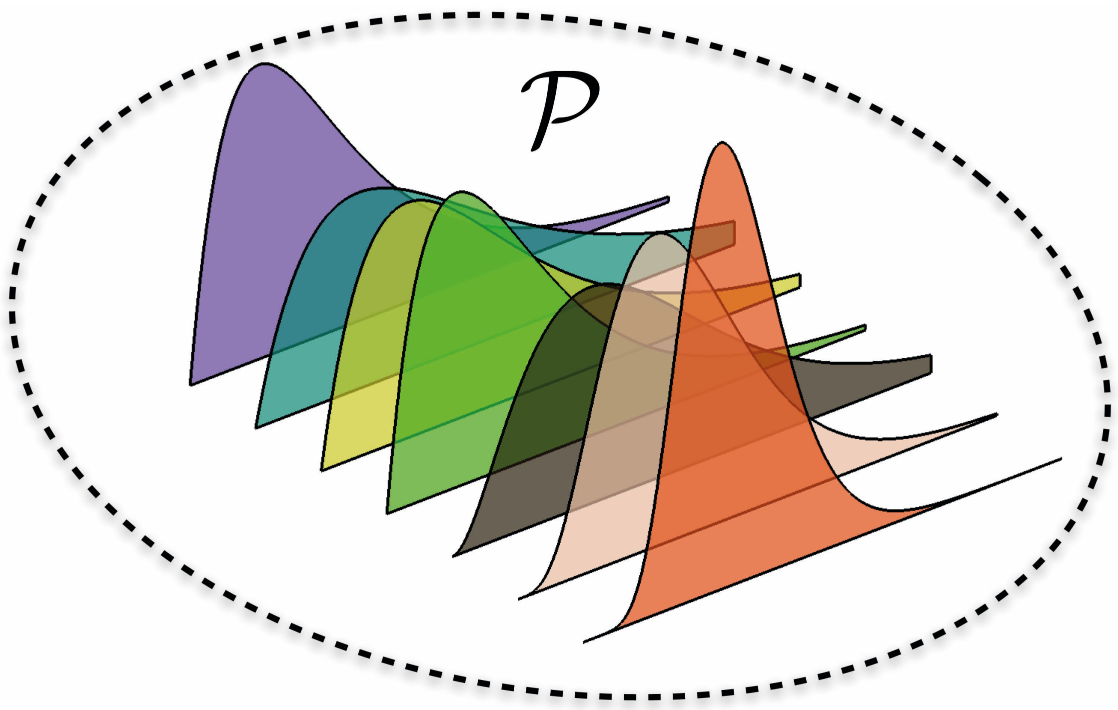
CHALMERS
UNIVERSITY OF TECHNOLOGY



Chance constraint optimization

Different approximation of the chance constraints and a case study on fuel cell buses.

Master's thesis in Systems, Control and Mechatronics

STEN ELLING TINGSTAD JACOBSEN

Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis 2019

Chance constraint optimization

Different approximation of the chance constraints
and a case study on fuel cell buses.

Sten Elling Tingstad Jacobsen

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2019

Chance constraint optimization
Approximation of the chance constraint
and a case study on fuel cell buses.
Sten Elling Tingstad Jacobsen

Supervisor: Nikolce Murgovski, Department of Electrical Engineering
Examiner: Balázs Adam Kulcsár, Department of Electrical Engineering

Master's Thesis 2019
Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Written permission to use front figure from Bart Van Parys, https://web.mit.edu/vanparys/www/topics/dro/.

Gothenburg, Sweden 2019

Chance constraint optimization
Different approximation of the chance constraint
and a case study on fuel cell buses.
Sten Elling Tingstad Jacobsen
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Data-driven optimization methods have become increasingly popular due to more powerful and affordable computers and sensors. Therefore a shift in optimization from more traditional methods to data-driven methods has occurred. Traditional methods often assume a perfect model of the system, which is rarely known. These methods are therefore more sensitive to perturbation and uncertainty.

In this thesis, the use of chance constraint optimization to solve problems with uncertainty is investigated. Chance constraint optimization gives guarantees on the probability of a random variable being above a certain value. Different relaxations of chance constraint optimization are derived. First, the exact mean and variance of a random variable are assumed known. Theory from Chebyshev inequality and Gauss inequality for univariate and multivariate random variables are used to approximate probability bounds. Sum of squares optimization as well as the scenario approach are also used to approximate probability bounds.

Secondly, the random variable is a data set and the mean and variance have to be approximated. Theory from distributional robust optimization is used to give an upper bound on the difference between the true and approximated mean and variance.

These methods are then used to minimize the fuel consumption for a fuel cell bus where the mass and speed are stochastic quantities.
This thesis project have been inspired by the master thesis 'Data-Driven Chance Constrained Optimization' by Bartolomeo Stellato, [1].

Keywords: Chance constraint optimization, Chebyshevs inequality, Gauss inequality, fuel cell bus, ambiguity set, unimodality, scenario approach, sum of square optimization.

Acknowledgements

I would like to express my deep gratitude to Balázs Kulcsár and Nikolce Murgovski, my examiner and supervisor, for their patient guidance, enthusiastic encouragement and useful critiques of this thesis work.

Sten Elling Tingstad Jacobsen, Gothenburg, June 2019

Contents

List of Figures xi

List of Tables xiii

1 Introduction 1
- 1.1 Aim . . . 2
- 1.2 Research questions . . . 3
- 1.3 Ethical and sustainability aspects . . . 3

2 Theory 5
- 2.1 Chebyshevs inequality for multivariate random variable. . . . 5
- 2.2 Gauss inequality for multivariate random variable. . . . 6
- 2.3 Conservatism of the bound approximations . . . 7
- 2.4 α-Unimodal bound . . . 9
  - 2.4.1 Choquet representation . . . 9
  - 2.4.2 α-Unimodality . . . 10
  - 2.4.3 α-Unimodal bound SDP theorem . . . 11
- 2.5 Minimum Volume Ellipsoid Approximation . . . 11
  - 2.5.1 Moment Uncertainty . . . 11
  - 2.5.2 Distributional Robust Optimization . . . 12
- 2.6 Multivariate Chebyshev Inequality with Estimated Mean and Variance 13
- 2.7 Scenario approach . . . 14
- 2.8 Sum of Squares optimization . . . 15
  - 2.8.1 Finding probability bound using SOS . . . 15

3 Applications 17
- 3.1 Fuel cell powertrain . . . 17
  - 3.1.1 Longitudinal vehicle dynamics . . . 18
  - 3.1.2 Fuel consumption and optimization formulation . . . 18
- 3.2 Optimization formulation . . . 19
- 3.3 Data from 20 runs on a specific bus route . . . 19
- 3.4 Results . . . 22
  - 3.4.1 Open loop optimization . . . 23
  - 3.4.2 Closed loop optimization . . . 27
  - 3.4.3 Discussion . . . 30

4 Conclusion 31

4.1 Future work 31

Bibliography **33**

List of Figures

2.1 Normal distribution of speed with mean 70 km/h and a standard deviation of 7 km/h. The blue area is the probability of the speed, v, being larger than 80 km/h and smaller than 60 km/h, i.e. $P(|v-\mu| \geq 10km/h)$. The probability of being outside the blue area is 0.8468 for a normal distribution, 0.5100 for Chebyshevs inequality and 0.7823 using Gauss inequality. 8
2.2 Chebyshevs inequality, Gauss inequality and cumulative distribution function of a normal distribution for $P(|X - \mu| \geq k)$ for varying k. . . 9

3.1 Series powertrain configuration for fuel cell vehicle, taken from [2]. P_F and P_B are power from fuel cells and the energy storage, i.e. the battery. P_{dem} is the power that is demanded by the electric motor. . . 17
3.2 The road altitude for the bus route. 20
3.3 The speed data for the different bus driving test. 20
3.4 The mass data for the different bus driving test. 21
3.5 The mean of the power demand, P_{dem}. 22
3.6 The standard deviation of the power demand, P_{dem}. 22
3.7 The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = n_\xi = 1$. The optimal is to use as much of the battery power as possible. 23
3.8 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 24
3.9 The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = \infty$. The optimal is to use as much of the battery power as possible. 24
3.10 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 25
3.11 The power from the battery and fuel cell delivered to the electric motor over distance when using SA. The optimal is to use as much of the battery power as possible. 25
3.12 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 26
3.13 The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = n_\xi = 1$ and exact mean and variance. The optimal is to use as much of the battery power as possible. 26

3.14 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 27
3.15 The power from the battery and fuel cell delivered to the electric motor over distance when using MCEMV. The optimal is to use as much of the battery power as possible. 28
3.16 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 28
3.17 The power from the battery and fuel cell delivered to the electric motor over distance when using SA. The optimal is to use as much of the battery power as possible. 29
3.18 The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value. 29

List of Tables

3.1 The fuel consumption in kJ for the different methods in open loop. The exact method is when the mean and variance are considered to be exactly known. Clearly the scenario approach gives the lowest fuel consumption. 27
3.2 The fuel consumption for the different methods in closed-loop. 29

1

Introduction

Recent developments in convex optimization, such as semidefinite programming [3], as well as faster and cheaper computers, have made it possible to solve large optimization problems in finite time. Computational heavy methods, such as model predictive control, can now be used on a system with fast dynamics. This in combination with the possibility to collect and store a large amount of data has led to a shift in optimization from more traditional methods to data-driven methods. Traditional methods assume a perfectly known model which is rarely available. This causes the optimal solution to be sensitive to perturbations in the model, [4]. A traditional method is convex optimization which is described in e.g., [5], and can be formulated as

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad i = 1, ..., m \\ & h_i(x) = 0 \quad i = 1, ..., p \end{aligned} \tag{1.1}$$

where $f(x)$, $g_i(x)$ are convex functions, $h_i(x)$ is an affine convex function and $x \in \mathbb{R}^n$ is the decision variable. $f(x)$ is called the objective function and the goal is to find an optimal value of x that minimizes the objective function, [5],

Different methods have been developed to be able to solve convex optimization problems with uncertainty. Robust optimization is such a method and can be formulated as:

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x, \xi) \leq 0 \quad \forall \xi \in \Xi \end{aligned} \tag{1.2}$$

where ξ is a random variable. Robust optimization considers all possible cases which may lead to a conservative solution. This conservatism is often due to a few extreme cases. If these extreme cases are discarded the solution often become less conservative. Chance constraint optimization is an optimization method that does not consider the worst cases. It only considers the uncertainty constraint to be valid for a percentage of the random variable, ξ. This is achieved by giving a probability guarantee to the constraint, $\mathbb{P}$. The probability of the constraint, $h(x, \xi) \leq 0$, has to be larger than some number, $1 - \epsilon$, where ϵ is close to 0. Which means that the constraint has a large probability of being fulfilled. The chance constraint linear program can be formulated as,

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c^\intercal x \\ \text{subject to} \quad & \mathbb{P}(h(x, \xi) \leq 0) \geq 1 - \epsilon \end{aligned} \tag{1.3}$$

In 1959 Charnes and Cooper first introduced chance constraint optimization (CCO) in [6]. A. Nemirovski and A. Shapiro discuss in their paper,[7], that CCO programs might be computationally intractable and L. G. Khachiyan shows that computing the probability of uniformly distributed variables is an NP-hard problem. CCO problems of the form (1.3) hence need to be approximated. Distributional robust optimization is a method to reformulate the chance constraint in a tractable way. By using information that is known about the distribution $\mathbb{P}$, one would like to minimize the set of possible distributions that $\mathbb{P}$ could be. This set is called the ambiguity set $\mathscr{P}$ and is used in distributional robust optimization,

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c^{\mathsf{T}} x \\ \text{subject to} \quad & \mathbb{P}(h(x,\delta) \leq 0) \geq 1-\epsilon, \forall \mathbb{P} \in \mathscr{P} \end{aligned} \tag{1.4}$$

If the first two moments of the distribution are known there exist methods to approximated the probability distribution. Chebyshev inequality is the most famous and was presented in [8]:

Theorem 1 (Chebyshev inequality[8]) *Let $X \in \mathbb{R}$ be a random variable with finite mean μ and finite non-zero variance σ^2. Then for any real number $k > 0$,*

$$\mathbb{P}(|X-\mu| > k\sigma) \leq \begin{cases} \frac{1}{k^2} & k > 1 \\ 1 & \textit{otherwise} \end{cases} \tag{1.5}$$

The Chebyshev inequality guarantees that for different probability distributions no more than $\frac{1}{k^2}$ of the values can be more than k standard deviations away from the mean, which gives a pessimistic bound. Chebyshev inequality considers all possible distribution in the ambiguity set.
Other probability inequalities are Cantelli inequality,[9], and Gauss inequality, [10]. Gauss inequality is similar to Chebyshev inequality but only unimodal distributions are considered in the ambiguity set:

Theorem 2 (Gauss inequality[10]) *Let $X \in \mathbb{R}$ be a unimodal distributed random variable with finite mean μ and finite non-zero variance σ^2. Then for any real number $k > 0$,*

$$\mathbb{P}(|X-\mu| > k\sigma) \leq \begin{cases} \frac{4}{9k^2} & k > \frac{2}{3} \\ 1-\frac{k}{\sqrt{3}} & \text{otherwise} \end{cases} \tag{1.6}$$

The Gauss inequality is hence less conservative compared to Chebyshev inequality, the bound is $\frac{4}{9}$ times less conservative.

1.1 Aim

The aim of this project was to learn about different relaxations of the probability bound and then test these relaxations on a practical dynamical case. Therefore an extensive literature study has been done which is presented in the theory section 2.

1.2 Research questions

- Theoretical, what are the strengths and weaknesses of different types of relaxation of the probability bounds?
- How are the different methods ranked in terms of conservatism?
- What are the strengths and weaknesses of the different chance constraint optimization linear programs for optimization of fuel economy of a hydrogen fuel cell vehicle?

1.3 Ethical and sustainability aspects

A motivation for this project from a sustainable and ethical point of view is that it will contribute to a future with more environmentally friendly cars. Cars stand for 12 % of EU's emission of carbon dioxide (CO_2) and hence are the EU restrictions on the emission of CO_2 getting stricter for every year. For 2021 a maximum limit for (CO_2) emission will be 95 g/km, which is a 40 % reduction compared to 2007, [11]. This contributes to a shift in the car industry from combustion engines to fuel cell and electric cars. This project will contribute to achieving United nations goal 13 on climate actions.

2

Theory

The theory chapter describes how the Chebyshev and Gauss inequality can be reformulated to also be valid for multivariate random variables. This is done by formulating semidefinite programs which find the optimal upper and lower bounds for a probability distribution.
The number of distributions that are included in the ambiguity set are important with respect to how conservative the bounds are. A method called α-unimodal bound is presented in which the variable α regulates what distributions are included in the ambiguity set.
Theory for two other ways of solving chance constraint programs are also presented. The first method is called the scenario approach and gives probability guarantee based on the number of samples of a random variable. The second is called sum of squares optimization and find an optimal bound based on the first two moments. Inspiration for this chapter have been taken from the master thesis 'Data-Driven Chance Constrained Optimization' by Bartolomeo Stellato, [1].

2.1 Chebyshevs inequality for multivariate random variable.

L. Vandenberghe, S. Boyd and K. Comanor have developed a method [12] to find the Chebyshev bound for multivariate random variables. Given an event set described by quadratic constraints, an exact mean and variance, the Chebyshev bound can be calculated by solving a semidefinite program. The results have been reformulated in [13] which results are presented here. The event set for this case is described by a polyhedron,

$$\Xi = \left\{ \xi \in \mathbb{R}^n : a_i^\top \xi < b_i \forall i = 1, ..., k \right\}, \tag{2.1}$$

where $a_i \in \mathbb{R}^n, a_i \neq 0$, $b_i \in \mathbb{R}$ and ξ is the random variable. The ambiguity set, $\mathscr{P}(\mu, S)$, is defined as,

$$\mathscr{P}(\mu, S) = \left\{ \mathbb{P} \in \mathscr{P}_\infty : \int_{\mathbb{R}^n} \xi \mathbb{P}(d\xi) = \mu, \int_{\mathbb{R}^n} \xi \xi^\top \mathbb{P}(d\xi) = S, \right\}, \tag{2.2}$$

where $\mu \in \mathbb{R}^n$, $S \in \mathbb{S}^n$ and the ambiguity set, $\mathscr{P}_\infty$, is the set of all distributions on $\mathbb{R}^n$. The following theorem describes the SDP to find the generalized Chebyshev bound,
Theorem 3 (Generalized Chebyshev bounds [12]) *If Ξ is a polytope of the form (3), the worst-case probability problem (P) with ambiguity set $\mathscr{P}(\mu, S)$, where μ and S are the mean respectively the covariance, is equivalent to a tractable SDP:*

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathscr{P}(\mu,S)} \quad & \mathbb{P}(\xi \notin \Xi) = \max \sum_{i=1}^{k} \lambda_i \\
\text{subject to} \quad & z_i \in \mathbb{R}^n, Z_i \in \mathbb{S}^n, \lambda_i \in \mathbb{R} && \forall i = 1, ..., k \\
& {a_i}^\top z_i \geq b_i \lambda_i && \forall i = 1, ..., k \\
& \sum_{i=1}^{k} \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} S & \mu \\ \mu^\top & 1 \end{bmatrix} \\
& \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \succeq 0 && \forall i = 1, ..., k.
\end{aligned}
\tag{2.3}
$$

Because of the exact SDP formulation the optimization problem, (2.3), is solvable in finite time using interior point methods,[14]. The SDP can be integrated into distributional robust programs and is therefore very useful. However, the Chebyshev bound is very conservative and therefore it might not be optimal for practical problems as Vandenberghe write in the paper [12]: 'In practical applications, the worst-case distribution will often be unrealistic, and the corresponding bound overly conservative. Improved bounds can be computed by further restricting the allowable distributions.' Therefore the number of possible distributions have to be smaller. Gauss inequality only considers unimodal distributions, which most common distributions are, such as normal distributions. The ambiguity set is smaller compared to the Chebyshev bound and therefore the bounds are less pessimistic.

2.2 Gauss inequality for multivariate random variable.

The ambiguity set for Gauss inequality only includes unimodal distribution which is defined as [15]:
Definition 1 (Univariate unimodality [15]) *A univariate distribution $\mathbb{P}$ is called unimodal with mode 0 if the mapping $t \mapsto \mathbb{P}(\xi \leq t)$ is convex for $t < 0$ and concave for $t > 0$.*
For multivarite distributions, unimodality is describe using star-shaped set [13]:
Definition 2 (Star-shaped sets [13]) *A set $B \subseteq \mathbb{R}^n$ is said to be star-shaped with center 0 if for every $\xi \in B$ the line segment $[0, \xi]$ is a subset of B*
Definition 3 (Star-unimodality [13]) *A distribution $\mathbb{P} \in \mathscr{P}_\infty$ is called star-unimodal with mode 0 if it belongs to the weak closure of the convex hull of all uniform distributions on star-shaped sets with center 0. The set of all star-unimodal distribution with mode 0 is denoted as $\mathscr{P}_*$*
In the case when the ambiguity set has a continuous probability density function $g(\xi)$ it can be proved that the probability set, $\mathbb{P} \in \mathscr{P}_\infty$, is star-unimodal iff $g(\xi)$ is non-increasing in t$\in (0, \infty)$ $\forall \xi \neq 0$. Therefore the distribution $\mathbb{P}$ is star-unimodal if $g(\xi)$ is non-increasing along any ray emanating from the orgin [15]. Definition 3 is also valid for distributions that do not have a density function.
Bart P.G. Van Prays et al. developed an SDP, given an star-unimodal ambiguity set and the two first moments of a distribution, that finds the maximal Gauss bound for multivariate random variables.

Theorem 4 (Generalized Gauss bounds [13]) *If Ξ is a polytope of the form in definition 3 with $0 \in \Xi$, the worst-case probability problem (P) with ambiguity set $\mathscr{P}_*(\mu, S)$ is equivalent to a tractable SDP:*

$$
\begin{aligned}
\sup_{\mathbb{P} \in \mathscr{P}_*(\mu,S)} \quad & \mathbb{P}(\xi \notin \Xi) = \max \sum_{i=1}^{k} \lambda_i - \tau_i \\
\text{subject to} \quad & z_i \in \mathbb{R}^n, Z_i \in \mathbb{S}^n, \lambda_i \in \mathbb{R}, \tau_i \in \mathbb{R}^{l+1} && \forall i = 1, ..., k \\
& {a_i}^\top z_i \geq b_i \lambda_i && \forall i = 1, ..., k \\
& \tau_i (a_i^\top z_i)^n \geq \lambda_i^{n+1} b_i^n && \forall i = 1, ..., k \\
& \sum_{i=1}^{k} \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \leq \begin{bmatrix} \frac{n+2}{n} S & \frac{n+1}{n} \mu \\ \frac{n+1}{n} \mu^\top & 1 \end{bmatrix} \\
& \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \geq 0 && \forall i = 1, ..., k.
\end{aligned}
\tag{2.4}
$$

where n is the dimension of the distribution and $l = \lceil log_2 n \rceil$. For proof of the theorem see [13].

The optimization formulation is a semidefinite program similar to theorem 3 and can easily be applied to an optimization problem. This theorem is less conservative compared to theorem 3 and can therefore be applied to practical examples with better results.

2.3 Conservatism of the bound approximations

To check the conservatism of Chebyshev inequality and Gauss inequality they can be compared to a normal distribution. Consider a normal distribution over the speed for a vehicle with mean 70 km/h and a standard deviation of 7 km/h, see Figure 2.1. The probability of the speed being between 60 km/h and 80 km/h is 0.8468 for a normal distribution, 0.5100 for Chebyshev inequality and 0.7823 using Gauss inequality.

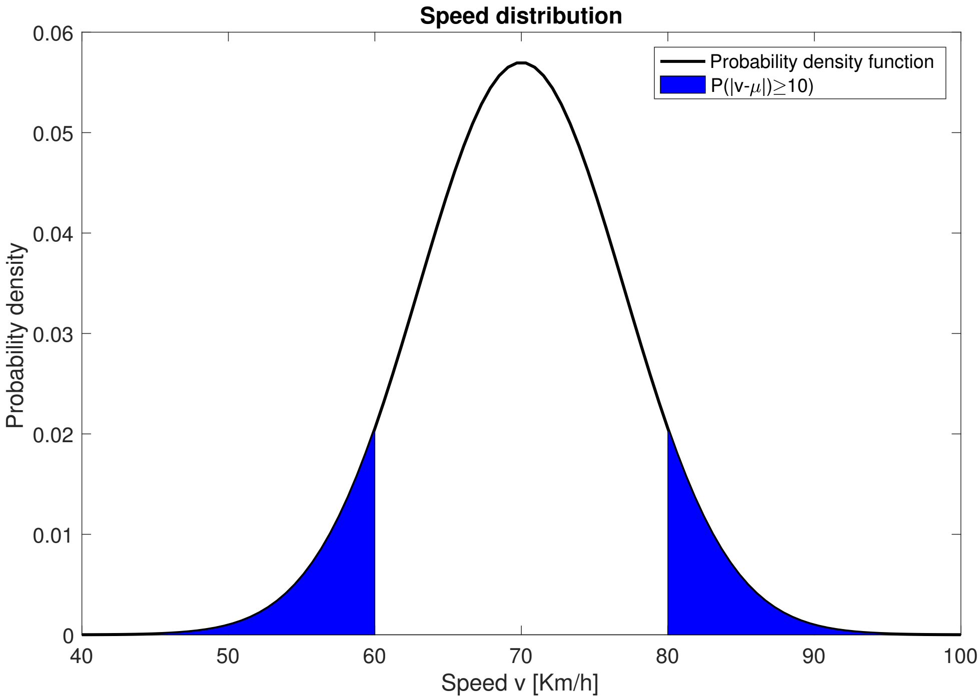


Figure 2.1: Normal distribution of speed with mean 70 km/h and a standard deviation of 7 km/h. The blue area is the probability of the speed, v, being larger than 80 km/h and smaller than 60 km/h, i.e. $P(|v-\mu| \geq 10km/h)$. The probability of being outside the blue area is 0.8468 for a normal distribution, 0.5100 for Chebyshevs inequality and 0.7823 using Gauss inequality.

The probability bound for Chebyshev inequality 2 and Gauss inequality 2 presented in the introduction are dependent on k. In Figure 2.2 these probability bounds are compared with the cumulative probability distribution for a normal distribution.

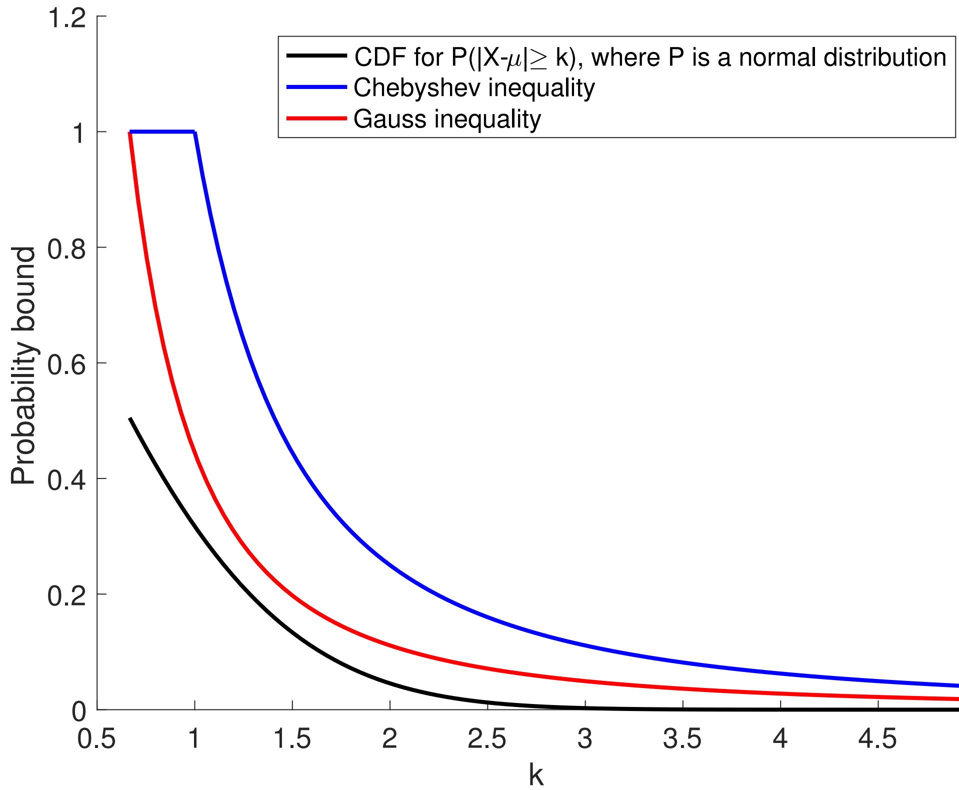


Figure 2.2: Chebyshevs inequality, Gauss inequality and cumulative distribution function of a normal distribution for $P(|X - \mu| \geq k)$ for varying k.

Gauss inequality is less conservative then the Chebyshev bound. To be precise Gauss inequality is $\frac{4}{9}$ times less conservative. The only difference between the two methods is the number of distributions that are considered in the ambiguity set. It seems like fewer distributions in the ambiguity set gives less conservative probability bounds. To be able to find a connection between Chebyshev and Gauss-inequality Van Parys et al. introduced α-Unimodal bound, which gives a variable, α, that tunes the number of distributions that are considered in the ambiguity set. This theorem is presented in Theorem 6 below.

2.4 α-Unimodal bound

The α-Unimodal bound was presented in the same paper as the Gauss inequality for multivariable random variables, [13]. The α-unimodality gives more freedom to specifying unimodality of distributions that are included in the ambiguity set. Before the α-unimodality is presented some definitions have to be explained.

2.4.1 Choquet representation

Choquet theory is way of representing a distributions within a set of distribution. This distribution is represented by extreme distributions,

Definition 4 (Extreme distributions [13]) *A distribution $\mathbb{P} \in \mathscr{P}_\infty$ is said to be an extreme point of an ambiguity set $\mathscr{P} \subseteq \mathscr{P}_\infty$ if it is not representable as a strict convex combination of two distinct distributions in $\mathscr{P}$. The set of all extreme points of $\mathscr{P}$ is denoted as ex$\mathscr{P}$.*

Definition 5 (Choquet representation) *A weakly closed and convex ambiguity set $\mathscr{P} \subseteq \mathscr{P}_\infty$ is said to admit a Choquet representation if for every $\mathscr{P} \subseteq \mathscr{P}_\infty$ there exists a Borel probability measure $\mathbb{m}$ on ex$\mathscr{P}$ with*

$$\mathbb{P}(\cdot) = \int_{\mathrm{ex}\mathscr{P}} e(\cdot)\mathbb{m}(\mathrm{d}e). \tag{2.5}$$

According to the Choquet representation a distribution $\mathbb{P} \in \mathscr{P}_\infty$ can be represented of the extreme points, $e(\cdot)$, of $\mathscr{P}_\infty$, where the extreme points are weighted. These weights are the so called mixtures, $\mathbb{m}$.

2.4.2 α-Unimodality

Definition 6 (α-Unimodality [15]) *For any fixed $\alpha > 0$, a distribution $\mathbb{P} \in \mathscr{P}_\infty$ is called α-unimodal with mode 0 if $t^\alpha(\mathbb{P}(B/t))$ is non-decreasing in $t \in (0, \infty)$ for every Borel set $B \in \mathscr{B}(\mathbb{R}^n)$. The set of all α-unimodal distributions with mode 0 is denoted as $\mathscr{P}_\alpha$*

Given this theorem a distribution $\mathbb{P} \in \mathscr{P}_\infty$ is α-unimodal iff

$$t^\alpha \int_{\mathbb{R}^n} g(\xi t)\mathbb{P}(d\xi) \tag{2.6}$$

is non-decreasing in $t \in (0, \infty)$. If the distributions are continuous the definition 6 means that the probability density function may increase along rays. However, the rate of how fast the pdf can increase is dependent on α. In fact the density function cannot increase faster than $\|\xi\|^{\alpha - n_\xi}$, where n_ξ is the dimension of the random variable. To be able to use this and formulate an SDP for α-unimodal bounds, radial α-unimodal distribution needs to be introduced.

Definition 7 (Radial α-unimodal distribution, [13]) *For any $\alpha > 0$ and $x \in \mathbb{R}^n$ we denote by $\delta^\alpha_{[0,x]}$ the radial distribution supported on the line segment $[0, x] \subset \mathbb{R}^n$ with property that*

$$\delta^\alpha_{[0,x]}([0, tx]) = t^\alpha \in [0, 1].$$

A radial distribution is the distribution between two points in space. If $\alpha = 1$ then the radial distribution is constant. Now definition 5 can be used for radial α-unimodal distribution,

Theorem 5 ([13], theorem 3) *For every $\mathbb{P} \in \mathscr{P}_\alpha$ there exists a unique mixture distribution $\mathbb{m} \in \mathscr{P}_\infty$ with*

$$\mathbb{P}(\cdot) = \int_{\mathbb{R}^n} \delta^\alpha_{[0,x]}(\cdot)\mathbb{m}(dx) \tag{2.7}$$

A distribution can hence be formulated as a mixture of radial distribution. What is important is that the α radial distributions are extremal in $\mathscr{P}_\alpha$. The following lemma is important for the α-unimodal bound,

Lemma 1 (Lemma 1 in [13]) *For any $\alpha > 0$ and $x \in \mathbb{R}^n$, the mean value and the second-order moment matrix of the radial distribution $\delta^\alpha_{[0,x]}$ are given by $\frac{\alpha}{\alpha+1}x$ and $\frac{\alpha}{\alpha+2}xx^\top$, respectively*

This results,i.e. lemma 1, is then used in combination with theorem 2.4 and gives the following theorem:

2.4.3 α-Unimodal bound SDP theorem

Theorem 6 (α-Unimodal bound [13]) *If Ξ is a polytope of the form (3) with 0 $\in \Xi$, the worst-case probability problem (P) with ambiguity set $\mathscr{P}_\alpha(\mu, S)$ is equivalent to:*

$$
\begin{array}{lll}
\sup\limits_{\mathbb{P}\in\mathscr{P}_\alpha(\mu,S)} & \mathbb{P}(\xi \notin \Xi) = \max \sum\limits_{i=1}^{k} \lambda_i - \tau_i & \\
\text{subject to} & z_i \in \mathbb{R}^n, Z_i \in \mathbb{S}^n, \lambda_i \in \mathbb{R}, \tau_i \in \mathbb{R}^{l+1} & \forall i = 1, ..., k \\
& {a_i}^\top z_i \geq 0, \tau_i \geq 0 & \forall i = 1, ..., k \\
& \tau_i(a_i^\top z_i)^\alpha \geq \lambda_i^{\alpha+1} b_i^\alpha & \forall i = 1, ..., k \\
& \sum\limits_{i=1}^{k} \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{bmatrix} & \\
& \begin{bmatrix} Z_i & z_i \\ {z_i}^\top & \lambda_i \end{bmatrix} \succeq 0 & \forall i = 1, ..., k.
\end{array}
\tag{2.8}
$$

Now there is a variable α that determines what distributions are included in the ambiguity set. When $\alpha \to \infty$ theorem 2.8 is the same as the Chebyshev inequality and $\alpha = n_\xi$ theorem 2.8 is the same as Gauss inequality.
Unfortunately, there is no method on how to find an optimal α but it seems like $\alpha = n_\xi$, i.e. Gauss inequality, is optimal.

2.5 Minimum Volume Ellipsoid Approximation

The minimum volume ellipsoid approximation (MVEA) was formulated in [1] and is a reformulation of 2.8 to a chance constrained linear program where the mean and variance are estimated. First, the theorem for bounds on the moment uncertainty is formulated and then the linear program is described.

2.5.1 Moment Uncertainty

In most practical problems the mean and the variance is estimated from data:

$$
\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \xi^i, \hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^{N} (\xi^i - \hat{\mu})(\xi^i - \hat{\mu})^\top \tag{2.9}
$$

where N is the number of data, $\hat{\mu}$ is the estimated mean and $\hat{\Sigma}$ is the estimated covariance. If these estimated means and covariance would be used for the previous describe theorems, (2.3) (2.4) (2.8), the probability bounds would not be correct. Van Parys et al. discuss this problem in the paper [13]. The authors write that the first and second-order moments can be assumed to belong to an SDP-representable confidence set $\mathscr{M}$, [13],

$$\begin{bmatrix} S & \mu \\ \mu^\top & 1 \end{bmatrix} \in \mathscr{M} \subseteq \mathbb{S}_+^{n+1} \tag{2.10}$$

This condition can then be added to theorem 6 and the ambiguity would handle uncertainty in both the mean and the covariance matrix. This results will however not be used in this thesis and instead we use another result that bound the norm of the difference between the true mean and the estimated mean respectively the true covariance and the estimated covariance [16],

Corollary 1 ([16]) *Let S be an N sample generated independently at random according to a distribution $\mathbb{P}$. Then with probability at least $1-\beta$ over the choice of S, we have:*

$$\|\hat{\mu} - \mu\|_2 \leq \frac{R}{\sqrt{N}}\Big(2 + \sqrt{2\ln\frac{1}{\beta}}\Big) = r_1 \tag{2.11}$$

Corollary 2 ([16]) *Let S be an N sample generated independently at random according to a distribution $\mathbb{P}$. Then with probability at least $1-\beta$ over the choice of S, we have:*

$$\|\hat{\Sigma} - \Sigma\|_{\mathrm{F}} \leq \frac{2R^2}{\sqrt{N}}\Big(2 + \sqrt{2\ln\frac{2}{\beta}}\Big) = r_2 \tag{2.12}$$

where R is the radius of the ball in feature space containing the support set of the distribution where the data is centered around 0 and that,

$$N > \Big(2 + \sqrt{2\ln\frac{4}{\beta}}\Big)^2$$

2.5.2 Distributional Robust Optimization

Distributional robust optimization is a way of reformulating chance constraint optimization. It considers all possible distributions in an ambiguity set with a given mean and covariance,

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & \mathbb{P}(g(x,\xi) \leq 0) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathscr{P}_\alpha(\mu, \Sigma) \end{aligned} \tag{2.13}$$

The ambiguity set here is the one described in (2.4). Assume that the uncertainty is linearly in a(ξ) and b(ξ) as,

$$a(\xi) = a_0 + \sum_{j=1}^{n_\xi} a_j \xi_j = a_0 + \hat{A}\xi \tag{2.14}$$

$$b(\xi) = b_0 + \sum_{j=1}^{n_\xi} b_j \xi_j = b_0 + \hat{B}\xi$$

From this the probability constraint in equation 2.13 can be formulated as,

$$b(x) - a(x)^\top \mu \geq \rho \|\Sigma^{1/2} a(x)\|_2 \tag{2.15}$$

where $\rho = (\frac{2}{\alpha+2})^{1/\alpha}(\frac{1}{\epsilon})^{1/2}$. If we use robust theory and Corollary 1 and 2 we get

$$\mu = \hat{\mu} + d, d \in \mathbb{R}^{n_\xi} : \|d\|_2 \leq r_1 \tag{2.16}$$

and

$$\Sigma = \hat{\Sigma} + \Delta, \Delta \in \mathbb{R}^{n_\xi \times n_\xi} : \|\Delta\|_F \leq r_2 \tag{2.17}$$

From this the distributional robust program can be written as,

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & b(x) - a(x)^\top \hat{\mu} - r_1 \|a_i(x)\|_2 \geq \rho \|(\hat{\Sigma} + r_2 I)^{1/2} a(x)\|_2, \quad i = 1, ..., n_g \end{aligned} \tag{2.18}$$

2.6 Multivariate Chebyshev Inequality with Estimated Mean and Variance

In the paper 'Multivariate Chebyshev Inequality with Estimated Mean and Variance', [17], a method to find the Chebyshev bound for approximated mean and variance, as in equation 2.9, is presented,

Theorem 7 ([17]) *Let* $\boldsymbol{\xi} \in \mathbb{E}^{n_\xi}$ *be a random variable and let* $N \in \mathbb{Z}_{\geq n_\xi}$. *Given N+1 i.i.d. samples* $\boldsymbol{\xi}^{(1)}, ..., \boldsymbol{\xi}^{(N)}, \boldsymbol{\xi}^{(N+1)} \in \mathbb{R}^{n_\xi}$, *if we assume that* $\boldsymbol{\Sigma_N}$ *is nonsigular, then for all* $\lambda \in \mathbb{R}_{>0}$ *it holds that:*

$$\mathbb{P}^{N+1}\left((\boldsymbol{\xi}^{N+1} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma_N^{-1}} (\xi^{N+1} - \boldsymbol{\mu}) \geq \lambda^2\right) \leq \min\left(1, \frac{n_\xi(N^2 - 1 + N\lambda^2)}{N^2\lambda^2}\right) \tag{2.19}$$

In the master thesis, 'Data-driven chance constraint optimization' [1], the author describes the uncertainty by formulating a ellipsoid Φ that is dependent on the first N samples,

$$\mathbb{P}^{N+1}\left(\xi^{N+1} \in \Phi\left(\boldsymbol{\xi}^{(1)}, ..., \boldsymbol{\xi}^{(N)}\right)\right) \geq 1 - \epsilon. \tag{2.20}$$

The ellipsoid is then defined using equation 2.19 as

$$\Phi\left(\boldsymbol{\xi}^{(1)}, ..., \boldsymbol{\xi}^{(N)}\right) = (\xi - \hat{\mu})^\top \frac{\hat{\Sigma}^{-1}}{\lambda^2} (\xi - \hat{\mu}) \leq 1 \tag{2.21}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the estimated mean and covariance from a dataset of N samples and λ is a changeable variable that determines the probability in 2.20. Now we would like λ to be dependent on ϵ and hence we use equation 2.19,

$$\frac{n_\xi(N^2 - 1 + N\lambda^2)}{N^2\lambda^2} \leq \epsilon \tag{2.22}$$

which gives the following equation for λ

$$\lambda = \sqrt{\frac{n_\xi(N^2 - 1)}{N(\epsilon N - n_\xi)}}. \tag{2.23}$$

This equation is valid as long as $N > \frac{n_\xi}{\epsilon}$ and the following SOCP can be formulated:

$$\begin{aligned} & \underset{x \in X}{\text{minimize}} && f(x) \\ & \text{subject to} && b_i(x) - a(x)^\top \hat{\mu} \geq \lambda \|\hat{\Sigma}^{(1/2)} a_i(x)\|_2 \end{aligned} \tag{2.24}$$

2.7 Scenario approach

The scenario approach is a different way of reformulating chance constraints. It was first introduced in [18] and has been further developed in [19]. The scanario approach gives probability guarantees based on a random program with N i.i.d. samples $\xi^1, ..., \xi^N \in \Xi \subseteq \mathbb{R}^{n_\epsilon}$,

$$\begin{aligned} & \text{RP}_N : \underset{x \in X}{\text{minimize}} && f(x) \\ & \text{subject to} && g(x, \xi^{(i)}) \leq 0 \quad \forall i = 1, ..., N. \end{aligned} \tag{2.25}$$

Before the theorem can be formulated some assumptions and definitions have to be presented first. For the optimization formulation in 2.25, the set X has to be convex and the function g has to be convex in x. The random program, 2.25, has to admit a unique solution with probability one. Now the definitions for violation probability, support constraint and Helly's dimension will be presented.

Definition 8 (Violation probability[19]) *The violation of an element $x \in \mathbb{R}^n$ is the probability that there exists an element $\delta \in \Delta$ for which the constraints are not satisfied.*

$$V(x) = \mathbb{P}(\xi \in \mathbb{R}^{n_\xi} : g(x, \xi) > 0) \tag{2.26}$$

Definition 9 (Support constraint [19]) *A constraint $\xi^{(r)}$ with $r \in \{1, ..., N\}$ is a support constraint for RP_N if its removal changes the solution of RP_N*

Definition 10 (Helly's dimension [19]) *Helly's dimension of RP_N is the least integer θ such that ess* $\sup \|Sc(RP_N)\|_{\xi \in \Xi} \leq \theta$ *holds for any finite N≥ 1*

The theorem is,

Theorem 8 ([19], Theorem 3.1) *Consider problem 2.25, then either the problem is unfeasible or it is feasible and the following holds for its optimal solution x^**

$$\mathbb{P}^{N+1}\left(\xi^1, ..., \xi^N \in \Xi : \mathbb{P}(V(x_N^*)) > \epsilon\right) \leq \sum_{j=1}^{\theta-1} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j} \tag{2.27}$$

This means that given a number of i.i.d. samples, there is a guarantee on the lower probability bound. Equation 2.27 can be reformulated,

$$N \geq \frac{e}{(e-1)\epsilon}\left(\theta - 1 + \ln \frac{1}{\beta}\right) \tag{2.28}$$

$$\theta = n_g(n_\xi + 1)$$

So the number of samples depends on ϵ^{-1} and on the $\ln \beta^{-1}$, where β is the confidence level. So we can choose a high confidence level and not increasing N too much, but if we choose ϵ we would need many samples. For example if $\epsilon = 0.01$ and $\beta = 0.01$ then we would need 886 samples.

2.8 Sum of Squares optimization

A sum of squares program is a linear program with constraints that are sum of squares, (SOS), polynomials. A multivariate polynomial, $p(x_1, ..., x_n)$, of degree 2n is a sum of square polynomial if it can be expressed as sum of polynomials, $f_1(x), f_2(x), ..., f_i(x)$ with degree n as:

$$p(x) = \sum_{i=1}^{m} f_i(x)^2 \tag{2.29}$$

This implies that every sum of square polynomial is positive polynomial. The polynomial 2.29 can also be expressed with the positive semidefinite matrix Q as,

$$p(x) = Z^\top(x) Q Z(x) \tag{2.30}$$

where $Z(x)$ is a vector of monomials. This method is also referred to as the Gram matrix method. SOS programs for finding the polynomials $f_i(x)$ are based on the Positivstellensatz. The constraint in the SOS program will be:

$$a_0(x) + \sum_{i=1}^{N} a_i(x) f_i(x) = 0 \tag{2.31}$$

where a_i are constants. The optimization can be formulated as:

$$\begin{aligned} &\text{Find polynomials} && f_i(x) \\ &\text{subject to} && a_0(x) + \sum_{i=1}^{N} a_i(x) f_i(x) = 0 \\ & && \sum_{i=1}^{m} f_i(x)^2 > 0 \end{aligned} \tag{2.32}$$

The optimization 2.33 can also be formulated to minimize a linear cost $c^\top w$ where w is a weight vector and c the unknown coefficients of the polynomials $f_i(x)$.

2.8.1 Finding probability bound using SOS

In a paper by Bertsimas and Popescu, [20], a SOS program for finding the bound of a univariate random program was described. The optimization program is formulated as:

$$\begin{aligned} \text{minimize} \quad & am_0 + bm_1 + cm_2 \\ \text{subject to} \quad & a + bx + cx^2 \geq 0 \ \forall x \in \mathbb{R} \\ & a + bx + cx^2 \geq 1 \ \forall x \notin \text{Event set} \end{aligned} \tag{2.33}$$

where $m_0 = 1, m_1 = \mu$ and $m_2 = \mu^2 + \sigma^2$. This is only valid for univariate variables and will therefore not be used for optimization of the fuel cell bus.

3

Applications

In this chapter different relaxations of the chance constraint are tested on a case. The case consider minimizing the fuel consumption of an electric bus that is powered by a battery and a fuel cell. A fuel cell has better efficiency, lower noise level and lower emission, actually zero emission, compared to a combustion engine, [21] and [22]. If the fuel cell is paired with a battery and if regenerative braking is used, the efficiency may be improved even more.
In this case the speed and the mass of the bus are stochastically uncertain. The speed is for example varying due to traffic and the mass is varying depending on the number of passengers. The fuel consumption is minimized by controlling the use of the fuel cell and the battery.

3.1 Fuel cell powertrain

In this thesis a series configuration for the fuel cell vehicle will be used, see Fig. 3.1.

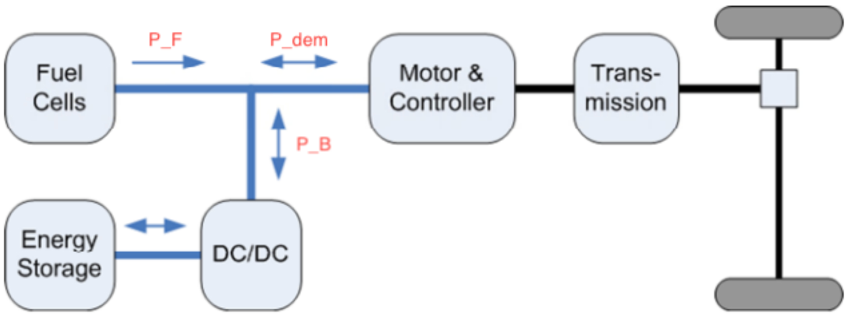


Figure 3.1: Series powertrain configuration for fuel cell vehicle, taken from [2]. P_F and P_B are power from fuel cells and the energy storage, i.e. the battery. P_{dem} is the power that is demanded by the electric motor.

The state equation for the battery is,

$$\begin{aligned} E_B(k+1) &= E_B(k) - P_B(k)\Delta t \\ \text{with } P_{dem} &= P_F + P_B \end{aligned}$$

where E_B is the energy of the battery, called state of charge (SOC), P_{dem} is the demanded power from the electric motor, P_B is the power output from the battery and P_F is the power from the fuel cell. If $P_B > 0$ the energy in the battery is decreasing and if $P_B < 0$, i.e. regeneration, the energy in the battery is increasing. Since the eigenvalue for the system is $\lambda = 1$ the system is marginally stable.

3.1.1 Longitudinal vehicle dynamics

The vehicle is modelled as a simple point mass moving in a hilly terrain. The velocity $v(t)$ and distance travelled $s(t)$ starting from the beginning of the horizon are

$$\dot{s}(t) = v(t) \tag{3.1}$$

$$m\dot{v}(t) = F_{\mathrm{w}}(t) + F_{\mathrm{brk}}(t) - F_{\alpha}(\alpha) - F_{\mathrm{air}}(v) \tag{3.2}$$

where F_{w} is wheel force delivered by the engine and $F_{\mathrm{brk}} \leq 0$ is the breaking force at the wheels. The forces $F_{\alpha}(\alpha)$ and $F_{\mathrm{air}}(v)$ are given as

$$F_{\alpha}(\alpha) = mg(\sin\alpha(\mathrm{s}) + c_{\mathrm{r}}\cos\alpha(\mathrm{s})) \tag{3.3}$$

$$F_{\mathrm{air}}(v) = \frac{\rho_{\mathrm{a}} c_{\mathrm{d}} A_{\mathrm{f}} v^2(t)}{2} \tag{3.4}$$

where m is vehicle mass, $g = 9.81m/s^2$ is gravitational acceleration, c_{r} is rolling friction coefficient between the road and wheels; ρ_{a}, c_{d} and A_{f} are air density, aerodynamic drag and vehicle's frontal area, respectively. The road slope $\alpha(s)$ is assumed to be given. Similar mathematical models are considered in [23]. The relationship between power and force is $P = Fv$ and therefore can the demanded power be formulated as:

$$P_{dem} = m\dot{v}v + \frac{\rho_{\mathrm{a}} c_{\mathrm{d}} A_{\mathrm{f}} v^3(t)}{2} + vmg(\sin\alpha(\mathrm{s}) + c_{\mathrm{r}}\cos\alpha(\mathrm{s})) \tag{3.5}$$

where the braking force is not included since full regeneration of braking is considered. Since the power demand is the only variable that is depended on the speed and the mass of the bus, it can be consider stochstically uncertain.

3.1.2 Fuel consumption and optimization formulation

The fuel consumption is described by the following equation,

$$F = a_0 + a_1 P_F + a_2 P_F^2$$

where F is the fuel consumption, a_0, a_1 and a_2 are constants. This formulation of the fuel consumption was also used in [24]. The optimization problem, i.e. to minimize the fuel consumption, can be formulated as:

$$\underset{P_F, E_B}{\text{minimize}} \int_0^{t_f} C_f(a_0 + a_1 P_F(t) + a_2 P_F(t)^2) - C_e E_B(t) dt \tag{3.6a}$$

subject to

$$E_B(0) = E_{B0}$$
$$\dot{E}_B = -P_B$$
$$P_{dem} = P_F + P_B$$
$$E_{Bmin} \leq E_B \leq E_{Bmax}$$
$$P_{Bmin} \leq P_B \leq P_{Bmax}$$

where C_f is the cost of hydrogen in currency/kg, C_B is the electric power cost in currency/kJ and t_f is the time when the route is completed. P_{dem} depends on the velocity of the vehicle which is non-deterministic and the road topography. A distribution for P_{dem} will be calculated from the bus line data and then different types of chance constraint programs will be used and compared to solve this minimization. For this case we simplify the fuel cost to a linear function,

$$F = a_1 P_F$$

where $a_1 = 17.75$. Since a_0 is close to zero, because of low fuel consumption when the vehicle is idling, this is ignored and will not affect the optimization.

3.2 Optimization formulation

The optimization formulation where P_B is decision variables and the integral, (3.6a), have been discretized using first order Euler discretization, can be formulated as:

$$\begin{aligned} \underset{P_B}{\text{minimize}} \quad & C_e \mathbf{P_B}^\intercal \mathbf{\Delta t} - C_e E_{B0} - a_1 C_f \mathbf{P_B}^\intercal \mathbf{\Delta t} \\ \text{subject to} \quad & \mathbf{E_B} = -M \cdot \mathbf{P_B} \cdot \mathbf{\Delta t} + E_{B0} \\ & E_{Bmin} \leq \mathbf{E_B} \leq E_{Bmax} \\ & P_{Bmin} \leq \mathbf{P_B} \leq P_{Bmax} \\ & \mathbf{P_B} \geq \mathbf{P_{dem}} - P_{Fmax} \\ & \mathbf{P_B} \leq \max(\mathbf{P_{dem}}, P_{Bmin}) \end{aligned} \tag{3.7}$$

$$M = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

where E_{Bmin} and E_{Bmax} are the lower and upper bound for the SOC of the battery, P_{Bmin} and P_{Bmax} are the lower and upper bound of the output power from the battery, P_{Fmax} is the maximal power from the fuel cell and $\mathbf{\Delta t}$ is the time interval between each sample. If the data is sampled in space then $\mathbf{\Delta t} = v/\mathbf{\Delta s}$, where $\mathbf{\Delta s}$ is the sample distance in space and v the speed in that sample interval. The only variable that is stochastically uncertain is the P_{dem} and hence all constraints with P_{dem} becomes chance constraints.

3.3 Data from 20 runs on a specific bus route

The data used in this thesis is from bus line 17 in Gothenburg. The data was gathered at 1 Hz and for 20 passes, [25]. The road topography of the route can be seen in Figure 3.2, the data gathered of the velocity can be seen in Figure 3.3 and the data of the varying mass, due to varying amount of passengers, can be seen in Figure 3.4.

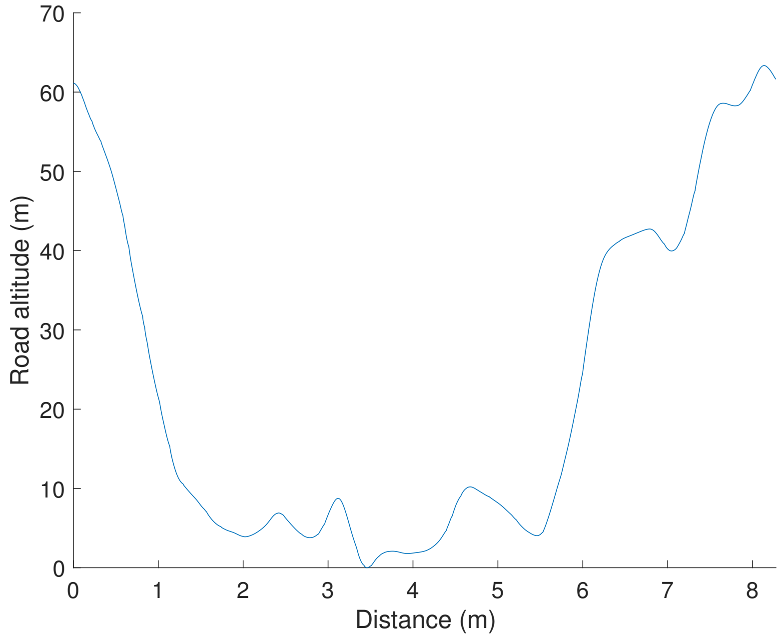


Figure 3.2: The road altitude for the bus route.

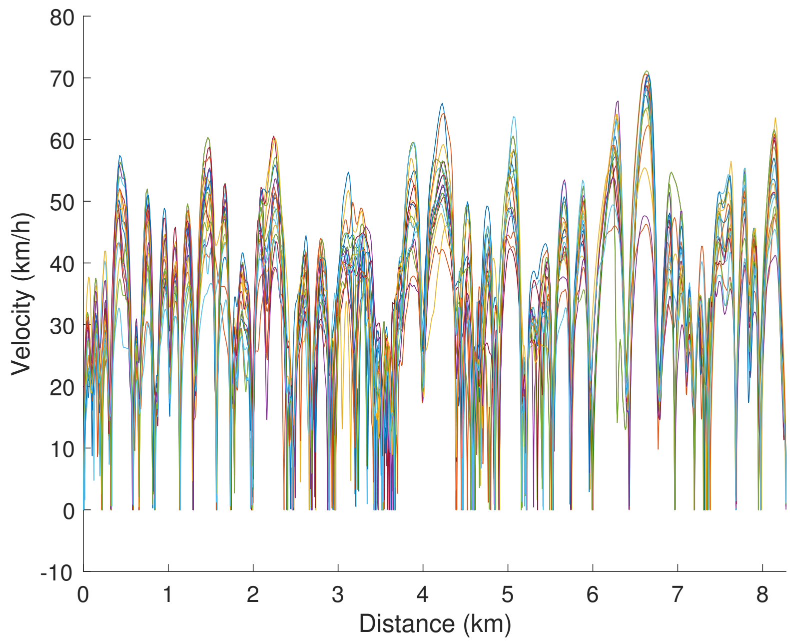


Figure 3.3: The speed data for the different bus driving test.

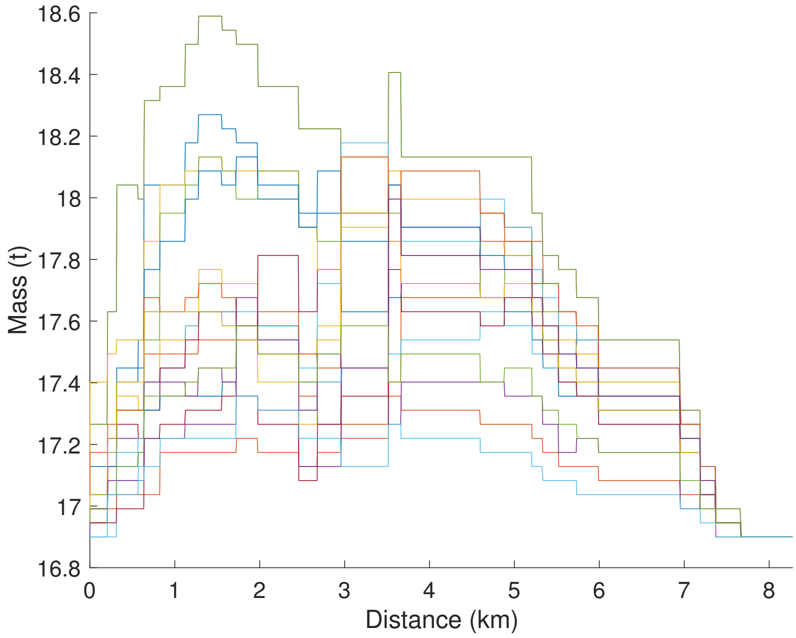


Figure 3.4: The mass data for the different bus driving test.

The data is gathered every second and is, therefore, time-dependent. However, the data is not comparable in time since the bus will be in different parts of the route at the same time. This is due to varying speed because of traffic and different bus drivers. Therefore the data is transformed to be distance dependent which can be done since the distance travelled was also measured. The transformation is calculated by first interpolating the data to get more data points and then all measurements for every 10 meter is gathered, i.e the sampling interval is 10 meters. The power demanded is calculated for each specific test. From these calculations, a mean and a variance for the power demand are calculated. These are then used to fit a normal distribution which i.i.d samples are drawn from.

The final data in this thesis is therefore artificial but is based on real data. There are two reason why this is done. The first is that the real data set is to small and gives a unrealistic high variance. The second reason is that the number of data points can be changed which gives extra freedom for the methods described in the theory section, 2. The mean and the standard deviation of the data can be seen in Figure 3.5 and Figure 3.6.

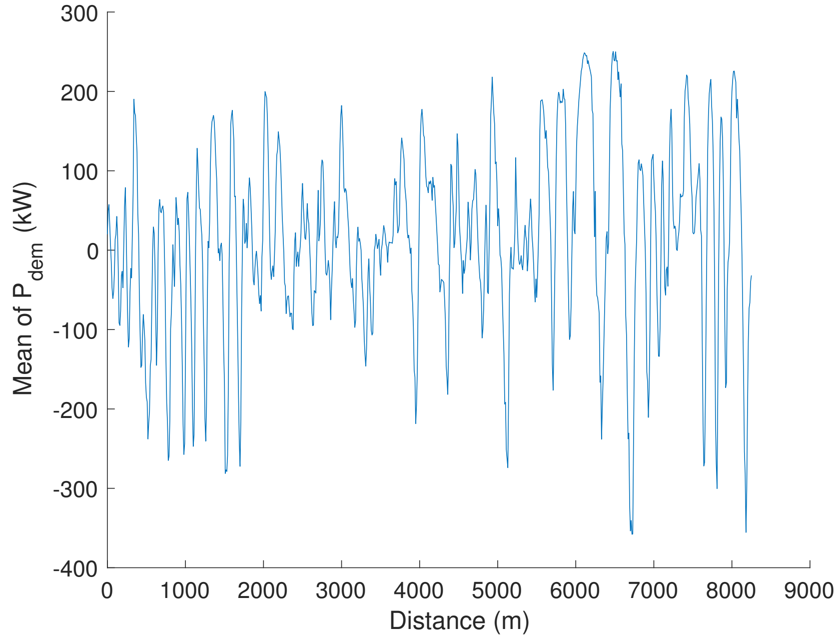


Figure 3.5: The mean of the power demand, P_{dem}.

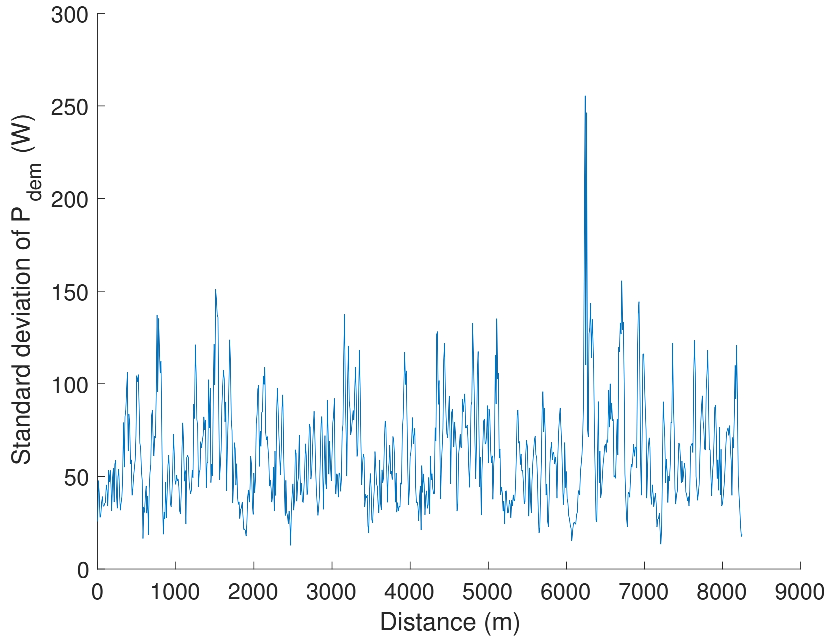


Figure 3.6: The standard deviation of the power demand, P_{dem}.

3.4 Results

In this section, the results for running the optimization methods from the theory chapter will be presented. First, the MVEA will be compared with the scenario approach for a control input with no feedback, i.e open-loop. Then the optimization is formulated as a model predictive control problem for the multivariate Chebyshev inequality with estimated mean and variance, MCEMV, and the scenario approach with a horizon of 100 samples, i.e. 1 km. The number of samples for each method is the same, $\epsilon = 0.1$ and the confidence level is $\beta = 0.01$. Therefore these results are the optimal solution for 90 % of the cases.

3.4.1 Open loop optimization

The results from the optimization are the power from the fuel cell and the battery, see Figures 3.13, 3.9 and 3.11, as well as the SOC of the battery, see figures 3.14, 3.10 and 3.12. The total power use from the fuel cell for the methods are shown in table 3.4.1. The scenario approach uses the least power from the fuel cell and thereby is the most effective method for open loop. The MVEA does not perform very well which might have to do with the fact that it depends on the dimension of the uncertainty which grows with the horizon.

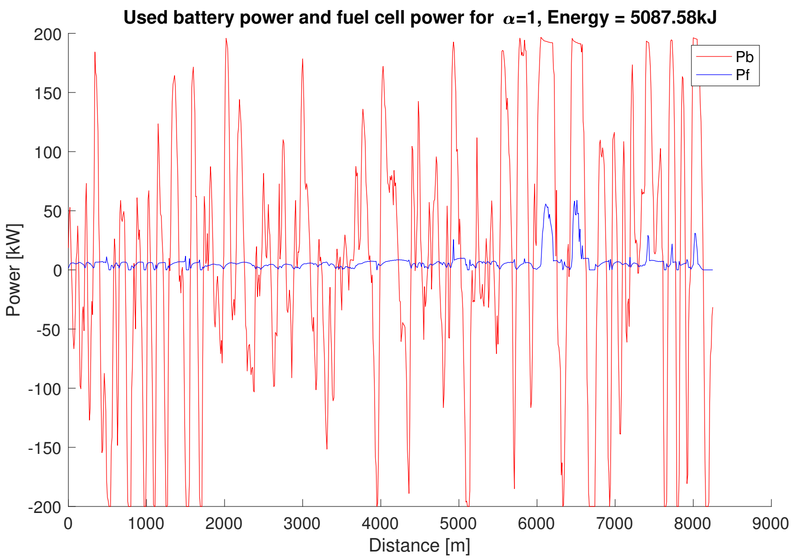


Figure 3.7: The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = n_\xi = 1$. The optimal is to use as much of the battery power as possible.

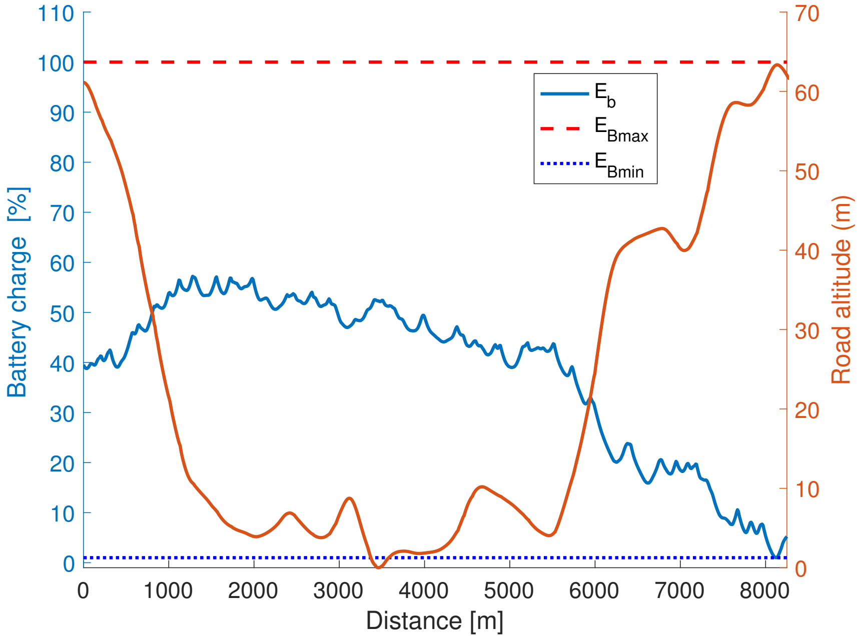


Figure 3.8: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

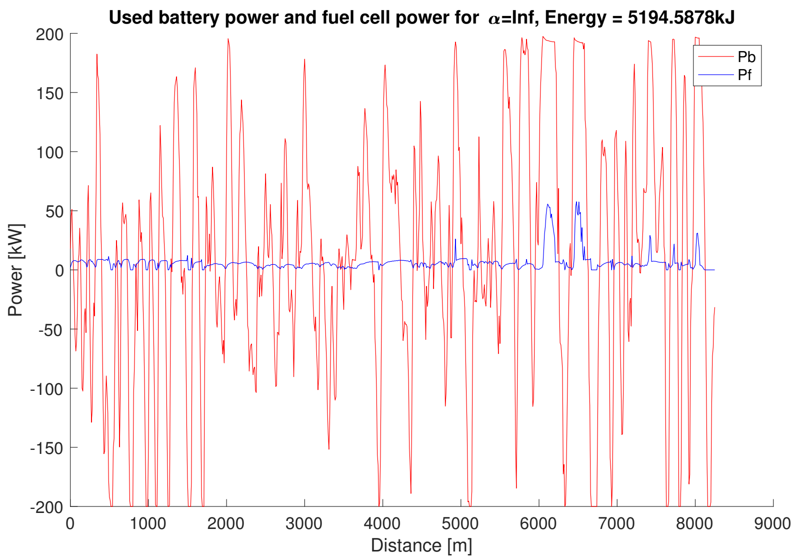


Figure 3.9: The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = \infty$. The optimal is to use as much of the battery power as possible.

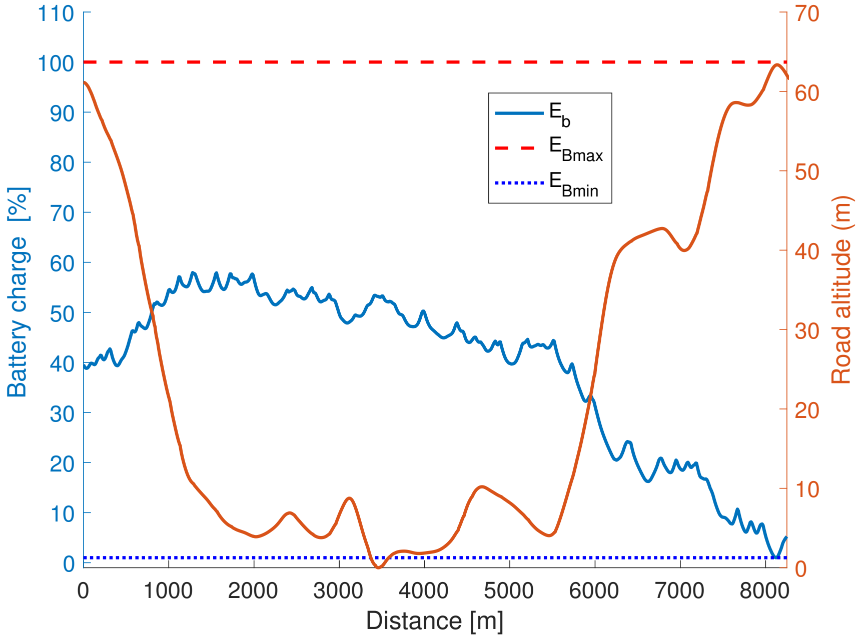


Figure 3.10: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

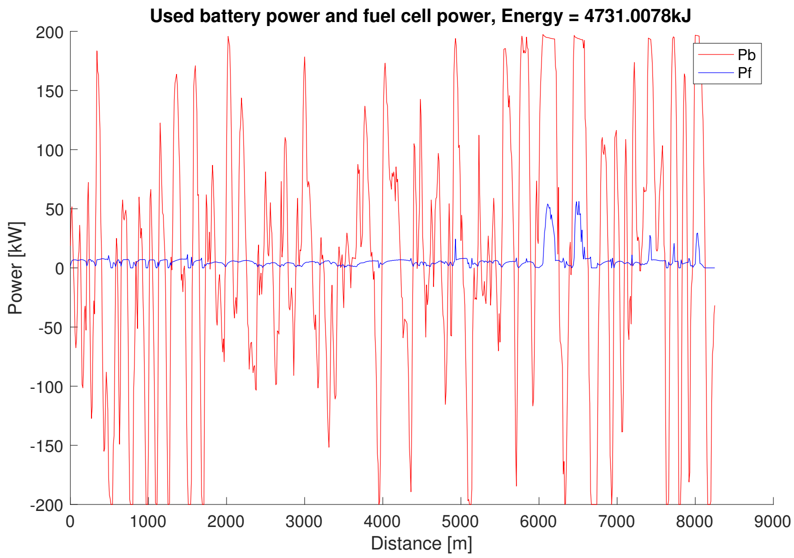


Figure 3.11: The power from the battery and fuel cell delivered to the electric motor over distance when using SA. The optimal is to use as much of the battery power as possible.

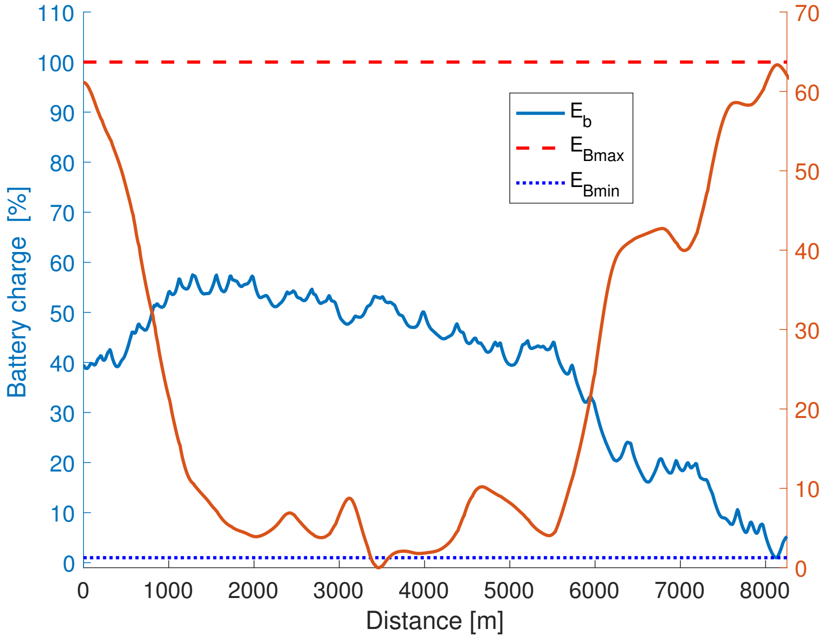


Figure 3.12: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

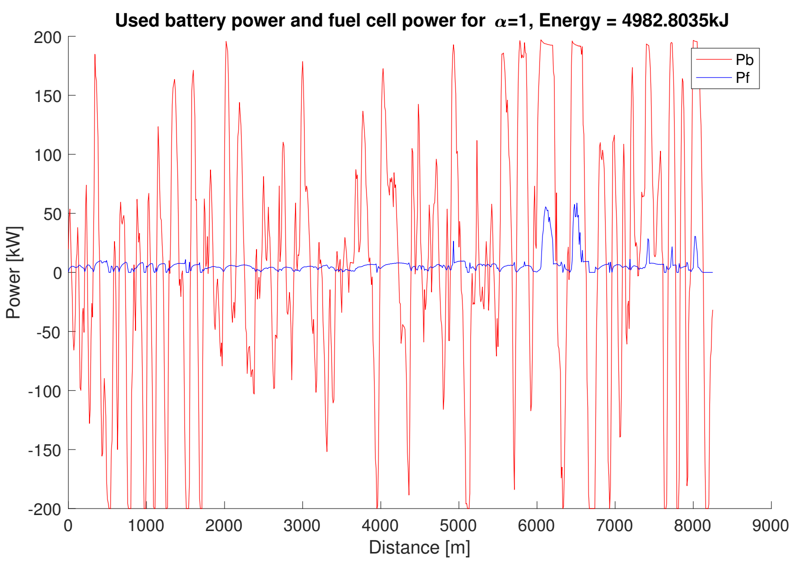


Figure 3.13: The power from the battery and fuel cell delivered to the electric motor over distance when using MVEA with $\alpha = n_\xi = 1$ and exact mean and variance. The optimal is to use as much of the battery power as possible.

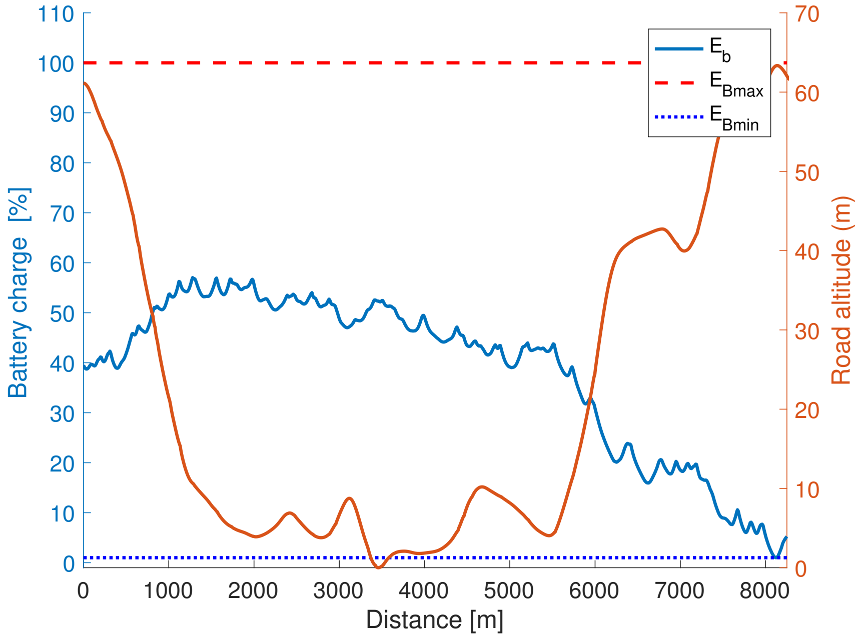


Figure 3.14: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

Method	MVEA, $\alpha = 1$	MVEA, $\alpha = \infty$	SA	Mean	Exact
Consumption [kJ]	5087	5194	4730	4500	4982

Table 3.1: The fuel consumption in kJ for the different methods in open loop. The exact method is when the mean and variance are considered to be exactly known. Clearly the scenario approach gives the lowest fuel consumption.

3.4.2 Closed loop optimization

Now the methods are compared in closed-loop. Since the closed loop contributes to a change in the dynamics of the system it might be hard to compare the results in this part. Also, what horizon is used might affect the solution differently for the MCEMV and the SA. The results for the closed-loop, therefore, have to be taken with a pinch of salt. The results from the optimization are the power from the fuel cell and the battery, see Figures 3.15 and 3.17, as well as the SOC of the battery, see figures 3.16 and 3.18. Due to the length of the horizon, too much power from the battery is used in the beginning and therefore more power from the fuel cell has to be used at the end of the driving cycle. This is less optimal, more fuel is used, then the solution for the open-loop optimization.

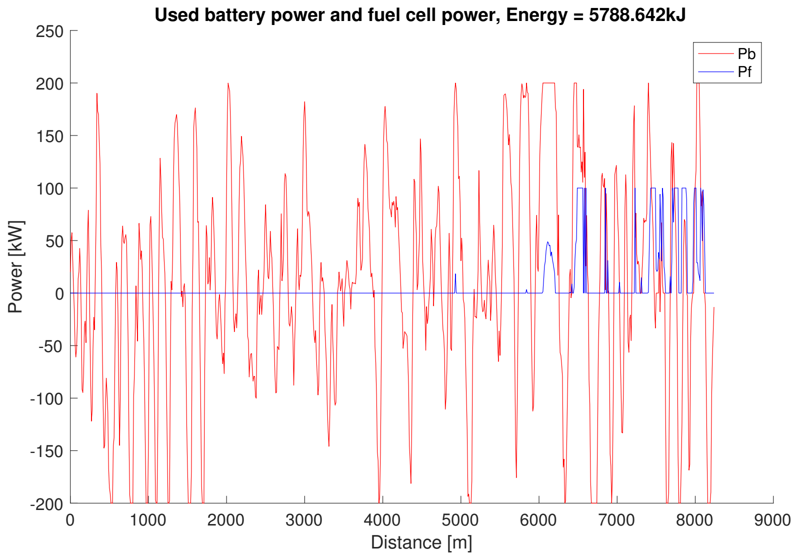


Figure 3.15: The power from the battery and fuel cell delivered to the electric motor over distance when using MCEMV. The optimal is to use as much of the battery power as possible.

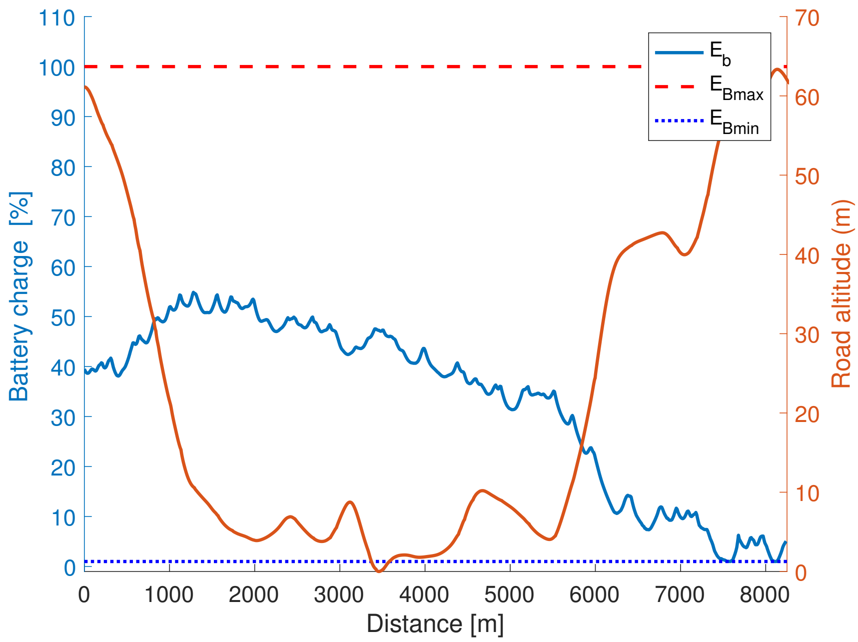


Figure 3.16: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

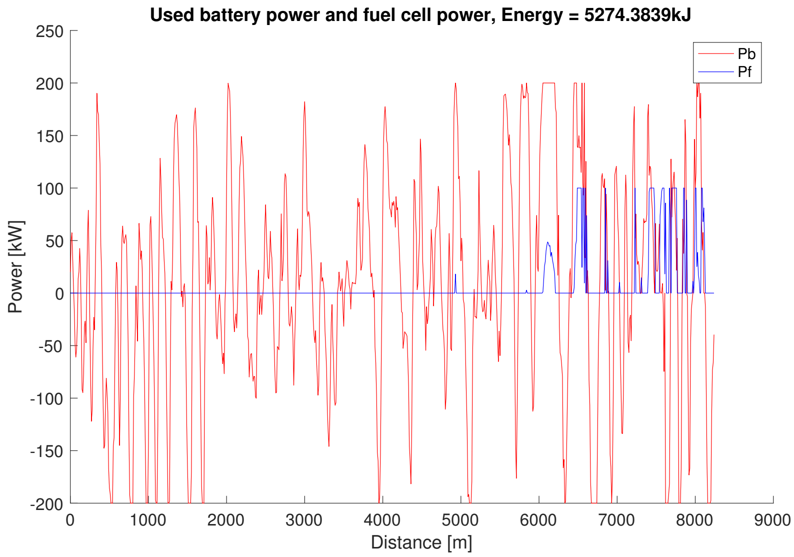


Figure 3.17: The power from the battery and fuel cell delivered to the electric motor over distance when using SA. The optimal is to use as much of the battery power as possible.

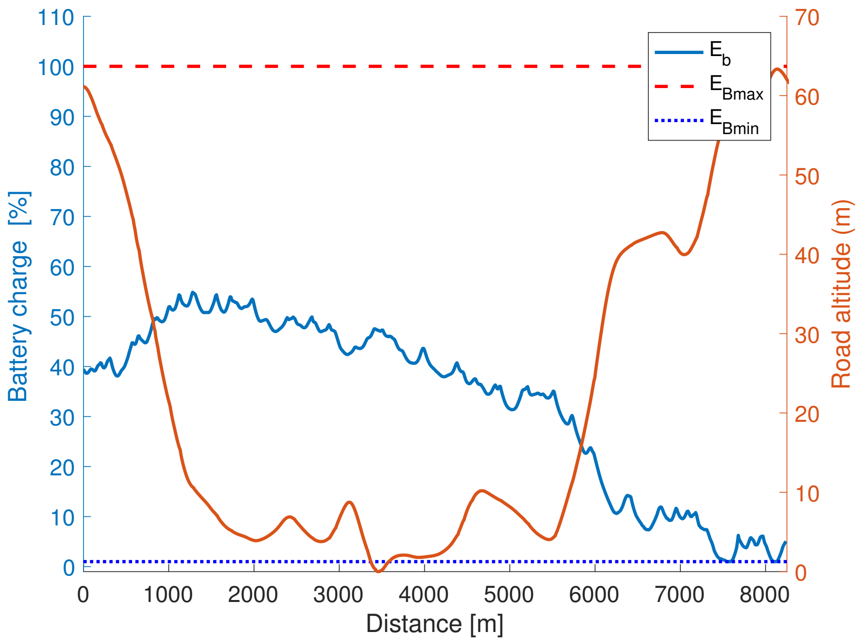


Figure 3.18: The state of charge of the battery. In the end of the driving cycle the battery reaches is lowest value.

Method	SA	MCEMV
Fuel consumption [kJ]	5274	5704

Table 3.2: The fuel consumption for the different methods in closed-loop.

3.4.3 Discussion

The scenario approach is the best method for both the open-loop and closed-loop cases. For the MVEA method $\alpha = 1$, i.e. Gauss inequality, gives better results compared to $\alpha = \infty$. This is expected since the Gauss inequality is less conservative compared to the Chebyshev inequality. The difference in consumption between the exact solution and MVEA is only 100-200 kJ which indicates that the estimation of the mean and variance are not too conservative. Also, the difference between a deterministic optimization, where the model is optimized with respect to the mean of P_{dem} is only 130-700 kJ better than the chance constraint methods. In general, the chance constraint methods are not too conservative. The computation times between the different methods are similar. One would expect that SA would be the worst, computationally-wise, because it has one constraint for each data point. But for this case only the maximal P_{dem} will be used as a constraint, i.e. the worst case, and the other constraints will be neglected.
It is also important to point out that for SA the number of data points determines what confidence is given to the chance constraint. When the number of data points are increasing the higher confidence and when $N \to \infty$ the scenario approach is equal to a robust optimization. However, for MVEA and MCEMV the confidence is fixed and the bounds become less conservative when the number of data points are increasing. Therefore would probably MVEA and MCEMV performs better compared to SA for a very high number of data points.

4
Conclusion

In this thesis different methods for finding the probability bound was presented. The methods are Chebyshev inequality, Gauss inequality, scenario approach and sum of squares optimization. Theorems for how these methods can be used for multivariate random variable as well as for estimated mean and variance was presented. These methods were then used for minimizing the fuel consumption for a bus with a fuel cell powertrain. The results show that the scenario approach gives the best results for both open-loop and closed-loop configuration. The difference in consumption between the exact solution and MVEA is only 100-200 kJ which indicates that the estimation of the mean and variance are not too conservative. Also, the difference between a deterministic optimization, where the model is optimized with respect to the mean of P_{dem} is only 130-700 kJ better than the chance constraint methods. In general, the chance constraint methods are not too conservative. The computation times between the different methods are similar. One would expect that SA would be the worst, computationally-wise, because it has one constraint for each data point. But for this case only the maximal P_{dem} will be used as a constraint, i.e. the worst case, and the other constraints will be neglected.

4.1 Future work

Future work can be done on other methods to find the probability bounds. Such as the sum of squares optimization and how to use that for multivariate random variables. If information of higher order moments can be included in these methods the bounds would probably be less conservative. There is no method that uses higher order moments and this would then be an important contribution.

There are no methods for finding an optimal α for the α-unimodal bound. From what can be concluded from this thesis it seems like $\alpha = n_\xi$ is optimal but there are no proofs for this.

It would also be interesting to apply these methods to a more complex case. What if the route is unknown for the vehicle, which is the case for most cars. Different types of configurations of the powertrain would be interesting to compare.

Bibliography

[1] Bartolomeo Stellato. Data-driven chance constrained optimization. *Automatic control laboratory, ETH*, September 2014.

[2] Hengbing Zhao and Andrew Burke. Effects of different powertrain configurations and control strategies on fuel economy of fuel cell vehicles. *The 25th World Battery, Hybrid and Fuel Cell Electric Vehicle Symposium Exhibition*, 2010.

[3] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[4] Lauren A. Hannah. Stochastic optimization. *Computational Statistics (Stat G6104),Columbia University*, 2014.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. 2004.

[6] A. Charnes and W. W. Cooper. Chance constraint optimization. *Management Science*, 6(1):73–79, 1959.

[7] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 2006.

[8] P. Chebyshev. Des valeurs moyennes. *Journal de Mathématiques pures et Appliquées*, 1867.

[9] F. P. Cantelli. Intorno ad un teorema fondamentale della teoria del rischio. *Bollettino dell'Associazione degli Attuari Italiani*, 1910.

[10] C. F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae, pars prior. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, 1821.

[11] European Commission. Regulation (eu) no 333/2014 of the european parliament and of the council. *Official Journal of the European Union*, 11 March 2014.

[12] S. Boyd L. Vandenberghe and K. Comanor. Generalized chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.

[13] Bart P. G. Van Parys, Paul J. Goulart, and Daniel Kuhn. Generalized gauss inequalities via semidefinite programming. *Mathematical Programming*, 156(1):271–302, Mar 2016.

[14] Yinyu Ye. *Interior Point Algorithms: Theory and Analysis*. John Wiley & Sons, Inc., New York, NY, USA, 1997.

[15] S. W. (Sudhakar Waman) Dharmadhikari and 1932 Joag-dev, Kumar. *Unimodality, convexity, and applications*. Boston : Academic Press, 1988. Includes indexes.

[16] John Shawe-Taylor and Nello Cristianini. Estimating the moments of a random vector with applications. Invited Talk, 2003.

[17] Bartolomeo Stellato, Bart P. G. Van Parys, and Paul J. Goulart. Multivariate chebyshev inequality with estimated mean and variance. *The American Statistician*, 71(2):123–127, 2017.

[18] Giuseppe Carlo Calafiore.

[19] Giuseppe Carlo Calafiore. Random convex programs. *SIAM Journal on Optimization*, 2010.

[20] Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15:780–804, 01 2005.

[21] Liangfei Xu, Li Jianqiu, Jianfeng Hua, Xiangjun Li, and Minggao Ouyang. Adaptive supervisory control strategy of a fuel cell/battery-powered city bus. *Journal of Power Sources - J POWER SOURCES*, 194:360–368, 10 2009.

[22] Ulrich Eberle, Bernd MÃŒller, and Rittmar von Helmolt. Fuel cell electric vehicles and hydrogen infrastructure: status 2012. *Energy Environ. Sci.*, 5:8780–8798, 2012.

[23] N. Murgovski, B. Egardt, and Magnus Nilsson. Cooperative energy management of automated vehicles. *Control Engineering Practice*, 57:84–98, 2016.

[24] Nikolce Murgovski, Xiaosong Hu, Lars Johannesson, and Bo Egardt. *Combined Design and Control Optimization of Hybrid Vehicles*, pages 1–14. American Cancer Society, 2015.

[25] Stefan Pettersson Lars Johannesson and Bo Egardt. Predictive energy management of a 4qt series-parallel hybrid electric bus. *Control Engineering Practice*, 2009.