# Uniform sampling of a feasible set of model parameters

**Wenyu Li**

Arun Hegde
Jim Oreluk
Andrew Packard
Michael Frenklach

# Acknowledgements

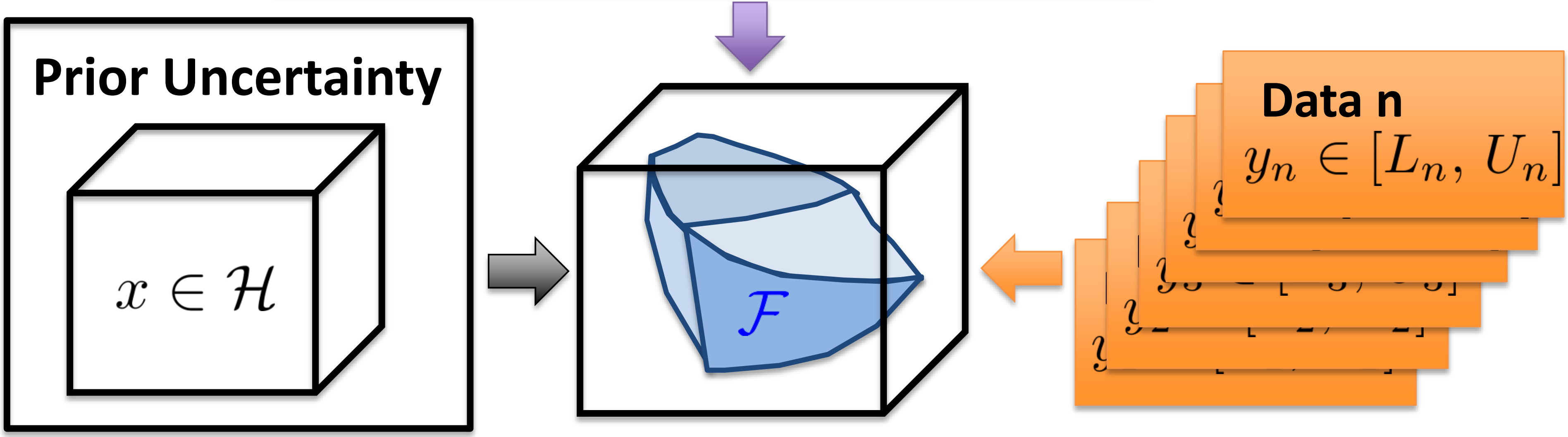
This work is supported as a part of the CCMSC at the University of Utah, funded through PSAAP by the National Nuclear Security Administration, under Award Number DE-NA0002375.

National Nuclear Security Administration

CARBON CAPTURE MULTIDISCIPLINARY SIMULATION CENTER

# Bound-to-Bound Data Collaboration (B2BDC)

**Model:** $M_e(x), \quad e = 1, 2, \ldots, n$




**Feasible set**

$$\{x \in \mathcal{H} : \; L_e \leq M_e(x) \leq U_e, \quad e = 1, 2, \ldots, n\}$$

# Uniform sampling

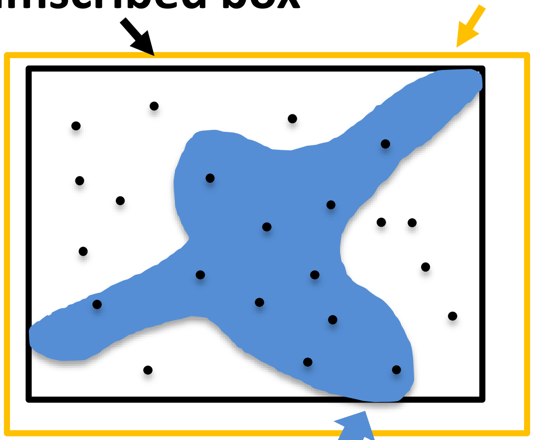
**Goal: uniform sampling of feasible set**

- Sampling is useful in providing information about $\mathcal{F}$
- B2BDC makes **NO** distribution assumptions, but as far as taking samples, uniform distribution of $\mathcal{F}$ is reasonable
- Applying Bayesian analysis with **specific prior assumptions** also leads to uniform distribution of $\mathcal{F}$ as posterior[1]

# Rejection sampling method with a box

**Procedure:**

- find a bounding box
  - available from B2BDC
- generate uniformly distributed samples in the box as candidates
- reject the points outside of feasible set



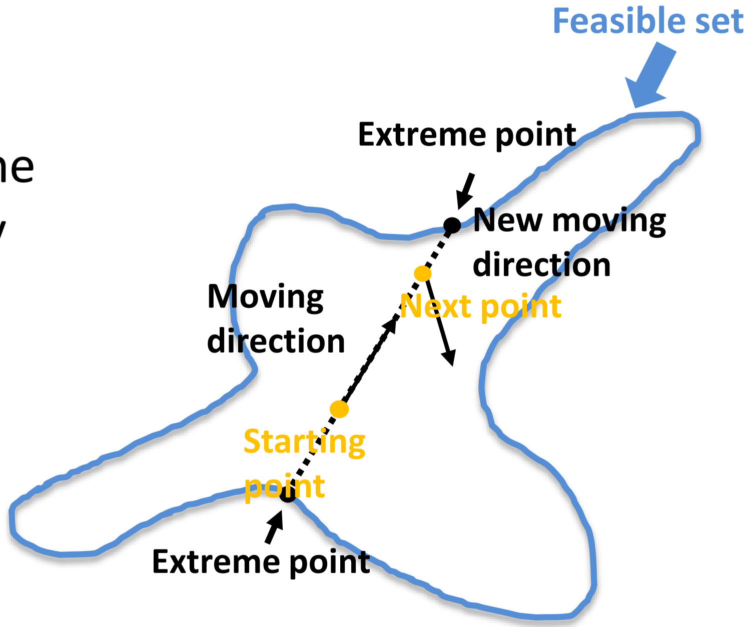

**Pros & Cons**

- **provably uniform** in the feasible set
- candidates can be drawn very efficiently
- **efficiency** drops quickly with increased dimension

# Random walk[2] (RW)

## Procedure:

- start from a feasible point
  - available from B2BDC
- select a random direction, calculate extreme points and choose the next point uniformly
- repeat the process



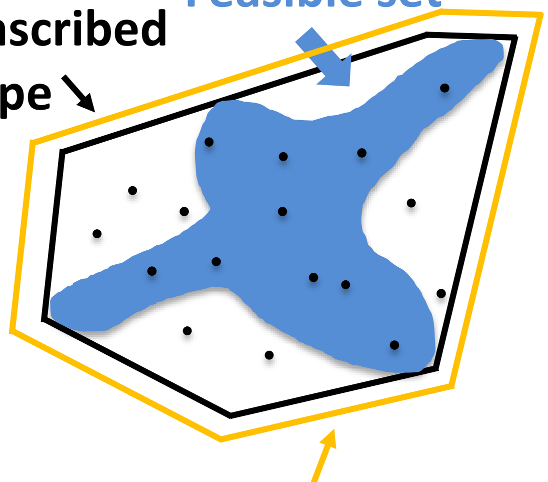

## Pros & Cons

- **NOT** limited by problem dimensions
- **NOT** necessarily uniform in the feasible set

# Rejection sampling method with a polytope

**Procedure:**

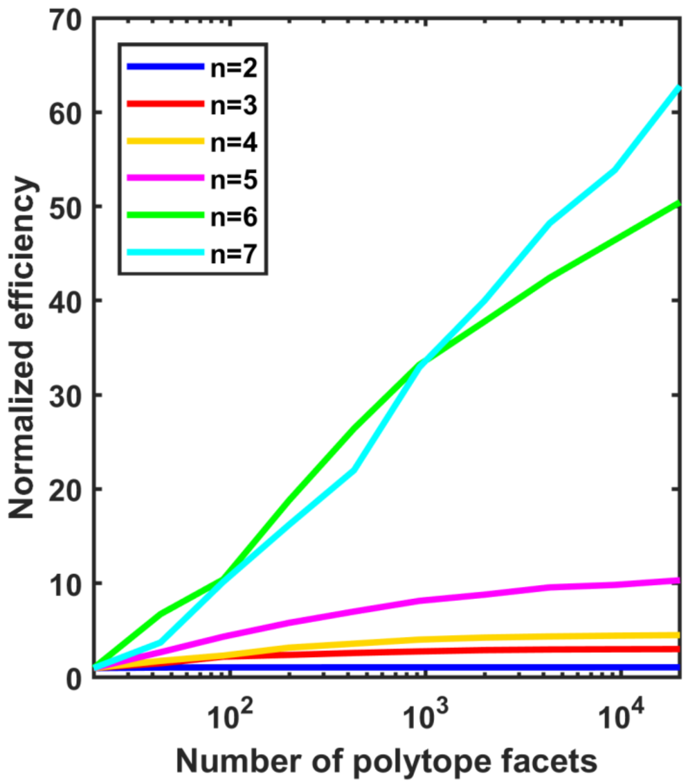
- find a convex bounding polytope
  - available from B2BDC
- generate candidate points by random walk
- reject the points outside of feasible set

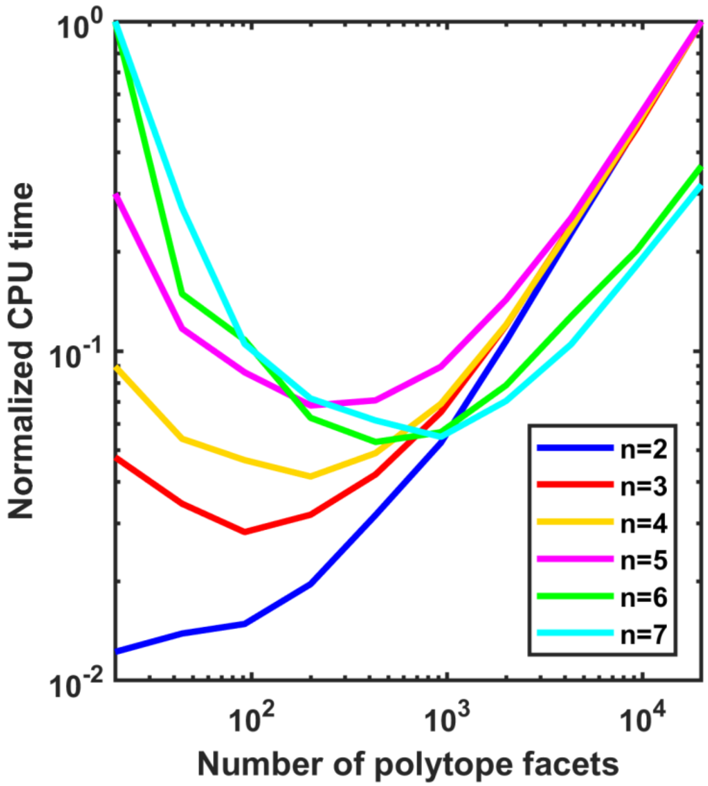



**Pros & Cons**

- **provably uniform** in the feasible set
- increased efficiency with more polytope facets
- **limited** by computational resource

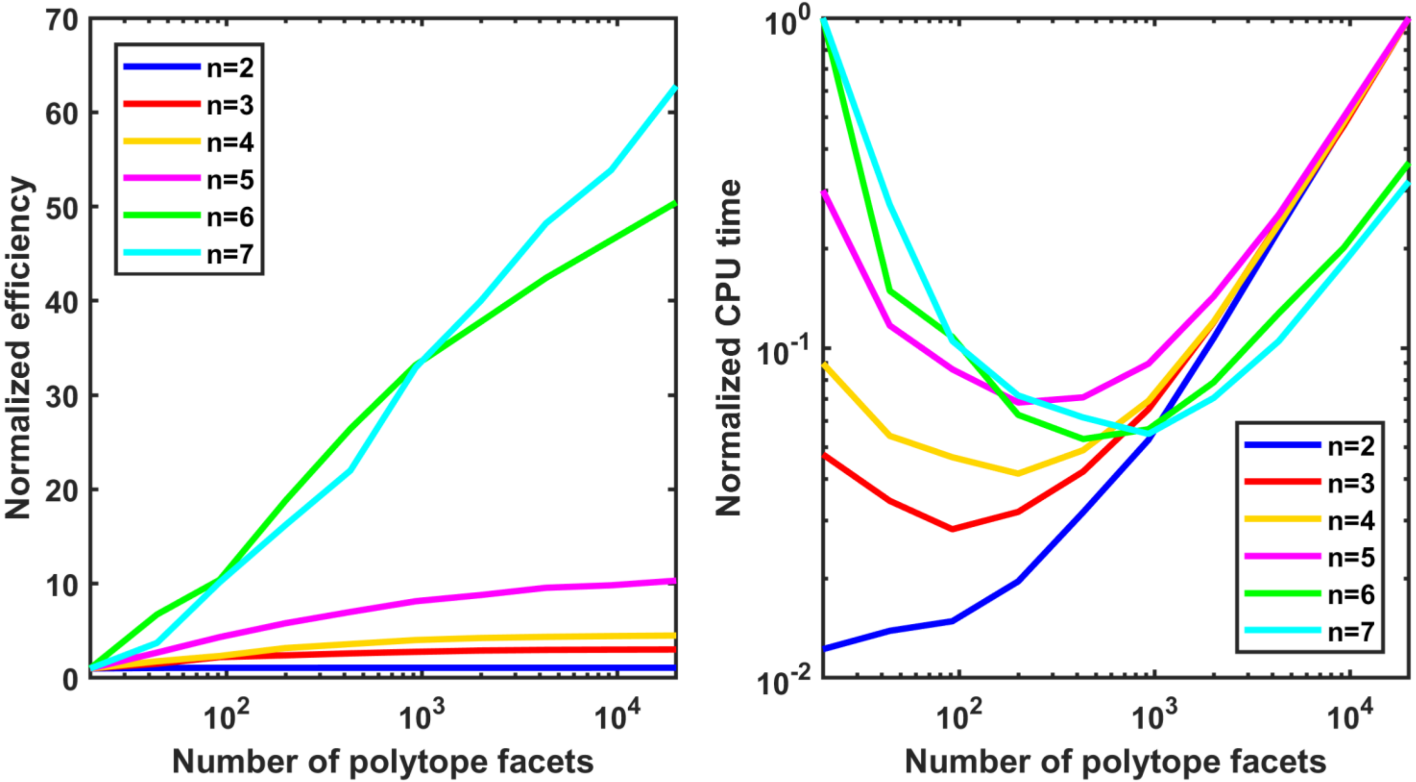
# Effect of polytope complexity




- Polytopes with different complexity are tested
- 5 million candidates are generated to calculate the efficiency and CPU time

# Effect of polytope complexity




- Sampling efficiency increases with more complex polytope
- The improvement is more significant at higher dimensions

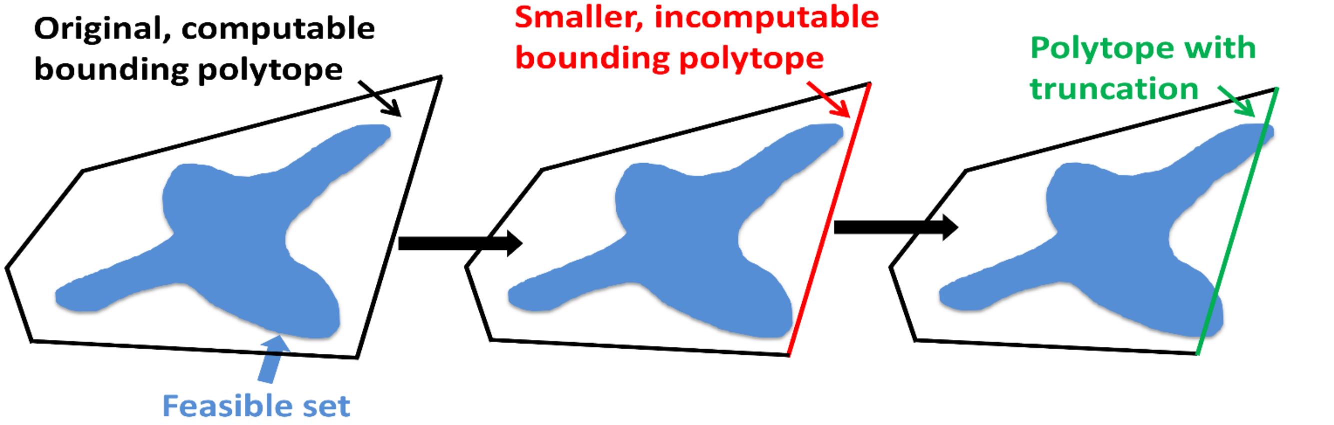
# Truncation strategy

## Motivations

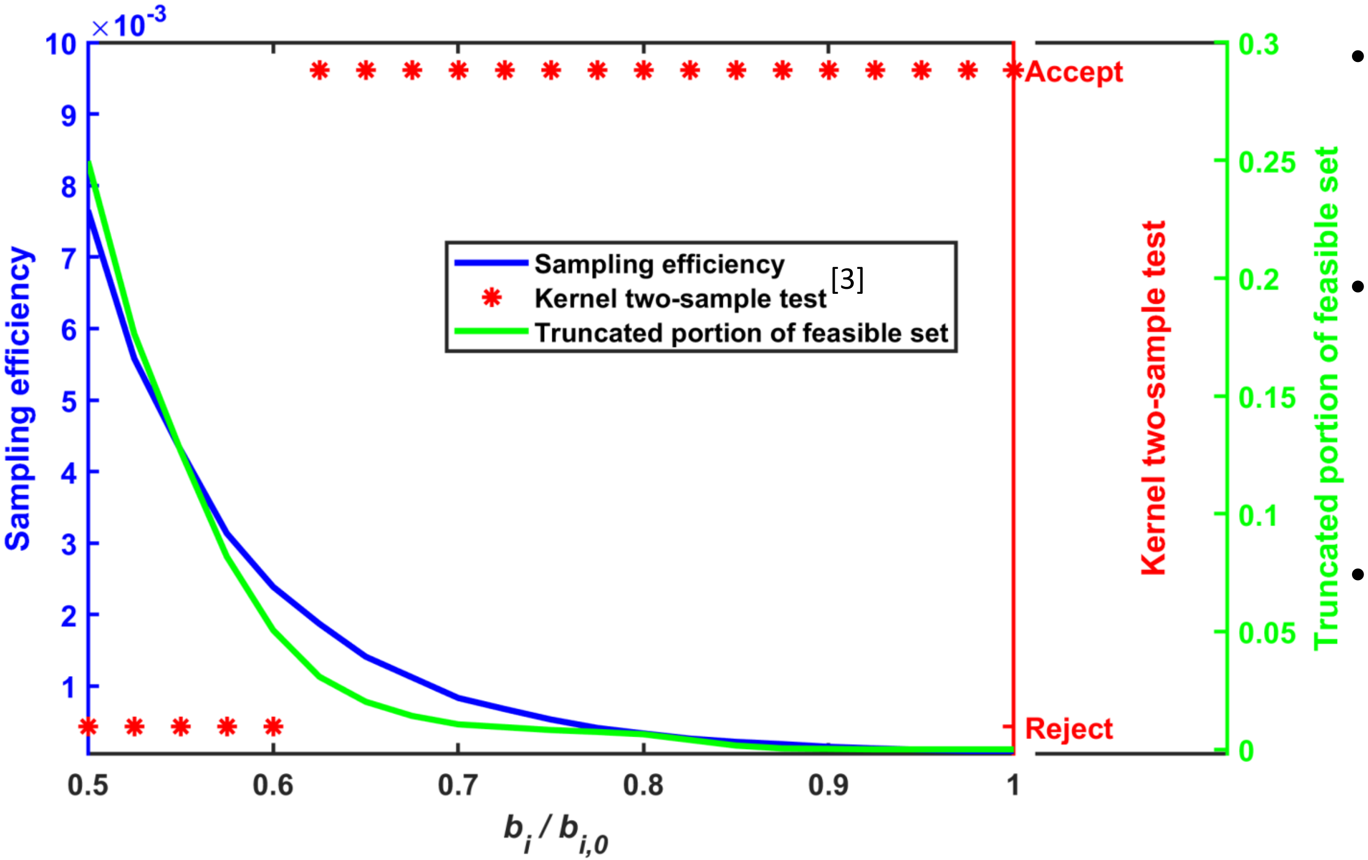
- difference between a bounding and circumscribed polytope
- existence of low-density tails along most of the directions

## Procedure

- start with a bounding polytope and shrink the polytope bounds
- recommended to stop when a practical efficiency is obtained



# Effect of truncation

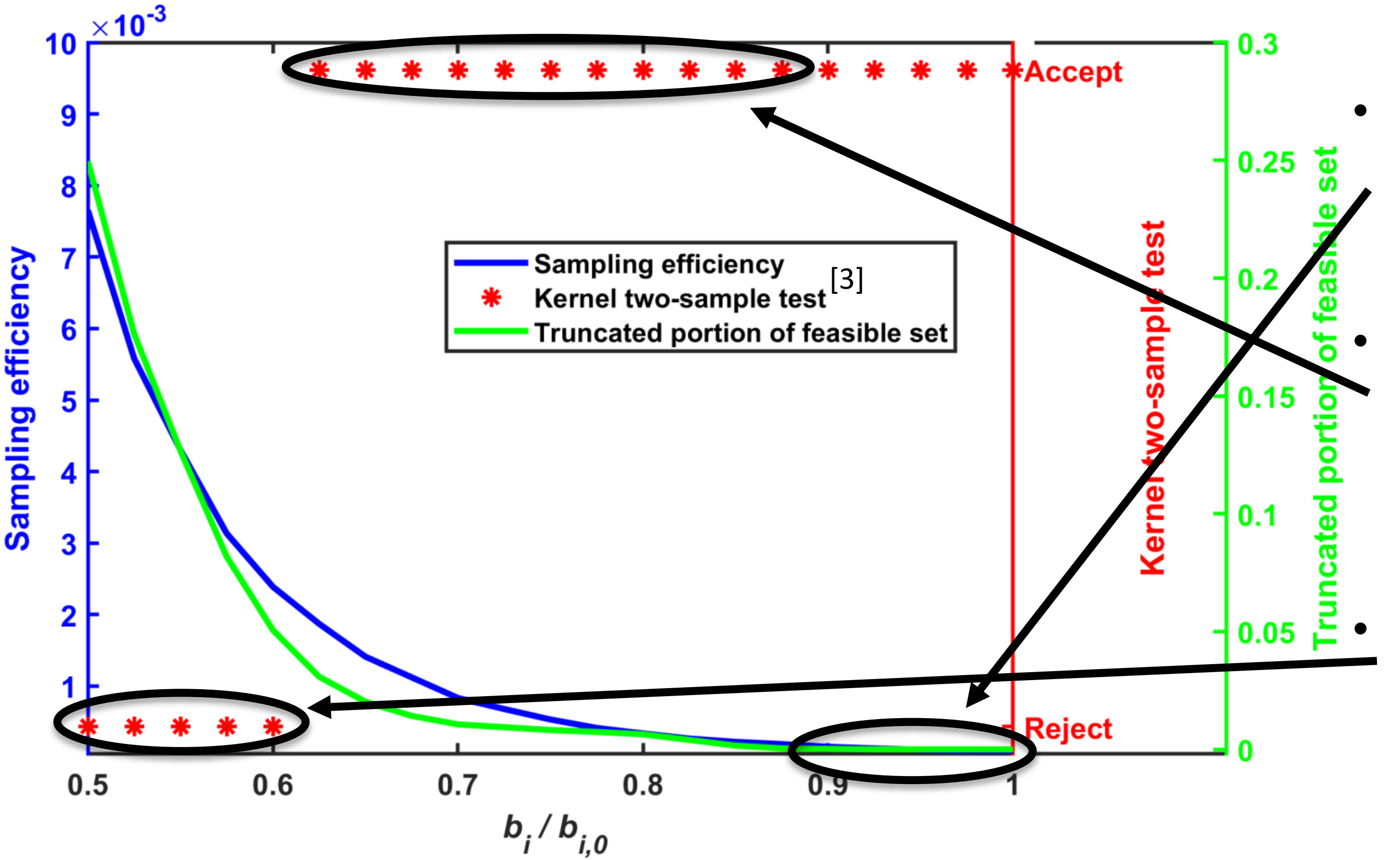



- A polytope is defined as:

$$a_i^T(x - x_0) \leq b_i$$
$$i = 1, 2, \ldots, n$$

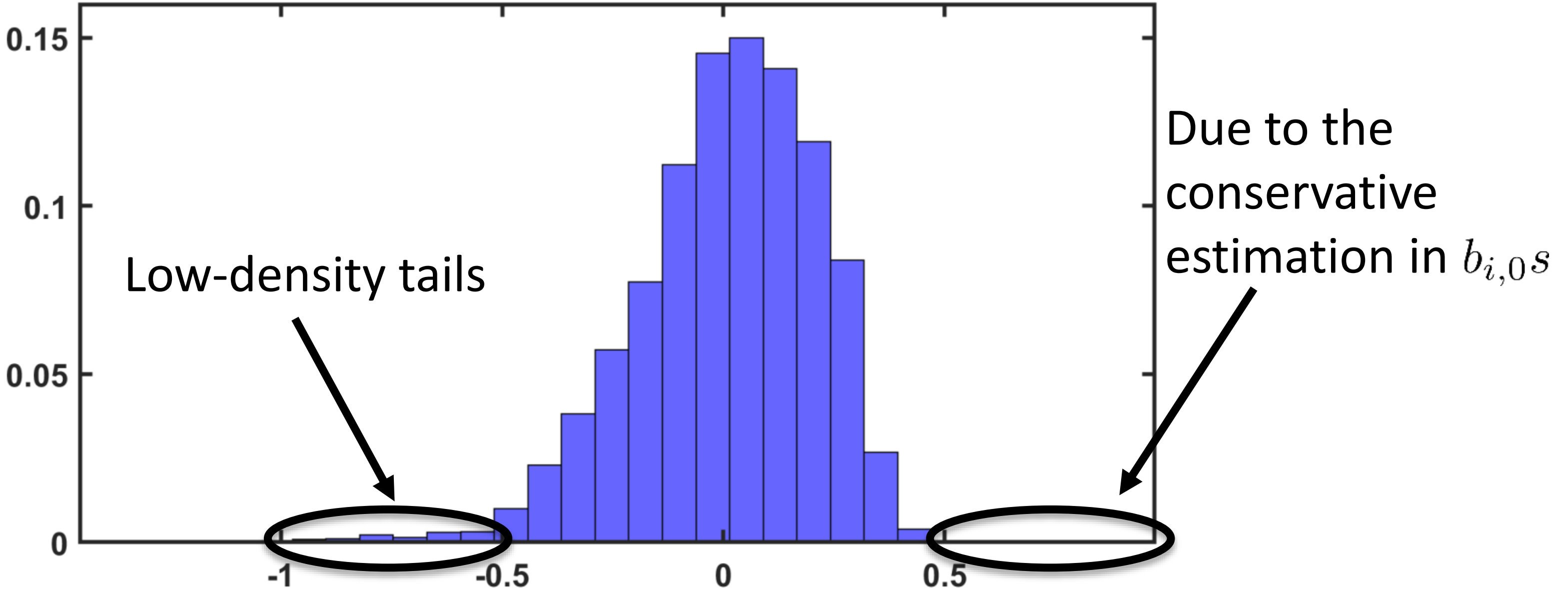
- The $b_{i,0}$s are calculated from B2BDC and represents a bounding polytope
- $b_i$s vary gradually to generate smaller polytopes

# Effect of truncation




- Region of no truncation, improved sampling efficiency.
- Region of increased truncation, improved efficiency but acceptable approximation
- Region of unacceptable approximation

# Check of directional histograms

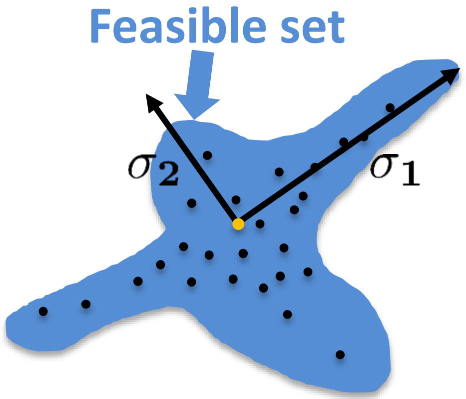



- This is observed along all the directions defining the polytope
- The distribution has zero-density regions
- The distribution has low-density tail regions

# Principal component analysis (PCA)

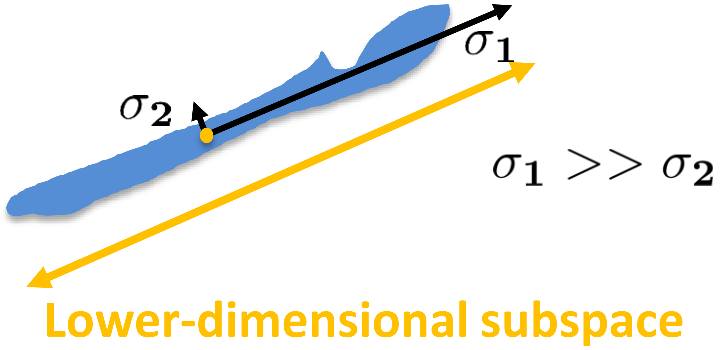
**Procedure:**

- collect RW samples from the feasible set
- conduct PCA on RW samples
- find a subspace based on PCA result
- generate uniform samples in the subspace

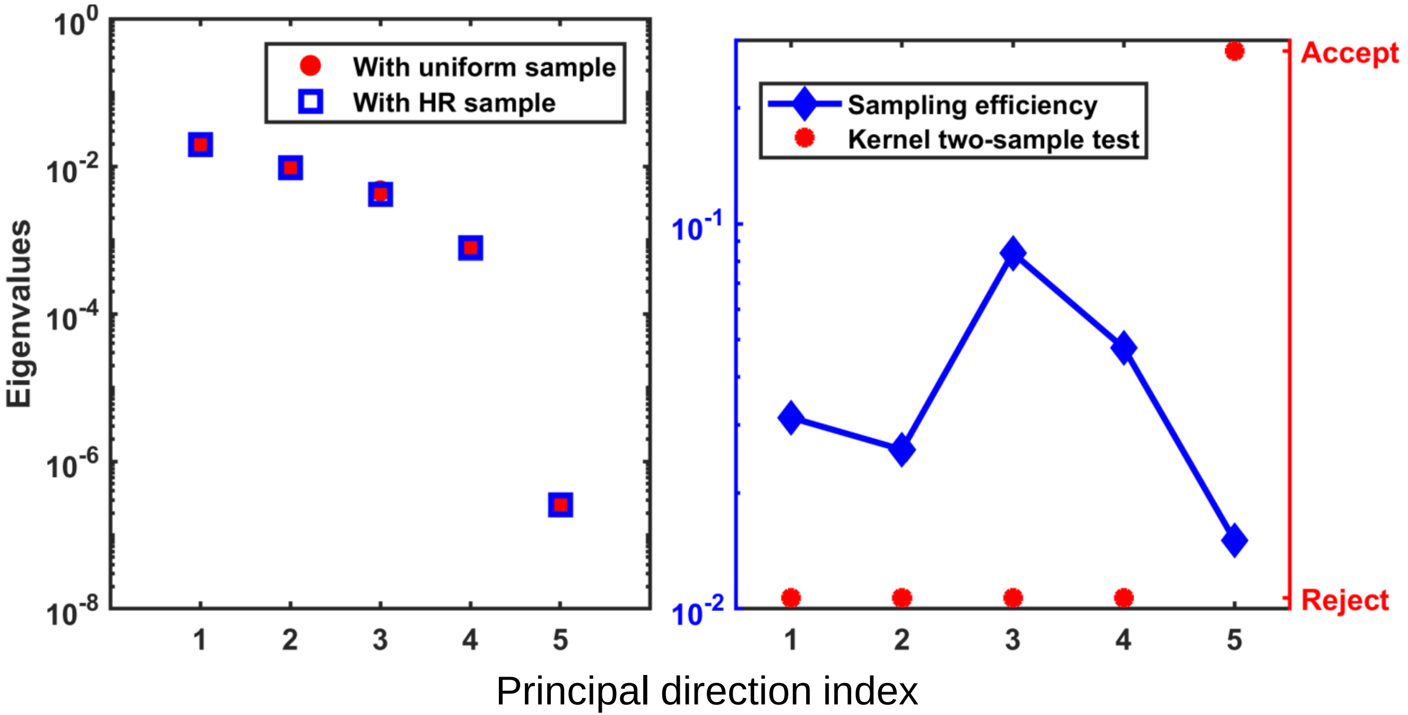



**Pros & Cons**

- improves sampling efficiency **significantly**
- works only if feasible set approximates a lower-dimensional manifold/subspace



# Effect of dimension reduction




- efficiency is affected mostly by problem dimension the (2.96e-5 in full dimension)
- returned samples approximate the desired distribution with acceptable accuracy only when the smallest principal direction is truncated

# Summary

- We developed methods to generate uniformly distributed samples of a feasible set
- Truncation strategy and PCA further improves the sampling efficiency of the method
- Numerical results support an advantageous efficiency-accuracy trade-off of the proposed approximation techniques

# Reference

[1] Frenklach, Michael, et al. "Comparison of statistical and deterministic frameworks of uncertainty quantification." SIAM/ASA Journal on Uncertainty Quantification 4.1 (2016): 875-901.

[2] Smith, Robert L. "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions." Operations Research 32.6 (1984): 1296-1308.

[3] Gretton, Arthur, et al. "A kernel two-sample test." Journal of Machine Learning Research 13.Mar (2012): 723-773.