Transport planning of the Dutch Gas Network

Finding challenging transport situations by a sufficient method

by

S. Maring

Literature Study

Student number:	4226518	
Project duration:	March 1, 2017 – December 1, 2017	
Thesis committee:	Prof. dr. ir. C. Vuik,	TU Delft, supervisor
	Dr. J.W. van der Woude,	TU Delft, supervisor
	Drs. H. Dijkhuis,	Gasunie Transport Services
	Dr. J.J. Steringa,	Gasunie Transport Services

TU Delft gasunie transport services

Abstract

This literature study is about finding detailed subjects for the research questions. At Gasunie, which is the Dutch transmission system operator of gas, a good planning of the gas flow through the network is important. By law, the Gasunie is required to manage a good transport of gas, which means a safe and reliable gas network. To make this possible, the network must be able to handle all possible realistic scenarios. To check if this is the case, a method must be used which quickly and accurately covers all scenarios.

There are many factors involved in the gas network. Therefore, the current method used for the research problem approximates the realistic scenarios and reduce the number of scenarios by taking only the severe ones. The scenarios involve only the capacities on entry points (where the gas flow is injected) and exit points (where the gas is taken off). If the severe situations can be handled by the gas network, the less severe ones are also satisfied by the gas network. After reducing the amount of situations, this amount is again reduced by taking out the ones that are close to each other. This is done by the distance called quadratic form distance. At Gasunie, the quadratic form distance is used with a parametrisation dependent on the mutual shortest paths between entry and exit points.

There is another distance for measuring the distance between two situations, called the earth mover's distance. This distance makes more use of the flow through the network, where the quadratic form distance only make use of the capacities on the entry points and the exit points.

Besides the method to compare situations described by capacities on entry and exit points, there are methods to compare situations based on the flow through each pipeline segment. This other description of the network leads to a graph with the original network points as nodes and the mutual pipelines as edges. The entry and exit points are represented by edges and two extra nodes. Such a network can be solved by linear programming, because it is an optimisation problem with a connected graph.

There are different algorithms to solve linear programming. Simplex is often used, because of its simplicity, however the worst-case scenario has a exponential time complexity. The interior-point algorithm is more advanced and has polynomial time complexity.

Pressure plays a big role in the gas transport network. With too less pressure, the gas cannot flow through the network. When the gas has to be transported over a long distance, the pressure decreases. The pressure equation is not linear, therefore the addition of pressure to the methods cause bigger computations.

Research questions involve the analysis of different measures for the severe situations of the current method. The accuracy, time of computation and memory load can be taken into account. Also, comparing the distances with different structures of the network is an interesting analysis.

The next part of the research questions are about the flow representation. How can this representation check whether the gas network suffices for all realistic scenarios? And which algorithm is most suitable for this problem?

Another subject in this project is the pressure drop in the network. It should be discussed how much value this quantity adds to the methods to solve the problem. The pressure equations or an approximation of it should be added to each method, so before answering the previous question, it should be considered how the pressure addition should be done.

Contents

1 Introduction

Gasunie is the gas Transmission system operator (TSO) of the Netherlands. The company is required to supply entry and exit capacity to the market. This is done by negotiating contracts with parties that supply or demand gas at the entry and exit points. These parties are called shippers, which can be companies from the Netherlands, but also from all over Europe, like Germany, the United Kingdom and Russia [17, 20]. A few well-known companies are Shell, NAM and Gazprom. Gasunie is responsible for 15.000 kilometer gas pipes [6], which transported 1236 Tera Watt hour in 2016 [8].

The entry and exit system of Gasunie differs in planning and optimisation of a system where capacity can be controlled [20]. The main planning issue here is to find the most challenging transport situations that may arise from the many possible entry and exit combinations. The optimisation lies between the costs of new projects and materials for the gas network, and to provide sufficient transport for all situations which may occur.

Sufficient capacity transport involves the need of gas of parties at the entry and exit points. The optimisation of this transport is a complex problem. Knowledge of scope, policy and behaviour of all the parties is needed. Dealing with this problem is done by mathematical theory, such as statistics, linear algebra and numerical analysis.

Gas transmission systems operate on an hourly timescale. Every shipper is required to notify their need of capacity within contractual limits. This notification is done a couple of hours in advance for every hour of the day. Each year has 8760 hours and for each hour billions of different transport situations exist, so the planning of the gas transmission involves a lot of scenarios. It is not desirable to analyse all possibilities, because it costs much time and many computations. There are ways to reduce the amount of transport situations and still base the gas network planning on all possible transport situations.

At Gasunie, there is already a method that reduces the possible transport situations. This is done by calculating an approximation of the gas flow through the network and only consider the most severe ones. After this selection, the amount of severe situations is reduced by taking out similar ones. To define whether a situation is similar to another, a measure is needed to compare the situations. Researchers of the planning department at Gasunie decided to use a certain parametrisation of the Quadratic Form Distance. In this paper this method is discussed and also other methods and measures are analysed. The goal of this project is to find the most suitable method to find the minimal amount of tests that have to be checked to ensure the quality of the gas network. Therefore an analysis of different measures and different approaches is needed.

In this report, only the High pressure grid (Dutch: Hoofdtransportleidingnet) (HTL) is considered for the gas network. This network is more complex for the planning of gas than the Intermediate pressure grid (Dutch: Regionale transportleidingnet) (RTL), because the pressure in the pipes is higher and the transport distance is larger. There are three different types of natural gas in the Dutch gas network. H-gas consists of a high calorific value, which contributes to high quality natural gas. L-gas is gas which has a lower caloric value and the lowest quality is called G-gas, Groningen gas. This quality of gas is used by households for cooking and the CV boiler.

The Dutch gas network consist not only on entry and exit points, but also of compressor stations, blending stations of different types of natural gas, pressure regulation-stations and nitrogen injections [1]. Compressor stations increases the pressure in the pipe such that there is enough pressure to transport the gas over a long distance. Blending stations are used for the conversion to G-gas. The nitrogen injection ensures the process of H-gas into G-gas. Pressure regulation-stations are used for the transition from the HTL-network to the RTL-network: the odorisation takes place and it is made sure that the pressure is at the right level such that the gas can safely go into the smaller pipes.

The total HTL-gas infrastructure of the Netherlands in 2014 is given on page 3. The yellow lines are the pipelines which transport high-calorific gas, also known as H-gas. The grey lines are the pipelines which transport G-gas. Furthermore, the stations described above are seen in the figure along with gas storage facilities.

In chapter 2, the transport situations of gas flow is explained. The quantity transport moment is introduced to define the stress test in section 2.1. The stress tests are currently measured by the quadratic form distance, considered in section 2.2. Another distance which can be used is the earth mover's distance (section 2.3). A third distance, the diffusion distance is discussed in section 2.4. Chapter 3 announce another representation of the gas transportation network. Here, linear programming (section 3.1) is described with the algorithms simplex and interior-point. Eventually in chapter 4, a summary of this report is given, which concludes in the research questions. These questions are guidelines of the continuation of this project.

Figure 1.1: Dutch HTL-gas network in 2014

2 Gas transport

Gasunie is obligated to ensure the safety and supply of the natural gas in the Netherlands by law. Therefore calculations must be made to ensure these regulations. This is done at the planning department of Gasunie Transport Services (GTS), where only the realistic situations of gas transport are taken into account (GTS is a subsidiary of Gasunie). These calculations answers questions like: is the current gas network sufficient for the supply or can pipelines be removed, is it necessary that some pipelines have to be added or replaced. Transport situations are realistic situations in a gas transmission system, where the entry and exit capacities are balanced. This means that the sum of the entry capacities equals the sum of the exit capacities, which is also known as the flow conservation law. An example of a transport situation is shown in figure 2.1 on page 6. This network is an approximation of the HTL gas network in the Netherlands, which has fictional flows in this example.

Realistic situations are situations that are likely to happen in practice. An assumption of these situations is that the shippers adhere to their contracts with Gasunie. So if a shipper has a contract for entry point A, there is a maximum and a minimum agreed on the capacity of the gas transported into the network at A. These bounds of entry and exit points are known to Gasunie to a certain extent. The assumptions made in this report are listed in appendix A on page 39.

It is sufficient to check the network for only the most severe situations, because if these situations are feasible for the network, the non-severe situations are also feasible for the network [19]. The severity of the situation can be given by the quantity transport moment, which approximates the transport load. The transport load is dependent of the pipe length, amount of flow, pressure requirements, compressor stations, temperature and more. The total transport moment T is calculated by the sum of the transport moment per pipeline segment. It is assumed that the pipe length and the relation to the amount of flow through that pipe are the most important quantities. If the transport distance is twice as long, the transport moment will be twice as high, and the same holds for the amount of flow. Therefore the transport moments are calculated by the product of the flow through that segment (q) and the length of that pipeline segment (d). The total transport moment formula is seen in equation 2.1, where E is the set of pipeline segments.

$$T_{\text{total}} = \sum_{e \in E} T(e) = \sum_{e \in E} q(e) \cdot d(e) \tag{2.1}$$

$q(e)$ is the flow through pipe $e \in E$ and $d(e)$ is the length of pipe $e \in E$.

In most articles on this subject, this equation is just given as $T = QD$.

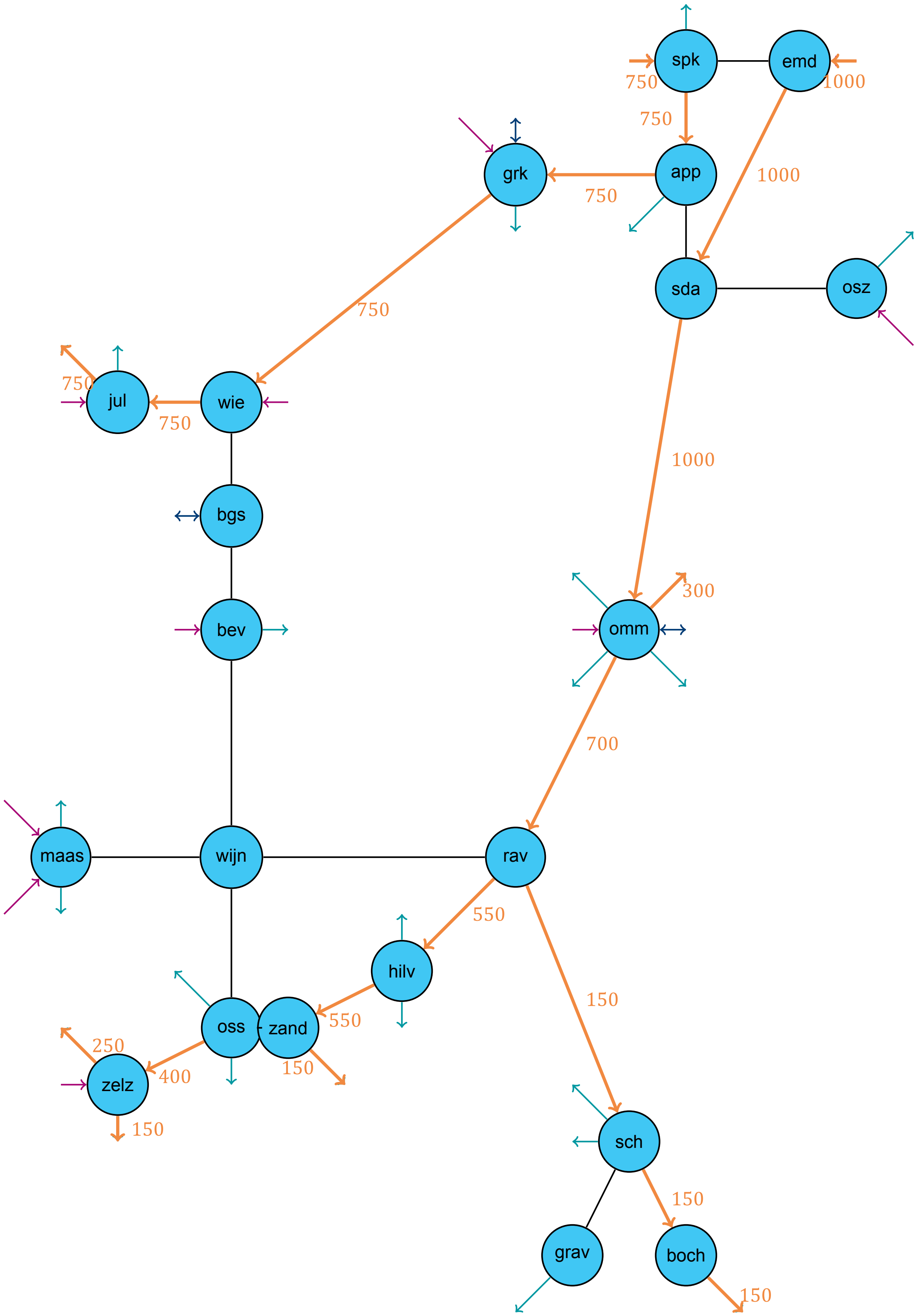


Figure 2.1: Example of a transport situation with fictional flows

2.1. Stress test

The transport moment of equation 2.1 is based on the flow in the network per situation. However, to compute the flow in the network, much computations are needed. Moreover, when the flow of each transport situation is calculated, the problem of the unknown quality of the gas network has been solved. Therefore, an approximation is made that is based on the capacities on the entry and exit points and the mutual distances. This approximation ensures that the transport situation is only described by the capacities at the end points, so the calculations of the flows in the network is not needed. The most severe transport situations that are balanced are called stress tests.

The transport moment based on entry and exit points is approximated by equation 2.2. X_i is exit point i, N_j is entry point j, the capacity is given by $c(\cdot)$ and the distance is given by $d(\cdot,\cdot)$. Distance d is the shortest path from one point to another. The severe situations are found by choosing the capacities on the entry and exit points so that the transport moment is maximised under the conditions of the conservation law and the capacity boundaries [19]. Let I be the amount of exit points and J is the amount of entry points. All situations are covered by taking every entry and exit point as anchor point a. It is not known what happens inside the network for the approximated transport moment, so it is only known what the capacity is on the entry and exit points and the spacing between entry and exit points. Therefore, all situations are covered by taking all entry and exit points as anchor point a for the approximation of the transport moment. In this way all directions of the flow are taken into account.

$$T(a) = \sum_{i=1}^{I} c(X_i)d(X_i, a) - \sum_{j=1}^{J} c(N_j)d(N_j, a) \tag{2.2}$$

Figure 2.2 is a network with a feasible transport situation. Assume that this is a severe situation for this network, then the stress test vector of this transport situation is stated below.

The capacity on the points $\begin{pmatrix} N_A \\ N_B \\ N_D \\ N_E \\ X_B \\ X_C \\ X_D \\ X_F \\ X_G \\ X_H \end{pmatrix}$ are $\begin{pmatrix} 400 \\ 200 \\ 100 \\ 300 \\ -50 \\ -150 \\ -100 \\ -200 \\ -300 \\ -200 \end{pmatrix}$ where the exits are given by a minus sign.

There are negative flows through the exit points, because the gas leaves the network at these points. It can be seen that the stress test is balanced, because the sum of the entry capacities equals the sum of the exit capacities.

2.1.1. Stress test algorithm

The algorithm to find the stress tests is given in algorithm 2.1, this is an adapted algorithm of the one by Steringa, et al [19]. The algorithm maximises the transport moment of equation 2.2 by iterating over each anchor point and adding capacity at the closest entry point such that the balance is preserved. The same is done for the exit capacity furthest away from the anchor point. By putting the exit capacity furthest away, the distance is maximised and by adding maximal capacity, the transport moment is maximised until some iteration in the algorithm. When putting more capacity to the network when the maximum is reached, the transport moment will decrease. The upper bound of an entry or exit point is given by $u(\cdot)$ and the lower bound is given by $l(\cdot)$.

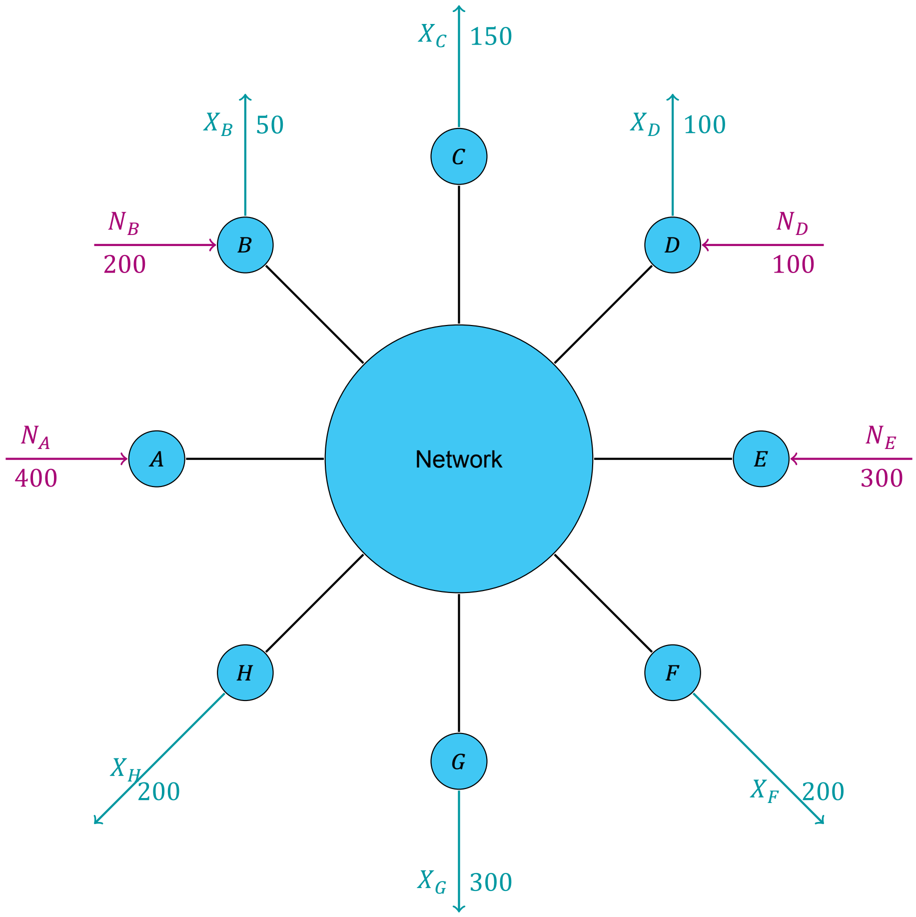


Figure 2.2: Example feasible situation

Algorithm 2.1 The Stress Test algorithm

Input: Upper bounds $u = \{u(N_1), \ldots, u(N_I), u(X_1), \ldots u(X_J)\}$, distance matrix $D \in \mathbb{R}^{(I+J)\times(I+J)}$
Output: Set of stress tests $S \in \mathbb{R}^{(I+J)\times(I+J)}$

intv1 $\leftarrow [1 : I]$
intv2 $\leftarrow [I+1 : I+J]$
for $k = 1, \ldots, (I+J)$ **do**
  $c \leftarrow$ zeros$(I+J, 1)$
  $c_{tmp} \leftarrow c$
  $T_1 \leftarrow -10^{-10}$
  $T_2 \leftarrow 0$
  while $T_1 < T_2$ and $\sum c == 0$ **do**
    $c \leftarrow c_{tmp}$
    $T_1 \leftarrow T_2$
    index$_N$ $\leftarrow$ index($\min\{D[k,\text{intv1}] : c_{tmp}[\text{intv1}] < u[\text{intv1}]\}$)
    index$_X$ $\leftarrow$ index($\max\{D[k,\text{intv2}] : c_{tmp}[\text{intv2}] < u[\text{intv2}]\}$)
    $u_{\min} \leftarrow \min\{u[\text{index}_N] - c_{tmp}[\text{index}_N], u[\text{index}_X] + c[\text{index}_X]\}$
    $c_{tmp}[\text{index}_N] \leftarrow c_{tmp}[\text{index}_N] + u_{\min}$
    $c_{tmp}[\text{index}_X] \leftarrow c_{tmp}[\text{index}_X] - u_{\min}$
    $T_2 \leftarrow D[k,:] \cdot -c_{tmp}$
  end while
  $S[:,k] \leftarrow c$
end for

2.1.2. Example

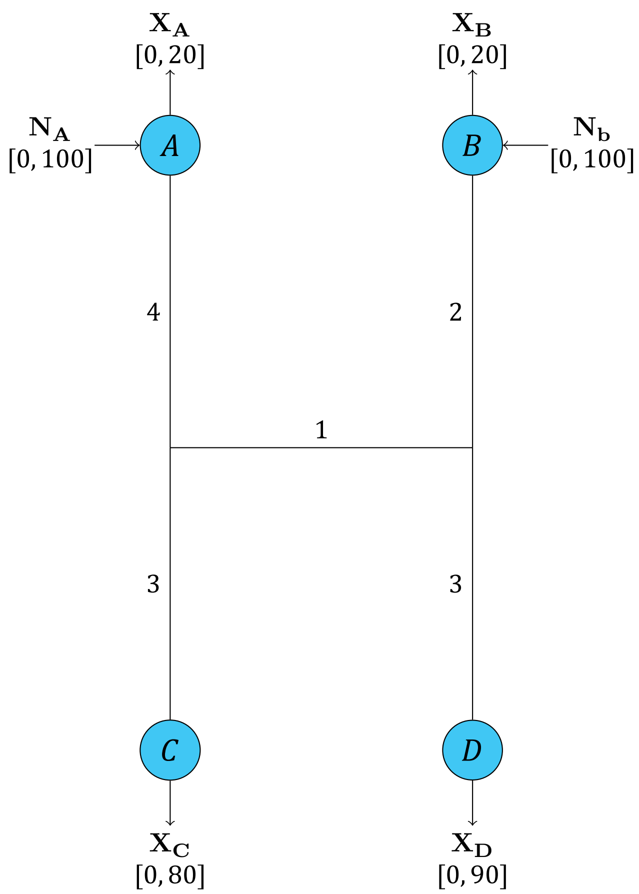


Figure 2.3: Example generating stress tests

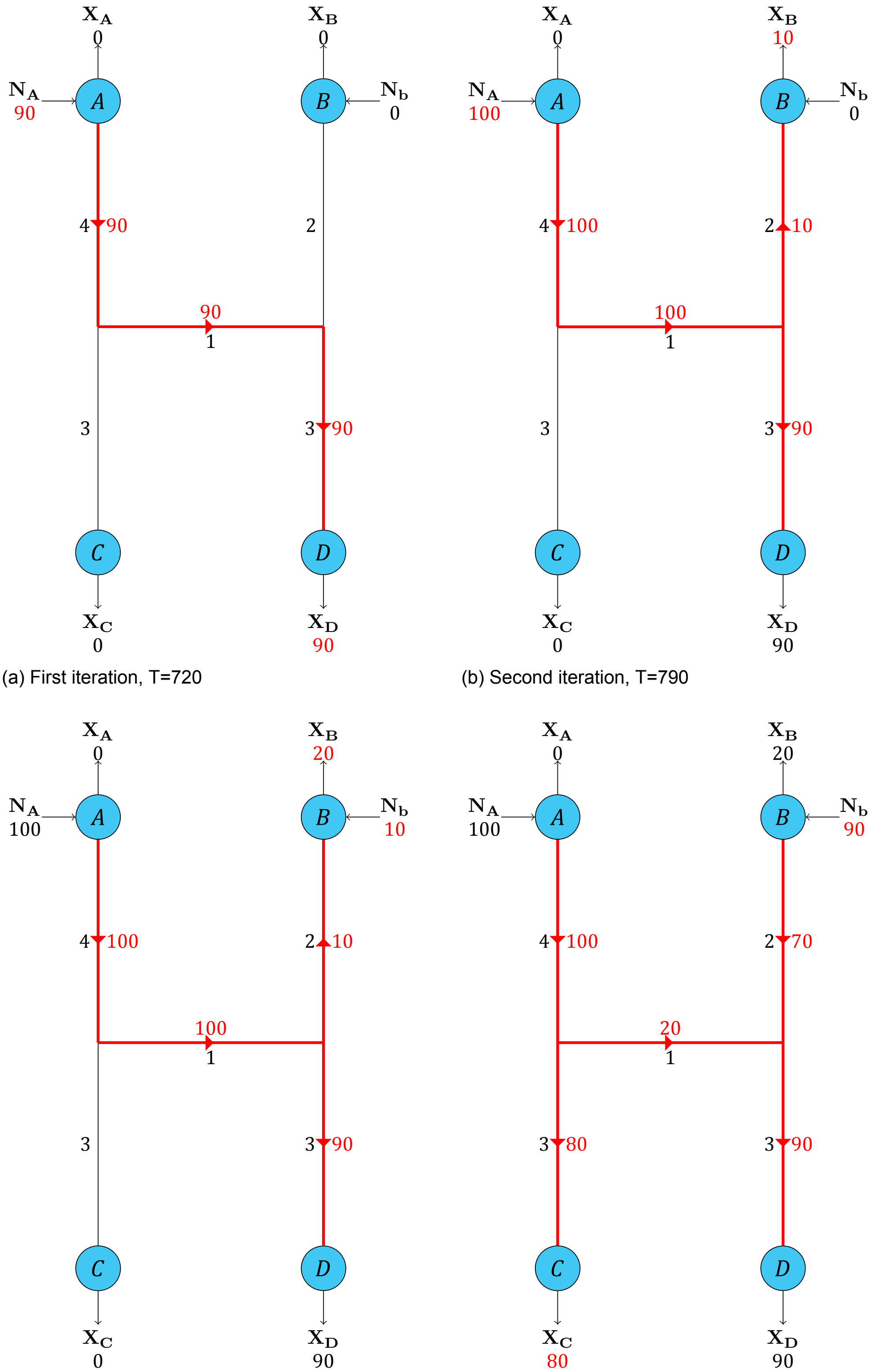
The network, considered for this example is given in figure 2.3. There are two entry points (N_A, N_B) and four exit points (X_A, X_B, X_C, X_D) and five pipelines. The length of each pipeline is given in the figure next to the relevant pipeline and the direction of the gas flow can be in both directions. The lower bound and the upper bound of each entry and exit is given at these points by [lb, ub], where lb is the lower bound and ub is the upper bound.

Often, there is an example used which has almost the same structure, but has no exits at points A and B, all pipe segments have the same length and the bounds on the entry and exit points are symmetric. In this example, the pipe lengths are not chosen symmetric, so that the calculations between different anchor points are more clear. The exit points at A and B are put into the network to show that some scenarios are similar, but not equal. The iterations of the algorithm are described below. The iterations of the first anchor point is explained in more detail than the other anchor points. If there is a choice to make between entries/ exits, the first entry/ exit in alphabetical order is chosen.

Anchor point N_A

An illustration is given to this anchor point in figure 2.4. In this figure is the flow shown by the red lines, however in the description of the calculations (explained below) are these flows not known. Only the mutual distances and the capacity on the entry and exit points are known. In this network it can be seen easily and it gives a better view on the transport situation when flow is indicated.

- The entry closest to anchor point N_A is N_A and exit furthest away to the anchor point is X_D. The minimum of the maximal capacities on N_A and X_D that can be used is $\min\{100, 90\} = 90$.



(a) First iteration, T=720

(b) Second iteration, T=790

(c) Third iteration, T=790

(d) Fourth iteration, T=790

Figure 2.4: Finding stress test for anchor point N_A

Therefore the transport moment on the capacities becomes:

$$\begin{aligned} T(N_A) &= \sum_{j=1}^{4} c(X_j)d(N_A, X_j) + \sum_{i=1}^{2} c(N_i)d(N_A, N_i) \\ &= 90 \cdot 8 - 90 \cdot 0 \\ &= 720 \end{aligned}$$

- More capacity on the entry and exit points can be added until the transport moment is decreasing. The entry point closest with capacity left is N_A and the exit point furthest away with capacity is X_B. The minimum of the maximal capacity that can be chosen on the entry and exit is 10. If this capacity is added, the transport moment becomes:

$$\begin{aligned} T(N_A) &= \sum_{j=1}^{4} c(X_j)d(N_A, X_j) + \sum_{i=1}^{2} c(N_i)d(N_A, N_i) \\ &= 10 \cdot 7 + 90 \cdot 8 - 100 \cdot 0 \\ &= 790 \end{aligned}$$

- The next closest entry point is N_B and exit point X_B has some capacity left. The capacity added to these points is 10. When adding this capacity, the transport moment is:

$$\begin{aligned} T(N_A) &= \sum_{j=1}^{4} c(X_j)d(N_A, X_j) + \sum_{i=1}^{2} c(N_i)d(N_A, N_i) \\ &= 20 \cdot 7 + 90 \cdot 8 - 100 \cdot 0 - 10 \cdot 7 \\ &= 790 \end{aligned}$$

- At exit point X_B all possible capacity is being used, so the next exit, which is far away from the anchor point and has capacity left is X_C. The entry point is N_B. The capacity which can be added on these points is 80.

$$\begin{aligned} T(N_A) &= \sum_{j=1}^{4} c(X_j)d(N_A, X_j) + \sum_{i=1}^{2} c(N_i)d(N_A, N_i) \\ &= 20 \cdot 7 + 90 \cdot 8 + 80 \cdot 7 - 100 \cdot 0 - 90 \cdot 7 \\ &= 790 \end{aligned}$$

- This fifth iteration will be the last iteration.

$$\begin{aligned} T(N_A) &= \sum_{j=1}^{4} c(X_j)d(N_A, X_j) + \sum_{i=1}^{2} c(N_i)d(N_A, N_i) \\ &= 10 \cdot 0 + 20 \cdot 7 + 90 \cdot 8 + 80 \cdot 7 - 100 \cdot 0 - 100 \cdot 7 \\ &= 790 \end{aligned}$$

This transport moment is decreased with respect to the transport moment before. So the addition of flow on exit X_A and entry N_B will stop the while loop and the stress test in equation 2.3 is found on the anchor point at N_A. In this simple example, the flows through the network can easily be generated when knowing the capacities on the entry and exit points. The fifth iteration (figure 2.5) is seen in figure 2.5 and it is seen that this is less severe than the situation at iteration 4.

$$\begin{pmatrix} 100 \\ 90 \\ -0 \\ -20 \\ -80 \\ -90 \end{pmatrix} \tag{2.3}$$

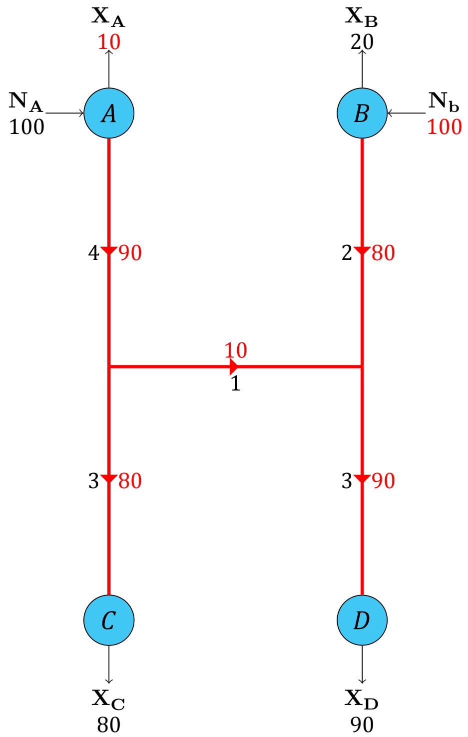


Figure 2.5: Fifth iteration, T=780

This stress test represents the capacity on the entry and exit points. It is seen that the stress test is balanced, because the sum of the capacities on the entries is 190 and the same capacity sum is found for the exit points.

Anchor point N_B

- Entry N_B, exit X_A, capacity that can be added is 20.

$$\begin{aligned} T(N_B) &= 20 \cdot 7 - 20 \cdot 0 \\ &= 140 \end{aligned}$$

- Entry N_B, exit X_C, capacity that can be added is 80.

$$\begin{aligned} T(N_B) &= 20 \cdot 7 + 80 \cdot 6 - 100 \cdot 0 \\ &= 620 \end{aligned}$$

- Entry N_A, exit X_D, capacity that can be added is 90.

$$\begin{aligned} T(N_B) &= 20 \cdot 7 + 80 \cdot 6 + 90 \cdot 5 - 90 \cdot 7 - 100 \cdot 0 \\ &= 440 \end{aligned}$$

This addition of capacity causes a smaller transport moment, therefore the last addition is not included for the stress test.

$$\begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix}$$

Anchor points X_A and X_B

The stress tests for the exit points X_A and X_B is the same as for N_A and N_B respectively, because they are located at the same point and therefore the same exit and entry points will be chosen. So for every point where the distance is zero, these cases can be considered as the same. Algorithm 2.1 will again do the same iterations as in N_A and N_B.

Anchor point X_C

- Entry N_B, exit X_A, capacity that can be added is 20.

$$\begin{aligned} T(X_C) &= 20 \cdot 7 - 20 \cdot 6 \\ &= 20 \end{aligned}$$

- Entry N_B, exit X_D, capacity that can be added is 80.

$$\begin{aligned} T(X_C) &= 20 \cdot 7 + 80 \cdot 7 - 100 \cdot 6 \\ &= 100 \end{aligned}$$

- Entry N_A, exit X_D, capacity that can be added is 10.

$$\begin{aligned} T(X_C) &= 20 \cdot 7 + 90 \cdot 7 - 10 \cdot 7 - 100 \cdot 6 \\ &= 100 \end{aligned}$$

- Entry N_A, exit X_B, capacity that can be added is 20.

$$\begin{aligned} T(X_C) &= 20 \cdot 7 + 20 \cdot 6 + 90 \cdot 7 - 30 \cdot 7 - 100 \cdot 6 \\ &= 80 \end{aligned}$$

This capacity addition causes a smaller transport moment, therefore the last addition is not included for the stress test.

$$\begin{pmatrix} 10 \\ 100 \\ -20 \\ -0 \\ -0 \\ -90 \end{pmatrix}$$

Anchor point X_D

- Entry N_B, exit X_A, capacity that can be added is 20.

$$\begin{aligned} T(X_D) &= 20 \cdot 8 - 20 \cdot 5 \\ &= 60 \end{aligned}$$

- Entry N_B, exit X_C, capacity that can be added is 80.

$$\begin{aligned} T(X_D) &= 20 \cdot 8 + 80 \cdot 7 - 100 \cdot 5 \\ &= 220 \end{aligned}$$

- Entry N_A, exit X_B, capacity that can be added is 20.

$$\begin{aligned} T(X_D) &= 20 \cdot 8 + 20 \cdot 5 + 80 \cdot 7 - 20 \cdot 8 - 100 \cdot 5 \\ &= 160 \end{aligned}$$

This capacity addition causes a smaller transport moment, therefore the last addition is not included for the stress test.

$$\begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix}$$

Finally, there are four different stress tests found, which are all balanced by definition of the algorithm. The stress tests of anchor point B and D are the same, because the order of nearest entry is the same at each anchor point. And the order of exit points farthest away is almost the same, it holds for the first two exit points: X_A and X_C. After using these two exit points in the algorithm, the transport moment of the capacities is decreased. Therefore, the stress tests of the anchor points B and D are the same.

$$s_A = \begin{pmatrix} 100 \\ 90 \\ -0 \\ -20 \\ -80 \\ -90 \end{pmatrix}, \quad s_B = \begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix}, \quad s_C = \begin{pmatrix} 10 \\ 100 \\ -20 \\ -0 \\ -0 \\ -90 \end{pmatrix}, \quad s_D = \begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix} \tag{2.4}$$

A comparison of the two different approximations of the transport moment is made by the stress tests in table 2.1.

Anchor point	T based on capacities	T based on flows
N_A, X_A	790	1070
N_B, X_B	620	620
X_C	100	520
X_D	200	620

Table 2.1: Two approximations of the transport moment of stress tests

It is seen in the table and figure 2.6 that there is no clear trend in these two transport moments.

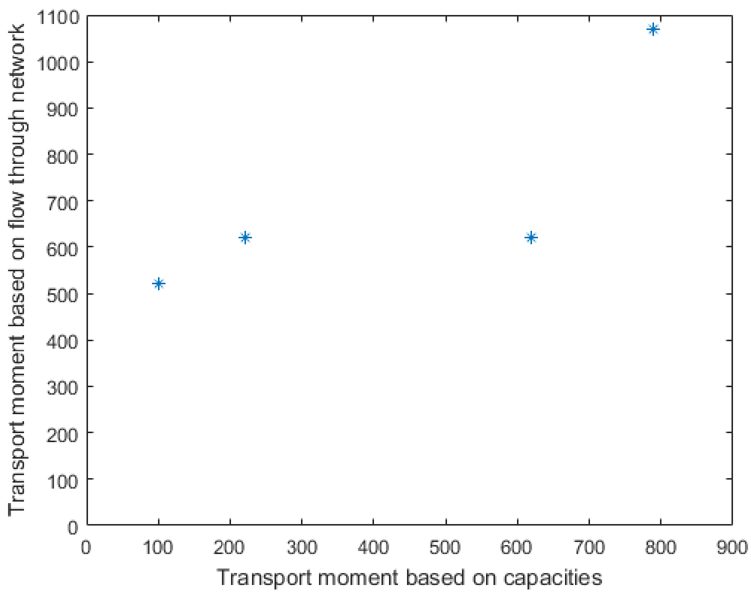
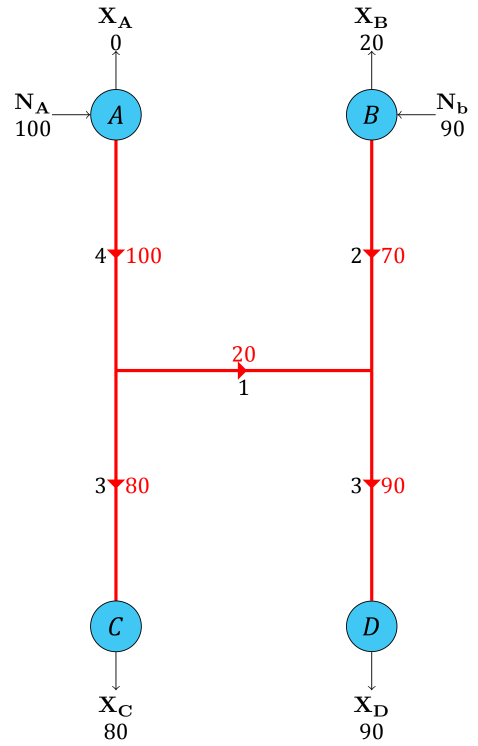
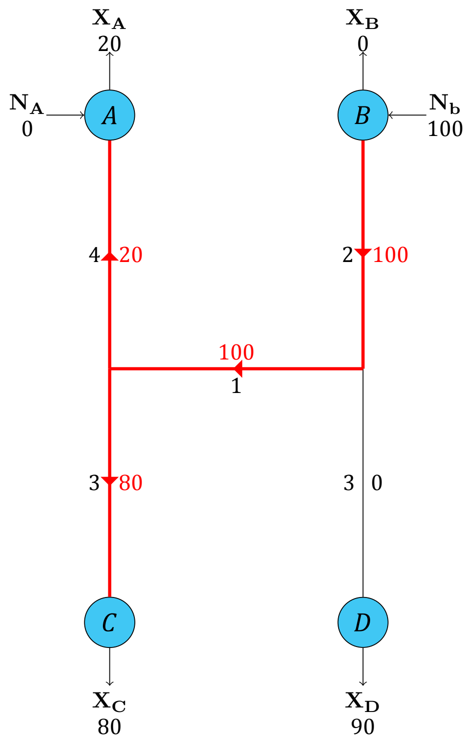


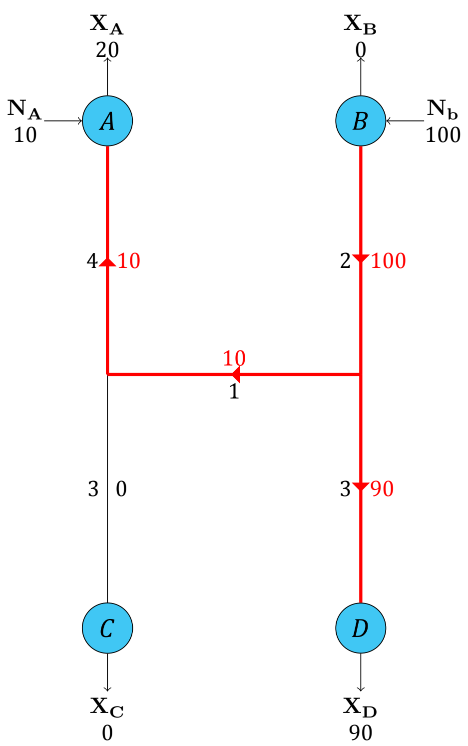
Figure 2.6: Two approximations of the transport moment on different stress tests



(a) Stress test, anchor point A



(b) Stress test, anchor point B



(c) Stress test, anchor point C

Figure 2.7: Flow on the stress tests

2.2. Quadratic form distance

In the previous section, the stress tests (severe situations) are found, but the amount of stress tests found is the amount of exit points (around 1100 in HTL-network) plus the amount of entry points (around 50 in HTL-network) [16]. To reduce the amount of stress tests, similarities can be found between the stress tests. To make this possible, a distance has to be defined to measure the difference between the stress tests and give an answer of which stress tests are similar. There is a distance which seems to work, however this cannot be proven mathematically (yet). This distance is called the quadratic form distance with a certain parametrisation.

In the article of Skopal, et al [18] images are compared to each other and with the Quadratic form distance (QFD) a rate of similarity is given. There are some techniques for comparing images by making histograms of the colours at every location of the image and comparing them with other images. However, if the image has some noise or is scaled or rotated, those techniques do not give a good similarity measure between the distorted image and the original. The QFD takes distortion of the image into account by increasing the dimensions of what the image is dependent of. The dimensions of an image can be the amount of red, green and blue of each pixel, so for m pixels, the dimension of the image is $3m$, where only colour is included. Texture for example is also a quantity which can be added to be a dimension.

This distance function can also be used for comparing stress tests, because besides the Euclidean distance (or other L_p-distances) between the vectors of stress tests, the geographical distance has also to be taken into account for comparing stress tests. When two entry points N_1 and N_2 are close to each other in the network, the transport load from N_1 to exit point X will be similar to the transport load from N_2 to X.

The aim is that the quadratic form distance will give stress tests that are both similar to each other. The stress tests can be denoted as n-dimensional vectors where n is the total number of entry and exit points. The capacity on the entry points is given in the vector with a plus sign and the capacity on the exit points is given in the vector with a minus sign, just as in previous section.

The QFD is given by

$$QFD_{\mathbf{A}}(u,v) = \sqrt{(u-v)^T A(u-v)} \tag{2.5}$$

where $\mathbf{A}$ is an $n \times n$ symmetric positive definite matrix [10] and u and v are the n-dimensional vectors which represents two stress tests. The transport moment represents the severeness of the situation, so the length of the transport must be included in the matrix $\mathbf{A}$.

In the article of Skopal et al the matrix A is defined by

$$a_{ij} = 1 - \frac{d_{ij}}{d_{\max}}, \quad \text{with } d_{\max} = \max_{i,j} d_{ij}, \quad i,j = 1,2,\dots,n, \tag{2.6}$$

where a_{ij} are the entries of matrix $\mathbf{A}$ and d_{ij} is the Euclidean distance between representatives of colours i and j.

In the master thesis of K. Lindenberg [10] the matrix A is defined as in equation 2.6, but with another distance d_{ij} than in colour comparison. In the gas network, there are no Euclidean distance necessarily, so the distance is chosen to be the shortest paths along the pipeline between point i and point j. These points are exit and exit points.

This quadratic form distance is more preferable than the L_p distance, because the QFD gives also the correlation between different dimensions, while the L_p distances give the combination of the distances of each dimension independently.

2.2.1. Example

Illustrating the previous theory, a simple example is given in this section. First a concrete example is given and the most severe situations is calculated with the quadratic form distance and matrix $\mathbf{A}$ as in equation 2.6 with D as the distance matrix of the shortest path through the pipeline. The network used in this section is given in figure 2.8. This network has one pipeline from a to b, two exit points (one at a and one at b) and one entry point, which is at point a. This network has only one pipeline, so the length of the pipeline can be arbitrary chosen as long as it is greater than zero. Call this length $L > 0$.

Figure 2.8: Simple network

The distance matrix is for the network above is for every capacity the same. The distance is the shortest path between entry and exit points.

$$D = \begin{pmatrix} 0 & 0 & L \\ 0 & 0 & L \\ L & L & 0 \end{pmatrix} \tag{2.7}$$

The matrix A can be computed from D:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

From the stress test algorithm, the most severe situation is found by adding some capacity to the network until the bounds are exceeded. However, in this simple network the most severe situation can be seen easily.

Consider the network with the following bounds on the entry and exit points:

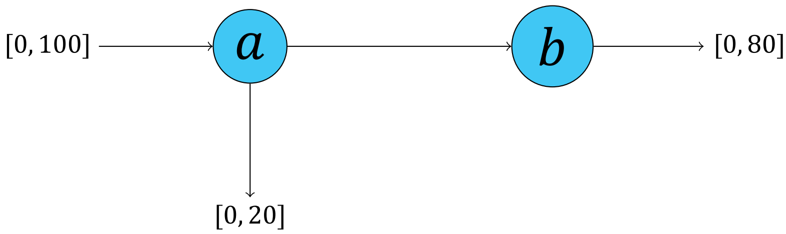


Figure 2.9: Simple example

Some of the severe situations can easily be seen in this example. Like the situation that there is 80 injected at a, no capacity leaves a directly through the exit point of a and 80 will leave at the exit point at b. This situation is represented in the following stress test vector $(80, 0, -80)^T$. At the entry points, the capacity is positive and at the exit the capacity is negative.

Another severe situation is $(100, -20, -80)^T$, so here 100 is injected to a and immediately 20 is leaving a and 80 leaves b at its exit point. In this example it is easily seen that these two situations give the same transport moment, because $T = QD = 80L$ for both situations.

In fact, all situations of the form $(80 + x, -x, -80)^T$ with $x \in [0, 20]$ will give the same transport moment. The QFD will show that they are similar, because the distance is zero. This is proven by taking two arbitrary stress tests of the following form $v_x = (80 + x, -x, -80)^T$ and $v_y = (80 + y, -y, -80)^T$.

$$\begin{aligned} QFD_{\mathbf{A}}(v_x, v_y) &= \sqrt{(v_x - v_y)^T \mathbf{A} (v_x - v_y)} \\ &= \sqrt{\begin{pmatrix} x-y & y-x & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x-y \\ y-x \\ 0 \end{pmatrix}} \\ &= \sqrt{\begin{pmatrix} x-y & y-x & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}} \\ &= 0 \end{aligned}$$

So for the gas network all situations which correspond to $(80 + x, -x, -80)^T$ with $x \in [0, 20]$ are similarly severe and most severe for the network with this bounds on the entry and exit points.

The network with general bounds is discussed in B.2 on page 43. The stress tests are found in the appendix and the criteria when two stress tests are similar are found and proven.

Using this relatively simple network the situations can be seen and proof can be made rather simply, but it already gives much computations. For more complex computations, the computer should do the work. This can be done by linear programming, because it is a maximisation problem with conditions. Linear programming is explained in section 3.1.

2.2.2. Second example

In this subsection, the distance of stress tests of the example in the section about stress tests (page 9) is calculated. The following network is used to get the stress tests in (2.9). The order of capacities in the stress tests is N_A, N_B, X_A, X_B, X_C, X_D.

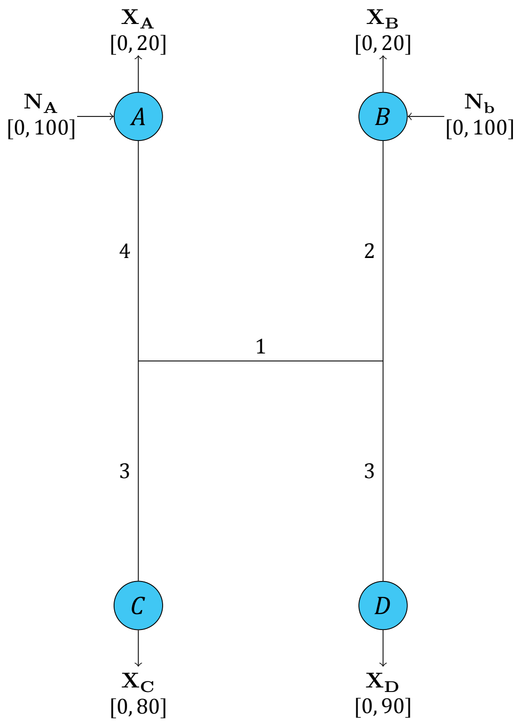


Figure 2.10: Example from the stress test section 2.1

$$s_A = \begin{pmatrix} 100 \\ 90 \\ -0 \\ -20 \\ -80 \\ -90 \end{pmatrix}, \quad s_B = \begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix}, \quad s_C = \begin{pmatrix} 10 \\ 100 \\ -20 \\ -0 \\ -0 \\ -90 \end{pmatrix} \tag{2.9}$$

Three distances can be computed from these stress tests. The matrices used for the QFD are:

$$D = \begin{pmatrix} 0 & 7 & 0 & 7 & 7 & 8 \\ 7 & 0 & 7 & 0 & 6 & 7 \\ 0 & 7 & 0 & 7 & 7 & 8 \\ 7 & 0 & 7 & 0 & 6 & 7 \\ 7 & 6 & 7 & 6 & 0 & 7 \\ 8 & 7 & 8 & 7 & 7 & 0 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & \frac{1}{8} & 1 & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{8} & 1 & \frac{1}{8} & 1 & \frac{1}{4} & \frac{1}{8} \\ 1 & \frac{1}{8} & 1 & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{8} & 1 & \frac{1}{8} & 1 & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & 1 & \frac{1}{8} \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{8} & 1 \end{pmatrix}$$

Now, the distances are calculated:

$$\begin{aligned} QFD_{\mathbf{A}}(s_A, s_B) &= \sqrt{(s_A - s_B)^T \mathbf{A} (s_A - S_B)} \\ &= \sqrt{23175} \\ &= 15\sqrt{103} \\ &\approx 152.2334 \end{aligned}$$

$$\begin{aligned} QFD_{\mathbf{A}}(s_A, s_C) &= \sqrt{(s_A - s_C)^T \mathbf{A} (s_A - S_C)} \\ &= \sqrt{17575} \\ &= 5\sqrt{703} \\ &\approx 132.5707 \end{aligned}$$

$$\begin{aligned} QFD_{\mathbf{A}}(s_B, s_C) &= \sqrt{(s_B - s_C)^T \mathbf{A} (s_B - S_C)} \\ &= \sqrt{13000} \\ &= 10\sqrt{130} \\ &\approx 114.0175 \end{aligned}$$

$QFD_{\mathbf{A}}(s_A, s_B) > QFD_{\mathbf{A}}(s_A, s_C) > QFD_{\mathbf{A}}(s_B, s_C)$, so the scenario generated from anchor point A is more similar to the scenario generated from C than from stress test with anchor point B.

2.2.3. Metric distance

The quadratic form distance looks like a well defined distance, but is it a metric distance as described by mathematicians? In this section it is shown that the QFD with the defined parametrisation is not always a metric distance. However, that is not needed for checking the similarity between stress tests, because when the distance between two stress tests is zero, then these two does not have to be the same situations. This can happen if an entry and exit point are at the same location. The situation when there is no flow in the network has distance zero to the situation when there is flow from the entry point to the exit point on the same location.

The matrix A is always symmetric if the entries are defined by $a_{ij} = 1 - \frac{d_{ij}}{d_{\max}}$, because $d_{ij} = d_{ji}$, thus $a_{ij} = a_{ji}$.

Assume matrix $A \in \mathbb{R}^{n \times n}$ is Symmetric positive definite (SPD), i.e. $x^T A x > 0$ for every non-trivial $x \in \mathbb{R}^n$. In this case, QFD_A is a metric distance, because the next four requirements hold [5]. The detailed proofs of these requirements are found in appendix B, section B.1. The first two requirements are easy to see, but the other are more difficult. For the last requirement, the triangular inequality, the theorem of Cauchy Schwarz is used for this special case.

- $d(u, v) \geq 0$
- $d(u, v) = d(v, u)$
- $d(u, v) = 0 \quad \Leftrightarrow \quad u = v$
- $d(u, v) \leq d(u, w) + d(v, w)$

In the simple example of previous subsection 2.2.2, the matrix A is not positive definite and this is not a necessarily desired property for measuring the similarity between stress tests. The QFD is a square root of a vector-matrix-vector multiplication, so it is desired that the vector-matrix-vector multiplication is non-negative. Therefore, the matrix A is required to be Symmetric positive semi-definite (SPSD), which is a less strict definition than SPD. If the matrix A is SPSD, the QFD becomes a semi-metric distance [5]. The definition of a SPSD-matrix $A \in \mathbb{R}^{n \times n}$ is $x^T A x \geq 0$ for every non-trivial $x \in \mathbb{R}^n$. The definition of a semi-metric distance is given below and the proofs are given in appendix B in section B.1.

- $d(u, v) \geq 0$
- $d(u, v) = d(v, u)$
- $d(u, u) = 0$
- $d(u, v) \leq d(u, w) + d(v, w)$

If A is a (semi-)negative definite matrix of an indefinite matrix, the QFD_A is not a distance which can be used for the similarity of stress tests.

2.2.4. Other definitions of the matrix

K. Lindenberg [10] tested some definitions of A on symmetric semi-positive definiteness, however she did not find a definition such that it is SPSD for every transport network. Perhaps flow conservation, pressure drop and/or the diameter of the pipeline should be taken into account.

The other definitions of A tested on Symmetric Positive Semi-Definiteness (SPSD) are :

$$a_{ij} = \exp\left\{\left(\frac{d_{ij}}{d_{\max}}\right)^2 - 1\right\},$$

$$a_{ij} = 1 - \sqrt{1 - \frac{d_{ij}}{d_{\max}}},$$

$$a_{ij} = 1 - \sqrt{\frac{d_{ij}}{d_{\max}}},$$

$$a_{ij} = 1 - \left(\frac{d_{ij}}{d_{\max}}\right)^2,$$

$$a_{ij} = \sqrt{1 - \frac{d_{ij}}{d_{\max}}}$$

$$a_{ij} = \exp\left\{-\frac{d_{ij}}{d_{\max}}\right\}$$

$$a_{ij} = \frac{1}{1 + \frac{d_{ij}}{d_{\max}}}$$

At Gasunie, they use $a_{ij} = 1 - \frac{d_{ij}}{d_{\max}}$ as the definition of A and they never had any problems with this measure. It is possible that some properties of the gas network is not included, which leads to a symmetric semi-positive definite matrix. These properties are not found, so another distance is examined in the next section.

2.3. Earth mover's distance

Another approach to the distance between situations is the Earth mover's distance (EMD). This distance is defined by the minimal cost to go from one situation to another. This is named the earth mover's distance, because it can be applied to the problem of earth moving: what costs more work to transport a pile of dirt of two kilo grams over 100 meters or a pile of dirt of three kilo grams over 70 meters. An illustration is seen at figure 2.11.

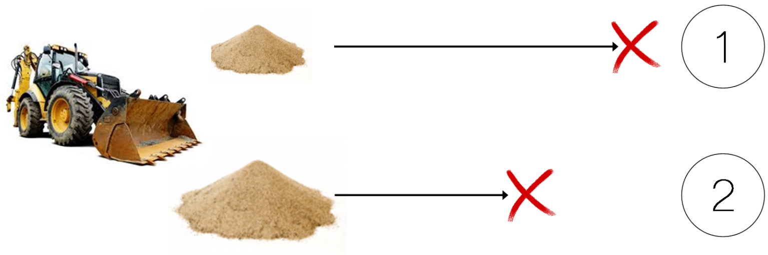


Figure 2.11: Illustration to the Earth Mover's Distance

A team at the computer science department of Stanford did a research for the application of the Earth Mover's Distance for content-based image retrieval [15]. The EMD is a metric between signatures for image retrieval in different feature spaces. A signature $\{s_j = (m_j, w_{m_j})\}$ represents a set of clusters where each cluster has a mean m_j and a fraction w_{m_j} of pixels that belong to that cluster.

Define two distributions P and Q, where $P = \{(p_1, w_{p_1})\}, \ldots, (p_m, w_{p_m})$ with p_i is the cluster representative and w_{p_i} is the weight of the cluster. $Q = \{(q_1, w_{q_1}), \ldots, (q_n, w_{q_n})\}$. The distance between p_i and q_j is d_{ij} which is given in the ground distance matrix D. $F = f_{ij}$ is the flow between p_i and q_j which is non-negative and minimises the overall cost. The total amount of work is dependent of the flow and the distance between the clusters: WORK$(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}$.

The problem given in mathematical form is:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}$$

subject to

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right)$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i} \qquad \forall i \in [1, m]$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j} \qquad \forall j \in [1, n]$$

$$f_{ij} \geq 0 \qquad \forall i \in [1, m],\ \forall j \in [1, n]$$

With this optimal flow the earth mover's distance is defined by:

$$EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{2.10}$$

The earth mover's distance is the total amount of work normalised by the total flow. The normalisation factor is the total weight of the smaller signature, this is caused by the first constraint above. In order to avoid favouring smaller signatures, the normalisation is needed when two signatures have different total weight.

2.3.1. Application

The earth mover's distance is applicable for the stress tests of the gas network. Assume there are two scenarios of which the distance between these scenarios is required. For the EMD a distinction is made between the entry and exit points. Call $N = \{N_1, \dots, N_I\}$ the set of entry points and $X = \{X_1, \dots, X_J\}$ the set of exit points. For the calculation of the EMD for stress test, the balanced capacities are known. The capacities of the difference scenario is given by the absolute difference of the capacity of each entry and exit point. The notation in equation 2.11: $c(N_i^k)$ is the capacity of entry point i in scenario k.

$$c(N_i) = |c(N_i^1) - c(N_i^2)| \qquad c(X_j) = |c(X_j^1) - c(X_j^2)| \tag{2.11}$$

The earth mover's distance is based on the flow through the network, but the flow is not known when the only given are the stress tests and properties of the network. Therefore the flow is found by minimising the amount of work under certain conditions given in (2.12). d_{ij} is the network distance (shortest path) from entry i to exit j, f_{ij} is the flow from entry i to exit j and $c(\cdot)$ is the capacity on entry or exit point of the difference scenario.

$$\min \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} f_{ij} \tag{2.12a}$$

subject to

$$\sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij} = \min \left\{ \sum_{i=1}^{I} c(N_i), \sum_{j=1}^{J} c(X_j) \right\} \tag{2.12b}$$

$$\sum_{j=1}^{J} f_{ij} \leq c(N_i) \qquad \forall i \in [1, I] \tag{2.12c}$$

$$\sum_{i=1}^{I} f_{ij} \leq c(X_j) \qquad \forall j \in [1, J] \tag{2.12d}$$

$$f_{ij} \geq 0 \qquad \forall i \in [1, I],\ j \in [1, J] \tag{2.12e}$$

The EMD is defined by the minimal amount of work that is needed to make scenario 1 from scenario 2. That follows equation 2.12a, where the amount of work is dependent on the mutual transport distances between entries and exits and the amount of flow of the difference of scenarios. Equation 2.12b ensures maximal amount of gas that flows through the network. The balance is preserved in the capacity difference, so the two sums in the minimum are the same. Equations 2.12c and 2.12d causes that the flow from entry (or exit) does not exceed the capacity on that entry (or exit). The last equation (2.12e) ensures that the flow goes from entries to exits and not the other way. In the difference scenario, the bounds of the exit and entry point does not have be taken into account, because this is not a real scenario, it is meant to calculate the distance between two scenarios.

Once the flow is found, the earth mover's distance of two stress tests u, v can be calculated by equation 2.13. The EMD is the total amount of work, normalised by the total amount of flow. Normalisation is needed, when comparing two distances with each other and there is different total flow in the scenarios.

$$EMD(u, v) = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} f_{ij}}{\sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij}} \tag{2.13}$$

On the other hand, it should be discussed whether the normalisation is needed. When no normalisation is done, the earth mover's distance equals the transport moment of the difference scenario, i.e. the amount of work to get one scenario from another.

A problem with this EMD is that the network is seen as entry and exit points with different pipelines between these points. However, there are mutual pipelines when finding the shortest path between entry and exit points. So in the network of previous section, if there is flow from point A to D and from

B to C, then the middle pipeline is used twice in both directions. In the EMD described above, the transport moment is the length of the pipe times the sum of both flows. However, the actual transport moment is product of the length of the pipeline and the difference of the flows. In order to calculate the earth mover's distance such that it is the approximated transport moment of (2.1) on page 5, the earth mover's distance should be adapted.

2.3.2. Example

The same network is given as in the stress test section (2.1.2 on page 9) and the quadratic form section (2.2.2 on page 18).

Now the shortest path matrix $D \in \mathbb{R}^{I \times J}$ is:

$$D = \begin{matrix} & X_a & X_b & X_c & X_d \\ N_a & 0 & 7 & 7 & 8 \\ N_b & 7 & 0 & 6 & 5 \end{matrix} \tag{2.14}$$

The EMD of the following stress tests are calculated:

$$s_A = \begin{pmatrix} 100 \\ 90 \\ -0 \\ -20 \\ -80 \\ -90 \end{pmatrix}, \quad s_B = \begin{pmatrix} 0 \\ 100 \\ -20 \\ -0 \\ -80 \\ -0 \end{pmatrix}, \quad s_C = \begin{pmatrix} 10 \\ 100 \\ -20 \\ -0 \\ -0 \\ -90 \end{pmatrix} \tag{2.15}$$

First, the absolute value of the stress test are given to find the flow through the difference scenarios. The definition of the i^{entry} element of the difference scenario of stress tests s_A and s_B is $v_AB(i) = |s_A - s_B|$. All the vectors of the the difference scenarios are seen in equation (2.16).

$$v_{AB} = \begin{pmatrix} 100 \\ 10 \\ 20 \\ 20 \\ 0 \\ 90 \end{pmatrix}, \quad v_{AC} = \begin{pmatrix} 90 \\ 10 \\ 20 \\ 20 \\ 80 \\ 0 \end{pmatrix}, \quad v_{BC} = \begin{pmatrix} 10 \\ 0 \\ 0 \\ 0 \\ 80 \\ 90 \end{pmatrix} \tag{2.16}$$

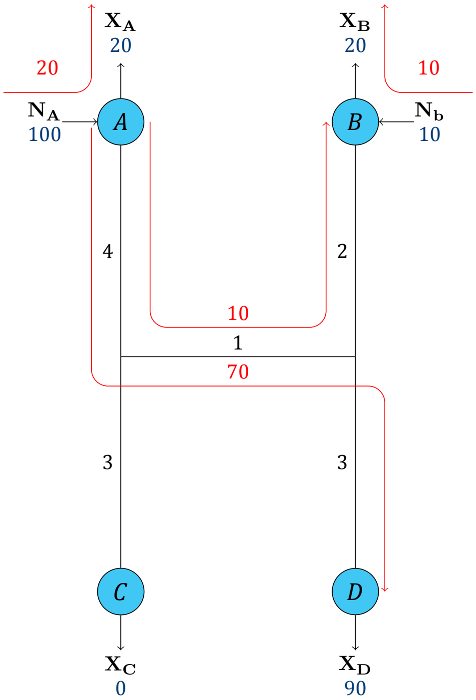
The difference of the scenarios is calculated by the absolute difference per element. Afterwards, the flow through the difference scenario can be calculated. In figure 2.12 on page 24 is the capacity from the difference scenarios given in blue at the entry and exit points. Because this is a relatively simple example, the flow under conditions (2.12) can be found by inspection. In more complex cases, the flow is found by the computer. The question how the computer can find this flow, is going to be answered after this literature study.

The flow matrices are:

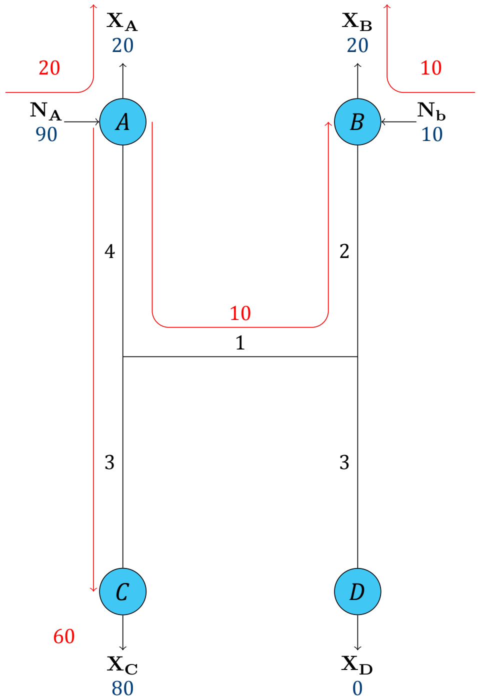
$$F_{AB} = \begin{pmatrix} 20 & 10 & 0 & 70 \\ 0 & 10 & 0 & 0 \end{pmatrix}, \quad F_{AC} = \begin{pmatrix} 20 & 10 & 60 & 0 \\ 0 & 10 & 0 & 0 \end{pmatrix}, \quad F_{BC} = \begin{pmatrix} 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

With these flow matrices, the earth mover's distance can be computed for each pair of stress tests. This is shown below:

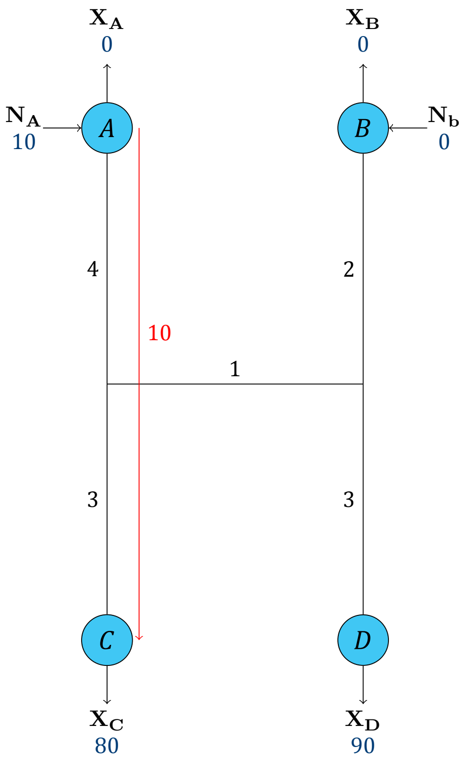
$$\begin{aligned} EMD(s_A, s_B) &= \frac{\sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij} f_{ij}}{\sum_{i=1}^{2}\sum_{j=1}^{4} f_{ij}} \text{ or } \sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij} f_{ij} \\ &= \frac{630}{110} \text{ or } 630 \\ &\approx 5.7273 \text{ or } = 630 \end{aligned}$$



(a) Difference of A and B



(b) Difference of A and C



(c) Difference of B and C

Figure 2.12: Difference scenarios for earth mover's distance

$$EMD(s_A, s_C) = \frac{\sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij}f_{ij}}{\sum_{i=1}^{2}\sum_{j=1}^{4} f_{ij}} \text{ or } \sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij}f_{ij}$$
$$= \frac{490}{100} \text{ or } 490$$
$$= 4.9 \text{ or } = 490$$

$$EMD(s_B, s_C) = \frac{\sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij}f_{ij}}{\sum_{i=1}^{2}\sum_{j=1}^{4} f_{ij}} \text{ or } \sum_{i=1}^{2}\sum_{j=1}^{4} d_{ij}f_{ij}$$
$$= \frac{70}{10} \text{ or } 70$$
$$= 7 \text{ or } 70$$

The normalisation is not needed, because the distance between $(100, 0, 0, -100)^T$ and $(0, 100, -100, 0)^T$ has to be higher than the distance between $(20, 0, 0, -20)^T$ and $(0, 20, -20, 0)^T$. This example will be further discussed in the next subsection such that it is compared with the quadratic form distance.

2.3.3. Comparison QFD and EMD

The quadratic form distance has the benefit that it does not need to know the flow pattern through the network, but the earth mover's distance calculate the distance with smaller matrices and vectors. Therefore, the final calculation of the transport moment of the EMD is less large than the calculation of the QFD.

Distance	$\mathbf{QFD_A}$		$\mathbf{EMD}$	
	exact	round-off	without norm	with norm
v_A, v_B	$5\sqrt{1455}$	190.7223	630	5.7273
v_A, v_C	$10\sqrt{130}$	145.5163	490	4.9
v_B, v_C	$55\sqrt{7}$	114.0175	70	7

Table 2.2: Comparison of distances

The EMD without normalisation has the correlation with QFD. However, this is a small set of comparison, so in future work, more stress tests are generated to compare these two measures with each other.

2.4. Diffusion distance

Another distance used for image retrieval is discussed in the article of Okada and Ling [11]. Just as in the quadratic form distance, the colours are represented in histogram bins. The diffusion distance is measured by the temperature field of each histogram bin. This distance is the sum of all dissimilarities.

This new approach handles the difference between two (multi-dimensional) histograms as a temperature field and uses a diffusion process on the field. Subsequently, the dissimilarity between the histograms is measured by the integration of a norm on the diffusion field. To create computational efficiency, a Gaussian pyramid is used to discretise the continuous diffusion process. A Gaussian pyramid [14] is a process of blurring an image. This process yields new images with less pixels than the previous image. Every new iteration of blurring produces a new image, which is called a pyramid layer. In numerical analysis, this is called the multigrid method.

The diffusion distance is defined by the sums of norms over all pyramid layers. The new distance allows cross-bin comparison, that means that it is robust for distortion (deformation, lightning change and noise, etc.). The size of the layers of the Gaussian pyramid decrease exponentially, so it is much faster than similar methods. A partial differential equation model of linear diffusion process has Gaussian convolution as a unique solution [9].

2.4.1. One dimensional space

Consider two one dimensional distributions $h_1(x)$ and $h_2(x)$ and its difference $d(x) = h_1(x) - h_2(x)$. There is not a direct metric on $d(x)$, but it is defined on an isolated temperature field $T(x,t)$. The begin temperature is $T(x,0) = d(x)$. It is known that the heat diffusion equation on an isolated field is

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} \tag{2.17}$$

with the unique solution

$$T(x,t) = T_0(x) \cdot \phi(x,t) = d(x) \cdot \phi(x,t) \tag{2.18}$$

where $\phi(x,t)$ is the Gaussian filter, which is defined by

$$\phi(x,t) = \frac{1}{\sqrt{2\pi}t} e^{-\frac{x^2}{2t^2}}. \tag{2.19}$$

The mean of the difference field is zero, so $T \to 0$ when $t \to \infty$. That is why $h_1(x)$ and $h_2(x)$ become equivalent after some time.

The distance between h_1 and h_2 is

$$\hat{K}(h_1,h_2) = \int_0^{\bar{t}} k(|T(x,t)|)\,\mathrm{d}t \tag{2.20}$$

where $\bar{t}$ is a positive constant upper bound of the integration, which can be infinity if the integral converges. The function $k(\cdot)$ is the norm that measures what $T(x,t)$ differs from zero. In this paper, the $\mathcal{L}_1$-norm is used, because it is simple in its computation and it performed good in performance of pilot studies.

Consider a simple case:

$$h_1(x) = \delta(x)$$
$$h_2(x) = \delta(x - \Delta)$$

where δ is the Dirac-delta function:

$$\delta(x) = \begin{cases} +\infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

So $d(x) = \delta(x) - \delta(x - \Delta)$, thus $d(x)$ has a positive impulse at zero and a negative impulse at Δ. For the temperature field the following expression is obtained:

$$
\begin{aligned}
T(x,t) &= T_0(x)\phi(x,t) \\
&= (\delta(x) - \delta(x-\Delta))\,\phi(x,t) \\
&= \phi(x,t) - \phi(x-\Delta,t)
\end{aligned}
$$

The $\mathcal{L}_1$ is used for $k(\cdot)$. Below is an outline of the proof to the simplified expression of $k(\cdot)$. For the detailed proof see section B.3 in Appendix B.

$$
\begin{aligned}
k(|T(x,t)|) &= \int_{-\infty}^{\infty} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \,\mathrm{d}x \\
&= 2\int_{-\infty}^{\frac{\Delta}{2}} \left(\left|\phi(x,t) - \phi(x-\Delta,t)\right|\right) \mathrm{d}x \\
&= 2\left(\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t)\,\mathrm{d}x - \int_{-\infty}^{-\frac{\Delta}{2}} \phi(x,t)\,\mathrm{d}x\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t)\,\mathrm{d}x - 1\right)
\end{aligned}
$$

2.4.2. Higher dimensions

$$\frac{\partial T}{\partial t} = \nabla^2 T \tag{2.21}$$

$$\phi(x,t) = \frac{1}{(2\pi)^{\frac{m}{2}}t} e^{-\frac{x^T x}{2t^2}} \tag{2.22}$$

The problem now is to compute $\hat{K}$. Direct computation is expensive, so use Gaussian pyramid for discretisation.

$$K(h_1,h_2) = \sum_{\ell=0}^{L} k(d_\ell(x)) \tag{2.23}$$

where

$$
\begin{aligned}
d_0(x) &= h_1(x) - h_2(x) \\
d_\ell(x) &= [d_{\ell-1}(x)\phi(x,\sigma)] \downarrow_2
\end{aligned}
$$

are different layers of the pyramid. $\downarrow_2$ is the half size down sampling, L is the number of pyramid layers and σ is the constant standard deviation for the Gaussian filter ϕ.

The $\mathcal{L}_1$-norm is used, so equation 2.23 is simplified:

$$K(h_1,h_2) = \sum_{\ell=0}^{L} \left|d_\ell(x)\right| \tag{2.24}$$

The computation complexity of $K(h_1,h_2)$ is $\mathcal{O}(N)$, where N is the number of bins in the histogram. The reason for this is that the size of d_ℓ reduces exponentially and only a small Gaussian filter ϕ is needed. Therefore, the convolution take time linear in the size of d_ℓ for each scale ℓ.

3

Gas transport based on flow

In the previous chapter, the stress tests are based on the capacities on the entry and exit points. However, the load of the gas transport is generally based on the flow through the pipelines. When describing the problem in terms of pipelines, the amount of data is increased. The contractual bounds on the entries and exits are still needed for the computations and the structure in the network is added to the known data. Apart from exit and entry points, there are also points in between which are connected to the network.

In the new representation, there are two nodes added to the problem which are the representatives of the entry and exit points. The flow from N (entry points) must equal the incoming flow at X (exit points). Below is a full description of this representation. Afterwards an example of the transformation is given (figure 3.1 and 3.2).

1. Take the nodes from the network and make edges between nodes which are connected.
2. The weight on each edge is the distance between the corresponding nodes.
3. The bounds of the flow is taken $[l, u] = [-\infty, \infty]$, because now it assumed that through each pipeline the flow can go both ways and there are no restrictions on the amount of flow (this bound will not be shown in the illustrations).
4. Connect each node which has an entry with N and put the capacity bounds of the corresponding entry point on the edge. These are the bounds of the flow through those edges.
5. Connect each node which has an exit point the same way as for entry points at X.
6. The edges connected to N and X have weight zero (this is not shown in the illustrations).

To illustrate this new representation of the gas network, the example of previous chapter is transformed into the new representation. In figure 3.1 the network is shown that was used in the last chapter and in figure 3.2 is the new representation. The two networks seem very different from each other, however they represent the same network.

The mathematical definitions describing the network are given below. These definitions will define the quality problem of the gas network.

- $V = \{N, X, P_1, \ldots, P_k\}$ is the set of nodes in the network.
- $E = \{(v_1, v_2)$: there is a direct pipeline between node v_1 and $v_2\}$ is the set of edges in the network.
- $G = (V, E)$ is the graph that describes the network by nodes and edges.
- $c : E \to \mathbb{R}$ is the function that describes the amount of flow through an edge or a set of edges.
- $l, u : E \to \mathbb{R}$ are the lower and upper bounds (respectively) on the amount of flow through the pipelines.

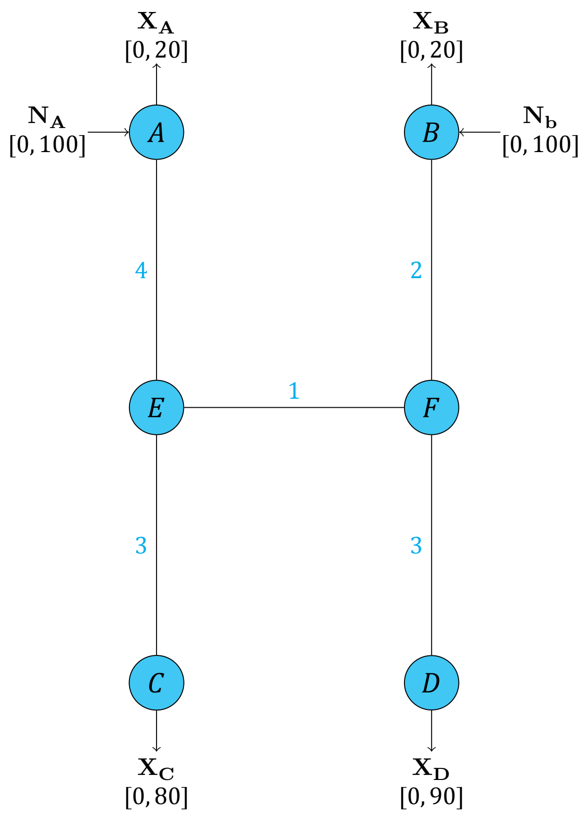


Figure 3.1: Network entry/exit representation

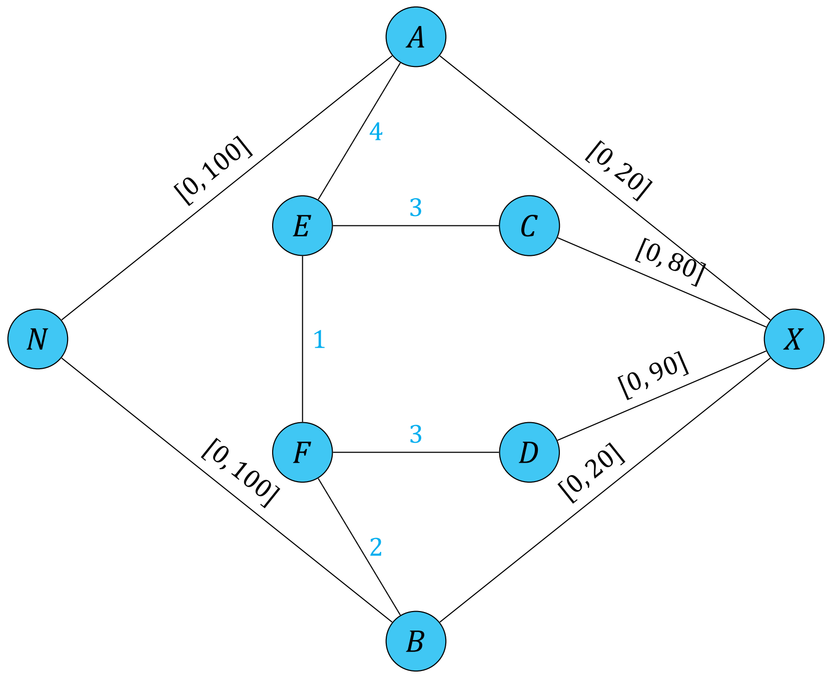


Figure 3.2: Network flow representation

- $d : E \to \mathbb{R}$ are the weights on an edges. In the gas network problem is the weight the length of the pipeline.
- For $A \subseteq V$ is $\delta^{in}(A) = \{e \in E : e$ has only one end in A where the flow towards A is positive $\}$ [2]
- For $A \subseteq V$ is $\delta^{out}(A) = \{e \in E : e$ has only one end in A where the flow leaving A is positive $\}$

The following constraints are necessary to model the gas network:

- $c(\delta^{out}(N)) = c(\delta^{in}(X))$. This ensures the flow conservation in the network.
- $c(\delta^{out}(v)) = c(\delta^{in}(v))$ for every $v \in V \backslash \{N, X\}$. The incoming flow must equal the outgoing flow of a node.
- $c(\delta^{in}(N)) = 0$. There is no incoming flow at entry points in the network.
- $c(\delta^{out}(X)) = 0$. There is no outgoing flow at exit points in the network.

Finding the severe situations by the description from above, a maximisation problem is constructed by several corresponding constraints. If these constraints are linear, then problems like this are handled by the optimisation department of mathematics by linear programming.

3.1. Linear programming

The constraints of the gas transport, which is given in previous paragraph, are all linear constraints. Linear programming (LP) models the problem of finding an certain vector x that maximises the linear function $c^T x$ under a number of linear inequalities which are given by the system $Ax \leq b$ [3]. A solution is feasible if the solution satisfies all the given constraints.

The set of solutions to the linear constraints is called a polyhedron [4]. If the polyhedron is bounded, then it is called a polytope. Thus a polyhedron $P \subseteq \mathbb{R}^n$ is a polytope if there exists lower and upper bounds $l, u \in \mathbb{R}^n$ such that $l \leq x \leq u$ for all $x \in P$.

Linear programming models a variety of practical problems and several fast solution methods are developed. The most well-known method is the simplex method, which is designed by G. B. Dantzig in the late 1940s. Furthermore, many definitions and theora are based on linear programming.

The program Matlab has an pre-programmed function called `linprog` in the optimisation toolbox [12]. This linear programming solver gives a solution to the following problem, where c, x, b, b_{eq}, l, u are vectors and A, A_{eq} are matrices.

$$\begin{array}{rrl} \min & c^T x & \\ \text{s.t.} & Ax & \leq b \\ & A_{eq}x & = b_{eq} \\ & l \leq & x \leq u \end{array} \tag{3.1}$$

The necessary inputs for the function `linprog` are c, A and b. The other vectors and matrix from above are optional. Moreover, there is the option to choose the method for solving the minimisation problem. One of the methods that can be chosen is the simplex method. The simplex method generates a sequence of feasible solutions.

3.1.1. The simplex method

The simplex method finds a solution to minimisation problems described by linear programming [3]. Despite the popularity and the simplicity of this method, the algorithm is not solvable in polynomial time in general. The worst case of the simplex algorithm has time complexity is exponential. In practice, this method is very effective and needs generally $2m$ to $3m$ iterations, where m is the number of equality constraints [21].

In Combinatorial Optimisation [3], the LP problems are of the form as in equation (3.2). The simplex algorithm can also be described by other forms of linear programming problems.

$$\text{(P)} \quad \begin{array}{rl} \max & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \tag{3.2}$$

A is an m by n matrix and rank(A) $= m$. Let T be the indices of the columns of matrix A and let $B \subseteq T$ be the indices of the basis of the columns of A, i.e. the columns of A which are given by B are linear independent and $|B| = m$. Define the matrix generated by B as A_B, the corresponding elements of c by c_B and the corresponding variables by x_B of x.

Note that there are no inequalities in the system above. Consider $A_i x \leq b_i$, where a_i is the i^{th} inequality, i.e. the i^{th} row of matrix A. In the simplex method, this inequality is replaced by $A_i x + s_i = b_i$, where s_i is a non-negative slack variable. This theory is used for the complementary slackness theorem 3.1.1, which is used for the simplex method.

Theorem (Complementary Slackness)**.** *Suppose x^* is a solution to the primal problem* (3.2) *and y^* is a solution to the dual problem* (3.3)*. Then these solutions are optimal with respect to their problem if and only if the complementary slackness holds:*

- *For every $i = 1, \ldots, m$ it holds that $y_i^* \left(b_i - a_i x^*\right) = 0$ and*
- *For every $j = 1, \ldots, n$ it holds that $x_j^* \left(a_j y^* - c_j\right) = 0$*

The simplex method uses the dual problem to find an optimal solution to the (primal) LP problem, described in equation (3.2).

$$\text{(D)} \quad \begin{array}{rl} \min & y^T b \\ \text{s.t.} & A^T y \geq c \end{array} \tag{3.3}$$

This method needs a system of inequalities and a initial solution to find a optimal solution. It is common to take $x_i = 0$ and $s_i = b_i$ as initial solution. The outline as in Combinatorial Optimisation is described in algorithm 3.1.

Algorithm 3.1 Simplex algorithm [3]

loop
  Find unique solution to $A_B^T y = c_B$
  if $A_i^T y \geq c_i$ for all $i \in T \backslash B$ **then**
    STOP, the current solution is optimal
  else
    Choose i such that $A_i^T y < c_i$
    Find the unique solution to $A_B z = A_i$
    Find the largest ε such that $x_B - \varepsilon z \geq 0$
    if ε does not exist **then**
      STOP, the LP problem is unbounded
    else
      Choose $j \in B$ such that $z_j > 0$ and the j-th component of $x_B - \varepsilon z$ is 0
      Replace B by $(B \cup \{i\}) \backslash \{j\}$ and x_B by $x_B - \varepsilon z$ and $x_i = \varepsilon$
    end if
  end if
end loop

3.1.2. Interior point algorithm

Another primal-dual method is the interior point method. Matlab can also use this method to solve a linear programming problem. The time complexity of the interior point method is in all cases polynomial time. This method takes feasible points inside the polytope, but never on the boundary. This method also works for non-linear programming [21].

The algorithm in Matlab has three steps [13]. First, in the presolve step, the unnecessary constraints are removed, the feasibility of the problem is checked and also it is checked if the problem is unbounded. In the next step, the initial point is generated. This step is important for solving the algorithm efficiently, but it can be time consuming. The last step is using the predictor-corrector to solve the Karush-Kuhn-Tucker (KKT) equations. This step is the most time consuming. The KKT conditions are a generalisation of the linear programming conditions, because these conditions are of the form $g(x) \leq 0$ and $h(x) = 0$, where $f(x)$ is optimised.

Solving the KKT problem, the Newton-Raphson method is used for the predictor step and then the corrector is calculated. The predictor-corrector algorithm iterates until a feasible point is found and where the relative step sizes are small. The default step size of the algorithm is 10^{-10}.

It should be analysed whether the interior point method is better than the simplex method for the capacity planning problem. Which of the methods is more simple to implement and which is faster. The simplex method can be solvable in polynomial time for this example.

3.2. Pressure drop

In the previous theory, it is assumed that correlation between the severeness of the transport of gas is linear with the length of the pipe. However, the pressure plays a part in this process as well. The equation related to this problem is

$$P_{in}^2 - P_{out}^2 = \frac{c \cdot L}{d^5} Q^2. \tag{3.4}$$

Most of the parameters in this equation are fixed and known [10]. d is the diameter of the gas pipe, which is $1.187m$ or smaller for HTL pipes. c is assumed to be a fixed constant, Q is the flow, that is chosen, thus fixed. The pressure at the beginning P_{in} is known, so the pressure at the end of the pipe P_{out} is now only dependent of the length of the pipe L. So the pressure drop $\Delta P = P_{in} - P_{out}$ depends only on the transport distance L.

In the article of K. Lindenberg there are plots shown which gives various relations that are given in table 3.1.

Quantity	Effect of the pressure drop by increasing quantity
Fixed flow Q	Occurs at a smaller transport distance
Diameter d	Occurs at a longer transport distance
Initial pressure P_{in}	Occurs at a longer transport distance

Table 3.1: Relation between pressure drop and transport distance with different values for quantities

An important conclusion on the plots is that the relation of transport distance and the pressure drop is not linear and thus the assumption made by Skopal et al [18] is not realistic. It follows from the figures of K. Lindenberg that linearity will only work for transport distances small enough.

This pressure drop could be taken in to account of the matrix A for a more realistic representation of the severity of the gas network. However, this problem becomes more complex, because the pressure drop equation is not linear. A solution for this non-linear problem is to linearise the pressure drop and thereby reducing the complexity.

Another addition which can be put into the model is the pressure restrictions. Besides the bounds on the capacity, in the contracts is also stated what the requirements of the pressure on the entry and exit point are. When applying the bounds of the pressure to the model, the answer on the severeness of situations can be different than the severeness based only on capacities. For the entry points of the gas network, there is an upper bound given for the pressure. The shippers inject their gas with pressure below the given upper bound, so the gas can be controlled by Gasunie in their network. At an exit point of the gas network, there is a lower and a upper bound of the pressure. Gasunie will deliver the gas with a pressure close to the lower bound, because then the energy needed to have a certain pressure at an end point is minimised.

4 Conclusions

The research problem of this project is to have a good and efficient method which computes if the gas network will suffice for every scenario which can occur. There are many factors which play a role in the scenarios. A lot of time and memory is needed to compute all scenarios.

4.1. Summary literature study

The current approach to the problem is to find stress test by taking anchor points and find the most severe situation seen from that anchor point. From that set of stress tests, the distance between those stress tests is calculated by the quadratic form distance (QFD). When two stress tests are close to each other, the set of stress tests can be reduced by taking out one of the two stress tests. By reducing the set, the amount of calculations is reduced, which cause less computation time. The computation of each scenario of the reduced set involves a lot of calculations, because many factors are involved in the gas network, like pressure, gas quality, compressor stations, diameter of the pipeline, etc.

The parametrisation of the quadratic form distance can be chosen differently then the current, however, that does not change the positive semi-definiteness of the parametrisation into strict positive definiteness. Although that means that the QFD_A is a semi-distance, this parametrisation works in practice.

There is another distance which can be used to calculate the distance between two stress tests called the earth mover's distance (EMD). This distance calculates the minimal amount of work to go from one situation to another. The research problem is about transport problems, so this distance should fit with it. However, in this distance normalises with the total amount of flow and that is not the distance that is preferred. So the normalisation should be left out. Another adjustment, which is not found yet, is that the EMD does not take different directions on a pipeline into account. This will cause in some cases a higher EMD than the transport moment, which should be equal to the EMD.

Besides the entry and exit representation of the network, the network can also be represented by a graph with the network points as nodes and the mutual pipelines as edges. The entry and exit points are represented by edges and two extra nodes. Such a network can be solved by linear programming, because it is an optimisation problem with a connected graph.

Pressure plays a big role in the gas transport network. With not enough pressure, the gas cannot flow through the network. When the gas has to be transported over a long distance, the pressure decreases. The pressure equation is not linear, therefore the addition of pressure to the methods will cause bigger computations.

4.2. Research questions

There are questions that arise from the research problem and the literature found with this subject. The research questions are about the methods which can be used for solving the research problem. The analysis in the research project involves the current method and comparison with other methods. Furthermore, it has to be discussed whether some dependencies and different structures of the transportation of gas has to be taken into account for the model or method. Some examples of network dependencies are the amount of flow, the quality of the gas, pressure, pipe length, boundaries on capacities of entry points, exit points and pipelines, etc. There are different structures in the gas network, for example loops and compressor stations. The outline of the research questions is below.

1. What is the current method to ensure the capacity of the gas network?
    - Are all the severe situations taken into account in the current selection method?
    - Which other measures can be used for this method?
    - What is the criterion of distance when two situations are similar?
2. Which other method(s) can be used for the problem?
    - What are the advantages and the disadvantages of the methods?
    - How can the computer find the flow through the network for the earth mover's distance?
3. What is the network structure of the Dutch gas network?
    - Which complex structures should be taken into account?
    - How are loops in the gas network handled in each method?
4. How can the description of the network transform from capacities on the exit en entry points to a description by flows on the network?
    - What are the advantages and the disadvantages of describing the problem in flows instead of capacities on the entries and exits?
    - Which algorithm is most suitable to check if the gas network suffices for each realistic scenario in the flow representation?
5. Which part does the pressure play in the gas network?
    - How can the dependency of pressure on the gas network be added to each method?
    - How can the pressure equations be changed such that it can be used for linear programming?

The focus on the research is on the quality of each method in comparison with other methods.

4.3. Continuation of the project

The continuation of this research project is to answer the research questions. That involves further research on details and applications of the methods found in this literature study.

The details and application of the earth mover's distance has to be investigated more, so that the directions in the network is taken into account.

The methods have to be programmed in Matlab. Then the analysis can be done on the comparison of the methods. The comparison will be done on different networks: is the difference between methods changed on bigger networks? This difference of methods can be measured in the time complexity, time of computation, memory load, simplicity and the ability to tune parameters.

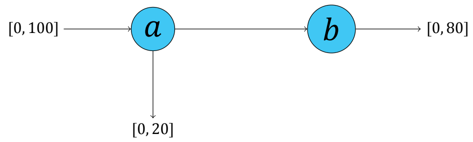
Furthermore, the application of the flow description to the network planning problem has to be done. The algorithms of linear programming gives one solution, but the stress test algorithm gives a set of solutions such that this set can be checked in detail.

The networks that are going to be used in the continuation of this project are given in figure 4.1.

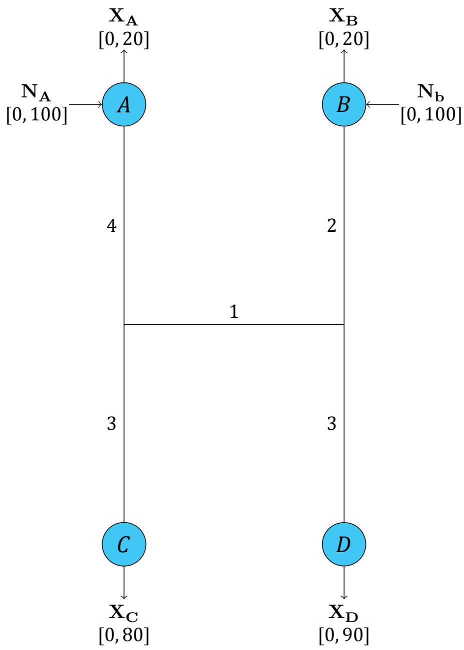
One pipeline network The one pipeline network has only one pipeline and has two exit points and one entry point. The bounds are known and given in the figure. The length of the pipeline is not important for finding the severeness of the transport situations and to check whether the situation is feasible.

H-network This network is a H-shaped network, where the gas flows can flow in different directions on the pipelines. There are two entry points at A and B and every point has an exit point. The lengths of each pipeline segment is given and the bounds on entries and exits.

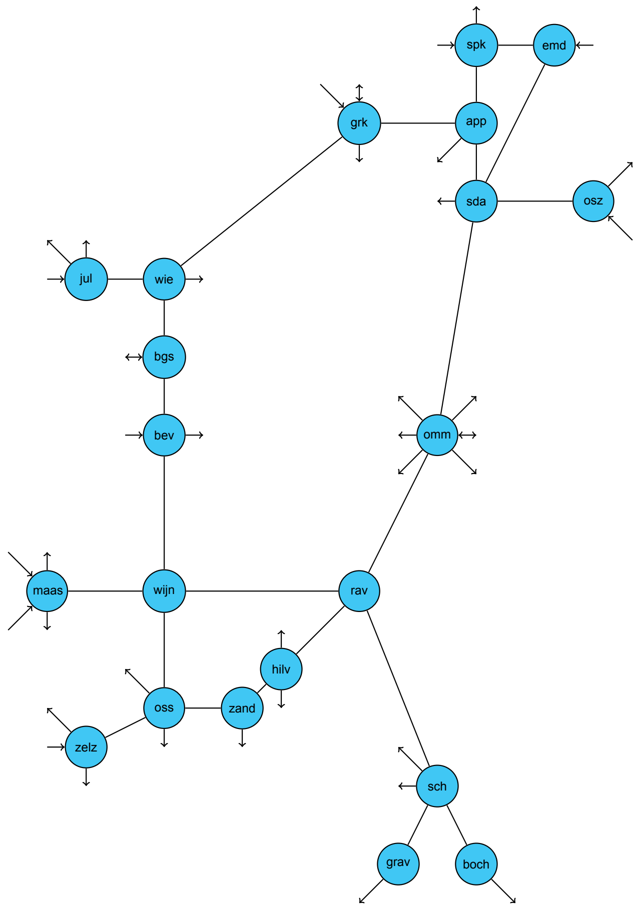
Shopping cart network This network is an approximation of the real Dutch gas network. In the figure, points correspond to important locations of the network. The lengths of the pipelines and bounds on the entry and exit points are not given in this network, because of clarity of the picture. Some points have more than one exit point, because of different markets. Some points have a double arrow, which means that there is a gas storage connected to that point. This is different than a separate entry and exit point, because there is no flow from the storage to the storage via the corresponding point. Furthermore, there are loops in this network, which makes the flow transportation more complex.



(a) One pipeline network



(b) H-network



(c) Shopping cart network

Figure 4.1: Standard gas network examples

A

Assumptions

- A network has no negative transport distances.
- The contracts with the shippers is not violated, this means that
  - there is a balance of feed in and take off of gas in the network, and
  - the minimal and maximal feed in and take off of gas is known and will not be violated.
- The length of the pipe and the amount of gas in the transport network are the most important quantities for measuring the severity of the gas transportation.
- The gas in a pipeline can flow in two directions and the quantities are the same in both directions. This means that the transport moments of both directions are equal.
- There are no bounds on the amount of flow through each pipeline segment.

B

Proofs

B.1. Proof of metric QFD

If A is a SPSD-matrix (or SPD-matrix):

$$(u-v)^T A(u-v) \geq 0 \qquad \text{(A is SPSD)}$$
$$\sqrt{(u-v)^T A(u-v)} \geq 0$$
$$QFD_A(u,v) \geq 0$$

So non-negativity holds for the quadratic form distance

If A is a SPSD-matrix (or SPD-matrix), then

$$\begin{aligned} QFD_A(u,v) &= \sqrt{(u-v)^T A(u-v)} \\ &= \sqrt{(v-u)^T A(v-u)} \\ &= QFD_A(v,u) \end{aligned}$$

So symmetry holds.

If A is a SPSD-matrix (or a SPD-matrix), then the theorem of Cauchy-Schwarz holds in this case (equation B.1).

$$x^T Ay \leq \sqrt{x^T Ax}\sqrt{y^T Ay} \tag{B.1}$$

The proof of this special case of the theorem is given below. Here it is used that a SPSD-matrix (or SPD) can be written as $A = B^T B$ and there exist a unique B such that it is symmetric [22].

$$\begin{aligned} 0 &\leq \left(\sqrt{x^T Ax}By - \sqrt{y^T Ay}Bx\right)^T \left(\sqrt{x^T Ax}By - \sqrt{y^T Ay}Bx\right) \\ &= x^T Ax(By)^T By - 2\sqrt{x^T Ax}\sqrt{y^T Ay}(Bx)^T By + y^T Ay(Bx)^T Bx \\ &= x^T Axy^T B^T By - 2\sqrt{x^T Ax}\sqrt{y^T Ay}x^T B^T By + y^T Ayx^T B^T Bx \\ &= x^T Axy^T Ay - 2\sqrt{x^T Ax}\sqrt{y^T Ay}x^T Ay + y^T Ayx^T Ax \\ &= 2x^T Axy^T Ay - 2\sqrt{x^T Ax}\sqrt{y^T Ay}x^T Ay \end{aligned}$$

This can be used for the triangular inequality $d(u,v) \leq d(u,w) + d(w,v)$:

$$\begin{aligned} (QFD_A(u,v))^2 &= (u-v)^T A(u-v) \\ &= (u-w+w-v)^T A(u-w+w-v) \\ &= (u-w)^T A(u-w) + 2(u-w)^T A(w-v) + (w-v)^T A(w-v) && \text{(A is symmetric)} \\ &\leq (u-w)^T A(u-w) + 2\sqrt{(u-w)^T A(u-w)}\sqrt{(w-v)^T A(w-v)} + (w-v)^T A(w-v) \\ & && \text{(Cauchy-Schwarz, see equation B.1)} \\ &= \left(\sqrt{(u-w)^T A(u-w)} + \sqrt{(w-v)^T A(w-v)}\right)^2 \end{aligned}$$

So the triangular inequality holds for the QFD with a symmetric (semi-)positive definite matrix.

If $A \in \mathbb{R}^{n \times n}$ is a positive definite matrix, then $QFD_A(u,v) = 0$ if and only if $u = v$ for every $u, v \in \mathbb{R}^n$. The proof is given below.

Assume $u = v$:

$$\begin{aligned} QFD_A(u,u) &= \sqrt{(u-u)^T A(u-u)} \\ &= 0 \end{aligned}$$

Assume $QFD_A(u,v) = 0$, then

$$\begin{aligned} QFD_A(u,v) &= 0 \\ &\Leftrightarrow \\ \sqrt{(u-v)^T A(u-v)} &= 0 \\ &\Leftrightarrow \\ (u-v)^T A(u-v) &= 0 \\ &\Leftrightarrow \\ (u-v)^T A(u-v) = 0 \ \forall u, v & \\ \text{or } u = v & \end{aligned}$$

As A is positive definite, the first possibility does not hold. Therefore $d(u,v) = 0 \quad \Leftrightarrow \quad u = v$.

If A is semi-positive definite, the equality $d(u,u) = 0$ holds for every u holds for the QFD_A.

B.2. Network with general bounds

Consider again the network from figure 2.8. The bounds for the capacities on the entry and exit points are now taken arbitrary, which lead to the following network (figure B.1). $l(P)$ is the lower bound of the capacity on entry or exit point P. The upper bound of the capacity of entry or exit point P is denoted as $u(P)$.

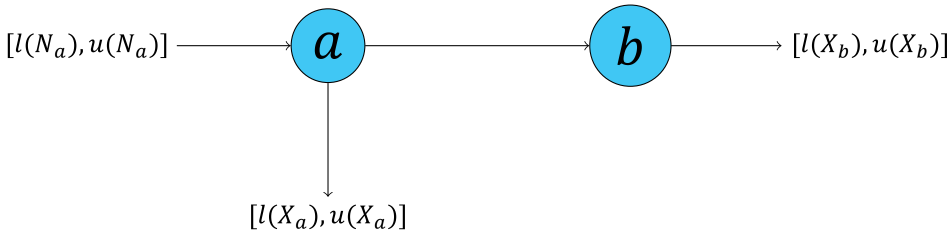


Figure B.1: Simple example, with general capacities

Define $q = \min\{u(N_a), u(N_b)\}$, such that it cause a most severe situation for this network. q is the maximal feasible flow on the pipeline from a to b. In figure B.2 the general feasible and most severe situation is seen. The network is feasible if the bounds of the capacities at the entry and exit points are satisfied. This means $x \in [l(X_a), u(X_a)]$, $q \in [l(X_b), u(N_b)]$ and $q + x \in [l(N_a), u(N_a)]$. Together with the conservation of flow and the definition of q, the following requirements are needed:

1. $x \in [l(X_a), u(X_a)]$
2. $q \in [l(X_b), u(N_b)]$
3. $q + x \in [l(N_a), u(N_a)]$
4. $c(X_a) + c(X_b) = c(N_a)$ (conservation of flow)
5. $q = \min\{u(N_a), u(N_b)\}$

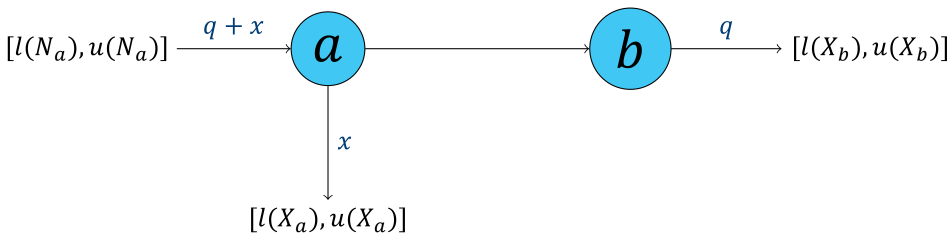


Figure B.2: Simple example, with general capacities and general situation

The bounds on $q + x$ are not necessary, if we change the constraint for x into the following bounds:

$$x \in [l(X_a), \min\{u(X_a), u(N_a) - q\}] = [l(X_a), \min\{u(X_a), \max\{0, u(N_a) - u(N_b)\}\}]$$

Let the following hold:

1. $x \in [l(X_a), \min\{u(X_a), \max\{0, u(N_a) - u(N_b)\}\}]$
2. $c(X_a) + c(X_b) = c(N_a)$
3. $q = \min\{u(N_a), u(N_b)\}$

Two cases are distinguished to prove that the bounds of $q + x$ and q are satisfied. First consider the bounds of $q + x$:

- Assume $u(N_a) \leq u(N_b)$

$$\begin{aligned} q + x &= u(N_a) + x \\ &\geq u(N_a) && \text{(capacities and their bounds are positive)} \\ &\geq l(N_a) \end{aligned}$$

  So the lower bound is satisfied for this case.

$$\begin{aligned} q + x &= u(N_a) + x \\ &\leq u(N_a) + \min\{c(X_a), \max\{0, u(N_a) - u(N_b)\}\} \\ &= u(N_a) + \min\{u(X_a), 0\} \\ &= u(N_a) + 0 \\ &= u(N_a) \end{aligned}$$

  And also the upper bound is satisfied in this case.

- Assume $u(N_a) > u(N_b)$

$$\begin{aligned} q + x &= u(N_b) + x \\ &\geq u(N_b) + u(X_a) \\ &\geq c(X_b) + c(X_a) \\ &= c(N_a) && \text{(conservation law)} \\ &\geq l(N_a) \end{aligned}$$

  The upper bound is satisfied in this case.

$$\begin{aligned} q + x &= u(N_b) + x \\ &\leq u(N_b) + \min\{u(X_a), \max\{0, u(N_a) - u(N_b)\}\} \\ &\leq u(N_b) + u(N_a) - u(N_b) \\ &= u(N_a) \end{aligned}$$

So $q + x \in [l(N_a), u(N_a)]$ is satisfied. The bounds of q is done the same way:

- Assume $u(N_a) \leq u(N_b)$

$$\begin{aligned} q &= u(N_a) \\ &\geq c(N_a) \\ &= c(X_a) + c(X_b) && \text{(flow conservation)} \\ &\geq c(X_b) \\ &\geq l(X_b) \end{aligned}$$

  The lower bound is satisfied for q in this case. Also the upper bound is satisfied:

$$\begin{aligned} q &= u(N_a) \\ &\leq u(N_b) && \text{(assumption)} \end{aligned}$$

- Assume $u(N_a) > u(N_b)$. Then $q = u(N_b)$, so it follows directly.

In conclusion, if this general network is satisfied by

1. $x \in [l(X_a), \min\{u(X_a), \max\{0, u(N_a) - u(N_b)\}\}]$
2. $c(X_a) + c(X_b) = c(N_a)$
3. $q = \min\{u(N_a), u(N_b)\}$

then all situations with the same q are similar severe and satisfy the requirements of the stress test.

B.3. Proof of simplified function in diffusion distance

$$
\begin{aligned}
k(|T(x,t)|) &= \int_{-\infty}^{\infty} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \, \mathrm{d}x \\
&= \int_{-\infty}^{\frac{\Delta}{2}} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \, \mathrm{d}x + \int_{\frac{\Delta}{2}}^{\infty} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \, \mathrm{d}x \\
&= \int_{-\infty}^{\frac{\Delta}{2}} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \, \mathrm{d}x + \int_{-\infty}^{\frac{\Delta}{2}} \left|\phi(x,t) - \phi(x-\Delta,t)\right| \, \mathrm{d}x \qquad \text{(Symmetric in } \tfrac{\Delta}{2}\text{)} \\
&= 2\int_{-\infty}^{\frac{\Delta}{2}} \left(\left|\phi(x,t) - \phi(x-\Delta,t)\right|\right) \mathrm{d}x \\
&= 2\int_{-\infty}^{\frac{\Delta}{2}} \left(\phi(x,t) - \phi(x-\Delta,t)\right) \mathrm{d}x \\
&= 2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - 2\int_{-\infty}^{\frac{\Delta}{2}-\Delta} \phi(x-\Delta,t) \, \mathrm{d}(x-\Delta) \\
&= 2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - 2\int_{-\infty}^{-\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x \\
&= 2\left(\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \int_{-\infty}^{-\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \left(\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x + \int_{-\infty}^{-\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x\right)\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \left(\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x + \int_{\frac{\Delta}{2}}^{\infty} \phi(x,t) \, \mathrm{d}x\right)\right) \qquad \text{(Symmetry in zero)} \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \int_{-\infty}^{\infty} \phi(x,t) \, \mathrm{d}x\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t^2}\right) \mathrm{d}x\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - \frac{1}{\sqrt{2\pi t}} \cdot \sqrt{2t^2}\sqrt{\pi}\right) \\
&= 2\left(2\int_{-\infty}^{\frac{\Delta}{2}} \phi(x,t) \, \mathrm{d}x - 1\right)
\end{aligned}
$$

C

List of Terms and Acronyms

Glossary

Calorific value energy value or volume of energy by gas (in MJ/m^3) [7]. 1

Shipper party who is recognised by the network operator of the national grid and consequently has programme responsibility [17]. 1

Stress test transport situation which is extreme for the transport of gas and is within the contractual bounds of the shippers. 7

Transport moment quantity of the transport load, dependent on the amount of flow through the pipes and the length of the pipes. 5

Transport situation balanced combination of the quantities on the entry and exit points, the capacity is the most common used quantity. 5

Acronyms

EMD Earth mover's distance. 21

G-gas Groningen gas (Wobbe index ≤ 44.4). 1

GTS Gasunie Transport Services. 5

H-gas Natural gas with a Wobbe index ≥ 49.0. 1

HTL High pressure grid (Dutch: Hoofdtransportleidingnet). 1, 5

KKT Karush-Kuhn-Tucker. 33

L-gas Natural gas with a Wobbe index between 44.4 and 47.2. 1

LP Linear programming. 32

QFD Quadratic form distance. 15

RTL Intermediate pressure grid (Dutch: Regionale transportleidingnet). 1

SPD Symmetric positive definite. 19

SPSD Symmetric positive semi-definite. 19, 20

TSO Transmission system operator. 1

Bibliography

[1] Emile J. L. Chappin. Aardgas en distributie, 2016. URL http://eduweb.eeni.tbm.tudelft.nl/TB141E/?aardgas-transport. Online; accessed 4-April-2017.

[2] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. Wiley Interscience, 1998. page 12.

[3] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. Wiley Interscience, 1998. page 325-335.

[4] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. Wiley Interscience, 1998. page 204-206.

[5] Michel Marie Deza and Elena Deza. Encyclopedia of distances. In *Encyclopedia of Distances*, pages 1–583. Springer, 2009.

[6] Gasunie. Netwerkkaart uk, 2016. Received on intranet Gasunie, accessed 4-April-2017.

[7] Gasunie. Blending station, 2017. Online; accessed 4-April-2017.

[8] Gasunie. Jaarverslag 2016, 2017. URL http://report2016.gasunie.nl/. Online; accessed 4-April-2017.

[9] Jan J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984. ISSN 1432-0770. doi: 10.1007/BF00336961. URL http://dx.doi.org/10.1007/BF00336961.

[10] Kimberley Lindenberg. Comparing severe gas transport situations through the network. Master's thesis, TU Delft, 2015.

[11] H. Ling and K. Okada. Diffusion distance for histogram comparison, May 11 2010. URL https://www.google.com/patents/US7715623. US Patent 7,715,623.

[12] Mathworks. Solve linear programming problems, 2017. URL https://nl.mathworks.com/help/optim/ug/linprog.html?requestedDomain=www.mathworks.com. Online; accessed on 17-May-2017.

[13] Mathworks. Linear programming algorithms, 2017. URL https://nl.mathworks.com/help/optim/ug/linear-programming-algorithms.html. Online; accessed on 17-May-2017.

[14] Ross Moore and Nikos Drakos. Gaussian and laplacian pyramids, 2002. URL https://www.cs.utah.edu/~arul/report/node12.html. Online; accessed 14-March-2017.

[15] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[16] Gasunie Transport Services. Entry- & exitcapaciteit, 2016. URL https://www.gasunietransportservices.nl/over-gts/gastransport/entry-exitcapaciteit. Online; accessed 27-March-2017.

[17] Gasunie Transport Services. Shipper worden, 2017. URL https://www.gasunietransportservices.nl/shippers/klant-worden/shipper-worden. Online; accessed 24-March-2017.

[18] Tomáš Skopal, Tomáš Bartoš, and Jakub Lokoč. On (not) indexing quadratic form distance by metric access methods. In *Proceedings of the 14th International Conference on Extending Database Technology*, pages 249–258. ACM, 2011.

[19] Jarig J. Steringa and Marco Hoogwerf. Generating network stress tests for hydraulic evaluation. *SIMONE Congress*, 13, apr 2016.

[20] Jarig J. Steringa, Marco Hoogwerf, Harry Dijkhuis, et al. A systematic approach to transmission stress tests in entry-exit systems. In *PSIG Annual Meeting*. Pipeline Simulation Interest Group, 2015.

[21] Eric W. Weisstein. Simplex method, 2006. URL http://mathworld.wolfram.com/SimplexMethod.html. Online; accessed on 17-May-2017.

[22] Eric W. Weisstein. Positive definite matrix, 2017. URL http://mathworld.wolfram.com/PositiveDefiniteMatrix.html. Online; accessed 28-March-2017.