 1 on 13 PGLib-OPF test cases up to 500 buses and input dimensionality $|\mathbf{x}| = d$ of 125
- We evaluate 2 on 39- and 162 bus test cases considering 5 line outages, 3 uncertain loads, and 3 uncertain renewable generators
- Sampling sizes:
  - $N_1 = 10^3$ potential hyperplanes
  - $N_2 = 10^4$ samples from convex polyhedron
  - $N_3 = 10^5$ samples from multivariate Gaussian distribution ($s_{\text{red}} = 0.25$)

Unclassified Input Volumes for PGLib-OPF Cases

case	Dim. d	V^{BT}	$\|HP\|$	V^{HP}
(1) Operational constraints				
case30_ieee	7	6.2e-03		
case73_ieee_rts	62	1.0e+00		
case300_ieee	125	1.0e-12		
Median 13 cases	23	8.6e-02		
(2) N-1 security and uncertainty				
case39_epri	25	2.6e-01		
case162_ieee_dtc	29	2.2e-04		

- **Bound tightening** allows to a-priori reduce the unclassified input space
- The reduction is **test-case dependent**

Unclassified Input Volumes for PGLib-OPF Cases

case	Dim. d	V^{BT}	$\|HP\|$	V^{HP}
		(1) Operational constraints		
case30_ieee	7	6.2e-03	61	8.8e-06
case73_ieee_rts	62	1.0e+00	608	6.1e-16
case300_ieee	125	1.0e-12	1000	3.4e-40
Median 13 cases	23	8.6e-02	271	7.0e-08
		(2) N-1 security and uncertainty		
case39_epri	25	2.6e-01	271	2.0e-05
case162_ieee_dtc	29	2.2e-04	394	6.0e-10

- Using the algorithm based on hyperplanes allows to **significantly reduce** the unclassified input space for **all** 15 test cases
- For some test cases, larger number N_1 could lead to further reductions

Created Datasets for PGLib-OPF Cases

Share of secure samples (%)

Power system case	Boundary Identification $N_2 = 10^4$	Inside (MVGD) $N_3 = 10^5$	Overall Secure
(1) Operational constraints			
case30_ieee	75.0%		
case73_ieee_rts	63.9%		
case300_ieee	50.0%		
Average all 13 cases	59.7%		
(2) N-1 security and uncertainty			
case39_epri	58.2%		
case162_ieee_dtc	50.0%		

- **Directly** sampling from convex polyhedron identifies secure samples

Created Datasets for PGLib-OPF Cases

Share of secure samples (%)			
Power system case	Boundary Identification $N_2 = 10^4$	Inside (MVGD) $N_3 = 10^5$	Overall Secure
(1) Operational constraints			
case30_ieee	75.0%	50.2%	
case73_ieee_rts	63.9%	51.1%	
case300_ieee	50.0%	32.6%	
Average all 13 cases	59.7%	44.2%	
(2) N-1 security and uncertainty			
case39_epri	58.2%	78.2%	
case162_ieee_dtc	50.0%	17.9%	

- Sampling from fitted multi-variate Gaussian distribution

Created Datasets for PGLib-OPF Cases

Share of secure samples (%)

Power system case	Boundary Identification $N_2 = 10^4$	Inside (MVGD) $N_3 = 10^5$	Overall Secure
(1) Operational constraints			
case30_ieee	75.0%	50.2%	54.0%
case73_ieee_rts	63.9%	51.1%	52.7%
case300_ieee	50.0%	32.6%	34.7%
Average all 13 cases	59.7%	44.2%	46.5%
(2) N-1 security and uncertainty			
case39_epri	58.2%	78.2%	75.2%
case162_ieee_dtc	50.0%	17.9%	23.2%

- We obtain **balanced** datasets for a range of test cases

Outline

1. Motivation and Scope of Our Work
2. Infeasibility Certificates
3. Efficient Algorithm to Create Datasets
4. Balanced Datasets for PGLib-OPF cases
5. **Conclusions**

Conclusions

- Proposed **infeasibility** certificates based on **hyperplanes**
- Introduced algorithm to **a-priori** minimize unclassified input space
  $\Rightarrow$ Reduced volumes ranging from 10^{-2} up to 10^{-40}.
- **Modular** and parallelizable algorithm created **balanced** datasets with detailed security **boundary description**
- We plan to soon release an **open-source implementation + datasets**!

Open Questions

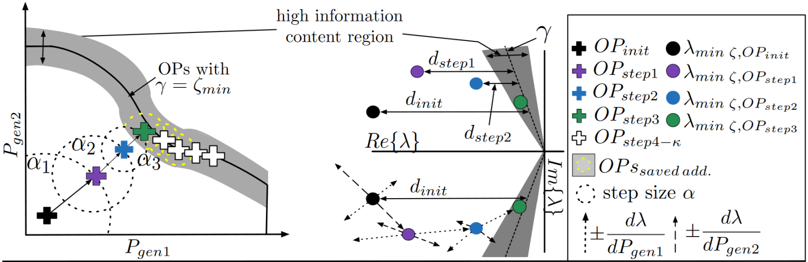
Thank you for your attention!

Do you have any questions?

source: www.DTU.dk

Appendix

Directed walks for dynamic security criteria



2-dim. subspace of multi-dim. space of potential generation patterns (remaining dim. held constant for illustration)

Eigenspace of all $\Gamma + 1$ analyzed systems (only lowest damped (pair) of eigenvalues of all systems shown)

- Security Critera: N-1 security + small-signal stability
- Illustration of the Directed Walk (DW) through a two dimensional space using varying step sizes, α_i, following the steepest descent of distance, d.
- **For dynamic security more details in our paper**: Thams, F., Venzke, A., Eriksson, R. and Chatzivasileiadis, S., 2019. Efficient database generation for data-driven security assessment of power systems. IEEE Transactions on Power Systems.

Full results – Unclassified input volumes

case	$\lvert x \rvert$	V^{BT}	$\lvert HP \rvert$	V^{HP}	$\frac{-\log_{10}(V)}{\lvert x \rvert}$
AC-OPF without N-1 security and without uncertainty					
case3_lmbd	4	6.3e-02	28	3.3e-02	37.0%
case5_pjm	7	1.0e+00	99	6.9e-03	30.9%
case14_ieee	6	2.4e-01	54	6.9e-04	52.7%
case24_ieee_rts	20	9.2e-01	184	2.3e-06	28.2%
case30_ieee	7	6.2e-03	61	8.8e-06	72.2%
case39_epri	19	9.9e-02	203	7.0e-08	37.7%
case57_ieee	10	3.8e-02	231	4.9e-06	53.1%
case73_ieee_rts	62	1.0e+00	608	6.1e-16	24.5%
case118_ieee	72	1.7e-02	1000	1.6e-17	23.3%
case162_ieee_dtc	23	6.1e-04	371	1.5e-11	47.1%
case200_tamu	69	9.3e-01	1000	6.0e-11	14.8%
case300_ieee	125	1.0e-12	1000	3.4e-40	31.6%
case500_tamu	111	8.6e-02	1000	5.4e-26	22.8%

Full results – Created datasets

Power system case	Boundary $N_2 = 10^4$	Inside (MVND) $N_3 = 10^5$	Overall Secure	Overall Points
AC-OPF without N-1 security and without uncertainty				
case3_lmbd	69.5%	36.5%	40.6%	114'389
case5_pjm	68.6%	69.4%	69.3%	125'432
case14_ieee	73.3%	59.0%	61.0%	147'047
case24_ieee_rts	66.8%	44.3%	48.7%	131'158
case30_ieee	75.0%	50.2%	54.0%	124'944
case39_epri	57.2%	29.9%	33.9%	154'635
case57_ieee	58.9%	35.2%	38.9%	150'865
case73_ieee_rts	63.9%	51.1%	52.7%	222'730
case118_ieee	53.2%	47.0%	47.6%	209'996
case162_ieee_dtc	50.0%	40.1%	41.7%	129'165
case200_tamu	50.2%	36.6%	38.1%	177'023
case300_ieee	50.0%	32.6%	34.7%	163'087
case500_tamu	50.0%	35.4%	37.1%	174'774

Accuracy of neural network classifiers

using the created datasets

case	full dataset	only boundary
AC-OPF without N-1 security and without uncertainty		
case14_ieee	78.2%	60.5%
case39_epri	74.6%	38.5%
case162_ieee_dtc	84.4%	49.8%
AC-OPF including N-1 security and uncertainty		
case39_epri	81.0%	80.4%
case162_ieee_dtc	93.4%	31.9%