

Comparing severe gas transport situations through the network

Similarity or reduction methods

K. Lindenberg

Master of Science Thesis



Comparing severe gas transport situations through the network

Similarity or reduction methods

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Applied Mathematics at Delft
University of Technology

K. Lindenberg
4087208

August 26, 2015

Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) · Delft
University of Technology

The work in this thesis was supported by GasUnie Transport Services - Groningen, The Netherlands. Their cooperation is hereby gratefully acknowledged.



Copyright © Delft Institute of Applied Mathematics
All rights reserved.

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING, MATHEMATICS
AND COMPUTER SCIENCE
DELFT INSTITUTE OF APPLIED MATHEMATICS

The undersigned hereby certify that they have read and recommend to the Faculty of
Electrical Engineering, Mathematics and Computer Science (EEMCS) for acceptance
a thesis entitled

COMPARING SEVERE GAS TRANSPORT SITUATIONS THROUGH THE NETWORK

by

K. LINDENBERG

in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE APPLIED MATHEMATICS

Dated: August 26, 2015

Supervisor(s):

Prof.dr.ir. C. Vuik

Drs. H. Dijkhuis

Dr. J.J. Steringa

Reader(s):

Dr. J.W. van der Woude

Abstract

A number of severe entry-exit combinations is selected, to investigate whether or not all possible future gas transport situations through the gas transport network of Gasunie Transport Services can be met. These severe entry-exit combinations are selected, such that if the transport capacity is manageable regarding the pipeline network for these entry-exit combinations, then it is possible to satisfy all possible entry-exit combinations. In short, these severe entry-exit combinations represent the most severe transport situations through the gas network within contractual limits, describing realistic market behavior, and determining the size and shape of the gas transport network. These are often called shipping variants or stress tests.

If a large set of stress tests is generated, then it takes a relatively large amount of time to evaluate these stress tests. The aim of this project is, therefore, to reduce this generated set as much as possible. When reducing this set, we have to make sure that we derive a minimal set, which is still a complete set. So, our goal is to delete identical stress tests from the generated set, but also less severe transport situations.

These stress tests are denoted by vectors which components represent the gas capacities at entry and exit points. In general, we deal with n -dimensional vectors, when we are considering a network of n entries and exits. We are always dealing with a balanced combination of entry and exit capacities, and therefore such vectors are called balanced vectors. This means that the components of such a vector add up to zero. So, when a stress test is represented by an n -dimensional vector x , then $\sum_i x_i = 0$ holds. The individual components of these vectors are correlated, because we want to take the mutual distance between the network points (following the pipelines) into account as well, besides the capacities on the entry and exit points. Thus, the vector space we are considering is not an Euclidean space.

This report focuses on the quadratic form distance, which can measure similarities between vectors for which the individual components are correlated. The quadratic form distance is a distance measure which uses a correlation matrix to determine the distance between two n -dimensional vectors. It is important that this correlation matrix is symmetric positive semidefinite, to ensure that the quadratic form distance is a well defined metric. A part of this thesis is dedicated to experiment with different choices of the correlation matrix, and to investigate their positive semidefiniteness. Finally, some generated sets of stress tests are reduced according to reducing criteria.

Table of Contents

Preface	xi
1 Introduction	1
1-1 Infrastructure	1
1-2 Gas transport network	2
1-3 Multi case approach	4
1-4 Outline of this thesis	5
2 Description of the stated problem	7
2-1 A small example	7
2-2 Algorithm to generate stress tests	10
2-2-1 The algorithm applied after clustering	13
2-3 Characterization of stress tests	14
2-3-1 Transporting gas through the network	15
2-3-2 Blending different types of gas	15
2-4 Research questions	17
3 Mathematical background	19
3-1 Quadratic form distance	19
3-1-1 Symmetric positive definiteness	20
3-1-2 A small example	21
3-1-3 A second example	23
3-1-4 A third example	24
3-1-5 Semi-norm	26
3-2 Positive semidefinite on a subspace	26
3-2-1 Subspace of balanced vectors	26
3-2-2 The distance matrix for the networks of GTS	33

4	Definitions correlation matrix	45
4-1	Relation to the pressure drop	45
4-1-1	Applying the new defined matrices on a small network	48
4-2	The quadratic form distance in MCA	49
4-3	Some analyzes	54
4-3-1	Deleting identical rows and columns	54
4-3-2	Rounding off lengths of pipelines	55
4-3-3	Reverse Cuthill-McKee ordering	56
4-3-4	Dependence on the diameter of the pipelines as well	57
4-4	Emphasizing the H-gas transport network	64
5	Reducing generated set of stress tests	69
5-1	Threshold value for the quadratic form distance	69
5-1-1	The distance between the capacity vectors	70
5-1-2	The angle between the capacity vectors	71
5-2	Conducted experiments	74
5-3	Analyzing the conducted experiments	78
5-4	Some more experiments	83
6	Conclusions and recommendations	85
6-1	Answers to the sub research questions	85
6-2	Future research	89
A	Figures	91
	Bibliography	117
	Glossary	119
	List of Acronyms	119
	List of Symbols	119

List of Figures

1-1	The gas transport grid of Gasunie.	3
2-1	An example of a simple network.	8
2-2	Stress test 1	8
2-3	Stress test 2	9
2-4	Stress test 3	9
2-5	An illustration of the definition of the mean transport distance.	10
2-6	An illustration of the reference point A and entry and exit centers N and X . . .	11
2-7	An example of a simple network, consisting of six points.	12
2-8	An example of a simple network divided into clusters (marked with yellow). . . .	14
2-9	The distances in one cluster of the simple network multiplied with 10.	14
2-10	The distances in the other cluster of the simple network multiplied with 10. . . .	14
3-1	An example of a simple network, consisting of three entry and exit points.	23
3-2	A small network to illustrate a needed property of the used distance.	24
3-3	A small network with capacities on the entry and exit points.	25
3-4	A network based on the distance matrix \mathbf{H}	32
3-5	The H-gas network of GTS.	34
3-6	A simplification of the network of GTS.	35
3-7	An expanded example network, in which G-gas is added.	36
3-8	An expanded example network, in which G-gas is added, with two blending stations. 38	
3-9	An expanded example network, in which G-gas is added, with two blending stations between entry/exit points.	40
3-10	An expanded example network, in which G-gas is added, with one blending stations connected to one entry/exit point.	41
3-11	A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with an indication of closed blending stations.	42

3-12	A simplification of the network of GTS, consisting of H-gas and G- and L-gas, in which some G-gas pipelines are deleted.	44
4-1	The relation between entry \mathbf{A}_{ij} and transportation distance d_{ij} with $d_{max} = 250$	52
4-2	The network of GTS.	53
4-3	A graph of the network points in the considered network.	61
4-4	A graph of the network points in the considered network, such that there exists a path between every pair of points.	62
4-5	A graph of the network points in the considered network, such that there exists a path between every pair of points.	63
4-6	A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with some of the names.	68
5-1	Some capacity vectors in \mathbb{R}^n , also called the capacity space.	69
5-2	Some capacity vectors in \mathbb{R}^n , also called the capacity space.	70
5-3	Two reduced sets, when the angle between two vectors is considered.	75
5-4	Two reduced sets, when the angle between two vectors is considered with variable inaccuracy δ_i	76
5-5	The reduced sets for the new diameters, when the angle between two vectors is considered with a 'constant' inaccuracy δ_i	77
5-6	The reduced sets for the new diameters, when the angle between two vectors is considered with variable inaccuracy δ_i	77
5-7	A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with some of the names and maximum flow through some pipelines.	82
A-1	The relation between entry \mathbf{A}_{ij} and transportation distance with $d_{max} = 250$	91
A-2	The relation between entry \mathbf{A}_{ij} and transportation distance with $d_{max} = 250$	92
A-3	The relation between entry \mathbf{A}_{ij} and transportation distance with $d_{max} = 250$	92
A-4	The relation between pressure drop and transportation distance for $Q = 1000 \text{ dam}^3/h$	93
A-5	The relation between pressure drop and transportation distance for $Q = 2000 \text{ dam}^3/h$	94
A-6	The relation between pressure drop and transportation distance for $Q = 3000 \text{ dam}^3/h$	95
A-7	The relation between pressure drop and transportation distance for $Q = 4000 \text{ dam}^3/h$	96
A-8	The relation between pressure drop and transportation distance for $Q = 5000 \text{ dam}^3/h$	97
A-9	The relation between pressure at the end of a pipe and transportation distance for $Q = 1000 \text{ dam}^3/h$	98
A-10	The relation between pressure at the end of a pipe and transportation distance for $Q = 2000 \text{ dam}^3/h$	99
A-11	The relation between pressure at the end of a pipe and transportation distance for $Q = 3000 \text{ dam}^3/h$	100

A-12	The relation between pressure at the end of a pipe and transportation distance for $Q = 4000 \text{ dam}^3/h$	101
A-13	The relation between pressure at the end of a pipe and transportation distance for $Q = 5000 \text{ dam}^3/h$	102
A-14	The Wobbe distribution of the first shipping variant of the H-gas network of GTS.	103
A-15	The Wobbe distribution of the second shipping variant of the H-gas network of GTS.	104
A-16	The Wobbe distribution of the third shipping variant of the H-gas network of GTS.	105
A-17	The Wobbe distribution of the fourth shipping variant of the H-gas network of GTS.	106
A-18	The Wobbe distribution of the fifth shipping variant of the H-gas network of GTS.	107
A-19	The Wobbe distribution of the sixth shipping variant of the H-gas network of GTS.	108
A-20	The Wobbe distribution of the seventh shipping variant of the H-gas network of GTS.	109
A-21	The Wobbe distribution of the eighth shipping variant of the H-gas network of GTS.	110
A-22	The visualization of the sparsity pattern of the matrix \mathbf{A} , applied on the H-gas network of GTS.	111
A-23	The visualization of the sparsity pattern of the matrix \mathbf{A} , applied on a simplified network of GTS.	112
A-24	The visualization of the sparsity pattern of the two new defined matrices, applied on a simplified network of GTS.	112
A-25	The contour plot of the original matrix and of the matrix on which the reverse Cuthill-McKee ordering is applied.	113
A-26	The contour plot of the levels between 0.8 and 1 of the original matrix and of the matrix on which the reverse Cuthill-McKee ordering is applied.	114
A-27	Another simplification of the network of GTS.	115

List of Tables

2-1	The distance table of entry and exit points with respect to point 1.	11
2-2	The distance table of entry and exit points with respect to point 5.	12
2-3	The distance table of entry and exit points with respect to point 4.	13
3-1	An overview of some choices for a_H and a_G and the resulting minimum and maximum eigenvalues of the distance matrix.	37
3-2	An overview of some choices for a and b and the resulting eigenvalues of the distance matrix.	39
4-1	The QFD, with different approaches, applied on some of the stress tests.	65
4-2	Some difference vectors of two different stress tests.	67
5-1	The eight mentioned difference vectors.	81
6-1	An overview of the findings of the QFD regarding the criteria.	86

Preface

The project I have worked on is carried out in collaboration with Gasunie Transport Services. The first three months were dedicated to a literature study and this report is a sequel of the literature study. This report is written to fulfill the requirements for the degree of Master of Science in Applied Mathematics at the Delft University of Technology.

Acknowledgements

I have worked on this project with great pleasure, due to the interesting and challenging assignments during this project, but due to my lovely colleagues as well! I would like to extend my sincere thanks to all of my colleagues here at Gasunie Transport Services. Especially, to my colleagues Patrick Tel, Jasper Stevens, Džanan Omanović, Jan Albert Laverman, Wilco Kelder and Jan de Boer.

In addition, I have experienced some nice activities during my internship, like some readings hold at Gasunie, a volleyball tournament with a few colleagues, a field trip to the site Wieringermeer, a visit the control command post of Gasunie in the basement, a drink with colleagues.

I would like to express my gratitude towards my supervisors Kees Vuik, Harry Dijkhuis and Jarig Steringa as well, and thank them for the opportunity to do my Master thesis internship at Gasunie Transport Services. I also want to thank them for their help, support, advice, feedback and guidance during my thesis project.

Furthermore, I want to thank Jacob van der Woude for his ideas and input during my project, because he helped me at a later stage of my internship with some of our challenging problems.

Finally, I would like to express my special gratitude and thanks to my family and my housemate Dorine Wijmenga for their encouragement, interest and support during my time in Groningen.

Groningen, July 2015

Kimberley Lindenberg

Chapter 1

Introduction

1-1 Infrastructure

Gasunie is one of the largest gas infrastructure companies in Europe which provides the transport of natural and green gas in the Netherlands and the Northern part of Germany. Gasunie has two subsidiaries which are responsible for the management, the operation and the development of the gas transmission grid, i.e.

- Gasunie Transport Services in the Netherlands for the national transmission grid. The Gasunie Transport Services is a wholly owned subsidiary of N.V. Nederlandse Gasunie, but their activities are performed independently as required by law;
- Gasunie Deutschland in Germany for a long-distance pipeline grid.

The core activity of GTS is the transport of gas in the Netherlands, and of Gasunie Deutschland in North Germany respectively.¹ The gas transport network of Gasunie, located in the Netherlands and in the Northern part of Germany, has a length of approximately 15,500 kilometers, conveys approximately 125 billion cubic meters of natural gas, and consists of multiple compressor stations, blending stations, metering and regulating stations, export stations, gas delivery stations, liquefied natural gas storages, air separation units, and an underground gas and nitrogen storage.

Firstly, compressor stations supply compression (pressure and capacity) and are necessary to keep the pressure constant when gas is being transported over long distances. As gas is transported through the pipeline, the pressure drops which leads to a reduction in capacity, and therefore, the pressure must be increased approximately every 100 kilometers to make it possible to transport greater volumes of gas at the correct pressure.

Secondly, blending stations ensure that the gas has the desired quality for Dutch households. For this reason, the quality of the transported gas is closely monitored. Natural gas can come

¹Information is taken from the Gasunie site <http://www.gasunie.nl/en/about-gasunie>.

from different sources and the composition of each source varies and, where necessary, these different gas types are blended such that the gas is suitable for domestic use. The gas quality is determined by the Wobbe index, calorific value and density of the gas.

The gas usually comes above ground again near to the place of destination, which often happens at a metering and regulating station. This is also where high pressure of the gas that was necessary for national transport, is brought down again. The familiar odor to the natural gas is added as well at a metering and regulating station, so that everyone can smell the unwanted presence of natural gas.

Export stations are the end points of this gas transport system at the borders and the function of such a station is, to measure the volume of natural gas supplied to customers in other countries.

In addition, gas delivery stations form the connection between this transport system and the transport system of a regional network operator or an industrial gas pipeline. Their main functions are to reduce the transport pressure (around 8 *bar*) and to measure the volume of natural gas supplied.

Finally, underground storages can be both entry and exit points and can be divided into three groups:

- season storages;
- short cycle storages;
- storages with a connection to other networks. [1]

A few of these storages are owned by Gasunie, like the one located at the *Maasvlakte* and *Zuidwending*. However, most of the storages are owned by other companies and these companies determine how much gas is injected into or removed from these storages.

The transport grid of Gasunie (the Netherlands and Northern part of Germany) is shown in Figure 1-1.²

In the Netherlands, the transmission networks of GTS are divided, according to pressure categories, into a high pressure grid and an intermediate pressure grid. The HTL network is also divided into two networks according to the type of gas flowing through the network, namely into a network transporting G-gas (the gray lines in the Netherlands in Figure 1-1) and a network transporting H-gas (the yellow lines in the Netherlands in Figure 1-1). The RTL network is pretty much used to transport G-gas.³

1-2 Gas transport network

GTS sells the available capacity in a reliable network with competitive conditions. This gas network consists of a number of entry and exit points. An entry point is a point where gas can be (physically) injected into the pipeline network of Gasunie and an exit point where

²Information and picture are taken from the Gasunie site <http://www.gasunie.nl/en/about-gasunie/infrastructure>.

³Information is taken from <http://www.gasunietransportservices.nl/en/transportinformation/the-transmission-network>.

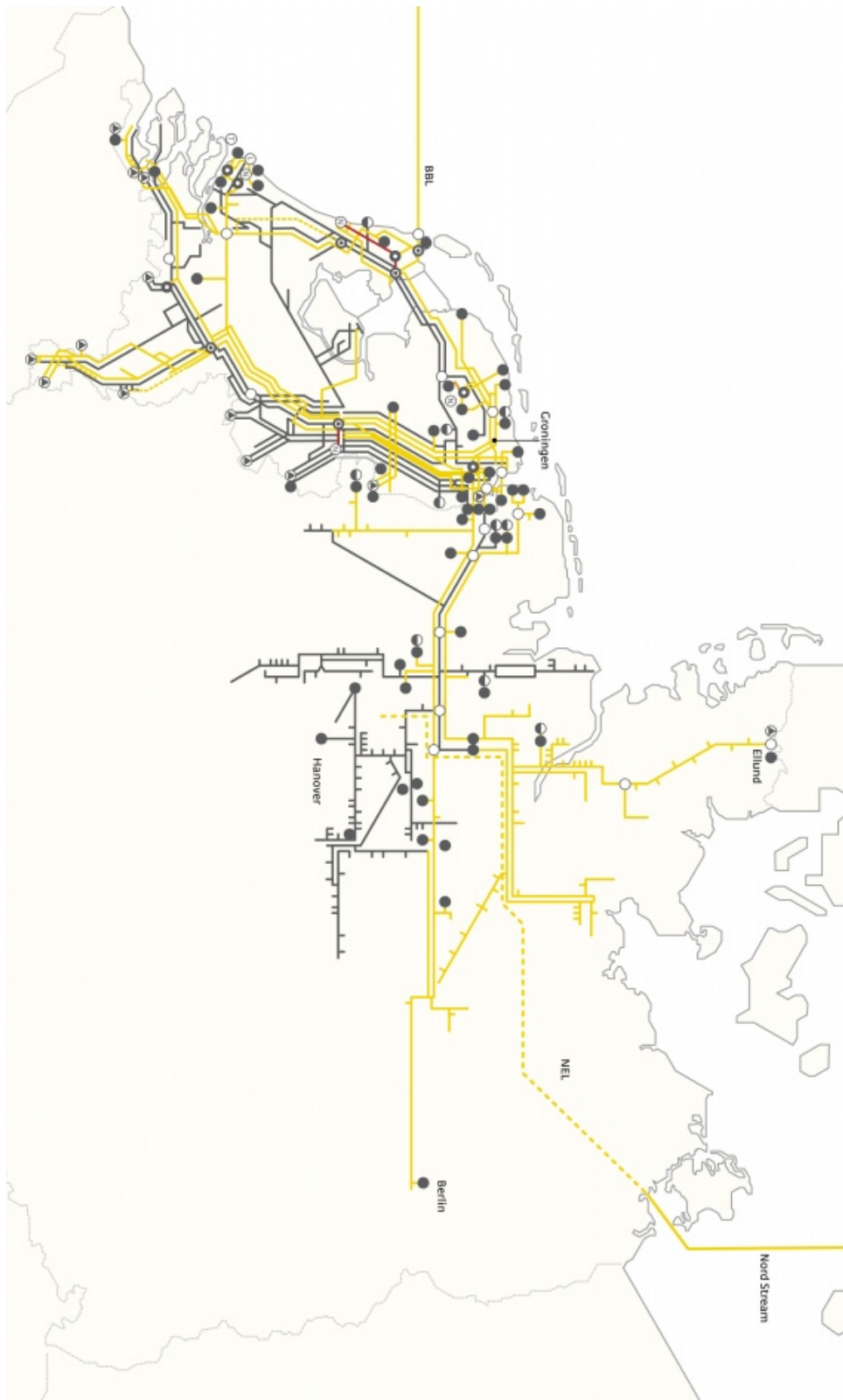


Figure 1-1: The gas transport grid of Gasunie.

gas can be (physically) removed from this pipeline network. The contractors of Gasunie, who are called shippers, buy gas capacity in advance at these entry and exit points. Thus, customers enter into contracts which allow them to book capacity at certain entry or exit points in the network for a certain period (year, month or day). For example, it is possible that a shipper injects gas into the network at *Oude Statenzijl* and removes gas from the network at *Zandvliet*. Shippers are free to choose the combinations of entry and exit points, but must ensure that a certain balance is maintained between the volume of gas injected into the system and the volume taken off. Gasunie Transport Services works with a balancing regime, that is the method by which the network can be kept at the right pressure, ensuring that, on balance, the same amount of gas is removed from the network as is injected into it. Besides this, customers can trade gas amongst themselves at a virtual market place which is called Title Transfer Facility. GTS wants a liquid and competitive capacity market, because it makes the GTS infrastructure more attractive to its customers. [2]

Thus, shippers independently contract the right to use the network on the entry and exit points and the combination of entries and exits is known only at the time of actual use. The main goal of GTS is to ensure that all shippers are getting enough gas at the right time and that enough gas is being injected into the network. A tool in this process can be, for example, remotely opening or closing valves.⁴

One of the questions is, whether or not all possible future transport which will be used by the shippers can be met. Gasunie already has a strategy to handle this problem by selecting a small number (around 50) of severe entry-exit combinations, which are called shipping variants or stress tests, such that if the transport capacity is manageable regarding the pipeline network for these shipping variants (stress tests), then it is possible to satisfy all entry-exit combinations. So, shipping variants represent severe transport situations through the pipeline network. If the network can manage the most severe transport situations, then the network is sufficient and no adjustments have to be made to the network. When determining these shipping variants, the contracts and realistic market behavior are considered. Summarizing the above, a shipping variant is a severe transport situation within contractual limits which describes realistic market behavior and determines the size and shape of the gas network, and therefore can be called a stress test as well.

1-3 Multi case approach

GTS uses simulation models to determine the transport capacity of the gas transport network. A commonly used simulation model is the model MCA, which stands for multi case approach and is an important tool of the network planning process. The transport calculations for given transport situations through the gas network are conducted in MCA. These transport calculations indicate how the gas is flowing through the gas transport network and calculate all relevant variables, like the flow, pressure, temperature and gas quality.[3]

⁴Information is taken from the GTS site <http://www.gasunietransportservices.nl/en/about-gts>.

1-4 Outline of this thesis

The outline of the remainder of this thesis is as follows:

Chapter 2 *Description of the stated problem.*

The concept ‘stress test’ is illustrated on the basis of a small network and an algorithm to generate stress tests, used by GTS, is posed as well. Besides, some characterizations of a stress test are summarized. In addition, the research questions for this project are formulated at the end of this chapter.

Chapter 3 *Mathematical background*

The dissimilarity measure, named quadratic form distance, is introduced in this chapter and some examples are discussed to illustrate this dissimilarity measure. In addition, some properties of the quadratic form distance are mentioned and dealt with. A great part of this chapter is dedicated to the performed investigation regarding positive semidefiniteness on a subspace which is related to the quadratic form distance.

Chapter 4 *Definitions correlation matrix*

The discussed dissimilarity measure, the quadratic form distance, uses a so-called correlation matrix and some different choices for this correlation matrix are summarized and tested on some gas networks of GTS.

Chapter 5 *Reducing generated set of stress tests*

This chapter explains some suggested ideas to reduce a generated set of stress tests. A few experiments regarding this reduction are conducted and summarized in the second section. In addition, its results are analyzed in the third section and some more experiments are performed and mentioned at the end of this chapter due to these results.

Chapter 6 *Conclusions and recommendations*

This final chapter gives an overview of the conclusions and recommendations of this thesis.

Description of the stated problem

Any balanced combination of entry and exit points defines a transport situation through the gas transmission network, as described in Chapter 1. We consider the most severe transport situations that can occur in this network, during this project. These severe transport situations which describe realistic market behavior and determine the network configuration within contractual limits, are called shipping variants or stress tests. Whether or not a transport situation is severe depends, for example, on the allowed pressure at the entry and exit points, the length and width (diameter) of the pipelines in the network. [4]

It takes a relatively large amount of time to evaluate these stress tests for a large set of stress tests. The aim of this thesis is therefore, to reduce the generated set of stress tests as much as possible. When reducing this set, we have to make sure that we derive an exhaustive or minimal subset such that if the transport capacity is manageable regarding the pipeline network for this smaller (sub)set, then it is possible to satisfy all entry-exit combinations.

First, let us consider a small gas transmission network in the next section. An algorithm to generate stress tests, given by Jarig Steringa, is described and applied on the small network after considering a small network. In addition, some background information about stress tests is given. Finally, the main research question and sub research questions for this thesis are formulated.

2-1 A small example

Suppose we have a small gas pipeline network, like the network drawn in Figure 2-1 where the five pipes all have the same length (L). In this network there exists two entry points (numbered as 1 and 2) and two exit points (numbered as 3 and 4) and 0 to 100 units of gas can be injected into the system at both entries but also removed from the system at the exit points. GTS works with a balancing regime and therefore the volume of injected gas needs to be equal to the volume of gas removed from the network in a gas transport situation.

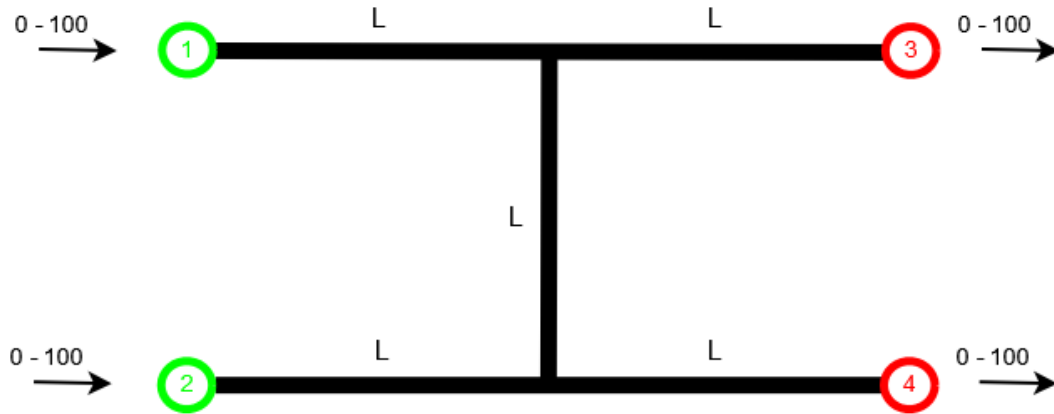


Figure 2-1: An example of a simple network.

First, we want to determine the possible severe transport situations (shipping variants or stress tests) in this case. In general a severe gas transport situation occurs, when a large volume of gas is transported over a long distance. Three severe transport situations can be identified for this network, i.e.

1. injecting 100 units of gas into the system at (entry) point 1 and removing 100 units of gas at (exit) point 4. This stress test is drawn in Figure 2-2.

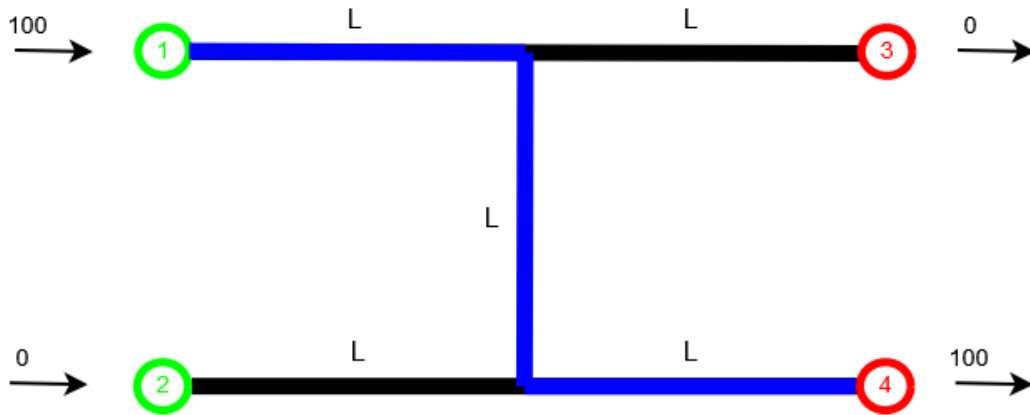


Figure 2-2: Stress test 1

We see that a maximal gas volume (100 units) is transported over the longest distance in this gas network ($3L$), for this stress test.

2. injecting 100 units of gas into the system at (entry) point 2 and removing 100 units of gas at (exit) point 3. This stress test is shown in Figure 2-3.

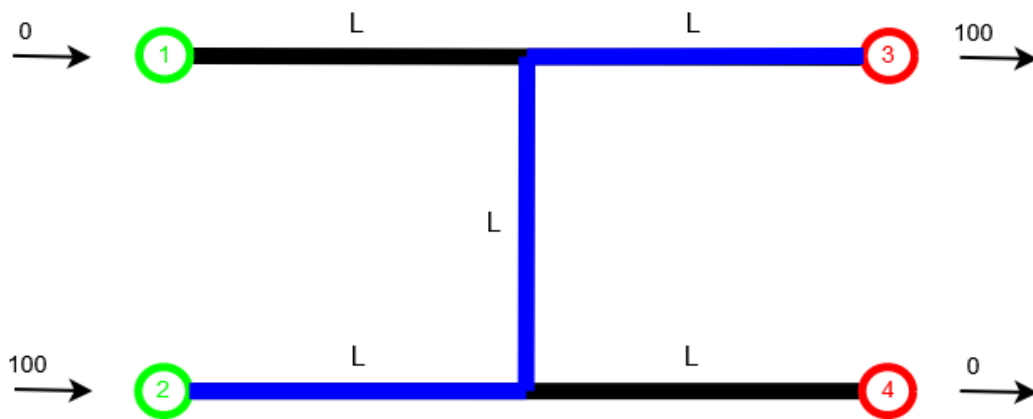


Figure 2-3: Stress test 2

3. injecting 100 units of gas into the system at both (entry) points 1 and 2 and removing 100 units of gas at both (exit) points 3 and 4. This stress test is drawn in Figure 2-4.

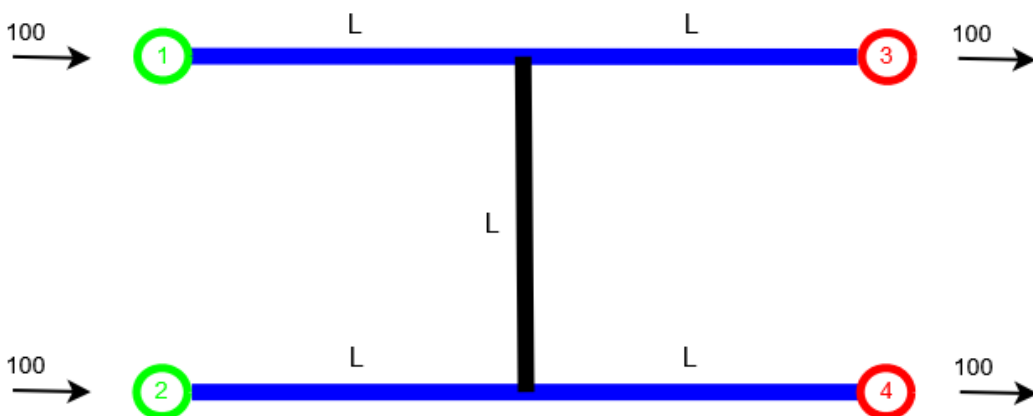


Figure 2-4: Stress test 3

We see that for a small example we have a small set (three) of stress tests. We will have a large set of stress tests for a larger and more complex network. Then it takes a relatively large amount of time to evaluate this compiled set of stress tests. Therefore, we want to reduce the set of stress tests as much as possible. A first step in this reducing process is to find a method which can compare the stress tests. Then it may be possible to eliminate stress tests based on (almost) similarity.

2-2 Algorithm to generate stress tests

GTS applies several steps to generate stress tests. Firstly, these steps are mentioned and secondly, the algorithm is applied on a small example to clarify these steps. The algorithm consists, roughly speaking, of the following steps: [5]

1. Choose any network point as reference (anchor) point (A). This point does not have to be an entry or exit point.
2. Draw up a distance table of entry and exit points with respect to this point along the network/graph. The distances between all entry/exit points and the reference point are listed in such a distance table.
3. Take list of all capacities on these entry (N_i) and exit (X_i) points.
4. We consider the transport moment T in this step, which is defined as the product of the system throughput Q and the mean transport distance D of the gas flowing through the system. The system throughput Q is the same as the total flow from entry to exit. The mean transport distance D is set to be equal to the distance between the center of gravity of the entry points and the center of gravity of the exit points in this algorithm, and is illustrated in Figure 2-5.

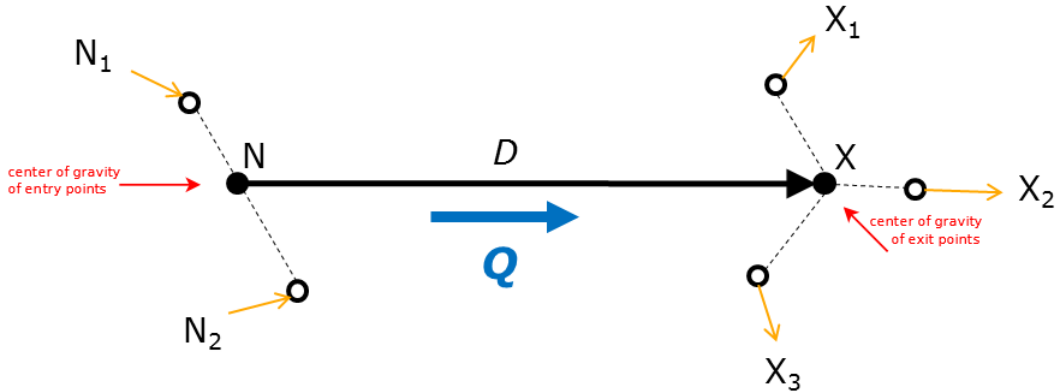


Figure 2-5: An illustration of the definition of the mean transport distance.

The location of the entry center N with respect to the reference point A is represented by its distance vector \mathbf{D}_N pointing from A to N , see Figure 2-6. In the same way, a distance vector \mathbf{D}_X is defined for the exit center X . These two distance vectors can be formulated in terms of the distance vectors of entry and exit points:

$$\mathbf{D}_N = \frac{\sum_i N_i \cdot \mathbf{D}_{N_i}}{\sum_i N_i} \quad \text{and} \quad \mathbf{D}_X = \frac{\sum_i X_i \cdot \mathbf{D}_{X_i}}{\sum_i X_i}.$$

The mean transport distance vector \mathbf{D} can be rewritten as $\mathbf{D} = \mathbf{D}_X - \mathbf{D}_N$ which follows from standard vector addition in Figure 2-6.

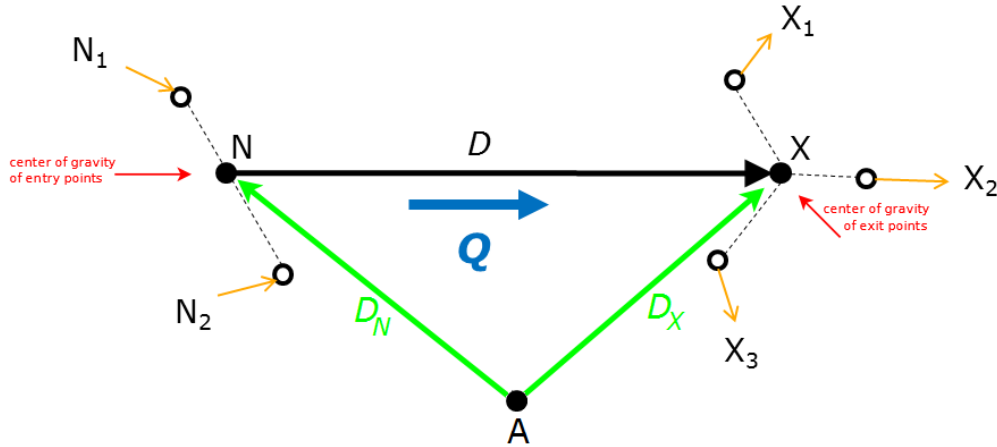


Figure 2-6: An illustration of the reference point A and entry and exit centers N and X .

The system throughput Q is, due to the required balance in the system, equal to $\sum_i N_i = \sum_i X_i$ and can be represented by a vector \mathbf{Q} as well. This vector \mathbf{Q} has the same direction as the vector \mathbf{D} and therefore the transport moment can be denoted as the inner product of \mathbf{Q} and \mathbf{D} . This inner product results in the following final form for the transport moment through the gas network

$$T = \sum_i X_i D_{X_i} - \sum_i N_i D_{N_i}$$

Then, we need to maximize this transport moment for the entry and exit centers subject to the following balance constraint in the fourth step of the algorithm

$$\sum_i X_i = \sum_i N_i.$$

5. Keep the resulting NX-combination, which is a stress test.
6. Repeat these five steps for each reference point.
7. Reduce the resulting set of stress tests, for example by identifying identical stress tests or based on similarities. However, the milder cases can be deleted as well.

So, we consider again the small network introduced in Section 2-1 which is displayed in Figure 2-1, in order to illustrate this algorithm. First we choose point 1 as reference point. The second step is to draw up a distance table of entry and exit points with respect to point 1 and this table can be found in Table 2-1.

Table 2-1: The distance table of entry and exit points with respect to point 1.

	Point 1 (entry D_{N_1})	Point 2 (entry D_{N_2})	Point 3 (exit D_{X_1})	Point 4 (exit D_{X_2})
Point 1	0	$3 \cdot L$	$2 \cdot L$	$3 \cdot L$

The capacities on these entry and exit points can be listed: $N_1 = 0 - 100$, $N_2 = 0 - 100$, $X_1 = 0 - 100$ and $X_2 = 0 - 100$. The transport moment needs to be maximized after determining the distances and the capacities and is defined in step 4 as

$$\begin{aligned} \sum_i X_i D_{X_i} - \sum_i N_i D_{N_i} &= X_1 D_{X_1} + X_2 D_{X_2} - N_1 D_{N_1} - N_2 D_{N_2} \\ &= X_1 \cdot 2L + X_2 \cdot 3L - N_1 \cdot 3L, \end{aligned} \quad (2-1)$$

but we have to keep the following balance constraint in mind while maximizing the transport moment in Equation (2-1),

$$X_1 + X_2 = N_1 + N_2. \quad (2-2)$$

Note that the transport moment in Equation (2-1) is maximized (subject to the balance constraint in Equation (2-2)) for

$$N_1 = X_2 = 100 \quad \text{and} \quad N_2 = X_1 = 0.$$

This corresponds with a flow of $100 \text{ dam}^3/h$ from point 1 to point 4 through the network which refers to the stress test in Figure 2-2.

We have to apply the same principle for the other points in the network according to step 6 of the algorithm. Before we conduct this step, note that the network displayed in Figure 2-1 actually consists of six points which represent two entry points, two exit points, and two 'connector' points (see Figure 2-7).

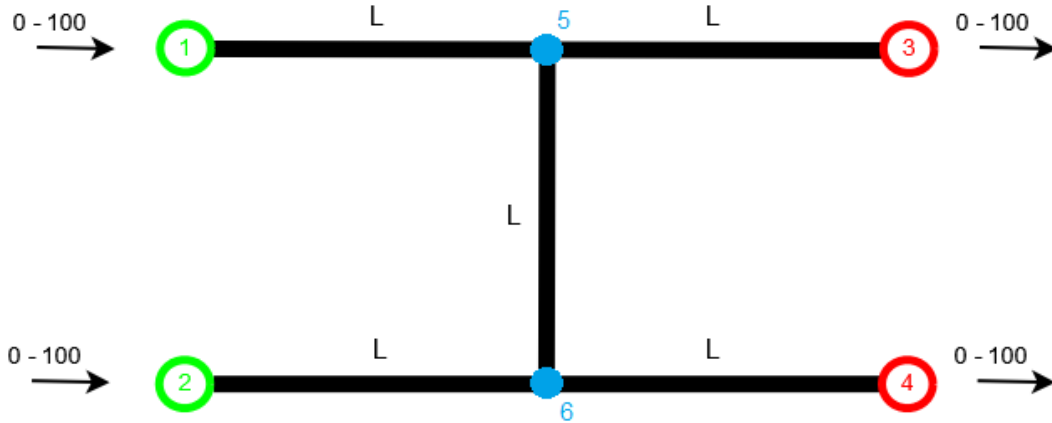


Figure 2-7: An example of a simple network, consisting of six points.

So, we can choose point 5 as reference point as well. Then Table 2-2 represents the distance table of entry and exit points with respect to this 'connector' point.

Table 2-2: The distance table of entry and exit points with respect to point 5.

	Point 1 (entry D_{N_1})	Point 2 (entry D_{N_2})	Point 3 (exit D_{X_1})	Point 4 (exit D_{X_2})
Point 5	L	$2 \cdot L$	L	$2 \cdot L$

The capacities on the entry and exit points are always the same, i.e. 0 till $100 \text{ dam}^3/h$. The transport moment is in this case equal to

$$X_1 \cdot L + X_2 \cdot 2L - N_1 \cdot L - N_2 \cdot 2L,$$

and is maximized (subject to the balance constraint) for

$$N_1 = X_2 = 100 \quad \text{and} \quad N_2 = X_1 = 0.$$

Again, this corresponds with a flow of $100 \text{ dam}^3/h$ from point 1 to point 4 through the network which is the same stress test as when point 1 is chosen as reference point.

Choosing point 3 as reference point will also result in the same stress test.

In the same way we can determine the other stress test which can be derived when point 2, 4 or 6 is chosen as reference point. Let us consider point 4 as reference point and apply steps 2 till 5 of the algorithm. The distance table is shown in Table 2-3 and the capacities are still the same.

Table 2-3: The distance table of entry and exit points with respect to point 4.

	Point 1 (entry D_{N_1})	Point 2 (entry D_{N_2})	Point 3 (exit D_{X_1})	Point 4 (exit D_{X_2})
Point 4	$3 \cdot L$	$2 \cdot L$	$3 \cdot L$	0

This results in a flow of $100 \text{ dam}^3/h$ from point 2 to point 3 ($N_2 = X_1 = 100$ and $N_1 = X_2 = 0$) which is the stress test displayed in Figure 2-3.

Finally, we end up with a set of six stress tests. The last step is to reduce this set and we can reduce the set to two (different) stress tests. It is still the question whether or not we can reduce the set some more.

Note that the stress tests in this example can also be found by searching for the nearest entry point and furthest exit point with respect to the chosen reference point and maximizing the flow between these points. For example, if we choose point 1 as reference point, then the nearest entry point is itself and the furthest exit point is point 4. The capacity (flow) is maximized in the next step, in this example up to $100 \text{ dam}^3/h$. This transport situation is a stress test and corresponds with the one drawn in Figure 2-2. If we choose point 3 or 5 as reference point, then we end up with the same stress test. Choosing point 2, 4 or 6 as reference point will result in the stress test drawn in Figure 2-3, because the nearest entry point is point 2, the furthest exit point is point 3 and the maximum capacity (flow) is 100.

2-2-1 The algorithm applied after clustering

Three stress tests are given in Section 2-1, but applying the algorithm results in only two (different) stress tests. We divide the network into clusters and then we search for severe transport situations locally, to derive the third one. Note that points 2 and 4 are the furthest points with respect to point 1, but also with respect to points 3 and 5. Therefore, points 2 and 4 can be called opposites of points 1, 3 and 5. Or the other way around, points 1 and 3 are also opposites of points 2, 4 and 6. Therefore, we can divide the points into two clusters, like it is drawn in Figure 2-8.

A second step is to multiply the distances within one cluster with a relative high factor, in this case for example with 10, which results in the two networks displayed in Figures 2-9 and

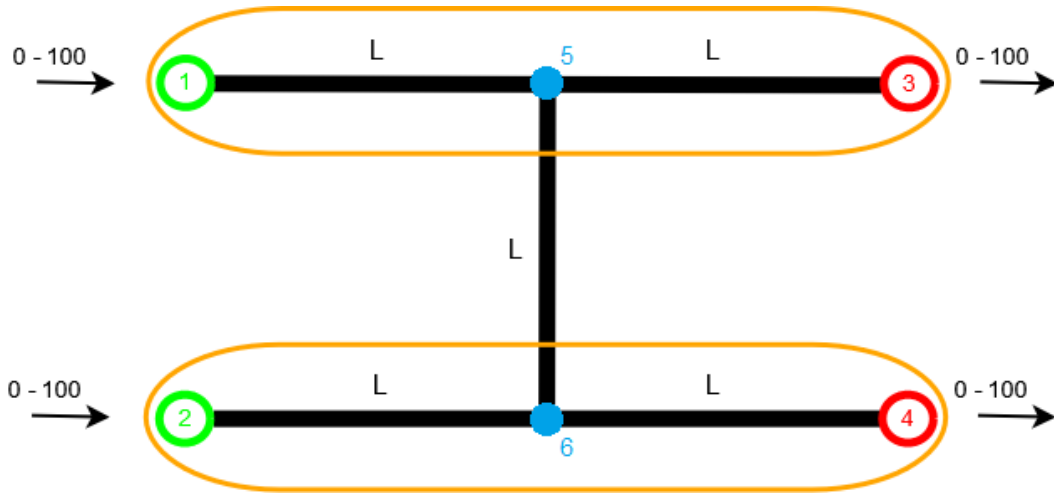


Figure 2-8: An example of a simple network divided into clusters (marked with yellow).

2-10. The algorithm can again be applied on the whole network after this multiplication, but we have to apply steps 1-4 twice (once per cluster) before determining a stress test. In that way we have the most severe situation. This results in the third stress test mentioned in Section 2-1.

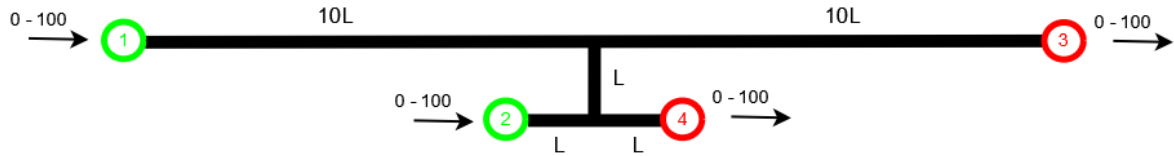


Figure 2-9: The distances in one cluster of the simple network multiplied with 10.

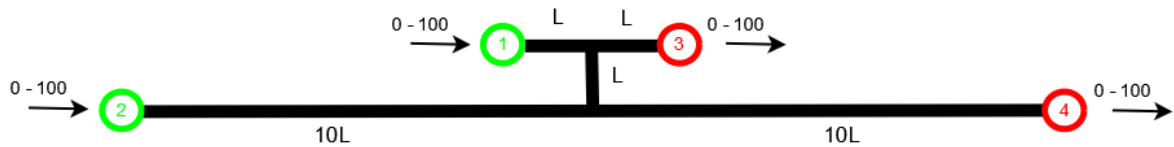


Figure 2-10: The distances in the other cluster of the simple network multiplied with 10.

It can be said that for the first two stress tests or stress tests the algorithm is globally (whole network) applied and for the third one locally (per cluster). It can be concluded that it is also important to apply the algorithm within a cluster in order to generate a complete set of stress tests.

2-3 Characterization of stress tests

Each stress test represents a specific transport situation through the network, often characterized by the use of a certain amount of transport means, like gas pipelines and compressors.

The gas pipelines and compressors are the most important transport means in the main transmission grid (HTL) of GTS, and connected to each other via pressure regulators, valves, non-return valves and other technical equipment by which different network configurations can be formed.[1]

It is important that a stress test seriously tests the capabilities or limits of the gas network in a realistic way, in order to say something about whether or not all possible future gas transport can be met. When stress tests are chosen severe enough regarding transportation load and regarding load caused by blending different types of gas, then the limits or capabilities of this network are really tested.[1]

2-3-1 Transporting gas through the network

During the project, stress tests are taken into account which are supposed to only depend on the transportation load. The load of the compressor stations and gas pipelines is considered in order to determine the transportation load of the network.[1]

When determining the load of the compressor stations, the compression power and pressure head of these stations can, for example, be investigated and for the pipelines, the pressure drop and gas flow. We are interested in transport situations which heavily load the compressor stations and pipelines for a stress test. The compression power, pressure drop and gas flow are mostly high for such a severe transport situation.[1]

Severe gas transport situations often occur, when a large amount of gas is transported over long distances through the gas network. Then it can happen that high pressure drops arise at certain transportation routes and these high pressure drops can be overcome by using a lot of compression. Therefore, the required compression power (in *MW*) can be used to measure the severity of a transport situation through the network.[1]

2-3-2 Blending different types of gas

It is also important to consider the blending of different types of gas in this network besides the transportation of natural gas through the network of GTS. The different types of gas transported through the system of GTS are:[1]

- G-gas, with a Wobbe index smaller than or equal to 44.4.
This type of gas refers to Groningen gas and is transported for the public use. The transported quantity depends strongly on the temperature.
- L-gas, with a Wobbe index between 44.4 and 47.2.
This type of gas is transported as export gas to Hilvarenbeek, Winterswijk and Zevenaar for, for example, Belgian and German public use. The L-gas demand is smaller in the summer period than in the winter period.
- H-gas, with a Wobbe index greater than or equal to 49.0.
This type of gas is the least temperature dependent of the three types and is transported to relatively large industries in the Netherlands which often need a constant gas capacity, or is exported.

A definition of the Wobbe index is the following one:

*“A comparative measure of thermal energy flow through a given size of orifice. A measure of the interchangeability of gases used for combustion. Gases which have the same Wobbe index can replace each other without a change in the relative air-fuel ratio at the same fuel metering settings.”*¹

The Wobbe index of a gas is defined as the following quotient:

$$W = \frac{H_s}{\sqrt{d}}, \quad (2-3)$$

with H_s the calorific value and d the density of the gas compared to air. This density d is formulated as $d = \frac{\rho_g}{\rho_l}$, where ρ_g and ρ_l is the density of the gas and air, respectively.[6]

Different types of gas are blended to make sure that the gas has the desired Wobbe index. For example, G-gas (with a low Wobbe index) can be enriched with H-gas which has a high Wobbe index, and H-gas can be impoverished by G-gas in order to get the desired Wobbe index. These processes take place at the blending stations of GTS. It is also possible to inject nitrogen into the H-gas (till the desired Wobbe index is achieved) to create G-gas. This is a different form of blending different types of gas.[1]

The load caused by blending different types of gas can be measured by the use of the blending stations and nitrogen installations. The H-gas overflow and the amount of injected nitrogen are considered to determine this load. The H-gas overflow represents the amount of H-gas injected into the network which does not directly flow to the H-gas market but to the L- and G-gas market (whether or not blended with nitrogen).

¹Cited from <http://www.contractorsunlimited.co.uk/cgi-bin/glossitemsrch.pl?header=Term&method=all&search=Wobbe+index>.

2-4 Research questions

The main research question during this project is:

Which techniques are available to sufficiently reduce the generated set of stress tests, to find an exhaustive subset of which the elements are mutually exclusive, given a required accuracy?

At the end of this project, we want to have found or developed such a method which also satisfies the following criteria

1. the physical characteristics of the transport network, like the pressure drop, are taken into account, when applying the method.
2. the method can distinguish between different stress tests.
3. the threshold value, used to determine which stress tests are almost equal or similar, has to relate to the uncertainty used during the generating of stress tests ($\pm 10 \text{ dam}^3/h$).
4. the parameters used in the method can be tuned.
5. the need to use specific transport physics of the gas network has to be as little as possible, before conducting the method. The motivation for this is the fact that transport calculations need to be done as well, when the (reduced) set is tested on the network. Then it has to be calculated twice, which is time consuming.
6. the method concludes, for example, that two stress tests are similar in the case one has capacities which are all the same (positive) fraction of the capacities of the other one. In this case, one stress test is less severe than the other and therefore can be removed from the generated set.
7. the method is applicable to reduce a set of stress tests depending on the load caused by blending different types of gas as well, because in general, stress tests are determined based on their transportation load and blending load.

These criteria give rise to (or correspond with) the following sub research questions:

1. Is it possible to take the physical characteristics of the transport network, for example the pressure drop, into account, when we apply the method?
2. Does the method distinguish between different stress tests or remains the calculated distance the same?
3. Which threshold value(s) should be used to label two stress tests as equal and can we relate this threshold value to the uncertainty used during the generating of stress tests ($\pm 10 \text{ dam}^3/h$)? This may be dependent on the applied method.
4. Is it possible to tune the parameters which are used in the method?
5. Do we need to know (some of) the transport physics of the gas network for a stress test or make transport calculation before applying the method?

6. Do we need to filter the set of stress tests before we apply the method? Or will the method conclude, for example, that two stress tests are similar in the case one has capacities which are all the same (positive) fraction of the capacities of the other one?
7. Is the examined method applicable to reduce a set of stress tests depending on the load caused by blending different types of gas or do we need to search for another method?

Mathematical background

Multiple methods which can be applied to reduce the generated set of stress tests, are examined during the literature study. One of the methods is the quadratic form distance and we have chosen to continue studying this method after conducting the literature study. Therefore, the quadratic form distance is first recapitulated in this chapter.

3-1 Quadratic form distance

We are searching for a way to measure the similarity between these stress tests, to reduce the set of stress tests. The stress tests can be represented by vectors and its individual dimensions are correlated, because we want to take the (mutual) distance between the network points into account as well, besides the capacities on the entry and exit points. It is possible to measure similarity between vectors for which the individual dimensions are correlated, according to the paper of Skopal et al. This similarity measuring is done with the quadratic form distance. The QFD is applied to search for similarities in a set of color images in the paper of Skopal et al. Color images can be compared by examining their color histograms. An image can be represented as a 256-dimensional histogram which describes the proportions of individual colors. Such a histogram consists of multiple bins (dimensions), in this case 256 bins, in which the i -th histogram bin represents the number of pixels in the color image having the i -th color. Because the individual dimensions (bins) are correlated, Skopal et al propose to apply the QFD to measure similarity of two histograms. Skopal et al state that it would be inappropriate to use an L_p distance, since it ignores the dimension ordering and it would even measure an image of a sunset (red tones) as more similar to a tennis ball (yellow tones) than to an orange fruit (orange tones), which is incorrect.[7]

Color histograms can also be denoted as a vector. We consider an image M as an illustration. If the colors of this image are mapped into a discrete color space containing n colors, then the color histogram of this image, denoted as $H(M)$, is a vector $(h_{c_1}, h_{c_2}, \dots, h_{c_n})$. The elements h_{c_j} of this vector represent the number of pixels of color c_j in the image M . It may be

assumed, without loss of generality, that all images contain N pixels and therefore, that the following condition holds

$$\sum_{i=1}^n h_{c_i} = N. \quad [8]$$

Several dissimilarity measures can be applied to compare two histograms H and K :

- bin-by-bin dissimilarity measures which only compare contents of corresponding histogram bins. Thus, bin-by-bin dissimilarity measures compare h_{c_i} and k_{c_i} for all i , but not h_{c_i} and k_{c_j} for $i \neq j$.
- cross-bin dissimilarity measures which compare non-corresponding bins as well. Therefore, cross-bin distances use the ground distance d_{ij} which is the distance between representative features for bin i and j . [9]

The QFD is an example of a cross-bin dissimilarity measure. We have to denote the stress tests as n -dimensional vectors in order to apply the QFD on the generated set of stress tests, and then the QFD for two n -dimensional vectors u and v is defined as

$$QFD_{\mathbf{A}}(u, v) = \sqrt{(u - v)^T \mathbf{A} (u - v)}, \quad (3-1)$$

with \mathbf{A} an $n \times n$ symmetric positive definite matrix. The matrix \mathbf{A} for RGB image histograms is defined in the paper of Skopal et al as

$$\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}, \quad \text{with } d_{max} = \max_{i,j} d_{ij}, \quad i, j = 1, 2, \dots, n, \quad (3-2)$$

with d_{ij} the Euclidean distance between representatives of colors i and j in the RGB color space.[7]

As mentioned above, the QFD is more preferable than the class of Minkowski (L_p) distances which is defined as

$$L_p(u, v) = \left(\sum_{i=1}^n |u_i - v_i|^p \right)^{1/p}, \quad p \geq 1 \quad (3-3)$$

for two vectors u and v . The QFD can deal with the correlation between the individual dimensions, while the L_p distances suppose all the vector space dimensions to be not correlated (independent).[7]

3-1-1 Symmetric positive definiteness

When applying the QFD defined in Equation (3-1), the matrix \mathbf{A} needs to be an $n \times n$ symmetric positive definite matrix. Then the following definitions should be considered for this $n \times n$ matrix \mathbf{A} :

- the matrix \mathbf{A} is symmetric $\Leftrightarrow \mathbf{A}^T = \mathbf{A}$; [10]
- the matrix \mathbf{A} is positive definite $\Leftrightarrow \forall w \in \mathbb{R}^n \setminus \{0\}: w^T \mathbf{A} w > 0$. [10]

It is also known that all the eigenvalues of a symmetric positive definite matrix are real and strictly positive.[10]

Firstly, the matrix \mathbf{A} can be defined in almost the same way for a network as for the color images, i.e.

$$\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}, \quad \text{with } d_{max} = \max_{i,j} d_{ij}, \quad i, j = 1, 2, \dots, n, \quad (3-4)$$

where d_{ij} is the distance measured in the network (length of the pipes) between the points i and j of the network.[7] The matrix \mathbf{A} is indeed symmetric according to this definition, because $d_{ij} = d_{ji}$ holds $\forall i, j$.

The Gershgorin circle theorem can be used to check whether or not the eigenvalues of the symmetric matrix \mathbf{A} are strictly positive. This theorem can be applied to bound the spectrum of the matrix \mathbf{A} and is stated as follows:

Every eigenvalue λ of \mathbf{A} is located in at least one of the n Gershgorin discs in the complex plane with center \mathbf{A}_{ii} and radius $\rho_i = \sum_{j=1, j \neq i}^n |\mathbf{A}_{ij}|$ for $i = 1, 2, \dots, n$. [10]

It follows from the definition of the matrix \mathbf{A} , Equation (3-4), that the center of each Gershgorin disc equals 1, because $d_{ii} = 0 \forall i$. The radius of these n Gershgorin discs are determined by

$$\rho_i = \sum_{j=1, j \neq i}^n |\mathbf{A}_{ij}| = \sum_{j=1, j \neq i}^n \left\{ 1 - \frac{d_{ij}}{d_{max}} \right\} = n - 1 - \frac{1}{d_{max}} \sum_{j=1, j \neq i}^n d_{ij}.$$

If $\rho_i < 1$ holds, then we know for certain that all the eigenvalues are positive, because the center \mathbf{A}_{ii} equals 1. Then the matrix \mathbf{A} is positive definite.

3-1-2 A small example

We consider again the small example discussed in Chapter 2, to illustrate the measuring with the QFD, described by Skopal et al [7]. We have derived the three stress tests of the example in Chapter 2:

1. injecting 100 units of gas into the system at (entry) point 1 and removing 100 units of gas at (exit) point 4 (Figure 2-2);
2. injecting 100 units of gas into the system at (entry) point 2 and removing 100 units of gas at (exit) point 3 (Figure 2-3);
3. injecting 100 units of gas into the system at both (entry) points 1 and 2 and removing 100 units of gas at both (exit) points 3 and 4 (Figure 2-4).

These three stress tests can be represented as vectors consisting of four elements. Here, the first element corresponds with the units of gas at point 1, the second at point 2, the third at point 3 and the fourth at point 4. Then, the three stress tests can be denoted as

1. $u_1 = (100, 0, 0, -100)$;
2. $u_2 = (0, 100, -100, 0)$;
3. $u_3 = (100, 100, -100, -100)$;

where -100 means that 100 units of gas are leaving the system at that point. The number of units of gas is also called the gas capacity.

In general, the class of Minkowski (L_p) distances is used to measure similarity of 4-dimensional vectors, but L_p distances assume that the dimensions of the vector space are independent (not correlated). In this case, the four points (two entries and two exits) of the network are correlated, because the distance with respect to each other plays an important role as well. Therefore, the four gas capacities are correlated and thus the four dimensions are correlated. Skopal et al suggest to use the QFD which is defined as

$$QFD_{\mathbf{A}}(u, v) = \sqrt{(u - v)^T \mathbf{A} (u - v)},$$

where \mathbf{A} is required to be a 4×4 symmetric positive definite matrix, to measure the similarity between these vectors.[7] This distance is also known as the \mathbf{A} -norm and can therefore be denoted as $QFD_{\mathbf{A}}(u, v) = \|u - v\|_{\mathbf{A}}$.

The matrix \mathbf{A} might again be defined as

$$\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}, \quad \text{with } d_{max} = \max_{i,j} d_{ij}, \quad i, j = 1, 2, 3, 4,$$

where d_{ij} is the distance measured in the network (length of the pipes) between the points i and j of the network.[7]

We can determine the elements of the matrix \mathbf{A} in our example where $d_{max} = 3 \cdot L$:

$$\begin{aligned} \mathbf{A}_{11} &= 1 - \frac{d_{11}}{d_{max}} = 1 - \frac{0}{3L} = 1 \\ \mathbf{A}_{22} = \mathbf{A}_{33} = \mathbf{A}_{44} &= 1 \\ \mathbf{A}_{21} = \mathbf{A}_{12} &= 1 - \frac{d_{12}}{d_{max}} = 1 - \frac{3L}{3L} = 0 \\ \mathbf{A}_{41} = \mathbf{A}_{14} &= 0 \\ \mathbf{A}_{32} = \mathbf{A}_{23} &= 0 \\ \mathbf{A}_{43} = \mathbf{A}_{34} &= 0 \\ \mathbf{A}_{31} = \mathbf{A}_{13} &= 1 - \frac{d_{13}}{d_{max}} = 1 - \frac{2L}{3L} = \frac{1}{3} \\ \mathbf{A}_{42} = \mathbf{A}_{24} &= \frac{1}{3} \end{aligned}$$

Here, $\mathbf{A}_{ii} = 1$ and $\rho_i = \sum_{j=1, j \neq i}^4 |\mathbf{A}_{ij}| = \frac{1}{3}$ for $i = 1, 2, 3, 4$, and $\mathbf{A}_{ij} = \mathbf{A}_{ji} \forall i, j$. Thus, the matrix \mathbf{A} in this example is indeed symmetric positive definite, because every eigenvalue of \mathbf{A} is real and lies within the Gershgorin disc $\left(1, \frac{1}{3}\right)$.

The distances can be calculated after compiling the matrix \mathbf{A} :

$$QFD_{\mathbf{A}}(u_1, u_2) = QFD_{\mathbf{A}}(u_2, u_1) \approx 230.94$$

$$QFD_{\mathbf{A}}(u_1, u_3) = QFD_{\mathbf{A}}(u_3, u_1) \approx 141.42$$

$$QFD_{\mathbf{A}}(u_2, u_3) = QFD_{\mathbf{A}}(u_3, u_2) \approx 141.42$$

We see that the distances between the first and third stress test and between the second and third stress test are the same and smaller than the distance between the first and second stress test from these results.

3-1-3 A second example

Let us consider a network with three entry and exit points as drawn in Figure 3-1.

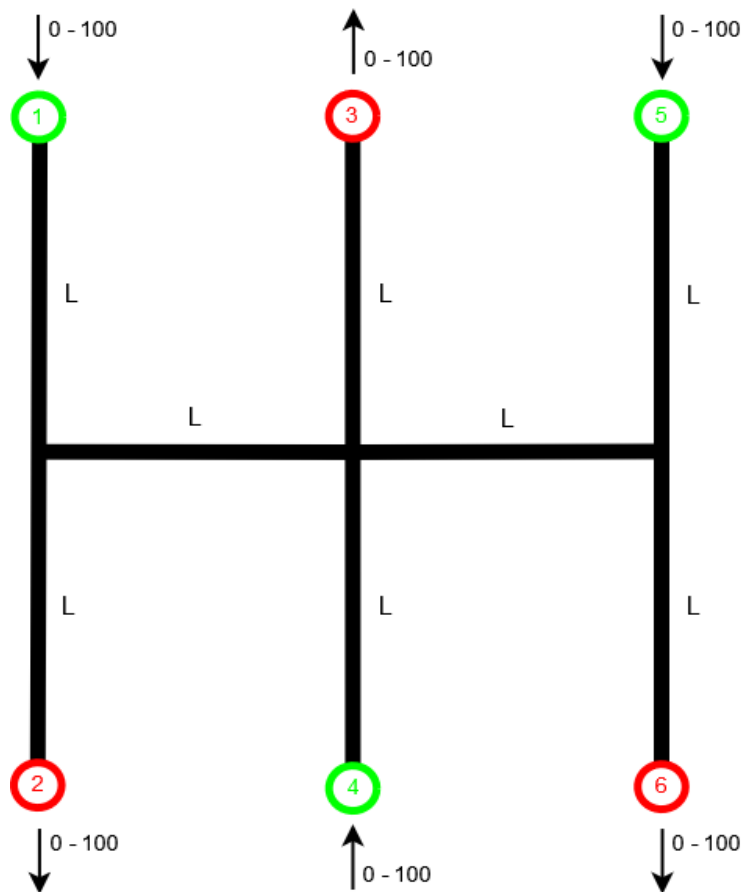


Figure 3-1: An example of a simple network, consisting of three entry and exit points.

Again every pipeline has length L and in this case d_{max} equals $4 \cdot L$. Then the matrix \mathbf{A} can be determined according to the definition in Equation (3-4):

$$\mathbf{A} = \begin{pmatrix} 1 & 0.5 & 0.25 & 0.25 & 0 & 0 \\ 0.5 & 1 & 0.25 & 0.25 & 0 & 0 \\ 0.25 & 0.25 & 1 & 0.5 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.5 & 1 & 0.25 & 0.25 \\ 0 & 0 & 0.25 & 0.25 & 1 & 0.5 \\ 0 & 0 & 0.25 & 0.25 & 0.5 & 1 \end{pmatrix}.$$

Applying the Gershgorin circle theorem results in the following two Gershgorin discs:

- a Gershgorin disc with center $\mathbf{A}_{ii} = 1$ and radius $\rho_i = 1$ for $i = 1, 2, 5, 6$;
- a Gershgorin disc with center $\mathbf{A}_{ii} = 1$ and radius $\rho_i = 1.5$ for $i = 3, 4$.

We see from this example that, it can not be concluded with the Gershgorin circle theorem that the matrix \mathbf{A} is positive definite, because it is possible that an eigenvalue is zero or real and negative. Therefore, we have to determine the eigenvalues of this matrix:

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.5, \quad \lambda_4 \approx 0.7929, \quad \lambda_5 = 1.5 \quad \text{and} \quad \lambda_6 \approx 2.2071.$$

All the eigenvalues are real and strictly positive and thus the matrix \mathbf{A} is SPD.

3-1-4 A third example

An interesting example to examine is the small network displayed in Figure 3-2, because it shows which condition must hold for the applied distance, in this case the QFD. This condition is related to the fact that two transport situations through the network can be the same, while the capacities on the entry and exit points deviate. The following example will illustrate this principle.



Figure 3-2: A small network to illustrate a needed property of the used distance.

This network consists of one pipeline of length L and two nodes of which the blue node is both an entry and exit point and the red node is an exit point. Three capacities can be defined for this network and thus the vectors representing these capacities are vectors from the vector space \mathbb{R}^3 . Let us consider the two situations displayed in Figure 3-3 where the entry point is numbered as 1 and the exit point of the blue node as 2.

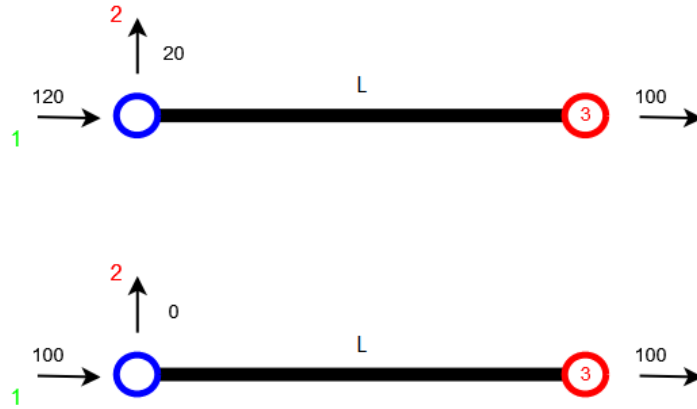


Figure 3-3: A small network with capacities on the entry and exit points.

100 dam^3/h gas is transported through the pipeline from the blue node to the red node in both situations. This transport situation is severe and thus a stress test. Both stress tests can be represented by a vector:

- in the first situation: $u = (120, -20, -100)$;
- in the second situation: $v = (100, 0, -100)$.

These two vectors are not the same in the vector space \mathbb{R}^3 , but these have to be similar with respect to the volume of transported gas through the network. Therefore, it is wanted that the distance between these two vectors is zero for a distance function (QFD).

The matrix \mathbf{A} can be compiled for this example with $d_{max} = L$ which results in

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

All the eigenvalues of \mathbf{A} are greater than or equal to 0 according to Gershgorin circle theorem. If we determine these eigenvalues which are

$$\lambda_1 = 0, \quad \lambda_2 = 1, \quad \text{and} \quad \lambda_3 = 2,$$

we see that one eigenvalue equals 0. This means that the matrix \mathbf{A} is (symmetric) positive semidefinite ($w^T \mathbf{A} w \geq 0 \forall w \in \mathbb{R}^n \setminus \{0\}$).

Let us determine the QFD:

$$QFD(u, v) = QFD(v, u) = \sqrt{(-20, 20, 0) \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -20 \\ 20 \\ 0 \end{pmatrix}} = 0,$$

which is the desired result.

So, you could say that the matrix \mathbf{A} needs to be symmetric positive semidefinite and not SPD, considering this example. Otherwise the QFD cannot be equal to zero.

3-1-5 Semi-norm

We have seen that the matrix \mathbf{A} needs to be an $n \times n$ symmetric positive definite matrix according to the literature. In addition, the correlation matrix \mathbf{A} needs to be even symmetric positive semidefinite, when we are applying the QFD on a gas transport network according to the last example. Why is it necessary that the used correlation matrix \mathbf{A} is positive semidefinite? We consider the definition of a semi-norm to answer this question:

A semi-norm $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$, is a function which satisfies the two properties:

1. $\|cu\| = |c| \cdot \|u\|$ for all scalars $c \in \mathbb{R}$, $u \in \mathbb{R}^n$
2. $\|u + v\| \leq \|u\| + \|v\|$ for all vectors $u, v \in \mathbb{R}^n$.[11]

The matrix \mathbf{A} needs to be positive semidefinite for the QFD to be real valued, because the QFD is defined as a square root (of a vector-matrix-vector multiplication).

3-2 Positive semidefinite on a subspace

Sometimes we encounter the problem that our defined matrices are not positive semidefinite and have relatively small negative eigenvalues. This motivates us to examine if the defined matrices are positive semidefinite on a subspace. Therefore, we want to consider the quadratic form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ on the subspace $\sum_i x_i = 0$. This subspace represents all the balanced gas transport situations through the network. If the quadratic form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is positive semidefinite on this subspace, then we can still use the defined matrices for the QFD. We only consider the balanced transport situations and then the QFD is a metric (restricted to this subset).

3-2-1 Subspace of balanced vectors

The following is stated in the article *Efficient Color Histogram Indexing for Quadratic Form Distance Functions* regarding this subject:

Consider a quadratic form $\mathbf{z}^T \mathbf{H} \mathbf{z}$, $\mathbf{H} = [h_{ij}]$, on the subspace $\sum_i z_i = 0$. Then this quadratic form is negative semidefinite ($\mathbf{z}^T \mathbf{H} \mathbf{z} \leq 0$) on the defined subspace, if each h_{ij} represents the distance between some points P_i and P_j in some finite dimensional L_1 or L_2 normed space. In particular, this quadratic form is negative semidefinite on the defined subspace, if the matrix \mathbf{H} satisfies the following conditions:

1. $h_{ii} = 0$
2. $h_{ij} = h_{ji}$
3. $h_{ij} \leq h_{ik} + h_{kj}$.

So, if the distance matrix $\mathbf{D} = [d_{ij}]$ satisfies the requirements stated above, then we get the following results for the matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ on the subspace $\sum_i x_i = 0$:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_i \sum_j x_i x_j \left(1 - \frac{d_{ij}}{d_{max}}\right) = \sum_i \sum_j x_i x_j - \sum_i \sum_j x_i x_j \frac{d_{ij}}{d_{max}},$$

which is equivalent with

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = - \sum_i \sum_j x_i x_j \frac{d_{ij}}{d_{max}},$$

because $\sum_i \sum_j x_i x_j = \sum_i x_i (\sum_j x_j) = \sum_i x_i \cdot 0 = 0$. We can write the last expression as

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = - \frac{1}{d_{max}} \cdot \mathbf{x}^T \mathbf{D} \mathbf{x}.$$

Then it holds that $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$, because $\mathbf{x}^T \mathbf{D} \mathbf{x} \leq 0$ on the subspace $\sum_i x_i = 0$. Thus, the matrix \mathbf{A} is positive semidefinite on this subspace, when \mathbf{D} satisfies the three stated requirements.[12]

The first question which arises, is: does the distance matrix $\mathbf{D} = [d_{ij}]$ satisfy the three required conditions? The first two conditions have been satisfied, because of the definition of d_{ij} . It represents the length of the pipeline between two network points i and j and therefore, $d_{ii} = 0$ and $d_{ij} = d_{ji}$ (matrix is symmetric). In addition, the total length of the pipelines between network points i and j is the sum of the two shortest pipelines between points i and k and between k and j , when the shortest length between i and j is represented by the ‘path’ (of pipes) from i to j along k . In this case, it holds that $d_{ij} = d_{ik} + d_{kj}$. However, it is possible that more ‘paths’ exist between i and j which do not pass point k . The value d_{ij} represents the shortest distance (‘path’) between points i and j . Therefore, the third condition $d_{ij} \leq d_{ik} + d_{kj}$ holds as well.

The second question which needs to be answered, is whether or not the statement presented in the article is true or the three requirements for $\mathbf{H} = [h_{ij}]$ are sufficient. We try to answer this question in the next subsection.

Quadratic forms and definite matrices

We are interested in the quadratic form $Q(\mathbf{z}) = \mathbf{z}^T \mathbf{H} \mathbf{z}$, where $\mathbf{H} = [h_{ij}]$ satisfies the conditions $h_{ii} = 0 \forall i$, $h_{ij} = h_{ji} \forall i, j$, and $h_{ij} \leq h_{ik} + h_{kj} \forall i, j$ and k .

First, we examine whether or not the matrix \mathbf{H} is negative semidefinite. If the matrix \mathbf{H} is not negative semidefinite in general, then we investigate the quadratic form $\mathbf{z}^T \mathbf{H} \mathbf{z}$ under the constraint $\sum_i z_i = 0$. We expect that the matrix \mathbf{H} is not negative semidefinite in general.

We want to apply the following theorem from Simon and Blume to determine whether or not the matrix \mathbf{H} is negative semidefinite:

Theorem *An $n \times n$ symmetric matrix \mathbf{H} is negative semidefinite if and only if every principal minor of odd order is ≤ 0 and every principal minor of even order is ≥ 0 . [13]*

Hereby is a k th order principal minor of \mathbf{H} the determinant of a k th order principal submatrix of \mathbf{H} which is a $k \times k$ submatrix of \mathbf{H} formed by deleting $n - k$ columns and the same $n - k$ rows from \mathbf{H} . We consider the following 4×4 matrix to illustrate what a k th order principal minor is:

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{pmatrix}.$$

Then there exists one fourth order principal minor which is $\det(\mathbf{H})$. We have four third order principal minors:

- $\begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{vmatrix}$, formed by deleting column 4 and row 4;
- $\begin{vmatrix} h_{11} & h_{12} & h_{14} \\ h_{21} & h_{22} & h_{24} \\ h_{41} & h_{42} & h_{44} \end{vmatrix}$, formed by deleting column 3 and row 3;
- $\begin{vmatrix} h_{11} & h_{13} & h_{14} \\ h_{31} & h_{33} & h_{34} \\ h_{41} & h_{43} & h_{44} \end{vmatrix}$, formed by deleting column 2 and row 2;
- $\begin{vmatrix} h_{22} & h_{23} & h_{24} \\ h_{32} & h_{33} & h_{34} \\ h_{42} & h_{43} & h_{44} \end{vmatrix}$, formed by deleting column 1 and row 1.

There exists six second order principal minors as well:

- $\begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}$, formed by deleting the last two columns and rows;
- $\begin{vmatrix} h_{22} & h_{23} \\ h_{32} & h_{33} \end{vmatrix}$, formed by deleting columns 1 and 4 and rows 1 and 4;
- $\begin{vmatrix} h_{11} & h_{13} \\ h_{31} & h_{33} \end{vmatrix}$, formed by deleting columns 2 and 4 and rows 2 and 4;
- $\begin{vmatrix} h_{22} & h_{24} \\ h_{42} & h_{44} \end{vmatrix}$, formed by deleting columns 1 and 3 and rows 1 3;
- $\begin{vmatrix} h_{11} & h_{14} \\ h_{41} & h_{44} \end{vmatrix}$, formed by deleting columns 2 and 3 and rows 2 and 3;
- $\begin{vmatrix} h_{33} & h_{34} \\ h_{43} & h_{44} \end{vmatrix}$, formed by deleting the first two columns and rows.

Finally, we have four first order principal minors, namely $|h_{11}|$, $|h_{22}|$, $|h_{33}|$ and $|h_{44}|$. So, we have $\binom{n}{k} = \binom{4}{k}$ k th order principal minors for a 4×4 matrix and the total number of principal minors is equal to $\sum_{k=1}^n \binom{n}{k} = 2^n - 1 = 2^4 - 1 = 15$. [13]

We first determine the first order principal minors of the $n \times n$ matrix \mathbf{H} : $|h_{11}| = |h_{22}| = \dots = |h_{nn}| = 0$. In addition, a second order principal minor of \mathbf{H} is $\begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} = \begin{vmatrix} 0 & h_{12} \\ h_{12} & 0 \end{vmatrix} = -h_{12}^2 \leq 0$, while it is a principal minor of even order. So, we cannot conclude with the theorem described above that the matrix \mathbf{H} is negative semidefinite. Therefore, we take the linear constraint $\sum_i z_i = 0$ into account as well.

One linear constraint

Note that the linear constraint $\sum_i z_i = 0$ can be written as

$$\mathbf{A}^T \mathbf{z} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = 0.$$

The next step is to check negative semidefiniteness of \mathbf{H} with the following theorem from an article of Debreu:

Theorem *If \mathbf{H} is symmetric and $|\mathbf{A}_m| \neq 0$, then $\mathbf{z}^T \mathbf{H} \mathbf{z} \leq 0$ for every \mathbf{z} such that $\mathbf{A}^T \mathbf{z} = 0$ if and only if $(-1)^r \begin{vmatrix} \mathbf{H}_r^\pi & \mathbf{A}_{r,m}^\pi \\ \mathbf{A}_{r,m}^{\pi T} & \mathbf{0} \end{vmatrix} \geq 0$ for $r = m + 1, \dots, n$, and for any π . [14]*

This theorem is, for example, also formulated (in other words) in the article of Ikramov and Savel'eva. [15]

Here, π denotes a permutation of the first n integers, \mathbf{H}^π represents the matrix obtained from \mathbf{H} by performing the permutation π on its columns and rows, and \mathbf{A}^π denotes the matrix obtained from \mathbf{A} by conducting the permutation π on its rows. [14] \mathbf{A} is an $n \times 1$ matrix in our case and thus m equals 1. Besides, a permutation conducted on the \mathbf{A} does not result in a different matrix, because all elements are 1. Thus $\mathbf{A}^\pi = \mathbf{A}$. Note that the matrix \mathbf{H}^π remains symmetric and has a diagonal consisting of zeros, because when two columns are switched, the same two rows are switched as well.

The matrix \mathbf{H} is symmetric, because $h_{ij} = h_{ji}$, and $|\mathbf{A}_m| = |\mathbf{A}_1| = 1 \neq 0$. So, we have to prove that $(-1)^r \begin{vmatrix} \mathbf{H}_r^\pi & \mathbf{A}_{r,1}^\pi \\ \mathbf{A}_{r,1}^{\pi T} & \mathbf{0} \end{vmatrix} \geq 0$ for $r = 2, \dots, n$, and for any π .

We get for $r = 2$

$$\begin{aligned} (-1)^2 \cdot \begin{vmatrix} h_{ii} & h_{ij} & 1 \\ h_{ji} & h_{jj} & 1 \\ 1 & 1 & 0 \end{vmatrix} &= \begin{vmatrix} 0 & h_{ij} & 1 \\ h_{ij} & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = \\ (-1)^{3+1} \begin{vmatrix} h_{ij} & 1 \\ 0 & 1 \end{vmatrix} + (-1)^{3+2} \begin{vmatrix} 0 & 1 \\ h_{ij} & 1 \end{vmatrix} &= 2h_{ij} \geq 0 \quad \forall i, j. \end{aligned}$$

It is greater than or equal to zero, because h_{ij} represents a distance between two points P_i and P_j and therefore is non-negative. So, the condition described in the theorem has been satisfied for $r = 2$.

We have for $r = 3$

$$\begin{aligned} (-1)^3 \cdot \begin{vmatrix} h_{ii} & h_{ij} & h_{ik} & 1 \\ h_{ji} & h_{jj} & h_{jk} & 1 \\ h_{ki} & h_{kj} & h_{kk} & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} &= - \begin{vmatrix} 0 & h_{ij} & h_{ik} & 1 \\ h_{ij} & 0 & h_{jk} & 1 \\ h_{ik} & h_{jk} & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} = \\ -h_{ij}^2 - h_{ik}^2 - h_{jk}^2 + 2h_{ij}h_{ik} + 2h_{ij}h_{jk} + 2h_{ik}h_{jk}. \end{aligned}$$

Now, we want to use the triangle inequality $h_{ik} + h_{kj} \geq h_{ij}$. Therefore, we write the part $2h_{ij}h_{ik} + 2h_{ij}h_{jk} + 2h_{ik}h_{jk}$ as $h_{ij}(h_{ik} + h_{jk}) + h_{ik}(h_{ij} + h_{jk}) + h_{jk}(h_{ij} + h_{ik})$ and apply the triangle inequality three times. Then $2h_{ij}h_{ik} + 2h_{ij}h_{jk} + 2h_{ik}h_{jk} \geq h_{ij}^2 + h_{ik}^2 + h_{jk}^2$ holds and we get the following result

$$(-1)^3 \cdot \begin{vmatrix} h_{ii} & h_{ij} & h_{ik} & 1 \\ h_{ji} & h_{jj} & h_{jk} & 1 \\ h_{ki} & h_{kj} & h_{kk} & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} \geq 0 \quad \forall i, j, k.$$

We have to calculate the following 5×5 determinant for $r = 4$

$$(-1)^4 \cdot \begin{vmatrix} h_{ii} & h_{ij} & h_{ik} & h_{il} & 1 \\ h_{ji} & h_{jj} & h_{jk} & h_{jl} & 1 \\ h_{ki} & h_{kj} & h_{kk} & h_{kl} & 1 \\ h_{li} & h_{lj} & h_{lk} & h_{ll} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix} = \begin{vmatrix} 0 & h_{ij} & h_{ik} & h_{il} & 1 \\ h_{ij} & 0 & h_{jk} & h_{jl} & 1 \\ h_{ik} & h_{jk} & 0 & h_{kl} & 1 \\ h_{il} & h_{jl} & h_{kl} & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix},$$

which equals

$$\begin{aligned} & -2h_{ij}^2 h_{kl} - 2h_{ij} h_{kl}^2 - 2h_{ik}^2 h_{jl} - 2h_{ik} h_{jl}^2 - 2h_{il}^2 h_{jk} - 2h_{il} h_{jk}^2 \\ & - 2h_{ij} h_{ik} h_{jk} - 2h_{ij} h_{il} h_{jl} - 2h_{ik} h_{il} h_{kl} - 2h_{jk} h_{jl} h_{kl} \\ & + 2h_{ij} h_{ik} h_{jl} + 2h_{ij} h_{ik} h_{kl} + 2h_{ij} h_{il} h_{jk} + 2h_{ij} h_{il} h_{kl} + 2h_{ij} h_{jk} h_{kl} + 2h_{ij} h_{jl} h_{kl} \\ & + 2h_{ik} h_{il} h_{jk} + 2h_{ik} h_{il} h_{jl} + 2h_{ik} h_{jk} h_{jl} + 2h_{ik} h_{jl} h_{kl} + 2h_{il} h_{jk} h_{jl} + 2h_{il} h_{jk} h_{kl}. \end{aligned}$$

We can apply the triangle inequality on the first six terms:

- $-2h_{ij}^2 h_{kl} \geq -h_{ij}(h_{ik} + h_{jk})h_{kl} - h_{ij}(h_{il} + h_{jl})h_{kl}$
 $= -h_{ij}h_{ik}h_{kl} - h_{ij}h_{jk}h_{kl} - h_{ij}h_{il}h_{kl} - h_{ij}h_{jl}h_{kl}$
- $-2h_{ij} h_{kl}^2 \geq -h_{ij}(h_{ik} + h_{il})h_{kl} - h_{ij}(h_{jk} + h_{jl})h_{kl}$
 $= -h_{ij}h_{ik}h_{kl} - h_{ij}h_{il}h_{kl} - h_{ij}h_{jk}h_{kl} - h_{ij}h_{jl}h_{kl}$
- $-2h_{ik}^2 h_{jl} \geq -h_{ik}(h_{ij} + h_{jk})h_{jl} - h_{ik}(h_{il} + h_{kl})h_{jl}$
 $= -h_{ij}h_{ik}h_{jl} - h_{ik}h_{jk}h_{jl} - h_{ik}h_{il}h_{jl} - h_{ik}h_{kl}h_{jl}$
- $-2h_{ik} h_{jl}^2 \geq -h_{ik}(h_{ij} + h_{il})h_{jl} - h_{ik}(h_{jk} + h_{kl})h_{jl}$
 $= -h_{ij}h_{ik}h_{jl} - h_{ik}h_{il}h_{jl} - h_{ik}h_{jk}h_{jl} - h_{ik}h_{kl}h_{jl}$
- $-2h_{il}^2 h_{jk} \geq -h_{il}(h_{ij} + h_{jl})h_{jk} - h_{il}(h_{ik} + h_{kl})h_{jk}$
 $= -h_{ij}h_{il}h_{jk} - h_{il}h_{jk}h_{jl} - h_{il}h_{ik}h_{jk} - h_{il}h_{kl}h_{jk}$
- $-2h_{il} h_{jk}^2 \geq -h_{il}(h_{ij} + h_{ik})h_{jk} - h_{il}(h_{jl} + h_{kl})h_{jk}$
 $= -h_{ij}h_{il}h_{jk} - h_{il}h_{jk}h_{jl} - h_{il}h_{ik}h_{jk} - h_{il}h_{kl}h_{jk}$

The sum of these six terms gives then $-2h_{ij}^2 h_{kl} - 2h_{ij} h_{kl}^2 - 2h_{ik}^2 h_{jl} - 2h_{ik} h_{jl}^2 - 2h_{il}^2 h_{jk} - 2h_{il} h_{jk}^2 \geq 2h_{ij} h_{ik} h_{jl} + 2h_{ij} h_{ik} h_{kl} + 2h_{ij} h_{il} h_{jk} + 2h_{ij} h_{il} h_{kl} + 2h_{ij} h_{jk} h_{kl} + 2h_{ij} h_{jl} h_{kl} + 2h_{ik} h_{il} h_{jk} + 2h_{ik} h_{il} h_{jl} + 2h_{ik} h_{jk} h_{jl} + 2h_{ik} h_{jl} h_{kl} + 2h_{il} h_{jk} h_{jl} + 2h_{il} h_{jk} h_{kl}$. Thus, we get

$$(-1)^4 \cdot \begin{vmatrix} h_{ii} & h_{ij} & h_{ik} & h_{il} & 1 \\ h_{ji} & h_{jj} & h_{jk} & h_{jl} & 1 \\ h_{ki} & h_{kj} & h_{kk} & h_{kl} & 1 \\ h_{li} & h_{lj} & h_{lk} & h_{ll} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix} \geq -2h_{ij} h_{ik} h_{jk} - 2h_{ij} h_{il} h_{jl} - 2h_{ik} h_{il} h_{kl} - 2h_{jk} h_{jl} h_{kl} \quad \forall i, j, k.$$

If we choose to apply the triangle inequality on different terms, we still end up with an expression unequal to zero or possible negative terms. This motivates us to look for a counterexample.

An example

Consider the matrix

$$\mathbf{H} = \begin{pmatrix} 0 & 222 & 396 & 220 & 493 \\ 222 & 0 & 246 & 442 & 286 \\ 396 & 246 & 0 & 335 & 532 \\ 220 & 442 & 335 & 0 & 273 \\ 493 & 286 & 532 & 273 & 0 \end{pmatrix}.$$

This matrix satisfies the three conditions, $h_{ii} = 0$, $h_{ij} = h_{ji}$ and $h_{ij} \leq h_{ik} + h_{kj}$, but is not negative semidefinite on the subspace $\sum_i z_i = 0$. We found a vector \mathbf{z} for which $z_1 + z_2 + z_3 + z_4 + z_5 = 0$ and $\mathbf{z}^T \mathbf{H} \mathbf{z} > 0$. This vector is the following one

$$\mathbf{z} = \begin{pmatrix} -0.427284376156840 \\ 0.565869316486029 \\ -0.284218864707640 \\ 0.523280609740979 \\ -0.377646685362528 \end{pmatrix}.$$

However, does this answer the question, whether or not the statement presented in the article is true? It seems that the three requirements for the matrix \mathbf{H} are not sufficient, but we need to take a closer look to the statement from the article. It is mentioned just above the three requirements for \mathbf{H} that h_{ij} needs to represent the distance between some point P_i and P_j in some finite dimensional L_1 or L_2 normed space.

First, let us draw a network or graph corresponding with the given matrix \mathbf{H} , see for example Figure 3-4.

We see that a shorter geographical/Euclidean distance exists between three pair of points, between points 1 and 5, 2 and 4, and 3 and 5. A next step is, to measure these three distances and adjust the matrix \mathbf{H} . Then the matrix equals

$$\mathbf{H}_{eucl} = \begin{pmatrix} 0 & 222 & 396 & 220 & 380 \\ 222 & 0 & 246 & 320 & 286 \\ 396 & 246 & 0 & 335 & 90 \\ 220 & 320 & 335 & 0 & 273 \\ 380 & 286 & 90 & 273 & 0 \end{pmatrix},$$

and calculating $\mathbf{z}^T \mathbf{H}_{eucl} \mathbf{z}$ results in a negative number. This result could indicate that an Euclidean ground distance should be used instead of the length of pipelines for the statement to hold.

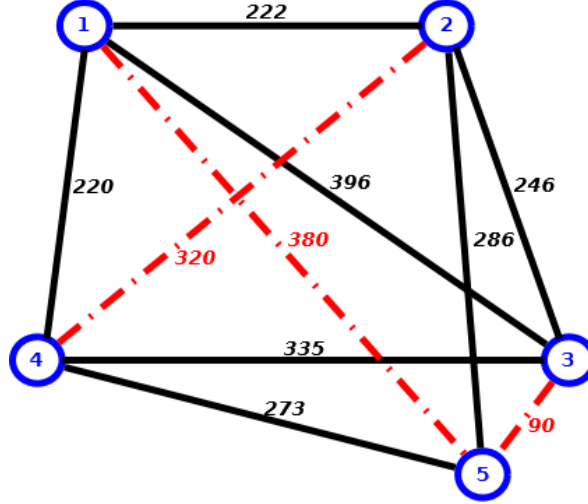


Figure 3-4: A network based on the distance matrix \mathbf{H} .

We see that the three conditions $h_{ii} = 0$, $h_{ij} = h_{ji}$ and $h_{ij} \leq h_{ik} + h_{kj}$ for the matrix $\mathbf{H} = [h_{ij}]$ are not sufficient to ensure that $\mathbf{z}^T \mathbf{H} \mathbf{z} \leq 0$ subject to $\sum_i z_i$ holds. However, we think that when h_{ij} represents the Euclidean (geographical) distance between points i and j , then $\mathbf{z}^T \mathbf{H} \mathbf{z} \leq 0$ will hold on the defined subspace. Then, we cannot use the statement from the article *Efficient Color Histogram Indexing for Quadratic Form Distance Functions*, because our ground distance d_{ij} is not an geographical/Euclidean distance, but a distance measured along the network (following the pipelines).

The example displayed above is a counterexample for our case in which h_{ij} would represent the total length of the pipelines between points i and j . This counterexample motivates us to investigate whether or not we can define a different subspace on which the distance matrix \mathbf{D} is negative semidefinite. So, an option is to examine the properties of a capacity vector. We know, for example, that a capacity vector is a balanced vector ($\mathbf{x} : \sum_i x_i = 0$). In addition, we know which vector components correspond to entry points and which to exit points for a specific network (and specific numbering of the network points). Then all ‘entry components’ should have the same sign (all positive or negative) and all ‘exit components’ an opposite sign with respect to the ‘entry components’. It is possible that some components are zero as well. Sometimes, a storage is linked to a network point and it can be an entry or exit. Thus the sign (positive, negative or zero) may vary for a storage. However, a difference of two capacity vectors does not have to satisfy these properties, only being balanced. Therefore, we conduct some experiments regarding the distance matrix \mathbf{D} of a simplified gas transport network of GTS and a more detailed network in the next subsection.

3-2-2 The distance matrix for the networks of GTS

The distance matrix \mathbf{D} is compiled and saved by *MCA*. Therefore, we can calculate the eigenvalues of this 40×40 matrix with *Matlab* and then we see that almost all eigenvalues are smaller than or equal to zero. There is one eigenvalue which is positive and equals approximately $8.8891 \cdot 10^3$. However, if we determine the correlation matrix \mathbf{A} according to Equation (3-4), then we see that all eigenvalues are greater than or equal to zero. Thus, the matrix \mathbf{A} is for the network displayed in Figure 3-5 positive semi-definite, while the distance matrix is not negative semi-definite.

So, if we determine a right eigenvector x corresponding to the eigenvalue $8.8891 \cdot 10^3$, then we notice that the sum of its vector components is not equal to zero (approximately 6.2738). It holds that $\mathbf{x}^T \mathbf{D} \mathbf{x} > 0$ for this eigenvector, which means that \mathbf{x} should represent a difference of two capacity vectors in our case. We know that a difference of two capacity vectors, which are balanced vectors, must be a balanced vector as well ($\sum_i x_i = 0$). The fact that it is not a well defined ‘difference vector’ and that there is only one positive eigenvalue could be an indication of why the correlation matrix \mathbf{A} is PSD, while there exists a vector \mathbf{x} such that $\mathbf{x}^T \mathbf{D} \mathbf{x} > 0$.

We can also examine the eigenvalues and eigenvectors of the distance matrix \mathbf{D} for a larger network, for example a simplified network of GTS which is displayed in Figure 3-6. We see from this figure that a part is added to the previous network and this part represents the simplified G-gas network of GTS. If we isolate the balanced eigenvectors, then we see that the corresponding eigenvalues are all approximately zero, which means that for these eigenvectors \mathbf{v} it holds that $\mathbf{v}^T \mathbf{D} \mathbf{v} = 0$. However, the matrix \mathbf{D} has three positive eigenvalues. If we determine the eigenvalues of the correlation matrix \mathbf{A} (according to Equation (3-4)), then this matrix has three negative eigenvalues. So, this matrix is no longer positive semi-definite for an expanded network displayed in Figure 3-6.

We could determine the corresponding eigenvectors with *Matlab* in order to say something about these negative eigenvalues. If we do this, we see that the sum of the components of the three corresponding eigenvectors is not equal to zero. This means that such an eigenvector cannot represent a difference of capacity vectors. However, we can combine these three eigenvectors (linear combination) into a new vector \mathbf{y} such that $\mathbf{y}^T \mathbf{A} \mathbf{y} < 0$ holds for this new vector \mathbf{y} . This vector \mathbf{y} is a well defined ‘difference vector’ and therefore it can be concluded that the ‘original’ correlation matrix is not PSD for this expanded network.

We are wondering if we can explain this transition from being PSD to not being PSD, when a G-gas network is added to the H-gas network. Maybe this transition is caused by the fact that some of the pipelines are intersecting/crossing each other in two dimensions (in space). Therefore, we want to see what happens if we combine two small network with each other such that we cannot draw the new network in two dimensions without crossing lines. We examine the H-shape network as example in the next subsection.

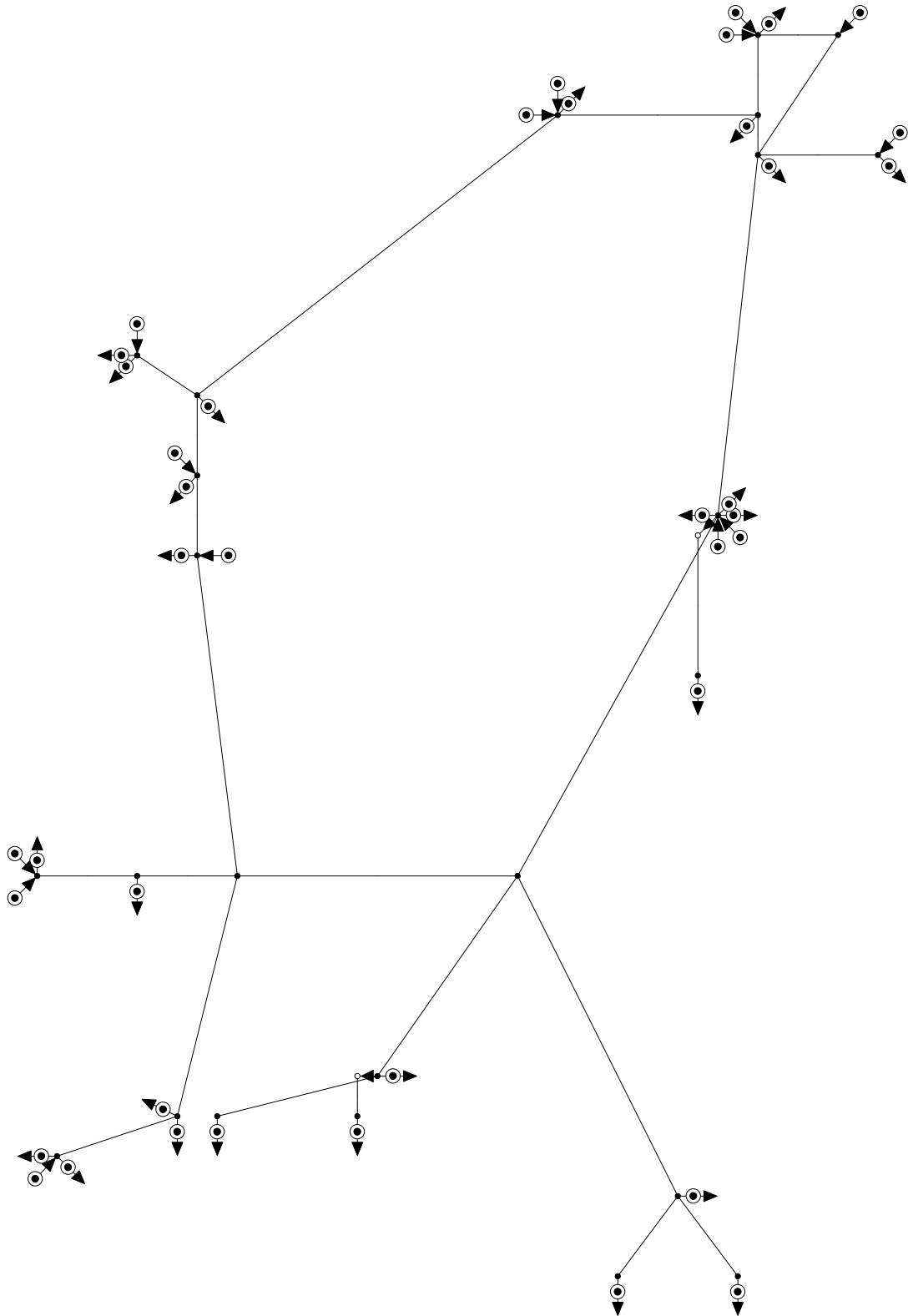


Figure 3-5: The H-gas network of GTS.

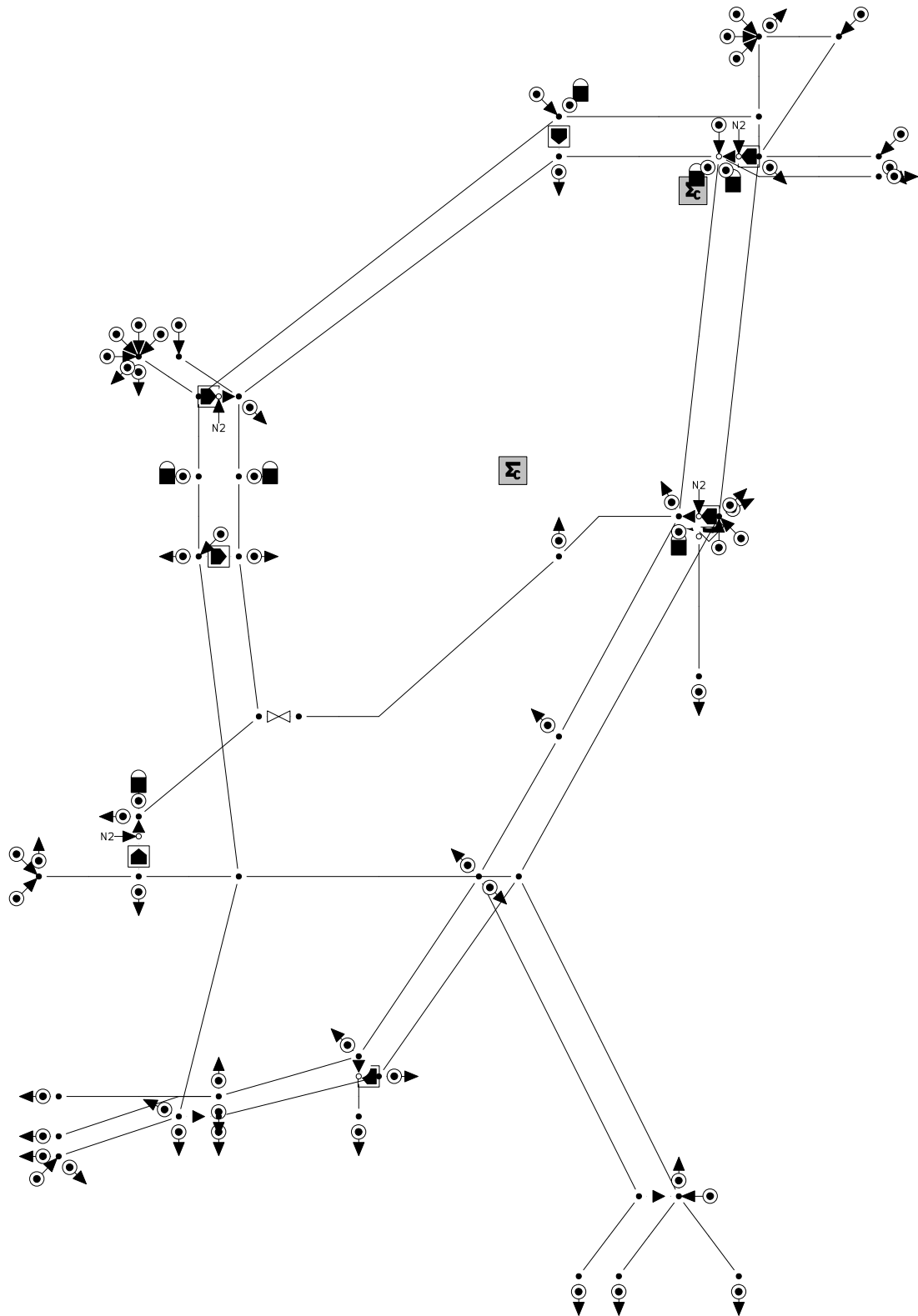


Figure 3-6: A simplification of the network of GTS.

Expanding a small network

We consider a small network, a H-shaped network, and expand this network with a second H-shaped network. We combine the two H-shape networks with a blending station and assume that H-gas is flowing through one of these networks and G-gas through the other one. This network is drawn in Figure 3-7. Every pipeline has length L and the horizontal pipelines of length L are divided in two parts by the blending station, $a_H \cdot L$ and $(1 - a_H) \cdot L$ (for the H-gas network) or $a_G \cdot L$ and $(1 - a_G) \cdot L$ (for the G-gas network).

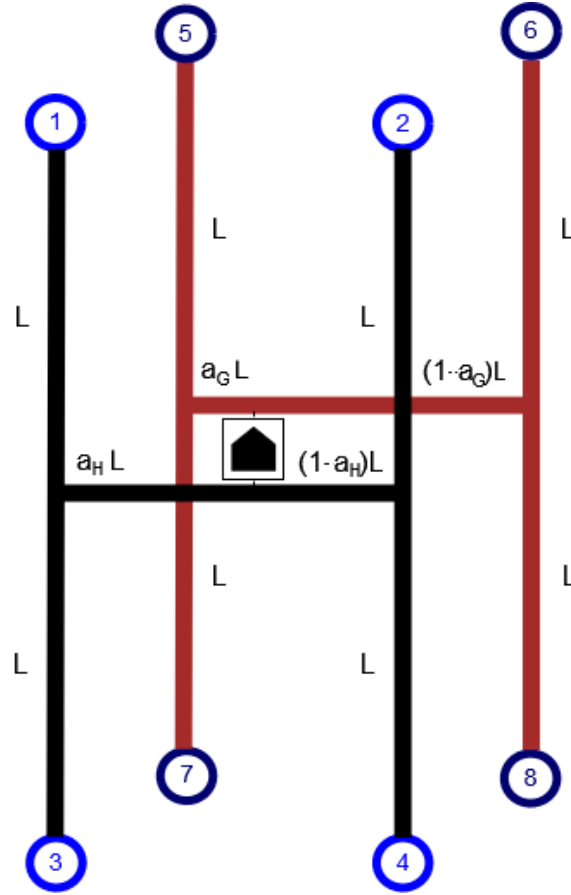


Figure 3-7: An expanded example network, in which G-gas is added.

We want to determine the distance matrix \mathbf{D} for this network. Notice that this distance matrix can be compiled as follows

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_H & \mathbf{D}_{blend} \\ \mathbf{D}_{blend}^T & \mathbf{D}_G \end{pmatrix},$$

with \mathbf{D}_H and \mathbf{D}_G the distance matrix of the decoupled H-gas network and G-gas network, respectively. Both networks have the same shape: the same number of network points and pipelines and all the pipelines have the same length. So, $\mathbf{D}_H = \mathbf{D}_G$ holds. The matrix \mathbf{D}_{blend} represents the added part as a result of the coupling of the two H-shaped networks. Thus,

we have

$$\mathbf{D}_H = \begin{pmatrix} 0 & 3L & 2L & 3L \\ 3L & 0 & 3L & 2L \\ 2L & 3L & 0 & 3L \\ 3L & 2L & 3L & 0 \end{pmatrix}$$

and

$$\mathbf{D}_{blend} = L \cdot \begin{pmatrix} 2 + a_H + a_G & 3 + a_H - a_G & 2 + a_H + a_G & 3 + a_H - a_G \\ 3 - a_H + a_G & 4 - a_H - a_G & 3 - a_H + a_G & 4 - a_H - a_G \\ 2 + a_H + a_G & 3 + a_H - a_G & 2 + a_H + a_G & 3 + a_H - a_G \\ 3 - a_H + a_G & 4 - a_H - a_G & 3 - a_H + a_G & 4 - a_H - a_G \end{pmatrix}.$$

Then we can vary the values of a_H and a_G , and calculate the eigenvalues of the correlation matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$. A few results are summarized in Table 3-1.

Table 3-1: An overview of some choices for a_H and a_G and the resulting minimum and maximum eigenvalues of the distance matrix.

Choices lengths		Eigenvalues \mathbf{A}	
a_H	a_G	Minimum	Maximum
1/2	1/2	0.6667	1.3333
2/3	1/2	0.6316	1.7362
3/4	1/2	0.6154	1.9221
4/5	1/2	0.6061	2.0292
1	1/2	0.5714	2.4268
0	0	0.5000	3.1180
0	1	0.5000	3.1180
1	0	0.5000	3.1180
1	1	0.5000	3.1180

The eigenvalues of the correlation matrix \mathbf{A} of one H-shaped network (consisting of four network points and pipes with the same length) are

$$\lambda_1 = \frac{2}{3}, \quad \lambda_2 = \frac{2}{3}, \quad \lambda_3 = 1\frac{1}{3} \quad \text{and} \quad \lambda_4 = 1\frac{1}{3}.$$

We have determined all eigenvalues with the program *Matlab* and we have seen that the correlation matrix \mathbf{A} also has these eigenvalues (four times $\frac{2}{3}$ and four times $1\frac{1}{3}$), when we choose $a_H = a_G = 1/2$ ('overlapping'). In addition we see that there is a difference in eigenvalues between a 'distinction' ($a_H = 0$ & $a_G = 0$, $a_H = 1$ & $a_G = 1$, $a_H = 0$ & $a_G = 1$ and $a_H = 1$ & $a_G = 0$) of the networks and 'overlapping' ($a_H = 1/2$ & $a_G = 1/2$). Then, the maximum eigenvalue becomes greater and the smallest eigenvalue becomes a little bit smaller.

A next step is, to examine what happens with the eigenvalues if we add another blending station. When the discussed network consists of two blending stations, then the configuration of this network is fixed, because we assume that a blending station does not have a length. An example of this network with two blending stations is drawn in Figure 3-8. Only the locations of the blending stations can vary and we use the parameters a and b for the different blending stations as can be seen in Figure 3-8.

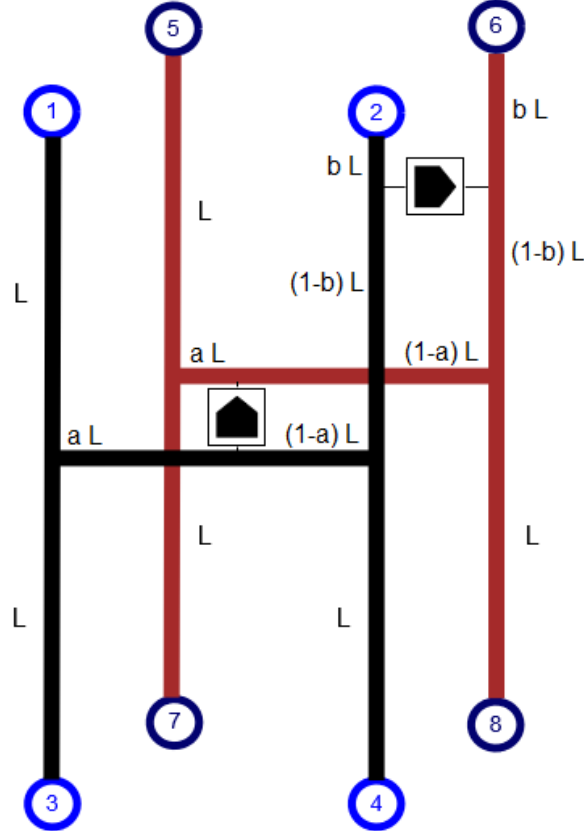


Figure 3-8: An expanded example network, in which G-gas is added, with two blending stations.

The distance matrix can be compiled in the same way as in the previous case. The distance submatrix \mathbf{D}_H remains the same, but the distance submatrix \mathbf{D}_{blend} changes, when a second blending station is added. Then we end up with the following \mathbf{D}_{blend} matrix

$$\mathbf{D}_{blend} = \begin{pmatrix} 2(1+a)L & 3L & 2(1+a)L & 3L \\ 3L & 2bL & 3L & 2L \\ 2(1+a)L & 3L & 2(1+a)L & 3L \\ 3L & 2L & 3L & \min\{2(2-b)L, 2(2-a)L\} \end{pmatrix}.$$

Again, we can vary the values of the parameters a and b and calculate the eigenvalues of the correlation matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$. A few results are summarized in Table 3-2.

Table 3-2: An overview of some choices for a and b and the resulting eigenvalues of the distance matrix.

Choices lengths		Eigenvalues \mathbf{A}							
a	b	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
1/2	1/2	0.3333	0.5880	0.6667	0.6667	1.0000	1.3333	1.3333	2.0787
2/3	1/2	0.3000	0.6000	0.6000	0.6066	0.8000	1.2544	1.4000	2.4389
3/4	1/2	0.2857	0.5714	0.5714	0.6088	0.7143	1.1719	1.4286	2.6479
4/5	1/2	0.2778	0.5556	0.5556	0.6092	0.6667	1.1210	1.4444	2.7698
1	1/2	0.2500	0.5000	0.5000	0.5000	0.6076	0.9281	1.5000	3.2143
0	0	0	0.3515	0.5000	0.5000	0.5000	1.0000	1.6066	3.5419
0	1	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	2.0000	2.0000
1	0	0	0.5000	0.5000	0.5000	0.6771	1.0000	1.5000	3.3229
1	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.8820	1.5000	3.1180
1/2	2/3	0.4444	0.6306	0.6667	0.6667	0.8889	1.3333	1.3333	2.0361
1/2	3/4	0.5000	0.6461	0.6667	0.6667	0.8333	1.3333	1.3333	2.0205
1/2	4/5	0.5333	0.6535	0.6667	0.6667	0.8000	1.3333	1.3333	2.0132
1/2	1	0.6667	0.6667	0.6667	0.6667	0.6667	1.3333	1.3333	2.0000
1/2	0	0	0.6667	0.6667	0.6667	1	1.3333	1.3333	2.3333

We see that the smallest determined eigenvalue equals 0 which occurs three times: when $a = 0$ & $b = 0$, $a = 1$ & $b = 0$ and $a = 1/2$ & $b = 0$. In addition, the first two combinations give the greatest determined eigenvalues, 3.5419 and 3.3229, respectively. We see that these combinations have something in common: $b = 0$ holds in all cases. This means that the added blending station is positioned between two entry/exit points.

Therefore, we are curious about what happens with respect to the smallest and greatest eigenvalues, when we have two blending stations positioned between entry/exit points as drawn in Figure 3-9.

We determine again the submatrix \mathbf{D}_{blend} which gives in this case

$$\mathbf{D}_{blend} = \begin{pmatrix} 0 & 3L & 2L & 3L \\ 3L & 0 & 3L & 2L \\ 2L & 3L & 4L & 5L \\ 3L & 2L & 5L & 4L \end{pmatrix}.$$

Calculating the eigenvalues with *Matlab* results in the following eight eigenvalues

$$\begin{aligned} \lambda_1 = 0, \quad \lambda_2 = 0, \quad \lambda_3 \approx 0.1119, \quad \lambda_4 \approx 0.4000, \\ \lambda_5 \approx 0.5528, \quad \lambda_6 \approx 1.2000, \quad \lambda_7 \approx 1.4472 \quad \text{and} \quad \lambda_8 \approx 4.2881. \end{aligned}$$

We see that there exists now two zero eigenvalues and the greatest eigenvalue equals 4.2881.

If we delete one of the blending stations, for example the right one (between points 2 and 6), then we end up with the following eight eigenvalues

$$\begin{aligned} \lambda_1 = 0, \quad \lambda_2 = 0.2505, \quad \lambda_3 \approx 0.3333, \quad \lambda_4 \approx 0.3333, \\ \lambda_5 \approx 0.4795, \quad \lambda_6 \approx 0.5655, \quad \lambda_7 \approx 1.8539 \quad \text{and} \quad \lambda_8 \approx 4.1841. \end{aligned}$$

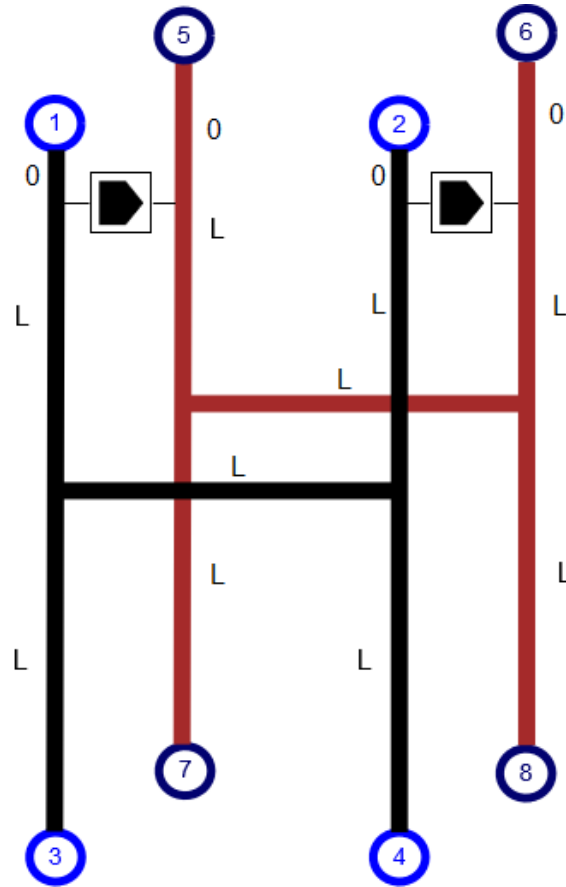


Figure 3-9: An expanded example network, in which G-gas is added, with two blending stations between entry/exit points.

What will happen, if the blending station is only connected to one entry/exit point? We consider the situation drawn in Figure 3-10 in order to answer this question. We determine the distance submatrix \mathbf{D}_{blend} as well:

$$\mathbf{D}_{blend} = L \cdot \begin{pmatrix} a & 3-a & 2-a & 3-a \\ 3+a & 6-a & 5-a & 6-a \\ 2+a & 5-a & 4-a & 5-a \\ 3+a & 6-a & 5-a & 6-a \end{pmatrix}.$$

This submatrix is only symmetric for $a = 0$, because $3 + a = 3 - a$ and $2 + a = 2 - a$ should hold. Therefore, it is not possible that a blending station is only connected to one entry/exit point for this H-shaped network.

The conducted experiments regarding the H-shaped network indicate that the smallest eigenvalue decreases the most, when one of more blending stations are positioned between entry/exit points.

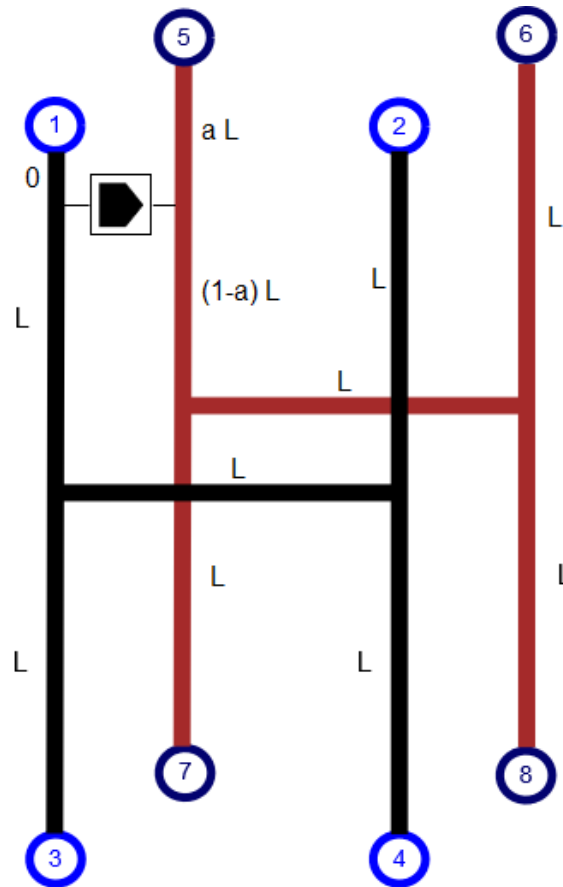


Figure 3-10: An expanded example network, in which G-gas is added, with one blending stations connected to one entry/exit point.

If we consider the network displayed in Figure 3-6, then we see that all blending stations are positioned between entry/exit points. So, this positioning could explain why we end up with three positive eigenvalues for the distance matrix \mathbf{D} , while the distance matrix \mathbf{D} of the network displayed in Figure 3-5 only has one positive eigenvalue.

So, it could be wise to conduct some experiments with the network displayed in Figure 3-6 in which the open blending stations are not positioned between entry/exit points, but between the pipelines. However, it is important that the distance matrix \mathbf{D} remains symmetric. It is shown in Figure 3-11 which blending stations are closed in the discussed network.

We have to move the open blending stations parallel to the pipelines such that the distance between the two pipelines equals zero, because it is assumed that a blending station has length zero. We start by moving one blending station, numbered as 1 in Figure 3-11. This blending station is for example moved down between two overlapping pipelines, one of length 24 km and one of 28 km. A next step is, to determine the eigenvalues of the new distance matrix \mathbf{D} . Still, we end up with three positive eigenvalues and thus three negative eigenvalues for the correlation matrix \mathbf{A} (checked with *Matlab*). We can move this blending station up as well, but that does not change the sign of the three positive eigenvalues of \mathbf{D} (or the three negative eigenvalues of \mathbf{A}).

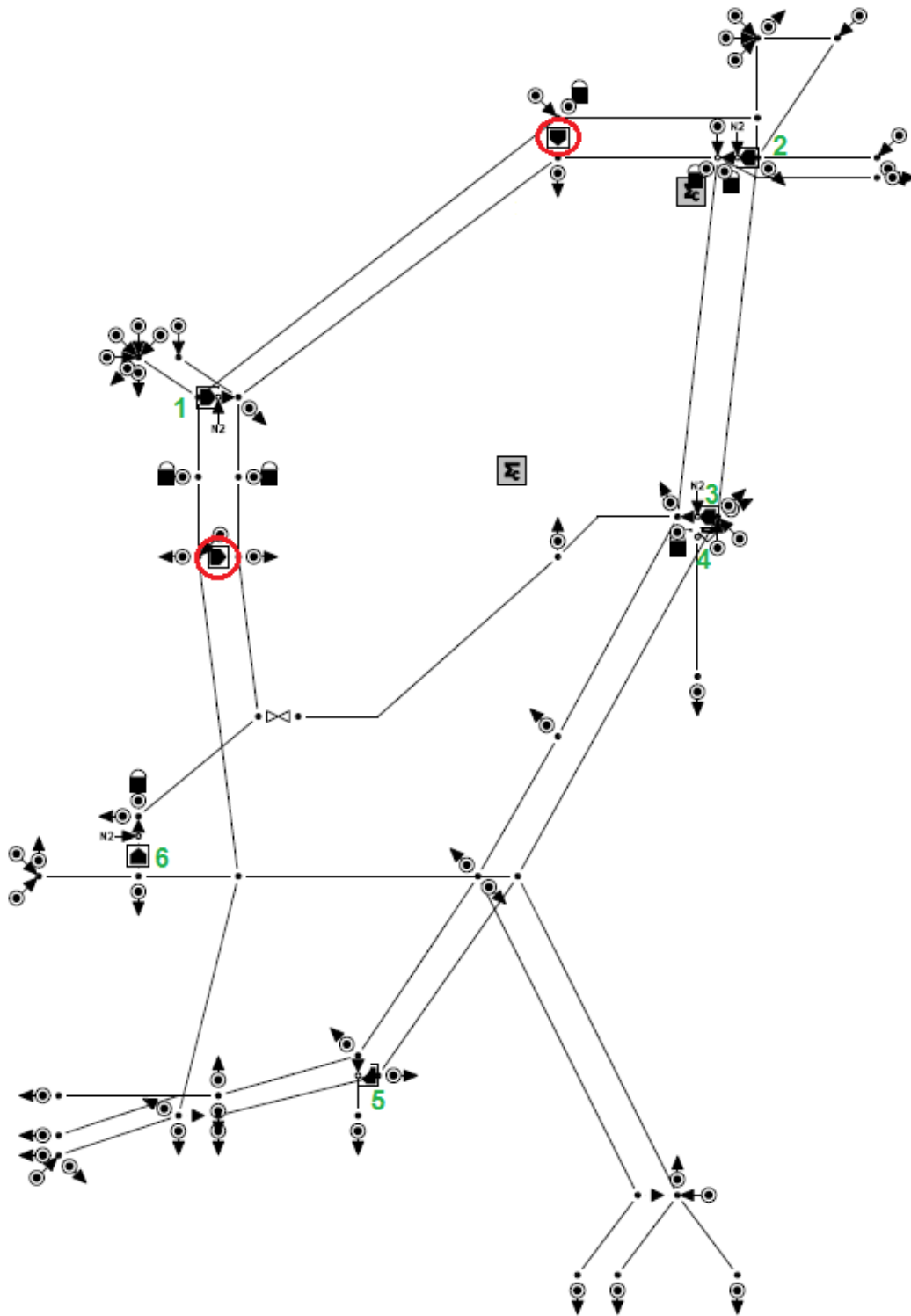


Figure 3-11: A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with an indication of closed blending stations.

In addition, we move blending stations 3 and 4 down as well. These blending stations are moved down parallel to a pipeline of length 70 *km* and one of 113 *km*. Again, a new distance matrix \mathbf{D} is compiled. However, the signs of the considered eigenvalues do not change. Then, we move blending station 5 parallel to two overlapping pipelines, one of length 55 *km* and one of length 65 *km*. Unfortunately, there is no improvement in the behavior of the considered eigenvalues.

The other open blending stations, numbered as 2 and 6 in Figure 3-11, cannot be moved parallel to two pipelines, because the G-gas and H-gas pipelines are not overlapping each other in the neighborhood of these blending stations.

So, it seems that moving the open blending stations does not give the desired result that the correlation matrix \mathbf{A} ‘becomes’ PSD on the subspace of balanced vectors ($\sum_i x_i = 0$).

We also wanted to investigate what happens with respect to the eigenvalues, if we delete some of the pipelines of the G-gas network which are intersecting/crossing a pipeline of the H-gas network. A part of the G-gas network is overlapping the H-gas network and we keep these overlapping pipelines. If we delete the G-gas pipelines which are crossing H-gas pipelines, then we need to delete four points as well. These four points are circled in red in Figure 3-12. It seems that more G-gas pipelines are crossing H-gas pipelines, but these G-gas pipelines can be repositioned such that they are overlapping H-gas pipelines or are not crossing H-gas pipelines.

We delete these four exit points by deleting the corresponding columns and rows in the distance matrix \mathbf{D} . Then we determine the 59×59 correlation matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ and calculate its eigenvalues. In this case, the correlation matrix \mathbf{A} has one negative eigenvalue, $\lambda_1 \approx -0.0533$, instead of three negative eigenvalues. We are interested in the right eigenvector corresponding to this negative eigenvalue and especially in the sum of the components of this corresponding eigenvector. We have determined this corresponding eigenvector with *Matlab* and added its components. The resulting sum is not equal to zero. So, this right eigenvector is not a balanced vector and therefore, we can say that it is not a well defined ‘difference vector’. We can conclude that the new correlation matrix \mathbf{A} is PSD on the subspace of balanced vectors.

Thus, deleting some G-gas pipelines which are crossing H-gas pipelines does give the desired result.

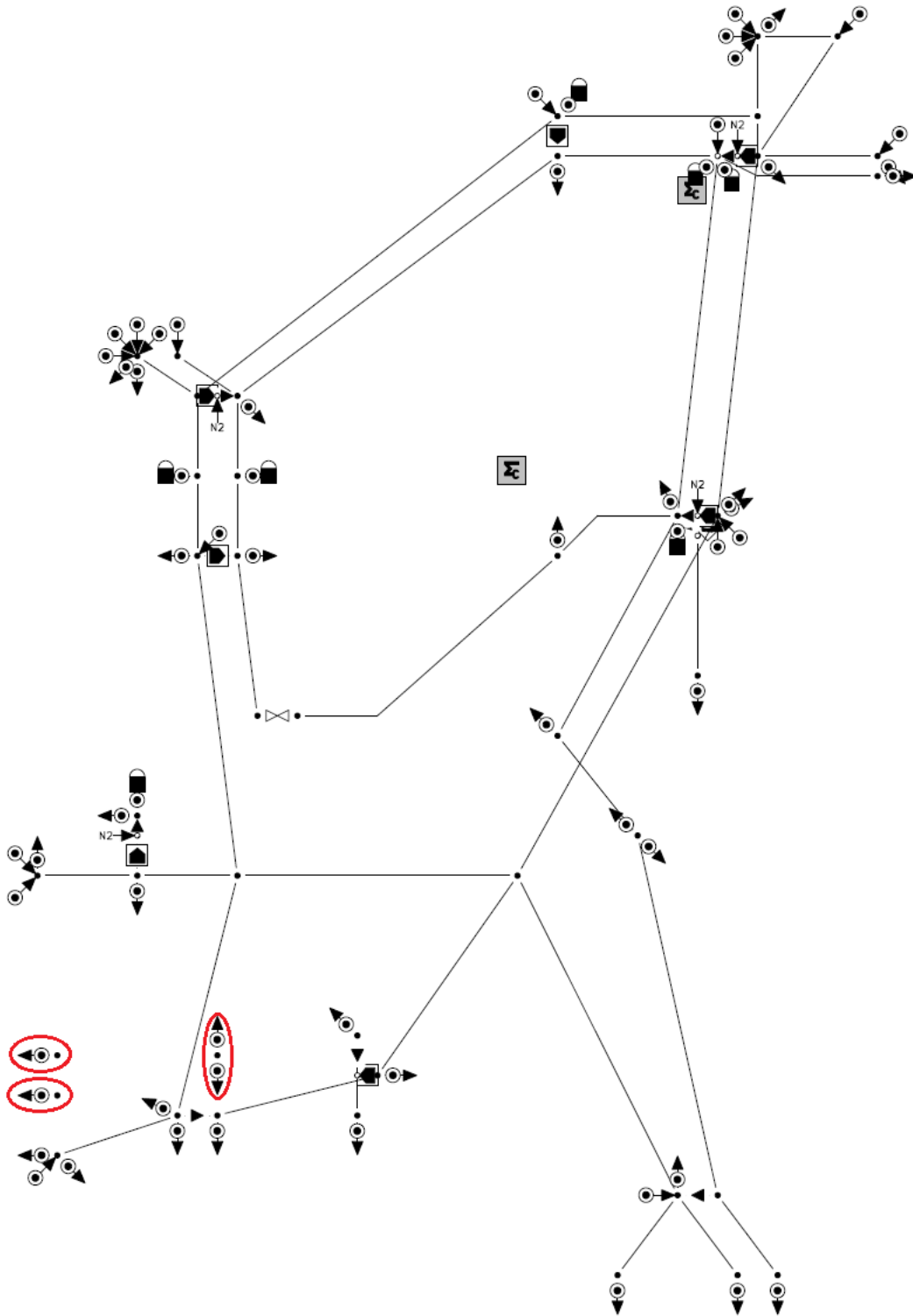


Figure 3-12: A simplification of the network of GTS, consisting of H-gas and G- and L-gas, in which some G-gas pipelines are deleted.

Definitions correlation matrix

This chapter deals with different choices for the correlation matrix \mathbf{A} and tries to motivate these choices. When choosing a definition for the correlation matrix, we want to take some of the transport physics into account, like the pressure drop. Therefore, the first section describes the pressure drop through a pipeline and offers some choices for the correlation matrix. Some of these choices are tested on the simplified and detailed transport networks of GTS in the next section.

4-1 Relation to the pressure drop

The pressure drop ΔP in a pipeline is defined as $P_{in} - P_{out}$ with P_{in} the pressure at the begin of the pipeline and P_{out} the pressure at the end of the pipeline. The pressure at the beginning and at the end of a pipeline are related according to the equation

$$P_{in}^2 - P_{out}^2 = \frac{c \cdot L}{D^5} \cdot Q^2. \quad (4-1)$$

It is known that gas flows through a pipeline from high to low pressure. Therefore, we can assume that the pressure at the beginning of a pipeline (P_{in}) is higher than the pressure at the end (P_{out}). Otherwise, gas cannot be flowing through this pipeline. However, if we want to take the direction of the flow into account, we can substitute Q^2 in Equation (4-1) by $Q|Q|$ [4, 16]:

$$P_{in}^2 - P_{out}^2 = \frac{c \cdot L}{D^5} \cdot Q \cdot |Q|. \quad (4-2)$$

In general, the pipelines of the HTL have a diameter of 1.187 m. If we assume the constant c to be a fixed constant and choose a fixed flow Q and pressure P_{in} at the begin of the pipe as well, then the pressure P_{out} at the end of the pipe depends only at the length L of the pipe (or transportation distance). In that case, the pressure drop ΔP only depends on the the distance as well. We vary the length of a pipe (transportation distance) in the program *MCA*¹

¹*MCA* is a simulation model of GTS which generates, for example, a set of shipping variants.

and let *MCA* determine the pressure at the end of this pipe in order to sketch the relation between the pressure drop and distance. Then this relation is plotted in Figures A-4, A-5, A-6, A-7 and A-8 for different choices of P_{in} , D and Q . We see the same relation between the pressure drop and the transportation distance in each plot.

The definition of the matrix \mathbf{A} according to the paper of Skopal et al defines a linear relation between the entries of this matrix (\mathbf{A}_{ij}) and the distance (d_{ij}) as can also be seen in Figure A-1. We have seen from the previous plots that the relation between the pressure drop and the transportation distance is not linear.

Therefore, we want to try some different definitions of the QFD matrix as a next step, for example

$$\mathbf{A}_{ij} = \exp \left\{ \left(\frac{d_{ij}}{d_{max}} \right)^2 - 1 \right\}. \quad (4-3)$$

The relation between the entry of the matrix \mathbf{A} and the distance is plotted in Figure A-2 for this choice. Another choice is the following one

$$\mathbf{A}_{ij} = 1 - \sqrt{1 - \frac{d_{ij}}{d_{max}}}, \quad (4-4)$$

and the relation between \mathbf{A}_{ij} and d_{ij} (for $d_{max} = 250$) is shown in Figure A-3.

We have to investigate whether or not the new defined matrices \mathbf{A} are SPSD as well. Both matrices are symmetric, because $d_{ij} = d_{ji}$ holds for all i and j . First we want to test both matrix definitions on the small network introduced in Subsection 3-1-4. Therefore, we consider again the two shipping variants from Subsection 3-1-4 represented by the two vectors $u = (120, -20, -100)$ and $v = (100, 0, -100)$.

The matrix according to Equation (4-3) can be compiled with $d_{max} = L$:

$$\mathbf{A} = \begin{pmatrix} e^{-1} & e^{-1} & 1 \\ e^{-1} & e^{-1} & 1 \\ 1 & 1 & e^{-1} \end{pmatrix}.$$

Calculating the QFD with the matrix \mathbf{A} defined as in Equation (4-3) gives the required result for this very small network: $QFD(u, v) = QFD(v, u) = 0$. A next step can be to determine the eigenvalues with, for example, the program *Matlab*. The eigenvalues of this matrix \mathbf{A} are

$$\lambda_1 \approx -0.8743, \quad \lambda_2 \approx 0 \quad \text{and} \quad \lambda_3 \approx 1.9779.$$

We can conclude that the matrix \mathbf{A} defined as in Equation (4-3) is not PSD, considering these eigenvalues.

Secondly, we determine the matrix \mathbf{A} according to the definition in Equation (4-4) with again $d_{max} = L$:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Again, we get that $QFD(u, v) = QFD(v, u) = 0$ for this choice of matrix \mathbf{A} , but we also want this matrix to be PSD. Thus, we calculate the eigenvalues:

$$\lambda_1 \approx -1.4142, \quad \lambda_2 = 0 \quad \text{and} \quad \lambda_3 \approx 1.4142,$$

and conclude that the matrix defined in Equation (4-4) is also not PSD.

It is still a question how we have to take into account the pressure drop. Do we have to use a similar relation between the entries of matrix \mathbf{A} and the distance as the relation between the pressure drop and the distance? The form of the curves drawn in Figures A-4 till A-8 have the same shape as the curve drawn in Figure A-3, but the corresponding matrix (defined in Equation (4-4)) does not satisfy all the wanted properties (being PSD).

We have also examined the following definitions for the matrix \mathbf{A}

$$\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}} \quad \text{or} \quad \mathbf{A}_{ij} = 1 - \left(\frac{d_{ij}}{d_{max}}\right)^2, \quad (4-5)$$

which give the following matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

for the small network mentioned in Subsection 3-1-4. Then using this matrix in the distance function QFD results again in $QFD(u, v) = QFD(v, u) = 0$ and the eigenvalues are already calculated in Subsection 3-1-4. The matrix for this very small network is indeed SPSD. We still have to investigate this matrix, because the matrix needs to be SPSD in general.

Another idea is to consider the relation between the pressure at the end of a pipeline P_{out} and the length of the pipeline L instead of the relation between the pressure drop and this length. We have plotted these relations as well, see Figures A-9 till A-13. The relations plotted are the same and correspond with the formula in Equation (4-2), because we can rewrite the pressure P_{out} as function of L :

$$P_{out} = \sqrt{P_{in}^2 - \frac{c \cdot L}{D^5} \cdot Q \cdot |Q|}.$$

Then it could be wise to use the following definition of the matrix \mathbf{A} for the distance function QFD:

$$\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}. \quad (4-6)$$

Note that we end up with the same matrix for the small network introduced in Subsection 3-1-4 as for the definitions in Equation (4-5). So, this definition satisfies the desired properties for the small network of Subsection 3-1-4.

4-1-1 Applying the new defined matrices on a small network

The distance function QFD with the new matrices defined in Equation (4-5) and (4-6) can be applied on the small network introduced in Section 2-1. The distance matrix for this small network equals

$$\mathbf{D} = [d_{ij}] = \begin{pmatrix} 0 & 3L & 2L & 3L \\ 3L & 0 & 3L & 2L \\ 2L & 3L & 0 & 3L \\ 3L & 2L & 3L & 0 \end{pmatrix},$$

and d_{max} is equal to $3L$. The first matrix, $\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}$, can be compiled:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 - \sqrt{\frac{2}{3}} & 0 \\ 0 & 1 & 0 & 1 - \sqrt{\frac{2}{3}} \\ 1 - \sqrt{\frac{2}{3}} & 0 & 1 & 0 \\ 0 & 1 - \sqrt{\frac{2}{3}} & 0 & 1 \end{pmatrix},$$

and then the distance between the three shipping variants can be calculated with the distance function QFD which results in

$$QFD_{\mathbf{A}}(u_1, u_2) = QFD_{\mathbf{A}}(u_2, u_1) \approx 217.58$$

$$QFD_{\mathbf{A}}(u_1, u_3) = QFD_{\mathbf{A}}(u_3, u_1) \approx 141.42$$

$$QFD_{\mathbf{A}}(u_2, u_3) = QFD_{\mathbf{A}}(u_3, u_2) \approx 141.42.$$

The second matrix, $\mathbf{A}_{ij} = 1 - \left(\frac{d_{ij}}{d_{max}}\right)^2$, can also be determined:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \frac{5}{9} & 0 \\ 0 & 1 & 0 & \frac{5}{9} \\ \frac{5}{9} & 0 & 1 & 0 \\ 0 & \frac{5}{9} & 0 & 1 \end{pmatrix},$$

and then applying the distance function QFD gives

$$QFD_{\mathbf{A}}(u_1, u_2) = QFD_{\mathbf{A}}(u_2, u_1) \approx 249.44$$

$$QFD_{\mathbf{A}}(u_1, u_3) = QFD_{\mathbf{A}}(u_3, u_1) \approx 141.42$$

$$QFD_{\mathbf{A}}(u_2, u_3) = QFD_{\mathbf{A}}(u_3, u_2) \approx 141.42.$$

Finally, the matrix according to the third definition, $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$, is equal to

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \sqrt{1 - \frac{2}{3}} & 0 \\ 0 & 1 & 0 & \sqrt{1 - \frac{2}{3}} \\ \sqrt{1 - \frac{2}{3}} & 0 & 1 & 0 \\ 0 & \sqrt{1 - \frac{2}{3}} & 0 & 1 \end{pmatrix},$$

and the distance between the three shipping variants are as follows (with the QFD)

$$QFD_{\mathbf{A}}(u_1, u_2) = QFD_{\mathbf{A}}(u_2, u_1) \approx 251.19$$

$$QFD_{\mathbf{A}}(u_1, u_3) = QFD_{\mathbf{A}}(u_3, u_1) \approx 141.42$$

$$QFD_{\mathbf{A}}(u_2, u_3) = QFD_{\mathbf{A}}(u_3, u_2) \approx 141.42.$$

We can check whether or not the eigenvalues of these three matrices are all greater than or equal to zero, besides calculating the distance with the help of the distance function QFD. This can be done with, for example, *Matlab* and indeed all the eigenvalues are positive. Thus, all three matrices are PD for this small network.

We see from these results that the distance between the first and third stress test and between the second and third stress test remains the same, and the distance between the first and second stress tests varies a little bit. However, the distance between the first and second stress test is still the greatest distance between the stress tests.

4-2 The quadratic form distance in MCA

The distance function QFD is already implemented in the program *MCA*. Thus, the QFD is calculated by *MCA* for each pair of stress tests after generating a set of stress tests for a network. Let us consider, for example, the network drawn in Figure 3-5.

A cluster level represents how much opposites the network points have in common within a cluster. When we set the cluster level on 7, we end up with eight different stress tests which Wobbe distribution can be found in the Appendix (Figures A-14 till A-21). So, if the cluster level is set on 7, then the network points with the same 7 opposites form a cluster. An opposite of a network point p is the furthest point in the network with respect to this network point p as stated in Subsection 2-2-1.

When these stress tests are determined, we see that the distance between the fourth and sixth stress test is the smallest one ($QFD \approx 698.07$) and between the first and seventh stress test the greatest one ($QFD \approx 9483.74$). Note that, if we take a look at the Wobbe distributions of the fourth and sixth stress test (Figures A-17 and A-19), then we see that the Wobbe distributions are pretty similar. It is still important to determine a threshold value to mark two stress tests as (almost) equal.

We can also set the cluster level equal to the number of network points, here 22. When we do this, then we end up with a set of 22 stress tests which consists of 9 different stress tests.

The distance matrix \mathbf{D} is compiled and saved as well, besides calculating the distance between the stress tests with the distance function QFD. Therefore, the matrix \mathbf{A} according to Equation (3-4) can be determined. Then we can calculate the eigenvalues of this 40×40 matrix with *Matlab* and we see that all eigenvalues are greater than or equal to zero. Thus, the matrix \mathbf{A} is for the network displayed in Figure 3-5 positive semi-definite.

The sparsity pattern of this matrix can be visualized besides calculating the eigenvalues with *Matlab*. This visualization can be found in Figure A-22.

The matrix \mathbf{A} defined in Equation (4-6) can also be applied on the network drawn in Figure 3-5 and again, the eigenvalues can be determined with *Matlab*. Then there exists three negative eigenvalues, $\lambda_1 \approx -0.6483$, $\lambda_2 \approx -0.1099$ and $\lambda_3 \approx -0.0456$. So, the matrix \mathbf{A} defined in Equation (4-6) is not PSD for this network.

A next step is to consider a larger network and choose different definitions for the QFD matrix \mathbf{A} and determine their eigenvalues. First, we choose to consider a simplified version of the GTS network which is already mentioned in Subsection 3-2-2 and displayed in Figure 3-6.

One of the outputs of *MCA* is the distance matrix \mathbf{D} for this network. Then we can compile the correlation matrix \mathbf{A} defined as in Equation (3-4). Thus, we can determine the eigenvalues of this 63×63 matrix with *Matlab* and see that the matrix has three negative eigenvalues

$$\lambda_1 \approx -0.1304, \quad \lambda_2 \approx -0.0446 \quad \text{and} \quad \lambda_3 \approx -0.0364.$$

We see that the original matrix is no longer positive semi-definite for an expanded network displayed in Figure 3-6, while the original matrix \mathbf{A} was SPSD for the network drawn in Figure 3-5.

We can determine the radius of the Gershgorin discs for this matrix as well (the center is always equal to one), besides determining the corresponding eigenvectors. Then, the maximum radius equals approximately 38.68 and the minimum radius approximately 17.46.

The eigenvalues of the matrix \mathbf{A} according to Equation (4-6) are determined as well, and there exists four negative eigenvalues in this case which are

$$\lambda_1 \approx -0.3531, \quad \lambda_2 \approx -0.3217, \quad \lambda_3 \approx -0.1404 \quad \text{and} \quad \lambda_4 \approx -0.0261.$$

The maximum radius of the Gershgorin discs for this matrix equals approximately 48.45 and the minimum radius is approximately 30.40. The sparsity pattern of these two matrices is visualized in Figure A-23 and we see that the two matrices only have two entries which are zero.

The maximum radius of 38.68 for the first matrix and of 48.45 for the second matrix are caused by rows 1, 14, 27, 42, 49 and 59 of the matrices. In other words,

$$\rho_1 = \rho_{14} = \rho_{27} = \rho_{42} = \rho_{49} = \rho_{59} = \begin{cases} 38.68 & \mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}} \\ 48.45 & \mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}} \end{cases}$$

The rows 1, 14, 27, 42, 49 and 59 are identical rows and thus, the corresponding rows of the distance matrix $\mathbf{D} = [d_{ij}]$ are identical as well. If we examine these rows of the distance matrix \mathbf{D} , then we see that the row consists of six elements which are zero (columns 1, 14, 27, 42, 49 and 59). So, there exists five elements in these rows which are equal to zero besides the diagonal element. Then these six rows in the matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ or $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$ have five off-diagonal elements which are equal to its diagonal element (center \mathbf{A}_{ii}). This means that it is not possible to have only Gershgorin discs with a radius smaller than or equal to \mathbf{A}_{ii} (all elements of \mathbf{A} are non-negative) for these two definitions of the matrix \mathbf{A} in this network.

Thus, we cannot adjust the two definitions for the matrix \mathbf{A} by using the Gershgorin circle theorem as a tool to ensure symmetric positive semi-definiteness.

If we take a closer look at the eigenvalues of both matrices \mathbf{A} , we see that

- the smallest eigenvalue of \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ equals approximately -0.1304 and the greatest eigenvalue approximately 34.8986 ;
- the smallest eigenvalue of \mathbf{A} defined as $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$ equals approximately -0.3531 and the greatest eigenvalue approximately 45.5706 .

So, we see that the negative eigenvalues are relatively small compared to the greatest eigenvalues.

It could be an alternative to consider the nearest correlation matrix which is the nearest SPSD matrix with ones on the diagonal. [17] Another possibility is to compute a nearest symmetric positive semidefinite matrix of these two matrices \mathbf{A} . [18]

If we choose one of these two matrices as QFD matrix for the network drawn in Figure 3-6, then the QFD is no longer a well defined distance function. This results in the fact that we cannot draw a conclusion about the comparison of two stress tests, when one of these matrices is used. This motivates us to take a look at different definitions of the matrix \mathbf{A} , for example the earlier mentioned definitions stated in Equation (4-5):

$$\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}} \quad \text{or} \quad \mathbf{A}_{ij} = 1 - \left(\frac{d_{ij}}{d_{max}}\right)^2.$$

Determining the eigenvalues of both matrices gives the following results:

- the matrix defined as $1 - \sqrt{\frac{d_{ij}}{d_{max}}}$ is SPSD, because it has no negative eigenvalues;
- the matrix defined as $1 - \left(\frac{d_{ij}}{d_{max}}\right)^2$ is not SPSD, because it has (ten) negative eigenvalues with $\lambda_1 \approx -1.4835$ the smallest one.

One matrix from the four examined matrices is SPSD and is defined as $\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}$. However, it is uncertain whether or not the relation between the pressure at the end of a pipeline P_{out} and the length of the pipeline L is taken into account. Initially, we want to approximate this relation by defining a matrix such that the same relation (curve) is visible between the entries of the matrix \mathbf{A}_{ij} and the distance d_{ij} , like the matrix defined in Equation (4-6). In that case, this matrix defined as $\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}$ is not a very good/logical choice. Therefore, some more possible definitions should be considered, for example

$$\mathbf{A}_{ij} = \exp\left\{-\frac{d_{ij}}{d_{max}}\right\}, \quad (4-7)$$

or

$$\mathbf{A}_{ij} = \frac{1}{1 + \frac{d_{ij}}{d_{max}}}. \quad (4-8)$$

The relation between the entries of the matrix and the distance of these two new matrices are plotted in Figure 4-1 together with entries $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$ and $\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}$.

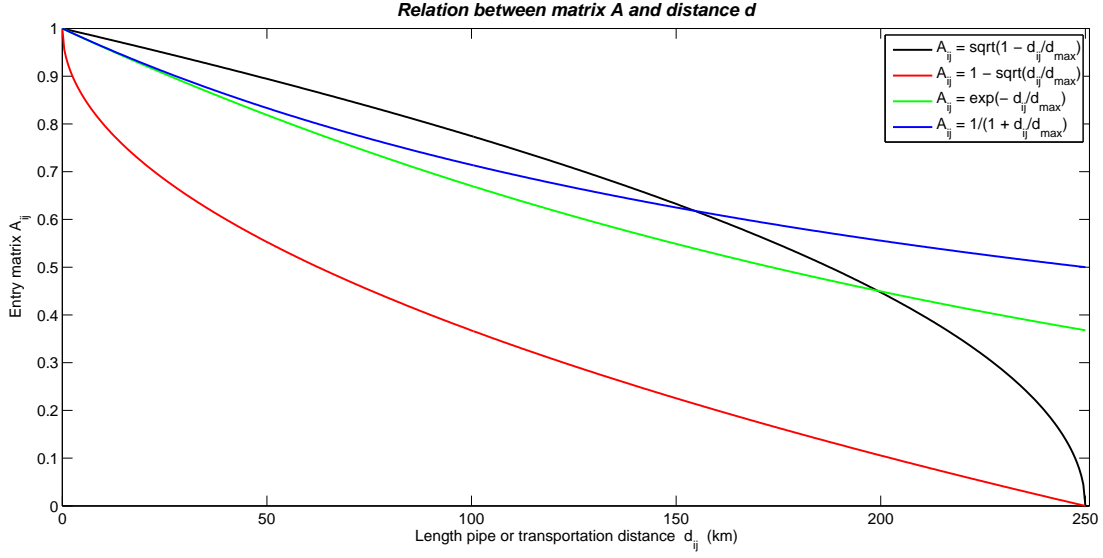


Figure 4-1: The relation between entry A_{ij} and transportation distance d_{ij} with $d_{max} = 250$.

We investigate whether or not these two new matrices are SPSD for the network shown in Figure 3-6, before we examine more possible matrices. All eigenvalues of both matrices are non-negative according to *Matlab* and thus, these matrices are SPSD. The sparsity pattern of these two matrices can be plotted as well. They have the same sparsity pattern and this pattern is visualized in Figure A-24.

The following three definitions for a QFD matrix can be used for the detailed network of GTS which can be found in Figure 4-2,

$$\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}, \quad \mathbf{A}_{ij} = \exp\left\{-\frac{d_{ij}}{d_{max}}\right\} \quad \text{and} \quad \mathbf{A}_{ij} = \frac{1}{1 + \frac{d_{ij}}{d_{max}}}.$$

If we determine the eigenvalues of these matrices for the detailed network of GTS, then we get some negative eigenvalues for the second and third definition:

for $\mathbf{A}_{ij} = \exp\left\{-\frac{d_{ij}}{d_{max}}\right\}$:

$$\lambda_1 \approx -0.0612, \quad \lambda_2 \approx -0.0329, \quad \lambda_3 \approx -0.0254, \quad \lambda_4 \approx -0.0096 \quad \text{and} \quad \lambda_5 \approx -0.0037.$$

for $\mathbf{A}_{ij} = \frac{1}{1 + \frac{d_{ij}}{d_{max}}}$:

$$\lambda_1 \approx -0.0354, \quad \lambda_2 \approx -0.0056, \quad \lambda_3 \approx -0.0033 \quad \text{and} \quad \lambda_4 \approx -2.5206 \cdot 10^{-4}.$$

So, the first defined matrix is positive semi-definite, but the last two defined matrices not.

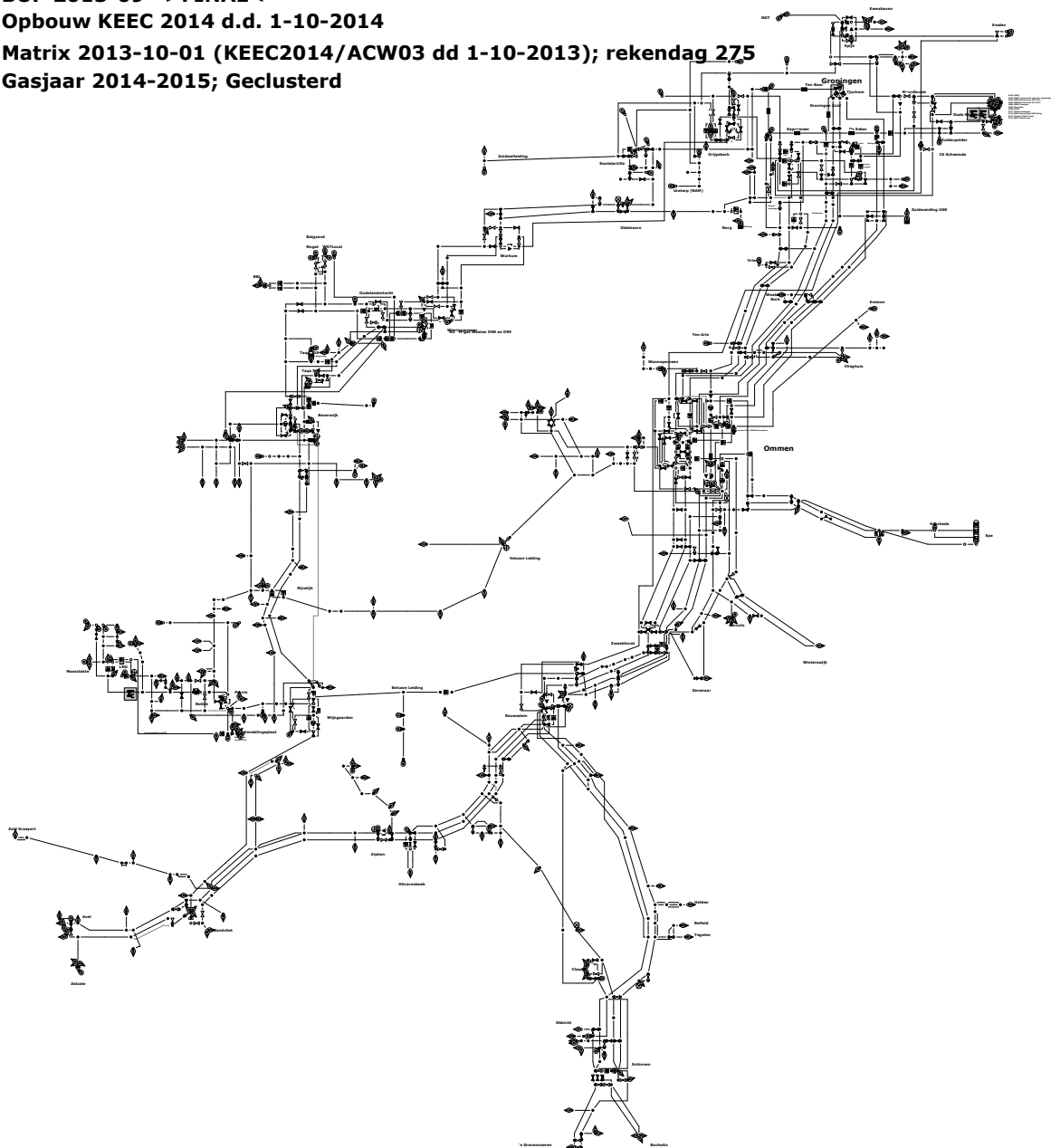
Hoofd Transport Leidingnet (HTL)

BUP 2013-09 >FINAL<

Opbouw KEEC 2014 d.d. 1-10-2014

Matrix 2013-10-01 (KEEC2014/ACW03 dd 1-10-2013); rekendag 275

Gasjaar 2014-2015; Geclusterd

Nieuwe Groningen modellering**Figure 4-2:** The network of GTS.

4-3 Some analyzes

When we are considering the network displayed in Figure 3-6, then some eigenvalues of the matrix defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ become negative. Therefore, some experiments are conducted regarding this matrix \mathbf{A} for the network displayed in Figure 3-6.

First, we remove the identical rows and columns in the matrix \mathbf{A} and then determine again the eigenvalues to see whether or not the smallest eigenvalues are still negative. This procedure corresponds to viewing the entry and exit points with mutual distance equal to zero as one point. In other words, only the network points are considered and not the entries and exits (entries and exits are linked to network points). Secondly, the length of the pipelines are rounded off. This rounding is repeated for different rounding numbers till the matrix \mathbf{A} has no longer negative eigenvalues. Finally, a different ordering of the rows and columns is considered which is called the Reverse Cuthill McKee ordering.

4-3-1 Deleting identical rows and columns

When we delete the identical rows and columns of the matrix $\mathbf{D} = [d_{ij}]$ with the following commands in *Matlab*

```
% Locating all zero elements of the distance matrix D_mat
[row, col] = find(~D_mat);
% Save the indices of all zero elements of the distance matrix Dmat_cl
D_zeros = zeros(length(row),2);
for i = 1 : length(row)
    D_zeros(i,1) = col(i);
    D_zeros(i,2) = row(i);
end
% The rows and columns, which have to be deleted, are numbered by hand
RowsColsToDelete = [63; 61; 59; 58; 57; 52; 51; 50; 49; 48; 47; 46; 44; 43; ...
    42; 41; 39; 38; 37; 36; 34; 32; 31; 30; 29; 27; 23; 22; 17; 14];
% The relevant rows and columns are deleted
for j = 1 : length(RowsColsToDelete)
    D_mat(RowsColsToDelete(j),:) = [];
    D_mat(:,RowsColsToDelete(j)) = [];
end
% The new matrix A1 is defined and the eigenvalues are calculated
A1 = 1 - D_mat/d_max;
Eig1 = eig(A1);
```

then the smallest eigenvalue equals approximately -0.0797 . We see that we still have negative eigenvalues and therefore we want to experiment with the lengths of the pipelines by rounding them off to the nearest defined multiple.

4-3-2 Rounding off lengths of pipelines

We want to examine what happens with the eigenvalues, when we are rounding off the length of the pipes. The results of this investigation are summarized in this section. We want to see whether or not the eigenvalues become non-negative during this experiment. Some pipeline lengths become zero, because of the rounding off and thus are removed from the network. Perhaps, a different network configuration results in a QFD matrix \mathbf{A} which is SPSD.

We start with the rounding of the length of the pipelines to nearest multiple of 10, then to the nearest multiple of 20, 25, 50 and 100. Then again the identical rows and columns of the matrix $\mathbf{D} = [d_{ij}]$ are deleted, the matrix \mathbf{A} is computed with the new distance matrix \mathbf{D} and the eigenvalues are calculated. We still have some negative eigenvalues in these five cases. Therefore, we round the lengths off to the nearest multiple of 110 and 115 as well. Then, there exists some negative eigenvalues in the first case (multiple of 110), but no negative eigenvalues in the second case (multiple of 115). So, the turning point (from negative eigenvalues to non-negative eigenvalues) is between 110 and 115. We examine the rounded lengths in the case of 110 and 115 and we see that two pipelines deviate in length. The original length of these pipelines are 55 and 56 km, respectively. These lengths are both rounded off to 110 in the case of 110 and to 0 in the case of 115. We are interested in the case in which one of the lengths is rounded off to zero and the other one to a number greater than zero. Therefore, we consider multiples of 112. Then a length of 55 is rounded off to zero and a length of 56 to 112. Then the smallest eigenvalue is approximately $-2.1351 \cdot 10^{-16} \approx 0$. The following steps are computed in *Matlab* to determine which rows and columns need to be deleted. Then they are deleted, the matrix \mathbf{A} is computed and the eigenvalues are determined.

```
% Locating all zero elements of the distance matrix Dmat_cl
[row_cl, col_cl] = find(~Dmat_cl);

% Save the indices of all zero elements of the distance matrix Dmat_cl
Dcl_zeros = zeros(length(row_cl),2);
for i = 1 : length(row_cl)
    Dcl_zeros(i,1) = col_cl(i);
    Dcl_zeros(i,2) = row_cl(i);
end

% The rows and columns, which have to be deleted, are numbered by an algorithm
Indices_cl = sparse(n,n);
for j = 0 : n-1
    Ind_cl = find(col_cl == n-j);
    if length(Ind_cl) > 1
        if row_cl(Ind_cl(1)) >= n-j
            for k = 1 : length(Ind_cl)-1
                Indices_cl(n-j,k) = row_cl(Ind_cl(1) + (length(Ind_cl) - k));
            end
        end
    end
end
```

```

        end
    else
        Indices_cl(n-j,:) = zeros(n,1);
    end
end
end

% Want to delete the rows and columns in descending order, otherwise the numbering of the rows
and columns changes
VectorIndices_cl = Indices_cl(:);
Sorting_cl = sort(VectorIndices_cl, 'descend');
Size_sortCl = find(Sorting_cl, 1, 'last');
IndDescCl = Sorting_cl(1:Size_sortCl);
% The relevant rows and columns are deleted
for j = 1 : Size_sortCl
    Dmat_cl(IndDescCl(j),:) = [];
    Dmat_cl(:,IndDescCl(j)) = [];
end
% The new matrix A1_cl is defined and the eigenvalues are calculated
A1_cl = 1 - Dmat_cl/d_maxcl;
Eig1_cl = eig(A1_cl);

```

So, we ended up with a network for which the QFD matrix \mathbf{A} is SPSD, but we need to figure out whether or not this experiment gives an indication for the matrix being SPSD. Does this remaining network give some information about when the defined QFD matrix \mathbf{A} is SPSD and when not?

4-3-3 Reverse Cuthill-McKee ordering

Another tool to investigate the ‘behavior’ of the eigenvalues was to change the ordering of the rows and columns of the QFD matrix \mathbf{A} and then determining again the eigenvalues. A different ordering should not have any effect on the eigenvalues (without any rounding errors), but it will probably reduce the influence of rounding errors.

An example of getting a different ordering is to apply the reverse Cuthill-McKee ordering. We can use the command $r = \text{symrcm}(\mathbf{A})$ in *Matlab* which returns the symmetric reverse Cuthill-McKee ordering of \mathbf{A} . Here, r is a permutation such that $\mathbf{A}(r,r)$ tends to have its nonzero elements closer to the diagonal. [19] We have noticed that the eigenvalues do not change very much, when we apply this ordering for the matrix \mathbf{A} . We only see a change in the 15th decimal (or greater, or sometimes in the 14th decimal). So, this slight change indicates the influence of the used rounding errors.

Therefore, the slight change in eigenvalues is a cause of the working precision of *Matlab*. This statement can be proved, because the following relation holds

$$\sigma(\mathbf{A}) = \sigma(\mathbf{P}^T \mathbf{A} \mathbf{P})$$

for \mathbf{P} a permutation matrix and with $\sigma(\mathbf{A})$ the spectrum of the matrix \mathbf{A} . In general, permutation matrices are orthogonal, which means that $\mathbf{P}^T \mathbf{P} = \mathbf{P} \mathbf{P}^T = \mathbf{I}$ holds for the permutation matrix \mathbf{P} . Therefore, it holds that $\mathbf{P}^T = \mathbf{P}^{-1}$ and thus we can write $\mathbf{P}^T \mathbf{A} \mathbf{P}$ as $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$. If we call this matrix \mathbf{B} ($= \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$), then we can say that the matrices \mathbf{A} and \mathbf{B} are similar, because \mathbf{P} is square and nonsingular. So, it indeed holds that $\sigma(\mathbf{A}) = \sigma(\mathbf{B})$, because of this similarity (between \mathbf{A} and \mathbf{B}).[20]

We have also applied this ordering on the distance matrix $\mathbf{D} = [d_{ij}]$ before determining the matrix \mathbf{A} . Again, there is a slightly change in eigenvalues (14th decimal or greater). We have noted that some identical rows and columns of the matrix \mathbf{A} are placed together due to this ordering, if the identical rows and columns are not deleted.

In addition, a contour plot is made for the original matrix and for the matrix which ordering is changed according to the reverse Cuthill-McKee ordering and can be found in the Appendix (Figures A-25 and A-26).

4-3-4 Dependence on the diameter of the pipelines as well

We can adjust the QFD matrix \mathbf{A} defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ by taking another property of the network into account, for example the diameter of the pipelines besides clustering some entries and exits or changing the ordering of the used matrices. We start with the diameter, because both the diameter and length of the pipelines are constants and do not change during a (or per) gas transport situation. The pressure at the beginning and at the end of a pipeline are related as followed

$$P_{in}^2 - P_{out}^2 = \frac{c \cdot L}{D^5} \cdot Q^2,$$

as mentioned in Section 4-1 (see Equation (4-1)). We have derived the relation between the pressure P_{out} and length of the pipeline L from this equation:

$$P_{out} = \sqrt{P_{in}^2 - \frac{c \cdot L}{D^5} \cdot Q^2},$$

which can be rewritten as follows

$$P_{out} = P_{in} \cdot \sqrt{1 - \frac{c \cdot Q^2}{P_{in}^2} \cdot \frac{L}{D^5}}.$$

So, we first want to consider

$$\mathbf{A}_{ij} = 1 - \frac{r_{ij}}{r_{max}} \quad \text{with} \quad r_{ij} = \sum_{k=i}^{j-1} r_{k,k+1} = \sum_{k=i}^{j-1} \frac{L_{k,k+1}}{D_{k,k+1}^5}, \quad r_{max} = \max_{i,j} r_{ij}, \quad (4-9)$$

with $L_{k,k+1}$ the length and $D_{k,k+1}$ the diameter of the pipeline between network point k and $k + 1$ as the new definition of the matrix \mathbf{A} .

First, the new matrix according to Equation (4-9) is computed for the network displayed in Figure 3-5. The length and diameter of the pipelines can be extracted from the program *MCA*. The length of all pipelines is given and almost every diameter, except for two pipelines which transport L-gas from *Beekse Bergen* to *Hilvarenbeek* and from *Ommen* to *Winterwijk-s/Zevenaer*. These two diameters are set equal to 1.5 m in order to compute the matrix \mathbf{A} for this network. *MCA* also gives an overview of the pipelines and between which two network points (names) it is placed. So, we only have to transform the names of these network points to numbers to generate a graph in *Matlab* with $\frac{L}{D^5}$ as weights on the edges. This graph is drawn in Figure 4-3. We see from this figure that a few network points are not connected to the rest of the network points. This is caused by the fact that there are no pipelines between P-OMML and P-OMMH and between P-BBRL and P-BBRH, but a checkvalve is placed between these two (pair of) points. In other words, the distance between network points P-OMML and P-OMMH is zero and between network points P-BBRL and P-BBRH as well. Fixing these network points results in the graph drawn in Figure 4-4.

The next step is, to calculate all shortest paths between every pair of network points. This is done in *Matlab* with the function `graphallshortestpaths(G)` in which matrix \mathbf{G} represents the graph. An element of the matrix \mathbf{G} , G_{ij} , equals the weight of the edge between i and j . When an element of this matrix equals zero, then there exists no edge between these points. This function gives a symmetric matrix as output and element (i, j) of this matrix represents the shortest path from point i to j . The maximum shortest path is the path from P-ZANDH (*Zandvliet*) to P-ZELZ (*Zelzate*).

We have to delete the rows and columns corresponding to the points P-WIJN and P-RAVH in order to create a ‘distance’ matrix $\mathbf{R} = [r_{ij}]$ for the entries and exits, because both points have no entries or exits. Then some rows and columns need to be duplicated to get the full ‘distance’ matrix \mathbf{R} , because some network points consist of more than one entry and/or exit. The new matrix \mathbf{A} according to Equation (4-9) can be determined after doing this duplications and the eigenvalues can be computed. Then we see that the eigenvalues are complex, but with a very small imaginary part (approximately zero). Therefore, we assume that all eigenvalues are real and the smallest eigenvalue is then approximately equal to $-1.3802 \cdot 10^{-29}$ which can be rounded off to zero because of the working precision of *Matlab*.

We can also determine the eigenvalues of the square root of this matrix ($\mathbf{A}_{ij} = \sqrt{1 - \frac{r_{ij}}{r_{max}}}$) and then the smallest eigenvalue is approximately equal to -0.1841 . Unfortunately, there exists still some negative eigenvalues in this case.

Secondly, the new matrix \mathbf{A} is computed for the network showed in Figure 3-6 as well. Again, the length and diameter of the pipelines can be extracted from the program *MCA*, but more diameters are unknown (L-gas and G-gas pipelines). The diameter of these pipelines are also set to 1.5 m in order to conduct the experiment for this network. The same principles as for the previous network are applied here. However, this network, consisting of pipelines transporting G-gas, H-gas and L-gas, is more complicated than the previous one. This network consists of blending stations and storages as well. A storage can be seen as both an entry and exit point and does not give any problems, but we need to mind the blending stations. When a blending station is placed between two network points, then the distance between these two network points is not equal to zero. However, when a blending station and a check valve is placed between two network points, then this distance does equal zero. If we take all the ‘zero

distances' between a pair of network points into account, then the graph displayed in Figure 4-5 represents the connection in the network.

We need to delete some of the network points in this graph, because there are no storages, entries or exits linked to these network points and the 'distance' matrix $\mathbf{R} = [r_{ij}]$ represents the 'distance' between the available storages, entries or/and exits. Again, some points need to be duplicated, because more storages, entries or exits are linked to such a network point. This means that some of the rows and columns of the generated 'shortest path' matrix (of the graph) needs to be duplicated. Then the matrix \mathbf{A} defined as in Equation (4-9) can be calculated and the corresponding eigenvalues can be determined. We see that also these eigenvalues have a very small imaginary part (approximately zero), and therefore we can assume that the computed eigenvalues are real (working precision of *Matlab*). The smallest eigenvalue is approximately equal to $-1.1511 \cdot 10^{-22}$ which can be assumed to be zero due to the precision of *Matlab*.

The smallest eigenvalue of the matrix defined as the square root ($\mathbf{A}_{ij} = \sqrt{1 - \frac{r_{ij}}{r_{max}}}$) equals approximately -0.0853 .

We have also based the diameter of the pipelines of G-gas and L-gas at a later stage on the diameter of the pipelines in another simulation network. This simulation network can be found in the Appendix (Figure A-27). A few of the pipelines are closed (for example the vertical pipelines at the right outside) and is thus similar to the network displayed in Figure 3-6. So, the pipelines with no known diameter get the same diameter as the pipelines in this new network instead of setting them equal to 1.5 m . Most pipelines have the same length as the pipelines in this new network, except for one pipeline. This pipeline is a pipe of length 40 km in the network displayed in Figure 3-6, but is a combination of two pipelines, one of length 15 km with a diameter of 1.85 m and one of length 25 km with a diameter of 1.57 m in the network displayed in Figure A-27. Then the mean diameter is determined for this pipeline as follows: $(15 \cdot 1.85 + 25 \cdot 1.57)/40 = 1.675$.

The eigenvalues are determined in the same way as the previous case. Then we see that the smallest eigenvalue of the 'original' correlation matrix ($\mathbf{A}_{ij} = 1 - \frac{r_{ij}}{r_{max}}$) approximately equals $-2.2584 \cdot 10^{-16}$ which is approximately zero, and of the square root matrix ($\mathbf{A}_{ij} = \sqrt{1 - \frac{r_{ij}}{r_{max}}}$) approximately equals -0.0916 .

We can conclude that the new definition of the QFD matrix \mathbf{A} introduced in this subsection (Equation (4-9)) leads to an improvement of the eigenvalues of the matrix, especially for the second network. The eigenvalues of the matrix representing the first network were all non-negative in an early experiment summarized in Section 4-2, but not for the second network. So, when only the length of the pipelines was considered, the matrix was no longer SPSD for the second network. However, when the diameter of the pipelines are taken into account as well, then we see that the matrix 'remains' SPSD.

When we take the diameter into account as well, then the ratios between the matrix components change because of the new definition depending on $r_{ij} = \sum_{k=i}^{j-1} r_{k,k+1} = \sum_{k=i}^{j-1} \frac{L_{k,k+1}}{D_{k,k+1}^5}$. This could explain why the correlation matrix is SPSD with diameters, but not SPSD without diameters.

This motivates us to consider the original definition of the correlation matrix defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ again and experiment with the value of the denominator, here d_{max} . We want to investigate whether or not the sign of the negative eigenvalues of the correlation matrix changes, when we increase the denominator. The denominator is originally defined as $\max_{i,j} d_{ij}$ which equals 485 in this case.

We start with three negative eigenvalues. When we increases d_{max} , then we see that the sign of one negative eigenvalue changes for $d_{max} = 609$. This means that the new defined correlation matrix with d_{max} has two negative eigenvalues instead of three negative eigenvalues (for $d_{max} \leq 608$).

If we experiment some more with this denominator, then we notice that the remaining two eigenvalues does not easily change sign. If we choose the denominator, for example, equal to 100000, then these two eigenvalues are still negative.

These experiments strengthens our assumption that the negative eigenvalues become non-negative due to the ratio change between the matrix components \mathbf{A}_{ij} .

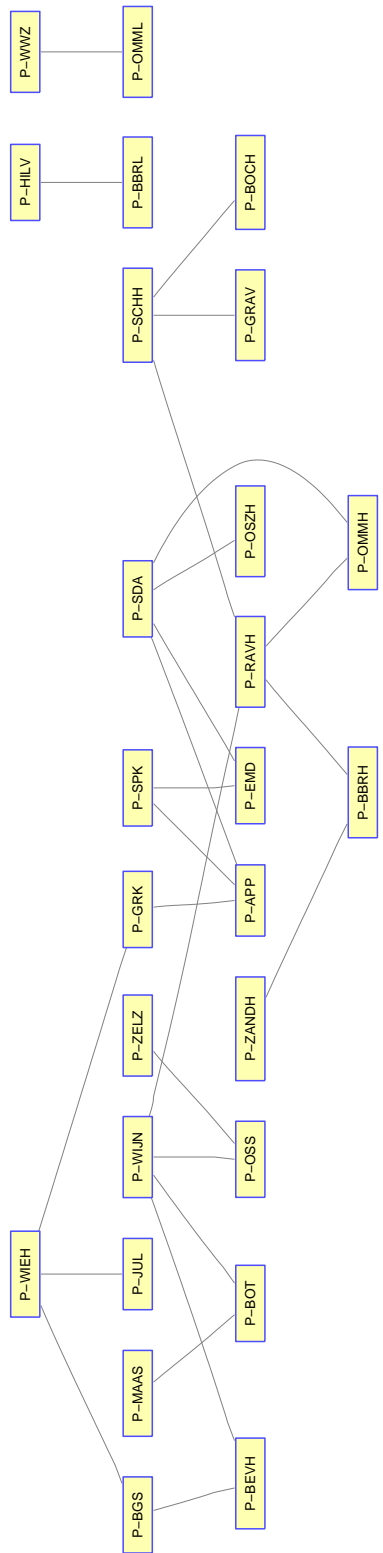


Figure 4-3: A graph of the network points in the considered network.

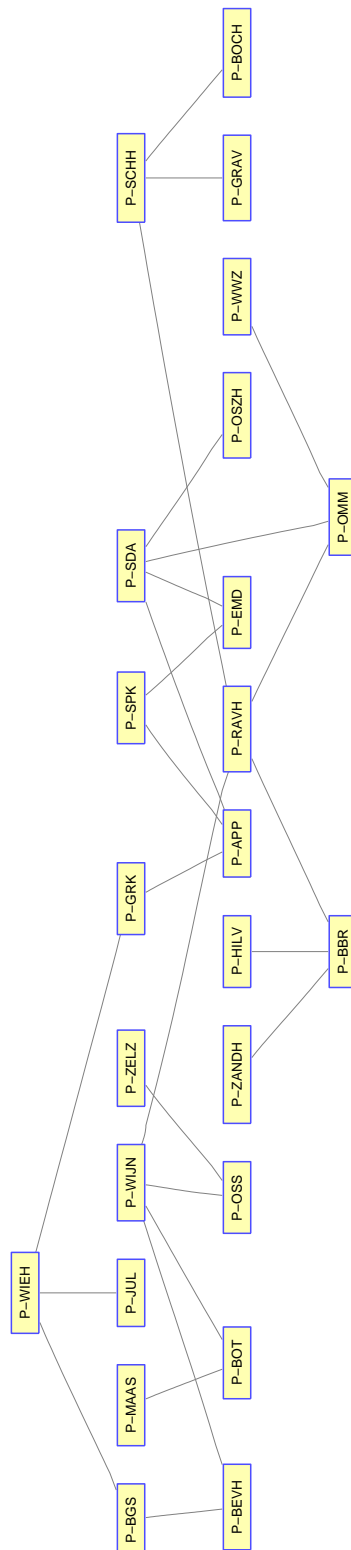


Figure 4-4: A graph of the network points in the considered network, such that there exists a path between every pair of points.

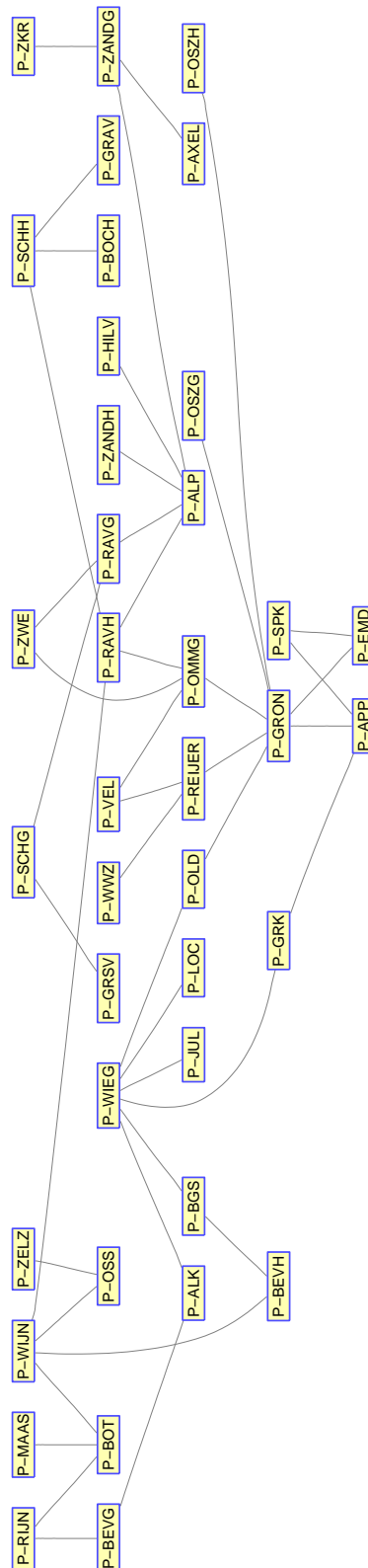


Figure 4-5: A graph of the network points in the considered network, such that there exists a path between every pair of points.

4-4 Emphasizing the H-gas transport network

In general, the differences in stress tests regarding G-gas and L-gas are not that important with respect to the differences regarding H-gas. Therefore, we want to emphasize the differences in the H-gas transport network and to minimize the influence of the differences in the G- and L-gas transport network.

The vectors representing the stress tests of the gas transport network can be ordered as follows

$$\underbrace{(x_1, x_2, \dots, x_n)}_{\text{H-gas}}, \underbrace{(x_{n+1}, \dots, x_m)}_{\text{G-}\&\text{L-gas}} \quad m \geq n.$$

A next step is to create a new vector representing the same stress test by adding the capacities of the entries and exits in the G- and L-gas transport network, $x_{GL} = \sum_{i=n+1}^m x_i$. The new vector is then denoted as

$$(x_1, x_2, \dots, x_n, x_{GL}).$$

In addition, we need to reformulate the corresponding QFD matrix \mathbf{A} . This matrix is also rearranged, just like the vectors representing the stress tests. We want to decouple the H-gas transport network and the G- and L-gas transport network besides rearranging the matrix. Then, the new matrix equals

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

and the zeros in the $(n + 1)$ th column and row represent the decoupling of the two gas transport networks.

We could also choose to not create a new vector, but only adjusting the matrix as follows

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & a_{n+1,n+1} & \cdots & a_{n+1,m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{m,n+1} & \cdots & a_{mm} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_H & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{GL} \end{pmatrix},$$

with \mathbf{A}_H the matrix representing the QFD matrix of the H-gas transport network and \mathbf{A}_{GL} of the G- and L-gas transport network.

The question is now: does this new representation give the desired result (a distinction between the H-gas transport network and G- and L-gas transport network with more emphasis on the differences in the H-gas transport network)?

We apply both approaches on the network displayed in Figure 3-6 which consists of H-gas and G- and L-gas, in order to answer the question stated above. First, we have generated a set of 66 stress tests for this network with the help of *MCA* in which the cluster level is set equal to 37. Secondly, we adjust the vectors representing the stress tests such that

- the vectors have the form $(x_1, x_2, \dots, x_n, x_{GL})$ for approach 1;
- the vectors are ordered as $(x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_m)$ with $x_i, i = 1, \dots, n$ the capacities in the H-gas network and $x_j, j = n + 1, \dots, m$ in the G- and L-gas network for approach 2.

In addition, the distance matrix $\mathbf{D} = [d_{ij}]$ is adjusted for

- approach 1 as
$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} & d_{max} \\ d_{21} & d_{22} & \cdots & d_{2n} & d_{max} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} & d_{max} \\ d_{max} & d_{max} & \cdots & d_{max} & 0 \end{pmatrix};$$
- approach 2 as
$$\begin{pmatrix} \mathbf{D}_H & \mathbf{D}_{max} \\ \mathbf{D}_{max} & \mathbf{D}_{GL} \end{pmatrix}$$
 with \mathbf{D}_{max} an $n \times (m - n - 1)$ matrix with d_{max} as elements.

Then, the QFD matrix \mathbf{A} is determined according to $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$ for each case and the quadratic form distance between every pair of vectors (stress tests) is calculated. A few results are summarized in Table 4-1. The corresponding difference vectors can be found in Table 4-2. Not all vector elements are written down in this table, because the vectors consist of 63 elements (second and original approach) or of 39 elements (first approach). However, the network points which are not mentioned in this table all have capacity zero. The names of the network points which are denoted in Table 4-2 are also displayed in the considered transport network, see Figure 4-6.

Table 4-1: The QFD, with different approaches, applied on some of the stress tests.

Quadratic form distance			
Applied on stress tests	Approach 1	Approach 2	Original
2 & 9	0.0100	282.3010	282.3010
2 & 35	0.0447	267.9711	267.9859
9 & 35	0.0500	205.3619	205.3858
8 & 53	0.0424	197.9043	197.9124
16 & 19	946.2644	946.2644	612.1948
12 & 45	504.7075	455.6138	295.5835
14 & 47	509.7577	457.7133	291.0260

We see that there is a difference between the two introduced approaches from the first four results showed in Table 4-1: the quadratic form distance is very small when the first approach is applied, but for the second approach the distances are almost equal to the QFD with original matrix. If we consider the difference of the second and ninth stress test, then we see that there are no differences in the H-gas network. However, there are some differences in the G-gas network for the exit *04-GGAS* and storage *B-NORG*. If we really want to emphasize

the difference occurring in the H-gas transport network and neglecting the differences in the G- and L-gas transport network, then we would prefer approach 1.

We see that the (quadratic form) distance between stress test 2 and 35 and 9 and 35 are very small for approach 1 and $QFD(9, 35) > QFD(2, 35)$, while $QFD(9, 35) < QFD(2, 35)$ when approach 2 or the original one is used. We see that both differences (2-35 and 9-35) result in a difference of approximately 0.04 on the exit *18-HGAS* in the H-gas network, but there are some differences in the G-gas network for *04-GGAS* (42.17 with respect to -198.02) and *B-NORG* (-358.57 with respect to -118.39). So, can we conclude which condition must hold, $QFD(9, 35) > QFD(2, 35)$ or $QFD(9, 35) < QFD(2, 35)$? Therefore, we take a look at the differences for x_{GL} . The x_{GL} of stress test 2 minus the x_{GL} of stress test 35 equals approximately -0.02 and for 9-35 approximately -0.03 . So, the overall difference in the G- and L-gas network is smaller for 2-35 than for 9-35. Thus, we observe that the condition $QFD(9, 35) > QFD(2, 35)$ must hold and that approach 1 is again preferable. In addition, it seems right that the QFD between 8 and 53 is smaller than the QFD between 2 and 35, seeing that the difference in the H-gas network is smaller.

The distances according to the second approach are in the last three results closer to the distances for the first approach than to the distances for the original approach. It seems right that $QFD(16, 19) > QFD(14, 47) > QFD(12, 45)$ considering the difference vectors in Table 4-2. So, in this case both approaches (1 and 2) can be chosen, but approach 1 is more favorable in general.

We can conclude from what is described above that approach 1 gives a distinction between the H-gas network and G- and L-gas network with more emphasis on the differences in the H-gas network.

Table 4-2: Some difference vectors of two different stress tests.

Difference stress tests	Network points H-gas (most nonzero)					Network points G-gas and L-gas (the most are nonzero)																
	IS-HGAS	B-GRIPPS	N-EMDEN	N-OSZH		01-GGAS	02-GGAS	03-GGAS	04-GGAS	05-GGAS	06-GGAS	07-GGAS	08-GGAS	09-GGAS	10-GGAS	11-GGAS	12-GGAS	13-GGAS	B-NORG	X-TEGEL	X-WWZ	
2 - 9	0	0	0	0	0	0	0	0	240.19	0	0	0	0	0	0	0	0	0	0	0	0	0
2 - 35	0.04	0	0	0	0	19.48	37.22	14.16	42.17	34.03	13.01	2.48	4.57	44.44	15.33	8.91	21.95	100.11	-240.18	0	0	0
9 - 35	0.04	0	0	0	0	19.48	37.22	14.16	-198.02	34.03	13.01	2.48	4.57	44.44	15.33	8.91	21.95	100.11	-358.57	0.69	0	0
8 - 53	0.03	0	0	0	0	-130.84	37.34	15.03	42.13	34.05	12.99	2.46	4.54	-106.75	14.78	8.91	22.40	99.86	-118.39	0.69	0	0
16 - 19	0	669.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-669.11
12 - 45	0.04	0	-359.02	0	0	19.48	37.22	14.16	42.17	34.03	13.01	2.48	4.57	44.44	15.33	8.91	22.4	100.11	0	0.69	0	0
14 - 47	0.04	0	0	-360.48	0	22.30	37.34	15.03	42.13	34.05	12.99	2.46	4.54	44.31	14.78	8.45	21.95	100.11	0	0	0	0

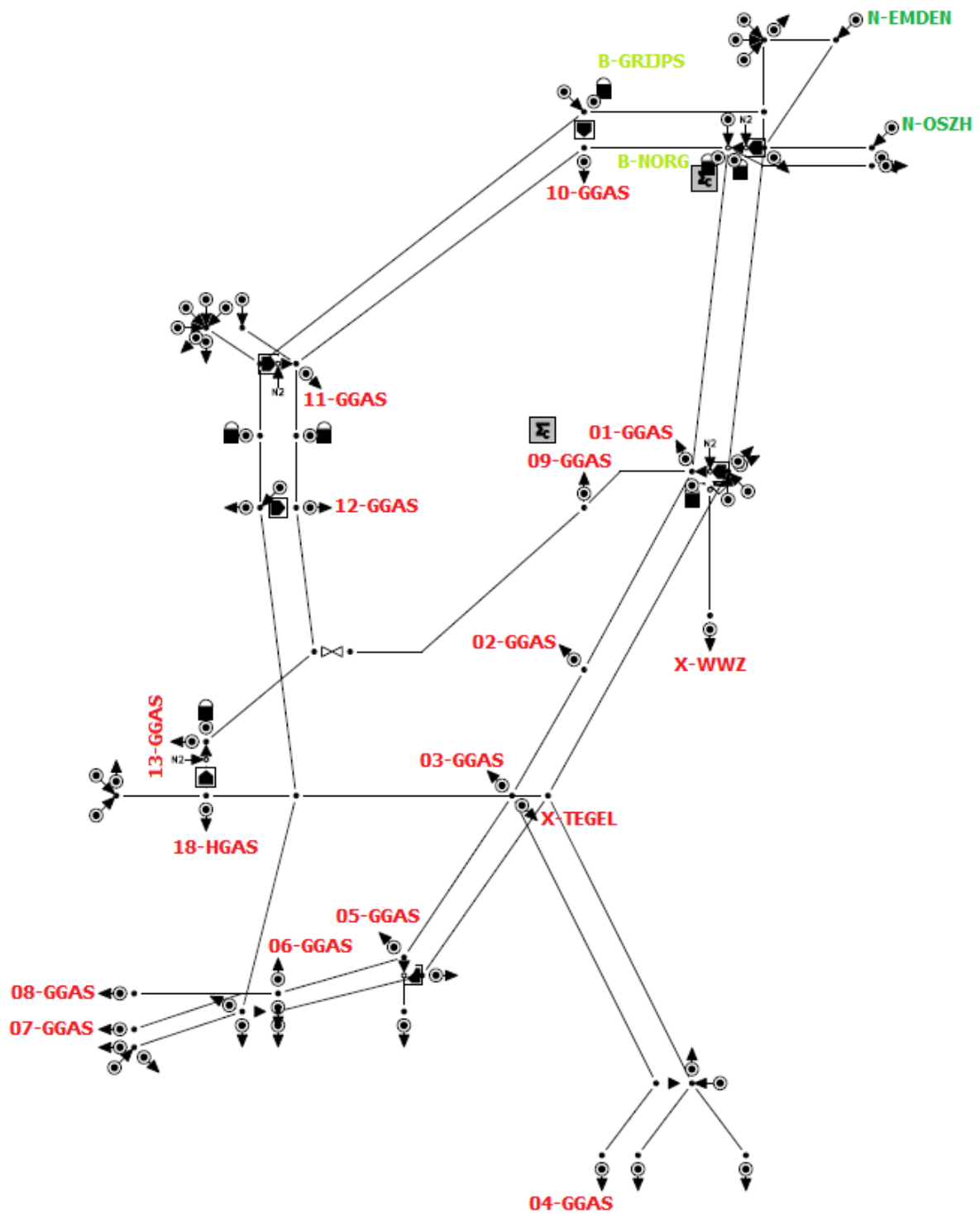


Figure 4-6: A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with some of the names.

Reducing generated set of stress tests

A next step is, to reduce the set of stress tests which are generated with *MCA*. This chapter describes some ideas for a threshold value for the quadratic form distance in the first section. In addition, some of these ideas are selected to experiment with in the second section and some generated sets are reduced.

5-1 Threshold value for the quadratic form distance

It is important to consider the so called capacity space to determine a threshold value to label two vectors as almost similar. This space is a QFD space and consists of the vectors representing the stress tests in \mathbb{R}^n . Several capacity vectors are drawn in Figure 5-1.

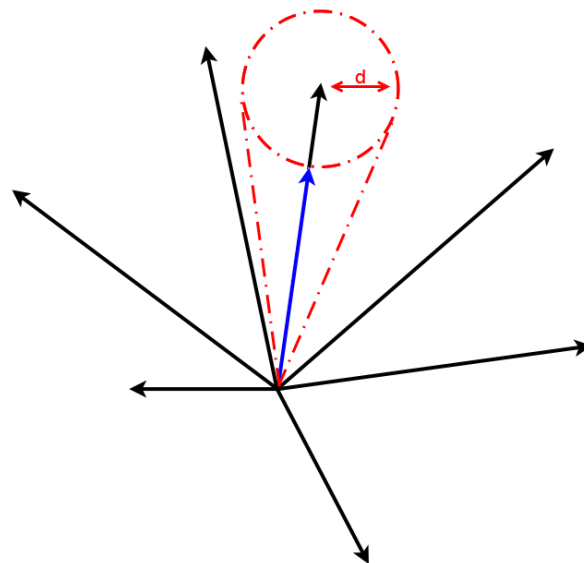


Figure 5-1: Some capacity vectors in \mathbb{R}^n , also called the capacity space.

5-1-1 The distance between the capacity vectors

It can be seen from this figure that a (QFD) sphere with radius d can be drawn around a capacity vector x . If another capacity vector y lies within this sphere, then we can call the two stress tests corresponding with these two capacity vectors almost equal and we can remove vector y from the set. A stress test corresponding with a vector with the same direction as the vectors within the sphere, but with a smaller length (see for example the blue vector in Figure 5-1), can also be called almost similar besides stress tests corresponding with these vectors. Thus, a region can be drawn to indicate which stress tests can be marked the same which is a region consisting of a cone and sphere both with radius d . So, it is important to determine this radius d which can be called a threshold value.

A possibility to determine the threshold value d is, to consider the two vectors in the capacity space and to take the inaccuracy of the generated stress tests into account. In general at GTS, these stress tests are compiled with an inaccuracy of $10 \text{ dam}^3/h$ per network point. In this case, we have a fixed inaccuracy and we wish the following condition to hold for each network point

$$|u_i| \leq |v_i \pm \delta_i| \quad \text{for } i = 1, 2, \dots, n,$$

such that the difference between the capacities is at most δ_i . Here, u_i is the capacity on point i for the stress test defined by the vector u , v_i the capacity on point i for the vector v and δ_i the permitted fixed inaccuracy ($\delta_i = 10 \text{ dam}^3/h$). If this condition holds, then the stress tests represented by the vectors u and v are similar enough to remove one of the stress tests from the generated set.

This condition results into the condition

$$\|u - v\|_{\mathbf{A}} = \|v - u\|_{\mathbf{A}} \leq \sqrt{\delta^T \mathbf{A} \delta} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \delta \mathbf{A}_{ij} \delta} = \delta \sqrt{\sum_{i=1}^n \sum_{j=1}^n \mathbf{A}_{ij}}.$$

for the distance function QFD. Thus, in this case d equals $\delta \sqrt{\sum_{i=1}^n \sum_{j=1}^n \mathbf{A}_{ij}}$.

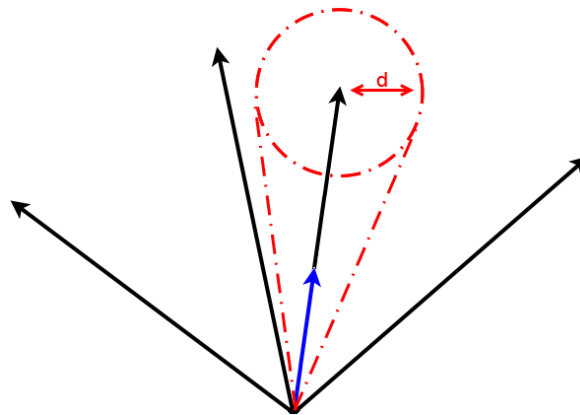


Figure 5-2: Some capacity vectors in \mathbb{R}^n , also called the capacity space.

However, the threshold value based on the radius d is not sufficient, because when a capacity vector has a much smaller length than another capacity vector but more or less the same

direction (see for example Figure 5-2), then the distance (QFD) between these vectors will be greater than the threshold d . If we consider the corresponding stress tests in this case, then we would conclude that the stress test represented by the vector with smaller length can be removed from the set, because the stress test is less severe than the other.

Therefore, it could be wise to consider the length of the vectors in combination with the QFD instead of the angle between the vectors. In that case, we have to adjust the length of the vectors such that every vector has the same length (for example, by multiplying each vector element with the same factor or normalizing each vector) and then apply the QFD.

However, another alternative is to consider the angle between two capacity vectors.

5-1-2 The angle between the capacity vectors

Note that the angle between the capacity vectors can be used as a threshold value besides this radius d or distance between the capacity vectors. If we consider, for example, two capacity vectors x and y , then the angle θ between these two vectors is calculated as

$$\cos \theta = \frac{\langle x, \mathbf{A}y \rangle}{\|x\|_{\mathbf{A}} \|y\|_{\mathbf{A}}} \Rightarrow \theta = \arccos \left(\frac{x^T \mathbf{A}y}{\sqrt{x^T \mathbf{A}x} \sqrt{y^T \mathbf{A}y}} \right). \quad (5-1)$$

If this angle θ equals zero, then the two capacity vectors are exactly the same or one capacity vector has a smaller length than the other but the same direction. If the angle θ is almost zero, then the two capacity vectors are almost similar. Consider the capacity vector x and its cone. Then the capacity vector y can be marked as almost the same (similar) as the vector x , when the vector y is lying within this cone. Therefore, it is important to determine the angle φ between the vector x and the boundary of its cone. Then it should hold that the angle θ between x and y according to Equation 5-1 is smaller than or equal to φ . Let us assume that an inaccuracy (fixed or variable) δ_i per vector component (for all components i) is permitted, which means that $x_i - \delta_i \leq x_i \leq x_i + \delta_i$ holds. We can store these inaccuracies in a vector δ . Then we want to find a vector $x + \delta$ with the greatest angle between this vector and x and this angle is then our wanted angle φ . This results in the following condition for the angle θ between x , an arbitrary vector y and $\delta = (\delta_1, \delta_2, \dots, \delta_n)^T$

$$\theta \leq \varphi = \max_{\delta} \left\{ \arccos \left(\frac{x^T \mathbf{A}(x + \delta)}{\sqrt{x^T \mathbf{A}x} \sqrt{(x + \delta)^T \mathbf{A}(x + \delta)}} \right) \right\}. \quad (5-2)$$

We consider a three dimensional example in the Euclidean space to illustrate how the angle between a capacity vector and the boundary of its cone can be determined:

$$x = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \quad \text{with a fixed inaccuracy} \quad \delta_i = \frac{1}{2}.$$

Then we can consider a few vectors y which satisfy the condition $x_i - \delta_i \leq y_i \leq x_i + \delta_i$ and calculate the angle between these vectors y and the vector x :

- $y = \begin{pmatrix} 2\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 19.47^\circ$;
- $y = \begin{pmatrix} 1\frac{1}{2} \\ -1\frac{1}{2} \\ -1\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 19.47^\circ$;
- $y = \begin{pmatrix} 2\frac{1}{4} \\ -1\frac{1}{2} \\ -1\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 8.05^\circ$;
- $y = \begin{pmatrix} 2\frac{1}{2} \\ -1\frac{1}{2} \\ -1\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 5.05^\circ$;
- $y = \begin{pmatrix} 2\frac{1}{2} \\ -1\frac{1}{4} \\ -1\frac{1}{4} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) = 0^\circ$;
- $y = \begin{pmatrix} 2\frac{1}{2} \\ -1\frac{1}{8} \\ -1\frac{3}{8} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 3.30^\circ$;
- $y = \begin{pmatrix} 2 \\ -1\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 16.10^\circ$;
- $y = \begin{pmatrix} 2\frac{1}{2} \\ -1\frac{1}{2} \\ -1 \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 6.59^\circ$;
- $y = \begin{pmatrix} 1\frac{1}{2} \\ -\frac{1}{2} \\ -1 \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 10.89^\circ$;
- $y = \begin{pmatrix} 1\frac{1}{2} \\ -1\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ with $\theta = \arccos\left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}\right) \approx 20.51^\circ$.

We suspect that φ equals approximately 20.51° judging the results above. We can generate a lot of random vectors y which satisfy $x_i - \frac{1}{2} \leq y_i \leq x_i + \frac{1}{2}$ with *Matlab* in the following way:

```
% Inaccuracy random components delta
nr = 1e5;      % number of random del
```

```

del_comp = 0.5*ones(3,1);      % fixed
% First component del
a1 = -del_comp(1);
b1 = del_comp(1);
r1 = a1 + (b1 - a1) .* rand(nr,1);
% Second component del
a2 = -del_comp(2);
b2 = del_comp(2);
r2 = a2 + (b2 - a2) .* rand(nr,1);
% Third component del
a3 = -del_comp(3);
b3 = del_comp(3);
r3 = a3 + (b3 - a3) .* rand(nr,1);
% Combine three components in a matrix
del_random = horzcat(horzcat(r1,r2),r3);
% Determine the angles between x and y = x+del (real part)
for i = 1 : nr
    Theta_xy(i) = real(Theta_eucl(x,(x+del_random(i,:))));
end
% Determine the maximum angle
with
function [ Theta_eucl ] = Theta_eucl( x, y )
    Theta_eucl = acosd(dot(x,y)/(norm(x)*norm(y)));
end

```

Running this code a few times always results in a maximum angle around the 20 °. So, φ is set equal to 20.51 °.

Another option is to determine the angle φ between a stress test (represented by a vector x) and the boundary of its cone with respect to a radius d

$$\sin \varphi = \frac{d}{\|x\|_{\mathbf{A}}} = \frac{\sqrt{\delta^T \mathbf{A} \delta}}{\sqrt{x^T \mathbf{A} x}}.$$

This results in the following condition for the angle θ between two stress tests

$$\theta = \arccos \left(\frac{\langle x, \mathbf{A} y \rangle}{\|x\|_{\mathbf{A}} \|y\|_{\mathbf{A}}} \right) \leq \varphi = \max_d \left\{ \arcsin \left(\frac{d}{\|x\|_{\mathbf{A}}} \right) \right\}. \quad (5-3)$$

If this condition holds, then the vectors x and y can be called almost similar. Again, we have to determine the greatest possible cone and thus, we have to maximize φ .

5-2 Conducted experiments

We apply the methods mentioned above on the network displayed in Figure 3-6 for which we generated a set of 66 stress tests. This set of generated stress tests consists of two subsets: a subset containing 33 stress tests in the case the temperature is -16 degree Celsius and a subset containing 33 stress tests with a temperature of -17 degree Celsius. Thus, we consider one of these subsets, for example the set of stress tests with a temperature of -17 degree Celsius which consists of 16 different stress tests. Thus, the reduced set must consist of maximum 16 stress tests. We want to reduce this set of 33 stress tests as follows:

1. Start with the vector with the greatest length, say x .
2. Calculate the angle φ between this vector and the boundary of its cone.
3. Apply the condition given in (5-3) and determine which other vectors lie within this cone.
4. Delete these vectors which are almost similar to this longest vector from the set.
5. Consider the next longest vector y ($y \neq x$) of the remaining set and apply steps **2** - **5** till the remaining set is empty.

First we conduct the experiment with the QFD matrix \mathbf{A} only depending on the length of the pipelines ($\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$). Then the reduced set equals $\{1, 5, 7, 10, 12, 14, 16, 17, 18, 25, 33\}$, whereby vectors 5, 10, 17 and 33 have no vectors in their cones, if we let *Matlab* generate 10 vectors δ besides $(10, 10, \dots, 10)^T$ and $(-10, -10, \dots, -10)^T$ or a few sets of 100 or 1000 vectors δ . The reducing step is done with the help of *Matlab* in the following way:

```
% Deleting based on colCone
% Start with the vectors inside the cone of the longest vector
for k = 1 : nIVs
    Length_IVs(k) = QFD(IVs(k,:), 0, A1);
end
[IVsL, IVsI] = sort(Length_IVs, 'descend');
for j = 1 : nIVs
    Ind = find(colCone == IVsI(j));
    Ind = sort(Ind, 'descend');
    if length(Ind) > 1
        for k = 1 : length(Ind)
            colToDel = find(colCone == rowCone(Ind(k)));
            colToDel = sort(colToDel, 'descend');
            if rowCone(Ind(k)) > IVsI(j)
                colCone(colToDel) = [];
                rowCone(colToDel) = [];
            elseif rowCone(Ind(k)) == IVsI(j)
```

```

rowCone = rowCone;
colCone = colCone;
else
colCone(colToDel) = [];
rowCone(colToDel) = [];
Ind = Ind - length(colToDel);
end
end
end
end
end
end

```

When the matrix depends on the diameter as well ($\mathbf{A}_{ij} = 1 - \frac{r_{ij}}{r_{max}}$, Equation (4-9)), then we have $\{1, 5, 7, 10, 12, 14, 16, 17, 26, 33\}$ as a reduced set with no vectors in the cones of vector 5, 17 and 33 for a different set of 10 extra vectors δ or a few sets of 100 or 1000. These two sets are visualized in Figure 5-3.

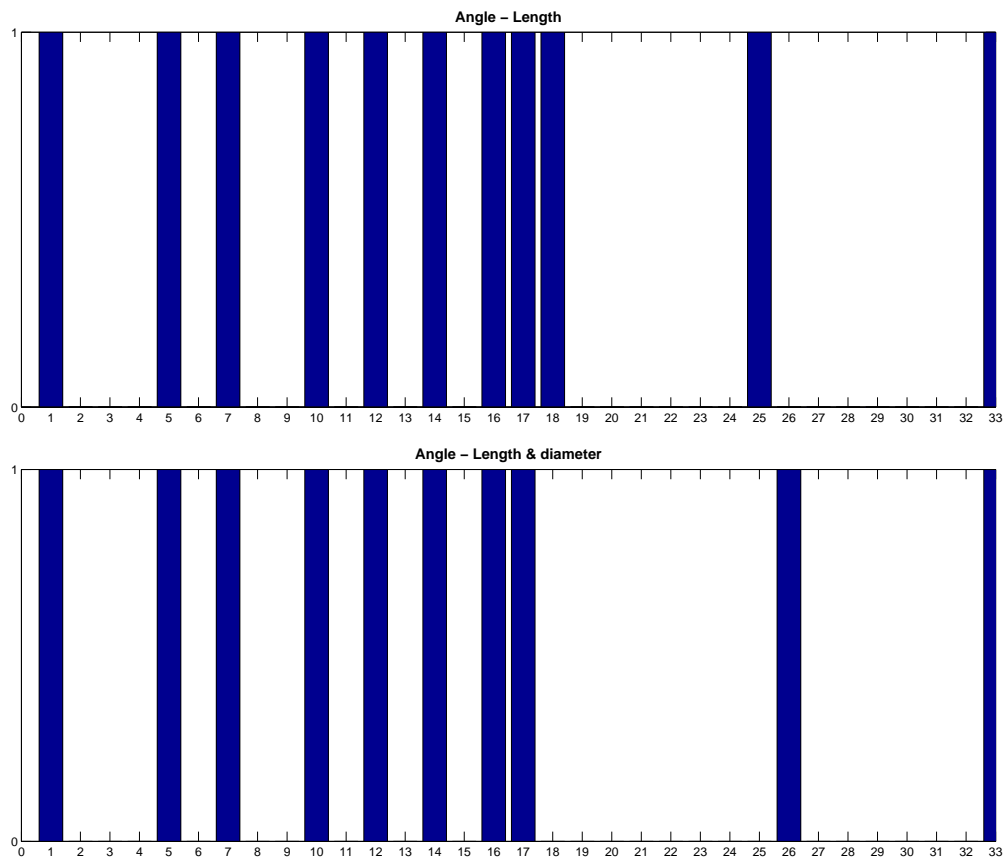


Figure 5-3: Two reduced sets, when the angle between two vectors is considered.

Another possibility is, to reduce the generated set of stress tests with the condition stated in Equation (5-2) and see if we end up with a different reduced set or not. Again, we start with the vector with the greatest length, calculate the angle φ between this vector and the boundary of its cone and then determine which other vectors lie within this cone by applying the other condition. This approach results indeed in the same two reduced sets.

In addition, an experiment in which a percentage of the capacity as inaccuracy is used is conducted with *Matlab*. If we implement this with, for example $\delta_i = 1\%$ in *Matlab*, then we end up with the set $\{1, 4, 5, 7, 10, 12, 14, 16, 17, 18, 19, 25, 33\}$ for the matrix only depending on the length of the pipelines with 10,000 generated vectors δ . When we apply the same experiment for the matrix depending on the diameter of the pipelines as well, then the set equals $\{1, 5, 7, 8, 10, 12, 14, 16, 17, 19, 26, 33\}$ for a set 10,000 generated vectors δ and $\{1, 5, 7, 8, 10, 12, 14, 16, 17, 26, 33\}$ in which vector 19 lies within the cone of vector 16 for a different set of 10,000 generated vectors δ . This last reduced set and the reduced set for the matrix depending only on the length are again visualized, see Figure 5-4.

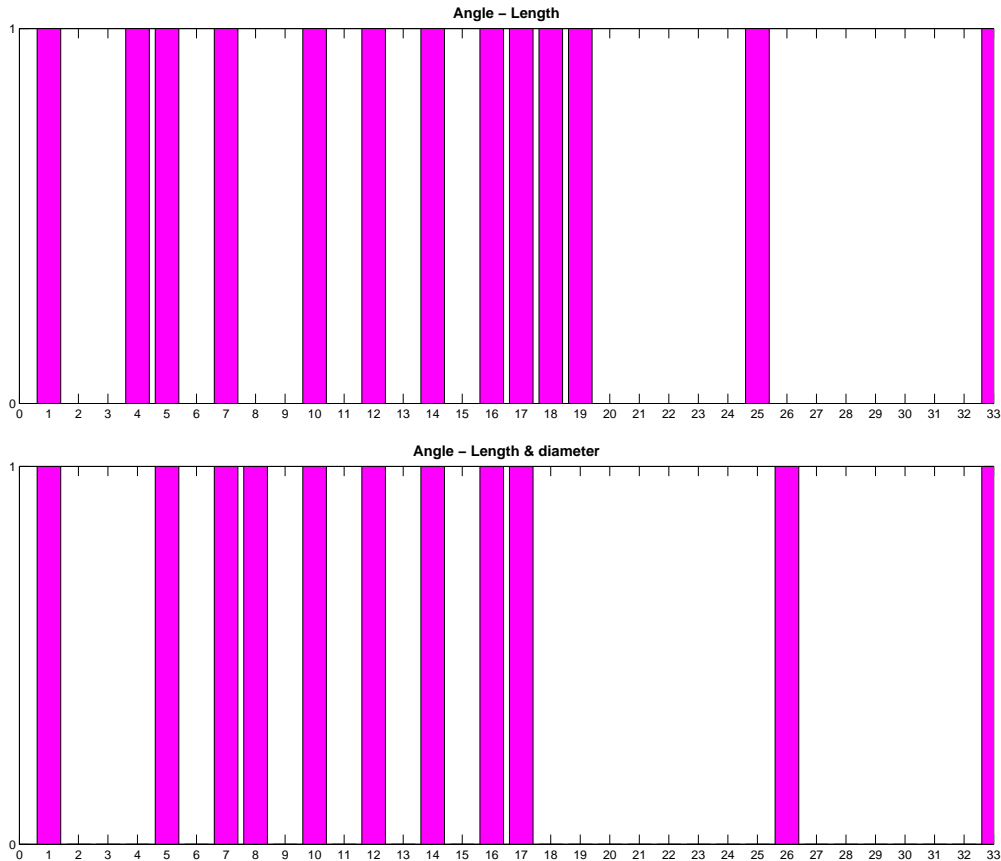


Figure 5-4: Two reduced sets, when the angle between two vectors is considered with variable inaccuracy δ_i .

At a later stage, we have adjusted the correlation matrix defined as $\mathbf{A}_{ij} = 1 - \frac{r_{ij}}{r_{max}}$ as is mentioned in Subsection 4-3-4. The unknown diameters of some pipelines in the network

displayed in Figure 3-6 were approximated by the diameters of corresponding pipelines in the network displayed in Figure A-27. If we apply our reducing algorithms, one with an inaccuracy of $10 \text{ dam}^3/h$ and one with an inaccuracy of 1% per network point capacity, then we end up with slightly different reduced sets:

- we get $\{1, 5, 7, 8, 12, 14, 16, 17, 33\}$, whereby vectors 10 and 26 are now lying in the cone of vector 8, for $\delta_i = 10 \text{ dam}^3/h$. This set is visualized in Figure 5-5.
- we get $\{1, 5, 7, 8, 10, 12, 14, 16, 17, 18, 33\}$, whereby vector 26 is now shorter than vector 18 and is lying in the cone of this vector, for $\delta_i = 1\%$. This set is visualized in Figure 5-6.

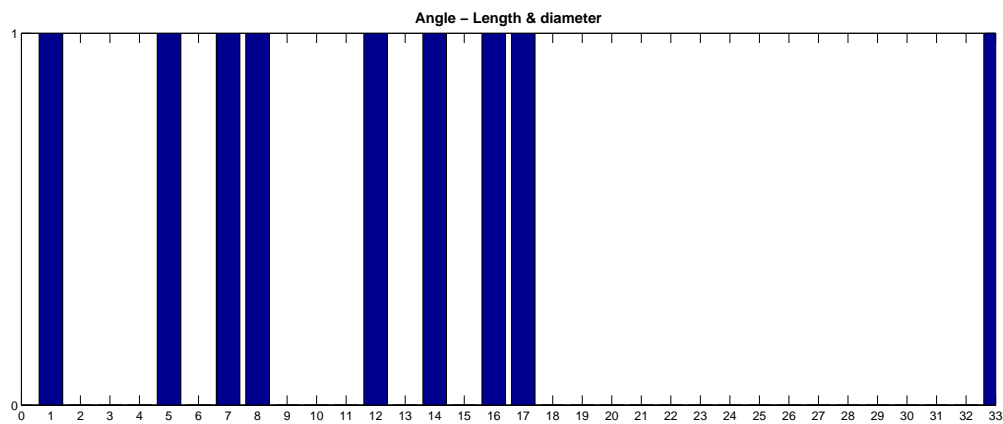


Figure 5-5: The reduced sets for the new diameters, when the angle between two vectors is considered with a 'constant' inaccuracy δ_i .

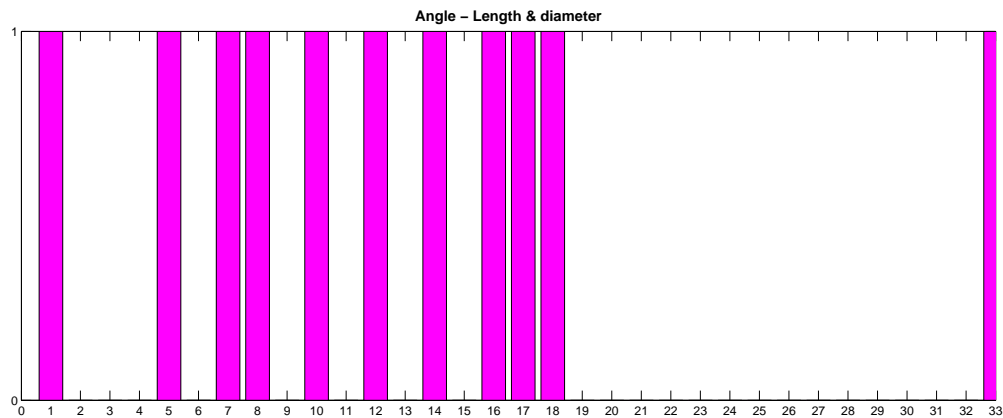


Figure 5-6: The reduced sets for the new diameters, when the angle between two vectors is considered with variable inaccuracy δ_i .

5-3 Analyzing the conducted experiments

When we are considering the two different reduced sets for which both the length and diameter are taken into account, then we see that the vector 26 is replaced by vector 18 for an inaccuracy of maximal 1% of its capacity per network point. This is due to the fact that the length of vector 18 is greater than the length of 26 with the adjusted correlation matrix, while first it was the other way around. So, we could say that these two reduced sets are much alike. When we apply an inaccuracy of $10 \text{ dam}^3/h$ per network point, then the new reduced set with respect to the adjusted correlation matrix regarding the diameter consists of nine vectors, while the previous reduced set consisted of ten vectors. If we take a closer look at these two reduced sets, then we see that vector 8 is added to the set and vectors 10 and 26 are deleted, because vectors 10 and 26 are now lying within the cone of vector 8.

A next step is to compare the reduced set for which the correlation matrix only depends on the length to the reduced set for which the correlation matrix also depends on the diameter. We will use the latest compiled reduced sets regarding the length and diameter for this comparison.

We have for $\delta_i = 10 \text{ dam}^3/h$

- the following similarities:
 - vectors 4 and 6 are lying in the cone of vector 1;
 - no vectors are lying in the cone of vector 5;
 - vectors 2, 3, 9, 13, 24, 29, 30 and 32 are lying in the cone of vector 7;
 - vector 23 is lying in the cone of vector 12;
 - vectors 15 and 27 are lying in the cone of vector 14;
 - vectors 19, 22 and 31 are lying in the cone of vector 16;
 - no vectors are lying in the cone of vector 17;
 - no vectors are lying in the cone of vector 33.

- the following differences:

first, vector 10 was not lying in a cone of a vector, while it is lying in the cone of vector 8 for the second reduced set. In addition, vectors 26 and 28 were lying in the cone of vector 18, but these three vectors are lying in the cone of vector 8 for the second set. Vectors 11, 20, 21 and 25 are also lying in the cone of vector 8, while first vectors 8, 11, 20 and 21 were lying in the cone of vector 25. However, the length of vector 25 became smaller than the length of 8 with a new correlation matrix (diameter added as well) which explains the switch from 25 to 8.

We have for $\delta_i = 1\%$

- the following similarities:
 - no vectors are lying in the cone of vector 5;
 - vectors 2, 3, 9, 13, 24, 29, 30 and 32 are lying in the cone of vector 7;
 - no vectors are lying in the cone of vector 10;
 - vector 23 is lying in the cone of vector 12;
 - vectors 15 and 27 are lying in the cone of vector 14;
 - vector 22 is lying in the cone of vector 16;
 - no vectors are lying in the cone of vector 17;
 - vectors 26 and 28 are lying in the cone of vector 18;
 - no vectors are lying in the cone of vector 33.

- the following differences:

vectors 4 and 6 are no longer lying in the cone of vector 1. Again, vector 8 has now a greater length than vector 25 and therefore, vectors 11, 20, 21 and 25 are lying in its cone instead of 8, 11, 20 and 21 in the cone of 25. Vectors 19 and 31 are now lying in the cone of 16 as well.

It needs to be investigated, whether or not one of the used inaccuracies, $\delta_i = 10 \text{ dam}^3/h$ and $\delta_i = 1\%$, is a ‘good’ threshold value to reduce the generated sets of stress tests and which correlation matrix needs to be applied, only depending on the length or also on the diameter.

An idea about how to judge if the difference of two capacity vectors is small enough to call the two vectors almost similar, is to consider the capacities of this difference vector and the allowed maximum flow through the pipelines according to *MCA*. If the difference vector has, for example a capacity of 1 at point *A* and a capacity of -1 at point *B*, then one volume of gas has to flow through the network from *A* to *B* and if the maximum flow through the pipeline between *A* and *B* is, for example 100, then we expect that we can call the two vectors corresponding with this difference vector almost similar.

First of all, let us list which capacity are exactly the same:

- vectors 2, 3, 13, 29, 30 and 32;
- vectors 4 and 6;
- vectors 7, 9 and 24;
- vectors 8, 11, 20 and 21;
- vectors 12 and 23;
- vectors 14, 15 and 27;
- vectors 16 and 22;

- vectors 19 and 31;
- vectors 26 and 28.

Therefore, it is important to investigate the difference between the following vectors

- 1 and 4;
- 2 and 7;
- 8 and 10;
- 8 and 18;
- 8 and 25;
- 8 and 26;
- 16 and 19;
- 18 and 26;

in order to say something about (almost) similarity. These differences are listed in Table 5-1. The gas transport network and the network points which are mentioned in Table 5-1 are displayed in Figure 5-7. In addition, the maximum flow according to *MCA* through the relevant pipelines is denoted in this figure as well.

If we look at the difference between stress tests 1 and 4, then this difference can be represented as a flow of 410.62 from *N-OSZH* to *23-HGAS*. There are two different ways to end at *23-HGAS*: through a pipeline with maximum flow 400 and one with maximum flow 2200. Following the route consisting of the last named pipeline (2200) does not give any problems. However, if we have to follow a route consisting of the pipe with maximum flow 400, then we do have a problem, because then it is not possible.

If we consider the difference between stress tests 2 and 7, then we do not encounter any problem. It is possible to let 240.15 volumes of gas flow through the system from *B-NORG* to *04-GGAS*. We do not encounter any problems for the differences 8 - 18, 8 - 25, 8 - 26, 16 - 19 and 18 - 26 as well.

The difference between stress tests 8 and 10 represents a more complicated situation. First of all, let us cluster the exit, entry and storage points *01-GGAS*, *15-HGAS*, *V-OMMEN* and *B-EPE* to one point. Note that we can transport 56.93 volumes of gas from *X-WWZ* to this cluster point without any problems. Then the situation is reduced to: transport 1936.43 volumes of gas from the cluster point to *B-BERGEN* and 151.06 gas volumes from *09-GGAS* to *11-GGAS* from which 44.61 volumes of gas needs to be transported further from *09-GGAS* to *B-BERGEN*. It is possible that the route of the cluster point to *B-BERGEN* consists of a pipeline with maximum flow 2000. Then a transport of 1936.43 gas volumes can be severe with respect to this pipeline.

So, we could say that we can likely call the vector pairs 2 - 7, 8 - 18, 8 - 25, 8 - 26, 16 - 19 and 18 - 26 almost similar based on the above idea.

Table 5-1: The eight mentioned difference vectors.

Difference stress tests	Entry and exit points (to characterize the differences)													
	01-GGAS	02-GGAS	03-GGAS	04-GGAS	09-GGAS	10-GGAS	11-GGAS	15-HGAS	23-HGAS					
1 - 4	0	0	0	0	0	0	0	0	0	0	0	0	0	-410.62
7 - 2	0	0	0	-240.15	0	0	0	0	0	0	0	0	0	0
8 - 10	190.56	0	0	0	151.06	0	-106.47	294.6	0	0	0	0	0	0
8 - 18	0	-277.32	-117.28	0	0	159.14	0	0	0	0	0	0	0	0
8 - 25	0	0	0	0	0	0	-56.93	0	0	0	0	0	0	0
8 - 26	0	-277.32	-117.28	0	0	159.14	0	0	0	0	0	0	0	0
16 - 19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18 - 26	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Difference stress tests	Entry and exit points (to characterize the differences)									
	B-BERGEN	B-EPE	B-GRIJPS	B-NORG	N-OSZH	V-GRFR	V-OMMEN	X-TEGEL	X-WWZ	
1 - 4	0	0	0	0	410.62	0	0	0	0	0
7 - 2	-1981.04	0	0	0	0	0	0	0	0	0
8 - 10	0	1300	0	0	0	0	94.34	0	0	56.93
8 - 18	0	0	848.65	0	0	0	0	-1.01	0	-612.18
8 - 25	0	0	0	0	0	0	0	0	0	56.93
8 - 26	0	0	670.56	0	0	178.09	0	-1.01	0	-612.18
16 - 19	0	0	847.2	0	0	-178.09	0	0	0	-669.11
18 - 26	0	0	-178.09	0	0	178.09	0	0	0	0

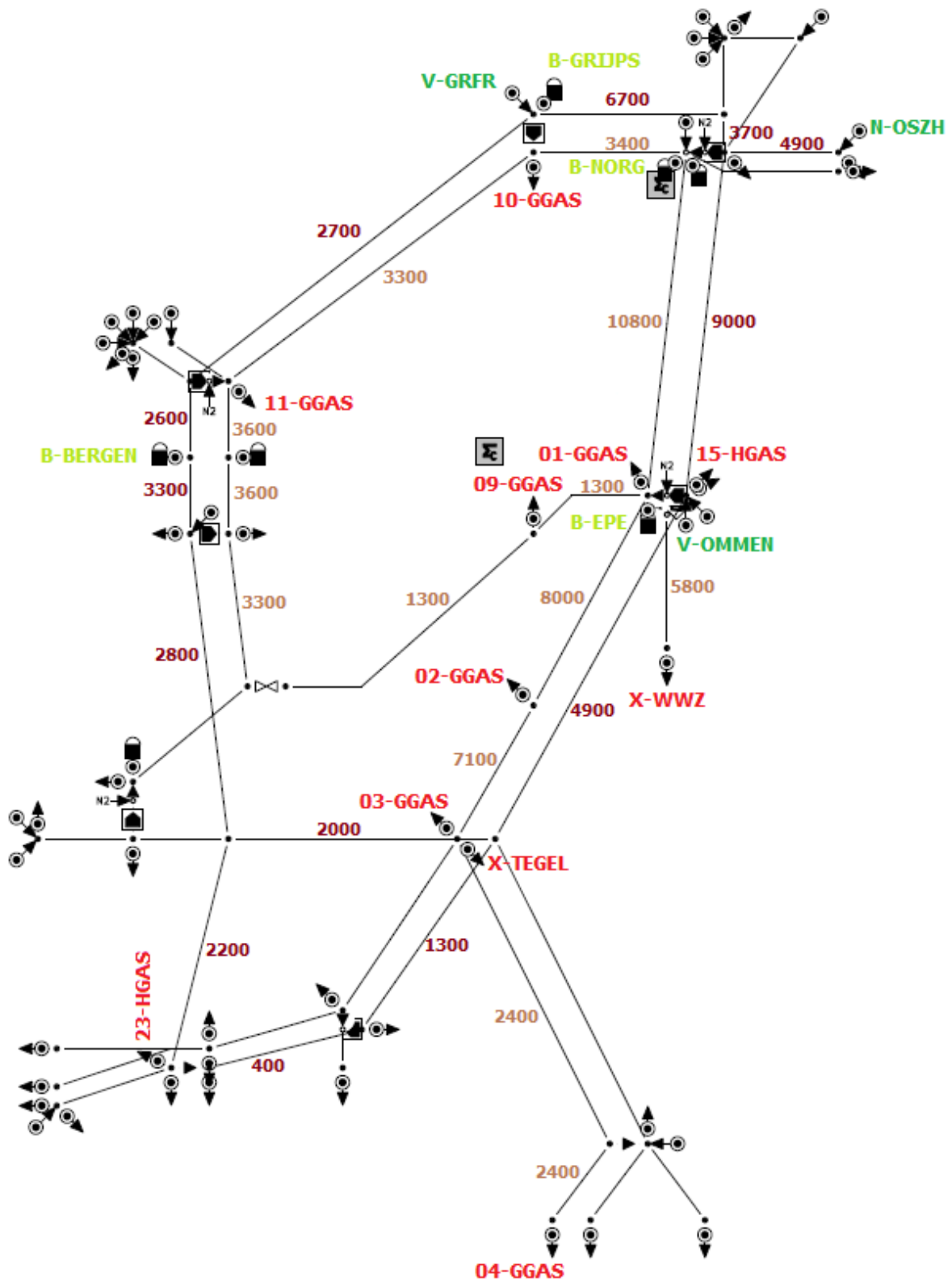


Figure 5-7: A simplification of the network of GTS, consisting of H-gas and G- and L-gas, with some of the names and maximum flow through some pipelines.

5-4 Some more experiments

Firstly, we adjusted the variable inaccuracy from 1% to 2% and conducted our reducing steps. Then we ended up with the reduced sets $\{1, 5, 7, 10, 12, 14, 16, 17, 19, 25, 33\}$ and $\{1, 5, 7, 8, 12, 14, 16, 17, 33\}$. We get the first set as result, when we are only considering the length of the pipelines and we get the second set as result, when we are also considering the diameter. A good result in the first set is the fact that vector 10 is not in the cone of another vector. However, it would be nice if vectors 16 and 19 were lying within each others cones. This is the case for the second set, but vector 10 is then lying within the cone of vector 8. In addition, vector 4 is in the cone of vector 1 in both sets.

Secondly, we used a variable inaccuracy of 1.5%, 1.25% or 1.2% during the reducing process which all result in the following two sets

- $\{1, 4, 5, 7, 8, 10, 12, 14, 16, 17, 18, 19, 33\}$, taking into account the length of the pipelines. It is nice that vector 4 is not lying within the cone of vector 1, but still vectors 16 and 19 are not in each others cone. Besides, vectors 18 and 25 are no longer lying within each others cones.
- $\{1, 5, 7, 8, 10, 12, 14, 16, 17, 33\}$, taking into account the length and diameter of the pipelines. An improvement is that vector 10 is now not lying within in a cone of another vector.

Thirdly, we tried an inaccuracy of 1.15% or 1.1% to reduce the generated set of stress tests. Then we get the set $\{1, 4, 5, 7, 8, 10, 12, 14, 16, 17, 18, 19, 33\}$, when we considering the length of the pipelines which is the same set as for an inaccuracy of 1.5%, 1.25% or 1.2%. In addition, we get the set $\{1, 5, 7, 8, 10, 12, 14, 16, 17, 26, 33\}$, when we are considering the diameter as well. Unfortunately, it does not give the desired sets according to our analysis in the previous section.

Finally, we have used some information from the detailed network of GTS which is displayed in Figure 4-2 to determine an inaccuracy. The stress tests are generated for this detailed network, which means that all capacities are determined. In general, these capacities are fixed capacities, except for 13 entry/exit points in the detailed network. It is possible that a somewhat higher capacity on these 13 specific entry/exit point is offered to the market. So, these 13 capacities can be considered as variables.

Therefore, we have determined which entry/exit points in the network drawn in Figure 3-6 are corresponding with these 13 entry/exit points in the detailed network:

- *20-HGAS* which is the 19th component of a vector representing a stress test.
- *N-EMDEN* which is the 35th component.
- *N-ZELZ* which is the 43rd component.

- *X-BBLDE* and *X-BBLGA* which are the 51st and 52nd components, respectively. These two exit points have a ‘shared’ inaccuracy. In other words, if the inaccuracy equals $10 \text{ dam}^3/h$ per point, then the total inaccuracy for these two exit points together equals 10. Then it is possible that the inaccuracy of one of the exit points is equal to 6 and then the other has an inaccuracy of 4.
- *X-GRAV* which is the 54th component.
- *X-HILV* which is the 55th component.
- *X-OSZG* which is the 56th component.
- *X-OSZH* which is the 57th component.
- *X-TEGEL* which is the 58th component.
- *X-WWZ* which is the 60th component.
- *X-ZANDH* which is the 62nd component.
- *X-ZELZ* which is the 63rd component.

First, we conducted a reducing with an inaccuracy of $10 \text{ dam}^3/h$ per point (based on the used step size in *MCA*), whereby we take the ‘shared’ inaccuracy of *X-BBLDE* and *X-BBLGA* into account. Secondly, we try a percentage as inaccuracy as well, 1% or 1.2%, which gave the same results. Besides, this fixed inaccuracy of $10 \text{ dam}^3/h$ and the variable inaccuracy 1% or 1.2% give the same reduced sets. We first take the length into account, then the diameter as well and then we get the following two sets:

- {1, 2, 4, 5, 7, 10, 12, 14, 16, 17, 18, 19, 25, 33};
- {1, 4, 5, 7, 8, 10, 12, 14, 16, 17, 18, 19, 33}.

If we compare these two sets, then the second set is more preferable, because it is plausible that vectors 2 and 7 are lying within each others cone.

Considering all conducted experiments, we have to reconsider the differences 16 - 19 and 8 - 18, because if we are trying to reduce the set such that 16 and 19 are in each others cone, then vectors 1 and 4 ‘end up’ in each others cone as well, which is most likely not desired. These two differences, 16 - 19 and 8 - 18, are pretty similar, but the difference 16 - 19 is smaller (669.11 volumes of gas with respect to 848.65 on the same route). So, it is expected that first vectors 16 and 19 are ‘ending up’ in each others cone and then vectors 8 and 18, when we adjust the inaccuracy such that the sets become more reduced.

One of the possible routes consists of pipelines with a maximum gas flow of 1300. Therefore, it could be wise to call stress tests 16 and 19 not almost similar considering 669.11 in relation to 1300 and thus, to be content with the reduced set {1, 4, 5, 7, 8, 10, 12, 14, 16, 17, 18, 19, 33}. Then, the vectors 1 and 4 are not almost similar nor vectors 8 and 10, which is desirable.

In this way, our choice is based on realistic inaccuracies, namely the step size of *MCA* on specific entry/exit points and on both the length and diameter.

Conclusions and recommendations

We have seen that stress tests can be denoted by vectors which elements represent the gas capacities at entry and exit points. In general, we deal with n -dimensional vectors, when we are considering a network of n entries and exits. In addition, we are always dealing with a balanced combination of entry and exit capacities and therefore we call these vectors balanced vectors. This means that the elements of such a vector sum up to zero. So, when a stress test is represented by an n -dimensional vector x , then $\sum_i x_i = 0$ holds. The individual dimensions of these vectors are correlated, because we want to take the mutual distance between the network points (following the pipelines) into account as well, besides the capacities on the entry and exit points. Thus, the vector space we are considering is not an Euclidean space.

Therefore, we have chosen to consider a cross-bin dissimilarity measure, called the quadratic form distance, in order to compare the stress tests which are already generated by Gasunie Transport Services. The quadratic form distance is a promising measure, because it can deal with correlation between the individual dimensions. It uses a correlation matrix in order to measure (dis)similarity between two stress tests represented as vectors. When we want to apply the QFD, then we have to make sure that this correlation matrix is symmetric positive semidefinite. Otherwise, the chosen (dis)similarity measure is not a well defined metric. We namely would like the QFD to be real valued and to have the possibility that the QFD is equal to zero (see the example of Subsection 3-1-4). We try to answer some of the sub research questions regarding the QFD in the next section.

6-1 Answers to the sub research questions

We have listed some sub research questions in Section 2-4 in order to test a proposed (dis)similarity measure. Thus, we have a compiled list of criteria that the chosen measure, the QFD, needs to satisfy. We have also partly examined the QFD during the literature study and our findings regarding these criteria are summarized in Table 6-1.

Table 6-1: An overview of the findings of the QFD regarding the criteria.

<i>Criteria</i>	1. Physical characteristics of the gas transport network	Probably
	2. Distinction between different stress tests	Yes
	3. Threshold value relates to the ‘generating inaccuracy’ of 10 dam^3/h	Yes
	4. Tuning the parameters	Yes → matrix A
	5. The need to use specific transport physics of the gas network	Low
	6. Similarity, when a stress test is less severe than the other	Not considering the threshold value d , but ‘yes’ for the threshold value φ
	7. Applicable for stress tests depending on blending load	Not examined

We expect that the QFD can take the physical characteristics of the transport network into account, because we can adjust the correlation matrix **A** in the QFD. Thus, it is possible to modify this matrix such that the physical characteristics of the gas transport network are included. We have tried for example to include the pressure drop in Section 4-1 by adjusting the relation between the components of the correlation matrix, \mathbf{A}_{ij} , and the distance between the network points, d_{ij} . This resulted in a different definition for the correlation matrix: $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$ besides the original definition $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$. A next step was to test both definitions on a few networks provided by Jarig Steringa which are the networks displayed in Figures 3-5 and 3-6, but the detailed network of GTS (Figure 4-2) as well. These tests indicate that the matrix defined as $\mathbf{A}_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{max}}}$ is not SPSD for one of these three networks. However the ‘original’ matrix ($\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$) is SPSD for the first network drawn in Figure 3-5, but it is no longer SPSD for the other two networks.

These results motivated us to search for a different definition for the correlation matrix such that this correlation matrix is SPSD for all three networks. A correlation matrix which satisfies this property is defined as $\mathbf{A}_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{max}}}$, but does not represent the desired relation (regarding the pressure drop) between its components \mathbf{A}_{ij} and the transportation distance d_{ij} .

So, a second option was to adjust the variables on which the components \mathbf{A}_{ij} depend, instead of changing its formula. In other words, we wanted to add a variable to the formula besides the variable d_{ij} based on the pressure drop equation. The length and the diameter of the pipelines do not change in this pressure drop equation during a or per gas transport situation and thus can both variables be seen as constants. Thus, we added the diameter to the formula as well. The components \mathbf{A}_{ij} are then defined as

$$\mathbf{A}_{ij} = 1 - \frac{r_{ij}}{r_{max}} \quad \text{with} \quad r_{ij} = \sum_{k=i}^{j-1} r_{k,k+1} = \sum_{k=i}^{j-1} \frac{L_{k,k+1}}{D_{k,k+1}^5}, \quad r_{max} = \max_{i,j} r_{ij},$$

with $L_{k,k+1}$ now representing the length and $D_{k,k+1}$ the diameter of the pipeline between network point k and $k + 1$. We have seen that this correlation matrix is SPSD for the two

networks drawn in Figures 3-5 and 3-6, which is a very positive result. However, we could not test this new definition on the detailed network of GTS displayed in the Figure 4-2, because it is a lot harder to take the diameter into account as well for this detailed network. In this detailed network there are a lot more network points which do not have an entry, exit point or storage linked to it and it is not possible to delete these points by hand, like is done for the second network (Figure 3-6).

We always have to keep in mind that the QFD must be a metric and therefore the correlation matrix \mathbf{A} needs to be symmetric positive semidefinite. So, we have to prove that the chosen matrix is indeed SPSD before we apply the QFD for a specific network. Because we have some freedom to define a correlation matrix \mathbf{A} , the QFD satisfies the fourth criterion as well.

We have also seen that the QFD can distinguish between different stress tests, for example in Subsection 3-1-2. It can also be concluded that there is a low need of using specific transport physics of the gas network, because we only need the capacities on the network points, the mutual distance between these points and in some cases the diameter of the pipelines as well. So, no transport calculations need to be done before conducting the QFD.

The threshold values d (radius) and φ (angle) defined in Section 5-1 relate to the ‘generating inaccuracy’ $\delta = 10 \text{ dam}^3/h$. This ‘generating inaccuracy’ is based on the used step size of $10 \text{ dam}^3/h$ in the program *MCA*. The threshold value φ especially depends on this step size, when we choose the inaccuracy vector δ as at the end of Section 5-4.

When we are applying the threshold value d , then the QFD does not satisfy criterion 6. The threshold value based on only the radius is not sufficient, because when a capacity vector has a much smaller length than another capacity vector but more or less the same direction, then the distance (QFD) between these vectors will be greater than the threshold d . Therefore, we have considered the angle between the stress tests as well.

This principle can be illustrated with the H-shaped network of Section 2-1. Suppose we also have a less severe transport situation than the first stress test u_1 , for example $v_1 = (50, 0, 0, -50)$. The distance between these two stress tests is then

$$\begin{aligned} QFD_{\mathbf{A}}(u_1, v_1) &= \sqrt{\begin{pmatrix} 50 & 0 & 0 & -50 \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 1 & 0 \\ 0 & \frac{1}{3} & 0 & 1 \end{pmatrix} \begin{pmatrix} 50 \\ 0 \\ 0 \\ -50 \end{pmatrix}} \\ &= \sqrt{\begin{pmatrix} 50 & -\frac{50}{3} & \frac{50}{3} & -50 \end{pmatrix} \begin{pmatrix} 50 \\ 0 \\ 0 \\ -50 \end{pmatrix}} = \sqrt{5000} \approx 70.71 > \frac{4}{3}\sqrt{3}. \end{aligned}$$

However, if we determine the angle between the two vectors, like is defined in Equation (5-1), then the angle between these vectors is zero. Thus, if we apply threshold value φ , then the QFD satisfies criterion 6.

So, the quadratic form distance is indeed a suitable (dis)similarity measure considering the above criteria. Unfortunately, we could not prove in general that the ‘original’ correlation matrix, $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$, is positive semidefinite on the subspace of balanced vectors, like is stated for color histograms in the article *Efficient Color Histogram Indexing for Quadratic Form Distance Functions*. In addition, the writers of this article claim that a proof is given in an unpublished internal note and thus, that the ‘original’ correlation matrix is positive semidefinite on the subspace of balanced vectors, when an Euclidean ground distance is applied. We could not take a look at this proof and therefore, we have tried to come up with our own proof in which we chose the applied ground distance d_{ij} as the length of the pipes between two network points i and j . During this process, we discovered a counterexample with the help of Jacob van der Woude instead of a proof.

Therefore, we restricted our experiments regarding proper correlation matrices on the subspace of balanced vectors to a few GTS networks. The correlation matrix depending on the length and diameter of the pipelines seems very suitable for the two smaller GTS networks and the ‘original’ correlation matrix seems only to be suitable when the network does not consist of crossing pipelines according to our experiments. Thus, the quadratic form distance with the correlation matrix depending on the length and diameter of the pipelines seems to be a proper method to use during the reducing process in practice.

When considering the experiments performed during this project, the best way to reduce a generated set of stress tests is to apply the following steps:

1. Start with the vector with the greatest length, say x .
2. Calculate the angle φ between this vector and the boundary of its cone defined by

$$\varphi = \max_{\delta} \left\{ \arccos \left(\frac{x^T \mathbf{A}(x + \delta)}{\sqrt{x^T \mathbf{A}x} \sqrt{(x + \delta)^T \mathbf{A}(x + \delta)}} \right) \right\},$$

with $\delta = (\delta_1, \delta_2, \dots, \delta_n)$, whereby $\delta_i = 0$ for all components, except the ones mentioned at the end of Section 5-4: $|\delta_j| \leq 10$ for $j = 19, 35, 43, 54, 55, 56, 57, 58, 60, 62, 63$ and $|\delta_{51} + \delta_{52}| \leq 10$. Hereby the correlation matrix \mathbf{A} depends on the length and diameter of the pipelines.

3. Determine the angle θ between this vector x and the other vectors in the set.
4. Apply the condition $\theta \leq \varphi$ to determine which vectors lie within the cone of x . These vectors are called almost similar to x .
5. Delete these vectors which are almost similar to this longest vector x from the set.
6. Consider the next longest vector y ($y \neq x$) of the remaining set and apply steps 2 - 6 till the remaining set is empty.

6-2 Future research

The recommendations for future research aim at further development of the used method, the QFD, in particular its correlation matrix or the posed reducing algorithm.

I would recommend to apply these steps of reducing on the detailed network of GTS in order to test this reducing process some more. This means that an algorithm or a method needs to be found such that the diameter can be taken into account as well. This is not done during this project, because the network points with no entry, exit points or storages linked to it pose a problem. We will call these network points ‘dummy points’ for simplicity. We cannot derive the needed information, the diameter of the pipelines between entry, exit points and storages, directly from *MCA*, but we have to adjust the network such that these dummy points are no longer part of the network. Another option is to put ‘dummy entries/exits’, which are entry/exit points with capacity zero, at these dummy points and then extract the information regarding the diameter of the pipelines. Then a new problem arises, because the distance between these dummy entries/exits and the other entry, exit points and storages need to be determined.

In addition, I recommend to investigate whether or not the first step of the posed reducing algorithm is the right one. It is possible that a different or smaller set can be found if we start with a different vector which has not the greatest length. So, the question is: does this posed reducing algorithm results in a minimal set, or can a different algorithm be found such that we end up with a smaller set?

I would also recommend to examine the threshold value φ . During this project, it is assumed that we can draw a region consisting of a cone and sphere both with radius d to indicate which stress tests can be marked the same, and that the angle between a capacity vector and its cone equals φ defined as in the posed reducing algorithm. A conjecture is that this definition results in a n -dimensional cube and not in a sphere (in n dimensions). Therefore, it seems a good idea to examine the definition of φ and its fit with respect to this region. It could also be wise to reconsider the shape of this region.

We have experimented with the network drawn in Figure 3-6 by deleting the G-gas pipelines which are crossing a pipeline of the H-gas network. The part of the G-gas network that is overlapping the H-gas network is kept. Then we have determined the ‘original’ correlation matrix defined as $\mathbf{A}_{ij} = 1 - \frac{d_{ij}}{d_{max}}$, calculated its eigenvalues and concluded that the new matrix is PSD. This result motivates me to recommend to extend this experiment to see the effect of adjusting the considered network (by deleting some of the pipelines) such that it can be drawn (topological) in two dimensions. The same experiment can for example be conducted on the detailed network of GTS displayed in Figure 4-2.

A next step can be to translate this network property (no crossing pipelines of to draw in two dimensions) into an additional condition besides the three conditions from Section 3-2, $h_{ii} = 0$, $h_{ij} = h_{ji}$ and $h_{ij} \leq h_{ik} + h_{kj}$, for the matrix $\mathbf{H} = [h_{ij}]$ to ensure that $\mathbf{z}^T \mathbf{H} \mathbf{z} \leq 0$ subject to $\sum_i z_i$ holds. If this is possible, then it is wise to examine if the counterexample displayed in Figure 3-4 is still a counterexample with the new condition(s) added.

During this project, we have only considered stress tests based on the transportation load. So, it could be good to involve the blending load as well. It could be wise to consider first only the blending load and then find a way to combine both methods in order to come up with a method for stress tests depending on both the transportation and blending load.

Appendix A

Figures

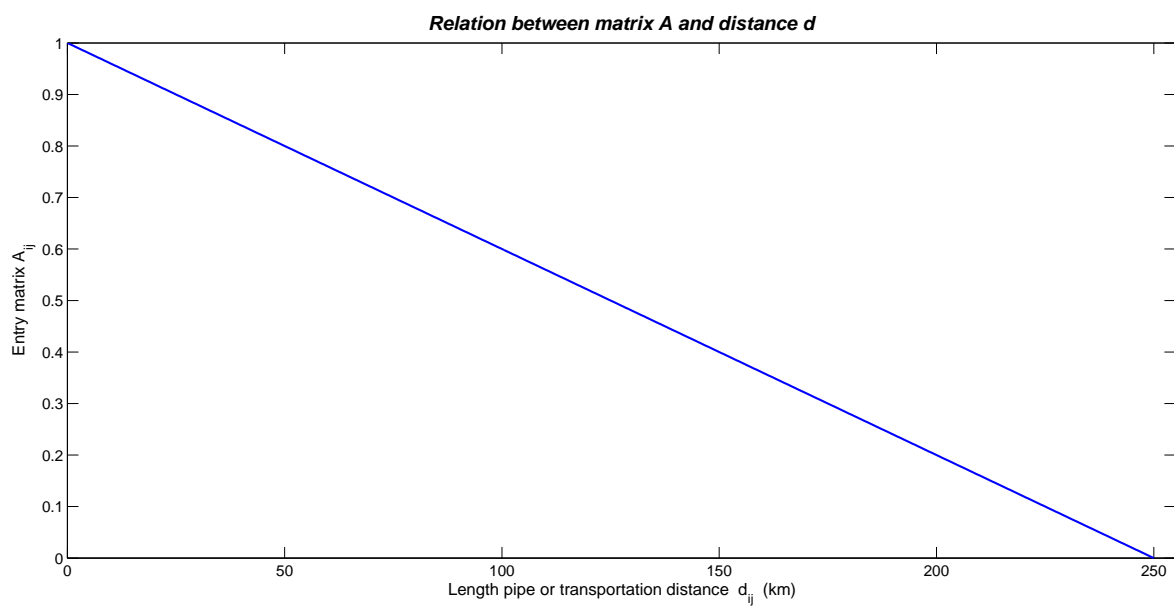


Figure A-1: The relation between entry A_{ij} and transportation distance with $d_{max} = 250$.

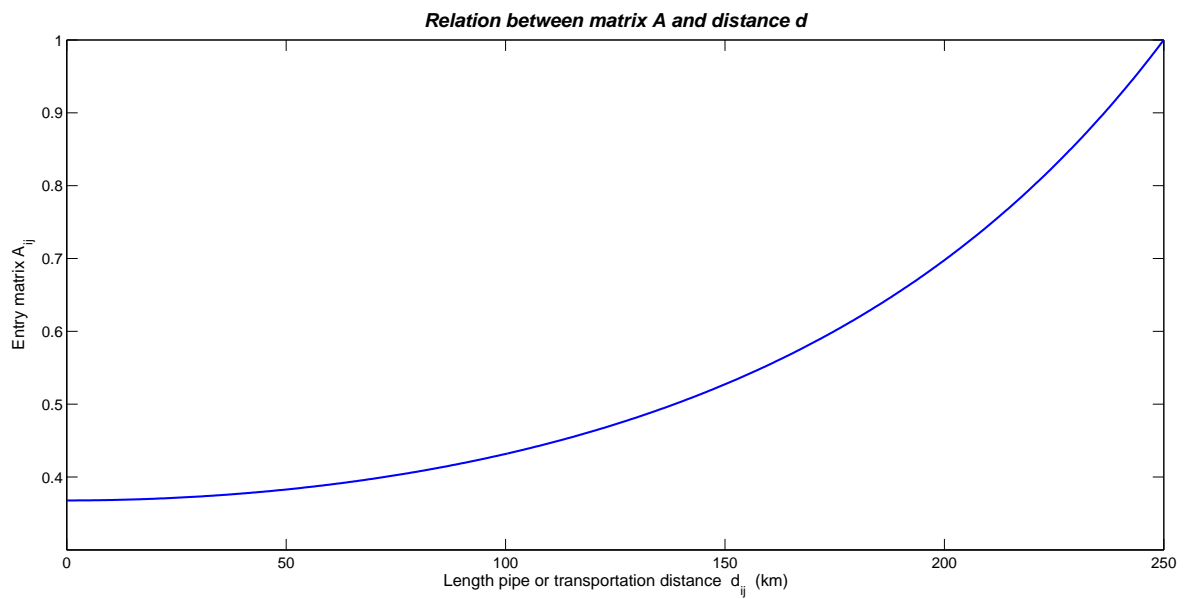


Figure A-2: The relation between entry A_{ij} and transportation distance with $d_{max} = 250$.

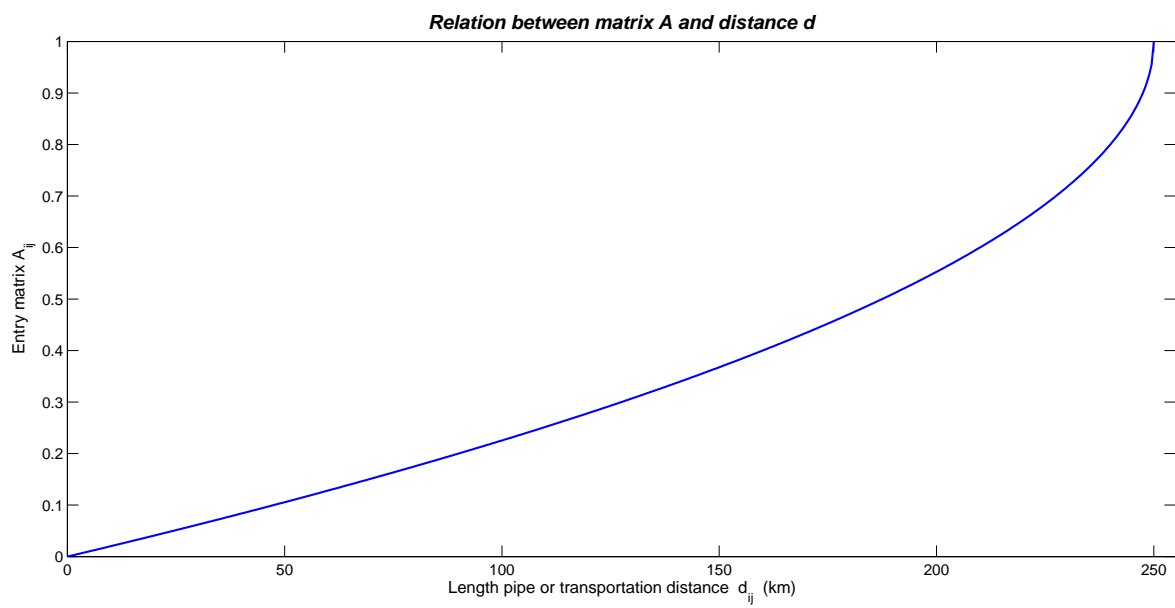


Figure A-3: The relation between entry A_{ij} and transportation distance with $d_{max} = 250$.

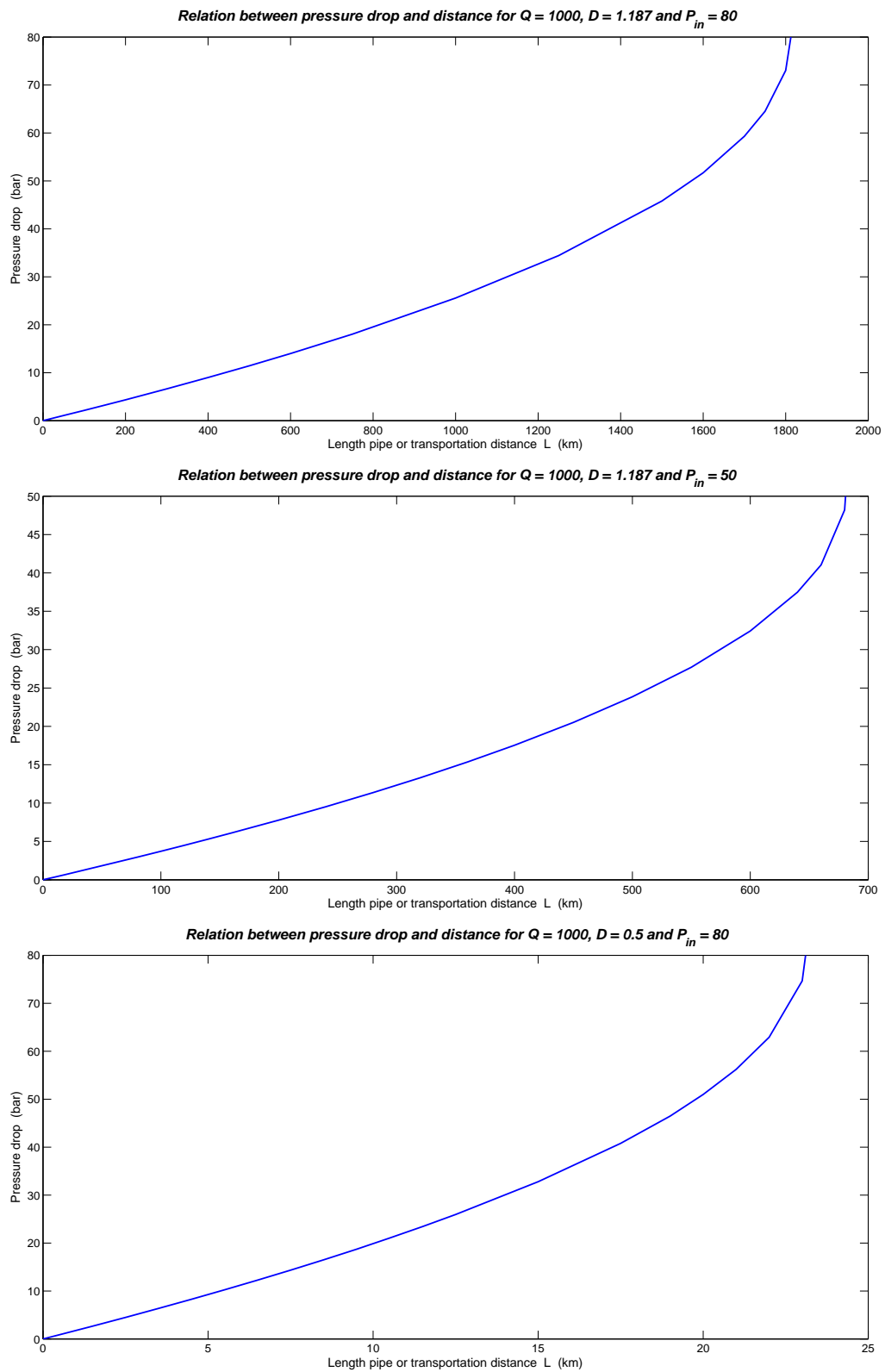


Figure A-4: The relation between pressure drop and transportation distance for $Q = 1000 \text{ dam}^3/\text{h}$.

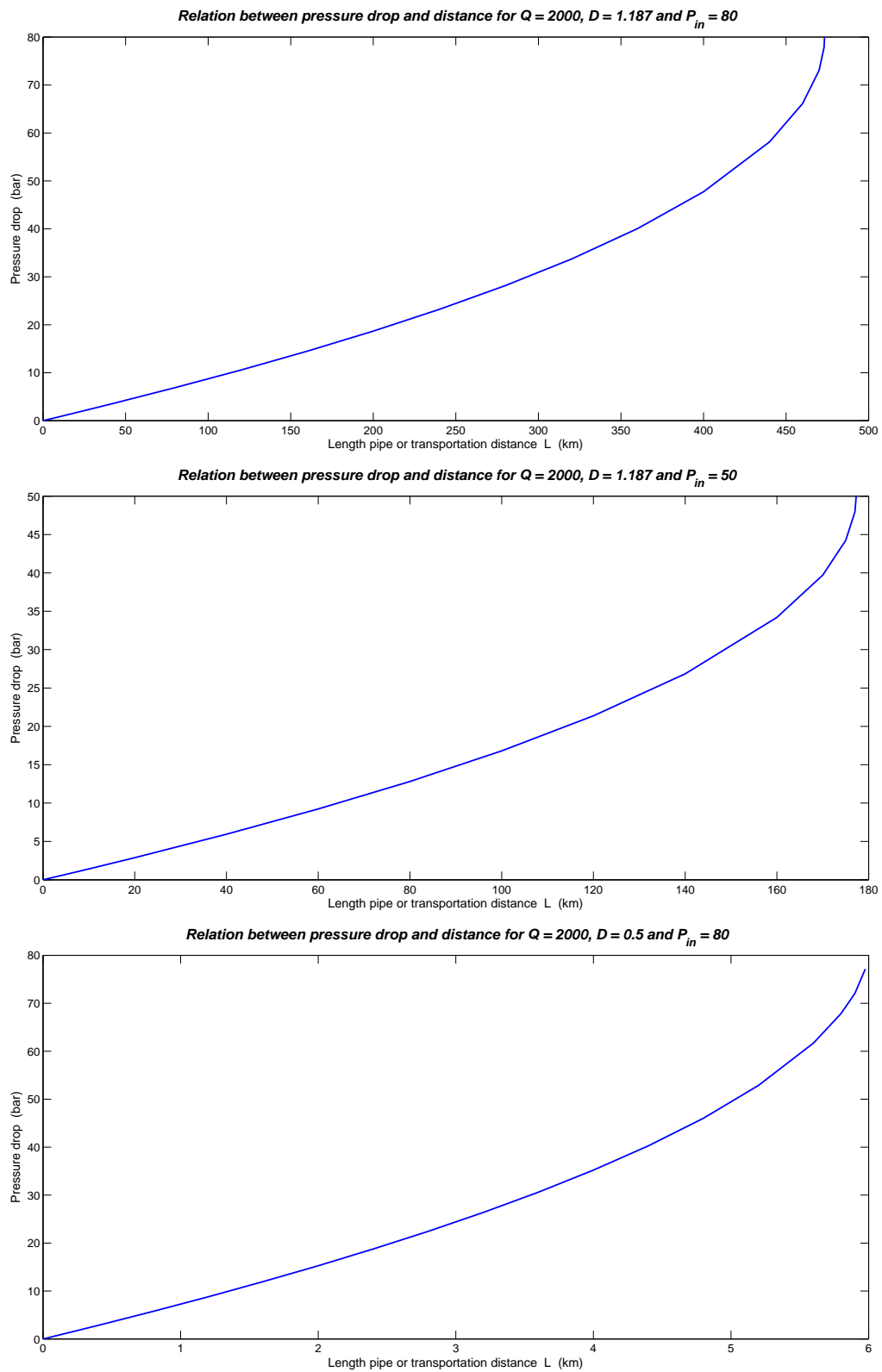


Figure A-5: The relation between pressure drop and transportation distance for $Q = 2000$ dam^3/h .

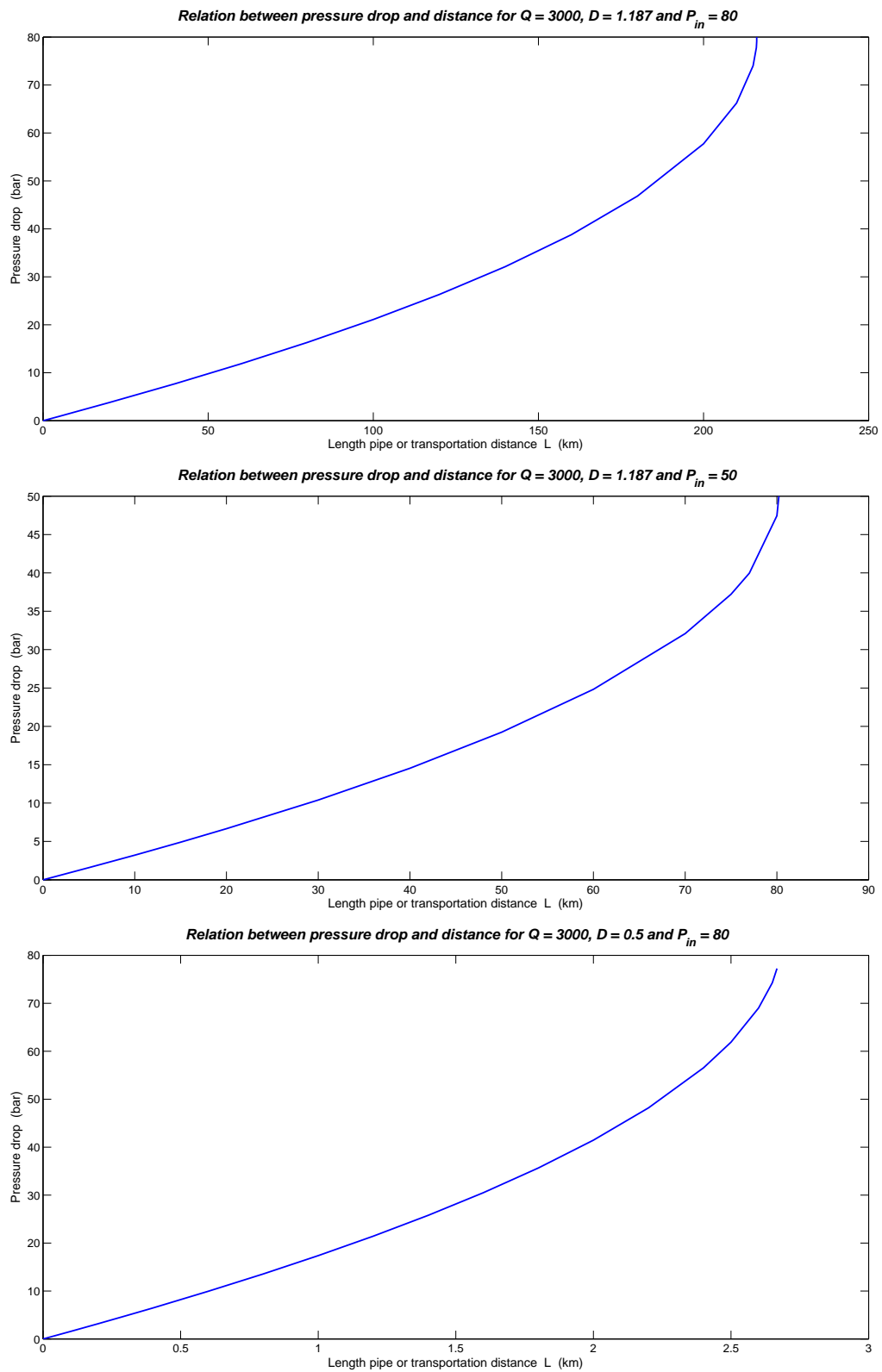


Figure A-6: The relation between pressure drop and transportation distance for $Q = 3000$ dam^3/h .

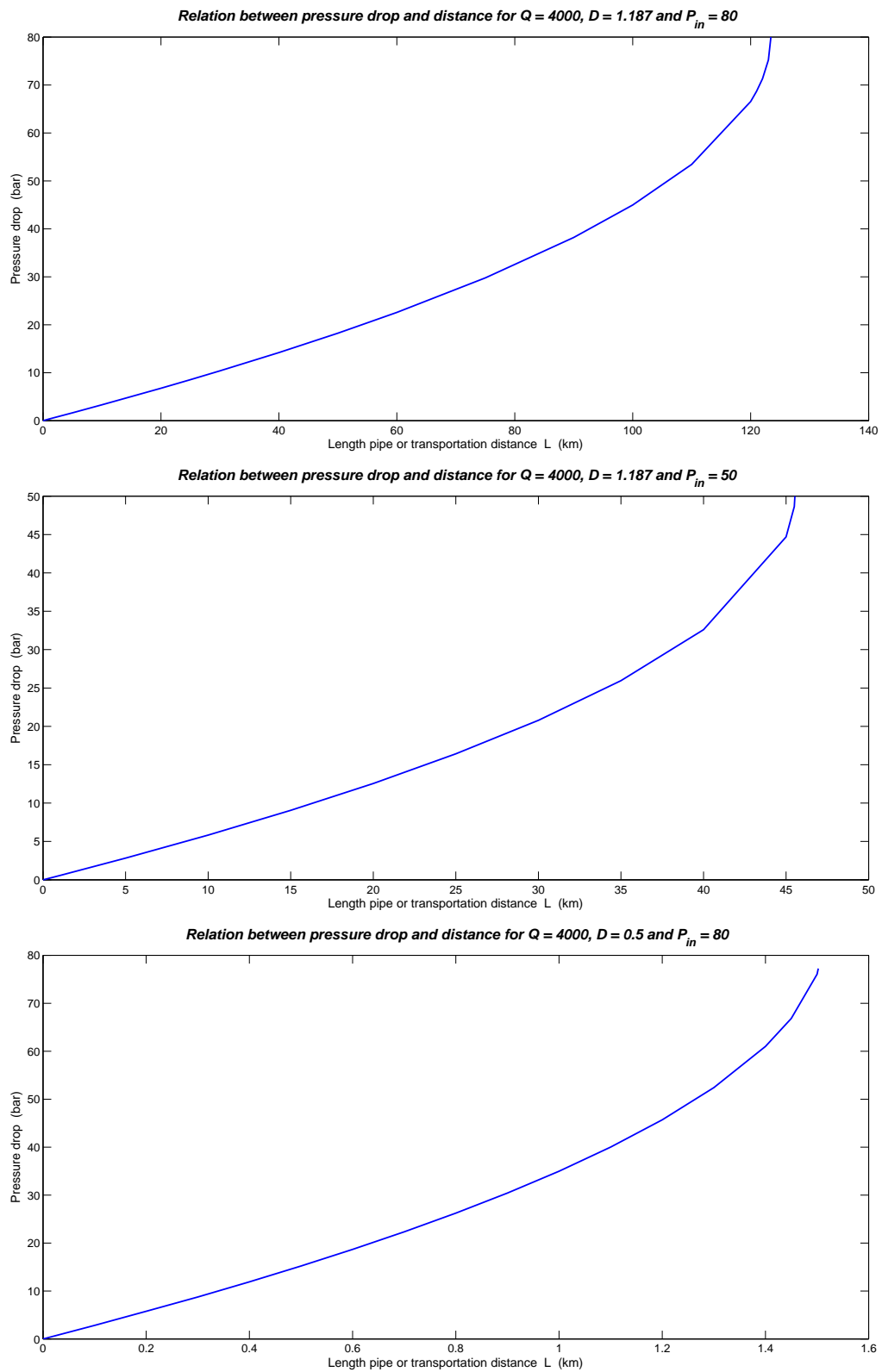


Figure A-7: The relation between pressure drop and transportation distance for $Q = 4000$ dam^3/h .

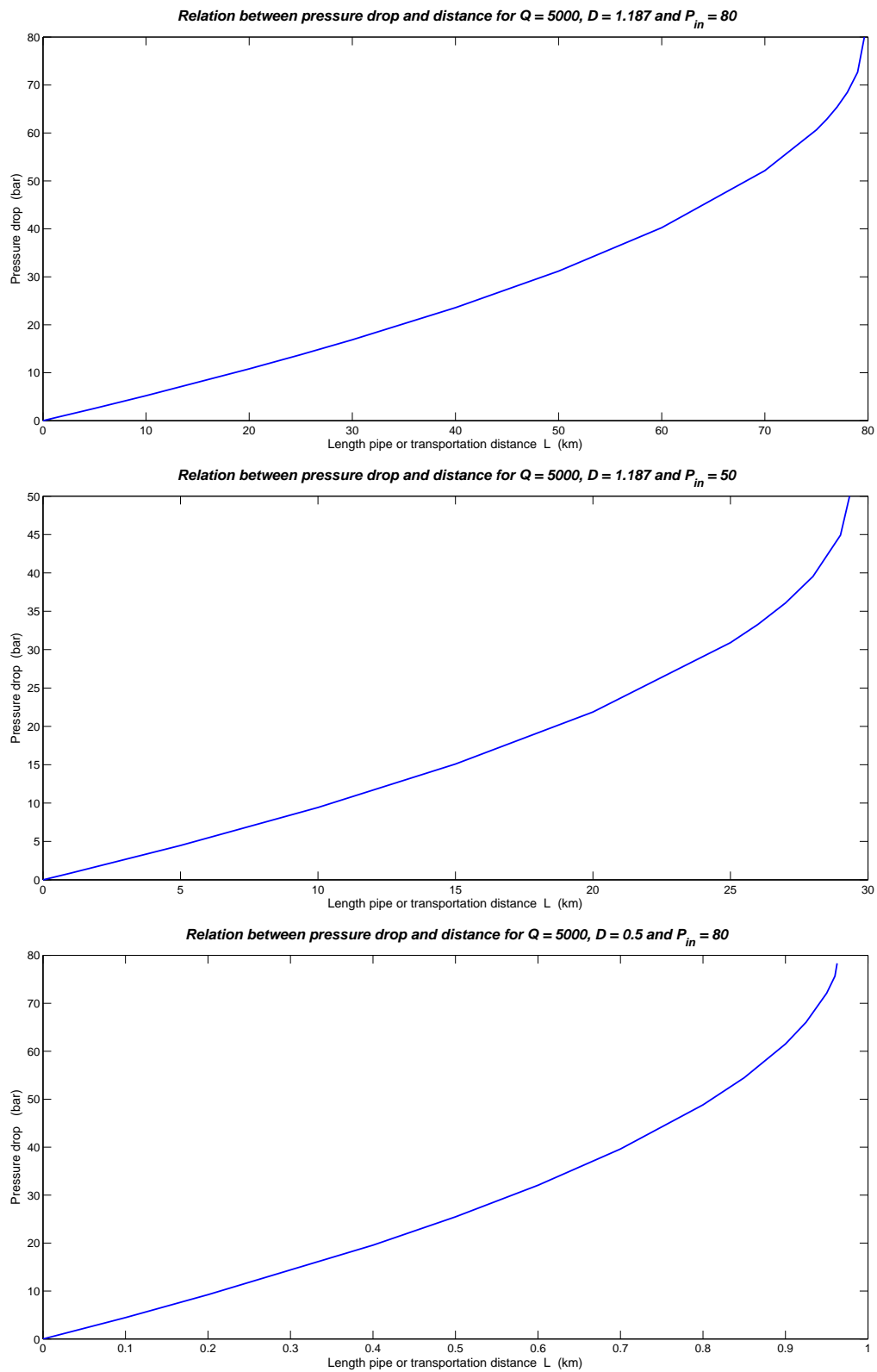


Figure A-8: The relation between pressure drop and transportation distance for $Q = 5000$ dam^3/h .

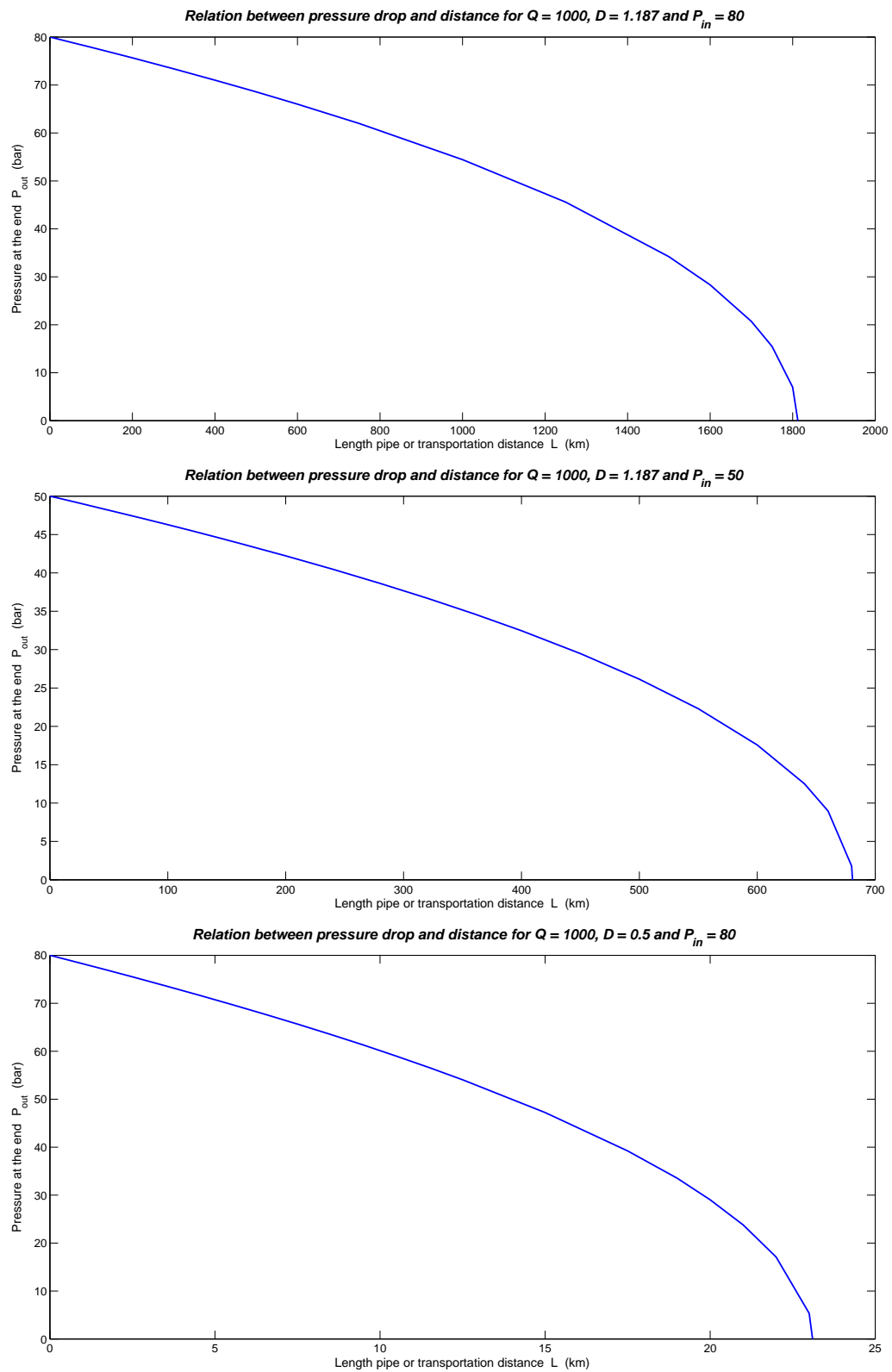


Figure A-9: The relation between pressure at the end of a pipe and transportation distance for $Q = 1000 \text{ dam}^3/\text{h}$.

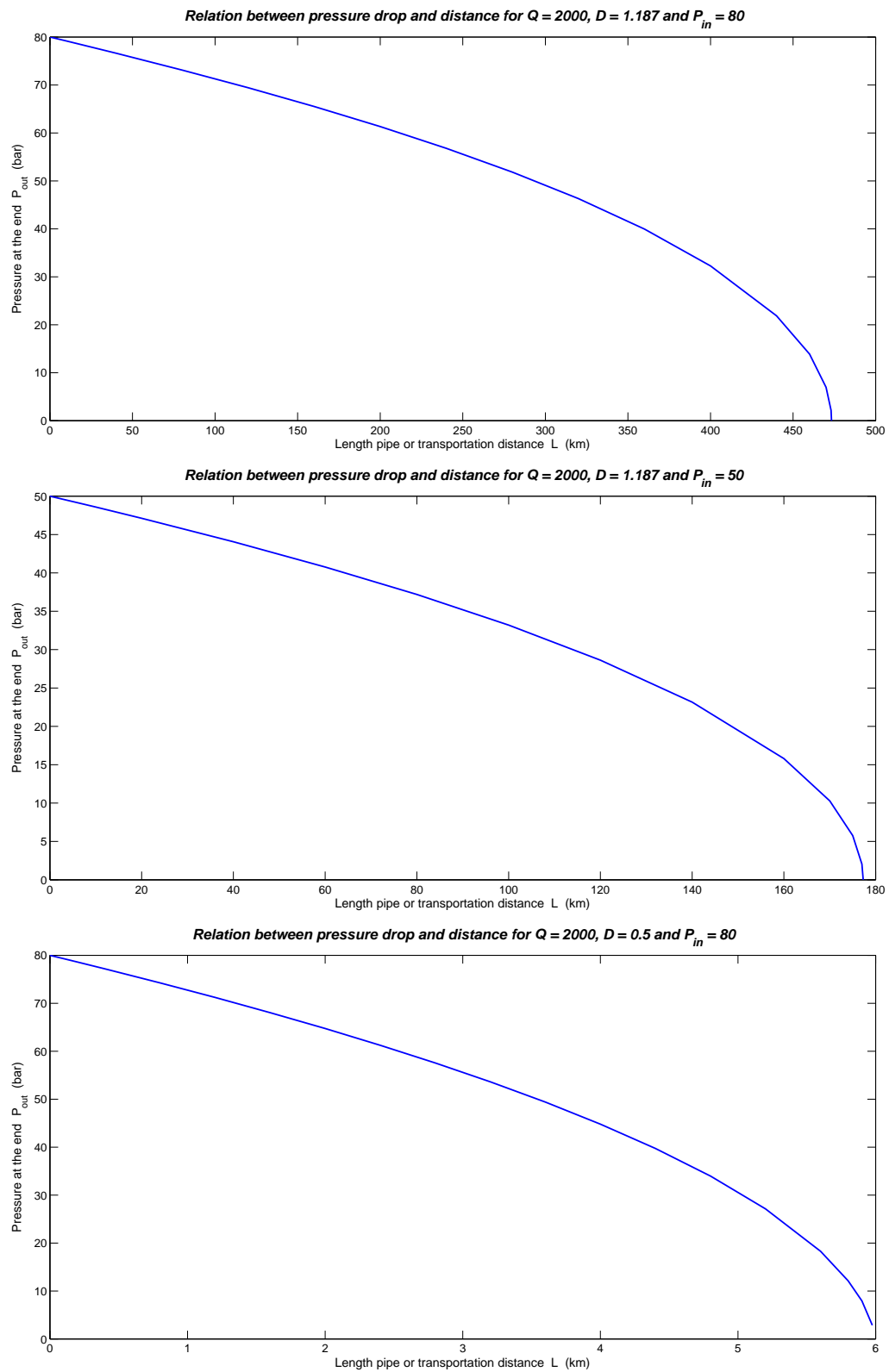


Figure A-10: The relation between pressure at the end of a pipe and transportation distance for $Q = 2000 \text{ dam}^3/\text{h}$.

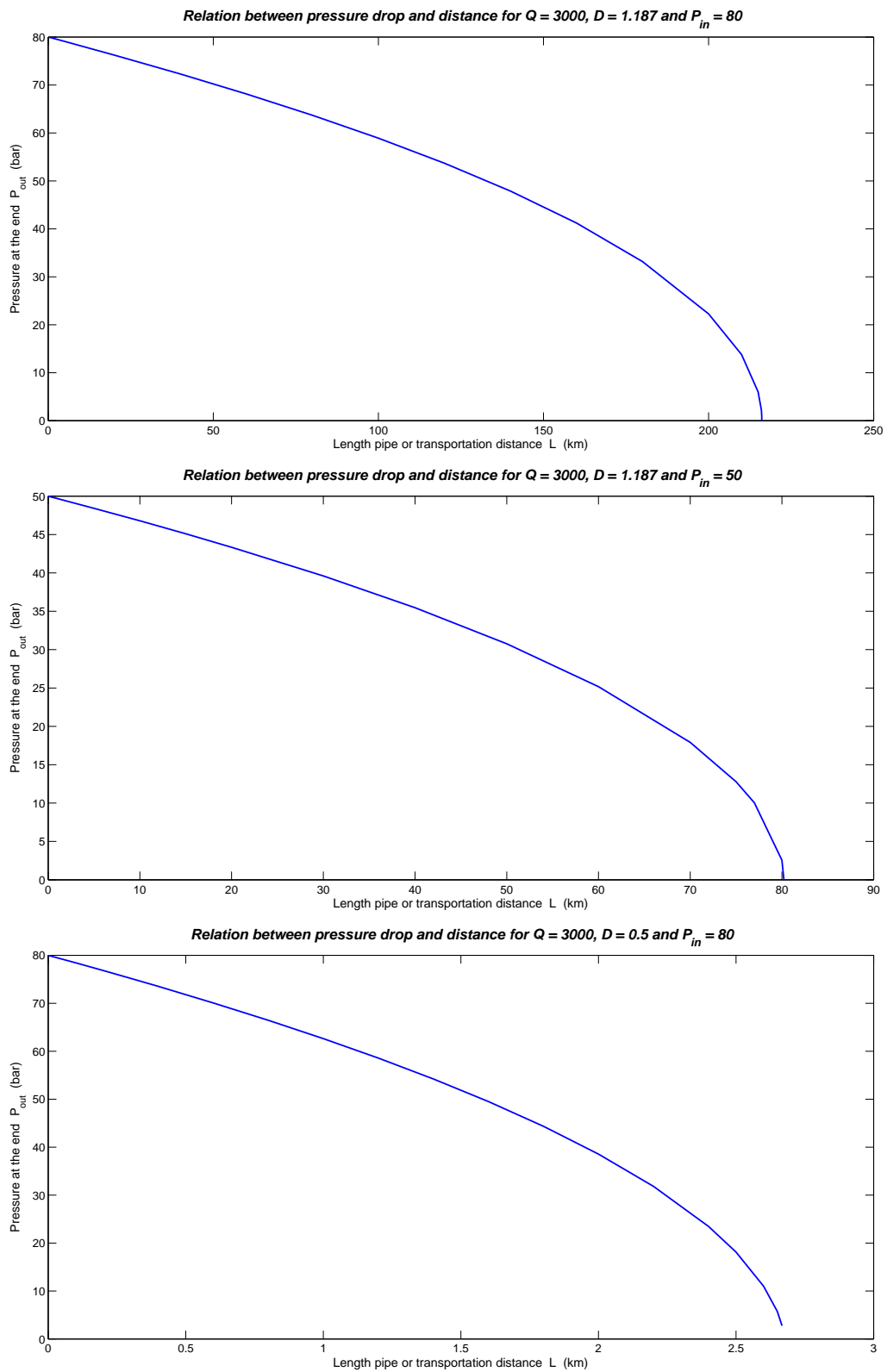


Figure A-11: The relation between pressure at the end of a pipe and transportation distance for $Q = 3000 \text{ dam}^3/\text{h}$.

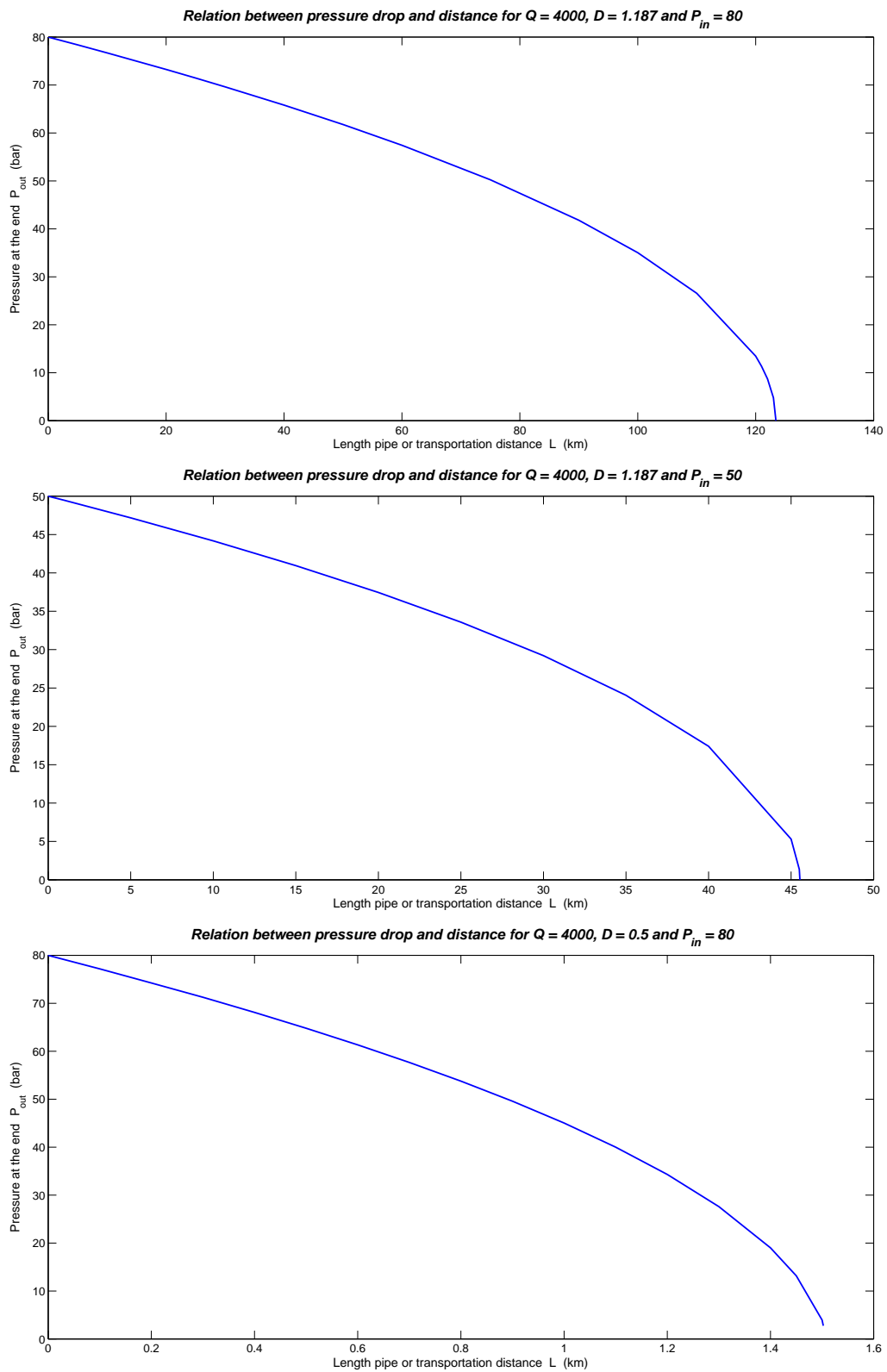


Figure A-12: The relation between pressure at the end of a pipe and transportation distance for $Q = 4000 \text{ dam}^3/\text{h}$.

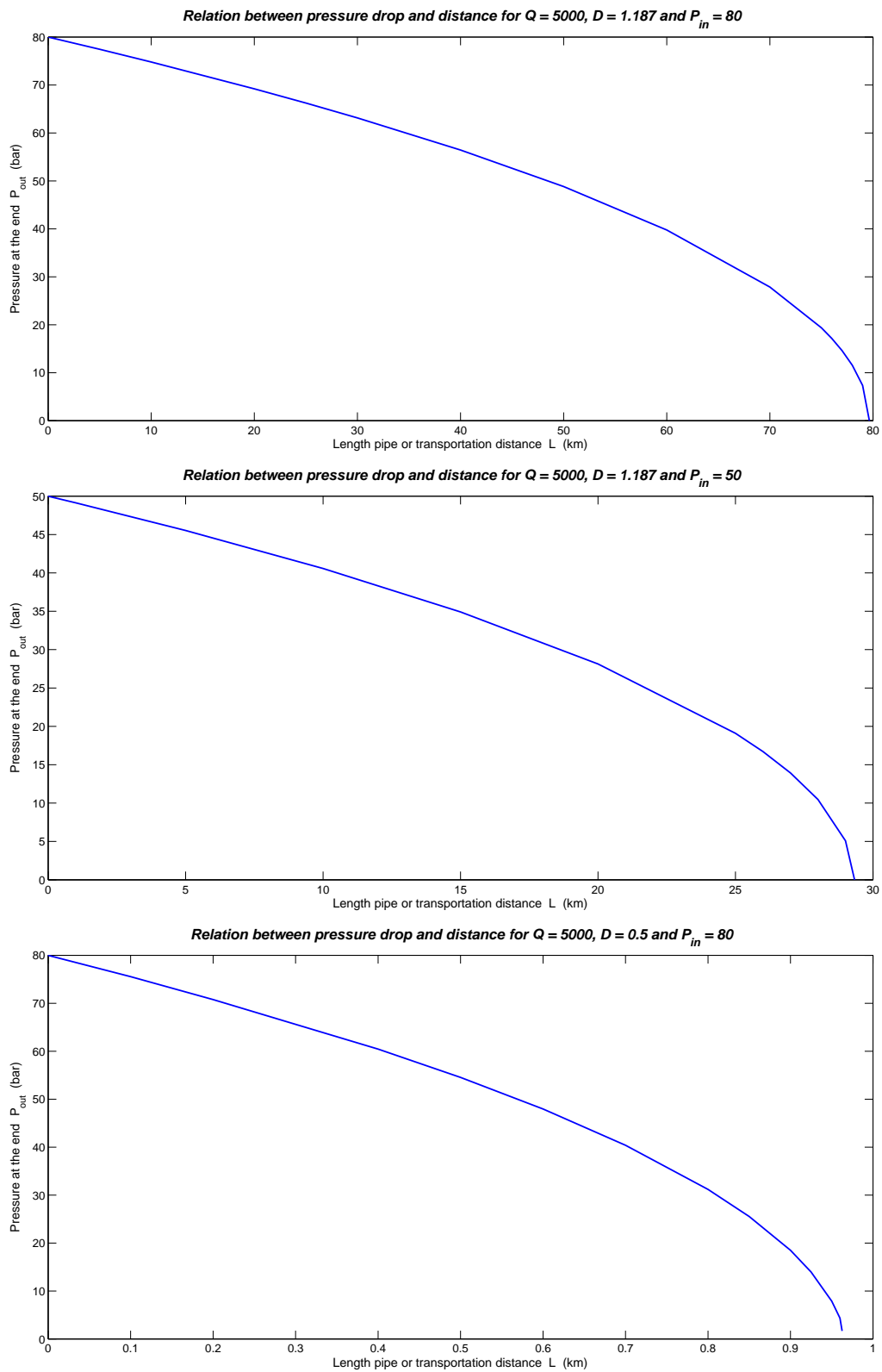


Figure A-13: The relation between pressure at the end of a pipe and transportation distance for $Q = 5000 \text{ dam}^3/\text{h}$.

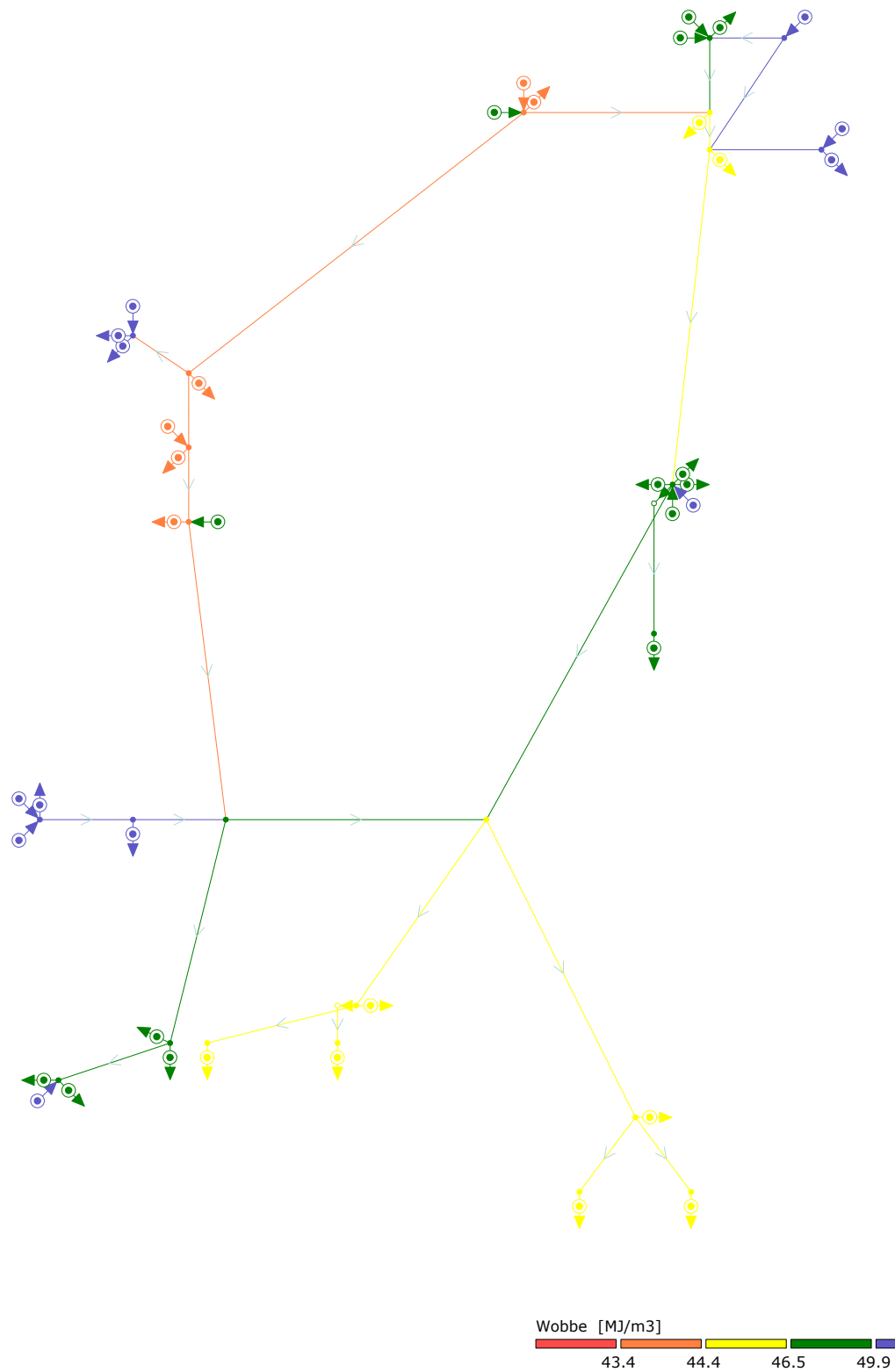


Figure A-14: The Wobbe distribution of the first shipping variant of the H-gas network of GTS.

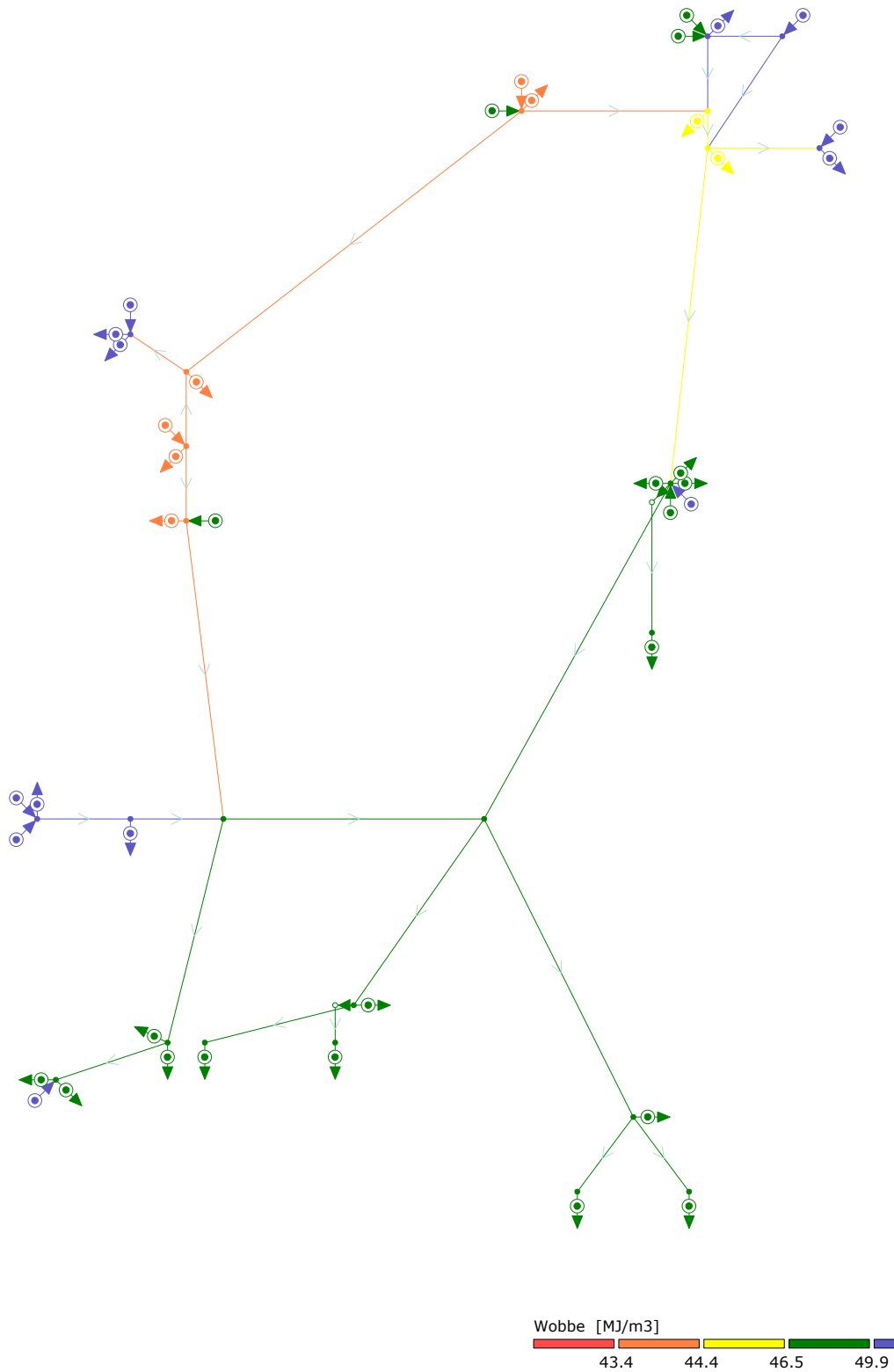


Figure A-15: The Wobbe distribution of the second shipping variant of the H-gas network of GTS.

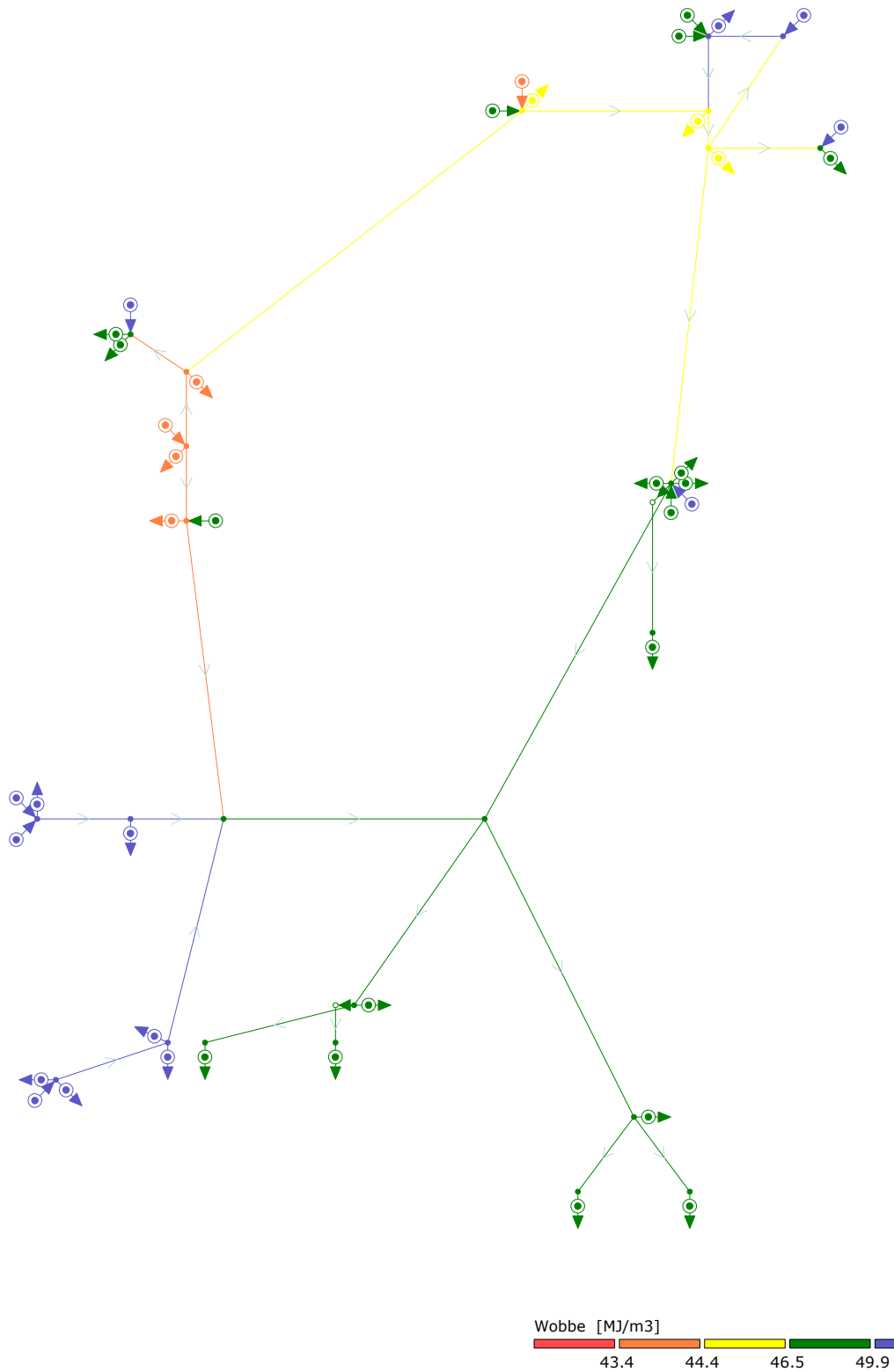


Figure A-16: The Wobbe distribution of the third shipping variant of the H-gas network of GTS.

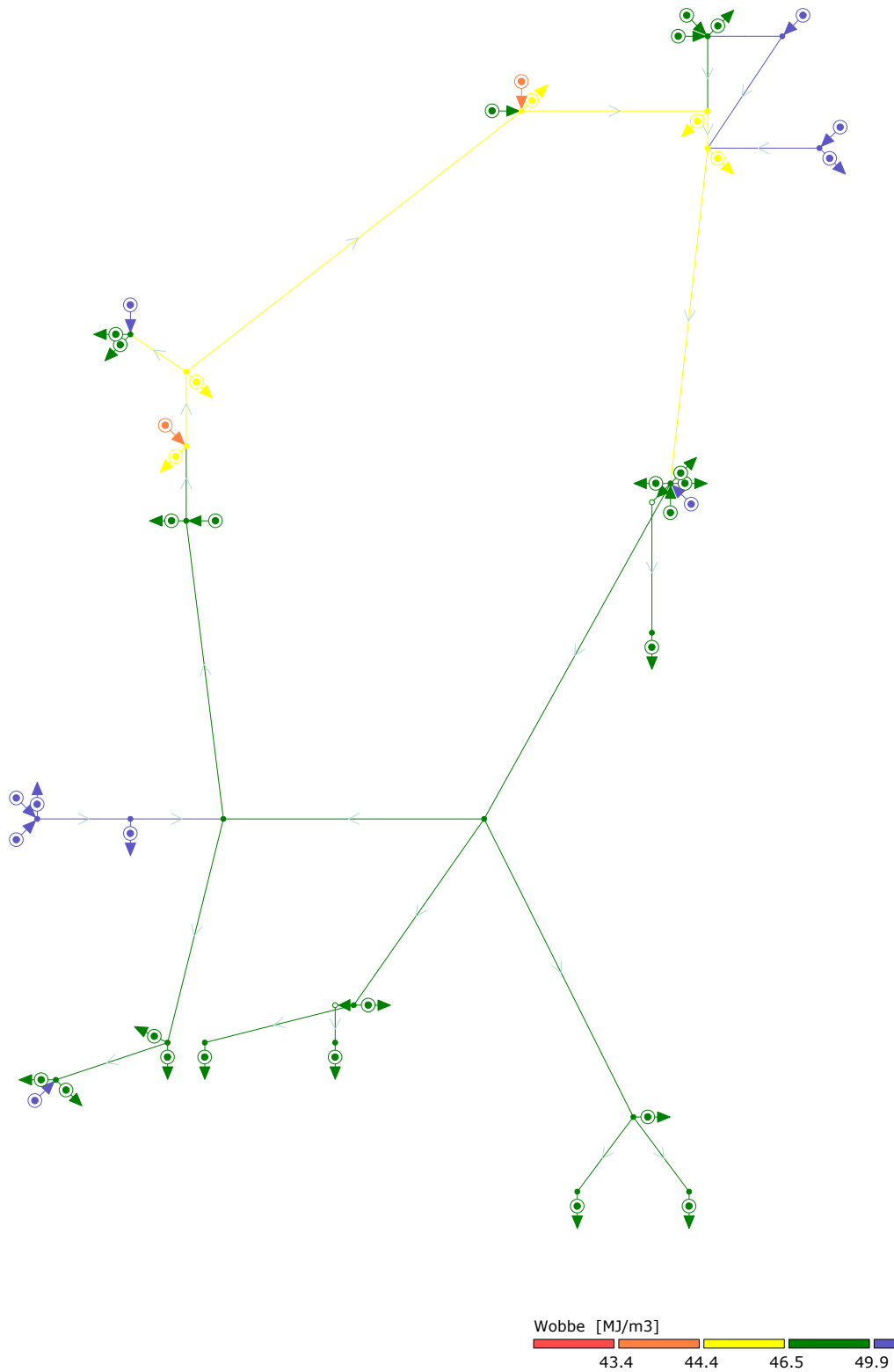


Figure A-17: The Wobbe distribution of the fourth shipping variant of the H-gas network of GTS.

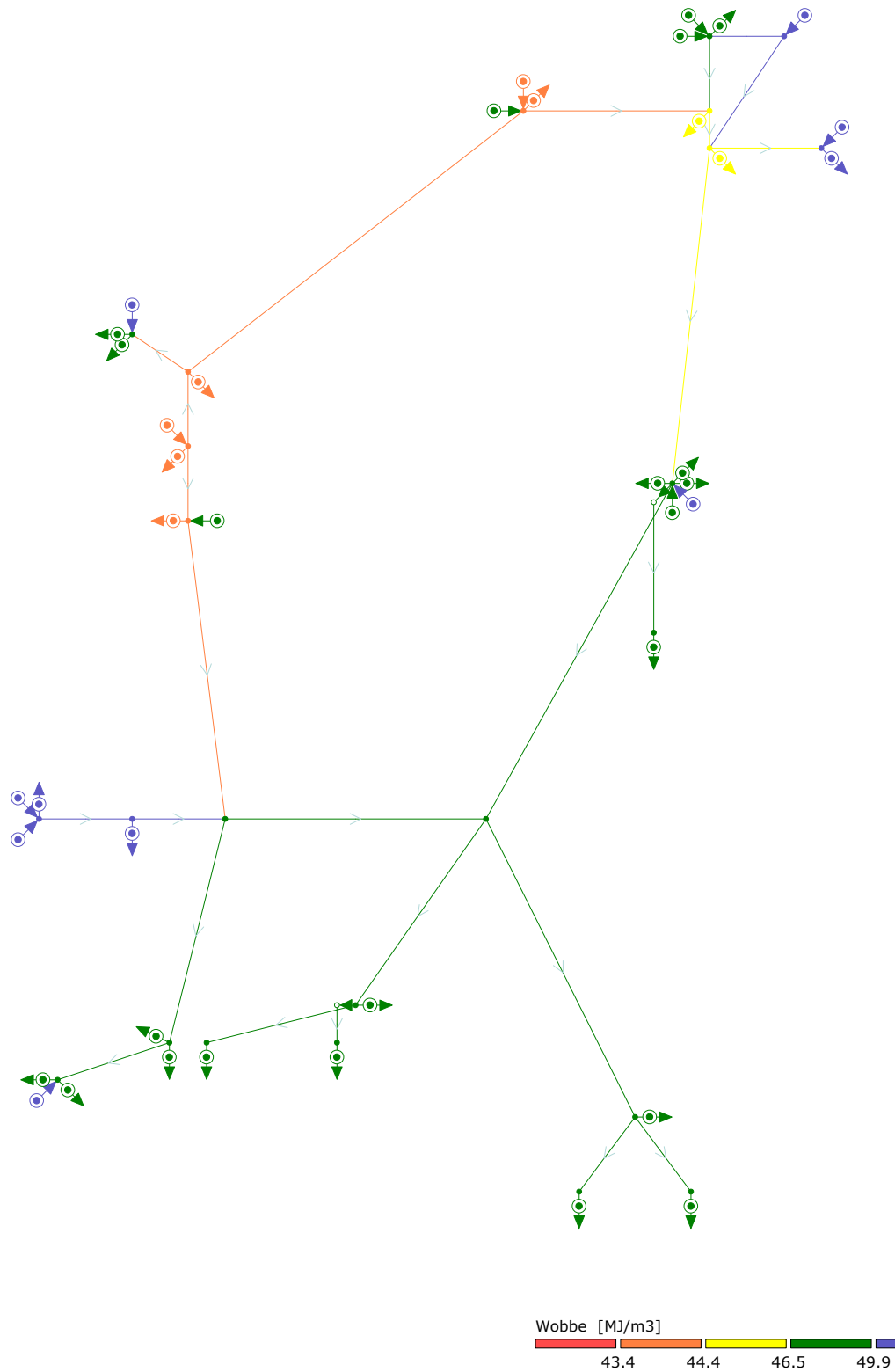


Figure A-18: The Wobbe distribution of the fifth shipping variant of the H-gas network of GTS.

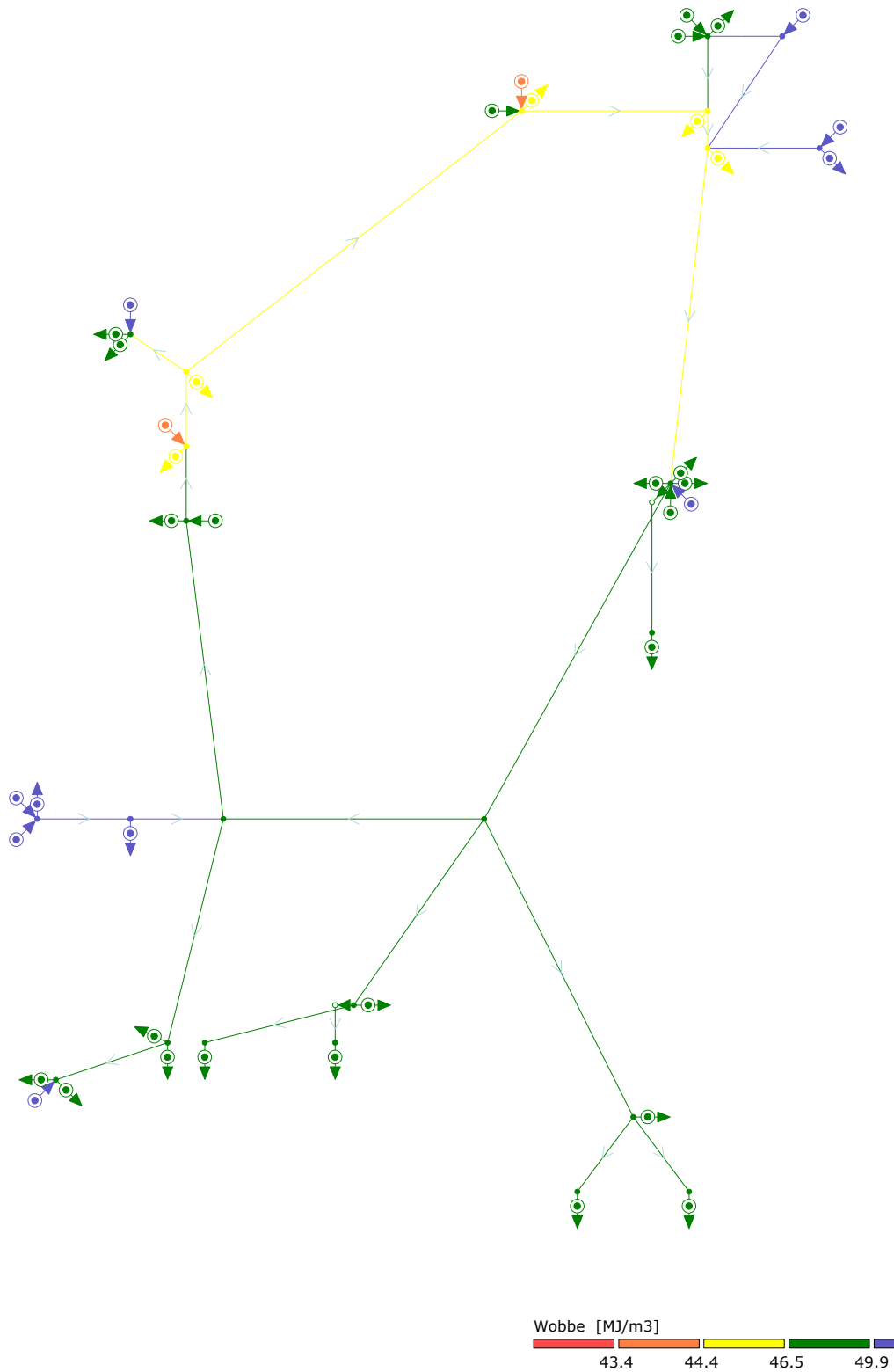


Figure A-19: The Wobbe distribution of the sixth shipping variant of the H-gas network of GTS.

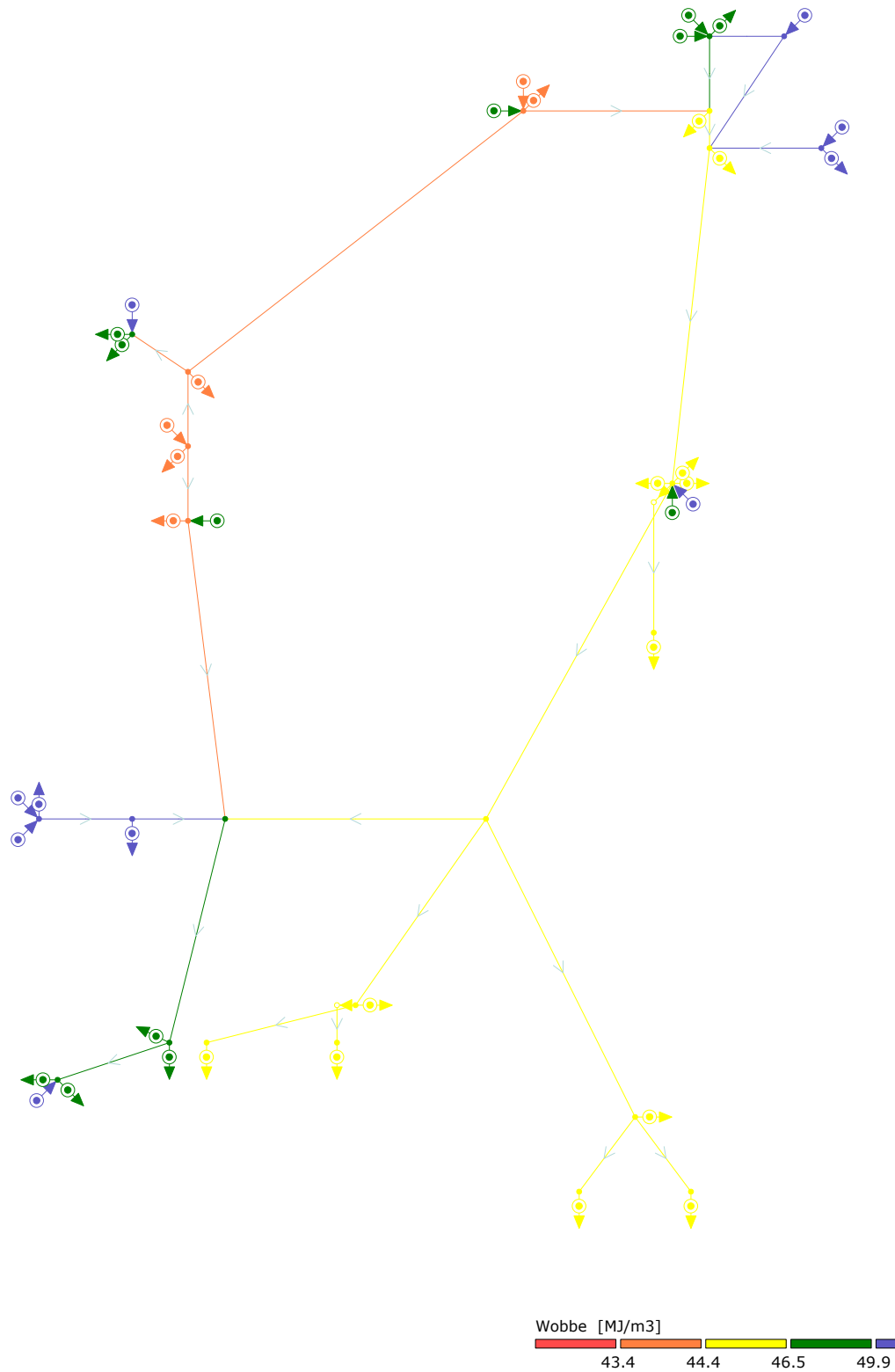


Figure A-20: The Wobbe distribution of the seventh shipping variant of the H-gas network of GTS.

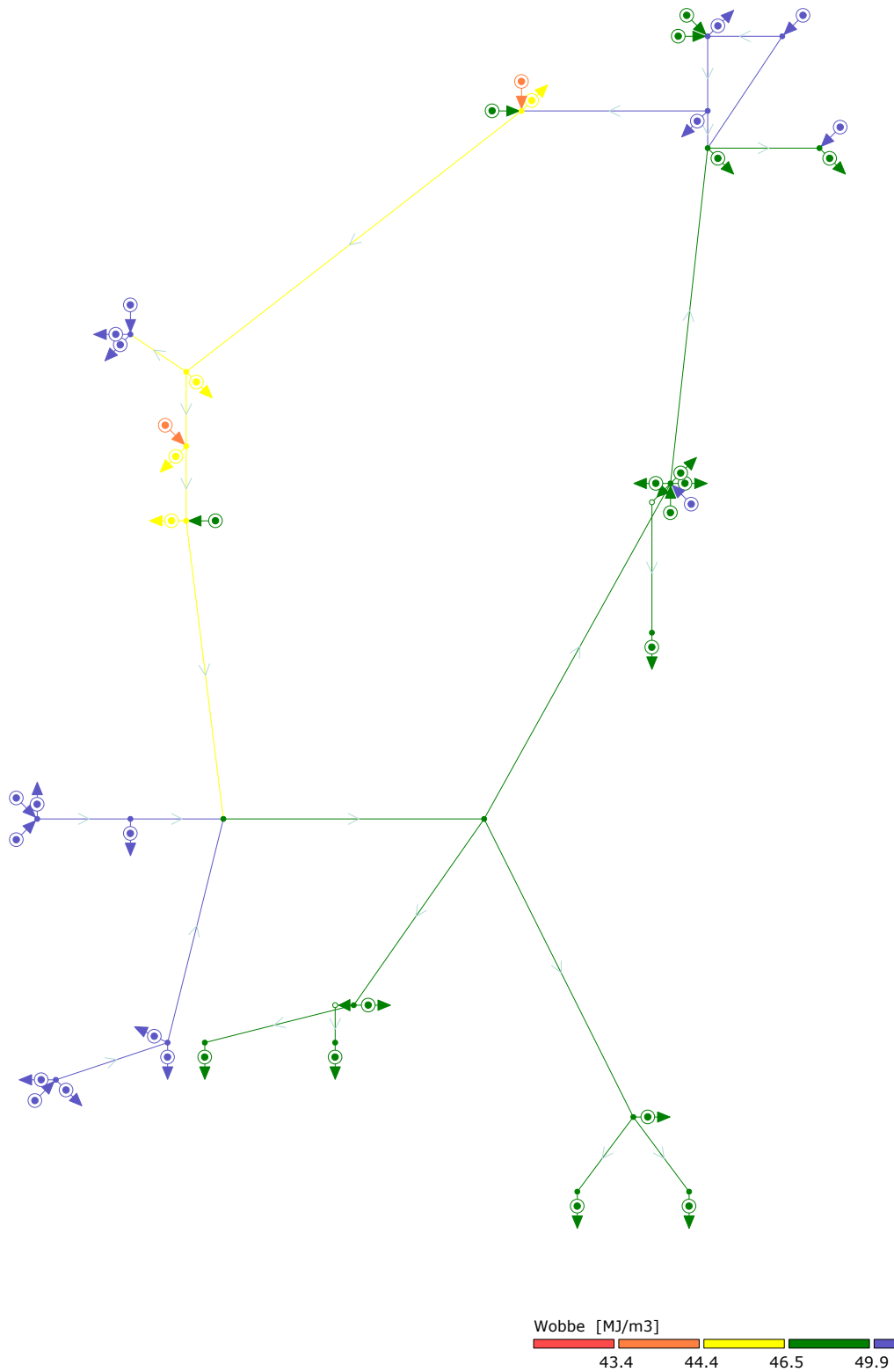


Figure A-21: The Wobbe distribution of the eighth shipping variant of the H-gas network of GTS.

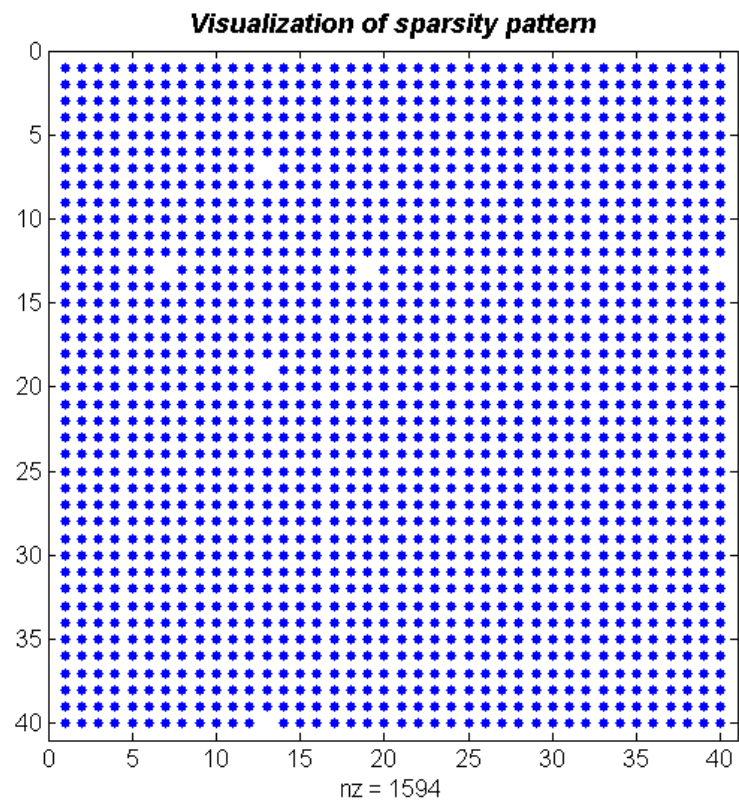


Figure A-22: The visualization of the sparsity pattern of the matrix \mathbf{A} , applied on the H-gas network of GTS.

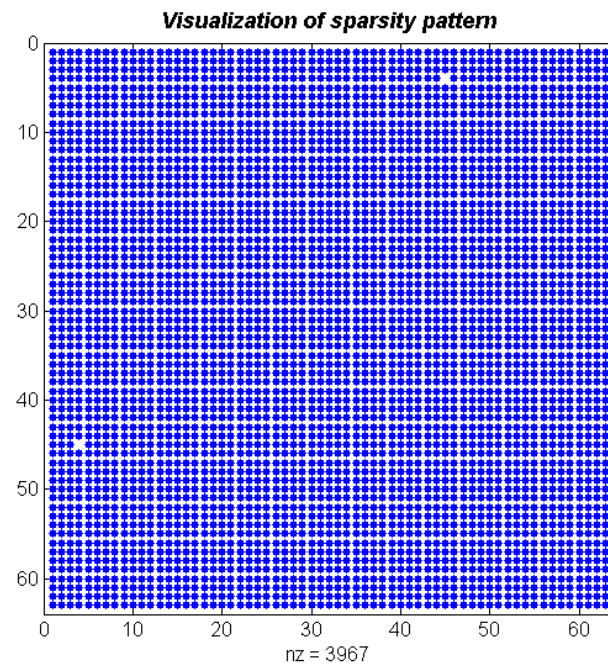


Figure A-23: The visualization of the sparsity pattern of the matrix A , applied on a simplified network of GTS.

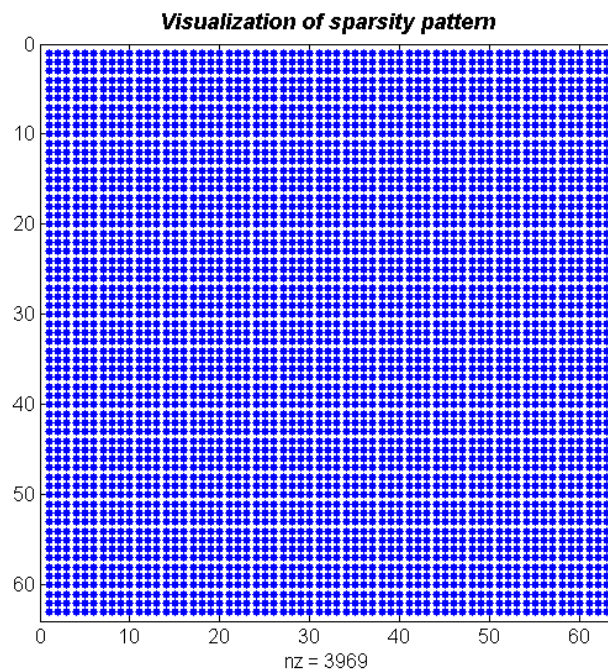


Figure A-24: The visualization of the sparsity pattern of the two new defined matrices, applied on a simplified network of GTS.

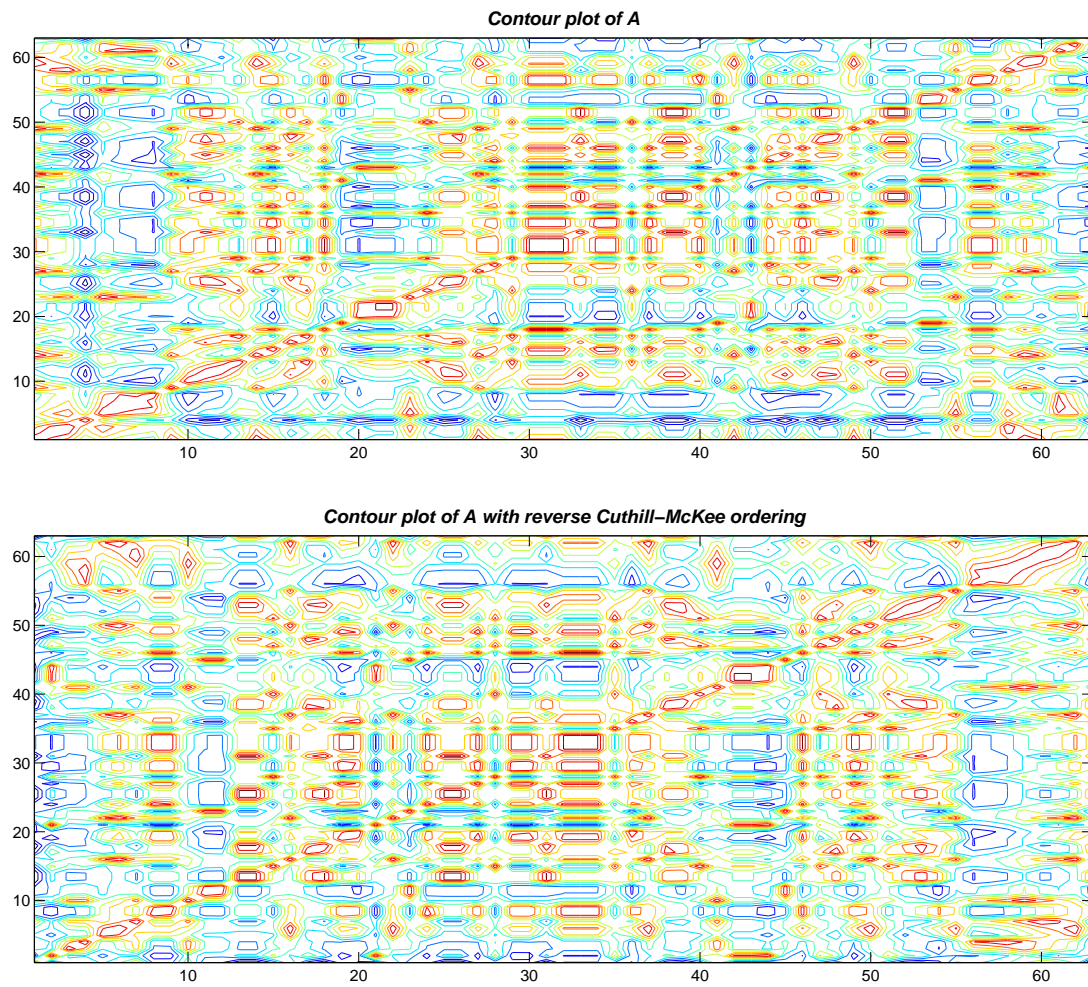


Figure A-25: The contour plot of the original matrix and of the matrix on which the reverse Cuthill-McKee ordering is applied.

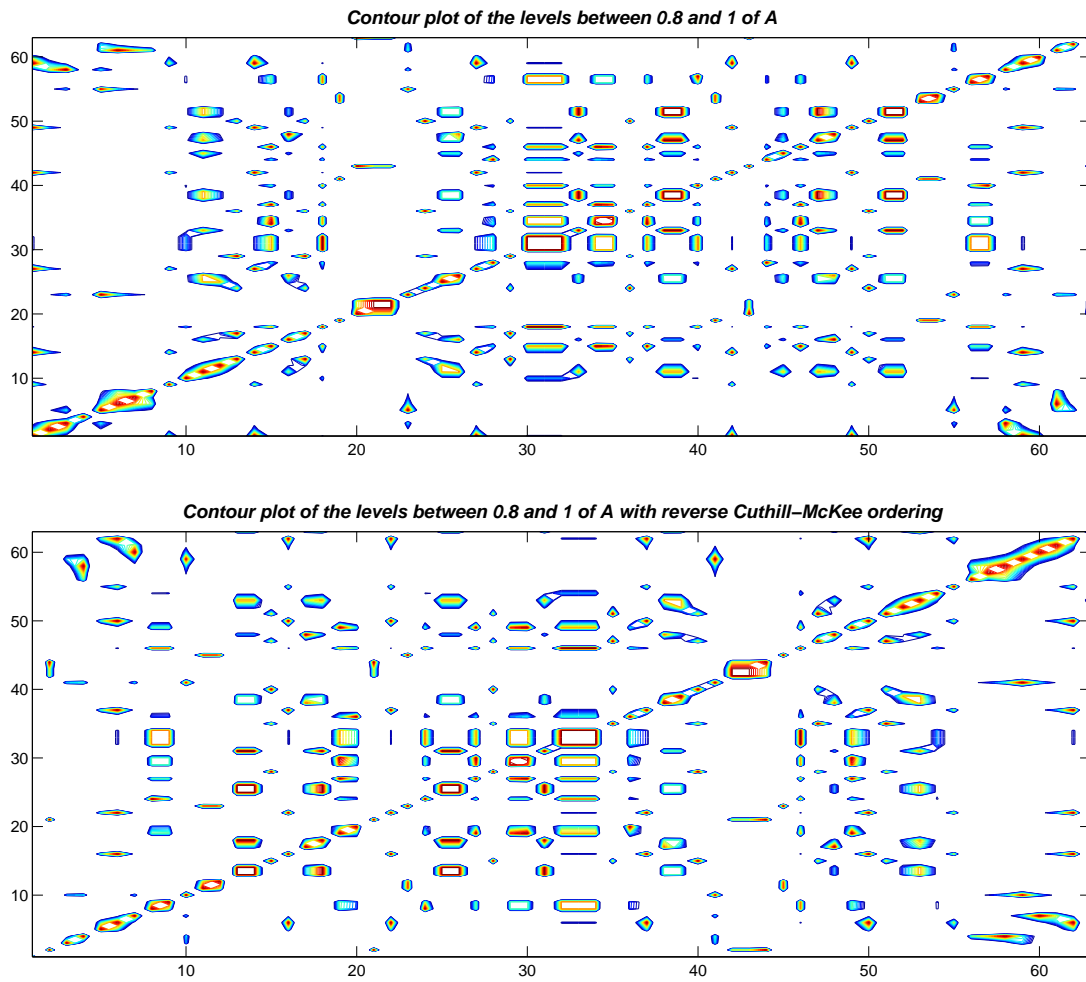


Figure A-26: The contour plot of the levels between 0.8 and 1 of the original matrix and of the matrix on which the reverse Cuthill-McKee ordering is applied.

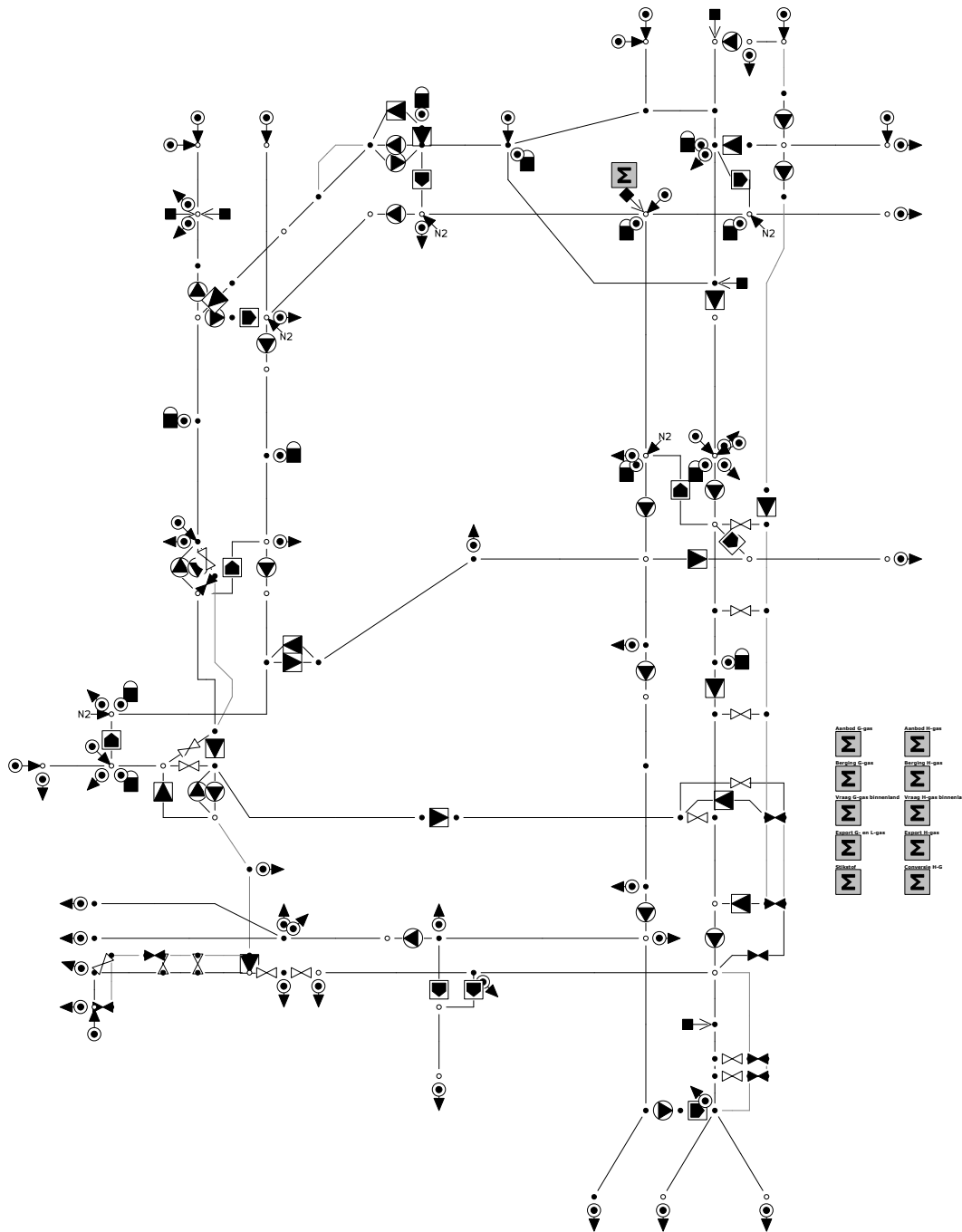


Figure A-27: Another simplification of the network of GTS.

Bibliography

- [1] J. Steringa, *Inzetvarianten voor het HTL*. Gas Transport Services B.V., 2011.
- [2] Gasunie, “Integrated annual report 2013,” tech. rep., N.V. Nederlandse Gasunie, 2013.
- [3] G. N. Configuration, “Ontwerp uitgangspunten transportsysteem,” tech. rep., N.V. Nederlandse Gasunie, 2014.
- [4] D. van Huizen, “Identifying severe transport situations in a gas transmission network,” internship thesis, University of Twente, 2014.
- [5] J. J. Steringa, M. Hoogwerf, and H. Dijkhuis, “A systematic approach to transmission stress tests in entry-exit systems,” in *Pipeline Simulation Interest Group*, 2015.
- [6] N.V. Nederlandse Gasunie, *basisgegevens aardgassen*, 1980.
- [7] T. Skopal, T. Bartoš, and J. Lokoč, “On (not) indexing quadratic form distance by metric access methods,” in *Proceedings of the 14th International Conference on Extending Database Technology*, 2011.
- [8] M. Stricker and M. Orengo, “Similarity of color images,” in *Proceedings of SPIE*, vol. 2420, pp. 381 – 392, 1995.
- [9] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99 – 121, 2000.
- [10] C. Vuik and D. J. P. Lahaye, “Reader scientific computing (wi4201),” 2013.
- [11] K. Yosida, *Functional Analysis*. Springer Berlin Heidelberg, 1968.
- [12] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, “Efficient color histogram indexing for quadratic form distance functions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729 – 736, 1995.

-
- [13] C. P. Simon and L. Blume, *Mathematics for Economists*. W.W. Norton & Company, first ed., 1994.
- [14] G. Debreu, “Definite and semidefinite quadratic forms,” *Journal of the Econometric Society*, vol. 20, no. 2, pp. 295 – 300, 1952.
- [15] K. D. Ikramov and N. V. Savel’eva, “Conditionally definite matrices,” *Journal of Mathematical Science*, vol. 98, no. 1, pp. 1 – 50, 2000.
- [16] T. van der Hoeven, *Math in Gas and the art of linearization*. PhD thesis, University of Groningen, 2004.
- [17] N. J. Higham, “Computing the nearest correlation matrix — a problem from finance,” *IMA Journal of Numerical Analysis*, vol. 22, no. 3, pp. 329 – 343, 2002.
- [18] N. J. Higham, “Computing a nearest symmetric positive semidefinite matrix,” in *Linear Algebra and its Applications*, vol. 103, pp. 103 – 118, Elsevier Science Publishing Co., 1988.
- [19] W.-H. Liu and A. H. Sherman, “Comparative analysis of the cuthill-mckee and the reverse cuthill-mckee ordering algorithms for sparse matrices,” *SIAM Journal on Numerical Analysis*, vol. 13, no. 2, pp. 198 – 213, 1976.
- [20] G. H. Golub and C. F. van Loan, *Matrix Computations*. The Johns Hopkins University Press, third ed., 1996.

Glossary


List of Acronyms


G-gas	Groningen gas (Wobbe index smaller than or equal to 44.4)
GTS	Gasunie Transport Services
H-gas	Gas with a Wobbe index greater than or equal to 49.0
HTL	High pressure grid (<i>Dutch: Hoofdtransportleidingnet</i>)
L-gas	Gas with a Wobbe index between 44.4 and 47.2
MCA	Multiple Case Analysis (or Approach)
PSD	Positive Semidefinite
QFD	Quadratic Form Distance
RTL	Intermediate pressure grid (<i>Dutch: Regionale transportleidingnet</i>)
SPD	Symmetric Positive Definite
SPSD	Symmetric Positive Semidefinite


List of Symbols

Q_i	flow in pipeline i
L_i	length of pipeline i
W	Wobbe index
H_s	higher calorific value
d	density of gas relative to air
ρ_g	density of gas
ρ_l	density of air
d_{ij}	distance between i and j
d_{max}	maximum distance
λ	eigenvalue
λ_i	eigenvalue i
ρ_i	radius of i -th Gershgorin disc
δ_i	permitted (fixed) inaccuracy for network point i
θ	angle between two vectors representing stress tests
P	pressure
P_{in}	pressure at the begin of the pipeline
P_{out}	pressure at the end of the pipeline
ΔP	pressure drop (in the pipeline)
c	a constant
L	length
D	diameter
Q	flow


Symbols used in MCA

• —  — • Blending station


• —  — • Checkvalve

•  • Entry point

•  • Exit point

•  • Nitrogen injection point

•  • Storage

• —  — • Valve