Generalized Gauss Inequalities in Optimization and Control

Paul Goulart

20 Aug, 2014

Automatic Control Laboratory, ETH Zürich

http://control.ee.ethz.ch

Some Collaborators

Bart Van Parys
ETH Zurich

Daniel Kuhn
EPFL

Manfred Morari
ETH Zurich

An Introduction

Controlling the probability and severity of extreme events

- For power systems, monetary policy, water distribution, building control...

How much information is required to make good decisions?

- Two classical results from the 19th century.

A Generalized Gauss bounding problem and its solution.

Some applications and future directions.

Simple Example: How Tall are the Swiss?

Problem : What percentage p of the Swiss are taller than 190cm?

Assume only:

- Mean : $\mu = 171.5cm$
- Standard deviation : $\sigma = 7.5cm$

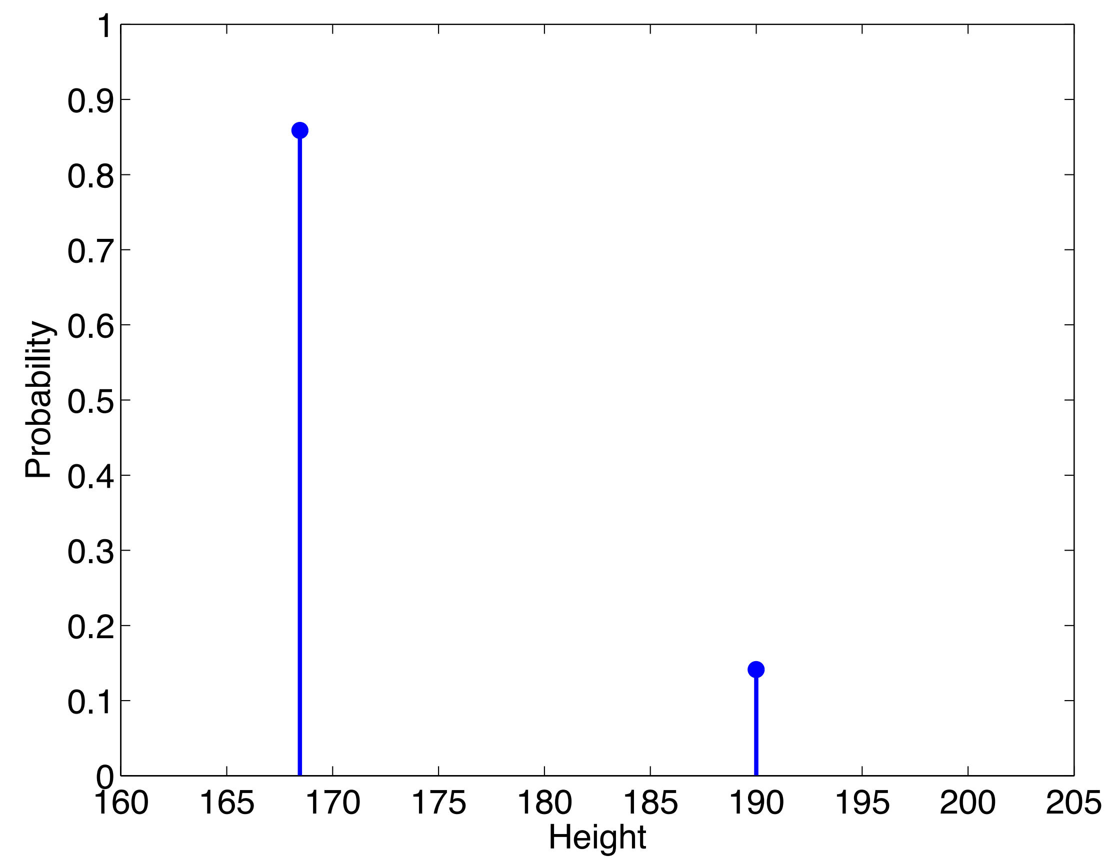
Statistics for social sciences dates to the astronomer Quetelet [1835].

Chebyshev (1867) : Can bound the worst case probability using the inequality:

$$\mathbb{P}\left(\mathbf{x} - \mu \geq k\sigma\right) \leq \frac{1}{1+k^2}$$

providing the worst-case estimate:

$$p \leq \frac{\sigma^2}{\sigma^2 + (190-\mu)^2} \approx 14.1\%$$



The Swiss according to Chebyshev

Image Courtesy of CSS Group

Simple Example: How Tall are the Swiss?

Problem : What percentage p of the Swiss are taller than 190cm?

Assume only:

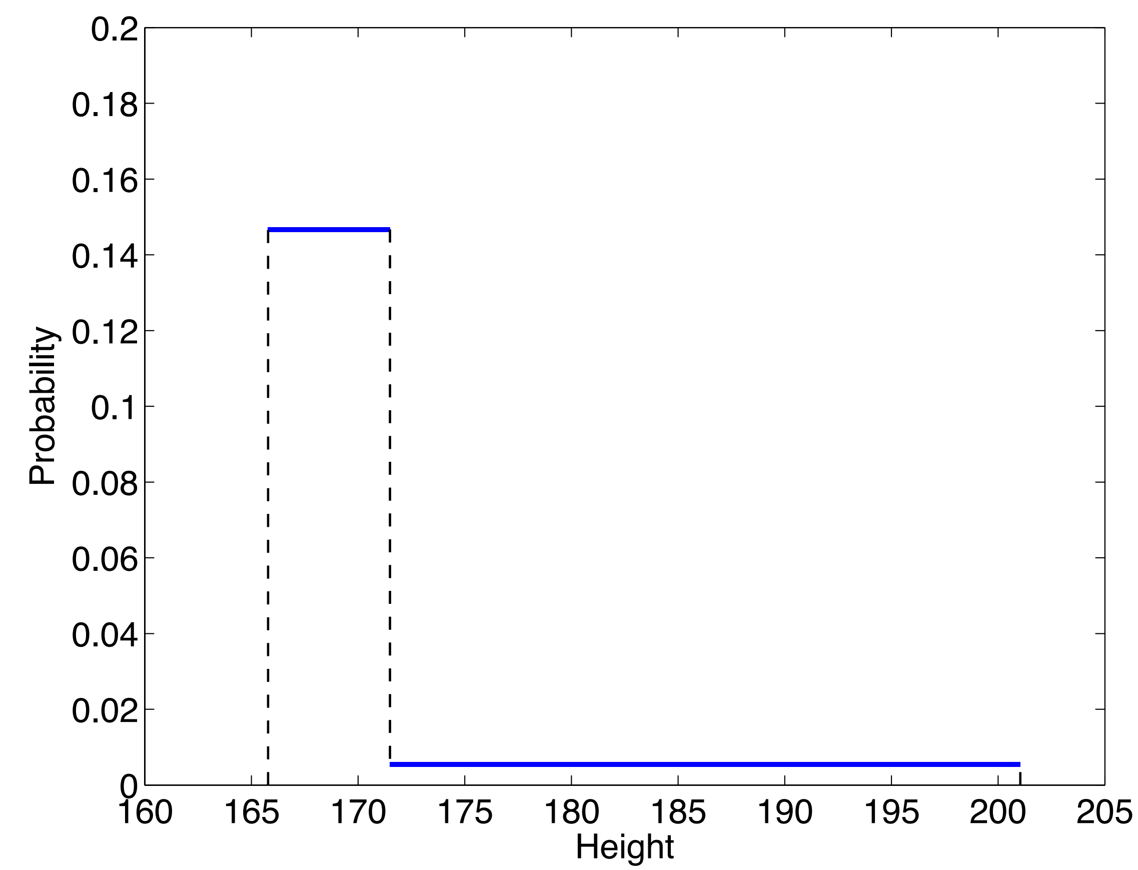
- Mean : $\mu = 171.5cm$
- Standard deviation : $\sigma = 7.5cm$

Statistics for social sciences dates to the astronomer Quetelet [1835].

Gauss (1821) : For a **unimodal** distribution, can bound the worst case probability using:

$$\mathbb{P}\left(\mathbf{x} - \mu \geq k\sigma\right) \leq \frac{4}{9}\frac{1}{1+k^2} = 6.3\%$$

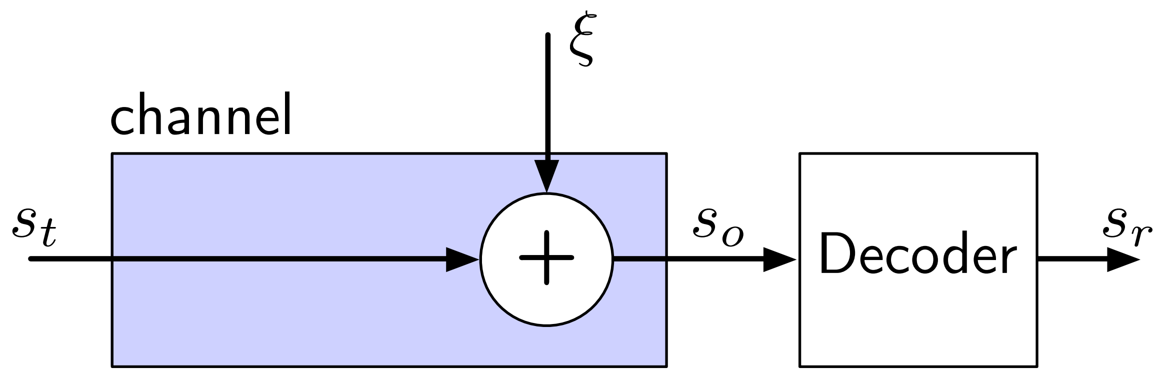
- A unimodal probability density function is greater when closer to the mode.



Example : Digital Communication Limits

Problem : Define a set of message $\mathcal{S} = \{s_1, \ldots, s_c\} \subseteq \mathbb{R}^2$. The messages are communicated over a channel with additive noise.

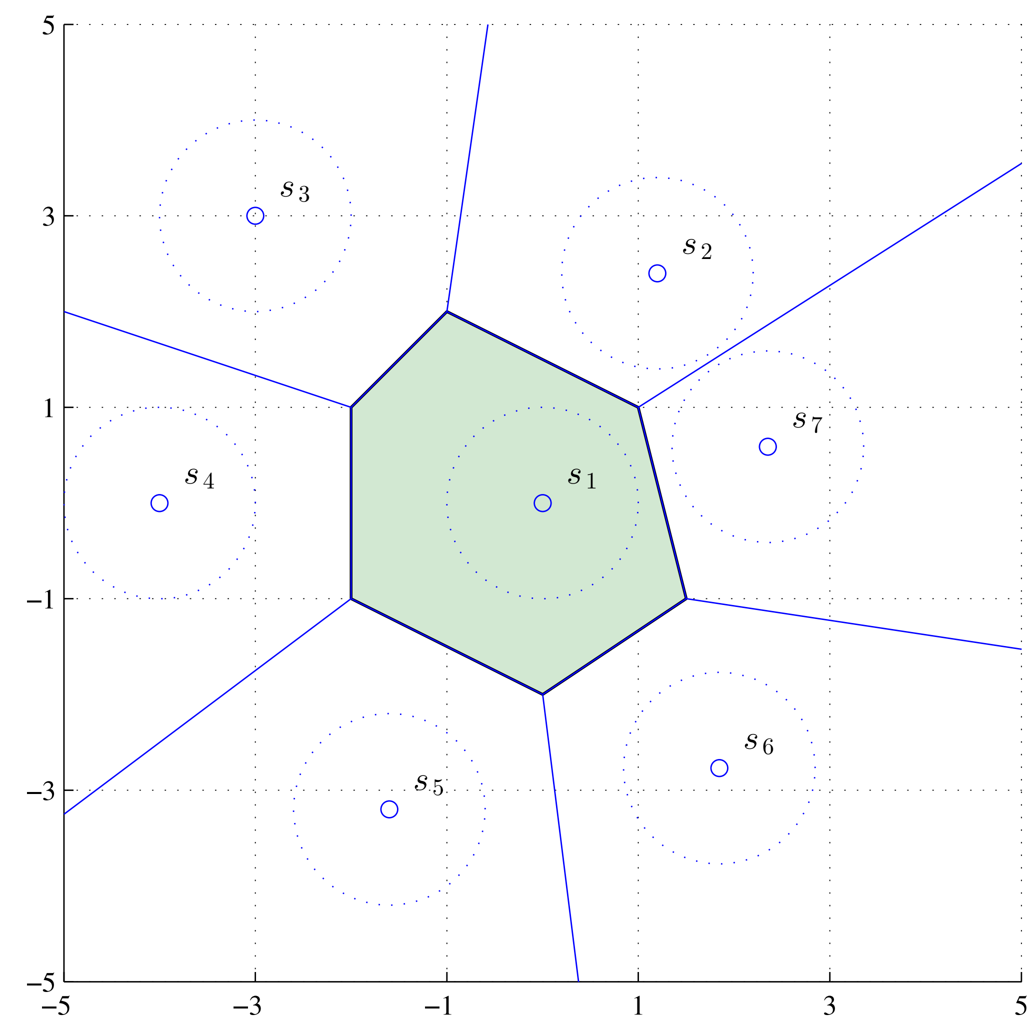
Find a bound on the rate of correct signal transmission.



Probability of correct transmission is :

$$p = 1 - \mathbb{P}\left[(s_i + \xi) \notin C_i\right]$$

where C_i is the set of outputs that should be decoded as s_i.

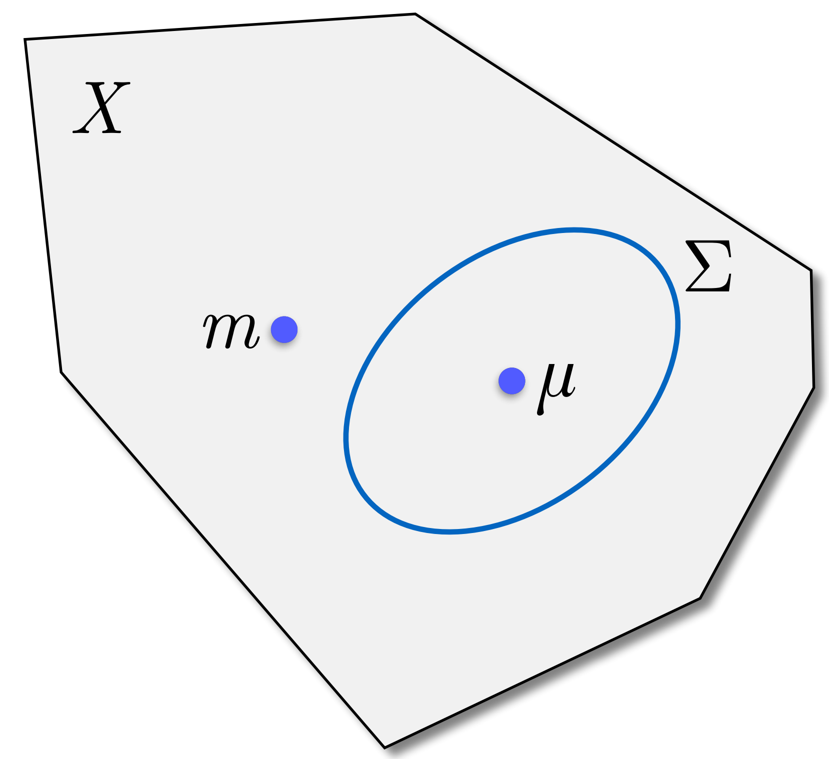


Generalized Chebyshev Bounds

Problem : What is the worst case probability that the random variable $\mathbf{x}$ falls outside of a set $X \subseteq \mathbb{R}^n$?

Assume only:

- Mean : $\mu \in \mathbb{R}^n$
- Second moment : $S = \Sigma + \mu\mu^\top$
- Open set : $X = \left\{x \;\middle|\; a_i^\top x < b_i,\; i = 1, \ldots k\right\}$



Classical Chebyshev inequality is a special case:

$$X = \{x \in \mathbb{R} \mid x \leq \kappa\sigma + \mu,\; -x \leq \kappa\sigma - \mu\}$$

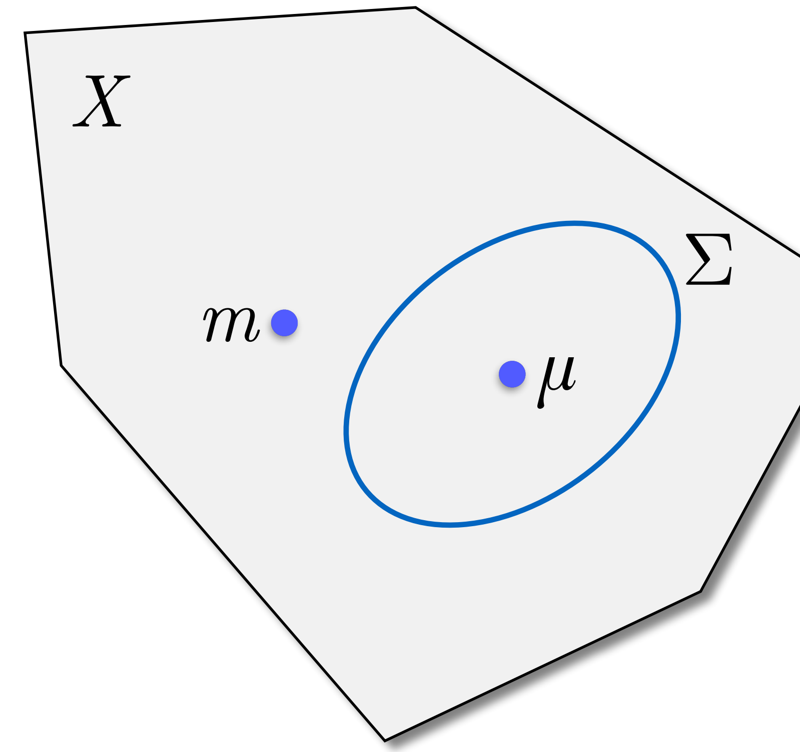
with $\Sigma = \sigma^2 > 0$.

Generalized Chebyshev Bounds

Problem : What is the worst case probability that the random variable $\mathbf{x}$ falls outside of a set $X \subseteq \mathbb{R}^n$?

Assume only:

- Mean : $\mu \in \mathbb{R}^n$
- Second moment : $S = \Sigma + \mu\mu^\top$
- Open set : $X = \left\{x \;\middle|\; a_i^\top x < b_i,\ i = 1, \ldots k\right\}$



An infinite-dimensional optimization problem:

$$\begin{aligned} \max_{\mathbb{P}} \quad & \mathbb{P}(\mathbf{x} \notin X) \\ \text{subject to:} \quad & \mathbb{P} \in \mathcal{P}(\mu, S) \end{aligned}$$

$\mathcal{P}(\mu, S) =$ Set of distributions consistent with moments (μ, S)

Example : Digital Communication Limits

Problem : Define a set of message $\mathcal{S} = \{s_1, \ldots, s_c\} \subseteq \mathbb{R}^2$. The messages are communicated over a channel with additive noise.

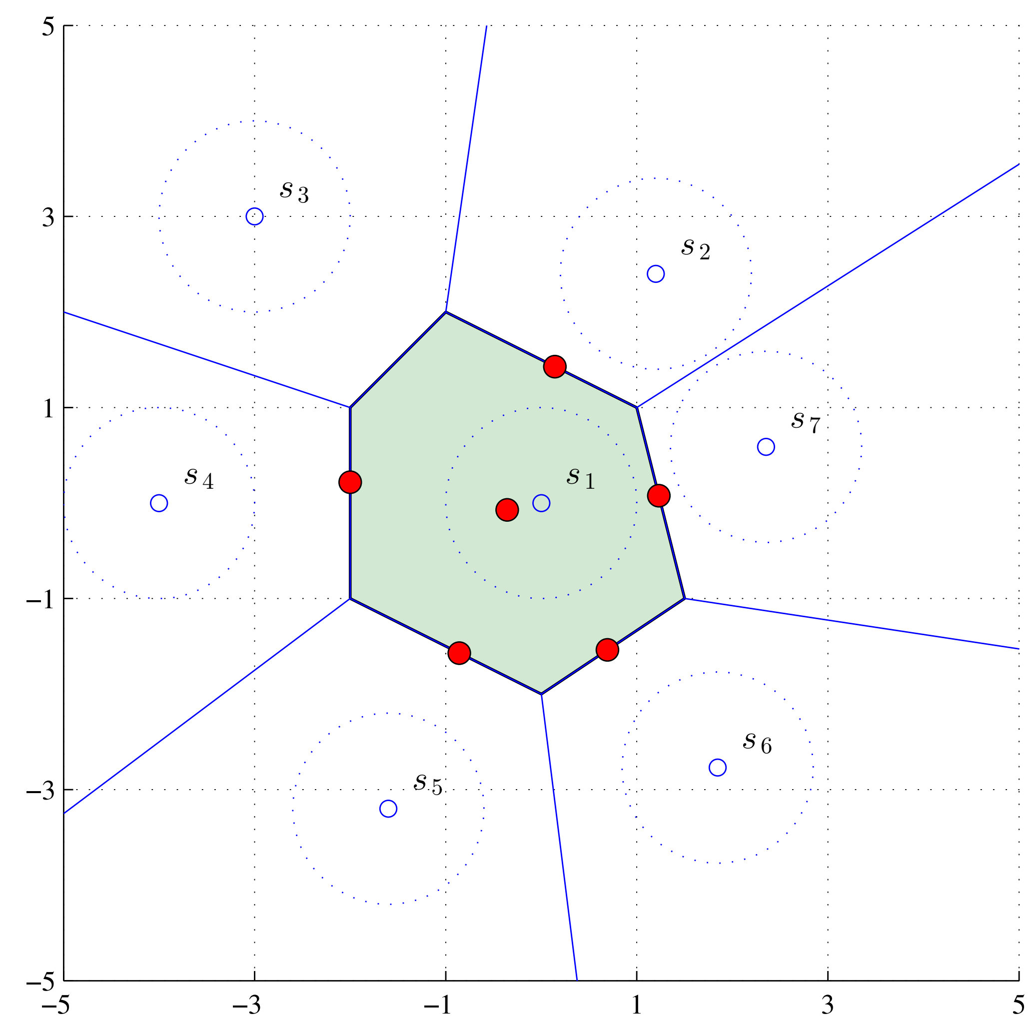
Find a bound on the rate of correct signal transmission.

Solution : The worst case error probability

$$p = 1 - \mathbb{P}\left[(s_i + \xi) \notin C_i\right]$$

is achieved by a point distribution.

- λ_i : Mass of each point
- z_i/λ_i : Location of point masses



Generalized Chebyshev Bounds (SDP form)

Theorem Vandenberghe and Boyd (2007)

The Chebyshev bounding problem :

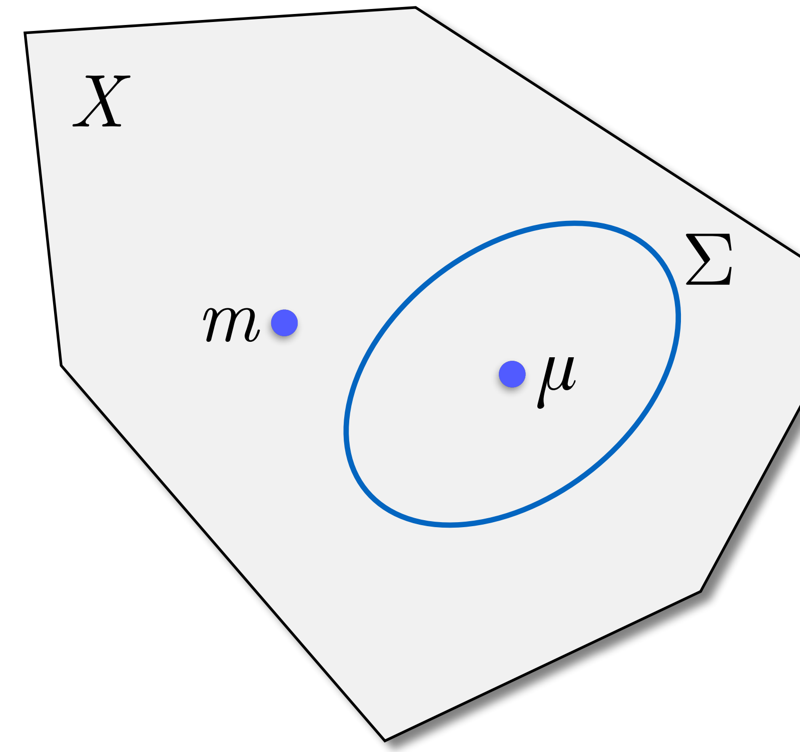
$$\begin{aligned} \max_{\mathbb{P}} \quad & \mathbb{P}(\mathbf{x} \notin X) \\ \text{subject to:} \quad & \mathbb{P} \in \mathcal{P}(\mu, S) \end{aligned}$$

is equivalent to a finite-dimensional convex problem:

$$\begin{aligned} \max_{\{Z_i, z_i, \lambda_i\}} \quad & \sum_{i=1}^{k} \lambda_i \\ \text{subject to:} \quad & \sum_{i=1}^{k} \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \\ & \left.\begin{aligned} \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 \\ a_i^\top z_i \geq \lambda_i b_i \end{aligned}\right\} \forall i = 1, \dots, k \end{aligned}$$

Generalized Gauss Bounds

Problem : What is the worst case probability that the random variable $\mathbf{x}$ falls outside of a set $X \subseteq \mathbb{R}^n$?



Assume only:

- Mean : $\mu \in \mathbb{R}^n$
- Second moment : $S = \Sigma + \mu\mu^\top$
- Open set : $X = \left\{x \;\middle|\; a_i^\top x < b_i,\ i = 1, \ldots k\right\}$
- Distribution is unimodal with mode m

An infinite-dimensional optimization problem:

$$\max_{\mathbb{P}} \quad \mathbb{P}(\mathbf{x} \notin X)$$

$$\text{subject to:} \quad \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_n$$

$\mathcal{P}(\mu, S) =$ Set of distributions consistent with moments (μ, S)

$\mathcal{P}_n =$ Set of unimodal distributions

Unimodal Distributions in $\mathbb{R}^n$

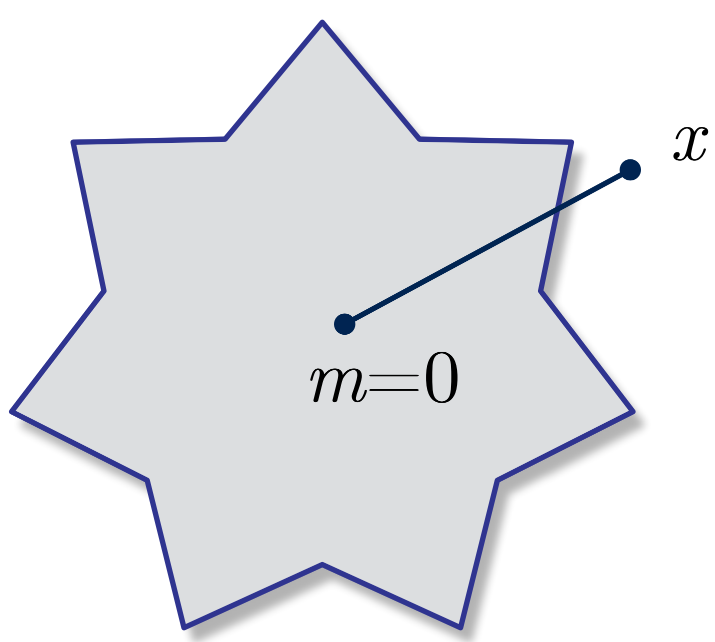
Definition : If $\mathbb{P}$ has mode 0 and continuous density f, then $\mathbb{P}$ is **unimodal** if

$$f(tx), \quad t \geq 0$$

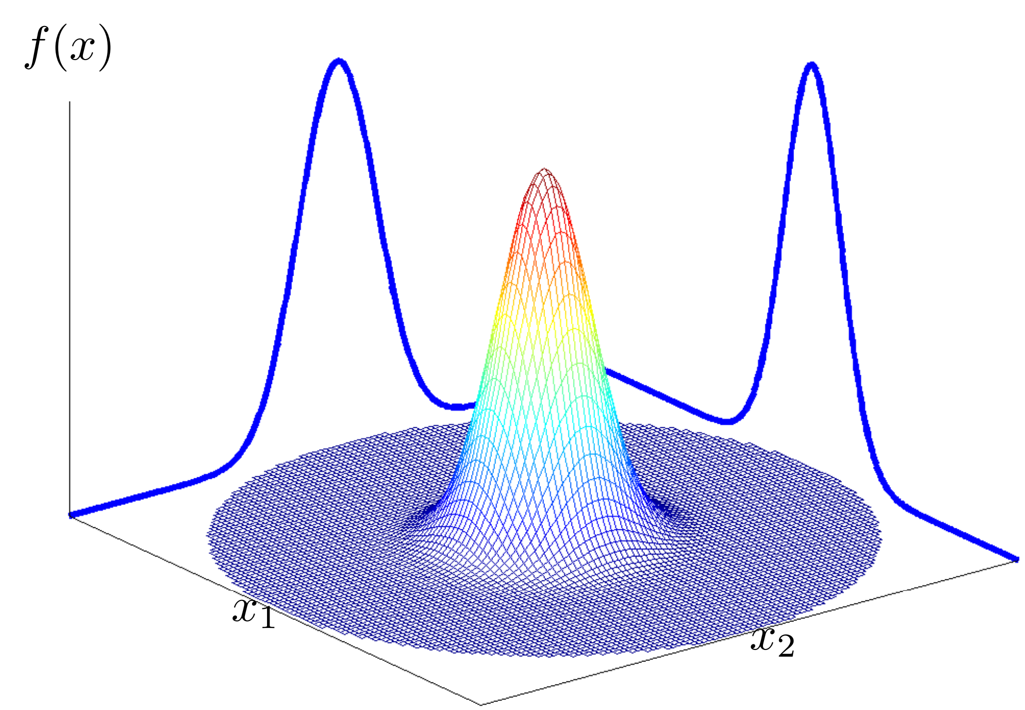
is **non-increasing** in t for any x. The set of such measures is called $\mathcal{P}_*$.

Most common distributions are unimodal :

- Beta, Chi, Dirichlet, Erlang, Fisher, Gamma, Hyperbolic, Inverse-Gauss, Laplace,....



Uniform distribution on a star-shaped set



Normal distribution in $\mathbb{R}^2$

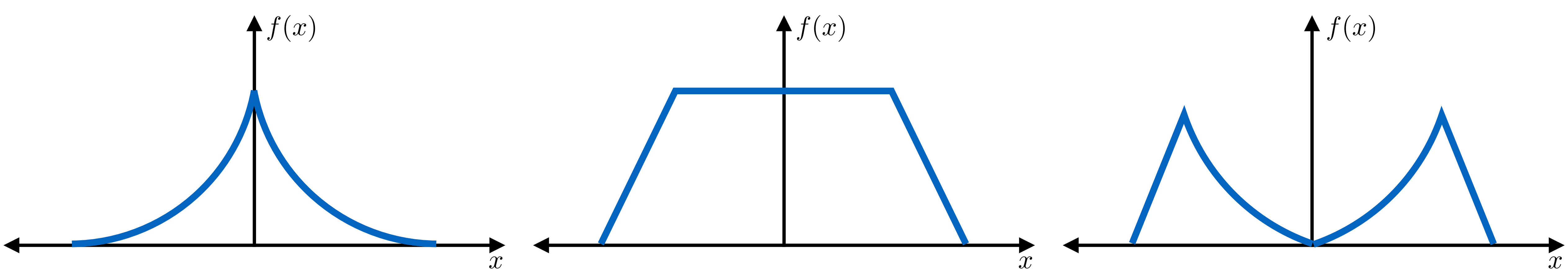
How Unimodal is a Measure?

Definition : If $\mathbb{P}$ has mode 0 and continuous density f, then $\mathbb{P}$ is **α-unimodal** if

$$\frac{f(tx)}{t^{\alpha-n}}, \quad t \geq 0$$

is **non-increasing** in t for any x. The set of such measures is called $\mathcal{P}_\alpha$.

- Rate of increase/decrease along rays is controlled by α
- $\mathcal{P}_n$: Worst-case uniform distributions (Gauss)
- $\lim_{\alpha\to\infty} \mathcal{P}_\alpha$: All possible distributions (Chebyshev)



$\alpha < n$ $\quad$ $\alpha = n$ $\quad$ $\alpha > n$

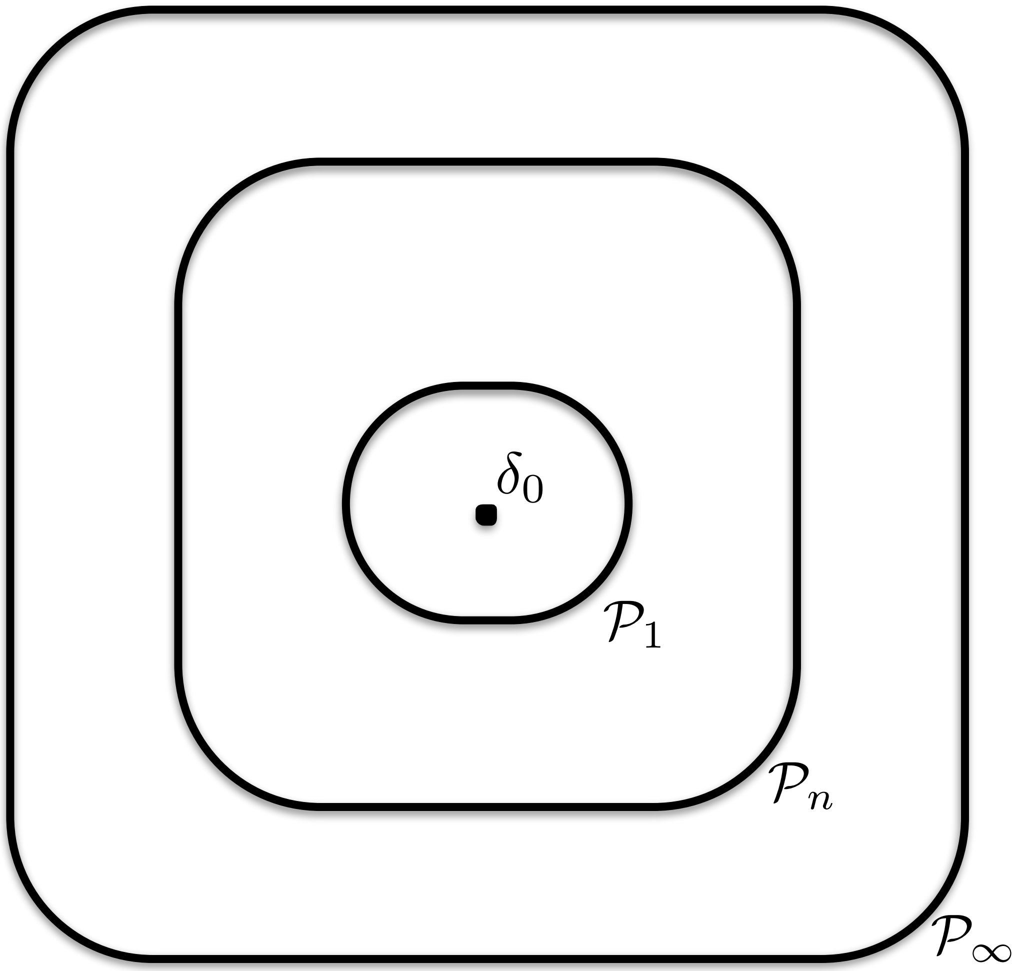
Some properties of α-unimodal measures

Some basic properties:

- $\mathcal{P}_\alpha$ is convex for every $\alpha \geq 1$
- The sets P_α are nested, i.e.

$$\mathcal{P}_\alpha \subset P'_\alpha \text{ for } \alpha \leq \alpha'$$

- δ_0 is the 'most unimodal' distribution
- P_∞ is the 'least unimodal' set of distributions



Some standard models:

- Deterministic problems : correspond to $\mathcal{P}_0$
- Robust optimisation problems : correspond to $\mathcal{P}_\infty$ with no moment information

From Infinite- to Finite-Dimensions

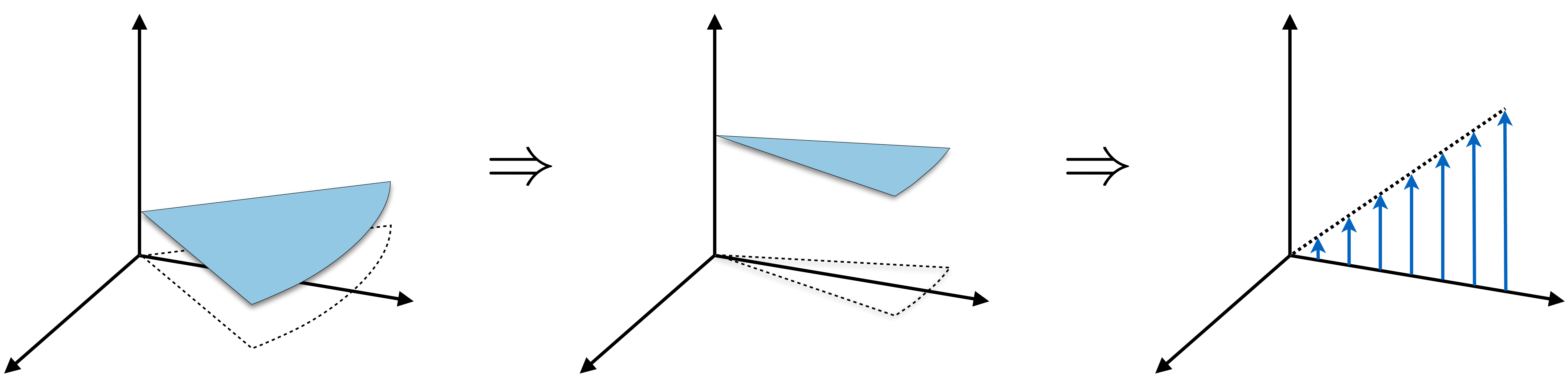
Problem is an infinite-dimensional linear program:

$$\max_{\mathbb{P}} \quad \mathbb{P}(\mathbf{x} \notin X)$$
$$\text{subject to:} \quad \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$$

Main Idea : Optimize over the **extreme points** of $\mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$

1) The extreme points of $\mathcal{P}_\alpha$ are measures along lines

$$\delta_x^\alpha([0, tx]) = t^\alpha, \quad \forall t \in [0,\ 1]$$



From Infinite- to Finite-Dimensions

Problem is an infinite-dimensional linear program:

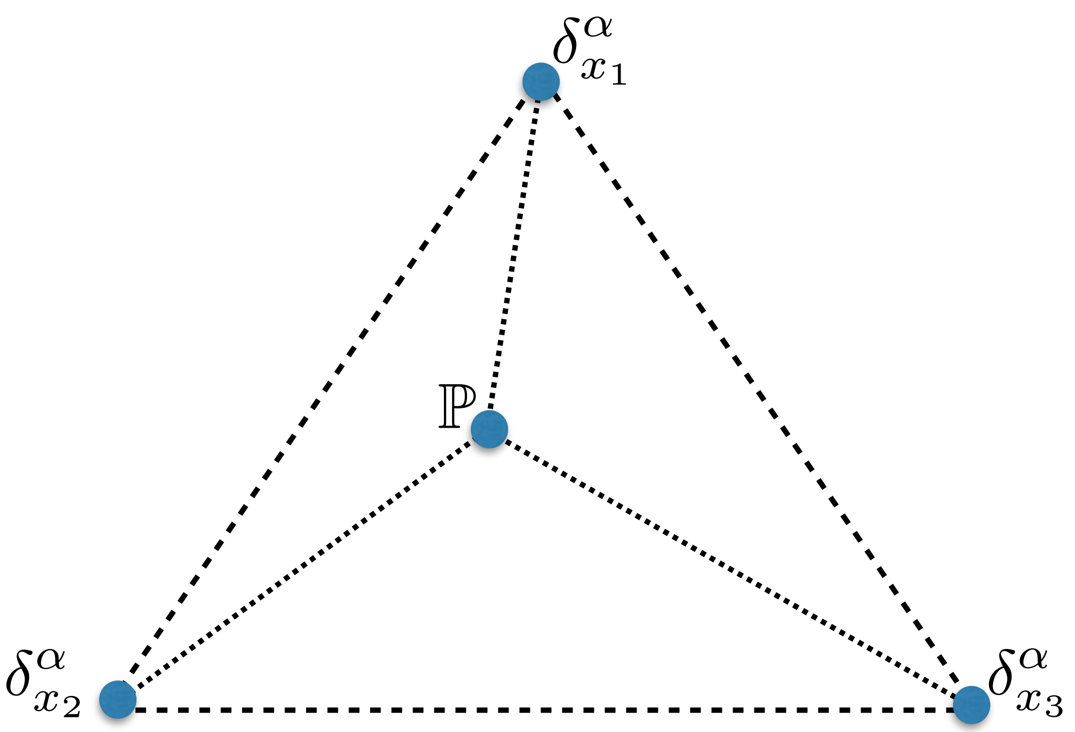
$$\max_{\mathbb{P}} \quad \mathbb{P}(\mathbf{x} \notin X)$$
$$\text{subject to:} \quad \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$$

Main Idea : Optimize over the **extreme points** of $\mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$

2) Every α-unimodal distribution is in the convex hull of the extreme points [Dharnadhikari, 1988]

$$\mathbb{P} \in \mathcal{P}_\alpha \iff \exists \mu \in \mathcal{P}_\infty \, : \, \mathbb{P}(\cdot) = \int \delta_x^\alpha(\cdot)\mu(\mathrm{d}x)$$

This is a **Choquet representation** for $\mathbb{P}$.



From Infinite- to Finite-Dimensions

Problem is an infinite-dimensional linear program:

$$\max_{\mathbb{P}} \quad \mathbb{P}(\mathbf{x} \notin X)$$
$$\text{subject to:} \quad \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$$

Main Idea : Optimize over the **extreme points** of $\mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$

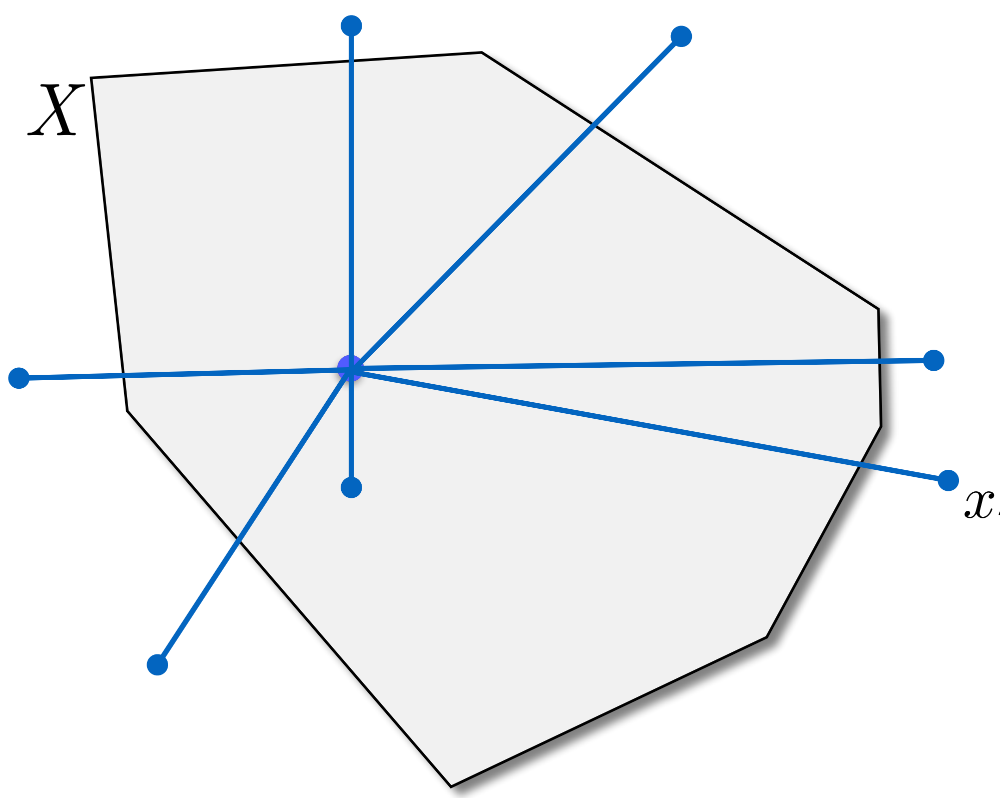
3) The worst-case distribution is supported on a finite set of extreme points

- Fix the distribution structure to

$$\mathbb{P} = \sum_{i=1}^{k} \lambda_i \cdot \delta_{x_i}^{\alpha}(x \notin X)$$

- Maximize the sum of the violations

$$\delta_{x_i}^{\alpha}(x \notin X) = 1 - \left(\frac{b_i}{a_i^\top x_i}\right)^{\alpha}$$



Chebyshev, Gauss and Everything in Between

Theorem (Van Parys, Goulart, Kuhn 2014)

For any $\alpha \geq 1$, the infinite-dimensional problem:

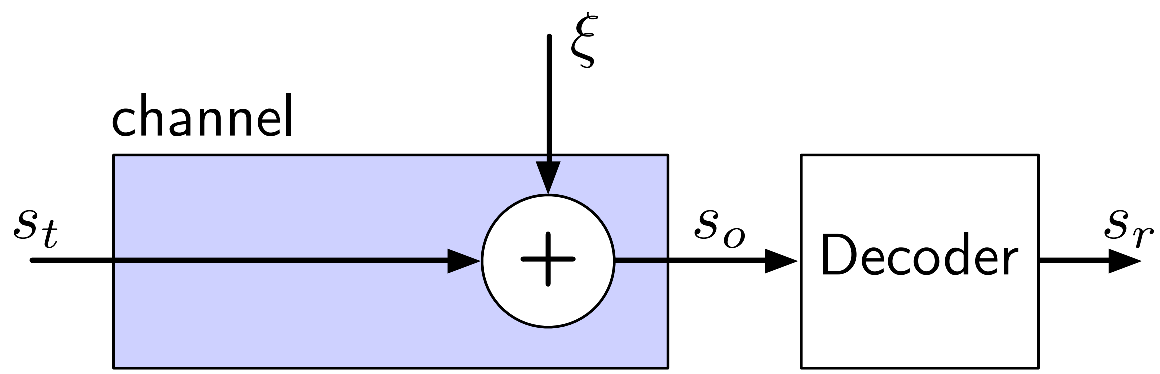
$$\begin{array}{ll} \max\limits_{\mathbb{P}} & \mathbb{P}(\mathbf{x} \notin X) \\ \text{subject to:} & \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha \end{array}$$

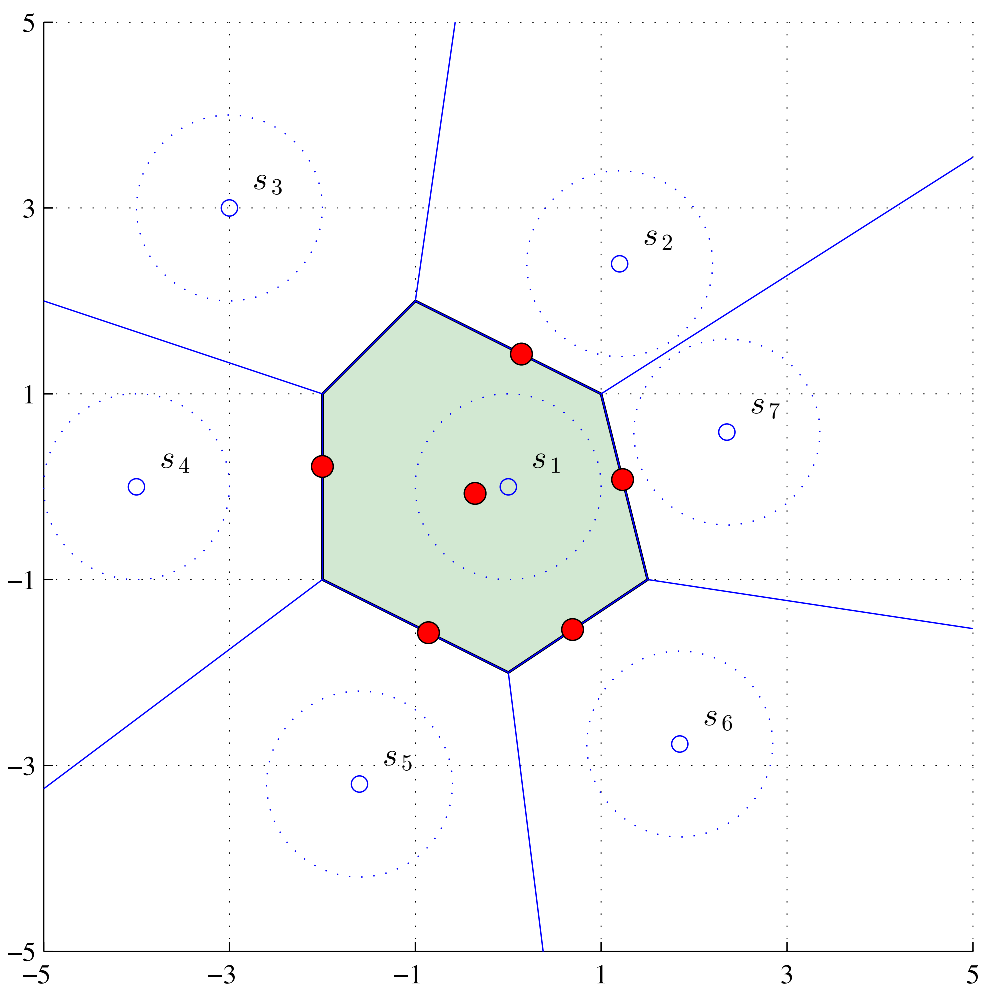
is equivalent to a finite-dimensional convex problem.

$$\begin{array}{ll} \max\limits_{\{Z_i, z_i, \tau_i, \lambda_i\}} & \sum\limits_{i=1}^{k} (\lambda_i - \tau_i) \\ \text{subject to:} & \sum\limits_{i=1}^{k} \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} \\ & \left.\begin{array}{c} \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 \\ a^\top z_i \geq 0 \\ \tau_i (a_i^\top z_i)^\alpha \geq \lambda_i^{\alpha+1} b_i^\alpha \end{array}\right\} \forall i = 1, \ldots, k \end{array}$$

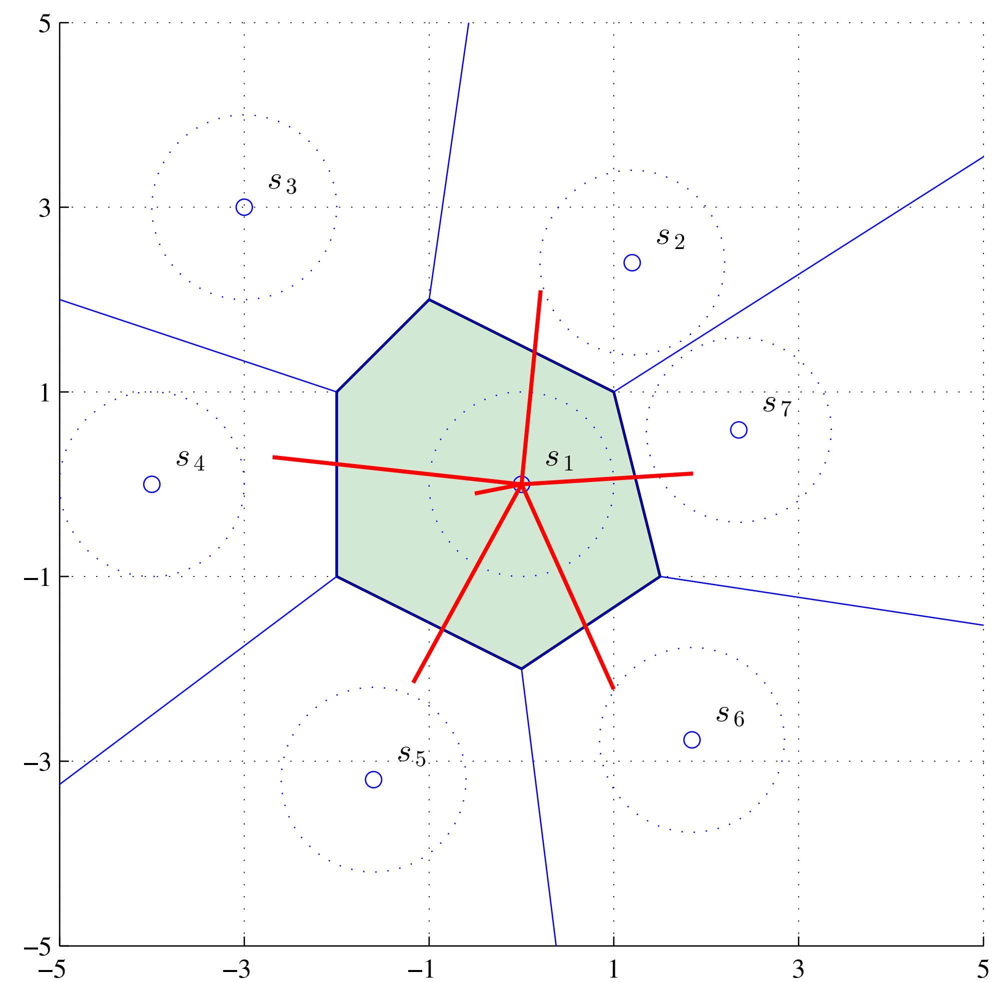
Example : Digital Communication Limits

Problem : What is the rate of correct signal detection for the central cell?





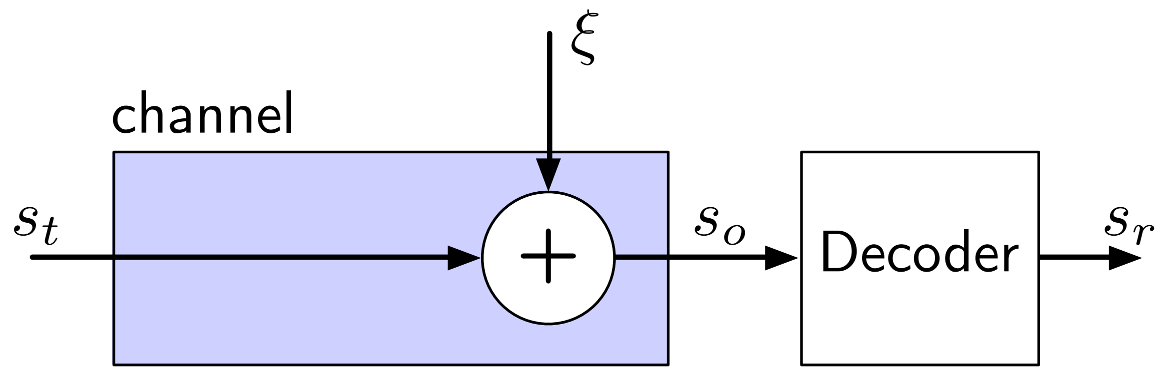
$\alpha = \infty$, (Chebyshev-like)

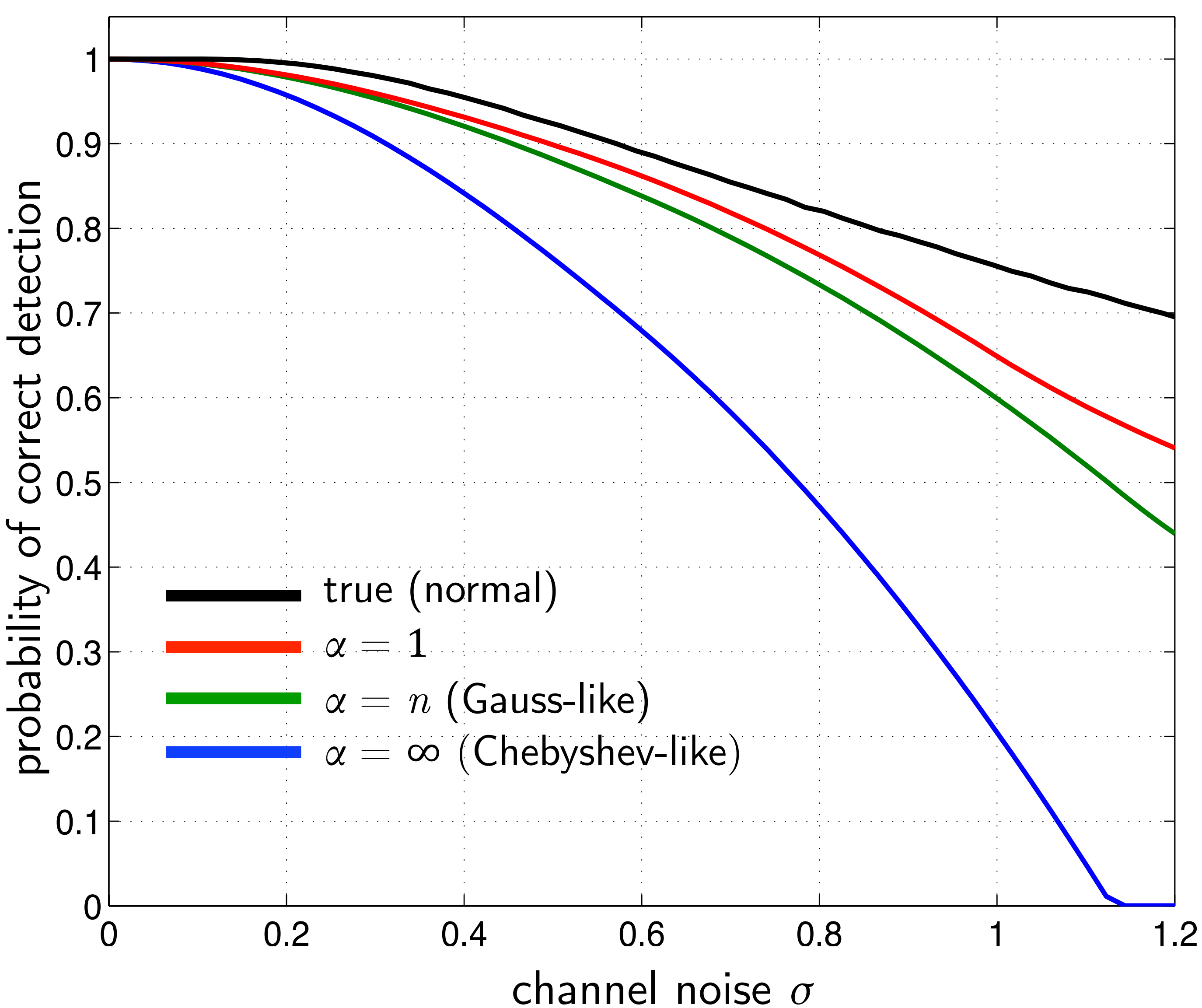


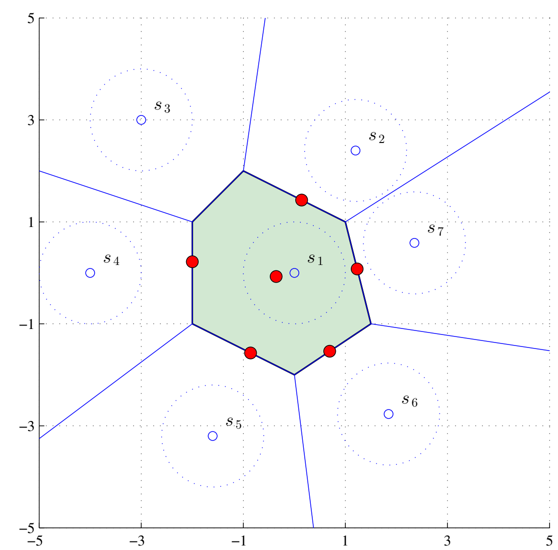
$\alpha = n$, (Gauss-like)

Example : Digital Communication Limits (cont)

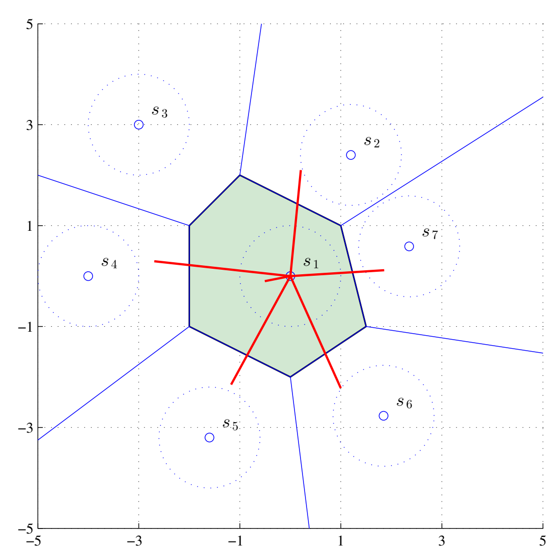
Problem : What is the rate of correct signal detection for the central cell?







worst-case noise
$\alpha = \infty$
(Chebyshev-like)



worst-case noise
$\alpha = n$
(Gauss-like)

Some immediate extensions

What other structural conditions can be added?

- **Moment ambiguity** : Moments (μ, S) need not be perfectly known

$$\begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \in \mathcal{M}$$

- **Multimodality** : Can model multiple peaks in the distribution at points m_i

$$\mathbb{P} = \sum_{i=1}^{m} \gamma_i \mathbb{P}_i, \quad \gamma \in \Gamma$$

- **Bounded support** : Can require measure to be zero outside set $\mathcal{B}$

$$\mathbb{P}(x \in \mathcal{B}) = 1$$

- **Symmetry** : Can restrict $\mathbb{P}$ to be a symmetric distribution

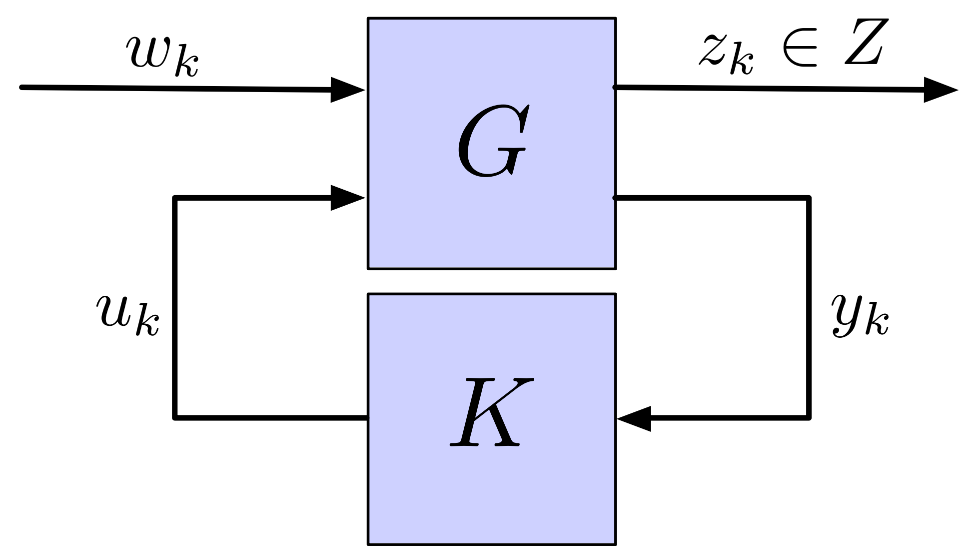
Applications: Controlling Extreme Events

Problem : Consider an LTI system driven by noise w_k with known moments.

Find a controller that bounds the probability of outliers at the output.

Current Application Areas :

- Control of Power Networks (with ABB)
- Control of Water Distribution (with IWB)
- Monetary Policy (with Swiss National Bank)
- Wind Turbine Control (with Imperial College)



Applications: Controlling Extreme Events

Problem : Consider an LTI system driven by noise w_k with known moments.

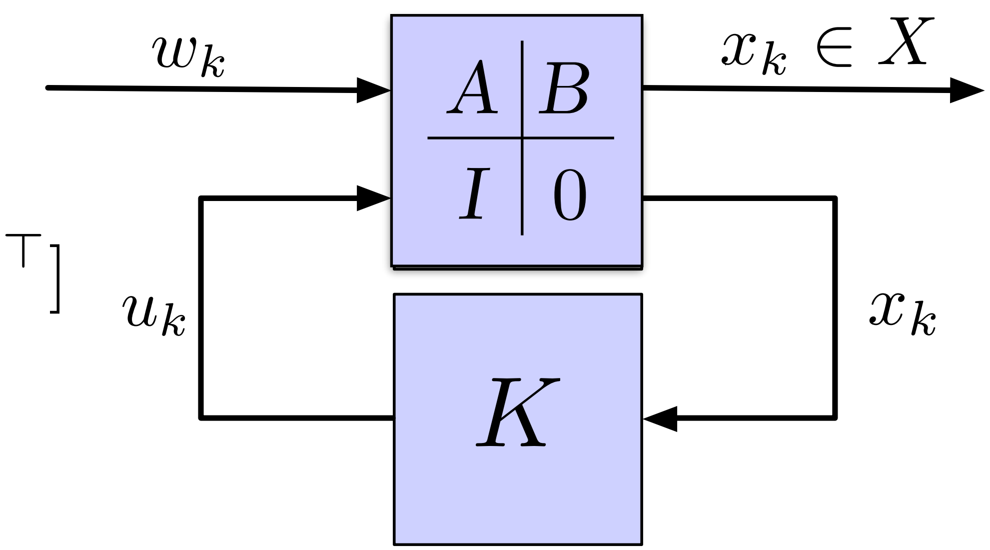
Find a controller that bounds the probability of outliers at the output.

State Feedback Version:

Steady-state moments satisfy relation

$$\mathbb{E}[xx^\top] = (A+BK)\mathbb{E}[xx^\top](A+BK)^\top + \mathbb{E}[ww^\top]$$

(if zero mean noise, K stabilizing).



Bounding steady-state rate of violation $\mathbb{P}(x \notin X)$ is a Chebyshev-like problem.

Applications: Controlling Extreme Events

Problem : Consider an LTI system driven by noise w_k with known moments.

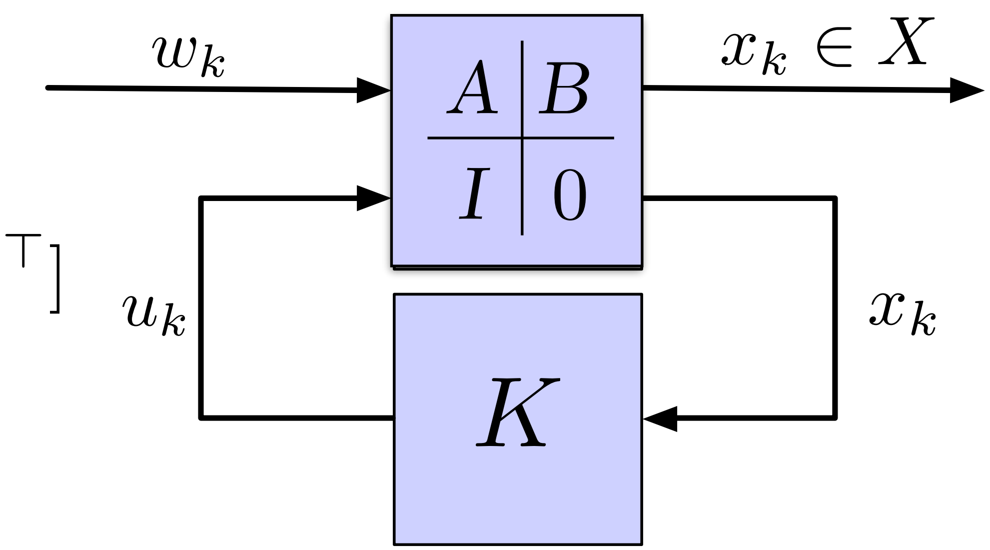
Find a controller that bounds the probability of outliers at the output.

State Feedback Version:

Steady-state moments satisfy relation

$$\mathbb{E}[xx^\top] = (A+BK)\mathbb{E}[xx^\top](A+BK)^\top + \mathbb{E}[ww^\top]$$

(if zero mean noise, K stabilizing).



What do we produce a bound for?

$$\sup_{\mathbb{P}\in\mathcal{P}_\infty}\left[\limsup_{k\to\infty}\ \mathbb{P}(x_k \notin X)\right]$$

The bounds produced are *tight*.

Applications : Machine Learning with SVMs

Problem : Given a collection of data points in $\mathbb{R}^n$ with associated labels, find a separating hyperplane (i.e. a linear classifier) with the lowest error rate.

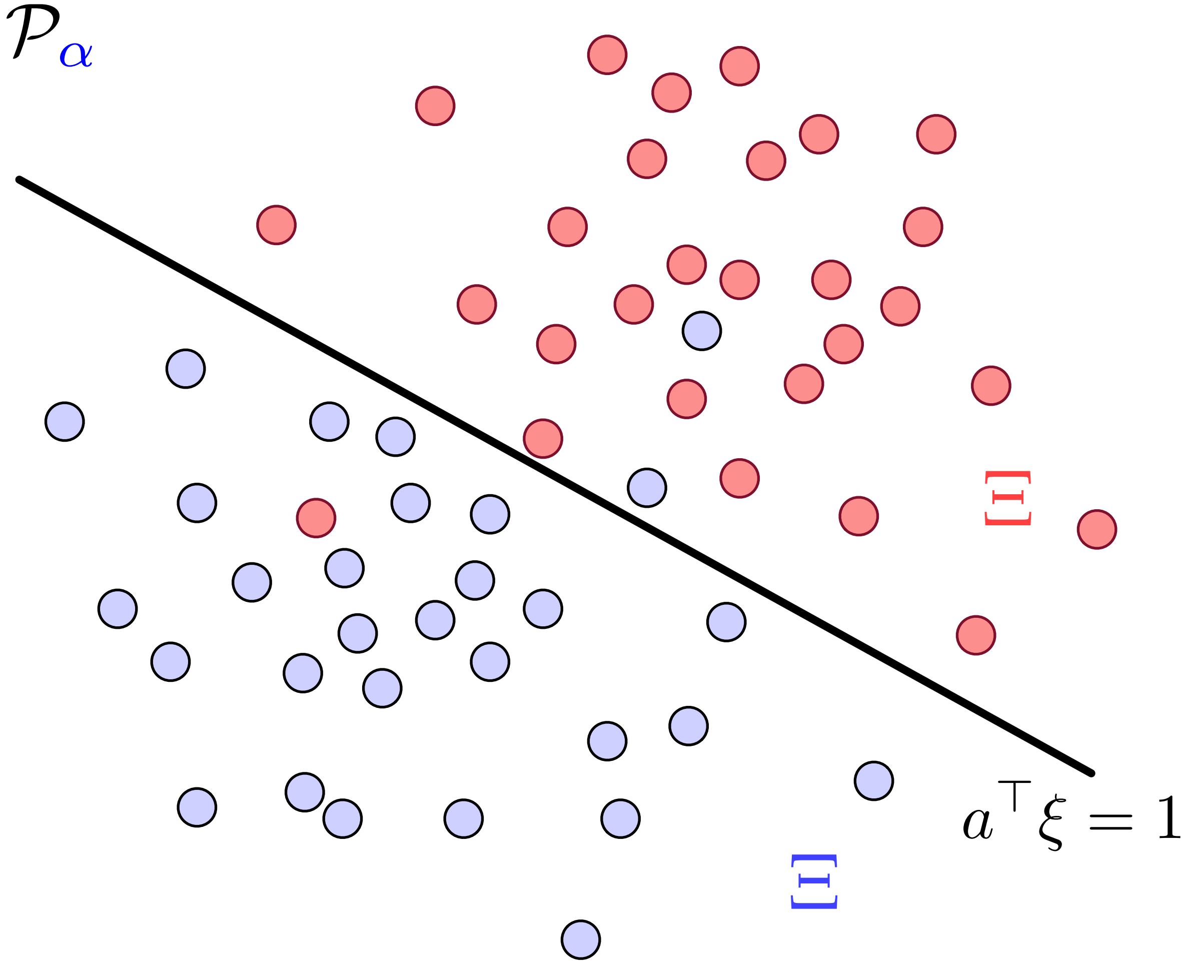
$$\min_{a,\epsilon} \quad \epsilon$$

subject to: $\mathbb{P}(\xi \notin \Xi) \leq \epsilon, \quad \forall \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$

$$\mathbb{P}(\xi \notin \Xi) \leq \epsilon, \quad \forall \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha$$

$$\Xi = \{\xi \mid a^\top \xi > 1\}$$

$$\Xi = \{\xi \mid a^\top \xi < 1\}$$



Current Application Area :

- Credit Card Fraud Detection (FP7 'Big Data' project with IBM / Feedzai)
- Blue/red data represent ≈1 billion credit card transactions, each with ≈40 unique data points

Modelling Severity of Violations

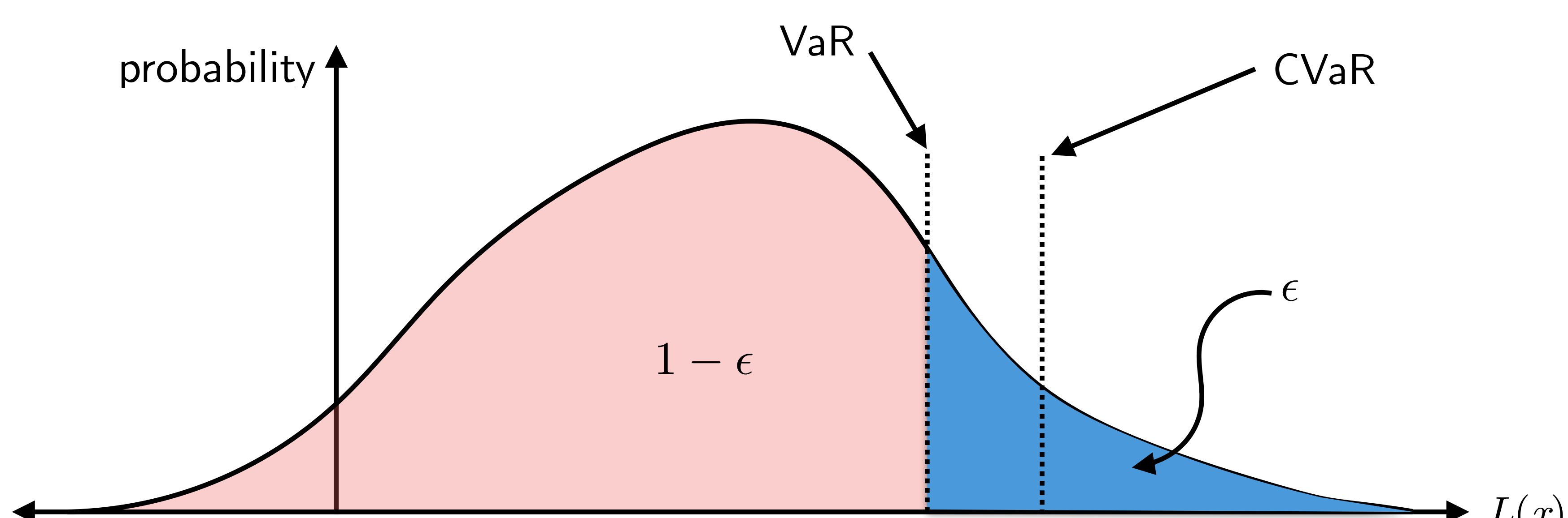
Bounds on Conditional Value at Risk (CVaR)

$$\begin{array}{rl} \max\limits_{\mathbb{P}} & \mathbb{P}\text{-CVaR}_\epsilon(L(x)) \\ \text{subject to:} & \mathbb{P} \in \mathcal{P}(\mu, S) \cap \mathcal{P}_\alpha \end{array}$$

CVaR computes the mean value of $L(x)$ in the worst $100 \cdot \epsilon\%$ of cases

CVaR ingredients :

- A *loss function* L measuring severity of outcomes
- A value ϵ characterising the fraction of outliers



Summary and Future Directions

A new fundamental result for bounding the probability of extreme events.

Extends and connects classical methods of Gauss and Chebyshev.

Applications in power systems, control design, machine learning, economics...

Many extensions and variations are possible:

- Moment Ambiguity, Multi-modality, Symmetry
- Bounds on severity of violations

Thanks