Increasing Energy Efficiency in Industrial Plants: Optimising Heat Exchanger Networks

Author:
Miten Mistry

Supervisor:
Dr Ruth Misener

Second Marker:
Dr Panos Parpas

Abstract

We live in a world where energy efficiency is of increasing importance. Heat exchanger networks form a fundamental part of bringing energy efficiency to many industrial processes but solvers can only guarantee local optima for these models and, due to numerical difficulties, approximations are regularly applied and are not accounted for as problem sizes are scaled up.

In this project we turn our attention to the log mean temperature difference, a defining function of heat exchangers but one that is often approximated and overlooked in practice. We identify that the reciprocal of this function is a better choice when modelling these problems and, on analysis of this function, we find that it possesses a mathematical property: convexity. With this property we develop an algorithm that overcomes the numerical difficulties and show that it will converge to the global optimum. The first of its kind. We also present an implementation of our algorithm. Our evaluation shows that we can place a significantly better lower bound on the solution compared to existing solvers and that for certain models convergence is quick.

Acknowledgements

I would like to thank:

- Dr Ruth Misener, my project supervisor, for her advice, guidance and time throughout the project. Her enthusiasm with regards to the project and passion for her own field of research made the project all the more enjoyable.
- Dr Panos Parpas, my second marker, for his advice and feedback on my work
- Dr Steffen van Bakel, my personal tutor, for his support and guidance throughout the past four years
- my friends for their advice and help throughout university
- my parents, Narendra and Prafulla, and my sister, Keerti, for their unconditional love and support during my studies

Contents

1 Introduction

1.1 Motivations

The heating and cooling of liquids is a staple part of many industrial processes. With a present day focus on the reduction of CO_2 emissions such as the Climate Change Act 2008 [8], the use of fossil fuels becomes a less viable option and the recycling of excess heat is favoured.

There are many heat intensive industries in the UK including glassmaking, refineries and food and drink however much of the energy required for industrial processes is ultimately emitted again to the environment in the form of heat [11]. In the UK, 73% of industrial demand is for heating [10]. The problem of wasted heat is not limited to the UK. Figure 1.1 shows that of the 24.7 Quads[1] used by US industries, 4.95 Quads, about 20%, is rejected.

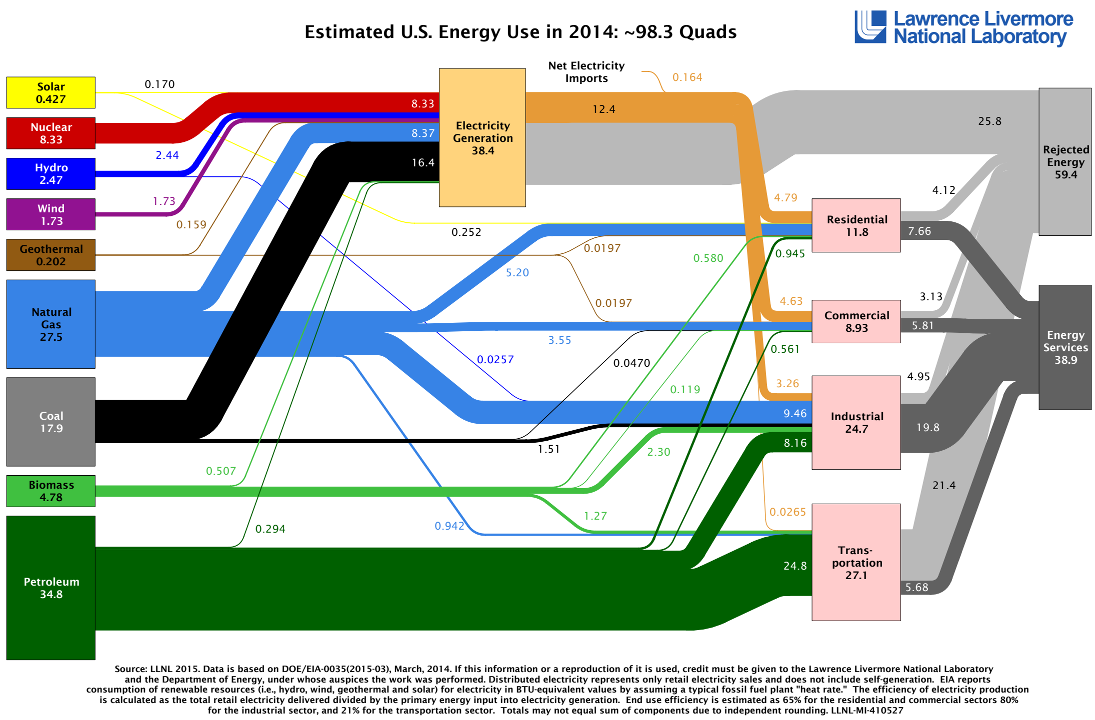


Figure 1.1: Estimated US energy use [30]

There are ways in which excess heat can be reused including

1. Heat re-use on site
2. Heat delivery over-the-fence, to another site
3. Conversion of heat to electricty

the above is ranked in order of CO_2 abatement potential [11]. On site re-use is favoured since it maximises abatement. We encounter heat losses in transport for over-the-fence

[1] 1 Quad $= 1.055 \times 10^{18}$ joules

heat delivery with additional power requirements for pumping, efficiency issues arise for conversion of heat to electricity. Heat exchangers form a integral part of on site re-use of heat in industrial processes.

The heating and cooling of streams found in industry come with associated operating costs. From a business perspective these costs should be minimised. Re-using heat from hot processes to heat the cold processes can reduce the amount spent on external heating and cooling, here we find that the investment into the recycling of excess heat agrees from both a business front and a environmental front.

The problem we have is finding the optimal pairings of hot and cold streams of a given process that allows us to minimise the running cost, this forms the study of heat exchanger network synthesis (HENS). These problems have been studied for many years with the introduction of the idea in 1944 [3] and early work in the field in 1965 [24]. The idea behind the problem is easy to grasp however finding the solution is hard, this is due to the combinatorial structure of the model. Numerical difficulties are encountered in the solving of the model as there is a function, the log mean temperature difference (LMTD), present that is indeterminate for certain values. Common approaches to overcoming numerical issues are to: use suitable (non-indeterminate) approximations e.g. the Chen approximation [5] or adding a small ε to parameters to prevent the indeterminate cases from arising [23] however both of these approaches introduce errors into the model.

1.2 Objectives

The HEN model belongs to the class of mixed integer nonlinear programs (MINLP) for which the solvers cannot guarantee the finding of global solutions, the best we have is a local solution which may be global but we cannot always verify this. Current approaches also make use of approximation methods surrounding LMTD however these approximations introduce errors which can become more significant as problem sizes scale up.

The class of mixed integer linear programs (MILP) have solvers that can guarantee global convergence. Our objectives are to analyse the model to **create a suitable MILP approximation** and, using the robustness of the associated solvers, **propose and implement an algorithm** that uses this approximation to solve the model iteratively with a MILP solver by improving the approximation including those involving LMTD.

1.3 Contributions

The main contributions made are:

- analysis of the HEN model with a justification for use of the reciprocal of LMTD opposed to LMTD
- mathematical analysis of LMTD and its reciprocal classifying results relating to bounds and shape. The significant results we prove are: concavity of LMTD and strict convexity of its reciprocal

- proposal of a new algorithm which we show to converge to the global solution of the HEN optimisation model proposed by Yee and Grossmann.
- an implementation of the algorithm using Python and the abstract modelling tool Pyomo.

The mathematical results we prove and the algorithm we propose form part of a a conference/workshop presentation, *Solving MINLP with Heat Exchangers: Special Structure Detection and Large-Scale Global Optimisation*, at:

- 13^{th} EUROPT Workshop on Advances in Continuous Optimisation[2] (07/2015)
- 22^{nd} International Symposium on Mathematical Programming[3] (07/2015)

[2] `http://www.ismp2015.org`
[3] `http://www.maths.ed.ac.uk/hall/EUROPT15/`

2 Background

2.1 Mathematical Analysis

In this section we will establish mathematical definitions and theorems that will be fundamental when it comes to proving results with rigour. The definitions and theorems presented here are common to most: analysis, calculus, linear algebra and optimisation; text books, we recommend the following books for reference: *Analysis: with an introduction to proof*, *Calculus*, *Introduction to Linear Algebra* and *An Introduction to Optimization* respectively.

2.1.1 Sets

Definition 2.1.
Let $\Omega \subseteq \mathbb{R}^n$. Ω is an open set if

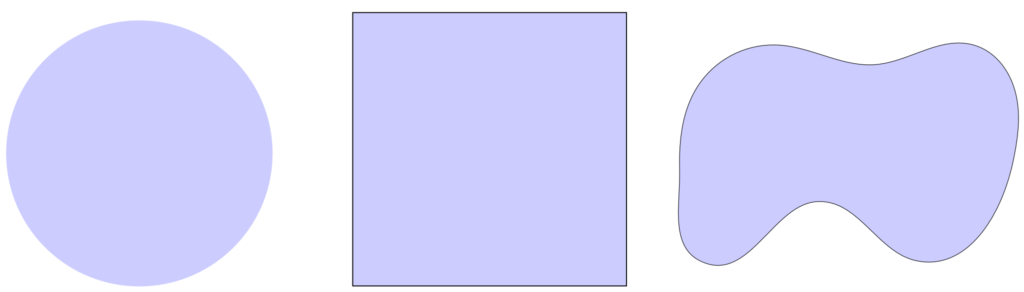
$$\forall\, \mathbf{x} \in \Omega\; \exists\, r > 0 \;\textit{s.t.}\; \forall\, \mathbf{y} \in \mathbb{R}^n : ||\, \mathbf{x} - \mathbf{y}\, || < r \implies \mathbf{y} \in \Omega$$

Definition 2.1 states that we can select a point x in our set and form a ball of non-zero radius r that is completely contained in our set. Intervals and circles form the balls of $\mathbb{R}$ and $\mathbb{R}^2$ respectively.

Definition 2.2.
Let $\Omega \subseteq \mathbb{R}^n$. Ω is a convex set if

$$\forall\, \lambda \in [0,1] : \mathbf{x}, \mathbf{y} \in \Omega \implies \lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in \Omega$$

Definition 2.2 states that if we have a convex set then the line segment between any two points contained in the set is contained in the set as shown by fig. 2.1. Convex sets play an important role when analysing in an optimisation context as if we are in the interior of a convex set, we can move in any direction.



(a) (b) (c)

Figure 2.1: Convex sets figs. 2.1a and 2.1b and a non convex set fig. 2.1c

2.1.2 Limits and Continuity

The idea of a limit is very important in analysis of functions. We have certain functions where the value at a given point is not obvious e.g. $\frac{\sin(x)}{x}$ at $x = 0$. The following definition allows us to evaluate whether the value of a function exists at a given point or not.

Definition 2.3.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set.
The limit of $f : \Omega \to \mathbb{R}$ at $\mathbf{a} \in \Omega$ is L if

$$\forall \varepsilon > 0 \, \exists \, \delta > 0 \text{ s.t. } \forall \mathbf{x} \in \Omega : 0 < ||\mathbf{x} - \mathbf{a}\,|| < \delta \implies |f(\mathbf{x}) - L| < \varepsilon$$

This is denoted

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = L$$

Definition 2.3 says if the limit of f is L at $\mathbf{a}$ then we can pick any path leading to $\mathbf{a}$ but not including $\mathbf{a}$ and as we get arbitrarily close to $\mathbf{a}$ the value of the function gets closer to L. The important part of the definition is that it says that all paths lead to the same value.

We don't need to prove every limit from first principles as in general we can evaluate the function directly. We require the further analysis for indeterminate values such as $\frac{0}{0}$.

Theorem 2.1.
Let $\Omega_1, \Omega_2 \subseteq \mathbb{R}$ be open, $f : \Omega_1 \to \mathbb{R}$ and $g : \Omega_2 \to \mathbb{R}$. If

$$\lim_{x \to c_1} f(x) = K$$
$$\lim_{y \to c_2} g(y) = L$$

where $c_1 \in \Omega_1$, $c_2 \in \Omega_2$, K and L are finite, then

$$\lim_{(x,y) \to (c_1,c_2)} [f(x)g(y)] = \left[\lim_{x \to c_1} f(x)\right] \left[\lim_{y \to c_2} g(y)\right] = KL$$

We now define continuity. In the univariate case, this can be loosely defined as being able to draw the function without having to lift our pen.

Definition 2.4.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set. $f : \Omega \to \mathbb{R}$ is continuous at $\mathbf{a} \in \Omega$ if

$$\forall \varepsilon > 0 \, \exists \, \delta > 0 \text{ s.t. } \forall \, \mathbf{x} \in \Omega : ||\mathbf{x} - \mathbf{a}\,|| < \delta \implies |f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$$

A function is called continuous over Ω if it is continuous $\forall \, \mathbf{x} \in \Omega$.

There is a subtle difference between definition 2.3 and definition 2.4. We have that definition 2.4 includes $\mathbf{a}$ in the set that we are evaluating over whereas definition 2.3 does not. This means that we have the following result

$$f \text{ is continuous at } \mathbf{a} \implies \text{the limit of } f \text{ at } \mathbf{a} \text{ is } f(\mathbf{a})$$

i.e. f is defined to be equal to its limit if it is continuous

If we have a function, f, that is continuous at all points except for some $\mathbf{x}_0$ and we can show that the limit at $\mathbf{x}_0$ exists and is L then adding $f(\mathbf{x}_0) = L$ to our definition of f gives a continuous definition.

2.1.3 Univariate Differentiability

Differentiability is very important in the study of optimisation as derivatives characterise the rate of change of a function. The value of the derivative of a function can not only be used to find optimal points but it can be used to reason other properties such as bounds on a function.

For the scope of this investigation we shall use the following theorem and the differentiation results mentioned after.

Theorem 2.2.
Let $\Omega \subseteq \mathbb{R}$ be open and let $f, g : \Omega \to \mathbb{R}$ be differentiable then we have

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x) \qquad \text{(Product Rule)}$$

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}, \qquad g(x) \neq 0 \qquad \text{(Quotient Rule)}$$

$$\left(\frac{f(g(x))}{g(x)}\right)' = g'(x)f'(g(x)) \qquad \text{(Chain Rule)}$$

Derivatives of common functions:

$$(\ln(x))' = \frac{1}{x}, \qquad x > 0$$

$$(e^x)' = e^x, \qquad x \in \mathbb{R}$$

$$(\sin(\theta))' = \cos(\theta), \qquad \theta \in (0, {}^{\pi}/_{2})$$

$$(\cos(\theta))' = -\sin(\theta), \qquad \theta \in (0, {}^{\pi}/_{2})$$

Any further derivatives we establish in this report can be found using the above results and theorem 2.2.

In section 2.1.2, we stated that the definition of a limit allows us to evaluate indeterminancies such as $\frac{0}{0}$. Proving this directly from definition can be hard, the following theorem allows us to check for the existence of a '$\frac{0}{0}$' limit and its value using derivatives.

Theorem 2.3 (L'Hôpital's Rule)**.**
Let $f, g : [a, b] \to \mathbb{R}$ be differentiable on (a, b) and let $c \in (a, b)$.

If

$$f(c) = g(c) = 0$$

then

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

The application of l'Hôpital's rule extends further than taking one derivative as we have that the derivatives of f and g can also be differentiable hence if $f'(c) = g'(c) = 0$ and both of these functions are differentiable then we can take derivatives once more and evaluate the limit using second order derivatives. The same process can be used again if necessary.

L'Hôpital's rule allows us to evaluate the quotient rule mentioned in theorem 2.2, if $g(c) = 0$ and $f'(c)g(c) - f(c)g'(c) = 0$.

2.1.4 Bivariate Differentiability

The definitions presented here have similar definitions for the multivariate case however we will only be analysing bivariate functions hence, for simplicity, only bivariate definitions are established.

Definition 2.5.
Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$. The first order partial derivatives at $\mathbf{x} \in \Omega$ with respect to x_1 and x_2 are

$$f'_{x_1}(x_1, x_2) = \lim_{h \to 0} \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h}$$
$$f'_{x_2}(x_1, x_2) = \lim_{h \to 0} \frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h}$$

respectively.
The gradient of f at $\mathbf{x} \in \Omega$ is defined as

$$\nabla f(x_1, x_2) = \begin{pmatrix} f'_{x_1}(x_1, x_2) \\ f'_{x_2}(x_1, x_2) \end{pmatrix}$$

When taking calculating f'_{x_i} we treat all variables x_j, $j \neq i$ as constant and take a univariate derivative with respect to variable x_i. The validity of this process can be seen by comparing the definition of a partial derivative with the definition of a univariate derivative.

For a univariate function, the derivative (also the gradient in this case) tells us the rate of change of a function as we increase x. The bivariate gradient ∇f can also be used to tell us the rate of change of function f.

For a bivariate function f the rate of change of f along distinct directions $\mathbf{d}_1$ and $\mathbf{d}_2$ is not necessarily the same.

Definition 2.6.
Let $\Omega \subseteq \mathbb{R}^2$, $f : \mathbb{R}^2 \to \mathbb{R}$, $\mathbf{x} \in \Omega$, $\mathbf{d} \in \mathbb{R}^2$ and $\mathbf{d} \neq \mathbf{0}$. The directional derivative of f at $\mathbf{x}$ along the direction $\mathbf{d}$ is

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

The directional derivative can be calculated easily if $\nabla f(\mathbf{x})$ is known as the following can be shown

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \mathbf{d}^T \nabla f(\mathbf{x})$$

If $||\mathbf{d}|| = 1$ then we define $\mathbf{d}^T \nabla f(\mathbf{x})$ as the rate of change of f along the direction $\mathbf{d}$.

The above relate to the bivariate generalisations of the derivative. The bivariate generalisation of the univariate second derivative is given by the Hessian.

Definition 2.7.
Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$. The second order partial derivatives at $\mathbf{x} \in \Omega$ with respect to x_i and x_j are

$$f''_{x_i x_j} = \left(f'_{x_i}\right)'_{x_j}, \qquad\qquad i, j \in \{1, 2\}$$

The Hessian of f at $\mathbf{x} \in \Omega$ is defined as

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} f''_{x_1 x_1}(x_1, x_2) & f''_{x_1 x_2}(x_1, x_2) \\ f''_{x_2 x_1}(x_1, x_2) & f''_{x_2 x_2}(x_1, x_2) \end{pmatrix}$$

For definition 2.7, we have that the Hessian matrix of f is symmetric if all of the second order partial derivatives are continuous (*Clairaut's theorem* or *Schwarz's theorem*).

Definition 2.8.
Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$. f is said to be n-times continuously differentiable on Ω if it is n-times differentiable on Ω and all partial derivatives of order n are continuous.

We write $f \in \mathcal{C}^n$ if it is n-times continuously differentiable.

We have only defined the second order partial derivatives therefore when it comes to continuous differentiability, we will consider at most the set $\mathcal{C}^2$. Note that $\mathcal{C}^n \subset \mathcal{C}^{n+1}$ i.e. p^{th} order continuous differentiability does not necessarily mean that we don't have $(p+1)^{\text{th}}$ order continuous differentiability.

If we show that $f \in \mathcal{C}^2$ then we know we have a symmetric Hessian.

2.1.5 Matrices

The Hessian of bivariate function f is a square matrix. The Hessian of a function can allow us to identify properties of f. The following definitions and theorem will help to show properties of the functions we will analyse.

Definition 2.9.
$\mathbf{A} \in \mathbb{R}^{n\times n}$ *is said to be:*

- *positive semi-definite if* $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$*, for all* $\mathbf{x} \in \mathbb{R}^n$
- *negative semi-definite if* $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$*, for all* $\mathbf{x} \in \mathbb{R}^n$
- *positive definite if* $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$*, for all* $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$
- *negative definite if* $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$*, for all* $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$

Definition 2.10.
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ *and* $0 < k \leq n$*.*

- *A minor of order k, denoted Δ_k is obtained by deleting: $(n-k)$ rows of $\mathbf{A}$ and $(n-k)$ columns of $\mathbf{A}$; and calculating the determinant of the underlying $k \times k$ matrix.*
- *Δ_k is principal if the set of deleted row indices is equal to the set of column indices.*
- *The leading principal minor of $\mathbf{A}$ of order k, denoted D_k, is the minor of order k obtained by deleting the last $(n-k)$ rows and the last $(n-k)$ columns.*

Theorem 2.4.
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric and D_k be the associated leading principal minor of order k. Then $\mathbf{A}$ is:

- *Positive definite* $\iff D_k > 0, \quad$ *for all* $1 \leq k \leq n$
- *Negative definite* $\iff (-1)^k D_k > 0, \quad$ *for all* $1 \leq k \leq n$

2.1.6 Convex Functions

Convex functions are very important in optimisation as their presence in a model can make finding a minimiser easier due to their structure. We will use the properties of convex functions to justify an approximation.

Definition 2.11.
Let $\Omega \subseteq \mathbb{R}^n$ be a convex set and let $f : \Omega \to \mathbb{R}$.
f is convex if

$$\forall\, \mathbf{x}, \mathbf{y} \in \Omega, \forall\, \lambda \in [0,1] : f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

f is strictly convex if

$$\forall\, \mathbf{x}, \mathbf{y} \in \Omega, \forall\, \lambda \in [0,1], \mathbf{x} \neq \mathbf{y} : f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

f is concave if $(-f)$ is convex

f is strictly concave if $(-f)$ is strictly convex

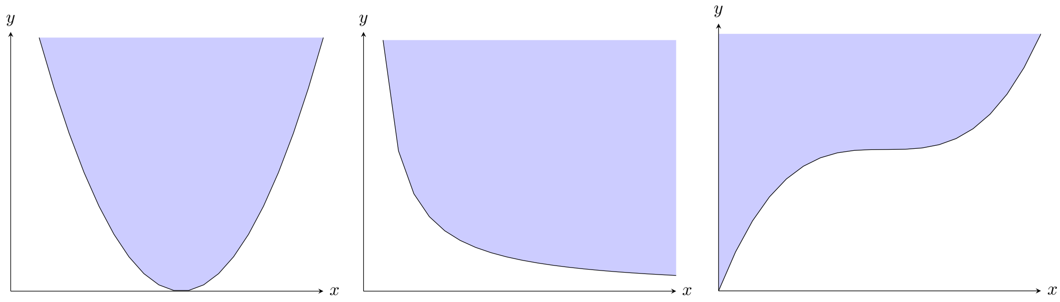
Since a concave function is the negation of a convex function, a general result for a convex function gives a general result for a concave function. Inituitively we can identify convex functions by inspecting their graphs. As the name suggests, there is a relation between a convex function and a convex set.

Definition 2.12.
Let $\Omega \subseteq \mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$. The epigraph of f is the set of points on or above its graph defined as

$$\operatorname{epi} f = \left\{ (x_1, ..., x_n, y)^T \;\middle|\; \mathbf{x} \in \Omega, y \in \mathbb{R}, y \geq f(\mathbf{x}) \right\}$$

Theorem 2.5.
Let $\Omega \subseteq \mathbb{R}^n$ be a convex set and $f : \Omega \to \mathbb{R}$.

$$f \text{ is convex } \iff \operatorname{epi} f \text{ is a convex set}$$



(a) (b) (c)

Figure 2.2: Graphs of convex functions figs. 2.2a and 2.2b, a non convex function fig. 2.2c and their epigraphs

Theorem 2.6.
Let $f_1, f_2, ..., f_n$ be convex functions, defined on a convex set $\Omega \subset \mathbb{R}^n$ then

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_n(\mathbf{x})\}$$

is convex over Ω

Theorem 2.7.
Let $\Omega \subset \mathbb{R}^n$ be an open convex set, $f : \Omega \to \mathbb{R}$ and $f \in \mathcal{C}^1$ then f is convex on Ω if and only if for all $\mathbf{x}, \mathbf{y} \in \Omega$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

f is strictly convex if and only if for all $\mathbf{x}, \mathbf{y} \in \Omega$, $\mathbf{x} \neq \mathbf{y}$

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$$

Proposition 2.1.
Let $(l, u) \subseteq \mathbb{R}$ be a convex set, $f : (l, u) \to \mathbb{R}$, $f \in \mathcal{C}^2$ and $a \in \Omega$.
If f satisfies the following two conditions

$$f'(a) = 0$$
$$\forall x \in (l, u) : \quad |x - a| > 0 \implies f''(x) > 0$$

then

$$\forall x \in (l, u) : \quad |x - a| > 0 \implies f(x) > f(a)$$

Theorem 2.8.
Let $\Omega \subset \mathbb{R}^n$ be an open convex set, $f : \Omega \to \mathbb{R}$ and $f \in \mathcal{C}^2$. If for all $\mathbf{x} \in \Omega, \nabla^2 f(\mathbf{x})$ is:

- *positive semi-definite then f is convex on Ω*
- *negative semi-definite then f is concave on Ω*
- *positive definite then f is strictly convex on Ω*
- *negative definite then f is strictly concave on Ω*

Theorem 2.7 relates a continuously differentiable convex function to its gradient stating that any tangent hyperplane will be less than or equal to the function.

Theorem 2.8 relates a twice continuous differentiable convex function to its Hessian. The result here is useful as we have an alternative approach to proving convexity, which can be hard to prove directly from definition.

2.2 Optimisation

We provide an overview of many fundamental optimisation concepts, we recommend the following text books for reference: *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications* and *An Introduction to Optimization*.

When we optimise a model we are looking to find an optimal value of some objective function f over a set Ω.

2.2.1 Models

There are many ways in which we can define an optimisation model. In general, an optimisation model is specified by:

- an objective function $f : \mathbb{R}^n \to \mathbb{R}$

- a vector of variables $\mathbf{x} \in \mathbb{R}^n$
- a set of inequality constraints $g(\mathbf{x}) \leq 0$ where $g : \mathbb{R}^n \to \mathbb{R}^p$
- a set of equality constraints $h(\mathbf{x}) = 0$ where $h : \mathbb{R}^n \to \mathbb{R}^m$

We have stated above that $\mathbf{x} \in \mathbb{R}^n$ however the constraints that we place may define a subset.

Optimisation problems can have multiple equivalent formulations however transformations allow us to create a standard form.
Given a general optimisation model we have the following equivalent forms

- $\max f(\mathbf{x}) \equiv -\min -f(\mathbf{x})$
- $g(\mathbf{x}) \geq k \equiv h(\mathbf{x}) \leq 0$, where $h(\mathbf{x}) = k - g(\mathbf{x})$

Definition 2.13.
Let $\mathbf{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$.

$$\begin{aligned} & \min f(\mathbf{x}) \\ s.t. & \\ & h(\mathbf{x}) = 0 \\ & g(\mathbf{x}) \leq 0 \end{aligned}$$

We define the above as the standard form of an optimisation problem

2.2.2 Continuous Optimisation

Definition 2.13 makes no assumptions about the nature of the functions f, g and h. Continuous optimisation is the study of optimisation problems consisting of: a continuous objective function and continuous constraints. i.e. f, g and h are continuous.

In the class of continuous optimisation problems, we get convex optimisation problems. A convex optimisation problem is given by definition 2.13 where f, g and h are all convex. This is a special case of continuous optimisation as we have that any local minimiser of the optimisation problem is a global minimiser.

We also get nonconvex continuous problems, this is where any of f, g or h are nonconvex. These problems can have many local optima and proving that an optimal solution is the global solution can be very hard. Being continuous, we can make use of properties of the gradient.

If we can reformulate the problem as a convex or series of convex problems through approximations or partitioning of the domain, finding a solution can be easier. This approach increases the complexity of the problem.

2.2.3 Mixed Integer Programs (MIPs)

A mixed integer program (MIP) is an optimisation problem in which we place integrality constraints on a subset of the variables, i.e. these variables belong to a discrete set.

First we formally define a mixed integer program

Definition 2.14.
Let $\mathbf{x} \in \mathbb{R}^n$, $\mathcal{I} \subseteq \{ 1, 2, ..., n \}$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$.

$$\begin{aligned} & \min f(\mathbf{x}) \\ s.t. & \\ & h(\mathbf{x}) = 0 \\ & g(\mathbf{x}) \leq 0 \\ & x_i \in \mathbb{Z}, \quad \forall i \in \mathcal{I} \end{aligned}$$

We define the above as a mixed integer program

Definition 2.14 differs from definition 2.13 by adding a restriction on the feasible space where a set of variables are restricted to the integers.

We can split the class of mixed integer programs into: mixed integer linear programs (MILPs) and mixed integer nonlinear programs (MINLPs).

A MILP is an optimisation problem that has a linear objective function and all constraints are linear. The linearities in a MILP allow us to formulate it in an alternative, more informative, form to that of definition 2.14 using matrix notation.

Definition 2.15.
Let $\mathbf{x}, \mathbf{c} \in \mathbb{R}^n$, $\mathcal{I} \subseteq \{ 1, 2, ..., n \}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$.

$$\begin{aligned} & \min \mathbf{c}^T \mathbf{x} \\ s.t. & \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & x_i \in \mathbb{Z}, \quad \forall i \in \mathcal{I} \end{aligned}$$

We define the above as a mixed integer linear program

A MINLP is an optimisation problem in which there exists nonlinearities in the cost function or any of the constraints. Definition 2.14 gives the general formulation of a MINLP.

2.2.4 Modelling Techniques in Integer Programming

Binary Variables

Definition 2.16.
z *is a binary variable if the only values it can take are in* $\{ 0, 1 \}$

Binary variables are very important concepts in MIPs as they allow us to represent choice and existence mathematically.

For example, assume that we have a model in which we have n options each with cost c_i $i \in \{1, ..., n\}$ and we are only allowed to make one choice. We can model this using binary variables. We formulate the model as

$$\min \sum_{i=1}^{n} c_i \cdot z_i$$

$$\text{s.t.}$$

$$\sum_{i=1}^{n} z_i = 1$$

$$z_i \in \{0, 1\}, i \in \{1, ..., n\}$$

In the above formulation we add the summation constraint, this says that our binary variables sum to 1 but they can only be 0 or 1 hence only one binary variable equals 1, characterising only being able to choose one option. We also have that there is a cost on with each choice, the objective summation gives the correct value as all other non active binary variable are equal to 0.

Binary variables have some meaning associated with them and we use this to form our constraints. While binary variables are a powerful tool they come with a lot of overhead. The above model can be solved by testing every binary variable and checking the solution, an exhaustive approach. This is what a search for a solution can become. Any practically applicable model will have a lot of binary variables and we may not have any restrictions placed on them e.g. no constraints like $\sum z_i = 1$, making them independent of each other.

The introduction of binary variables create discontinuities in a model therefore reasoning the switching of binary variables based on the continuous functions is not an easy task. In general, finding a solution becomes a combinatorial problem and while there are methods to improve our search, they may not succeed in doing so, in this case we reduce to the exhaustive approach which can lead to a large number of separate cases e.g. A MIP with 15 binary variables that are independent of each other results in $2^{15} = 32768$ combinations, each of which correspond to a different problem. Since binary variables cause us to search branches of a tree without necessarily providing a guarantee of reaching the global solution on the current branch, binary variables are the bottleneck in finding global optima.

Big-M Constraints

The assignment of $z = 0$ to a binary variable usually means that the object we associate with the binary variable is inactive. There are still associated constraints in the model that we want to ensure are feasible as the actual values applied to the constraints are irrelevant under the assignment of $z = 0$. Here big-M constraints are used and they take the form

$$g(\mathbf{x}) \leq M(1 - z)$$

where M is chosen so that we always have $g(\mathbf{x}) \leq M$. Naturally the use of big-M constraints require an upper bound on g.

2.2.5 Solving MIPs

The naive approach to solving would be to check every possible combination of integer values and solve the underlying continuous problem. In the worst case a solver will do this. A common way in which trees are searched in mixed integer optimisation is the branch and bound method.

The branch and bound method uses relaxation which where a discrete variable is allowed to become continuous.

Definition 2.17.
Let $x \in \mathbb{Z}$ be an integer variable bounded from below and above by l and u respectively. We relax the variable x by letting $x \in \mathbb{R}$ and $l \leq x \leq u$

For example, we make the following transformation

$$\begin{aligned} \text{Integrality constraint: } & x_i \in \{1, 2, 3, 4, 5\} \\ \text{Relaxed constraint: } & x_i \in [1, 5] \end{aligned}$$

The branch and bound process is given by:

1. Relax integer variables and solve - if this is infeasible so is the original problem since the relaxed formulation includes the original formulation.
2. In the relaxed optimal solution, if there is a non integral value for an integer variable, x_i, say c solve two subproblems with the added constraints $l \leq x_i \leq \lfloor c \rfloor$ and $\lceil c \rceil \leq x_i \leq u$ respectively. (Branching)
3. Solve subproblems using the same proces

If a feasible solution is found in a branch this solution is compared with the best optimal value (initially ∞) and the lower is kept as the best solution. If we find an infeasible solution that has a value that is greater than the optimal value, we know that on branching the solution will only get worst therefore we no longer search and discard the branch this is called pruning. Pruning is what allows us to find a solution in less time.

The process mentioned above is an overview of the process and there are further improvements that are made to the algorithm such as selecting the 'best' variable to branch on.

Problems with MINLPs
When we relax the problem or fix a set of integer variables we have a continuous problem to solve. If we have a linear problem then it can be solved using the simplex method [6]. If we have a nonlinear problem and it is convex then a local solution is a global solution. Issues are encountered when the problem is nonconvex this is given by having nonconvexities in the objective or any of the constraints e.g. $x \cdot y$ is nonconvex. A solver cannot always guarantee convergence to a global optimum only a local one. This is a problem as there can be a significant difference between the global optimum and a local optimum.

When we solve these problems, we want to find the global optimum however the current methods in MINLP solvers are not always able to provide this. One the other hand, MILP solvers can guarantee convergence to a global solution and robust MILP solvers exist. For this reason, a MILP approximation of a MINLP may find a better solution. A MILP approximation for a MINLP is given by all nonlinearities of the MINLP being replaced by linear approximations. We do introduce errors through reformulation into a MILP but tighter approximations can allow us to represent the MINLP better and therefore converge to the global solution.

2.2.6 Approximations

Sometimes an approximation to a function is easier to work with than the function itself. for example we have that

- a convex MINLP is easier to solve than a nonconvex MINLP therefore we may make convex approximations to the nonconvex functions
- a function may have indeterminacies over the feasible space therefore an approximation without this property may prove to be a better choice to avoid 'NaN' errors

Approximations to functions come with an approximation error which affects the solution we find. There are many different kinds of approximations including:

1. piecewise linear functions - this is where we partition the domain and use a linear approximation for each partition
2. outer approximation - this is used for a convex function where we approximate f with $\hat{f}$ with $\hat{f} \leq f$
3. convex hull - this is where we model the function as a convex set and allow the value of the function to be in the set, an example of a convex hull is given by the McCormick hull discussed in section 2.2.9

Error reduction can prove to be hard when we approximate functions however we get special cases that make it easier to reason how the error reduces as we improve our approximation these are under and over estimators.

Definition 2.18.
Let $\Omega \subseteq \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ and $f \in \mathcal{C}$. We call an estimator, $\hat{f}$, for f a

- *over estimator if $\hat{f}(\mathbf{x}) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$*
- *under estimator if $\hat{f}(\mathbf{x}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$*

2.2.7 Piecewise Linear Functions

A piecewise linear approximation for a univariate function f is formed by selecting a set of breakpoints and approximating the function by using the line segments between consecutive breakpoints. This idea can be extended to a bivariate function where line segments become triangles.

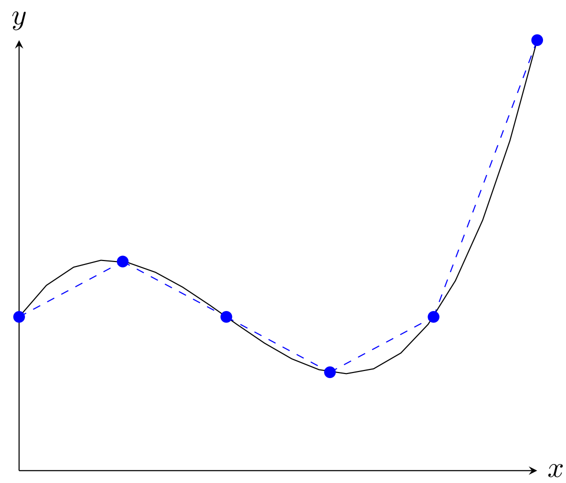


Figure 2.3: A piecewise linear function (dashed) with the original function (solid). Break points are indicated by the points

As fig. 2.3 shows, a piecewise linear function can have a different definition depending on which interval we are in. Binary variables are a natural fit when it comes to modelling piecewise linear functions as we want to model the fact that we are in a specific interval.

We will show how a univariate piecewise linear function is modelled and then extend this to the bivariate case.

There are a number of ways in which we can model a univariate piecewise linear function, we define the convex combination method as we will present the bivariate form of this method. The formulations and diagrams given here are similar to those presented by Geißler et al.

Definition 2.19.
Let $\mathbf{x}_i \in \mathbb{R}^n$, $i \in \{1, ..., m\}$. A convex combination of vectors $\mathbf{x}_i$ is the vector $\mathbf{x}$ such that

$$\mathbf{x} = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i$$
$$\sum_{i=1}^{m} \lambda_i = 1$$
$$\lambda_i \geq 0, \quad \forall i \in \{1, ..., m\}$$

The convex combination method for creating a piecewise linear function, $\hat{f}$, for $f : [l, u] \to \mathbb{R}$ over a set of ordered breakpoints $\{x_0, x_1, ..., x_n\}$ is formed by introducing binary variables z_i, $i \in \{1, ..., n\}$ and adding the constraints

$$x = \sum_{i=0}^{n} \lambda_i x_i, \quad \sum_{i=0}^{n} \lambda_i = 1, \quad \lambda \geq 0 \tag{2.1}$$

$$\hat{f}(x) = \sum_{i=0}^{n} \lambda_i f(x_i) \tag{2.2}$$

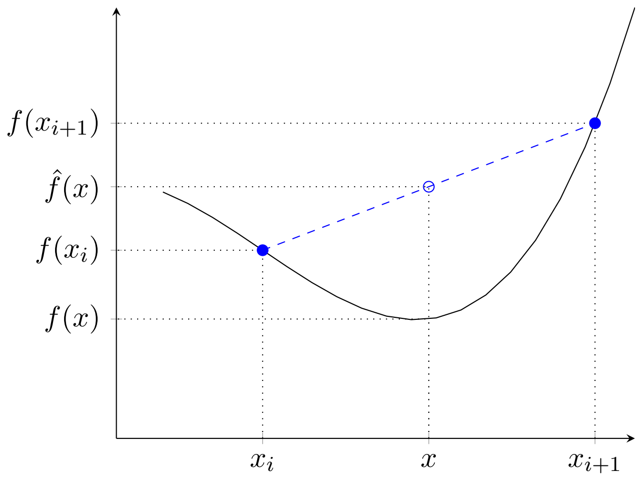


Figure 2.4: How the values of f compare with the values of $\hat{f}$ in interval i

$$
\begin{aligned}
&\lambda_0 \leq z_1 \\
\forall\, i \in \{\, 1, ..., n-1 \,\} : \quad &\lambda_i \leq z_i + z_{i+1} \\
&\lambda_n \leq z_n \\
&\sum_{i=1}^{n} z_i \leq 1
\end{aligned}
\tag{2.3}
$$

Figure 2.4 shows how a univariate piecewise linear approximation is constructed with respect to the function f. A piecewise approximation always gives intermediate values that are inside the endpoints but as we see from the figure this property is not necessarily a property of the function. A piecewise approximation is always correct at the breakpoints. In the formulation, we assign interval i with variable z_i if z_i is active then we are in interval i. We can only ever be in a single interval by the summation in eq. (2.3). The remaining constraints of eq. (2.3) allow the correct λ_i to vary thereby ensuring that our piecewise function does look something like fig. 2.3.

Piecewise linear functions are extended to the bivariate case using triangles opposed to lines. The definitions and formulations presented here have been adapted from an equivalent multivariate definition [17].

Definition 2.20.
Let $\mathcal{D} \subset \mathbb{R}^2$ be a compact set. A continuous bivariate function $\phi : \mathcal{D} \to \mathbb{R}$ is called piecewise linear if it can be written in the form

$$\phi(x) = \phi_S(x) \quad \text{for } x \in S \quad \forall\, S \in \mathcal{S}$$

with affine functions ϕ_S for a finite set of triangles $\mathcal{S}$ that partitions $\mathcal{D}$

Before we define the constraints that form the bivariate piecewise linear function we define:

- $\mathcal{S}$, the set of triangles that partition $\mathcal{D}$

- $n = |\mathcal{S}|$, the number of triangles
- $\mathcal{V}(T) = \left\{ \overline{x}_0^T, \overline{x}_1^T, \overline{x}_2^T \right\}$, the set of vertices of triangle T
- $\mathcal{V}(\mathcal{S}) = \left\{ \overline{x}_1^{\mathcal{S}}, ..., \overline{x}_m^{\mathcal{S}} \right\} = \bigcup_{T \in \mathcal{S}} \mathcal{V}(T)$, the entire set of vertices of $\mathcal{S}$
- $z_1, ..., z_n$, binary variables to decide which triangle we are in

A piecewise linear approximation $\hat{f}$ for bivariate function f using the convex combination method is given by

$$x = \sum_{j=1}^{m} \lambda_j \overline{x}_j^{\mathcal{S}}, \quad \sum_{j=1}^{m} \lambda_j = 1, \quad \lambda_j \geq 0 \tag{2.4}$$

$$\hat{f}(x) = \sum_{j=1}^{m} \lambda_j f(\overline{x}_j^{\mathcal{S}}) \tag{2.5}$$

$$\lambda_j \leq \sum_{\left\{ i \,\middle|\, \overline{x}_j^{\mathcal{S}} \in \mathcal{V}(T_i) \right\}} z_i, \qquad j = 1, ...m \tag{2.6}$$

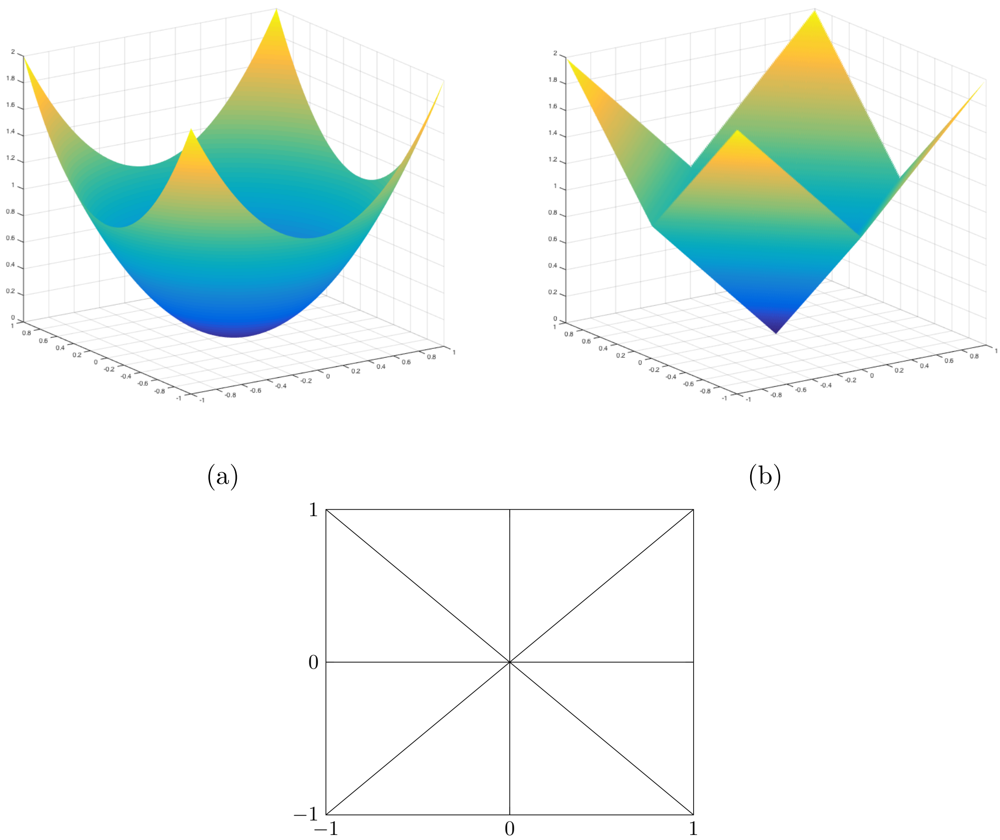
$$\sum_{i=1}^{n} z_i \leq 1 \tag{2.7}$$

Equations (2.4) and (2.5) are analogous to the univariate case, we take a convex combination of the vertices of the active triangle to form x, this same convex combination is applied to the function evaluated at the vertices of the same triangle. A vertex can be associated with more than one triangle as shown by fig. 2.5c where the central vertex is a member of all triangles.

A piecewise linear function is an over estimator for a convex function and an under estimator for a concave function. Geißler et al. showed the following

Proposition 2.2.
If f is convex over S, then $\epsilon_o(f, S)$ can be obtained by solving a convex optimisation problem. If f is concave over S, then $\epsilon_o(f, S) = 0$ holds.

In proposition 2.2, S is the set over which we have defined f and ϵ_o is the maximum over approximation error between a piecewise linear approximation, ϕ (for f) and convex function f. This proposition clearly shows that a piecewise linear approximation can have a over estimation error therefore any piecewise linear approximation can partially over approximate f. Negating everything in proposition 2.2, tells us that the maximum underapproximation error, ϵ_u, is 0 for a convex function therefore a piecewise linear approximation is always an overestimator. Similar reasoning shows that a piecewise linear approximation is always an underestimator for a concave function.



(a) (b)

(c)

Figure 2.5: A bivariate function, a piecewise linear function for it and the partitioning used to generate the piecewise function, each diamond in fig. 2.5b is two triangles

2.2.8 Outer Approximations

An outer approximation [18] is applicable to a convex/concave function and can be seen as the opposite of a piecewise linear approximation for these functions. An outer approximation needn't be linear however we will treat all outer approximations as linear.

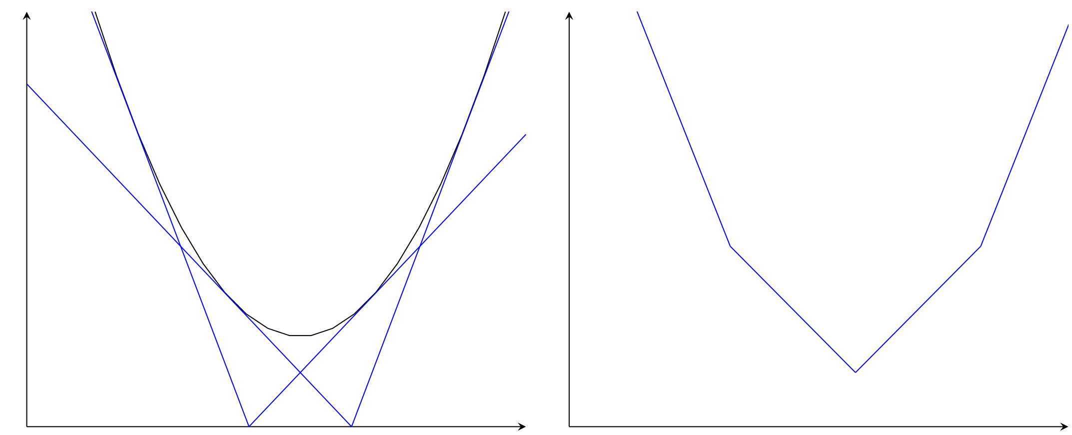
We form an outer approximation for a convex function using linear supports.

Definition 2.21.
A linear support for a convex function f at the point $\mathbf{x}_0$ is a linear function $\hat{f}$ with the properties

$$\begin{aligned} f(\mathbf{x}) &\geq \hat{f}(\mathbf{x}), \quad \forall \mathbf{x} \neq \mathbf{x}_0 \\ f(\mathbf{x}_0) &= \hat{f}(\mathbf{x}_0) \end{aligned}$$

Definition 2.21 implies that $\operatorname{epi} f \subseteq \operatorname{epi} \hat{f}$.

An outer approximation is created by taking the maximum of a set of linear support functions.



(a) A set of linear support functions and the function they support

(b) The outer approximation generated by the linear support functions

Figure 2.6: Linear support functions and the outer approximation they generate

Definition 2.22.
Let $\hat{f}_1, \hat{f}_2, ..., \hat{f}_n$ be linear support functions for convex function f. We define an outer approximation, $\hat{f}$, for f as

$$\hat{f}(\mathbf{x}) = \max \left\{ \hat{f}_1(\mathbf{x}), \hat{f}_2(\mathbf{x}), ..., \hat{f}_n(\mathbf{x}) \right\}$$

In fig. 2.6, the linear support functions are generated by taking tangent hyperplanes. This is how we always generate an outer approximation for a convex function. This result is simply theorem 2.7, although we do require $f \in \mathcal{C}^1$ to apply this.

The use of an outer approximation does not necessarily have to use binary variables. If we assume that the structure of the model requires for the result of convex function f to be minimised. We can model f as a series of $\geq$ constraints allowing for the result to be in the epigraph of the outer approximation but since we are trying to minimise, the result will always be on the boundary of the epigraph i.e. on one of the hyperplanes.

2.2.9 Bilinearities

As mentioned in section 2.2.2, a non-convex continuous problem can make finding the global solution hard. This property extends to MINLPs since if we have an efficient way to try all integer combinations in the mixed integer formulation the resulting non-convex problem can be hard to solve.

There are many types of non-convexities that we can encounter in optimisation, the non-convexity we will encounter is bilinearities.

Definition 2.23.
Let $x_l \leq x \leq x_u$ and $y_l \leq y \leq y_u$. We say that $f : [x_l, x_u] \times [y_l, y_u] \to \mathbb{R}$ is a bilinear

function if

$$f(x, y) = b \cdot x \cdot y$$

for some $b \in \mathbb{R}, b > 0$

For simiplicity we have assumed that $b > 0$ in definition 2.23.

A bilinear term has a simple definition but it can be shown to be neither convex nor concave everywhere. We overcome this non-convexity by introducing the McCormick hull [31, 27] in its place.

Let x and y be as defined in definition 2.23, the McCormick hull is defined by its

- convex envelope (under estimator)

$$b \cdot \max\{\, y_l \cdot x + x_l \cdot y - x_l \cdot y_l, y_u \cdot x + x_u \cdot y - x_u \cdot y_u \,\} \tag{2.8}$$

- concave envelope (over estimator)

$$b \cdot \min\{\, y_u \cdot x + x_l \cdot y - x_l \cdot y_u, y_l \cdot x + x_u \cdot y - x_u \cdot y_l \,\} \tag{2.9}$$

We relax the bilinear terms by allowing the result to be a member of a set of points, we introduce the variable z to replace the bilinear occurrences. The McCormick relaxation is added to an optimisation model by adding the following constraints

$$z \geq b\,(y_l \cdot x + x_l \cdot y - x_l \cdot y_l) \tag{2.10}$$
$$z \geq b\,(y_u \cdot x + x_u \cdot y - x_u \cdot y_u) \tag{2.11}$$
$$z \leq b\,(y_u \cdot x + x_l \cdot y - x_l \cdot y_u) \tag{2.12}$$
$$z \leq b\,(y_l \cdot x + x_u \cdot y - x_u \cdot y_l) \tag{2.13}$$

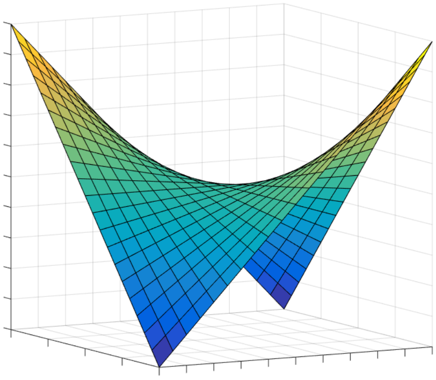
all occurrences of $b \cdot x \cdot y$ are then replaced by z.

This formulation is indeed a relaxation as it includes the values that $b \cdot x \cdot y$ can take. We have that the boundary set by eq. (2.8) is represented by eqs. (2.10) and (2.11) the correctness can be easily reasoned, if we assume that the right hand side of eq. (2.10) is larger than the right hand side of eq. (2.11) then we have that eq. (2.8) is equal to the right hand side of eq. (2.10) and z has to be larger than the right hand side of eq. (2.10) showing that z is bounded by the McCormick hull. Reasoning the other cases is similar. The McCormick envelopes and hull are shown by fig. 2.7.

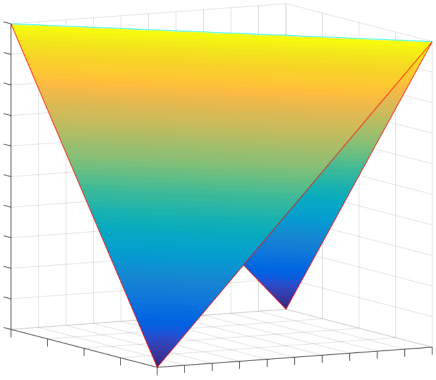
Use of the McCormick hull introduces an error, Androulakis, Maranas, and Floudas showed that this error is bounded with upper bound equal to

$$|b| \cdot \frac{(x_u - x_l)\,(y_u - y_l)}{4}$$

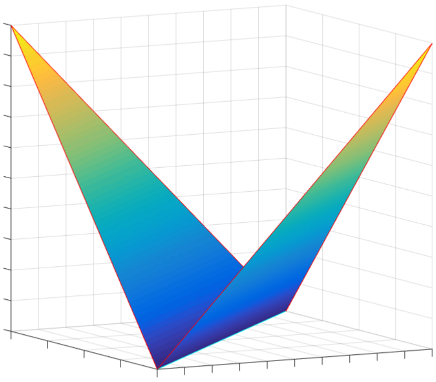
the maximal error occurs when x and y are equal to their midpoints.



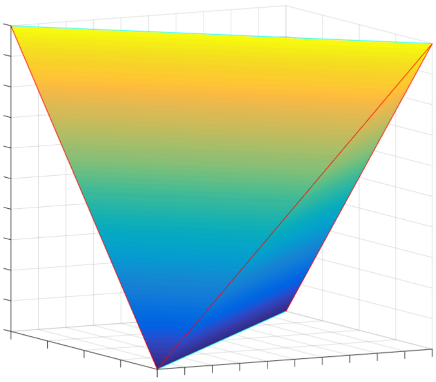
(a) A bilinear function where the domain is square over the same intervals



(b) The McCormick concave envelope of fig. 2.7a



(c) The McCormick convex envelope of fig. 2.7a



(d) The McCormick hull of fig. 2.7a

Figure 2.7: A bilinear function, its McCormick over and under estimators and its McCormick hull. All of these figures are viewed from the same orientation.

2.2.10 Cutting Plane Methods

A cutting plane method is the process of converging to the solution of a MIP by adding linear constraints that tighten the feasible set by 'cutting' part of it off while retaining all integral solutions. The process was introduced by Gomory to be used in the context of integer programs but has since evolved for use with convex MINLPs as discussed by Belotti et al.

We will adapt the cutting plane method to solve a non-convex MINLP. By making suitable relaxations, a non-convex MINLP can be estimated by a convex MINLP if it can be shown that the addition of a cut doesn't remove any feasible solutions we can add the cutting contraint without the need to add any additional binary variables (which would increase the size of the search tree).

2.3 Heat Exchanger Networks

2.3.1 Heat Exchangers

Here we simply state the formulas we use. For derivations and further reading we recommend *Fundamentals of Heat and Mass Transfer*.

Wherever there exists a temperature difference between two mediums, heat transfer must occur. This simple property sits at the heart of heat exchangers.

Definition 2.24.
A heat exchanger is a device that transfers heat from one liquid to another without allowing them to mix.

Heat exchangers have a number of applications including: household fridges and the cooling of liquids in industrial plants. The design of heat exchangers varies however for the purpose of this investigation the types of heat exchangers considered are those that are used in industry. A typical heat exchanger in industry would be able to handle high temperatures e.g. $1034°C$ in case study 04 run by Escobar and Trierweiler.

2.3.2 Calculating Heat Transfer

When working with heat exchangers, the following formula is used to calculate the total heat transfer rate, q.

$$q = U \cdot A \cdot \Delta T_{LMTD} \tag{2.14}$$

In the above:

- q is the total heat transfer rate [kW]
- U is the overall heat transfer coefficient [kW/m^2K]
- A is the surface area over which the heat exchange takes place [m^2]
- ΔT_{LMTD} is the log mean temperature difference [$°C$ or $°K$]

2.3.3 Log Mean Temperature Difference

In a heat exchanger, the change in temperature of a given stream does not generally follow a linear trend. This property is characterised by ΔT_{LMTD} [26].

$$\begin{aligned}\Delta T_{LMTD} &= \frac{\Delta T_1 - \Delta T_2}{\ln(\Delta T_1) - \ln(\Delta T_2)} \\ &= \frac{\Delta T_1 - \Delta T_2}{\ln\left(\frac{\Delta T_1}{\Delta T_2}\right)}\end{aligned}$$

2.3.4 Heat Exchanger Networks

A heat exchanger network (HEN) is a collection of heat exchangers that operate over a number of hot and cold streams to cool and heat the streams respectively.

For industrial applications, the scenario of a single hot stream and single cold stream does not occur. Generally there are several hot and cold streams therefore networks of heat exchangers are used to make use of the residual heat that would be otherwise wasted. For example an industrial HEN can have 22 hot streams and 17 cold streams as shown in case study 05 run by Escobar and Trierweiler.

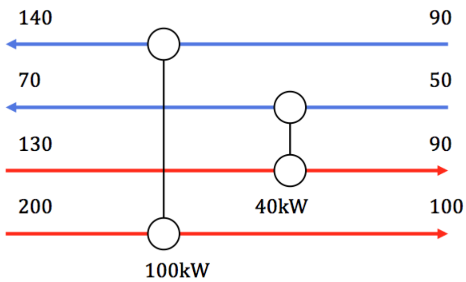


Figure 2.8: An example of how a heat exchanger network with two heat exchangers may look

2.4 Modelling a Heat Exchanger Network

The model formulation being used is formulated in appendix B and is identical to the model proposed by Yee and Grossmann. This formulation models the network in stages. At each stage, a potential match between a hot stream and cold stream is sought. The model includes hot and cold utilities which are used if no further matches between streams can be found and a stream still requires heating/cooling. The model is a non-convex MINLP.

2.4.1 Model Properties

The only integer variables we encouter in this model are binary variables, each of which corresponds to the existence of a heat exchanger.

The heat exchanger model has the following nonlinearities:

- bilinear constraints to model energy balances
- bilinear constraints to calculate the area of a heat exchanger
- log mean temperature difference calculations
- concave functions in the objective function

The model is non-convex because of the bilinear constraints.

The log mean temperature also has a set of indeterminacies, these are found when $\Delta T_1 = \Delta T_2$, direct evaluation gives

$$\frac{\Delta T_1 - \Delta T_1}{\ln(\Delta T_1) - \ln(\Delta T_1)} = \frac{0}{0}$$

this can cause numerical issues when given to a solver, making the task of finding a solution even harder.

The area of a heat exchanger is calculated as in eq. (2.14) but with A as the subject. Therefore the constraints we have take the form

$$A = q \cdot U^{-1} \cdot (\Delta T_{LMTD})^{-1}$$

The objective function, TAC, is the total annual cost of the running the heat exchanger network therefore it is a minimisation problem. The variables that contribute to TAC are

- q_{cui}, q_{huj} - the heat loads for the cold and hot utilities respectively. This represents the amount of external energy we require in order to get the streams to their respective outlet temperatures
- $z_{ijk}, z_{cui}, z_{huj}$ - the binary variables representing the existence of heat exchanger, a heat exchanger has an associated fixed cost
- $A_{ijk}, A_{cui}, A_{huj}$ - the area of the heat exchangers, these are all raised to β where $\beta \in (0, 1]$. These are the functions that give the objective its concave shape

2.4.2 Related Work

Heat exchanger network synthesis (HENS) has been studied for many years with the first mentions of the term occurring in 1944 [3]. Furman and Sahinidis reviewed the contributions over the 20$^{\text{th}}$ century categorising solution methods to:

1. Sequential Synthesis - These methods divide the problem into subproblems to reduce the computational requirements. The problems are solved in order of decreasing significant rules based on heuristic rules. Generally achieved by satisfying: the minimum utility usage, the minimum number of exchangers and the minimum cost of the network.
2. Simultaneous Synthesis - The idea here is to find the optimal network without the need to split the problem. These methods are primarily in the MINLP domain and it is a simultaneous synthesis model that we use.

Escobar and Trierweiler compared sequential synthesis and simultaneous synthesis approaches to HEN synthesis, the models used were

- LP Transshipment Model, MILP Transshipment Model [34], NLP Superstructure Model [15] - a sequential synthesis method designed to first minimise the utility cost (LP transshipment model) followed by minimising the the number of matches (MILP transshipment model) and then minimising the investment cost (NLP superstructure model)

- MINLP Hyperstructure Model [7] - a simultaneous synthesis model built from the transshipment and superstructure models mentioned above
- MINLP SYNHEAT Model [40] - a simultaneous synthesis model (the model we use) that simplifies a HEN design by modelling the exchanger in stages

Of the 5 case studies chosen, the simultaneous synthesis approaches were collectively better than the sequential synthesis approach.

Our approach to solving the model will be associated with the log mean temperature difference. In the above approaches, to avoid numerical difficulties, the Chen approximation to the log mean temperature difference was used.

The Chen approximation [5] to the log mean temperature difference is given by

$$\Delta T_{CHEN}(x, y) = \left(x \cdot y \cdot \frac{x+y}{2} \right)^{\frac{1}{3}}$$

the Chen approximation can be shown to correctly approximate LMTD when $x = y$ (analysis of LMTD at $x = y$ shows that the limit does exist).

It is quite common to see the Chen approximation used in place of LMTD as

- there are no indeterminacies over the domain used
- the error is small around $x = y$ where solutions do tend to settle
- the relative error can be shown to be small [5] for common values of x and y that appear in HEN synthesis
- while there are other approximations such as the Paterson approximation [35] the Chen approximation has an element of simplicity while still being able to accurately estimate LMTD

Huang, Al-mutairi, and Karimi proposed two alternative approaches to avoid the numerical difficulties that come with LMTD, these are

1. rewrite the LMTD constraint as

$$dt_{ijk} = dt_{ijk+1} + LMTD_{ijk} \cdot \ln \left(\frac{dt_{ijk}}{dt_{ijk+1}} \right) \tag{2.15}$$

$$LMTD_{ijk} \leq \frac{1}{2} \left(dt_{ijk} + dt_{ijk+1} \right) \tag{2.16}$$

2. rewrite the LMTD constraint as

$$LMTD_{ijk} = \left(dt_{ijk} - dt_{ijk+1} + \varepsilon \right) / \ln \left(\frac{dt_{ijk} + \varepsilon}{dt_{ijk+1}} \right) \tag{2.17}$$

The first approach ensures that we don't encounter numerical errors through the reformulation of the constraint, eq. (2.15), and enforcing bounds, eq. (2.16), ensures that the value of $LMTD_{ijk}$ is correct when $dt_{ijk} = dt_{ijk+1}$. These reformulations do increase the number of constraints which could otherwise be moved into the objective.

The second approach gives $LMTD_{ijk} = dt_{ijk}$ when $dt_{ijk} = dt_{ijk+1}$ for small ε although errors are introduced into the model.

2.5 Optimisation Tools

2.5.1 Solvers

For this investigation we will be using the Gurobi MILP solver [21]. Gurobi is a leading MILP solver used by many companies.

There are other solvers available such as

- CPLEX [25]
- XPRESS [39]
- CBC [9]

The choice of Gurobi was made due to its robustness and speed which is shown by the benchmarking tests [28] run by Mittelmann, from these results Gurobi can be shown to be among the top 3 MILP solvers.

2.5.2 Algebraic Modelling

Mathematical models are presented in an algebraic form so that they can be analysed generally. This abstraction is applied to modelling software where the concrete instances of a model specified by the parameters are separated from the abstract model.

For this investigation Pyomo [22] was used to create the algebraic model. Pyomo is an open source set of optimisation tools and libraries built on the programming language Python. Hart, Watson, and Woodruff discuss the design and implementation choices of Pyomo, these form the basis for choosing Pyomo over similar alternatives (e.g. GAMS or AMPL) for this investigation including

- leveraging a modern programming language - Python has a set of libraries that can be used to easily pipeline solver results and manipulate them as we need such as the `matplotlib` library for plotting results
- documentation - Python is well documented making it easy to learn and to allow for others to easily build upon the model we create.

3 Analysis

3.1 Overview

In this section, we analyse the nonlinearities in the model, with extensive analysis of the reciprocal of the log mean temperature difference (LMTD). The reciprocal of LMTD is analysed here as we identify that, under the structure of the model, it is better to approximate this function opposed to LMTD. A similar analysis for LMTD has been done, this is presented in appendix C. Some of the results presented here can be derived from proofs given by Zavala-Río, Femat, and Santiesteban-Cos, we present an alternative proof. The proofs we present were derived prior to finding the proofs given by Zavala-Río, Femat, and Santiesteban-Cos.

Throughout this section we make use of the auxilary variable w defined as

$$w = \frac{x}{y}$$

3.2 Nonlinearities in the Model

We want to create a MILP approximation to the HEN model. To do this we need to be able to form linear approximations of all of the nonlinearities in the model. The nonlinearities are:

- the heat load bilinearities
- LMTD
- the area bilinearities
- the nonlinear concave functions in the objective

The heat load bilinearities are defined by variables for which the bound is known. These can easily be relaxed to form the McCormick hull.

The nonlinear concave functions take the form

$$A^{\beta}$$

where $0 < \beta \leq 1$.

This function can be modelled using a piecewise linear function however to do so we need to calculate the bounds. The lower bound is known to be 0 therefore we only need to calculate the upper bound, this does involve LMTD. Hence to calculate the upper bound for the area, we need to classify the bounds for LMTD.

The cross sectional area for a heat exchanger is calculated by:

$$A = q \cdot U^{-1} \cdot (\Delta T_{LMTD})^{-1}$$

where

- q is the total heat transfer rate $[\ kW\ ]$

- U is the overall heat transfer coefficient $[\ kW/m^2K\ ]$
- A is the surface area over which the heat exchange takes place $[\ m^2\ ]$
- ΔT_{LMTD} is the log mean temperature difference $[\ ^\circ C\ ]$

but we are not simply modelling a single heat exchanger therefore, in the model, the area constraints take the form

$$\begin{aligned} A_{ijk} &= q_{ijk}\left(h_i^{-1}+h_j^{-1}\right)/LMTD_{ijk}, && i\in HP, j\in CP, k\in ST\\ A_{cui} &= q_{cui}\left(h_i^{-1}+h_{CU}^{-1}\right)/LMTD_{cui}, && i\in HP\\ A_{huj} &= q_{huj}\left(h_j^{-1}+h_{HU}^{-1}\right)/LMTD_{huj}, && j\in CP \end{aligned}$$

where

- i, j and k represent a hot stream, a cold stream and a stage respectively.
- h_i, h_j, h_{CU} and h_{HU} are constants, the sums involving these make U.

As this constraint is bilinear, we will relax it using the McCormick hull however to do so we need to classify the bounds of q and *LMTD*. Since q is calculated from the heat load bilinearities, we will know the bounds on the variable. We need further analysis for the log mean temperature difference.

In the area constraints, the bilinear part is

$$q\cdot\left(\frac{1}{LMTD}\right)$$

since we are going to divide by *LMTD*, the error in our approximation will be cascaded into the model by an additional layer. To minimise the cascades of errors, a better function to approximate would be the reciprocal of LMTD. This will be the function that we analyse and approximate.

The structure of the model states that the area be as small as possible (since the objective function increases as the area increases), this means that the reciprocal of LMTD should be as small as possible. We see from Figure 3.1 that the reciprocal of LMTD has a shape that looks convex (where defined, the epigraph looks convex), we will not only prove that this is true but we will use this to form a suitable approximation. We will also classify the bounds of the function as this will be required to form a McCormick relaxation for the area constraints.

3.3 Reciprocal of LMTD: Function Definition

We begin by defining the reciprocal of LMTD, for simplicity we will use the label m when referring the function.

First we define the following set, the domain of m.

$$S^* = \left\{\ (x,y)^T\ \middle|\ x,y>0,\ x\neq y\ \right\}$$

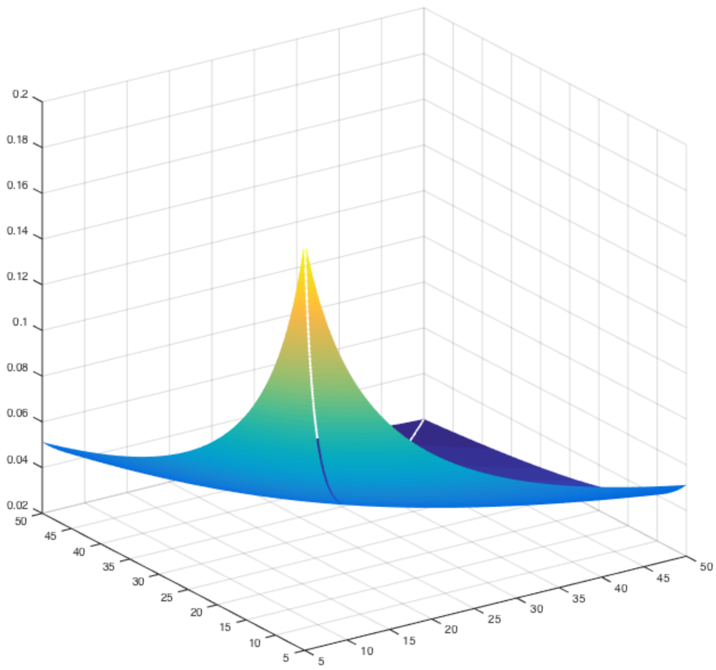


Figure 3.1: The reciprocal of LMTD, m

The reciprocal of LMTD, $m : S^* \to \mathbb{R}$ is defined as

$$m\left(x, y\right) = \frac{\ln\left(\frac{x}{y}\right)}{x - y}$$

We also define the gradient and Hessian of m, the derivations for these functions is presented in appendix G.

The gradient of m, $\nabla m : S^* \to \mathbb{R}^2$ is defined as

$$\nabla m\left(x, y\right) = \frac{1}{\left(x - y\right)^2} \begin{pmatrix} 1 - \ln\left(w\right) - w^{-1} \\ 1 + \ln\left(w\right) - w \end{pmatrix}$$

The Hessian of m, $\nabla^2 m : S^* \to \mathbb{R}^{2\times 2}$ is defined as

$$\nabla^2 m\left(x, y\right) = \frac{1}{\left(x - y\right)^3} \begin{pmatrix} 2\ln\left(w\right) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln\left(w\right) \\ w - w^{-1} - 2\ln\left(w\right) & 2\ln\left(w\right) - 4w + w^2 + 3 \end{pmatrix}$$

3.3.1 Indeterminate Evaluations

We have not defined m, ∇m and $\nabla^2 m$ over

$$S' = \left\{ \left(t, t\right)^T \,\middle|\, t > 0 \right\}$$

this is because when attempt to directly evaluate these functions we get an indeterminate form, e.g. for m we get the following:

Let $(c, c)^T \in S'$

$$m(c, c) = \frac{c - c}{\ln({}^{c}/_{c})} = \frac{0}{\ln(1)} = \frac{0}{0} \tag{3.1}$$

We get a similar result for all of the elements of ∇m and $\nabla^2 m$ when they are evaluated over S'.

Equation (3.1) does not mean that the limit of m over S' does not exist, it means that we require further analysis. It is here that we begin our analysis of m.

3.4 The Limits of the Reciprocal of LMTD

We will prove that the limits of m, ∇m and $\nabla^2 m$ over the set

$$S = \left\{ (x, y)^T \;\middle|\; x, y > 0 \right\}$$

do exist and we will state their values.

This is done by analysing S' as the function and its derivatives are well defined and continuous over S^*.

Figure 3.1 shows the indeterminate set of points. Intuitively, we establish that the limit appears to exist as the defined values of m seem to converge to the same point as we approach $(c, c)^T$, $c > 0$, we will show this rigourously.

Evaluating the limit

$$\lim_{(x,y)^T \to (c,c)^T} m(x, y)$$

is a problem as the definition of a limit states that we must prove the limiting value is the same when approached from any direction. The sequence

$$\left(c + \frac{1}{n}, c + \frac{1}{n}\right)^T, \quad n \in \mathbb{N} \setminus \{ 0 \}$$

gives a valid path on which we cannot evaluate the limit, hence this approach cannot be used.

3.4.1 Polar Coordinates

Polar coordinates allow us to model the function using magnitude and angle as parameters. Representing the function in polar form sees the entire indeterminate set of points from the cartesian system collapse to a single angle in the polar system as shown by fig. 3.2.

We redefine m in terms of polar coordinates, this is done by making the following

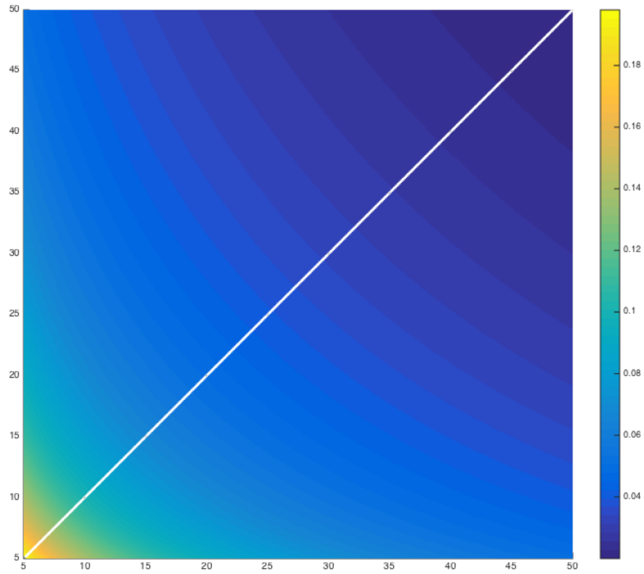


Figure 3.2: m viewed as a heat map in the x, y plane

transformations:

$$x \twoheadrightarrow r\cos(\theta) \tag{3.2}$$
$$y \twoheadrightarrow r\sin(\theta) \tag{3.3}$$
$$w = \frac{x}{y} \twoheadrightarrow \frac{r\cos(\theta)}{r\sin(\theta)} = \frac{\cos(\theta)}{\sin(\theta)} = \cot(\theta) \tag{3.4}$$

where

$$r = \sqrt{x^2 + y^2}$$
$$\theta = \arctan\left(\frac{y}{x}\right)$$

Under these transformations the analogues of S, S^* and S' in the polar coordinate system are (respectively)

$$S_{r,\theta} = \left\{ (r,\theta)^T \,\middle|\, r > 0, 0 < \theta < \frac{\pi}{2} \right\}$$
$$S^*_{r,\theta} = \left\{ (r,\theta)^T \,\middle|\, r > 0, 0 < \theta < \frac{\pi}{2}, \theta \neq \frac{\pi}{4} \right\}$$
$$S'_{r,\theta} = \left\{ (r,\theta)^T \,\middle|\, r > 0, \theta = \frac{\pi}{4} \right\}$$

The limit we will show (via the polar coordinate system) is:

$$(x, y)^T \to (c, c)^T$$

this equates to

$$\begin{aligned}\theta &\to \arctan\left(\frac{c}{c}\right) = \arctan(1) = \frac{\pi}{4} \\ r &\to \sqrt{c^2 + c^2} = c\sqrt{2} = r_c\end{aligned}$$

where $c > 0$
Hence the polar limit is

$$(r, \theta)^T \to \left(r_c, \frac{\pi}{4}\right)^T$$

3.4.2 Evaluation of the Limit

We will now evaluate the limit this will be done by making use of theorems 2.1 and 2.3.

Polar Representation of m
We derive the polar representation of m by making the transformations defined by eqs. (3.2) to (3.4)

$$\begin{aligned}m^{[r,\theta]}(r, \theta) &= m(r\cos(\theta), r\sin(\theta)) \\ &= \frac{\ln(\cot(\theta))}{r\cos(\theta) - r\sin(\theta)} \\ &= r^{-1} \cdot \frac{\ln(\cot(\theta))}{\cos(\theta) - \sin(\theta)} \\ &= p(r) \cdot \frac{q_1(\theta)}{q_2(\theta)} \qquad (3.5)\end{aligned}$$

$p(r)$ is clearly well defined for all $r > 0$.
The function $\frac{q_1(\theta)}{q_2(\theta)}$ is well defined except possibly for

$$\theta = \frac{\pi}{4}$$

On inspection we find that the θ parametrised parts of m at $\frac{\pi}{4}$ evaluates to a fraction of the form $\frac{0}{0}$, this is where we shall use theorem 2.3.

In eq. (3.5), we see that the resulting polar representations give a set of separable functions parametrised by independent variables, r and θ, this is where we shall use theorem 2.1.

Evaluation of the Derivatives
The derivation of the following derivatives can be found in appendix I.

We begin our analysis with $q_2(\theta)$ as this function indicates how many derivatives we will require from $q_1(\theta)$ if a limit is to exist i.e. we continue to take derivatives until the first non-zero is found.

$$\begin{aligned}
q_2(^{\pi}/_4) &= \cos(^{\pi}/_4) - \sin(^{\pi}/_4) \\
&= \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \\
&= 0 \\
q_2'(^{\pi}/_4) &= -\sin(^{\pi}/_4) - \cos(^{\pi}/_4) \\
&= -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \\
&= -\sqrt{2}
\end{aligned}$$

Here we see that the only necessary requirement on q_1 is

$$q_1(^{\pi}/_4) = 0$$

On evaluation we see that this property holds.

$$\begin{aligned}
q_1(^{\pi}/_4) &= \ln\left(\cot(^{\pi}/_4)\right) \\
&= \ln(1) \\
&= 0 \\
q_1'(^{\pi}/_4) &= -\csc(^{\pi}/_4)\sec(^{\pi}/_4) \\
&= -\sqrt{2}\cdot\sqrt{2} \\
&= -2
\end{aligned}$$

The Limit

By theorem 2.3 we get the following results:

$$\frac{q_1(^{\pi}/_4)}{q_2(^{\pi}/_4)} = \frac{q_1'(^{\pi}/_4)}{q_2'(^{\pi}/_4)} = \frac{-2}{-\sqrt{2}} = \sqrt{2}$$

Hence the limit of $\frac{q_1(\theta)}{q_2(\theta)}$ is well defined for all $\theta \in (0, \frac{\pi}{2})$.

We can now evaluate the limit of m using theorem 2.1.

$$\begin{aligned}
\lim_{(r,\theta)\to(r_c,^{\pi}/_4)} p(r)\cdot\frac{q_1(\theta)}{q_2(\theta)} &= \left[\lim_{r\to r_c} p(r)\right]\left[\lim_{\theta\to^{\pi}/_4}\frac{q_1(\theta)}{q_2(\theta)}\right] \\
&= r_c^{-1}\cdot\sqrt{2} \\
&= \frac{1}{c\sqrt{2}}\cdot\sqrt{2} \\
&= \frac{1}{c}
\end{aligned}$$

But c was simply the value of x (and y) hence we get the following result

$$\forall c > 0, \qquad \lim_{(x,y)^T \to (c,c)^T} m(x,y) = \frac{1}{c}$$

Let

$$S = \left\{ (x,y)^T \;\middle|\; x, y > 0 \right\}$$

The reciprocal of the log mean temperature difference, $m : S \to \mathbb{R}$, is defined as

$$m\left(x,y\right) = \begin{cases} {}^{1}\!/_{x}, & x = y \\ \dfrac{\ln\left({}^{x}\!/_{y}\right)}{x-y}, & x \neq y \end{cases}$$

Using a similar process, we can evaluate the limits of $\nabla m(x,y)$ and $\nabla^2 m(x,y)$, this is presented in appendix E.

3.5 Well Defined Formulation of the Reciprocal of LMTD

Having calculated the limits of m, ∇m and $\nabla^2 m$ at any indeterminate points and showing that they exist we can define the well defined formulations for these functions.

Definition 3.1.
Let

$$S = \left\{ (x,y)^T \;\middle|\; x, y > 0 \right\}$$

The reciprocal of the log mean temperature difference

$$m : S \to \mathbb{R}$$

its gradient

$$\nabla m : S \to \mathbb{R}^2$$

and its Hessian

$$\nabla^2 m : S \to \mathbb{R}^{2\times 2}$$

are defined as

$$m\left(x,y\right) = \begin{cases} {}^{1}\!/_{x}, & x = y \\ \dfrac{\ln\left({}^{x}\!/_{y}\right)}{x-y}, & x \neq y \end{cases}$$

$$\nabla m\left(x,y\right) = \begin{cases} \begin{pmatrix} {}^{-1}\!/_{2x^2} \\ {}^{-1}\!/_{2x^2} \end{pmatrix}, & x = y \\ \dfrac{1}{\left(x-y\right)^2} \begin{pmatrix} 1 - \ln\left(w\right) - w^{-1} \\ 1 + \ln\left(w\right) - w \end{pmatrix}, & x \neq y \end{cases}$$

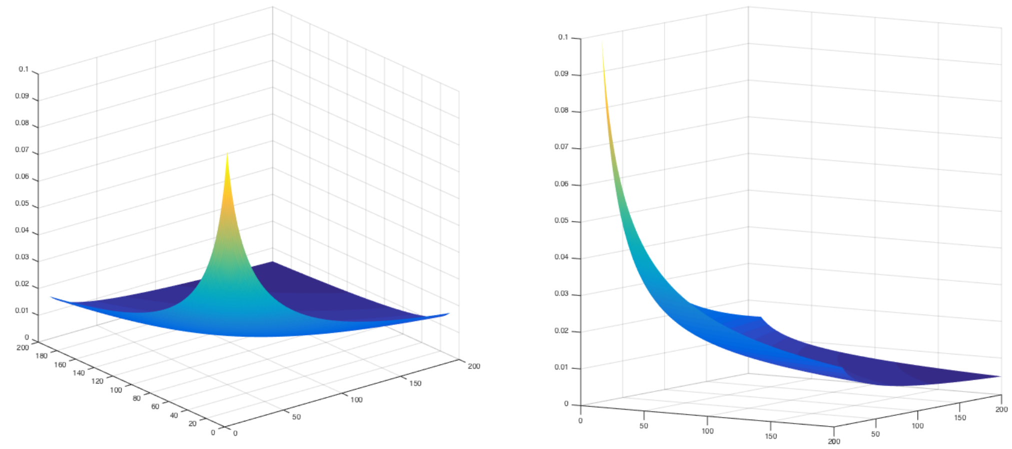


Figure 3.3: Plots of the well defined m

$$\nabla^2 m(x,y) = \begin{cases} \dfrac{1}{3x^3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, & x = y \\ \dfrac{1}{(x-y)^3}\begin{pmatrix} 2\ln(w) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln(w) \\ w - w^{-1} - 2\ln(w) & 2\ln(w) - 4w + w^2 + 3 \end{pmatrix}, & x \neq y \end{cases}$$

The above definition has been shown to conform to the limits of the respective: function definition, derived gradient and derived Hessian of m i.e. the function evaluates to its limit for all points of its domain, we therefore conclude that

$$m \in \mathcal{C}^2$$

that is: m belongs to the set of twice continuously differentiable functions.

A similar result holds for LMTD and its derivatives, presented in appendix C.

3.6 Properties of the Reciprocal of LMTD

Having a continuously differentiable definition of m over a convex domain, we now classify some properties of the function. The continuously differentiable result is important as it allows us to justify the use of certain methods when it comes to linearisation.

For the following properties of m that we prove for, we can prove similar results for LMTD, presented in appendix C.

From fig. 3.3, we can identify properties relating to: symmetry, bounds and shape.

3.6.1 Symmetry

We can see that the function is symmetric about the line $y = x$. A symmetric property about its derivative can also be shown.

Lemma 3.1.
For all $(x,y)^T \in S$

$$m(y,x) = m(x,y)$$
$$\nabla m(y,x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \nabla m(x,y)$$

Proof.
The $x = y$ cases are trivial. We only need to show the lemma for $x \neq y$

For all $(x,y)^T \in S$

$$\begin{aligned} m(y,x) &= \frac{\ln({}^{y}/_{x})}{y-x} \\ &= \frac{-\ln({}^{x}/_{y})}{-(x-y)} \\ &= \frac{\ln({}^{x}/_{y})}{x-y} \\ &= m(x,y) \end{aligned}$$

$$\begin{aligned} \nabla m(y,x) &= \frac{1}{(y-x)^2} \begin{pmatrix} 1 - \ln\left(w^{-1}\right) - \left(w^{-1}\right)^{-1} \\ 1 + \ln\left(w^{-1}\right) - w^{-1} \end{pmatrix} \\ &= \frac{1}{\left(-\left(x-y\right)\right)^2} \begin{pmatrix} 1 + \ln\left(w\right) - w \\ 1 - \ln\left(w\right) - w^{-1} \end{pmatrix} \\ &= \frac{1}{(x-y)^2} \begin{pmatrix} 1 + \ln\left(w\right) - w \\ 1 - \ln\left(w\right) - w^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \left(\frac{1}{(x-y)^2} \begin{pmatrix} 1 - \ln\left(w\right) - w^{-1} \\ 1 + \ln\left(w\right) - w \end{pmatrix} \right) \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \nabla m(x,y) \end{aligned}$$

□

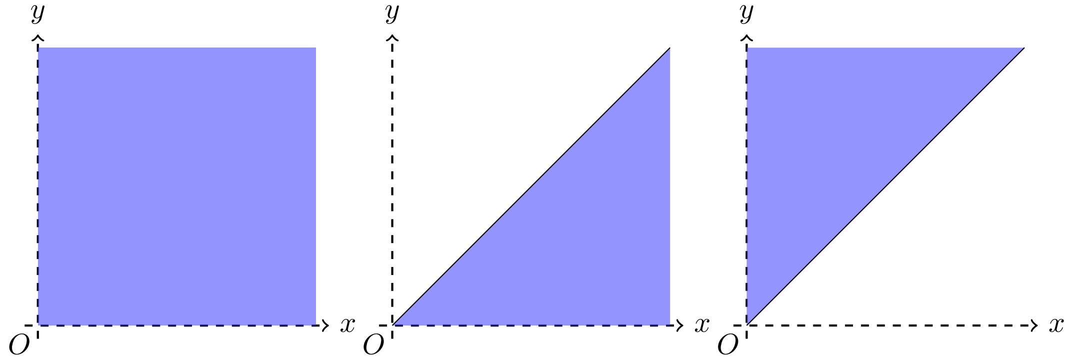
The symmetric property is useful when establishing some of the other properties as when we prove results over the set

$$S = \left\{ (x,y)^T \;\middle|\; x,y > 0 \right\}$$

the symmetric nature of the function sometimes allows us to reduce our proof to either of the sets:

$$S^{\mathrm{L}} = \left\{ (x,y)^T \;\middle|\; 0 < y \leq x \right\}$$
$$S^{\mathrm{U}} = \left\{ (x,y)^T \;\middle|\; 0 < x \leq y \right\}$$

the reduced sets split the original set in half as shown in fig. 3.4.



(a) The set S (b) The set S^L (c) The set S^U

Figure 3.4: The sets over which we apply our analysis to m

3.6.2 Bounds

We classify the bounds of m as we will require these to form the McCormick relaxation when formulating the area constraints. From fig. 3.3, we see a decreasing trend and what appears to be an asymptotic approach to towards 0 as x and y increase.

Proposition 3.1.

$$\forall (x,y)^T \in S, x \neq y \; \exists \, t > 0 \quad s.t \quad m(t,t) < m(x,y)$$

Proof.
By lemma 3.1 we only need to show the result over S^L.

Let $(x,y)^T \in S^L$ with $x \neq y$. We will look at the directional derivative along the direction $\mathbf{d} = (-1,1)^T$.

The directional derivative is

$$\begin{aligned}\mathbf{d}^T \nabla m(x,y) &= \begin{pmatrix} -1 \\ 1 \end{pmatrix} \cdot \left(\frac{1}{(x-y)^2} \begin{pmatrix} 1 - \ln(w) - w^{-1} \\ 1 + \ln(w) - w \end{pmatrix} \right) \\ &= \frac{1}{(x-y)^2} \left(2\ln(w) + w^{-1} - w \right)\end{aligned}$$

We have that

$$\forall (x,y)^T \in S^L,\, x \neq y, \quad \frac{1}{(x-y)^2} > 0$$

hence it only acts as a scaling factor.

Let

$$p_1(w) = 2\ln(w) + w^{-1} - w$$

Since $x > y$, $w > 1$, therefore we have $w = e^n$ where $n = \ln(w) > 0$, this gives

$$\begin{aligned} p_2(n) = p_1(e^n) &= 2n + e^{-n} - e^n \\ &= 2n - 2\left(\frac{e^n - e^{-n}}{2}\right) \\ &= 2n - 2\sinh(n) \end{aligned}$$

Taking derivatives, we get

$$\begin{aligned} p_2'(n) &= 2 - 2\cosh(n) \\ p_2''(n) &= -2\sinh(n) \end{aligned}$$

This means

$$\forall n > 0,\ p_2''(n) < 0$$

i.e. p_2 has a strict concave structure for all $n > 0$.
We also have

$$\begin{aligned} p_2(0) &= 0 \\ p_2'(0) &= 0 \end{aligned}$$

Hence we find

$$\forall n > 0,\ p_2(n) < 0 \tag{3.6}$$

Equation (3.6) shows us that m is strictly decreasing along the direction $\mathbf{d}$, for all $(x, y)^T \in S^L$, $x \neq y$.

Moving along $\mathbf{d}$ decreases the value of x and increases the value of y and we have that $x > y$, this implies

$$\exists\, \alpha, t > 0 \quad \text{s.t.} \quad (x, y)^T + \alpha\mathbf{d} = (t, t)^T \tag{3.7}$$

By continuous differentiability of m and eq. (3.7) we conclude that

$$\forall (x, y)^T \in S^L, x \neq y\ \exists t > 0 \quad \text{s.t} \quad m(t, t) < m(x, y)$$

By lemma 3.1, the result follows.

□

Corollary 3.1.

$$\forall (x, y)^T \in S, \quad m(x, y) > 0$$

Proof.
We have

$$\forall t > 0, \quad m(t, t) = \frac{1}{t} > 0$$

By proposition 3.1, we have that

$$\forall (x,y)^T \in S \ \exists t > 0 \quad \text{s.t} \quad m(x,y) \geq m(t,t) > 0$$

□

Proposition 3.2.
Let $(x,y)^T \in S$.

- *If we fix y and let x vary then $m(x,y)$ is strictly decreasing*
- *If we fix x and let y vary then $m(x,y)$ is strictly decreasing*

Proof.
We only need to show the first condition as the second condition follows by lemma 3.1 (The gradient result).

If we fix $y = c$ ($c > 0$), then the m becomes a univariate function of x, the derivative for which is given by the first element of ∇m.

We see that the derivative is stictly decreasing when $x = c$

$$\forall c > 0, \quad -\frac{1}{2c^2} < 0$$

therefore we only need to show that the derivative is negative when $x \neq c$ as we know that this function is continuous.

The derivative for $x \neq c$ is

$$p_1(x) = \frac{1}{(x-c)^2}\left(1 - \ln(w_c) - w_c^{-1}\right)$$

where $w_c = \frac{x}{c}$.

Since $c > 0$ and $x > 0$, we have $w_c = \frac{x}{c} > 0$. We also have that $x \neq c$ therefore $\frac{1}{(x-c)^2} > 0$ so we require

$$1 - \ln(w_c) - w_c^{-1} < 0,$$

We can make the following substituition

$$w_c = e^n$$

where $n = \ln(w_c) \neq 0$

Let

$$\begin{aligned} p_2(n) &= 1 - \ln(e^n) - (e^n)^{-1} \\ &= 1 - n - e^{-n} \end{aligned}$$

We will show that $p_2(n) < 0$ for all $n \neq 0$.

Taking derivatives, we get

$$p_2'(n) = e^{-n} - 1$$

$$p_2''(n) = -e^{-n}$$

from this we conclude

$$\forall n \in \mathbb{R}, \quad p_2''(n) < 0$$

This the same as saying that p_2 is strictly concave, therefore if a maximum exists it is unique. If there is a maximum, p_2' will evaluate to 0 at this point.

At $n = 0$ we get the following

$$p_2'(0) = 0$$
$$p_2(0) = 0$$

Hence $p_2(0) = 0$ is a unique maximum which gives

$$p_2(0) < 0, \quad \forall n \neq 0$$

proving the first condition.

The second condition follows from lemma 3.1

□

Proposition 3.3.
Let $l, u \in \mathbb{R}$ such that $0 < l < u$

Over the set

$$S_{[l,u]} = \left\{ (x,y)^T \;\middle|\; l \leq x, y \leq u \right\}$$

the unique bounds of m are given by

$$\min_{(x,y)^T \in S_{[l,u]}} m(x,y) = m(u,u) = \frac{1}{u}$$
$$\max_{(x,y)^T \in S_{[l,u]}} m(x,y) = m(l,l) = \frac{1}{l}$$

Proof.
We show these results using proposition 3.2.

First we define the following set generator defined over $S_{[l,u]}$

$$S(x,y) = \left\{ (s,t)^T \in S_{[l,u]} \;\middle|\; s \geq x,\ t \geq y \right\}$$

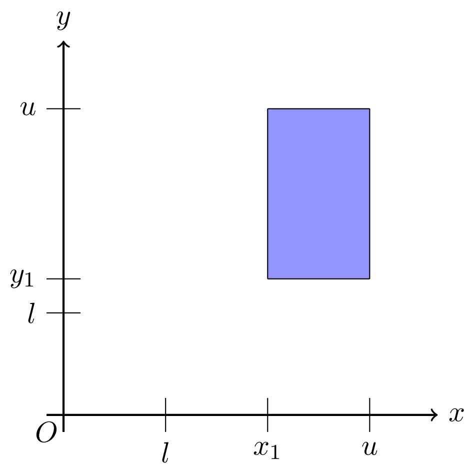


Figure 3.5: The set $S(x_1, y_1)$ w.r.t. $S_{[l,u]}$

The set $S(x, y)$ simply gives us the points in $S_{[l,u]}$ whose first and second coordinate are greater than or equal to x and y respectively as shown by fig. 3.5.

By proposition 3.2, for a given $(x, y)^T \in S_{[l,u]}$, we have

$$\forall (s,t)^T \in S(x,y), \quad m(x,y) \geq m(s,t) \tag{3.8}$$

We also have that $(l, l)^T$ is the unique coordinate $(x, y)^T$ for which the following holds

$$S(x,y) = S_{[l,u]}$$

By eq. (3.8), we get

$$\forall (s,t)^T \in S_{[l,u]}, \quad m(l,l) \geq m(s,t)$$

Hence the unique maximum is given by

$$\max_{(x,y)^T \in S_{[l,u]}} m(x,y) = m(l,l) = \frac{1}{l}$$

For the minimal value, we look at the following set

$$S_{\text{low}} = \bigcap_{(x,y)^T \in S_{[l,u]}} S(x,y)$$

By proposition 3.2, the following holds

$$\forall (s,t)^T \in S_{[l,u]}, (x,y)^T \in S_{\text{low}} : \quad m(s,t) \geq m(x,y)$$

It is also clear that

$$S_{\text{low}} = \left\{ \, (u,u)^T \, \right\}$$

Hence we get that the unique minimum is given by

$$\min_{(x,y)^T \in S_{[l,u]}} m(x,y) = m(u,u) = \frac{1}{u}$$

□

3.6.3 Strict Convexity

One of the most noticable features of m is its shape, we can see that the function possesses a convex structure. We will show that m is strictly convex over its domain.

Proposition 3.4.
The reciprocal of the log mean temperature difference

$$m(x,y) = \frac{\ln\left(\frac{x}{y}\right)}{x-y}$$

is strictly convex over the sets

$$S_1 = \left\{ (x,y)^T \;\middle|\; 0 < y < x \right\}$$
$$S_2 = \left\{ (x,y)^T \;\middle|\; 0 < x < y \right\}$$

Proof.
We use the substitution $w = \frac{x}{y}$, this means that S_1 and S_2 can be defined as

$$S_1 = \{ w \mid w > 1 \}$$
$$S_2 = \{ w \mid 0 < w < 1 \}$$

The Hessian of m is

$$\nabla^2 m(x,y) = \frac{1}{(x-y)^3} \begin{pmatrix} 2\ln(w) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln(w) \\ w - w^{-1} - 2\ln(w) & 2\ln(w) - 4w + w^2 + 3 \end{pmatrix}$$

For strict convexity, we require for $\nabla^2 m(x,y)$ to be positive definite.
Since $w \neq 1$ $(x \neq y)$, $\nabla^2 m(x,y)$ is well defined.

We have that

$$\frac{1}{(x-y)^3} > 0, \qquad x > y \tag{3.9}$$
$$\frac{1}{(x-y)^3} < 0, \qquad x < y \tag{3.10}$$

Equations (3.9) and (3.10) correspond to the case of $w \in S_1$ and $w \in S_2$ respectively therefore for $\nabla^2 m(x,y)$ to be positive definite, we require that the matrix

$$\mathbf{A} = \begin{pmatrix} 2\ln(w) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln(w) \\ w - w^{-1} - 2\ln(w) & 2\ln(w) - 4w + w^2 + 3 \end{pmatrix}$$

to be positive definite for $w \in S_1$ and negative definite for $w \in S_2$.

We show these conditions using theorem 2.4, this means that we require for $\mathbf{A}$ to have the following properties:

$$D_1 > 0, \qquad w \in S_1 \tag{3.11}$$
$$D_1 < 0, \qquad w \in S_2 \tag{3.12}$$
$$D_2 > 0, \qquad w \in S_1 \tag{3.13}$$
$$D_2 > 0, \qquad w \in S_2 \tag{3.14}$$

Since $w > 0$ we can use the following facts:

$$\exists\, n > 0 \text{ s.t. } w = e^n, \qquad w \in S_1$$
$$\exists\, n < 0 \text{ s.t. } w = e^n, \qquad w \in S_2$$

in both cases, $n = \ln(w)$.

We will begin by showing eqs. (3.11) and (3.12).

$$D_1 = d_1\left(w\right) = 2\ln\left(w\right) + 4w^{-1} - w^{-2} - 3$$

Applying the substitution $w = e^n$, we get

$$\begin{aligned} p_1(n) = d_1\left(e^n\right) &= 2\ln\left(e^n\right) + 4e^{-n} - e^{-2n} - 3 \\ &= 2n + 4e^{-n} - e^{-2n} - 3 \end{aligned}$$

Taking the derivatives of $p_1(n)$ we get:

$$\begin{aligned} p_1'(n) &= 2 - 4e^{-n} + 2e^{-2n} \\ p_1''(n) &= 4e^{-n} - 4e^{-2n} \\ &= 4e^{-n}\left(1 - e^{-n}\right) \end{aligned}$$

We see that

$$p_1''(n) > 0, \qquad n > 0$$
$$p_1''(n) < 0, \qquad n < 0$$

also note that $p_1'(0) = 0$ and $p_1(0) = 0$ hence by proposition 2.1 we find that there is a point of inflection at $n = 0$ giving

$$p_1(n) > 0, \qquad n > 0$$
$$p_1(n) < 0, \qquad n < 0$$

i.e.

$$D_1 = d_1\left(w\right) > 0, \qquad w \in S_1$$
$$D_1 = d_1\left(w\right) < 0, \qquad w \in S_2$$

Remark: Applying proposition 2.1 to $-p_1(n)$ for $n < 0$ gives the required condition

We now show eqs. (3.13) and (3.14).

Let

$$\mathbf{A} = \begin{pmatrix} a(w) & b(w) \\ b(w) & c(w) \end{pmatrix}$$

where

$$\begin{aligned} a(w) &= 2\ln(w) + 4w^{-1} - w^{-2} - 3 \\ b(w) &= w - w^{-1} - 2\ln(w) \\ c(w) &= 2\ln(w) - 4w + w^2 + 3 \end{aligned}$$

We have

$$\begin{aligned} D_2 = d_2(w) &= \det(\mathbf{A}) \\ &= a(w)\, c(w) - b(w)^2 \end{aligned}$$

The first product is

$$\begin{aligned} a(w)\, c(w) = &\ \Big(2\ln(w) + 4w^{-1} - w^{-2} - 3\Big)\Big(2\ln(w) - 4w + w^2 + 3\Big) \\ = &\ 4\ln(w)^2 - 8w\ln(w) + 2w^2\ln(w) + 6\ln(w) \\ &+ 8w^{-1}\ln(w) - 16ww^{-1} + 4w^{-1}w^2 + 12w^{-1} \\ &- 2w^{-2}\ln(w) + 4ww^{-2} - w^{-2}w^2 - 3w^{-2} \\ &- 6\ln(w) + 12w - 3w^2 - 9 \\ = &\ 4\ln(w)^2 - 8w\ln(w) + 8w^{-1}\ln(w) + 2w^2\ln(w) \\ &- 2w^{-2}\ln(w) + 16w + 16w^{-1} - 3w^2 - 3w^{-2} - 26 \end{aligned}$$

The second product is

$$\begin{aligned} b(w)^2 &= \Big(w - w^{-1} - 2\ln(w)\Big)^2 \\ &= w^2 + w^{-2} + 4\ln(w)^2 - 2ww^{-1} - 4w\ln(w) + 4w\ln(w) \\ &= w^2 + w^{-2} + 4\ln(w)^2 - 4w\ln(w) + 4w^{-1}\ln(w) - 2 \end{aligned}$$

Hence we get

$$\begin{aligned} d_2(w) &= a(w)\, c(w) - b(w)^2 \\ &= \ 4\ln(w)^2 - 8w\ln(w) + 8w^{-1}\ln(w) + 2w^2\ln(w) \\ &\quad - 2w^{-2}\ln(w) + 16w + 16w^{-1} - 3w^2 - 3w^{-2} - 26 \\ &\quad - \Big(w^2 + w^{-2} + 4\ln(w)^2 - 4w\ln(w) + 4w^{-1}\ln(w) - 2\Big) \\ &= \ 4w^{-1}\ln(w) - 4w\ln(w) + 2w^2\ln(w) - 2w^{-2}\ln(w) \\ &\quad + 16w + 16w^{-1} - 4w^2 - 4w^{-2} - 24 \end{aligned}$$

Once again we shall use the substitution

$$w = e^n \quad \text{where} \quad n = \ln(w)$$

This results in

$$\begin{aligned} p_2(n) = d_2\left(e^n\right) = &\ 4e^{-n}\ln\left(e^n\right) - 4e^n\ln\left(e^n\right) + 2e^{2n}\ln\left(e^n\right) - 2e^{-2n}\ln\left(e^n\right) \\ &+ 16e^n + 16e^{-n} - 4e^{2n} - 4e^{-2n} - 24 \\ = &\ 4ne^{-n} - 4ne^n + 2ne^{2n} - 2ne^{-2n} \\ &+ 16e^n + 16e^{-n} - 4e^{2n} - 4e^{-2n} - 24 \\ = &\ -8n\left(\frac{e^n - e^{-n}}{2}\right) + 4n\left(\frac{e^{2n} - e^{-2n}}{2}\right) \\ &+ 32\left(\frac{e^n + e^{-n}}{2}\right) - 8\left(\frac{e^{2n} + e^{-2n}}{2}\right) - 24 \\ = &\ 4n\sinh(2n) - 8n\sinh(n) + 32\cosh(n) - 8\cosh(2n) - 24 \end{aligned}$$

We want to show

$$\forall\, n \in \mathbb{R} \setminus \{\, 0 \,\} \quad p_2(n) > 0 \tag{3.15}$$

First notice that $p_2(n)$ is an even function

$$\begin{aligned} p_2(-n) &= 4(-n)\sinh(-2n) - 8(-n)\sinh(-n) + 32\cosh(-n) - 8\cosh(-2n) - 24 \\ &= 4(-n)\left(-\sinh(2n)\right) - 8(-n)\left(-\sinh(n)\right) + 32\cosh(n) - 8\cosh(2n) - 24 \\ &= 4n\sinh(2n) - 8n\sinh(n) + 32\cosh(n) - 8\cosh(2n) - 24 \\ &= p_2(n) \end{aligned}$$

This means to show eq. (3.15), we only need to show

$$\forall\, n > 0 \quad p_2\left(n\right) > 0 \tag{3.16}$$

We show eq. (3.16) by analysing the derivatives

$$\begin{aligned} p_2'(n) &= 4\sinh(2n) + 8n\cosh(2n) - 8\sinh(n) - 8n\cosh(n) + 32\sinh(n) - 16\sinh(2n) \\ &= 8n\cosh(2n) - 8n\cosh(n) - 12\sinh(2n) + 24\sinh(n) \\ p_2''(n) &= 8\cosh(2n) + 16n\sinh(2n) - 8\cosh(n) - 8n\sinh(n) - 24\cosh(2n) + 24\cosh(n) \\ &= 16n\sinh(2n) - 8n\sinh(n) - 16\cosh(2n) + 16\cosh(n) \\ p_2^{(3)}(n) &= 16\sinh(2n) + 32n\cosh(2n) - 8\sinh(n) - 8n\cosh(n) - 32\sinh(2n) + 16\sinh(n) \\ &= 32n\cosh(2n) - 8n\cosh(n) - 16\sinh(2n) + 8\sinh(n) \\ p_2^{(4)}(n) &= 32\cosh(2n) + 64n\sinh(2n) - 8\cosh(n) - 8n\sinh(n) - 32\cosh(2n) + 8\cosh(n) \\ &= 64n\sinh(2n) - 8n\sinh(n) \\ &= 8n\left(8\sinh(2n) - \sinh(n)\right) \end{aligned}$$

Applying the identity

$$\sinh(2n) = 2\sinh(n)\cosh(n)$$

$p_2^{(4)}(n)$ becomes

$$p_2^{(4)}(n) = 8n\left(8\sinh(2n) - \sinh(n)\right)$$

$$= 8n\,(16\sinh(n)\cosh(n) - \sinh(n))$$
$$= 8n\sinh(n)\,(16\cosh(n) - 1)$$

Since $\cosh(n) \geq 1$ for all n we find

$$\forall n \in \mathbb{R}, \qquad 16\cosh(n) - 1 \geq 15 > 0$$

and for all $n > 0$, $\sinh(n) > 0$, hence we also find

$$\forall n > 0, \qquad 8n\sinh(n) > 0$$

therefore we conclude

$$\forall n > 0, \qquad p_2^{(4)}(n) > 0$$

Evaluating the other derivatives at $n = 0$ gives

$$\begin{aligned}
p_2(0) &= 4 \cdot 0 \cdot \sinh(2 \cdot 0) - 8 \cdot 0 \cdot \sinh(0) + 32\cosh(0) - 8\cosh(2 \cdot 0) - 24 \\
&= 32 - 8 - 24 \\
&= 0 \\
p_2'(0) &= 8 \cdot 0 \cdot \cosh(2 \cdot 0) - 8 \cdot 0 \cdot \cosh(0) - 12\sinh(2 \cdot 0) + 24\sinh(0) \\
&= 0 \\
p_2''(0) &= 16 \cdot 0 \cdot \sinh(2 \cdot 0) - 8 \cdot 0 \cdot \sinh(n) - 16\cosh(2 \cdot 0) + 16\cosh(0) \\
&= 16 - 16 \\
&= 0 \\
p_2^{(3)}(0) &= 32 \cdot 0 \cdot \cosh(2 \cdot 0) - 8 \cdot 0 \cdot \cosh(0) - 16\sinh(2 \cdot 0) + 8\sinh(0) \\
&= 0
\end{aligned}$$

Using proposition 2.1 and

$$\begin{aligned}
\forall n > 0, \qquad p_2^{(4)}(n) &> 0 \\
p_2^{(3)}(0) &= 0
\end{aligned}$$

we get

$$\forall n > 0, \qquad p_2''(n) > p_2''(0) = 0$$

We can use proposition 2.1 again with

$$\begin{aligned}
\forall n > 0, \qquad p_2''(n) &> 0 \\
p_2'(0) &= 0
\end{aligned}$$

we get

$$\forall n > 0, \qquad p_2(n) > p_2(0) = 0$$

showing eq. (3.16) and

$$\begin{aligned} D_2 &> 0, & w &\in S_1 \\ D_2 &> 0, & w &\in S_2 \end{aligned}$$

and therefore we have shown that $\nabla^2 m\,(x, y)$ is positive definite.

□

Proposition 3.5.
The reciprocal of the log mean temperature difference

$$m\,(x, y) = \frac{x - y}{\ln\left(\frac{x}{y}\right)}$$

is strictly convex over its entire domain, S:

$$S = \left\{ \, (x, y)^T \,\middle|\, x, y > 0 \, \right\}$$

Proof.
We have by proposition 3.4 that m is strictly convex over

$$\begin{aligned} S_1 &= \left\{ \, (x, y)^T \,\middle|\, 0 < y < x \, \right\} \\ S_2 &= \left\{ \, (x, y)^T \,\middle|\, 0 < x < y \, \right\} \end{aligned}$$

and we already know that

$$m \in \mathcal{C}^2$$

therefore only strict convexity over the set

$$S^* = \left\{ \, (x, x)^T \,\middle|\, x > 0 \, \right\}$$

remains to be shown, we do this by analysing $\nabla^2 m(x, y)$ over S^*.

For $m(x, y)$ to be strictly convex over S^*, we require for $\nabla^2 m(x, y)$ to be positive definite over S^*, this can be shown directly

$$\begin{aligned} \forall\, \mathbf{v} \in \mathbb{R}^2 \setminus \{\, \mathbf{0} \,\}, \qquad \mathbf{v}^T \nabla^2 m(x, x) \mathbf{v} &= \frac{1}{3x^3} \cdot \mathbf{v}^T \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \mathbf{v} \\ &= \frac{1}{3x^3} \left(2v_1^2 + 2v_1 v_2 + 2v_2^2 \right) \\ &= \frac{1}{3x^3} \left(v_1^2 + 2v_1 v_2 + v_2^2 + v_1^2 + v_2^2 \right) \\ &= \frac{1}{3x^3} \left[(v_1 + v_2)^2 + v_1^2 + v_2^2 \right] \\ &> 0 \end{aligned}$$

The inequality holds as $x > 0$ and $\mathbf{v} \neq \mathbf{0}$.

Hence, we have shown that $\nabla^2 m(x, y)$ is positive definite over S^*.

This result along with proposition 3.4 and the twice continuous differentiability property of m shows that $m(x, y)$ is strictly convex over S.

□

4 Approaches

4.1 Overview

We have a MINLP model in which nonlinearities arise from:

- bilinear constraints
- nonlinear convex functions (reciprocal of LMTD) in the constraints
- concave functions in the objective (A^β)

and we want to create a MILP formulation that we can iteratively reformulate and resolve to provide an estimated solution for the MINLP model.

We will look at two approaches to approximating the reciprocal of LMTD, $m(x, y)$, these are: a piecewise linear approximation and an outer approximation. We will discuss which approach is better given the structure of the model.

4.2 Piecewise Linear Functions (PLFs)

By the strict convex property of m, we have that any piecewise linear approximation, $\hat{m}_p$, will form an over estimator for the function. As the function is bivariate, a piecewise linear approximation means that we select a set of breakpoints and generate a mesh of triangles (section 2.2.7). An example is shown in figs. 4.1c and 4.1d where:

- four triangles are plotted
- the selected breakpoints are symmetric about $y = x$ to maintain the symmetry of m

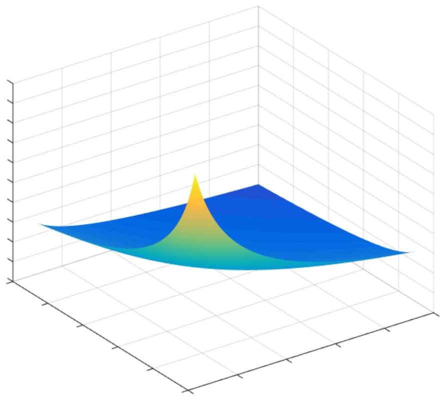
We can see that this piecewise linear approximation has a similar shape to that of m. It is important to note that maintaining the symmetry of the function is important as we would like to model m closely and this property is easily modelled in our approximation. This also makes error analysis easier as we only need to reason the maximal error over half of the domain e.g. over S^L in fig. 3.4.

4.3 Outer Approximations (OAs)

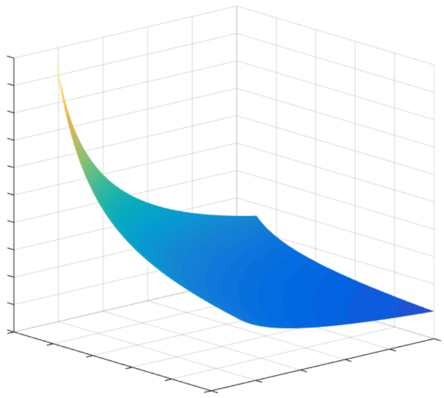
By the strict convex property of m we have that any outer approximation will form an under estimator for the function. Since all outer approximations are under estimators, we can define $\hat{m}_o$ to be the maximum of these at any given point hence $\hat{m}_o$ is also an under approximator.

We select a set of points at which we generate $\hat{m}_o$ by taking the tangent to the function. The tangent is indeed a linear support function due to the strict convex nature of the function (proposition 3.5) and theorem 2.7. An example is shown in figs. 4.1e and 4.1f where

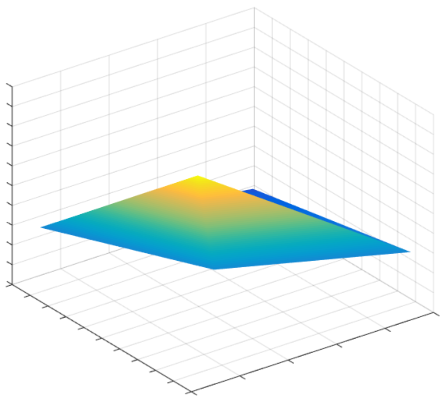
- five tangents are plotted



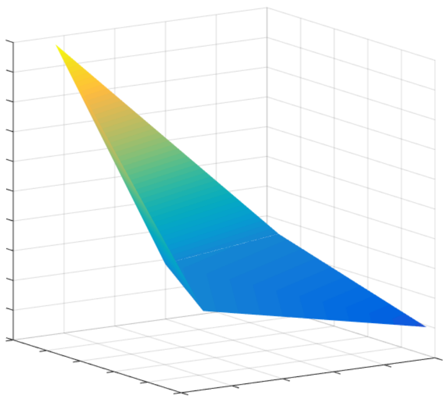
(a) m from orientation a



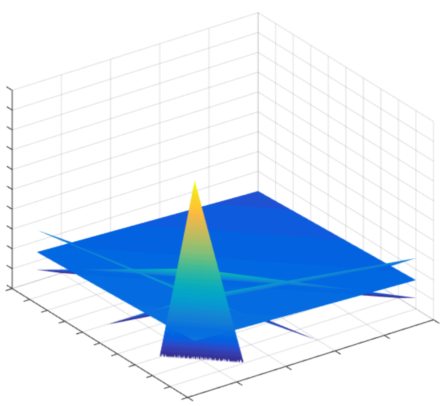
(b) m from orientation b



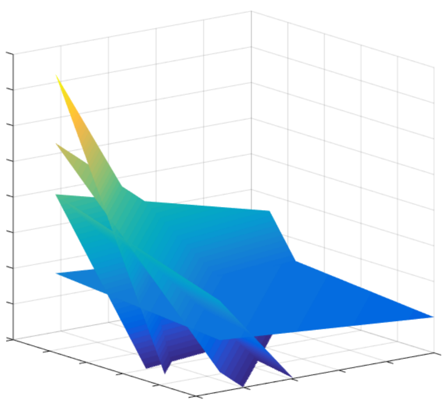
(c) piecewise linear approximation from orientation a



(d) piecewise linear approximation from orientation b



(e) outer approximation from orientation a



(f) outer approximation from orientation b

Figure 4.1: Plots of: m, a piecewise linear approximation of m and a set of outer approximations of m viewed from similar orientations: a is looking down along the $(1,1)^T$ direction - the 'front', b is looking along the $(-1,1)^T$ direction - the 'side'.

- to preserve symmetry, if a tangent is generated at (x, y) with $x \neq y$ then a tangent is generated at (y, x)

The similarity between the outer approximation and m is harder to see however if we look at only the top most tangents, we see that the colours (function value) are similar to that of the plot of m .

Once again symmetry is maintained to model the symmetric property found from m.

4.4 Comparison of the Methods

We compare the two approaches to linearising m under a simplified model. The only expectation that we have is: the results from using PLFs should be larger than the corresponding results from using outer approximations.

4.4.1 Simplifying the Non-LMTD Constraints

The model is simplified by relaxing any bilinearities in the model to their entire convex hull specified by the McCormick envelopes (section 2.2.9). The bilinearities in the model correspond to the energy balance constraints and heat exchanger area calculations. In general we add the McCormick hull by applying the following.

Let s and t be variables of the model such that $s \in [s_l, s_u]$ and $t \in [t_l, t_u]$. We incorporate the McCormick convex hull of $s \cdot t$ into the model by introducing the auxilary variable r into the model and the following constraints

$$
\begin{aligned}
r &\geq s_l\, t + t_l\, s - s_l\, t_l \\
r &\geq s_u\, t + t_u\, s - s_u\, t_u \\
r &\leq s_l\, t + t_u\, s - s_l\, t_u \\
r &\leq s_u\, t + t_l\, s - s_u\, t_l
\end{aligned}
$$

we then replace any occurrence of $s \cdot t$ with r.

In the heat exchanger model, we assign the following auxiliary variables to the respective bilinear terms, generating the McCormick hull constraints as above

$$
\begin{aligned}
b^{H,\text{in}}_{ijk} &\twoheadrightarrow f^H_{ijk} t_{ik} \\
b^{H,\text{out}}_{ijk} &\twoheadrightarrow f^H_{ijk} t^H_{ijk} \\
b^{C,\text{in}}_{ijk} &\twoheadrightarrow f^C_{ijk} t_{jk+1} \\
b^{C,\text{out}}_{ijk} &\twoheadrightarrow f^C_{ijk} t^C_{ijk}
\end{aligned}
$$

and replace the associated model constraints as follows

$$
\begin{aligned}
q_{ijk} = f^H_{ijk}\left(t_{ik} - t^H_{ijk}\right) \quad &\twoheadrightarrow \quad q_{ijk} = b^{H,\text{in}}_{ijk} - b^{H,\text{out}}_{ijk} \\
q_{ijk} = f^C_{ijk}\left(t^C_{ijk} - t_{jk+1}\right) \quad &\twoheadrightarrow \quad q_{ijk} = b^{C,\text{out}}_{ijk} - b^{C,\text{in}}_{ijk}
\end{aligned}
$$

	Model 1	Model 2	Model 3
Two Triangle Model	449135.87	INFEASIBLE	899320.95
Four Triangle Model	257821.95	710365.80	285685.84
Single Tangent Model	114118.92	594277.11	57499.24
Multiple Tangents Model	114118.92	601172.84	57595.36
Global Solutions	154895.93	634200.38	64898.23

Table 4.1: The results of the models we compared and the global solutions found for these models

Since we are using the reciprocal of LMTD now, we replace the area constraints by

$$\begin{aligned}
A_{ijk} &= q_{ijk}\left(h_i^{-1}+h_j^{-1}\right)/LMTD_{ijk} &\twoheadrightarrow\quad A_{ijk} &= q_{ijk}\left(h_i^{-1}+h_j^{-1}\right) RecLMTD_{ijk}\\
A_{cui} &= q_{cui}\left(h_i^{-1}+h_{CU}^{-1}\right)/LMTD_{cui} &\twoheadrightarrow\quad A_{cui} &= q_{cui}\left(h_i^{-1}+h_{CU}^{-1}\right) RecLMTD_{cui}\\
A_{huj} &= q_{huj}\left(h_j^{-1}+h_{HU}^{-1}\right)/LMTD_{huj} &\twoheadrightarrow\quad A_{huj} &= q_{huj}\left(h_j^{-1}+h_{HU}^{-1}\right) RecLMTD_{huj}
\end{aligned}$$

where $RecLMTD_{ijk}, RecLMTD_{cui}$ and $RecLMTD_{huj}$ are the variables we use to model the reciprocal of LMTD.

Since the area contraint is a bilinear equality we don't need to introduce an auxilary variable as the area variable can be used.

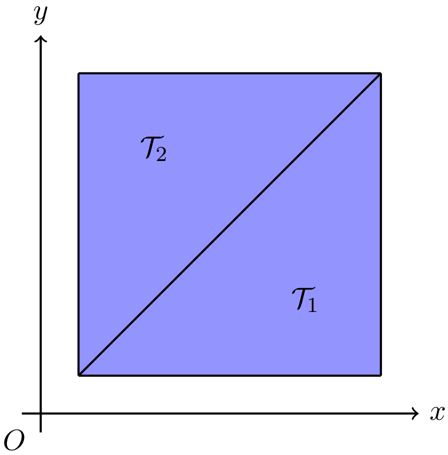
The model also contains a univariate nonlinear concave function in the objective function. This is linearised by taking a single line segment between function evaluated at the endpoints of its domain.

4.4.2 Results of the Simplified Model

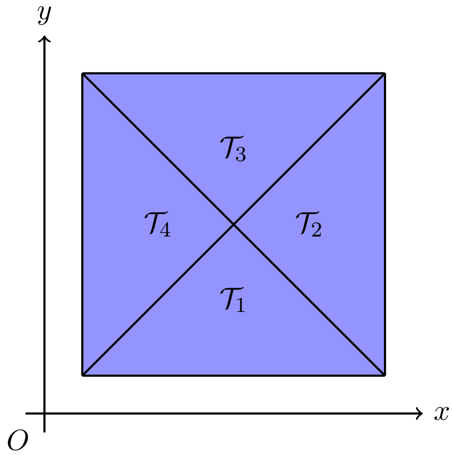
Here we will compare the PLF and OA approach by under the simplified model described in section 4.4.1.

For the piecewise linear approach we partition the domain into triangles, $\mathcal{T}_i$. The vertices are assigned the value of m that they correspond to. Any intermediate point inside the triangle can be represented as a convex combination of the vertices. m is estimated by taking the same convex combination of the values assigned to the vertices. We represent being in a specfic triangle using binary variables. The 'two triangle' and 'four triangle' model were run, these are shown in figs. 4.2a and 4.2b respectively.

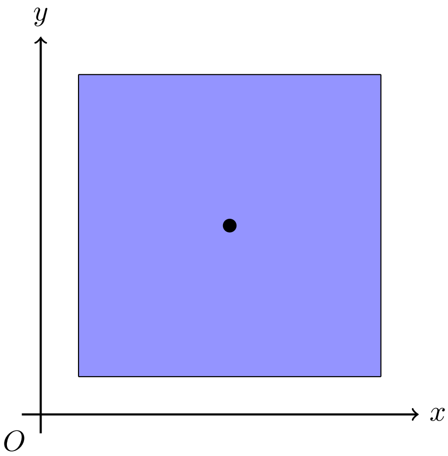
For the tangent approach we select a set of points in the domain and generate the equation of the tangent at that point. We then add a $\geq$ constraint per tangent point. The structure of model dictates that m be minimised hence the value assigned to m will be on the top most tangent (we have $\geq$ constraints). We don't need to add binary variables for this approach. A 'single tangent' and 'multiple tangents' model were run, these are shown in figs. 4.2b and 4.2c respectively.
From table 4.1, we see the expected result that the PLF models result in larger values than that of the outer approximation. We also see that the results are on either side of the global solutions.



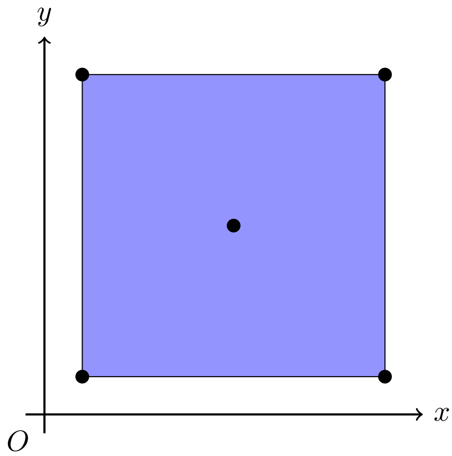
(a) The 'two triangle' PLF estimation



(b) The 'four triangle' PLF estimation

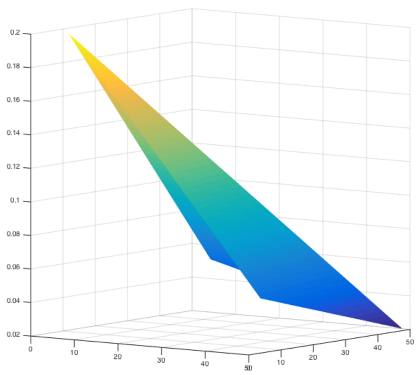


(c) Single tangent point

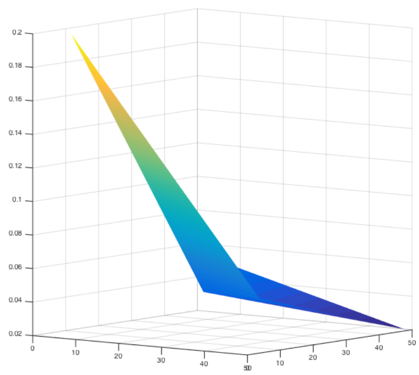


(d) Multiple tangent points

Figure 4.2: The locations of our piecewise linear functions (top) and our tangent locations (bottom) to run our comparison of the two approaches



(a) The two triangle estimation



(b) The four triangle estimation

Figure 4.3: The side views of the PLF estimations made

In both cases, we did not add any intelligent method of breakpoint/tangent selection, the points we selected simply partition the domain equally, in the case of PLF, and are equally spaced apart, in the case of the outer approximation. If these approaches are as good as each other we expect to find similarities in the results, this is not the case. For the outer approximation we see a small increase in the solutions as we add more tangents (tighter constraints) whereas for the PLF approach we see: an infeasibilty for the two triangle model and drastic drops in the solution as we add a small number of breakpoints. Here we question why the PLF solutions are so different.

4.4.3 Modelling the Reciprocal of LMTD

We want our linear approximation to have similar properties to that of m, the easiest property to incorporate is symmetry, this is done by:

- defining triangles in S^L, fig. 3.4b, then choosing the reflections for the remaining triangles when creating a PLF e.g. in fig. 4.2b, we define $\mathcal{T}_1$ and $\mathcal{T}_2$, the remaining triangles are reflections represented by $\mathcal{T}_4$ and $\mathcal{T}_3$ respectively.
- letting our set of tangent points, $\mathcal{G}$, have the following property

  $$(x, y)^T \in \mathcal{G} \implies (y, x)^T \in \mathcal{G}$$

  when creating an outer approximation. (This process creates a symmetric approximation by lemma 3.1). This is shown in fig. 4.2d where we have a set of tangent points on the $y = x$ line, these points are trivially in $\mathcal{G}$, the remaining two points are reflections of each other.

Therefore we see that a symmetric estimation can be created for both approaches.

Positivity of m is also important. A PLF approach will always result in the following condition:

$$\min \hat{m}_p(x, y) = \min m(x, y) > 0$$

as the function value at the vertices are exact. The positivity of $\hat{m}_o$ does not necessarily hold. A sufficient condition for positivity of $\hat{m}_o$ is

$$(u, u)^T \in \mathcal{G}$$

where u is the upperbound on x and y. We can also ensure positivity by setting the bounds in the implementation i.e. an additional constraint.

Another property that we would like to have is a convex shape, this does not always occur for the PLF approach, it is clear from fig. 4.3a that we can construct a PLF for m that is concave, it is also clear from fig. 4.3b that we can construct a PLF that is convex. We can also create a PLF that is neither convex or concave as we add more breakpoints. This creates an added layer of complexity when choosing breakpoints. For the outer approximation, we take the maximum value of all of the linear estimates generated by our set $\mathcal{G}$, a linear function is convex therefore by theorem 2.6, the outer approximation is convex.

4.4.4 Advantages of the Outer Approximation

In section 4.4.3, we saw that we can model the symmetry, bounds and shape of m using both approaches. However, the outer approximation proves to be a better choice given the context of the other nonlinear approximations and the way in which breakpoints would have to be generated in the PLF approach. Here we discuss some of these advantages.

A PLF approach would have to use binary variables. On further analysis, the required amount of binary variables may prove to be small however as the size of the model grows, the number of potential heat exchangers increases, we have an associated m function per heat exchanger hence the introduction of an extra stream in the model could see a large number of binary variables introduced with it. Modelling with binary variables should be avoided as they are the bottleneck when it comes to solving a mixed integer program. An outer approximation does not use binary variables.

In PLFs, breakpoint generation would require a complete reformulation of our approximation. If we simply added a breakpoint where we found our solution, we would only make the solution better in the same location hence the model will converge on the same solution with a better objective value. This tells us nothing about the location of the global solution. If we want to find the global solution using the PLF approach, we want to keep the maximum error constant across each piece to create an element of fairness among the pieces. This means that we require a complete reformulation of the approximation on consecutive solves hence one estimation can be completely different from the next. If we can remove a set of triangles, we would create a speed up however we are estimating other functions hence reasoning the rejection of a set of triangles is hard. An outer approximation tightens the constraints hence we can get a jump to a different solution (by different we mean a different set of active heat exchangers), therefore adding tangents moves us closer to the global minimum without the need to completely reformulate the approximation.

Finally, the outer approximation is a better fit given the other estimations we make. We use the McCormick hull for the bilinearities, these form an under approximator as they extend the feasible region. We also use a piecewise linear function (section 2.2.7) to estimate a concave function in the objective, this is also an under approximator. Adding the outer approximation for m makes a MILP model using the outer approximation an under estimator of the MINLP model hence tightening any of our estimations only increases the objective value i.e. the solution moves in one direction as we improve our estimate. If we use PLFs we are mixing an over estimator with under estimators, therefore we may get an under estimated solution if the PLF error is low enough with respect to the other estimates.

5 Implementation

5.1 Overview

Here we will present an algorithm that reformulates the MINLP heat exchanger model (appendix B) into a MILP estimation. The algorithm has a priori partitioning to form an outer approximation and iteratively generates cuts to converge to a solution within a set of termination criteria. The linear approximations we make are

- McCormick relaxation for bilinearities (section 2.2.9)
- outer approximation for the reciprocal of LMTD (section 2.2.8)
- a piecewise linear approximation for the concave nonlinearities in the objective function (section 2.2.7)

5.2 Reformulating the Model

The algorithm reformulates the MINLP model into a MILP approximation. This is then solved multiple times with each iteration having a progressively tighter approximation.

We make references to model variables and parameters the descriptions for these can be found in section B.1.

5.2.1 Approximating the Concave Functions

The objective function, TAC, is:

$$\begin{aligned} TAC = \min \ & c_{CU} \sum_{i \in HP} q_{cui} + c_{HU} \sum_{j \in CP} q_{huj} + c \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} z_{ijk} + c \sum_{i \in HP} z_{cui} \\ & + c \sum_{j \in CP} z_{huj} + \alpha \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} A_{ijk}^{\beta} + \alpha \sum_{i \in HP} A_{cui}^{\beta} + \alpha \sum_{j \in CP} A_{huj}^{\beta} \end{aligned}$$

of the above terms the nonlinearities are: $A_{ijk}^{\beta}, A_{cui}^{\beta}$ and A_{huj}^{β} where $0 < \beta \leq 1$.

These are approximated using piecewise linear functions as follows.

We will use the variable A to describe the general case. The entire model is bounded hence we can derive bounds for the variables used. Specifically, this means that we have that $A \in [0, A_{upper}]$ where $A_{upper} > 0$ (we know that the lower bound of A is 0).

We form a n-piece PLF for A^{β} by spliting the domain into n equal intervals. To do this we define the constant ΔA

$$\Delta A = \frac{A_{upper}}{n}$$

We only a single interval to be active in any solution therefore we add n binary variable,

$z_i^{(A)}$, to the model with the constraint

$$\sum_{i=1}^{n} z_i^{(A)} = 1$$

We define the respective lower and upper bounds of interval i as

$$A_i^{(L)} = \Delta A(i-1)$$
$$A_i^{(U)} = \Delta A \cdot i$$

To ensure that A stays inside interval i by defining the following lower and upper bounds respectively

$$A_l = \sum_{i=1}^{n} A_i^{(L)} \cdot z_i^{(A)}$$
$$A_u = \sum_{i=1}^{n} A_i^{(U)} \cdot z_i^{(A)}$$

and adding the constraint

$$A_l \leq A \leq A_u$$

Finally, we want to calculate the estimated value of A^{β}. To do this we define gradients, $g_i^{(A)}$, that we associate with interval i

$$g_i^{(A)} = \frac{\left(A_i^{(U)}\right)^{\beta} - \left(A_i^{(L)}\right)^{\beta}}{A_i^{(U)} - A_i^{(L)}}$$

our estimate, $\hat{A}^{(\beta)}$, is formed by adding the constraint

$$\hat{A}^{(\beta)} = \sum_{i=1}^{n} z_i^{(A)} \left(\left(A_i^{(L)}\right)^{\beta} + g_i^{(A)} \cdot (A - A_l)\right)$$

We have a degenerate case of $\beta = 1$, here our approximation is exact when $n = 1$ hence we reduce the above to a single constraint:

$$\hat{A}^{(\beta)} = A$$

and make any other constraints feasible by default.

5.2.2 Reformulating the Bilinearities

Using the entire McCormick hull for a given bilinear term has too large an error associated with it making it hard for us to reason correctness of a solution. We improve our bilinear estimate by partitioning the domain of one of the variables into multiple equally sized intervals.

The maximum error of using the McCormick hull [1] to estimate $b \cdot x \cdot y$ (b constant) is

$$\frac{|b|}{4}(x_u - x_l)(y_u - y_l) \tag{5.1}$$

if we partition on x we have that $(x_u - x_l)$ decreases and is bounded by 0 hence the error does for an intermetediate McCormick hull decreases. This is shown by fig. 5.1 where a partitioning system captures the shape of the function more accurately.

This partitioning process described here and used in the algorithm is the **nf4r** scheme for relaxing bilinear terms [20]. The presentation here is similar to the presentation given by Misener, Thompson, and Floudas.

Here we will assume that our bilinear constraint takes the form

$$\begin{gathered} x \cdot y \\ x_l \leq x \leq x_u \\ y_l \leq y \leq y_u \end{gathered}$$

and that we are creating N partitions on the domain of x, any scalar multipliers can be applied to the introduced constraints after formulation.

First we introduce the following constants and variables

- the interval size of all partitions, a

$$a = \frac{x_u - x_l}{N}$$

- binary variables, λ_i, $i \in \{ 1, ..., N \}$, to indicate which interval x is in
- continuous variables, Δy_i, $i \in \{ 1, ..., N \}$

We define constraints to enforce that:

- only one interval (λ_i) can be active

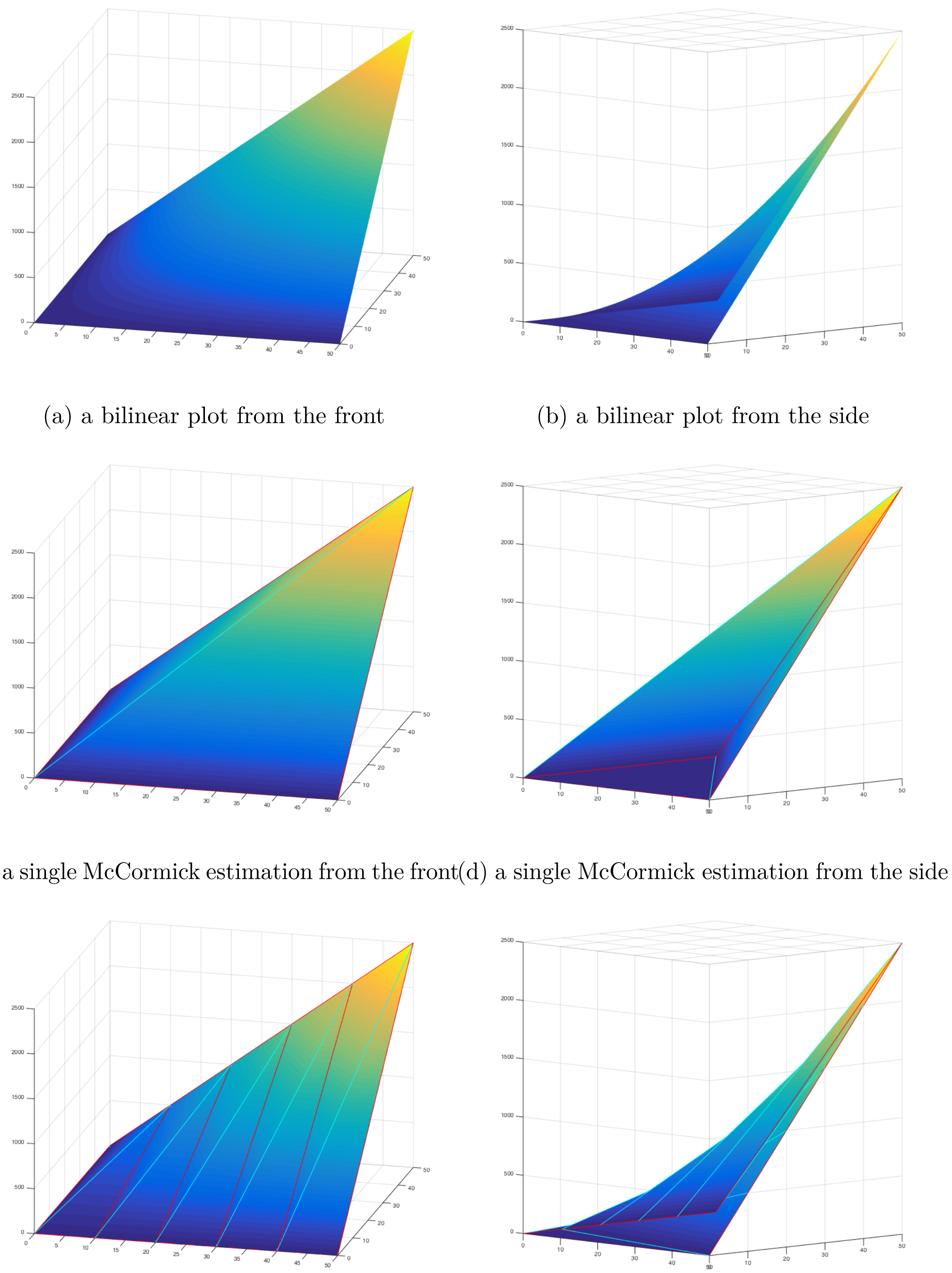
$$\sum_{i=1}^{N} \lambda_i = 1$$

- x must be in the active interval

$$x_l + \sum_{i=1}^{N} a(i-1) \cdot \lambda_i \leq x \leq x_l + \sum_{i=1}^{N} a \cdot i \cdot \lambda_i$$

We also require that Δy_i be equal to $(y - y_l)$ when $\lambda_i = 1$ and 0 otherwise and for $\Delta y_i \in [0, y_u - y_l]$. This is captured by the following constraints [38]:

$$\begin{gathered} y = y_l + \sum_{i=1}^{N} \Delta y_i \\ 0 \leq \Delta y_i \leq (y_u - y_l)\lambda_i, \quad \forall i \in \{ 1, ..., N \} \end{gathered}$$



(a) a bilinear plot from the front

(b) a bilinear plot from the side

(c) a single McCormick estimation from the front

(d) a single McCormick estimation from the side

(e) multiple disjoint McCormick estimations from the front

(f) multiple disjoint McCormick estimations from the side

Figure 5.1: Plots of: a bilinear function, the associated McCormick hull and a set of McCormick hulls constructed by partitioning the domain of one of the variables (the red and cyan lines indicate the edges of the McCormick hulls)

We construct the McCormick convex hull of the bilinear term over the reduced interval using the following constraints, the bilinearities in the model are then replaced by z.

$$z \geq x \cdot y_l + \sum_{i=1}^{N} [x_l + a(i-1)] \cdot \Delta y_i \tag{5.2}$$

$$z \geq x \cdot y_u + \sum_{i=1}^{N} [x_l + a \cdot i\,] \cdot [\Delta y_i - (y_u - y_l) \cdot \lambda_i] \tag{5.3}$$

$$z \leq x \cdot y_l + \sum_{i=1}^{N} [x_l + a \cdot i\,] \cdot \Delta y_i \tag{5.4}$$

$$z \leq x \cdot y_u + \sum_{i=1}^{N} [x_l + a(i-1)] \cdot [\Delta y_i - (y_u - y_l) \cdot \lambda_i] \tag{5.5}$$

The correctness of the above formulation can be seen by directly comparing against the general McCormick formulation i.e. it can be shown to be equivalent to the McCormick hull over the reduced interval.

We show equivalence for eq. (5.2):

If we assume that we are in interval k, $k \in \{\, 1, ..., N \,\}$ then the lower bound of x is $x_l + a(k-1) = \hat{x}_{l,k}$. We also know that $\Delta y_i = 0$ for all $i \neq k$ hence we get

$$\begin{aligned} z &\geq x \cdot y_l + \sum_{i=1}^{N} [x_l + a(i-1)] \cdot \Delta y_i \\ &= x \cdot y_l + \hat{x}_{l,k} \cdot \Delta y_k \end{aligned}$$

But $\Delta y_k = (y - y_l)$ therefore we get

$$\begin{aligned} z &\geq x \cdot y_l + \hat{x}_{l,k} \cdot \Delta y_k \\ &= x \cdot y_l + \hat{x}_{l,k} \cdot (y - y_l) \\ &= x \cdot y_l + \hat{x}_{l,k} \cdot y - \hat{x}_{l,k} \cdot y_l \end{aligned}$$

which is equivalent to one of the constraints of the McCormick hull over the reduced interval. The remaining constraints can be shown to be equivalent using similar reasoning.

5.2.3 Reformulating the Reciprocal of LMTD

We reformulate the reciprocal of LMTD by creating an outer approximation. This can be done easily because of the strict convex property (section 3.6.3). For a strictly convex function we have the following property:

Let $\Omega \subseteq \mathcal{R}^n$, $f : \Omega \to \mathbb{R}$ be strictly convex then

$$\forall \mathbf{x}, \mathbf{y} \in \Omega, \mathbf{x} \neq \mathbf{y} : f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \tag{5.6}$$

We select a set of tangent points, $\mathcal{G}$, and add a constraint for each tangent point as follows.

For each $(t_1, t_2)^T \in \mathcal{G}$ add the constraint

$$\hat{m}(x, y) \geq m(t_1, t_2) + \nabla m(t_1, t_2)^T((x, y)^T - (t_1, t_2)^T)$$

Let $\mathcal{G}^{\text{init}}$ be the initial set of tangent points. We model symmetry and positivity by respectively having the following conditions on $\mathcal{G}^{\text{init}}$.

Assuming the domain of m is $\left\{ (x, y)^T \;\middle|\; l \leq x, y \leq u \right\}$

$$(x, y)^T \in \mathcal{G}^{\text{init}} \implies (y, x)^T \in \mathcal{G}^{\text{init}}$$
$$(u, u)^T \in \mathcal{G}^{\text{init}}$$

5.3 The Algorithm

We formulate an algorithm to solve the HEN model iteratively. We create a model with a priori partitioning and solve it iteratively where we add cutting planes on each iteration.

We make use of the following:

- the set of hot streams, HP
- the set of cold streams, CP
- the set of stages, ST
- the set of stream to stream heat exchangers, $HX = \left\{ (i, j, k)^T \;\middle|\; i \in HP, j \in CP, k \in ST \right\}$

5.3.1 Initialisation

The algorithm is initialised with

- the model parameters required as specified by the general model
- bilinear constraints replaced by the **nf4r** scheme for disjoint McCormick hulls
- an initial outer approximation for all reciprocal of LMTD constraints
- a piecewise linear approximation for all nonlinear concave functions in the objective

For the above, we assume that the number of pieces:

- N_q, used to partition the heat load bilinear constraints
- N_A, used to partition the area bilinear constraints
- N_β, used to partition the concave functions in the objective

are passed to the model as parameters, this leaves the outer approximation initialisation.

Initialising the Outer Approximation for Stream to Stream Heat Exchangers
For a given stream to stream heat exchanger $(i, j, k)^T \in HX$ the domain of m_{ijk} is:

$$S_{ij} = \left\{ (x, y)^T \;\middle|\; l_{ij} \leq x, y \leq u_{ij} \right\}$$

where

$$l_{ij} = \Delta T_{\min}$$
$$u_{ij} = T_i^{\text{in}} - T_j^{\text{in}}$$

for the above the stage is irrelevant.

Using proposition 3.3, we have the following bounds on m_{ijk}.

$$\min_{(x,y)^T \in S_{ij}} m_{ijk}(x, y) = \frac{1}{u_{ij}}$$
$$\max_{(x,y)^T \in S_{ij}} m_{ijk}(x, y) = \frac{1}{l_{ij}}$$

By proposition 3.2, we can ensure that the outer approximation, $\hat{m}_{ijk}$, is also bounded as above. This is done by asserting the following condition

$$\left\{ (l_{ij}, l_{ij})^T, (u_{ij}, u_{ij})^T \right\} \subseteq \mathcal{G}_{ijk}^{\text{init}} \tag{5.7}$$

where $\mathcal{G}_{ijk}^{\text{init}}$ is the set of initial tangent points used to create the outer approximation.

The bounds mentioned above differ among heat exchangers, we use a generalised process of initialisation across all stream to stream heat exchangers using convex combinations. It is assumed that we have a set, $\mathcal{P}^{\text{init}} \subset \mathbb{R}^3$, satisfying:

$$\mathcal{P}^{\text{init}} \subset \left\{ (\lambda_a, \lambda_b, \lambda_c)^T \;\middle|\; \lambda_a + \lambda_b + \lambda_c = 1;\ \lambda_a, \lambda_b, \lambda_c \geq 0 \right\}$$

The initial set of tangent points is calculated using three of the vertices of S_{ij} (the set is square)

$$\mathbf{a} = (l_{ij}, l_{ij})^T, \quad \mathbf{b} = (u_{ij}, u_{ij})^T, \quad \mathbf{c} = (u_{ij}, l_{ij})^T$$

For each vector $(\lambda_a, \lambda_b, \lambda_c)^T \in \mathcal{P}^{\text{init}}$, we calculate

$$(x, y)^T = \lambda_a \mathbf{a} + \lambda_b \mathbf{b} + \lambda_c \mathbf{c}$$

then to preserve symmetry add the follwing tangent points to $\mathcal{G}_{ijk}^{\text{init}}$

$$\left\{ (x, y)^T, (y, x)^T \right\}$$

Equation (5.7) is asserted by having

$$\left\{ (1, 0, 0)^T, (0, 1, 0)^T \right\} \subseteq \mathcal{P}^{\text{init}}$$

The outer approximation is formed by adding the following constraint for each $(g_1, g_2)^T \in \mathcal{G}_{ijk}^{\text{init}}$

$$\hat{m}_{ijk}(x,y) \geq m(g_1,g_2) + \nabla m(g_1,g_2)^T \left((x,y)^T - (g_1,g_2)^T\right)$$

Initialising the Outer Approximation for Utility Heat Exchangers
For the cold and hot utilities, adding the tangents is simpler as we have that the second parameter of m is fixed therefore the function becomes univariate in the first parameter.

For a given hot stream $i \in HP$ and cold stream $j \in CP$, the respective fixed second parameter values for the cold and hot utilities are

$$\begin{aligned} c_{cui} &= T_i^{\text{out}} - T_{CU}^{\text{in}} \\ c_{huj} &= T_{HU}^{\text{in}} - T_j^{\text{out}} \end{aligned}$$

The initialisation process for outer approximations of m is similar to the stream to stream case.

Our domains for the univariate m are

$$\begin{aligned} S_{cui} &= \{ \, x \mid l_{cui} \leq x \leq u_{cui} \, \} \\ S_{huj} &= \{ \, x \mid l_{huj} \leq x \leq u_{huj} \, \} \end{aligned}$$

where

$$\begin{aligned} l_{cui} &= \Delta T_{\min} \\ u_{cui} &= T_i^{\text{in}} - T_{CU}^{\text{out}} \\ l_{huj} &= \Delta T_{\min} \\ u_{huj} &= T_{HU}^{\text{out}} - T_j^{\text{in}} \end{aligned}$$

The remaining process is only shown for the cold utility, the hot utility initialisation is done in the same we only need to change the subscripts.

Using proposition 3.2, we have the following bound on m_{cui}

$$\begin{aligned} \min_{x \in S_{cui}} m_{cui}(x, c_{cui}) &= m(u_{cui}, c_{cui}) \\ \max_{x \in S_{cui}} m_{cui}(x, c_{cui}) &= m(l_{cui}, c_{cui}) \end{aligned}$$

By proposition 3.2 again, we can ensure that the outer approximation, $\hat{m}_{cui}$ is also bounded as above. This is done by asserting the following condition

$$\{ \, l_{cui}, u_{cui} \, \} \subseteq \mathcal{G}_{cui}^{\text{init}} \tag{5.8}$$

where $\mathcal{G}_{cui}^{\text{init}}$ is the initial tangent points used to create the outer approximation for the cold utility.

As the bounds mentioned above differ among heat exchangers, we use a generalised process of initialisation across all of the cold utility heat exchangers. It is assumed that we have set, $\mathcal{P}^{\text{init}}_{CU} \subset [0,1]$ this is the cold utility analogue to $\mathcal{P}^{\text{init}}$.

For each $\lambda \in \mathcal{P}^{\text{init}}_{CU}$, we add to $\mathcal{G}^{\text{init}}_{cui}$ the calculated value

$$\lambda l_{cui} + (1-\lambda) u_{cui}$$

Equation (5.8) is asserted by having

$$\{\,0,1\,\} \subseteq \mathcal{P}^{\text{init}}_{CU}$$

The outer approximation is formed by adding the following constraint for each $g \in \mathcal{G}^{\text{init}}_{cui}$

$$\hat{m}_{cui}(x) \geq m(g, c_{cui}) + m'_x(g, c_{cui})\,(x-g)$$

Initialising the Model

Assuming that we have defined the sets $\mathcal{P}^{\text{init}}, \mathcal{P}^{\text{init}}_{CU}$ and $\mathcal{P}^{\text{init}}_{HU}$ the initialisation process is:

begin
  model = model with linear constraints from generalised formulation
  for $(i,j,k)^T \in HX$ **do**
    add **nf4r** McCormick hull contraints for q_{ijk}
    add **nf4r** McCormick hull contraints for A_{ijk}
    add piecewise linear constraints for A^{β}_{ijk}
    add initial $RecLMTD_{ijk}$ outer approximation constraints
  end
  for $i \in HP$ **do**
    add **nf4r** McCormick hull contraints for A_{cui}
    add piecewise linear constraints for A^{β}_{cui}
    add initial $RecLMTD_{cui}$ outer approximation constraints
  end
  for $j \in CP$ **do**
    add **nf4r** McCormick hull contraints for A_{huj}
    add piecewise linear constraints for A^{β}_{huj}
    add initial $RecLMTD_{huj}$ outer approximation constraints
  end
  return *model*
end

Algorithm 1: function: *CreateInitialModel()*

5.3.2 The Iterative Process

The idea behind the algorithm is that we can use the robustness of MILP solvers to converge to the global solution of the MINLP model. Parameters N_q, N_A and N_β are provided and we generate linear constraints to approximate the nonlinearities that these constants correspond to. Hence after initialisation these constraints are taken to be fixed and further approximations are not made to them. As for the reciprocal of LMTD, we have initialisation conditions where each outer approximation is constructed with a similar structure. The initial model is solved. The iterative process takes the solution to the previously solved model, identifies the active heat exchangers and generates further outer approximation constraints at the corresponding points at which the solution settled with respect to the reciprocal of LMTD. This is characterised by alg. 2.

```
begin
    model = CreateInitialModel()
    currentSolution = solve model
    while currentSolution doesn't satisfy termination criteria do
        activeHeatExchangers = active heat exchangers of currentSolution
        for activeHeatExchanger in activeHeatExchangers do
            (t1, t2)^T = solution of activeHeatExchanger
            Add tangent constraint corresponding to (t1, t2)^T to model
            Add tangent constraint corresponding to (t2, t1)^T to model
        end
        currentSolution = solve model
    end
    return currentSolution
end
```

Algorithm 2: The cutting plane algorithm

5.3.3 Termination Criteria

A simple condition for termination would be to run the algorithm until there are no further tangents to generate i.e. the solution has settled where we have already placed a tangent for every approximation. The problem with this approach is that we do not know how quickly the solution is approached or if it is ever reached. We can illustrate this with the following example.

For simplicity, we assume:

- on each iteration the active set of heat exchangers is the same
- on each iteration we are only adding further outer approximations to the same heat exchanger, h, i.e. all other heat exchangers have tangent points placed at their global solution values and they are not moving
- h is a utility heat exchanger
- c is the global solution point of h and we have not placed a tangent at c

For each iteration i assume that we add a tangent point for h at $c + \frac{1}{i}$ this is a sequence that is converging to the global solution however the algorithm will not terminate.

As we cannot guarantee that the algorithm will stop adding constraints, we add termination criteria to allow for the algorithm to terminate to an epsilon-accuracy. As we keep the non-LMTD approximations fixed the termination criteria we add to the model is

$$E_m(x, y) = |m(x, y) - \hat{m}(x, y)| \leq \varepsilon_m$$

for all active heat exchangers, for some $\varepsilon_m > 0$.

As we are not concerned with the inactive heat exchangers, these are not included in our termination criteria.

5.4 Convergence

We will show that the algorithm converges to the global solution. First we analyse how the errors cascade throughout the model then, by placing appropriate bounds we will show how errors in the approximations affect the total error in the objective function. With this, we will prove convergence of the algorithm by showing that: as we reduce our allowed error, representing the error in the solution with respect to the global solution, we can construct a model that will give a solution within ε of the global solution. We will then relate this back to the algorithm proving that it converges.

We will use the following notation:

- Real and approximated values - the real value of variable x will have approximated value $\hat{x}$
- Reciprocal of LMTD - We take *RecLMTD* to be the variable in the model however for clarity, we shall use m to denote the real value and $\hat{m}$ to denote the approximated value. The relevant subscripts will be attached.

5.4.1 Cascading Errors

In the model, we are approximating many different nonlinearities. We will analyse how the approximation errors are related to each other and how they affect the objective value.

The approximated values are the:

- stream to stream heat loads, $\hat{q}_{ijk}$
- stream to stream reciprocal of LMTDs, $\hat{m}_{ijk}$
- stream to stream areas, A_{ijk}
- stream to stream area scalings, A^{β}_{ijk}
- cold utility reciprocal of LMTDs, $\hat{m}_{cui}$
- cold utility areas, A_{cui}
- cold utility area scalings, A^{β}_{cui}
- hot utility reciprocal of LMTDs, $\hat{m}_{huj}$
- hot utility areas, A_{huj}
- hot utility area scalings, A^{β}_{huj}

Of the above, some of the error in approximations are carried forward into further approximations e.g. when we calculate A^{β}_{ijk} we are using A_{ijk} which already carries an approximation error. This creates a cascade of errors which affects the solution as shown by fig. 5.2. To ensure that the algorithm does approximate the MINLP model we need to make sure that the algorithm does converge and that systematic errors are not created.

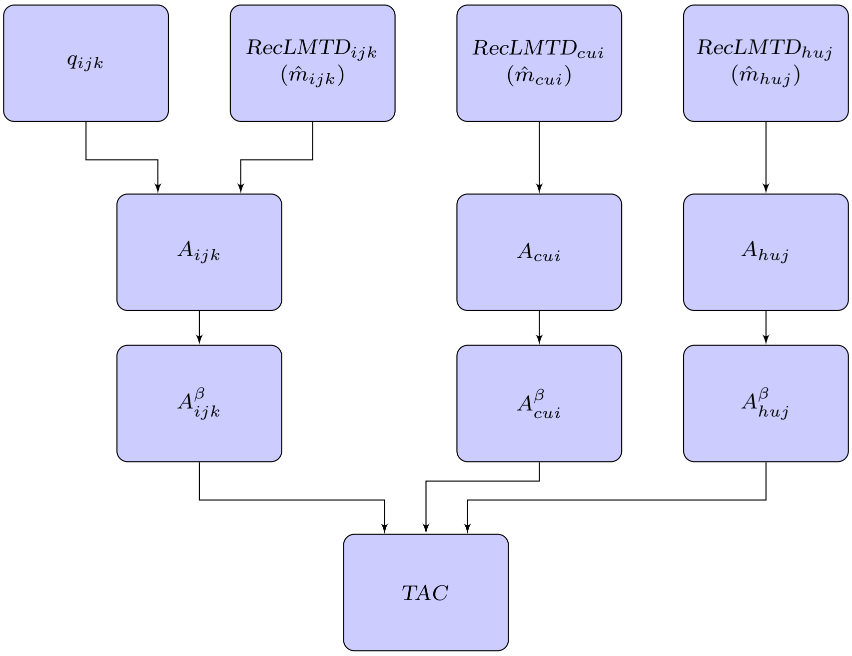


Figure 5.2: How the errors get cascaded into the objective function from the approximations made. At each non-*TAC* node, approximations are made using the nodes before them.

5.4.2 Stream to Stream Errors

The stream to stream errors are shown by the left column of fig. 5.2 (including the $RecLMTD_{ijk}$ node). We will analyse how these errors affect *TAC*.

We begin by assuming

$$\max\ m_{ijk} - \hat{m}_{ijk} \leq \varepsilon^{m}_{ijk}$$

The calculation of q_{ijk} involves four bilinear terms

$$q_{ijk} = f^{H}_{ijk}\left(t_{ik} - t^{H}_{ijk}\right) \tag{5.9}$$

$$q_{ijk} = f^{C}_{ijk}\left(t^{C}_{ijk} - t_{jk+1}\right) \tag{5.10}$$

which are replaced by auxiliary variables $b^{H,\text{in}}_{ijk}, b^{H,\text{out}}_{ijk}, b^{C,\text{out}}_{ijk}$ and $b^{C,\text{in}}_{ijk}$ as described by section 4.4.1. For the variables forming the bilinearities we have the following domains

$$\begin{aligned} t_{ik}, t^{H}_{ijk} &\in [T^{\text{out}}_{i}, T^{\text{in}}_{i}] \\ f^{H}_{ijk} &\in [0, F_i] \\ t^{C}_{ijk}, t_{jk} &\in [T^{\text{in}}_{j}, T^{\text{out}}_{j}] \\ f^{C}_{ijk} &\in [0, F_j] \end{aligned}$$

We are using the **nf4r** scheme to reduce the error in relaxation. We partition the temperature domains into $N_{q,ijk}$ pieces hence we have that the corresponding maximum errors are (the hot and cold bilinearities have the same bounds)

$$\varepsilon_{ijk}^{q(H)} = \frac{\left(T_i^{\text{in}} - T_i^{\text{out}}\right) F_i}{4N_{q,ijk}}$$
$$\varepsilon_{ijk}^{q(C)} = \frac{\left(T_j^{\text{out}} - T_j^{\text{in}}\right) F_j}{4N_{q,ijk}}$$

As q_{ijk} is calculated by taking the difference of the two bilinearities, by the triangle inequality, we define the upper bound of the error as

$$\varepsilon_{ijk}^{q} = \max\left\{ 2\varepsilon_{ijk}^{q(H)}, 2\varepsilon_{ijk}^{q(C)} \right\}$$

The definitions of ε_{ijk}^{m} and ε_{ijk}^{q} form the foundations to our analysis of the errors. The approximations $\hat{q}_{ijk}$ and $\hat{m}$ are used to calculate A_{ijk}.

We get

$$\hat{A}_{ijk} = \hat{q}_{ijk}\left(h_i^{-1} + h_j^{-1}\right)\hat{m}_{ijk} \tag{5.11}$$

where h_i and h_j are constants

This is another bilinearity therefore we linearise it in the same way as before, the difference this time is that we: do not have any sums and do have approximation errors in our variables. Before we derive an upper bound on the approximation error for $\hat{A}_{ijk}$ we need to classify the bounds of both the: approximations and variables we are approximating. We have that the bounds of the heat load and its approximation are the same and we can construct our outer approximation such that the bounds of the approximation also match that of the function (section 5.3.1). Therefore we have that

$$q_{ijk}^{U} = \min\left\{ F_i\left(T_i^{\text{in}} - T_i^{\text{out}}\right), F_j\left(T_j^{\text{out}} - T_j^{\text{in}}\right) \right\}$$
$$q_{ijk}, \hat{q}_{ijk} \in \left[0, q_{ijk}^{U}\right]$$
$$m_{ijk}, \hat{m}_{ijk} \in \left[{}^{1}\!/_{u_{ij}}, {}^{1}\!/_{l_{ij}}\right] = \left[m_{ijk}^{L}, m_{ijk}^{U}\right]$$

Using the triangle inequality we get

$$\hat{q}_{ijk} = \left|q_{ijk} + \hat{\varepsilon}_{ijk}^{q}\right| \le q_{ijk} + \left|\hat{\varepsilon}_{ijk}^{q}\right| \le q_{ijk} + \varepsilon_{ijk}^{q} \tag{5.12}$$
$$\hat{m}_{ijk} = \left|m_{ijk} - \hat{\varepsilon}_{ijk}^{m}\right| \le m_{ijk} + \hat{\varepsilon}_{ijk}^{m} \le m_{ijk} + \varepsilon_{ijk}^{m} \tag{5.13}$$

where $-\varepsilon_{ijk}^{q} \le \hat{\varepsilon}_{ijk}^{q} \le \varepsilon_{ijk}^{q}$ and $0 < \hat{\varepsilon}_{ijk}^{m} \le \varepsilon_{ijk}^{m}$.

We are also approximating eq. (5.11) using the **nf4r** scheme for which let $N_{A,ijk}$ be the number of pieces that we partition q_{ijk} on. From this approximation we get

$$\hat{A}_{ijk} = \hat{q}_{ijk} \cdot \hat{m}_{ijk} \pm \hat{\varepsilon}_{ijk}^{A}$$

for some

$$-\frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}} \leq \hat{\varepsilon}^A_{ijk} \leq \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}$$

hence by the triangle inequality we have

$$\begin{aligned}
\hat{A}_{ijk} &= \left|\left(h_i^{-1} + h_j^{-1}\right)\left[\hat{q}_{ijk} \cdot \hat{m}_{ijk} \pm \hat{\varepsilon}^A_{ijk}\right]\right| \\
&\leq \left(h_i^{-1} + h_j^{-1}\right)\left[\hat{q}_{ijk} \cdot \hat{m}_{ijk} + \left|\hat{\varepsilon}^A_{ijk}\right|\right] \\
&\leq \left(h_i^{-1} + h_j^{-1}\right)\left[\hat{q}_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right]
\end{aligned}$$

Using eqs. (5.12) and (5.13), we can represent $\hat{A}_{ijk}$ in terms of the correct values q_{ijk} and m_{ijk}

$$\begin{aligned}
\hat{A}_{ijk} &\leq \left(h_i^{-1} + h_j^{-1}\right)\left[\hat{q}_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \\
&\leq \left(h_i^{-1} + h_j^{-1}\right)\left[\left(q_{ijk} + \varepsilon^q_{ijk}\right) \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \\
&= \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk} \cdot \hat{m}_{ijk} + \varepsilon^q_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \\
&\leq \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk}\left(m_{ijk} + \varepsilon^m_{ijk}\right) + \varepsilon^q_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \\
&= \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk} \cdot m_{ijk} + q_{ijk} \cdot \varepsilon^m_{ijk} + \varepsilon^q_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \qquad (5.14)
\end{aligned}$$

But we also have

$$\begin{aligned}
\hat{A}_{ijk} &= A_{ijk} + \psi^{A,ijk} \\
A_{ijk} &= \left(h_i^{-1} + h_j^{-1}\right) q_{ijk} \cdot m_{ijk}
\end{aligned}$$

for some $\psi^{A,ijk}$

Substituting into eq. (5.14), we get

$$\begin{aligned}
A_{ijk} + \psi^{A,ijk} &\leq A_{ijk} + \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk} \cdot \varepsilon^m_{ijk} + \varepsilon^q_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \\
\iff \quad \psi^{A,ijk} &\leq \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk} \cdot \varepsilon^m_{ijk} + \varepsilon^q_{ijk} \cdot \hat{m}_{ijk} + \frac{q^U_{ijk}\left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}}\right] \qquad (5.15)
\end{aligned}$$

The way in which we calculated our bound on the error means that the negation of the right hand side of eq. (5.15) gives a lower bound on the error giving

$$\left|\psi^{A,ijk}\right| \leq \left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk} \cdot \varepsilon_{ijk}^{m} + \varepsilon_{ijk}^{q} \cdot \hat{m}_{ijk} + \frac{q_{ijk}^{U}\left({}^{1}/{}_{l_{ij}} - {}^{1}/{}_{u_{ij}}\right)}{4N_{A,ijk}}\right] \tag{5.16}$$

In eq. (5.16), we have that ε_{ijk}^{m} and ε_{ijk}^{q} contribute to the error. In the above formulation, the values of ε_{ijk}^{m} and ε_{ijk}^{q} are independent of the choice of $N_{A,ijk}$. Simply increasing $N_{A,ijk}$ does not necessarily make the error converge to 0, we make the following assumptions to make this the case.

$$\varepsilon_{q}^{ijk} \leq \frac{q_{ijk}^{U}}{N_{A,ijk}} \tag{5.17}$$

$$\varepsilon_{ijk}^{m} \leq \frac{m_{ijk}^{U}}{N_{A,ijk}} = \frac{1}{l_{ij} \cdot N_{A,ijk}} \tag{5.18}$$

The above simply mean that our previous approximations have been chosen such that the conditions specified by eqs. (5.17) and (5.18) hold, giving our bound on the error ε_{ijk}^{A}

$$\begin{aligned}
\varepsilon_{ijk}^{A} &= \left(h_i^{-1} + h_j^{-1}\right)\left[\frac{q_{ijk}^{U}\left({}^{1}/{}_{l_{ij}} - {}^{1}/{}_{u_{ij}}\right)}{4N_{A,ijk}} + \frac{q_{ijk}^{U}}{N_{A,ijk}} \cdot \frac{1}{l_{ij}} + q_{ijk}^{U} \cdot \frac{1}{l_{ij} \cdot N_{A,ijk}}\right] \\
&= \left(h_i^{-1} + h_j^{-1}\right)\left[\frac{q_{ijk}^{U}\left({}^{1}/{}_{l_{ij}} - {}^{1}/{}_{u_{ij}}\right)}{4N_{A,ijk}} + \frac{2q_{ijk}^{U}}{l_{ij} \cdot N_{A,ijk}}\right]
\end{aligned}$$

ε_{ijk}^{A} is then cascaded into the piecewise linear calculation of A_{ijk}^{β}.

We assume that we have partitioned the domain of A_{ijk}^{β} into $N_{\beta,ijk}$ equally sized intervals over which the maximum error between the function and the approximation is E_{ijk}. The bounds on A_{ijk} and $\hat{A}_{ijk}$ are

$$\begin{aligned}
A_{ijk}^{U} &= \left(h_i^{-1} + h_j^{-1}\right) \cdot q_{ijk}^{U} \cdot \frac{1}{l_{ij}} \\
A_{ijk}, \hat{A}_{ijk} &\in [0, A_{ijk}^{U}]
\end{aligned}$$

The bounds are the same because the values of the approximations are exact at the endpoints by construction. The same property holds for the range of the following piecewise linear approximation.

We make a similar assumption to those given by eqs. (5.17) and (5.18), as we want to limit the actual value of A_{ijk} to being in adjacent intervals to the interval defined by $\hat{A}_{ijk}$, the assumption is

$$\varepsilon_{ijk}^{A} \leq \frac{A_{ijk}^{U}}{N_{\beta,ijk}} \tag{5.19}$$

i.e. $N_{A,ijk}$ is large enough such that eq. (5.19) holds.

We have that the maximum error of the approximation is E_{ijk} but we want to know the error with respect to the actual value of A_{ijk}. We need to bound the error we encounter due to using $\hat{A}_{ijk}$.

To do this we define the set $R = \left\{ A \,\middle|\, 0 \leq A \leq A^U_{ijk} - {}^{A^U_{ijk}}\!/_{N_{\beta,ijk}} \right\}$. The maximum error we get due to the previous approximations is

$$E^A_{ijk} = \max_{A \in R} \left[\left(A + \frac{A^U_{ijk}}{N_{\beta,ijk}} \right)^{\beta} - A^{\beta} \right]$$

Hence, under these assumptions, the upper bound on the error that we get is

$$\varepsilon^{A^\beta}_{ijk} = E_{ijk} + E^A_{ijk}$$

therefore the maximum total error contributed to TAC by stream to stream heat exchangers is

$$\sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} E_{ijk} + E^A_{ijk}$$

5.4.3 Utility Errors

The utility errors are shown by the right two columns of fig. 5.2. We will only analyse the cold utility cascade of errors as the equivalent analysis for the hot utility is similar.

The difference between the stream to stream errors and the utility errors is that the heat load calculation for the utilities and a given stream is linear and can be calculated without the introduction of errors hence we only have the reciprocal of LMTD as the source of any errors.

As for the stream to stream case, we begin by assuming that there is a set of tangent constraints placed such that the maximum error between the outer approximation and the function is

$$\max \; m_{cui} - \hat{m}_{cui} \leq \varepsilon^m_{cui}$$

This approximation is then used to calculate $\hat{A}_{cui}$ forming the bilinearity

$$\hat{A}_{cui} = q_{cui} \left(h_i^{-1} + h_{CU}^{-1} \right) \hat{m}_{cui} \tag{5.20}$$

where h_i and h_{CU} are constants.

We still calculate this value using the **nf4r** scheme therefore to calculate the error we need to know the bounds of the variables

$$\begin{aligned} q^U_{cui} &= F_i \left(T^{\text{in}}_i - T^{\text{out}}_i \right) \\ m^L_{cui} &= m(u_{cui}, c_{cui}) \\ m^U_{cui} &= m(l_{cui}, c_{cui}) \end{aligned}$$

$$
\begin{aligned}
q_{cui} &\in \left[0, q_{cui}^U\right] \\
m_{cui}, \hat{m}_{cui} &\in \left[m_{cui}^L, m_{cui}^U\right]
\end{aligned}
$$

We can represent eq. (5.20) in terms of the actual value of m and the error giving

$$
\begin{aligned}
\hat{A}_{cui} &= \left(h_i^{-1} + h_{CU}^{-1}\right) q_{cui} \left(m_{cui} - \varepsilon_{cui}^m\right) \\
&= \left(h_i^{-1} + h_{CU}^{-1}\right) \left[q_{cui} \cdot m_{cui} - q_{cui} \cdot \varepsilon_{cui}^m\right]
\end{aligned}
$$

We assume that the domain of q_{cui} is partitioned into $N_{A,cui}$ pieces and using a similar reasoning as in the stream to stream case (section 5.4.2) we derive the following

$$
\hat{A}_{cui} \leq \left(h_i^{-1} + h_{CU}^{-1}\right) \left[q_{cui} \cdot m + q_{cui} \cdot \varepsilon_{cui}^m + \frac{q_{cui}\left(m_{cui}^U - m_{cui}^L\right)}{4N_{A,cui}}\right] \tag{5.21}
$$

But

$$
\begin{aligned}
\hat{A}_{cui} &= A_{cui} + \psi_{cui}^A \\
A_{cui} &= \left(h_i^{-1} + h_{CU}^{-1}\right) q_{cui} \cdot m_{cui}
\end{aligned}
$$

for some ψ_{cui}^A

Substituting into eq. (5.21), we get

$$
\begin{aligned}
A_{cui} + \psi_{cui}^A &\leq A_{cui} + \left(h_i^{-1} + h_{CU}^{-1}\right) \left[q_{cui} \cdot \varepsilon_{cui}^m + \frac{q_{cui}\left(m_{cui}^U - m_{cui}^L\right)}{4N_{A,cui}}\right] \\
\iff \qquad \psi_{cui}^A &\leq \left(h_i^{-1} + h_{CU}^{-1}\right) \left[q_{cui} \cdot \varepsilon_{cui}^m + \frac{q_{cui}\left(m_{cui}^U - m_{cui}^L\right)}{4N_{A,cui}}\right]
\end{aligned}
$$

Once again, the way in which our bound is calculated means that the bound we derived on the error is a bound on the magnitude hence

$$
\left|\psi_{cui}^A\right| \leq \left(h_i^{-1} + h_{CU}^{-1}\right) \left[q_{cui} \cdot \varepsilon_{cui}^m + \frac{q_{cui}\left(m_{cui}^U - m_{cui}^L\right)}{4N_{A,cui}}\right] \tag{5.22}
$$

In eq. (5.22), we have that ε_{cui}^m contributes to the error. In the above formulation, the value of ε_{cui}^m is independent of the choice of N_A. Simply increasing $N_{A,cui}$ does not necessarily make the error converge to 0, we make the following assumption to make this the case.

$$
\varepsilon_{cui}^m \leq \frac{m_{cui}^U}{N_{A,cui}} \tag{5.23}
$$

The above simply means that our previous approximation has been chosen such that the condition specified by eq. (5.23) holds, giving our bound on the error ε^A_{cui}

$$\begin{aligned}\varepsilon^A_{cui} &= \left(h_i^{-1} + h_j^{-1}\right)\left[q^U_{cui} \cdot \frac{m^U_{cui}}{N_{A,cui}} + \frac{q^U_{cui}\left(m^U_{cui} - m^L_{cui}\right)}{4N_{A,cui}}\right] \\ &= \left(h_i^{-1} + h_j^{-1}\right)\left[\frac{q^U_{cui} \cdot m^U_{cui}}{N_{A,cui}} + \frac{q^U_{cui}\left(m^U_{cui} - m^L_{cui}\right)}{4N_{A,cui}}\right]\end{aligned}$$

This error is cascaded into the piecewise linear calculation of A^β_{cui}.

We assume that the domain of A^β_{cui} has been partitioned into $N_{\beta,cui}$ equally sized intervals over which the maximum error between the function and the approximation is E_{cui}. The domain of A^β_{cui} and $\hat{A}^\beta_{cui}$ is

$$\begin{aligned}A^U_{cui} &= \left(h_i^{-1} + h_{CU}^{-1}\right) \cdot q^U_{cui} \cdot m(l_{cui}, c_{cui}) \\ A_{cui}, \hat{A}_{cui} &\in \left[0, A^U_{cui}\right]\end{aligned}$$

As we did for eq. (5.19), we place a constraint on the size of the error ε^A_{cui} to limit the actual value of A_{cui} to being in adjacent intervals to the interval defined by $\hat{A}_{cui}$, the assumption is

$$\varepsilon^A_{cui} \leq \frac{A^U_{cui}}{N_{\beta,cui}} \tag{5.24}$$

i.e. $N_{A,cui}$ is large enough such that eq. (5.24) holds.

We define the set $R = \left\{ A \;\middle|\; 0 \leq A \leq A^U_{cui} - {A^U_{cui}}/{N_{\beta,cui}} \right\}$ The maximum error we get due to the previous approximations is

$$E^A_{cui} = \max_{A \in R} \left[\left(A + \frac{A^U_{cui}}{N_{\beta,cui}}\right)^\beta - A^\beta\right]$$

Hence, under these assumptions, the upper bound on the error that we get is

$$\varepsilon^{A^\beta}_{cui} = E_{cui} + E^A_{cui}$$

therefore the maximum total error contributed to TAC by cold utility heat exchangers is

$$\sum_{i \in HP} E_{cui} + E^A_{cui}$$

Similarly, for the hot utility, we get the following results/assumptions

- the assumption on the reciprocal of the log mean temperature difference

$$\max\ m_{huj} - \hat{m}_{huj} \leq \varepsilon^m_{huj}$$

- the assumption on the bound of ε^m_{huj}

$$\varepsilon^m_{huj} \leq \frac{m^U_{huj}}{N_{A,huj}} \tag{5.25}$$

- the upper bound on the error of $\hat{A}_{huj}$

$$\varepsilon^A_{huj} = \left(h_i^{-1} + h_j^{-1}\right)\left[\frac{q^U_{huj} \cdot m^U_{huj}}{N_{A,huj}} + \frac{q^U_{huj}\left(m^U_{huj} - m^L_{huj}\right)}{4N_{A,huj}}\right]$$

- the assumption of $\hat{A}_{huj}$ being bounded by the interval length

$$\varepsilon^A_{huj} \leq \frac{A^U_{huj}}{N_{\beta,huj}}$$

- the upper bound on the error of $\hat{A}^\beta_{huj}$

$$\varepsilon^{A^\beta}_{huj} = E_{huj} + E^A_{huj}$$

- the upper bound on the total error contributed to TAC

$$\sum_{j \in CP} E_{huj} + E^A_{huj}$$

Hence for the maximum error in m

5.4.4 A Converging Model

We will show that we can construct a MILP model that converges to a MINLP solution given some allowance for error. Using this property we will establish convergence to the global solution.

What we showed in sections 5.4.2 and 5.4.3 is that if we have constructed a model, with some additional assumptions, how the errors affect the errors in the objective solution. We will now do the opposite and show that if we are given some convergence criteria that we can construct a model that will converge as required.

Assumptions:

1. we have a MILP solver that always gives us the global solution of any MILP model we provide
2. we are given ξ_{TAC} the maximum error allowed in the objective

We will use the results from sections 5.4.2 and 5.4.3. We made assumptions to derive these results and we will show that if these assumptions are satisfied, we can construct a model that converges to the global MINLP solution.

The maximum error in TAC is given by

$$Error_{TAC} = \sum_{i\in HP}\sum_{j\in CP}\sum_{k\in ST} E_{ijk} + E^A_{ijk} + \sum_{i\in HP} E_{cui} + E^A_{cui} + \sum_{j\in CP} E_{huj} + E^A_{huj} \tag{5.26}$$

We want

$$\xi_{TAC} \geq Error_{TAC} \tag{5.27}$$

We will create a convex combination of the term ξ_{TAC} and assign it to the group of summations in eq. (5.26).

Let

$$\lambda_{stream} = \frac{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}$$
$$\lambda_{CU} = \frac{N_{\text{Hot}}}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}$$
$$\lambda_{HU} = \frac{N_{\text{Cold}}}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}$$

Clearly, $\lambda_{stream} + \lambda_{CU} + \lambda_{HU} = 1$ therefore we have that the following agree with eq. (5.27)

$$\sum_{i\in HP}\sum_{j\in CP}\sum_{k\in ST} E_{ijk} + E^A_{ijk} \leq \lambda_{stream} \cdot \xi_{TAC} \tag{5.28}$$
$$\sum_{i\in HP} E_{cui} + E^A_{cui} \leq \lambda_{CU} \cdot \xi_{TAC} \tag{5.29}$$
$$\sum_{j\in CP} E_{huj} + E^A_{huj} \leq \lambda_{HU} \cdot \xi_{TAC} \tag{5.30}$$

We can satisfy the conditions specified in eqs. (5.28) to (5.30) using $N_{\beta,ijk}$, $N_{\beta,cui}$ and $N_{\beta,huj}$.

If we ignore the indices we have that as N_β increases $E + E^A$ decreases therefore we can choose values $N^*_{\beta,ijk}$, $N^*_{\beta,cui}$ and $N^*_{\beta,huj}$ such that eqs. (5.28) to (5.30) hold. We can simplify this by defining

$$N'_\beta = \max\left\{ \max_{(i,j,k)^T\in HX} N^*_{\beta,ijk}, \max_{i\in HP} N^*_{\beta,cui}, \max_{j\in CP} N^*_{\beta,huj} \right\}$$

Under the partitioning of N'_β pieces across all 'β' approximations, we define

$$E_{\min} = \min\left\{ \min_{(i,j,k)^T\in HX} E_{ijk}, \min_{i\in HP} E_{cui}, \min_{j\in CP} E_{huj} \right\}$$
$$E^A_{\min} = \min\left\{ \min_{(i,j,k)^T\in HX} E^A_{ijk}, \min_{i\in HP} E^A_{cui}, \min_{j\in CP} E^A_{huj} \right\}$$

and let N^*_β be a partitioning value such that

$$E^A_{\min} \geq E^A_{ijk}, \qquad \forall i \in HP, j \in CP, k \in ST$$

$$
\begin{aligned}
&E^A_{\min} \geq E^A_{cui}, && \forall i \in HP \\
&E^A_{\min} \geq E^A_{huj}, && \forall j \in CP \\
&E_{\min} \geq E_{ijk}, && \forall i \in HP, j \in CP, k \in ST \\
&E_{\min} \geq E_{cui}, && \forall i \in HP \\
&E_{\min} \geq E_{huj}, && \forall j \in CP
\end{aligned}
$$

then eqs. (5.28) to (5.30) reduce to

$$
\begin{aligned}
N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T \cdot \left(E_{\min} + E^A_{\min}\right) &\leq \lambda_{stream} \cdot \xi_{TAC} \\
N_{\text{Hot}} \cdot \left(E_{\min} + E^A_{\min}\right) &\leq \lambda_{CU} \cdot \xi_{TAC} \\
N_{\text{Cold}} \cdot \left(E_{\min} + E^A_{\min}\right) &\leq \lambda_{HU} \cdot \xi_{TAC}
\end{aligned}
$$

and from the definition of λ_{stream}, λ_{CU} and λ_{HU} we get

$$
E_{\min} + E^A_{\min} \leq \frac{\xi_{TAC}}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}
$$

Note that as N^*_β increases the interval size decreases therefore by the continuity and concavity of A^β we get that $E_{\min} \to 0$ and $E^A_{\min} \to 0$. Also $N^*_\beta \geq N'_\beta$.

We will now analyse how this selection of N^*_β affects the other approximations. We will look at the utility heat exchangers first. We defined the upper bound on the error allowed for the area approximations as:

$$
\begin{aligned}
\varepsilon^A_{ijk} &\leq \frac{A^U_{ijk}}{N_{\beta,ijk}} \\
\varepsilon^A_{cui} &\leq \frac{A^U_{cui}}{N^*_\beta} \\
\varepsilon^A_{huj} &\leq \frac{A^U_{huj}}{N^*_\beta}
\end{aligned}
$$

We want the upper bound to be equal across the utilities so that we can define a single constant for all of the approximations therefore we take

$$
\varepsilon^{A^*} = \frac{\min\left\{ \min\limits_{(i,j,k)^T \in HX} A^U_{ijk}, \min\limits_{i \in HP} A^U_{cui}, \min\limits_{j \in CP} A^U_{huj} \right\}}{N^*_\beta}
$$

We now see how this bound affects the calculation of the utility areas, we will look at a single cold utility heat exchanger, the reasoning is similar for a hot utility heat exchanger.

We have for some cold utility heat exchanger cui, assuming the condition eq. (5.23)

$$
\varepsilon^{A^*} = \left(h_i^{-1} + h_j^{-1}\right) \left[\frac{q^U_{cui} \cdot m^U_{cui}}{N_{A,cui}} + \frac{q^U_{cui}\left(m^U_{cui} - m^L_{cui}\right)}{4N_{A,cui}} \right]
$$

this means that (since $h_i, h_{CU} > 0$)

$$\frac{\varepsilon^{A^*}}{\left(h_i^{-1} + h_{CU}^{-1}\right)} = \frac{4q_{cui}^U \cdot m_{cui}^U + q_{cui}^U \left(m_{cui}^U - m_{cui}^L\right)}{4N_{A,cui}}$$

$$\iff \qquad N_{A,cui} = \frac{\left(h_i^{-1} + h_{CU}^{-1}\right)\left[4q_{cui}^U \cdot m_{cui}^U + q_{cui}^U \left(m_{cui}^U - m_{cui}^L\right)\right]}{4\varepsilon^{A^*}} \tag{5.31}$$

Similarly, for a hot utility exchanger we get

$$N_{A,huj} = \frac{\left(h_j^{-1} + h_{HU}^{-1}\right)\left[4q_{huj}^U \cdot m_{huj}^U + q_{huj}^U \left(m_{huj}^U - m_{huj}^L\right)\right]}{4\varepsilon^{A^*}} \tag{5.32}$$

We now analyse the stream to stream case. For some stream to stream heat exchanger ijk, assuming conditions eqs. (5.17) and (5.18) we have

$$\varepsilon^{A^*} = \left(h_i^{-1} + h_j^{-1}\right)\left[\frac{q_{ijk}^U \left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right)}{4N_{A,ijk}} + \frac{2q_{ijk}^U}{l_{ij} \cdot N_{A,ijk}}\right]$$

this means (since $h_i, h_j^{-1} > 0$)

$$\frac{\varepsilon^{A^*}}{\left(h_i^{-1} + h_j^{-1}\right)} = \frac{q_{ijk}^U \left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right) + 8 \cdot {}^{q_{ijk}^U}/_{l_{ij}}}{4N_{A,ijk}} \tag{5.33}$$

$$\iff \qquad N_{A,ijk} = \frac{\left(h_i^{-1} + h_j^{-1}\right)\left[q_{ijk}^U \left({}^1/_{l_{ij}} - {}^1/_{u_{ij}}\right) + 8 \cdot {}^{q_{ijk}^U}/_{l_{ij}}\right]}{4\varepsilon^{A^*}} \tag{5.34}$$

We want to pick a single value N_A^* to use across all approximations hence we have

$$N_A^* \geq \max\left\{\max_{(i,j,k)^T \in HX} N_{A,ijk}, \max_{i \in HP} N_{A,cui}, \max_{j \in CP} N_{A,huj}\right\}$$

Now we simply construct the initial approximations according to eqs. (5.17), (5.18), (5.23) and (5.25).

Equations (5.18), (5.23) and (5.25) correspond to the construction of the outer approximation for m, we have that our upper bounds on the errors are (with N_A^* substituted)

$$\varepsilon_{ijk}^m = \frac{1}{l_{ij} \cdot N_A^*}, \qquad \forall\, (i,j,k)^T \in HX$$

$$\varepsilon_{cui}^m = \frac{m_{cui}^U}{N_A^*}, \qquad \forall\, i \in HP$$

$$\varepsilon_{huj}^m = \frac{m_{huj}^U}{N_A^*}, \qquad \forall\, j \in CP$$

therefore a single upper bound across all approximations would have to satisfy

$$\varepsilon^{m^*} \leq \min\left\{ \min_{(i,j,k)^T \in HX} \left[\frac{1}{l_{ij} \cdot N_A^*}\right], \min_{i \in HP} \left[\frac{m_{cui}^U}{N_A^*}\right], \min_{j \in CP} \left[\frac{m_{huj}^U}{N_A^*}\right] \right\}$$

Equation (5.17) corresponds with the construction of the McCormick hulls in the heat load calculations, we have that the upper bound on the error is (with N_A^* substituted)

$$\varepsilon_q^{ijk} = \frac{q_{ijk}^U}{N_{A,ijk}}, \qquad \forall (i,j,k)^T \in HX$$

using this we can calculate the values $N_{q,ijk}$. Since the sum of two approximations are taken when we calculate $\hat{q}$, we halve the error. The definition of $\hat{q}$ also involves making two approximations equal hence we have to take this into account. This all gives

$$\frac{\varepsilon_q^{ijk}}{2} = \frac{\left(T_i^{\text{in}} - T_i^{\text{out}}\right) F_i}{4N_{q,ijk}^{(H)}}$$

$$\frac{\varepsilon_q^{ijk}}{2} = \frac{\left(T_j^{\text{out}} - T_j^{\text{in}}\right) F_j}{4N_{q,ijk}^{(C)}}$$

The above result in

$$N_{q,ijk}^{(H)} = \frac{\left(T_i^{\text{in}} - T_i^{\text{out}}\right) F_i}{2\varepsilon_q^{ijk}}$$

$$N_{q,ijk}^{(C)} = \frac{\left(T_j^{\text{out}} - T_j^{\text{in}}\right) F_j}{2\varepsilon_q^{ijk}}$$

giving an lower bound the number of pieces for exchanger ijk

$$N_{q,ijk}^* = \max\left\{ N_{q,ijk}^{(H)}, N_{q,ijk}^{(C)} \right\}$$

therefore a single value to use for all approximations is

$$N_q^* = \max_{(i,j,k)^T \in HX} N_{q,ijk}^*$$

We have shown that given $\xi_{TAC} > 0$ we can construct a model that results in a solution that is within ξ_{TAC} of the correct value. The model we constructed has the property that as $\xi_{TAC} \to 0$ all of the errors reduce to 0 showing that we don't necessarily have to provide any other bounds, ξ_{TAC} is sufficient for convergence.

We also assumed that we have a MILP solver that gives us the global solution of any MILP that we provide. As $\xi_{TAC} \to 0$ our approximation gets better therefore the MILP formulation get closer to the MINLP formulation. As all errors are also getting smaller, the global solution of the MILP formulation converges to the global solution of the MINLP formulation.

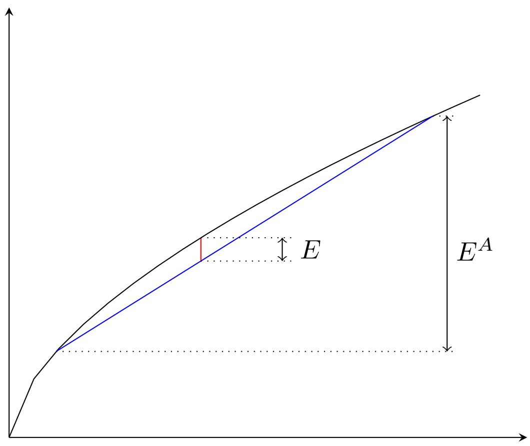


Figure 5.3: How the value of E compares with E^A in the concave function

5.4.5 Convergence of the Algorithm

Convergence of the algorithm is a direct consequence of section 5.4.4. We treat all approximations except the outer approximations as fixed, therefore to guarantee convergence within some absolute error tolerence we would have to provide N_q^*, N_A^* and N_β^*. We have shown how we can construct these constants hence we can provide suitable values for these constants. The values of N_q^* and N_A^* depend on N_β^*, we can define a value for N_β^* as follows.

We have that $E_{\text{min}} \leq E_{\text{min}}^A$ this follows from the fact that A^β is strictly increasing therefore the error between the function and approximation is bounded by the difference between the values at the end points as shown by fig. 5.3. Therefore the taking

$$E_{\text{min}}^A = \frac{1}{2} \cdot \frac{\xi_{TAC}}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}$$

gives the required condition. The structure of the function is such that the following satisfies the above

$$\left(\frac{A^U}{N_\beta}\right)^\beta = \frac{1}{2} \cdot \frac{\xi_{TAC}}{N_{\text{Hot}} \cdot N_{\text{Cold}} \cdot N_T + N_{\text{Hot}} + N_{\text{Cold}}}$$

since E^A is maximised at the lowest point and we know that the function is bounded from below by 0.

All that remains is the tangent points. In the worst case the algorithm will add tangents in different neighbourhoods but this reduces the overall error and therefore the maximum error will eventually reduce.

The values N_q^*, N_A^* and N_β^* may prove to be large enough such that the error in the outer approximation is larger than what we defined previously but this is not an issue as the absolute error does make the error go to zero. As N_q^*, N_A^* and N_β^* increase the error contributed by the outer approximation will become more significant hence more tangents will be required i.e. the error decresases. Assuming we have a solver that gives us the global solution of the MILPs at each iteration our solution will be within the required error tolerance and therefore converging to the global solution of the MINLP.

6 Evaluation

6.1 Overview

Having created an implementation of the algorithm, we now evaluate the results we achieve.

We used the solver Gurobi (version 6.0.3) to test our algorithm.

The specifications of machines on which we have run our evaluation tests are:

- Model: HP EliteDesk 800 G1 TWR
- CPU: Intel Core i7-4770 3.40GHz
- Memory: 16GB RAM
- OS: Linux Ubuntu 14.04.2 LTS, 64-bit

We have 3 models [12] that we use to evaluate the algorithm the input data for each are shown in table 6.2. Across all models, the minimum temperature approach for a given heat exchanger, ΔT_{min}, is equal to 10. Note that model 1 (table 6.2a) has $\beta = 1$ therefore the objective function is linear.

We compare against these models as feasible solutions are known therefore we gain a metric to compare our algorithm against.

Model	Best Solution	Lower Bound (Solver)
Model 1[4]	154895.93	100501.00 (LINDO)
Model 2[5]	634200.38	585255.58 (BARON)
Model 3[6]	64898.22	64898.22 (SCIP)

Table 6.1: The best solutions and lower bounds for the models as found on MINLP World

6.2 Numerical Analysis

We will evaluate our algorithm numerically, assessing the time taken to find a solution and the number of iterations required to find this solution.

As the algorithm has a priori partitioning, we have to define the number of breakpoints and partitions we make before we start the algorithm and the separation that they create will remain the same (section 5.3). We ran tests with a constant number, N, of breakpoints and partitions i.e. $N = N_A = N_\beta = N_q$.

We change the value of N and see how this affects the behaviour of our algorithm. The algorithm was run with a maximum absolute error tolerance of $\epsilon = 10^{-6}$

[4] `http://www.gamsworld.org/minlp/minlplib2/html/heatexch_gen1.html`
[5] `http://www.gamsworld.org/minlp/minlplib2/html/heatexch_gen2.html`
[6] `http://www.gamsworld.org/minlp/minlplib2/html/heatexch_gen3.html`

Stream	T^{in} (K)	T^{out} (K)	F $(kW \cdot K^{-1})$	h $(kW \cdot m^2 \cdot K^{-1})$
H1	650	370	10	1
H2	590	370	20	1
C1	410	650	15	1
C2	350	500	13	1
CU	300	320	-	1
HU	680	680	-	5

Cost of Heat Exchangers $(\$ \cdot yr^{-1}) = 5500 + 150[\text{Area } (m^2)]^1$
Cost of Cooling Utility $(\$ \cdot kW^{-1} \cdot yr^{-1}) = 15$
Cost of Heating Utility $(\$ \cdot kW^{-1} \cdot yr^{-1}) = 80$

(a) Model 1 (Escobar and Grossmann)

Stream	T^{in} (K)	T^{out} (K)	F $(kW \cdot K^{-1})$	h $(kW \cdot m^2 \cdot K^{-1})$
H1	500	320	6	2
H2	480	380	4	2
H3	460	360	6	2
H4	380	360	20	2
H5	380	320	12	2
C1	290	660	18	2
CU	300	320	-	1
HU	700	700	-	2

Cost of Heat Exchangers $(\$ \cdot yr^{-1}) = 5500 + 1200[\text{Area } (m^2)]^{0.6}$
Cost of Cooling Utility $(\$ \cdot kW^{-1} \cdot yr^{-1}) = 10$
Cost of Heating Utility $(\$ \cdot kW^{-1} \cdot yr^{-1}) = 140$

(b) Model 2 (Escobar and Grossmann)

Table 6.2: The input data for the models we ran

Stream	T^{in} (K)	T^{out} (K)	F $(kW \cdot K^{-1})$	h $(kW \cdot m^2 \cdot K^{-1})$
H1	160	93.3	8.8	1.7
H2	248.9	137.8	10.6	1.7
H3	226.7	65.6	14.8	1.7
H4	271.1	148.9	12.6	1.7
H5	198.9	65.6	17.7	1.7
C1	60	160	7.6	1.7
C2	115.6	221.7	6.1	1.7
C3	37.8	211.1	8.4	1.7
C4	82.2	176.7	17.3	1.7
C5	93.3	204.4	13.9	1.7
CU	25	40	-	1.7
HU	240	240	-	3.4

Cost of Heat Exchangers ($\$ \cdot yr^{-1}$) $= 4000 + 146[\text{Area } (m^2)]^{0.6}$
Cost of Cooling Utility ($\$ \cdot kW^{-1} \cdot yr^{-1}$) $= 10$
Cost of Heating Utility ($\$ \cdot kW^{-1} \cdot yr^{-1}$) $= 200$

(c) Model 3 (Escobar and Grossmann)

Table 6.2: The input data for the models we ran

6.2.1 Scalability

From table 6.3, we get an idea of how termination time scales with problem size. For $N \in \{ 10, 20, ..., 50 \}$ model 1 finishes in under 3 minutes (180s) whereas model 3 takes over an hour (3600s) to terminate for $N = 10$. This is due to the search tree growing exponentially as we add more streams.

The problems we test our algorithm against are small. Model 3 has a total of 10 streams but in industrial processes larger problems occur e.g. an ethylene product recovery process which can have 41 streams [4]. Clearly scalability is a necessity if we want to handle 'real world' problems however table 6.3 shows that this causes issues in solution finding time. We also want our algorithm to be within close proximity to the global solution but for this to occur we have to be willing to increase the number of binary variables we use for our approximations but, as table 6.3 shows, increasing the number of binary variables also has an adverse affect on termination time.

We find that, as it stands, the algorithm fails to scale in the directions that we want it to scale. The major issue here is the binary variables. For scaling of the problem size, there is not a lot that we can do about the extra variables as they are a consequence of the model however for the approximations we need a better approach to placing our breakpoints or we need to improve the non-LMTD approximations with each iteration opposed to partitioning before hand.

Model	N	Solution	# iterations	Time Taken (s)
	1	120669.59	5	0.15
	10	150340.41	6	2.09
Model 1	20	152846.01	7	12.24
	30	153845.10	6	26.89
	40	154057.37	7	76.07
	50	154269.90	10	148.24
	1	604958.96	4	0.16
	10	634383.18	2	5.36
Model 2	20	634598.67	2	22.16
	30	634766.08	2	57.55
	40	634705.76	2	49.9
	50	634727.03	2	207.99
	1	58930.17	5	11.29
	10	63847.37	5	3806.93
Model 3	20	64097.60	8	11618.66
	30	64106.66	7	19536.24
	40	64121.11	6	19882.15
	50	64116.25	6	32223.98

Table 6.3: The results of running each model as we vary N

6.2.2 Rate of Convergence

The rate of convergence is given by the number of iterations we need before the algorithm terminates. The results are shown in fig. 6.1. As we can see the rate of convergence appears to be quite quick with largest number of iterations being 10 in model 1 (fig. 6.1a).

We assess rate of convergence by the number of iterations and the change in objective value between consecutive iterations.

Firstly, we notice that the shapes of the results for varying N are similar for a given model, this shows that algorithm follows a similar pattern in placement of tangents as we increase the value of N.

For model 2 (fig. 6.1b), we see that the algorithm terminates in two iterations however this is not reflective of the algorithm's ability to converge, we have that the data for model 2 is such that there is a clear solution therefore even with a small N we get a jump to termination. Assessing the rate of convergence for models 1 and 3 give a more informative result.

For model 1 (fig. 6.1a), the algorithm converges in at most 10 iterations (for $N = 50$) showing that the rate of convergence is quick. Another noticable feature for model 1 is that there is a steep rise in the objective value after the first iteration and thereafter the change in objective value between iterations is small, relative to the initial change. We see that, while the algorithm terminates in at most 10 iterations, after iteration 5 the graphs appear to be flat i.e. the objective value isn't changing by much.

For model 3 (fig. 6.1c), the algorithm converges in at most 8 iterations (for $N = 30$). Like model 1, we see similarities in the shape of the graphs for varying N however we don't get an immediate steep rise in the objective value, the steepest rise we see occurs between iteration 2 and 3 and thereafter the algorithm converges in a similar pattern to that of model 1. Once again we see that the increase in N doesn't have an adverse effect on the path followed by algorithm other than a slight upwards shift in the plot. A noticable feature of model 3 is that for $N = 40$ and $N = 50$ the objective values at each iteration are more or less identical indicating that an increase in N would result in a similar plot.

The rate of convergence we see in these models give promising results however we must take into account the time taken to achieve these solutions as the practical applicability of this algorithm depends on achieving a solution in a 'reasonable' amount of time. We have that model 3 is the closest to a practical application due to its size. For $N = 40$ the time taken to terminate was 19882.15s which is about 5.5 hours and for $N = 50$ the time taken to terminate was 32223.98s which is almost 9 hours. In the algorithm, we partition on the heat load variables q which can have a large range e.g. $[0, 1634]$. These variables are also estimated therefore there are already errors in the model we then approximate q hence to achieve a low overall error we want to use a larger value for N but for model 3 this will only scale the problem up and we may find that the N we want to use will not terminate within what we choose to define as a 'reasonable' amount of time.

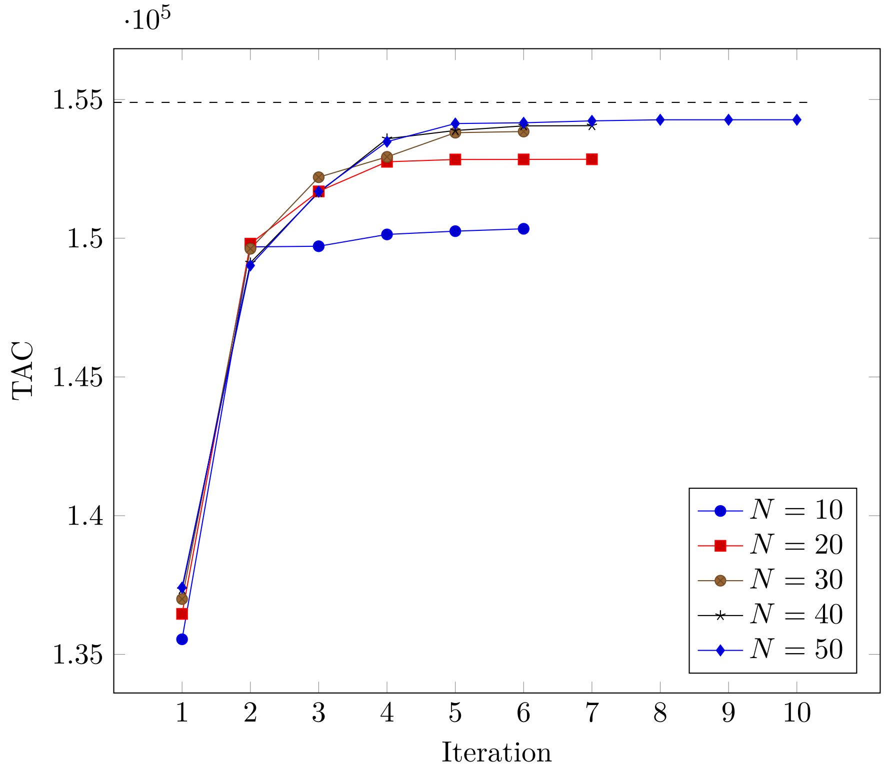
6.2.3 Bounds

A defining features of our algorithm is that it is an outer approximation to the entire model. The solutions that we get from our algorithm are likely to be infeasible however we know that the global solution will always be larger than or equal to any solution we find (assuming that the MILP solver we use is giving us the global solution of the outer approximation we form). The largest lower bounds achieved by MINLP solvers is given in table 6.1.

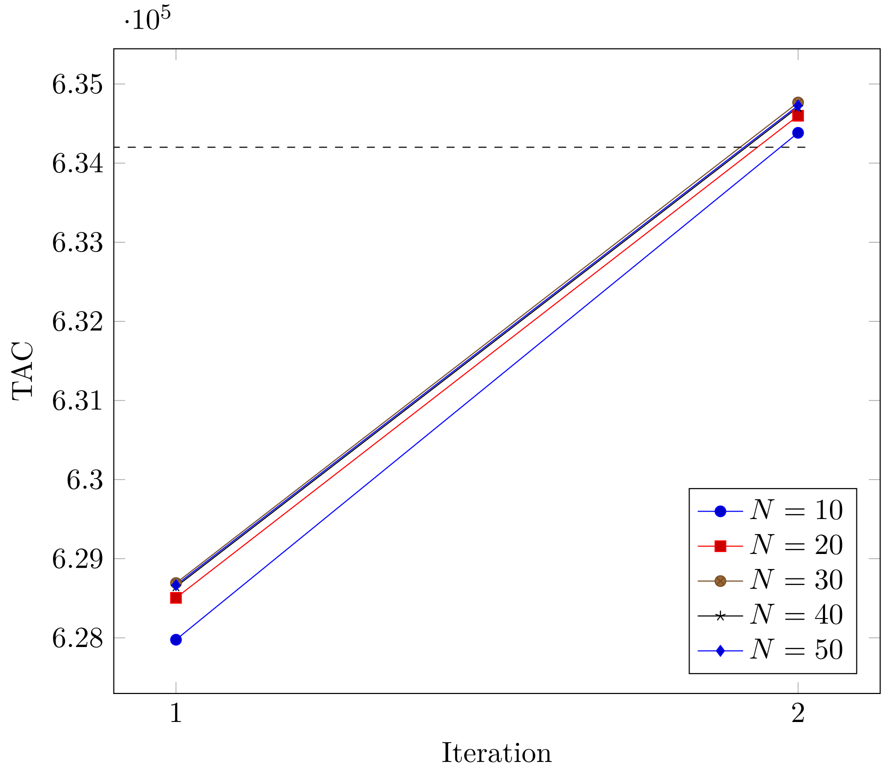
We have that the lower bound achieved by MINLP solvers for model 3 is equal to the global solution therefore we cannot do better than this. We turn out attention to the lower bounds of models 1 and 2. We can see that for both of these models, the solvers have a large gap between the lower bound and the solution found, this is an issue because we don't know whether there is a better solution between these values.

The weakest model we can form for our algorithm is one where $N = 1$ i.e. we don't partition anywhere (weak approximation). The results show that our approximated solution for $N = 1$ for models 1 and 2 are larger than the lower bounds given in Table 6.1. This increase comes from our handling of LMTD; since we are exposing its convex structure to form cuts, we get a progressive rise in our approximated solution. This makes our algorithm, at the very least, a means by which we can form bounds on the model we are trying to solve. For example, model 1 has a best solution of 154895.93 and our algorithm appears to converge to that value as shown by fig. 6.1a.

For model 2, our approximated solutions are larger than the best found solutions contradicting the statement that any solution we find results in a lower bound. The solution we achieve is always greater for the case of model 2 however it is not larger by an extreme amount. We put this down to numerical error as we see that our solution is

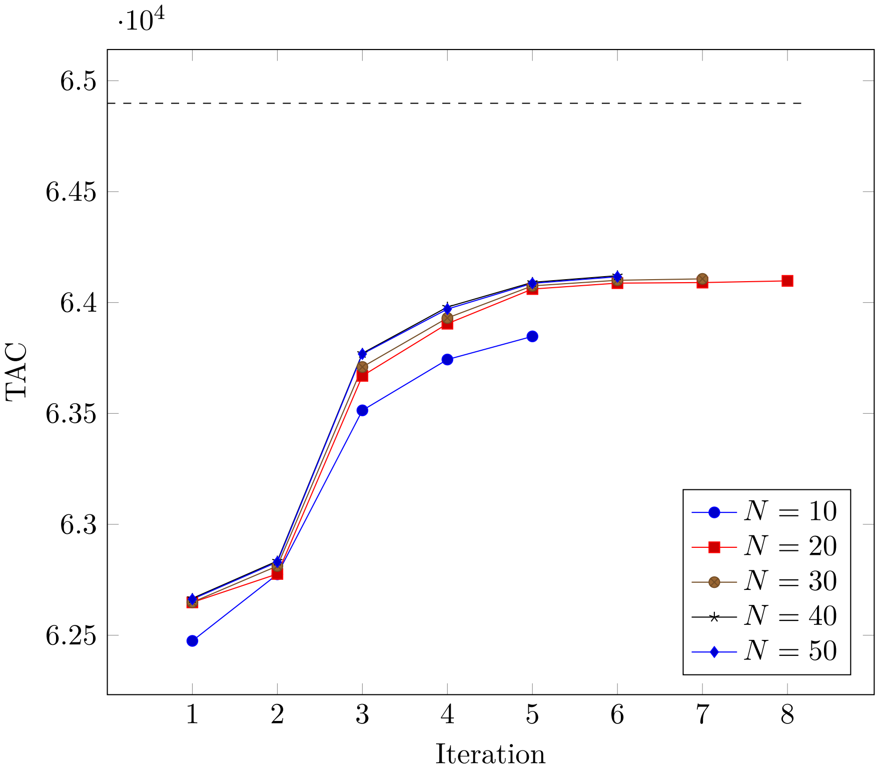


(a) Model 1



(b) Model 2

Figure 6.1: The objective solutions we get plotted against the iteration for $N \in \{10, 20, 30, 40, 50\}$ the dashed line is the best known solution.



(c) Model 3

Figure 6.1: The objective solutions we get plotted against the iteration for $N \in \{10, 20, 30, 40, 50\}$ the dashed line is the best known solution.

not increasing by a lot between $N = 40$ and $N = 50$ and the solution we achieve is very close to the global solution.

6.2.4 Convergence Issues

We see, from fig. 6.1c, that our algorithm is converging to a solution that is not equal to the global solution (table 6.1) but the global solution is equal to the MINLP lower bound for model 3.

We get this error due to the approximations we are making. The fixed approximations in the algorithm mean we converge to the global solution of the approximated model. Our fixed approximations partition the domains of our functions equally, this does not reflect the nonlinear nature of some of the functions we approximate. Our partitioning for the concave function in the objective is poor. A better approach to partitioning this function would be to place more break points towards lower values as the function increases more here.

6.3 Heuristic Approach

Clearly the time taken is an issue in the algorithm. One way of handling this is to improve our placement of breakpoints or dynamically generate them. Either way we are adding binary variables to the model and it is likely that we will still have a large number of binary variables added to the model.

Here we see if we can converge to a solution quicker, not necessarily the global optimum, using an appropriate heuristic. This approach is highly model specific however if we can identify speed ups we have the potential to scale the model up for certain models.

The heuristic we use is adding the following constraint to the model:

$$\sum_{k \in ST} z_{ijk} \leq 1, \qquad\qquad i \in HP,\ j \in CP$$

the above constraint says that for any hot stream i and cold stream j there can only ever be a single heat exchanger in the model. This is a heuristic based on the idea that there will be at most one heat exchanger in the model for a pair of streams, we know this holds for the global solutions given in table 6.1.

From table 6.4, we see that the addition of this constraint gives a significant improvement in termination time for models 1 and 2 with model 1 showing a 71% decrease in termination time and model 2 showing a 82% decrease in termination time for $N = 50$. These models were terminating quickly without the additional constraint, our interests lie with model 3. For model 3, the results don't show any significant improvement with a increase in termination time for $N = 50$. We do see some improvement for other values of N however they do not show improvement to the same scale as that of models 1 and 2. This indicates that there may be larger models than models 1 and 2 for which this additional constraint improves the termination time.

The use of this heuristic is not unreasonable as it can have applications to other models for example it is more likely to hold in a model where building a heat exchanger is very expensive. For example in model 2, building a heat exchanger has a fixed cost and the

Model	N	Solution	Heuristic Time (s)	Non-Heuristic Time (s)
	10	5	1.44	2.09
	20	6	4.45	12.24
Model 1	30	5	5.64	26.89
	40	6	21.74	76.07
	50	9	41.9	148.24
	10	2	2.07	5.36
	20	2	6.06	22.16
Model 2	30	2	11.58	57.55
	40	2	28.07	49.9
	50	2	36.07	207.99
	10	5	3001.06	3806.93
	20	8	10755.51	11618.66
Model 3	30	7	16082.15	19536.24
	40	6	19532.65	19882.15
	50	6	32566.33	32223.98

Table 6.4: The results of running each model with the added constraint as we vary N, the non-heuristic times are the same as those in table 6.3

cost per unit area is high therefore one heat exchanger is prefered over two (one utility heat exchanger vs stream to stream and a utility heat exchanger).

It is important to note that this is still only a heuristic therefore, unless further analysis is done, adding the constraint could exclude the global solution from the model even if we have a high cost per heat exchanger.

There may other heuristics under which the algorithm shows improvements in performance.

7 Conclusion

In this investigation, we have taken an alternative approach to solving the simultaneous synthesis model proprosed by Yee and Grossmann and have provided a method by which we can converge to the global solution of this model.

We chose to tackle the log mean temperature difference in the model, a function that is often overlooked in these models and replaced by an approximation. The first step we took was to analyse the reciprocal of LMTD as an approximation of this function fits well when looking to handle errors. We then analysed the function classifying: symmetry, bounds and shape. Here we discovered that the function is is convex, a significant result as convexity gives us a means by which we can model the function and guarantee reduction of error. This formed the foundations on which we developed our algorithm.

The main contribution we make is the development of an algorithm that makes use of the more robust MILP solvers to solve the model. One of the nicest properties of the algorithm is that we do not need to worry about what effect our approximation has on the solution as the algorithm iteratively reduces it. The algorithm was not only presented but it was also shown to converge to the global solution and with an appropriate partitioning we can use it to find a solution.

7.1 Future Work

As shown in section 6, the implementation that we present is not perfect and encounters issues when attempting to scale problem sizes up. Below are some suggestions for possible directions that can be taken in relation to this project:

- **Adaptive Refinement**
  This extension to the algorithm would strengthen it by having dynamic generation of breakpoints in the non-LMTD approximations. This would be a significant addition to the implementation as we would no longer have to provide any partitioning constants only maximum error tolerances.
- **Spatial Branch and Bound** [2]
  In our algorithm we generate a MILP approximation and solve it using a MILP solver we then tighten our constraints and solve again. The problem here is that we don't know the path that the solver followed or what additional properties the solver may have learnt about the model we are solving. Spatial branch and bound allows us to control the search tree on each iteration we can make a more intelligent choice on where to move in the search tree unlike the solver which resolves the problem.
- **Strengthening a MINLP solver**
  MINLP solvers struggle with the indeterminacies found in LMTD however a reformulation with the reciprocal gives a convex function. We have handled LMTD by linearising however isn't necessary. Since the function is convex, taking convex approximations of the non-LMTD functions means that we can formulate a convex optimisation problem for which we can use methods to solve the model. This is stronger than our approach as it doesn't require an approximation for

LMTD i.e. one less set of approximations therefore less constraints. Using convex NLP methods we should be able to converge on the solution.

The problem of finding the optimal arrangement of heat exchangers in a network of streams to minimise costs belongs to the class of MINLP. Many members of this class of problems are some of the hardest to solve. We have shown that through rigorous analysis we can expose the structure of a model to find a direction to the global solution through the easier to solve class of MILP. Using this same approach we may be able to extend this idea to many other hard problems.

Appendices

A Notation & Abbreviations

Notation	
x	*variable of* $\mathbb{R}$
$\mathbf{x}$	*vector of* $\mathbb{R}^n$, $n \geq 1$
$\mathbf{A}$	*matrix of* $\mathbb{R}^{m \times n}$
$\mathbf{x}^T$	*transpose of* $\mathbf{x}$
f'	*derivative of univariate function* f
f''	*second derivative of univariate function* f
$f^{(n)}$	n^{th} *derivative of univariate function* f
f'_{x_1}	*first partial derivative of multivariate function* f *with respect to* x_1
$f''_{x_1 x_2}$	*second partial derivative of multivariate function* f *with respect to* x_1 *and* x_2
∇f	*gradient of multivariate function* f
$\nabla^2 f$	*Hessian of multivariate function* f
l	*function label for the log mean temperature difference*
m	*function label for the reciprocal of the log mean temperature difference*

Abbreviation	
LP	*Linear Program*
NLP	*Nonlinear Program*
MIP	*Mixed Integer Program*
MILP	*Mixed Integer Linear Program*
MINLP	*Mixed Integer Nonlinear Program*
HEN	*Heat Exchanger Network*
HENS	*Heat Exchanger Network Synthesis*
LMTD	*Logarithmic Mean Temperature Difference*

B General MINLP Model for HEN Synthesis

General Mixed-Integer Nonlinear Programming Model for Optimal Simultaneous Synthesis of Heat Exchangers Network

Miten Mistry

The following model is identical to the model formulated by Escobar and Grossmann with changes made to a number of equations which were noticed to be incorrect and were found to have: discrepancies when comparing with the GAMS specifications and typographical errors. The changes made can be found in section B.3.

The model described here consists of a general straightforward formulation for a given stage-wise superstructure, resulting in a nonconvex MINLP with a nonconvex objective function and several nonconvex constraints. In addition, it involves the logarithmic mean temperature that can result in numerical difficulties when the approach temperatures of both sides of the heat exchanger are equal (it can cause an indetermination, i.e. zero divided by zero).

B.1 Nomenclature

Sets	
HP	*Set of hot process streams i*
CP	*Set of cold process streams j*
HU	*Set of hot utility*
CU	*Set of cold utility*
ST	*Set of stages in the superstructure*

Parameters	*Units*	
α	[\$/yr]	*Factor for the area cost*
β	-	*Exponent for the area cost*
c	[\$/yr]	*Fixed charge for exchangers*
c_{CU}	[\$/kW · yr]	*Utility cost coefficient for cooling utility*
c_{HU}	[\$/kW · yr]	*Utility cost coefficient for heating utility*
F_i	[kW/K]	*Flow capacity of hot stream i*
F_j	[kW/K]	*Flow capacity of cold stream j*
h_i	$\left[kW/\mathrm{m}^2K \right]$	*Heat transfer coefficient for hot stream i*
h_j	$\left[kW/\mathrm{m}^2K \right]$	*Heat transfer coefficient for cold stream j*
h_{CU}	$\left[kW/\mathrm{m}^2K \right]$	*Heat transfer coefficient for cold utility*
h_{HU}	$\left[kW/\mathrm{m}^2K \right]$	*Heat transfer coefficient for cold utility*
N_T	-	*Number of stages*
T_i^{in}	[K]	*Inlet temperature of hot stream i*
T_j^{in}	[K]	*Inlet temperature of cold stream j*
T_{CU}^{in}	[K]	*Inlet temperature of cold utility*

Parameters	*Units*	
T_{HU}^{in}	$[\,K\,]$	*Inlet temperature of hot utility*
T_{i}^{out}	$[\,K\,]$	*Outlet temperature of hot stream i*
T_{j}^{out}	$[\,K\,]$	*Outlet temperature of cold stream j*
T_{CU}^{out}	$[\,K\,]$	*Outlet temperature of cold utility*
T_{HU}^{out}	$[\,K\,]$	*Outlet temperature of hot utility*
Ω	$[\,kW\,]$	*Upper bound for heat exchangers loads*
Γ	$[\,^{\circ}C\,]$	*Upper bound for temperature difference*
ΔT_{min}	$[\,^{\circ}C\,]$	*Lower bound for temperature difference*

Positive Variables	*Units*	
A_{ijk}	$\left[\,\text{m}^2\,\right]$	*Area of heat exchanger between hot stream i and cold stream j at stage k*
A_{cui}	$\left[\,\text{m}^2\,\right]$	*Area of heat exchanger between hot stream i and cold utility*
A_{huj}	$\left[\,\text{m}^2\,\right]$	*Area of heat exchanger between cold stream j and hot utility*
f_{ijk}^{H}	$[\,kW/K\,]$	*Flow rate of the hot side of exchanger i,j,k*
f_{ijk}^{C}	$[\,kW/K\,]$	*Flow rate of the cold side of exchanger i,j,k*
t_{ijk}^{H}	$[\,K\,]$	*Outlet temperature of the hot side of exchanger i, j, k*
t_{ijk}^{C}	$[\,K\,]$	*Outlet temperature of the cold side of exchanger i, j, k*
dt_{ijk}	$[\,K\,]$	*Temperature approach between hot stream i and cold stream j at location k*
dt_{cui}	$[\,K\,]$	*Temperature approach between hot stream i and cold utility*
dt_{huj}	$[\,K\,]$	*Temperature approach between cold stream j and hot utility*
$LMTD_{ijk}$	$[\,K\,]$	*Log mean temperature difference between hot stream i and cold stream j at stage k*
$LMTD_{cui}$	$[\,K\,]$	*Log mean temperature difference between hot stream i and the cold utility*
$LMTD_{huj}$	$[\,K\,]$	*Log mean temperature difference between cold stream j and the hot utility*
q_{ijk}	$[\,kW\,]$	*Heat load between hot stream i and cold stream j at stage k*
q_{cui}	$[\,kW\,]$	*Heat load between hot stream i and cold utility*
q_{huj}	$[\,kW\,]$	*Heat load between cold stream j and hot utility*
t_{ik}	$[\,K\,]$	*Temperature of hot stream i at hot end of stage k*
t_{jk}	$[\,K\,]$	*Temperature of cold stream j at hot end of stage k*

Binary Variables		
z_{ijk}	-	*Existence of the match between hot stream i, cold stream j, at stage k*
z_{cui}	-	*Existence of the match between hot stream i, and cold utility*
z_{huj}	-	*Existence of the match between cold stream j, and hot utility*

B.2 Model Formulation

The model formulation described here is based in a postulated stage-wise superstructure depicted in fig. B.1 Yee and Grossmann, 1990. The number of stages N_T, is normally set to $\max\{N_H, N_C\}$, where N_H and N_C are the number of hot streams and the number of cold streams respectively. For each stage, the corresponding stream is split and directed to an exchanger for each potential match between each hot and each cold stream. The outlet temperatures are mixed, which then defines the stream for the next stage. The outlet temperatures of each stage are treated as variables in the formulation.

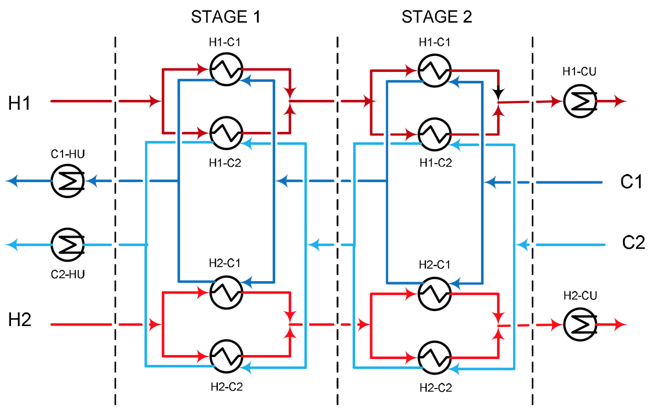


Figure B.1: Two-stage Superstructure for two hot and two cold streams.

In the model formulation, an overall heat balance is needed to ensure sufficient heating or cooling of each process stream:

$$\sum_{k \in ST} \sum_{j \in CP} q_{ijk} + q_{cui} = F_i \left(T_i^{\text{in}} - T_i^{\text{out}} \right), \qquad i \in HP \tag{B.1}$$

$$\sum_{k \in ST} \sum_{i \in HP} q_{ijk} + q_{huj} = F_j \left(T_j^{\text{out}} - T_j^{\text{in}} \right), \qquad j \in CP \tag{B.2}$$

For each split, mass balances are performed to define the flowrates to each heat exchanger:

$$\sum_{j \in CP} f_{ijk}^H = F_i, \qquad i \in HP, k \in ST \tag{B.3}$$

$$\sum_{i \in HP} f_{ijk}^C = F_j, \qquad j \in CP, k \in ST \tag{B.4}$$

Energy balances around each heat exchanger are performed in order to define the outlet temperatures of heat exchanger, which leads to equations with bilinear terms:

$$q_{ijk} = f_{ijk}^H \left(t_{ik} - t_{ijk}^H \right), \qquad i \in HP, j \in CP, k \in ST \tag{B.5}$$

$$q_{ijk} = f_{ijk}^C \left(t_{ijk}^C - t_{jk+1} \right), \qquad i \in HP, j \in CP, k \in ST \tag{B.6}$$

Energy balances around each mixer define the inlet temperatures of the stages, which also involve bilinear terms:

$$F_i t_{ik+1} = \sum_{j \in CP} f_{ijk}^H t_{ijk}^H, \qquad i \in HP, k \in ST \tag{B.7}$$

$$F_j t_{jk} = \sum_{i \in HP} f_{ijk}^C t_{ijk}^C, \qquad j \in CP, k \in ST \tag{B.8}$$

According to the superstructure, the assignment of the inlet temperatures is as follows:

$$t_{i1} = T_i^{\text{in}}, \qquad i \in HP \tag{B.9}$$
$$t_{j,N_T+1} = T_j^{\text{in}}, \qquad j \in CP \tag{B.10}$$

Energy balances for the final utility units define the utility loads:

$$q_{cui} = F_i \left(t_{i,N_T+1} - T_i^{\text{out}} \right), \qquad i \in HP \tag{B.11}$$
$$q_{huj} = F_j \left(T_j^{\text{out}} - t_{j1} \right), \qquad j \in CP \tag{B.12}$$

The constraints eqs. (B.13) to (B.16) are used to ensure feasibility of temperatures. They specify monotonic decrease in the temperatures along the stages:

$$t_{ik+1} \leq t_{ik}, \qquad i \in HP, k \in ST \tag{B.13}$$
$$t_{jk+1} \leq t_{jk}, \qquad j \in CP, k \in ST \tag{B.14}$$
$$t_{i,N_T+1} \leq T_i^{\text{out}}, \qquad i \in HP \tag{B.15}$$
$$t_{j1} \geq T_j^{\text{out}}, \qquad j \in CP \tag{B.16}$$

Upper bound constraints are needed to relate the heat loads q with the binary variables z. In the equations below Ω is an upper bound for the corresponding heat load. If the heat load is not equal to zero the corresponding binary variable is set to one, otherwise the binary variable can be 0 or 1, but the objective function forces the variable to be zero in order to minimize the number of units.

$$q_{ijk} - \Omega_{ij} z_{ijk} \leq 0, \qquad i \in HP, j \in CP, k \in ST \tag{B.17}$$
$$q_{cui} - \Omega_i z_{cui} \leq 0, \qquad i \in HP \tag{B.18}$$
$$q_{huj} - \Omega_j z_{huj} \leq 0, \qquad j \in CP \tag{B.19}$$

In addition, big-M constraints are needed to ensure that the temperature approaches only hold if the heat exchanger exists. The parameter Γ is an upper bound for the temperature difference. If the binary variable is equal to zero the equations are ensured to be feasible. On the other hand, if the binary variable is equal to one, the temperature differences are forced to act as an equality constraint in order to minimize the areas in the objective function.

$$dt_{ijk} \leq t_{ik} - t_{jk} + \Gamma_{ij} \left(1 - z_{ijk}\right), \qquad i \in HP, j \in CP, k \in ST \tag{B.20}$$
$$dt_{ijk+1} \leq t_{ik+1} - t_{jk+1} + \Gamma_{ij} \left(1 - z_{ijk}\right), \qquad i \in HP, j \in CP, k \in ST \tag{B.21}$$
$$dt_{cui} = t_{i,N_T+1} - T_{CU}^{\text{out}} + \Gamma_i \left(1 - z_{cui}\right), \qquad i \in HP \tag{B.22}$$
$$dt_{huj} = T_{HU}^{\text{out}} - t_{j1} + \Gamma_j \left(1 - z_{huj}\right), \qquad j \in CP \tag{B.23}$$

The trade-off between investment costs and operating costs are considered by adding the following constraints:

$$dt_{ijk} \geq \Delta T_{\min}, \qquad i \in HP, j \in CP, k \in ST \tag{B.24}$$
$$dt_{cui} \geq \Delta T_{\min}, \qquad i \in HP \tag{B.25}$$
$$dt_{huj} \geq \Delta T_{\min}, \qquad j \in CP \tag{B.26}$$

The next equations are necessary to calculate the area of each unit. The driving forces are calculated by the logarithmic mean temperature difference for each heat exchanger:

$$LMTD_{ijk} = \frac{dt_{ijk} - dt_{ijk+1}}{\ln\left(dt_{ijk}/dt_{ijk+1}\right)}, \qquad i \in HP, j \in CP, k \in ST \tag{B.27}$$

And the area of process-process heat exchangers can be calculated by the following equation:

$$A_{ijk} = q_{ijk}\left(h_i^{-1} + h_j^{-1}\right)/LMTD_{ijk}, \qquad i \in HP, j \in CP, k \in ST \tag{B.28}$$

Similarly for the utility exchangers, the temperature differences are calculated as follows:

$$LMTD_{cui} = \frac{dt_{cui} - \left(T_i^{\text{out}} - T_{CU}^{\text{in}}\right)}{\ln\left(dt_{cui}/\left(T_i^{\text{out}} - T_{CU}^{\text{in}}\right)\right)}, \qquad i \in HP \tag{B.29}$$
$$LMTD_{huj} = \frac{dt_{huj} - \left(T_{HU}^{\text{in}} - T_j^{\text{out}}\right)}{\ln\left(dt_{huj}/\left(T_{HU}^{\text{in}} - T_j^{\text{out}}\right)\right)}, \qquad j \in CP \tag{B.30}$$

And the areas of the utility units by the following equations:

$$A_{cui} = q_{cui}\left(h_i^{-1} + h_{CU}^{-1}\right)/LMTD_{cui}, \qquad i \in HP \tag{B.31}$$
$$A_{huj} = q_{huj}\left(h_j^{-1} + h_{HU}^{-1}\right)/LMTD_{huj}, \qquad j \in CP \tag{B.32}$$

Finally, the objective function for the total annual cost (TAC) and the complete model are as follows:

$$\begin{aligned} TAC = \min c_{CU} \sum_{i \in HP} q_{cui} + c_{HU} \sum_{j \in CP} q_{huj} + c \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} z_{ijk} + c \sum_{i \in HP} z_{cui} + c \sum_{j \in CP} z_{huj} \\ + \alpha \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} A_{ijk}^{\beta} + \alpha \sum_{i \in HP} A_{cui}^{\beta} + \alpha \sum_{j \in CP} A_{huj}^{\beta} \end{aligned} \tag{B.33}$$

The resulting MINLP model eqs. (B.1) to (B.33) is nonconvex because the block of eqs. (B.5) to (B.8) involve bilinear terms, and also the eqs. (B.27) to (B.32) are non-linear and the objective function eq. (B.33) is also nonlinear. In addition, the logarithmic mean temperature difference can cause numerical difficulties when driving forces are the same at both side of the heat exchanger, then LMTD reduces to zero by zero division, which is not determined.

B.3 Changes

- Addition of variables to table of parameters - The parameters $\{T_{CU}^{\text{in}}, T_{HU}^{\text{in}}, T_{CU}^{\text{out}}, T_{HU}^{\text{out}}, \Delta T_{\min}\}$ were added to the table as they were not previously present while references in the model were made.

- Removal of variables from the model - The variables $\{th_{ik}, tc_{jk}\}$ were removed as their semantic meaning was identical to that of $\{t_{ik}, t_{jk}\}$ respectively (The latter has been kept in the model)

- Notation change in table of variables. $\{fc_{ijk}, fh_{ijk}\} \rightarrow \left\{f_{ijk}^{C}, f_{ijk}^{H}\right\}$ respectively.

- Equations (B.3) and (B.4) - Removal of outer summation. Summation set placed as a quantifier $[k \in ST]$.

  From:

$$\sum_{k \in ST} \sum_{j \in CP} f_{ijk}^{H} = F_i, \qquad i \in HP$$
$$\sum_{k \in ST} \sum_{i \in HP} f_{ijk}^{C} = F_j, \qquad j \in CP$$

  To (respectively):

$$\sum_{j \in CP} f_{ijk}^{H} = F_i, \qquad i \in HP, k \in ST$$
$$\sum_{i \in HP} f_{ijk}^{C} = F_j, \qquad j \in CP, k \in ST$$

- Equation (B.6) - Variable change $[tc_{jk+1} \rightarrow t_{jk+1}]$. Order of subtraction changed, original equation implies $q_{ijk} \leq 0$

  From:

$$q_{ijk} = f_{ijk}^{C}\left(tc_{jk+1} - t_{ijk}^{C}\right), \qquad i \in HP, j \in CP, k \in ST$$

  To:

$$q_{ijk} = f_{ijk}^{C}\left(t_{ijk}^{C} - t_{jk+1}\right), \qquad i \in HP, j \in CP, k \in ST$$

- Equation (B.7) - Removal of quantifier $[j \in CP]$

  From:

$$F_i t_{ik+1} = \sum_{j \in CP} f_{ijk}^{H} t_{ijk}^{H}, \qquad i \in HP, j \in CP, k \in ST$$

  To:

$$F_i t_{ik+1} = \sum_{j \in CP} f_{ijk}^{H} t_{ijk}^{H}, \qquad i \in HP, k \in ST$$

- Equation (B.8) - Summation set changed [$j \in CP \rightarrow i \in HP$], removal of quantifier [$i \in HP$]

  From:

$$F_j t_{jk} = \sum_{j \in CP} f^C_{ijk} t^C_{ijk}, \qquad i \in HP, j \in CP, k \in ST$$

  To:

$$F_j t_{jk} = \sum_{i \in HP} f^C_{ijk} t^C_{ijk}, \qquad j \in CP, k \in ST$$

- Equation (B.19) - Reordering of equation (Semantics unchanged)

  From:

$$q_{huj} \leq \Omega_j z_{huj}, \qquad j \in CP$$

  To:

$$q_{huj} - \Omega_j z_{huj} \leq 0, \qquad j \in CP$$

- Equations (B.22) and (B.23) - Variable change due to inconsistent naming $\left[\left\{T^{CU}_{\text{out}}, T^{HU}_{\text{out}}\right\} \rightarrow \{T^{\text{out}}_{CU}, T^{\text{out}}_{HU}\}\right]$ respectively (Semantics unchanged)

  From:

$$\begin{aligned} dt_{cui} &= t_{i,NT+1} - T^{CU}_{\text{out}} + \Gamma_i \left(1 - z_{cui}\right), & i \in HP \\ dt_{huj} &= T^{HU}_{\text{out}} - t_{j1} + \Gamma_j \left(1 - z_{huj}\right), & j \in CP \end{aligned}$$

  To (respectively):

$$\begin{aligned} dt_{cui} &= t_{i,N_T+1} - T^{\text{out}}_{CU} + \Gamma_i \left(1 - z_{cui}\right), & i \in HP \\ dt_{huj} &= T^{\text{out}}_{HU} - t_{j1} + \Gamma_j \left(1 - z_{huj}\right), & j \in CP \end{aligned}$$

- Equations (B.24) to (B.26) - Changed inequality directions [$\leq \rightarrow \geq$]

  From:

$$\begin{aligned} dt_{ijk} &\leq \Delta T_{\min}, & i \in HP, j \in CP, k \in ST \\ dt_{cui} &\leq \Delta T_{\min}, & i \in HP \\ dt_{huj} &\leq \Delta T_{\min}, & j \in CP \end{aligned}$$

  To (respectively):

$$\begin{aligned} dt_{ijk} &\geq \Delta T_{\min}, & i \in HP, j \in CP, k \in ST \\ dt_{cui} &\geq \Delta T_{\min}, & i \in HP \\ dt_{huj} &\geq \Delta T_{\min}, & j \in CP \end{aligned}$$

- Equation (B.33) - Replacement of variables $[\{q_{cui}, q_{huj}\} \rightarrow \{z_{cui}, z_{huj}\}]$ respectively

  From:

$$TAC = \min c_{CU} \sum_{i \in HP} q_{cui} + c_{HU} \sum_{j \in CP} q_{huj} + c \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} z_{ijk} + c \sum_{i \in HP} q_{cui} + c \sum_{j \in CP} q_{huj}$$
$$+\alpha \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} A_{ijk}^{\beta} + \alpha \sum_{i \in HP} A_{cui}^{\beta} + \alpha \sum_{j \in CP} A_{huj}^{\beta}$$

  To:

$$TAC = \min c_{CU} \sum_{i \in HP} q_{cui} + c_{HU} \sum_{j \in CP} q_{huj} + c \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} z_{ijk} + c \sum_{i \in HP} z_{cui} + c \sum_{j \in CP} z_{huj}$$
$$+\alpha \sum_{i \in HP} \sum_{j \in CP} \sum_{k \in ST} A_{ijk}^{\beta} + \alpha \sum_{i \in HP} A_{cui}^{\beta} + \alpha \sum_{j \in CP} A_{huj}^{\beta}$$

C Analysis of LMTD

Here we carry out a similar proces to that found in section 3. Some of the results presented here have been proven before by Zavala-Río, Femat, and Santiesteban-Cos these are the limit of l, the gradient of l and the continuously differentiable property of l, we present an alternative proof. The proofs we present were derived prior to finding the proofs given by Zavala-Río, Femat, and Santiesteban-Cos.

We will prove existence of limits and that the log mean temperature difference, l, is concave.

C.1 Function Definition

We begin by defining the log mean temperature difference, for simplicity we will use the label l when referring the function.

First we define the following set, the domain of l.

$$S^* = \left\{ (x, y)^T \;\middle|\; x, y > 0,\, x \neq y \right\}$$

The log mean temperature difference, $l : S^* \to \mathbb{R}$ is defined as

$$l\left(x, y\right) = \frac{x - y}{\ln\left(\frac{x}{y}\right)}$$

We also define the gradient and Hessian of l, the derivations for these functions is presented in appendix F.

The gradient of l, $\nabla l : S^* \to \mathbb{R}^2$ is defined as

$$\nabla l\left(x, y\right) = \ln\left(w\right)^{-2} \begin{pmatrix} w^{-1} + \ln\left(w\right) - 1 \\ w - \ln\left(w\right) - 1 \end{pmatrix}$$

The Hessian of l, $\nabla^2 l : S^* \to \mathbb{R}^{2\times 2}$ is defined as

$$\nabla^2 l\left(x, y\right) = y^{-1} \ln\left(w\right)^{-2} k\left(w\right) \begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}$$

where

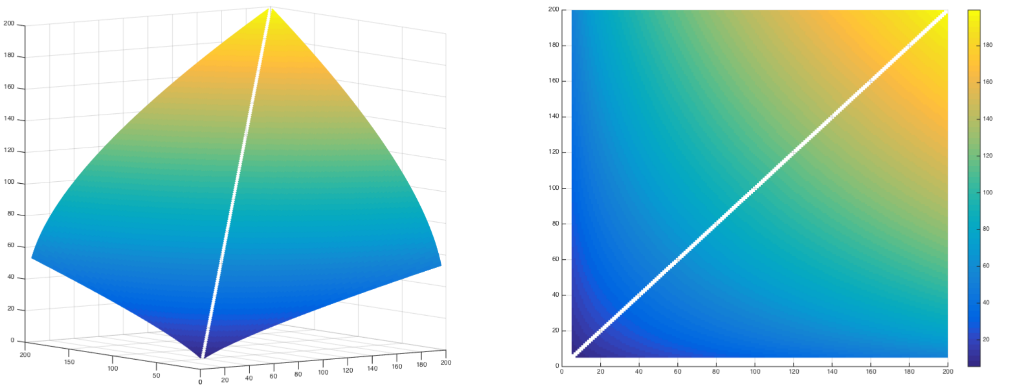
$$k\left(w\right) = 1 + w^{-1} + 2w^{-1} \ln\left(w\right)^{-1} - 2\ln\left(w\right)^{-1}$$

C.2 The Limits of LMTD

We will prove that the limits of l, ∇l and $\nabla^2 l$ over the set

$$S = \left\{ (x, y)^T \;\middle|\; x, y > 0 \right\}$$

do exist and we will state their values.



(a) LMTD, l (b) l viewed as a heat map in the x, y plane

Figure C.1: Plots of the log mean temperature difference

We can see, from fig. C.1a, the indeterminate set of points. Intuitively, we establish that the limit appears to exist as the defined values of l seem to converge to the same point as we approach $(c, c)^T$, $c > 0$. fig. C.1b shows that if we represent the function in polar form, the indeterminate set of points in the cartesian system collapse to a single angle in the polar system.

The proof presented here is similar to the proof given in section 3.4 and the limit will be proven via polar coordinates, where the limit being proven is:

$$(r, \theta)^T \to \left(r_c, \frac{\pi}{4}\right)^T$$

where $r_c = c\sqrt{2} > 0$

C.2.1 Evaluation of the Limit

We will now evaluate the limit this will be done by making use of theorems 2.1 and 2.3.

Polar Representation of l

We derive the polar representation of l by making the transformations defined by eqs. (3.2) to (3.4).

$$\begin{aligned} l_{r,\theta}(r, \theta) &= l\left(r\cos(\theta), r\sin(\theta)\right) \\ &= \frac{r\cos(\theta) - r\sin(\theta)}{\ln\left(\cot(\theta)\right)} \\ &= r \cdot \frac{\cos(\theta) - \sin(\theta)}{\ln\left(\cot(\theta)\right)} \\ &= f(r) \cdot \frac{g_1(\theta)}{g_2(\theta)} \end{aligned} \tag{C.1}$$

$f(r)$ is clearly well defined for all $r > 0$.

The function $\frac{g_1(\theta)}{g_2(\theta)}$ is well defined except possibly for

$$\theta = \frac{\pi}{4}$$

On inspection we find that the θ parametrised parts of l at $\frac{\pi}{4}$ evaluates to a fraction of the form $\frac{0}{0}$, this is where we shall use theorem 2.3.

In eq. (C.1), we see that the resulting polar representations give a set of separable functions parametrised by independent variables, r and θ, this is where we shall use theorem 2.1.

Evaluation of the Derivatives
The derivation of the following derivatives can be found in appendix H.

We begin our analysis with $g_2(\theta)$ as this indicates how many derivatives we will require from $g_1(\theta)$ if a limit is to exist i.e. we continue to take derivatives until the first non-zero is found.

$$\begin{aligned} g_2(^\pi/_4) &= \ln\left(\cot(^\pi/_4)\right) = 0 \\ g_2'(^\pi/_4) &= -\csc(^\pi/_4)\sec(^\pi/_4) = -2 \end{aligned}$$

Here we see that the only necessary requirement on g_1 is

$$g_1(^\pi/_4) = 0$$

On evaluation we see that this property holds.

$$g_1(^\pi/_4) = \cos(^\pi/_4) - \sin(^\pi/_4) \quad = 0 \tag{C.2}$$

$$g_1'(^\pi/_4) = -\sin(^\pi/_4) - \cos(^\pi/_4) = -\sqrt{2} \tag{C.3}$$

The Limit
By theorem 2.3 we get the following results:

$$\frac{g_1(^\pi/_4)}{g_2(^\pi/_4)} = \frac{g_1'(^\pi/_4)}{g_2'(^\pi/_4)} = \frac{-\sqrt{2}}{-2} = \frac{1}{\sqrt{2}}$$

Hence the limit of $\frac{g_1(\theta)}{g_2(\theta)}$ is finite (well defined) at $\frac{\pi}{4}$.

We can now evaluate the limit of l using theorem 2.1.

$$\begin{aligned} \lim_{(r,\theta)\to(r_c,^\pi/_4)} f(r)\cdot\frac{g_1(\theta)}{g_2(\theta)} &= \left[\lim_{r\to r_c} f(r)\right]\left[\lim_{\theta\to^\pi/_4}\frac{g_1(\theta)}{g_2(\theta)}\right] = c\sqrt{2}\cdot\frac{1}{\sqrt{2}} \\ &= c \end{aligned}$$

But c was simply the value of x (and y) hence we get the following result

$$\forall c > 0, \qquad \lim_{(x,y)^T\to(c,c)^T} l(x,y) = \frac{1}{c}$$

Let

$$S = \left\{ (x, y)^T \;\middle|\; x, y > 0 \right\}$$

The log mean temperature difference, $l : S \to \mathbb{R}$, is defined as

$$l(x, y) = \begin{cases} x, & x = y \\ \dfrac{x - y}{\ln\left({}^{x}/_{y}\right)}, & x \neq y \end{cases}$$

Using a similar process, we can evaluate the limits of $\nabla l(x, y)$ and $\nabla^2 l(x, y)$, this is presented in appendix D.

C.3 Well Defined Formulation of LMTD

Having calculated the limits of l, ∇l and $\nabla^2 l$ at any indeterminate points and showing that they exist we can define the well defined formulations for these functions.

Definition C.1.
The log mean temperature difference

$$l : S \to \mathbb{R}$$

its gradient

$$\nabla l : S \to \mathbb{R}^2$$

and its hessian

$$\nabla^2 l : S \to \mathbb{R}^{2\times 2}$$

are defined as

$$l(x, y) = \begin{cases} x, & x = y \\ \dfrac{x - y}{\ln\left({}^{x}/_{y}\right)}, & x \neq y \end{cases}$$

$$\nabla l(x, y) = \begin{cases} \begin{pmatrix} {}^{1}/_{2} \\ {}^{1}/_{2} \end{pmatrix}, & x = y \\ \ln(w)^{-2} \begin{pmatrix} w^{-1} + \ln(w) - 1 \\ w - \ln(w) - 1 \end{pmatrix}, & x \neq y \end{cases}$$

$$\nabla^2 l(x, y) = \begin{cases} \dfrac{1}{6x} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, & x = y \\ y^{-1} \ln(w)^{-2} k(w) \begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}, & x \neq y \end{cases}$$

where

$$k(w) = 1 + w^{-1} + 2w^{-1} \ln(w)^{-1} - 2 \ln(w)^{-1}$$

The above definition has been shown to conform to the limits of the respective: function definition, derived gradient and derived Hessian of l i.e. the function evaluates to its limit for all points of its domain, we therefore conclude that

$$l \in \mathcal{C}^2$$

that is: l belongs to the set of twice continuously differentiable functions.

C.4 Properties of LMTD

Having a continuously differentiable definition of l over a convex domain, we can classify some properties of the function. Unless stated otherwise the properties presented here have a similar proof to an equivalent property for m presented with proof in section 3.6.

C.4.1 Symmetry

Lemma C.1.
For all $(x, y)^T \in S$

$$l(y, x) = l(x, y)$$
$$\nabla l(y, x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \nabla l(x, y)$$

C.4.2 Bounds

Proposition C.1.

$$\forall (x, y)^T \in S, x \neq y \; \exists \, t > 0 \quad s.t \quad l(t, t) > l(x, y)$$

Proposition C.2.

$$\forall (x, y)^T \in S, \quad l(x, y) > 0$$

Proof.
Since

$$l(x, y) = \frac{1}{m(x, y)}$$

the result follows from corollary 3.1.

□

Proposition C.3.
Let $(x, y)^T \in S$.

- *If we fix* y *and let* x *vary then* $l(x, y)$ *is strictly increasing*
- *If we fix* x *and let* y *vary then* $l(x, y)$ *is strictly increasing*

Proposition C.4.
Let $l, u \in \mathbb{R}$ such that $0 < l < u$

Over the set

$$S_{[l,u]} = \left\{ (x,y)^T \;\middle|\; l \leq x, y \leq u \right\}$$

the unique bounds of l are given by

$$\min_{(x,y)^T \in S_{[l,u]}} l(x,y) = l(l,l) = l$$
$$\max_{(x,y)^T \in S_{[l,u]}} l(x,y) = l(u,u) = u$$

C.4.3 Concavity

Proposition C.5.
The log mean temperature difference

$$l\left(x,y\right) = \frac{x-y}{\ln\left(\frac{x}{y}\right)}$$

is concave over the sets

$$S_1 = \left\{ (x,y)^T \;\middle|\; 0 < y < x \right\}$$
$$S_2 = \left\{ (x,y)^T \;\middle|\; 0 < x < y \right\}$$

Proof.
We use the substitution $w = \frac{x}{y}$, this means that S_1 and S_2 can be defined as

$$S_1 = \{ w \mid w > 1 \}$$
$$S_2 = \{ w \mid 0 < w < 1 \}$$

The union of these sets is

$$S = \{ w \mid w > 0,\, w \neq 1 \}$$

The hessian of l is

$$\nabla^2 l\left(x,y\right) = y^{-1} \ln\left(w\right)^{-2} k\left(w\right) \begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}$$

where

$$k\left(w\right) = 1 + w^{-1} + 2w^{-1}\ln\left(w\right)^{-1} - 2\ln\left(w\right)^{-1}$$

For concavity, we require for $\nabla^2 l\left(x,y\right)$ to be negative semi-definite.
Since $w \neq 1$ we have that $\nabla^2 l\left(x,y\right)$ is well defined.

As $y^{-1}\ln\left(w\right)^{-2} > 0$, for $l\left(x,y\right)$ to be concave the following matrix has to be negative semi-definite.

$$\mathbf{A} = k\left(w\right)\begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}$$

$\mathbf{A}$ is negative semi-definite if $-\mathbf{A}$ is positive semi-definite.

$$-\mathbf{A} = k\left(w\right)\begin{pmatrix} w^{-1} & -1 \\ -1 & w \end{pmatrix}$$

The matrix

$$\mathbf{B} = \begin{pmatrix} w^{-1} & -1 \\ -1 & w \end{pmatrix}$$

is positive semi-definite, this can be shown directly

$$\begin{aligned}\mathbf{v}^T\mathbf{B}\mathbf{v} &= w^{-1}v_1^2 + wv_2^2 - 2v_1v_2 \\ &= \left(w^{-\frac{1}{2}}v_1 - w^{\frac{1}{2}}v_2\right)^2 \\ &\geq 0, \quad \forall \mathbf{v} \in \mathbb{R}^2\end{aligned}$$

As $-\mathbf{A} = k\left(w\right)\mathbf{B}$ we require that

$$\forall w \in S, \quad k\left(w\right) \geq 0 \tag{C.4}$$

To show this we make a substitution for w.
As $w > 0$ we know the following will be true

$$\exists\, n \in \mathbb{R} \setminus \{0\} \;\text{ s.t. } w = e^n$$

i.e. $n = \ln\left(w\right)$
Hence

$$k\left(w\right), \quad w \in S$$

is equivalent to

$$k\left(e^n\right) \quad \text{where} \quad n = \ln\left(w\right)$$

We wish to show that $k\left(e^n\right) \geq 0, \quad \forall n \in \mathbb{R} \setminus \{0\}$
We have

$$\begin{aligned}k_1(n) = k\left(e^n\right) &= 1 + e^{-n} + 2e^{-n}\ln\left(e^n\right)^{-1} - 2\ln\left(e^n\right)^{-1} \\ &= 1 + e^{-n} + 2e^{-n}n^{-1} - 2n^{-1}\end{aligned}$$

If we separate $n > 0$ and $n < 0$ we get the following cases:

- Case 1: $n > 0$

$$k_1(n) = 1 + e^{-n} + 2e^{-n}n^{-1} - 2n^{-1}$$

$$= n^{-1}e^{-n}\left(ne^{n} + n + 2 - 2e^{n}\right)$$

As $n > 0$, we have

$$n^{-1}e^{-n} > 0$$

Hence to show that $k_1(n) \geq 0$ for $n > 0$ we need to show that

$$\forall n > 0 \qquad ne^{n} + n + 2 - 2e^{n} \geq 0 \tag{C.5}$$

- Case 2: $n < 0$
  If we let $p = -n$, we have $p > 0$ and

$$\begin{aligned} k_1(n) = k_1(-p) &= 1 + e^{p} + 2e^{p}\left(-p\right)^{-1} - 2\left(-p\right)^{-1} \\ &= 1 + e^{p} - 2e^{p}p^{-1} + 2p^{-1} \\ &= p^{-1}\left(pe^{p} + p + 2 - 2e^{p}\right) \end{aligned}$$

  As $p > 0$, we have

$$p^{-1} > 0$$

  Hence to show that $k_1(-p) \geq 0$ for $p > 0$ we need to show that

$$\forall p > 0 \qquad pe^{p} + p + 2 - 2e^{p} \geq 0 \tag{C.6}$$

The resulting conditions, eqs. (C.5) and (C.6) are identical.
We will show that

$$\forall n > 0 \qquad t(n) = ne^{n} + n + 2 - 2e^{n} \geq 0$$

First we note that

$$t(0) = 0 \tag{C.7}$$

The derivatives of t are

$$\begin{aligned} t'(n) &= e^{n} + ne^{n} + 1 - 2e^{n} \\ &= ne^{n} - e^{n} + 1 \\ t''(n) &= e^{n} + ne^{n} - e^{n} \\ &= ne^{n} \end{aligned}$$

We have that

$$\forall n > 0 \qquad t''(n) > 0 \tag{C.8}$$

and

$$t'(0) = 0 \tag{C.9}$$

Using proposition 2.1 we get

$$\forall n > 0 \qquad t(n) > t(0) = 0 \tag{C.10}$$

This shows eq. (C.4) and therefore that $\nabla^2 l\,(x, y)$ is negative semi-definite.

□

Proposition C.6.
The log mean temperature difference

$$l\,(x, y) = \frac{x - y}{\ln\left(\frac{x}{y}\right)}$$

is concave over its entire domain, S:

$$S = \left\{ \, (x, y)^T \,\middle|\, x, y > 0 \, \right\}$$

Proof.
We have by proposition C.5 that l is concave over

$$S_1 = \left\{ \, (x, y)^T \,\middle|\, 0 < y < x \, \right\}$$
$$S_2 = \left\{ \, (x, y)^T \,\middle|\, 0 < x < y \, \right\}$$

We already know that

$$l \in \mathcal{C}^2$$

therefore only concavity over the set

$$S^* = \left\{ \, (x, x)^T \,\middle|\, x > 0 \, \right\}$$

remains to be shown, we do this by analysing $\nabla^2 l(x, y)$ over S^*.

For $l(x, y)$ to be concave over S^*, we require for $\nabla^2 l(x, y)$ to be negative semi-definite over S^*.

This can be shown directly,

$$\begin{aligned}
\forall\, \mathbf{v} \in \mathbb{R}^2, \qquad \mathbf{v}^T \nabla^2 l(x, x) \mathbf{v} &= \frac{1}{6x} \cdot \mathbf{v}^T \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{v} \\
&= \frac{1}{6x} \left(-v_1^2 + 2v_1 v_2 - v_2^2 \right) \\
&= -\frac{1}{6x} \left(v_1^2 - 2v_1 v_2 + v_2^2 \right) \\
&= -\frac{1}{6x} \left(v_1 - v_2 \right)^2 \\
&\leq 0
\end{aligned}$$

The inequality holds as $x > 0$

Hence, we have shown that $\nabla^2 l(x, y)$ is negative semi-definite over S^*.

This result along with proposition C.5 and the twice continuous differentiability property of l shows that $l(x, y)$ is concave over S.

□

D Limits of l Gradient and Hessian

The proofs presented here are similar to the proof given in section 3.4 and the limit will be proven via polar coordinates, where the limit being proven is:

$$(r, \theta)^T \to \left(r_c, \frac{\pi}{4}\right)^T$$

where $r_c = c\sqrt{2} > 0$

D.1 Limit of ∇l

The gradient of l is defined as

$$\begin{aligned}\nabla l\,(x, y) &= \ln\,(w)^{-2} \begin{pmatrix} w^{-1} + \ln\,(w) - 1 \\ w - \ln\,(w) - 1 \end{pmatrix} \\ &= \begin{pmatrix} l'_x\,(x, y) \\ l'_y\,(x, y) \end{pmatrix}\end{aligned}$$

Polar Representation

We derive the polar representations of $l'_x\,(x, y)$ and $l'_y\,(x, y)$ separately by making the transformations defined by eqs. (3.2) to (3.4)

$$\begin{aligned}l'^{[r,\theta]}_x(r, \theta) &= l'_x\,(r\cos(\theta), r\sin(\theta)) \\ &= \frac{\tan(\theta) + \ln\,(\cot(\theta)) - 1}{\ln\,(\cot(\theta))^2} \\ &= \frac{g^{[x]}(\theta)}{g(\theta)} \\ l'^{[r,\theta]}_y(r, \theta) &= l'_y\,(r\cos(\theta), r\sin(\theta)) \\ &= \frac{\cot(\theta) - \ln\,(\cot(\theta)) - 1}{\ln\,(\cot(\theta))^2} \\ &= \frac{g^{[y]}(\theta)}{g(\theta)}\end{aligned}$$

These functions are well defined except possibly for

$$\theta = \frac{\pi}{4}$$

Evaluation of the Derivatives

The derivation of the following derivatives can be found in appendix H

We begin our analysis with $g(\theta)$ as this indicates (for both functions) how many derivatives we require if a limit is to exist.

$$
\begin{aligned}
g(\pi/4) &= \ln\left(\cot(\pi/4)\right)^2 \\
&= 0 \\
g'(\pi/4) &= -2\csc(\pi/4)\sec(\pi/4)\ln\left(\cot(\pi/4)\right) \\
&= 0 \\
g''(\pi/4) &= 2\csc^2(\pi/4)\ln\left(\cot(\pi/4)\right) - 2\sec^2(\pi/4)\ln\left(\cot(\pi/4)\right) + 2\csc^2(\pi/4)\sec^2(\pi/4) \\
&= 8
\end{aligned}
$$

Here we see that the following are necessary requirements in order for a limit to be proven to exist by l'Hôpital's rule.

$$
\begin{aligned}
g^{[x]}(\pi/4) &= 0, & g^{[x]\prime}(\pi/4) &= 0 \\
g^{[y]}(\pi/4) &= 0, & g^{[y]\prime}(\pi/4) &= 0
\end{aligned}
$$

On evaluation we see that these properties hold.

Evaluation of derivatives of $g^{[x]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
g^{[x]}(\pi/4) &= \tan(\pi/4) + \ln\left(\cot(\pi/4)\right) - 1 &&= 0 \\
g^{[x]\prime}(\pi/4) &= \sec^2(\pi/4) - \csc(\pi/4)\sec(\pi/4) &&= 0 \\
g^{[x]\prime\prime}(\pi/4) &= 2\sec^2(\pi/4)\tan(\pi/4) + \csc^2(\pi/4) - \sec^2(\pi/4) &&= 4
\end{aligned}
$$

Evaluation of derivatives of $g^{[y]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
g^{[y]}(\pi/4) &= \cot(\pi/4) - \ln\left(\cot(\pi/4)\right) - 1 &&= 0 \\
g^{[y]\prime}(\pi/4) &= -\csc^2(\pi/4) + \csc(\pi/4)\sec(\pi/4) &&= 0 \\
g^{[y]\prime\prime}(\pi/4) &= 2\csc^2(\pi/4)\cot(\pi/4) - \csc^2(\pi/4) + \sec^2(\pi/4) &&= 4
\end{aligned}
$$

The Limit

By l'Hôpital's rule we get the following results:

$$
\begin{aligned}
\frac{g^{[x]}(\pi/4)}{g(\pi/4)} &= \frac{g^{[x]\prime\prime}(\pi/4)}{g''(\pi/4)} = \frac{4}{8} = \frac{1}{2} \\
\frac{g^{[y]}(\pi/4)}{g(\pi/4)} &= \frac{g^{[y]\prime\prime}(\pi/4)}{g''(\pi/4)} = \frac{4}{8} = \frac{1}{2}
\end{aligned}
$$

These limits are respectively the limits of l'_x and l'_y as $(x, y) \to (c, c)$.

Hence we can define the value of ∇l as follows

Let

$$S = \left\{ (x,y)^T \;\middle|\; x,y > 0 \right\}$$

The gradient of the log mean temperature difference, $\nabla l : S \to \mathbb{R}^2$, is defined as

$$\nabla l\left(x,y\right) = \begin{cases} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, & x = y \\ \ln\left(w\right)^{-2} \begin{pmatrix} w^{-1} + \ln\left(w\right) - 1 \\ w - \ln\left(w\right) - 1 \end{pmatrix}, & x \neq y \end{cases}$$

D.2 Limit of $\nabla^2 l$

The Hessian of l is defined as

$$\nabla^2 l\left(x,y\right) = y^{-1} \ln\left(w\right)^{-2} k\left(w\right) \begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}$$

where

$$k\left(w\right) = 1 + w^{-1} + 2w^{-1} \ln\left(w\right)^{-1} - 2\ln\left(w\right)^{-1}$$

Polar Representation

We derive the polar representations of l''_{xx}, l''_{xy} and l''_{yy} separately by making the transformations defined by eqs. (3.2) to (3.4)

$$\begin{aligned}
l''^{[r,\theta]}_{xx}(r,\theta) &= l''_{xx}\left(r\cos(\theta), r\sin(\theta)\right) \\
&= -r^{-1}\sec(\theta)\ln\left(\cot(\theta)\right)^{-3}\left(\ln\left(\cot(\theta)\right) + \tan(\theta)\ln\left(\cot(\theta)\right) + 2\tan(\theta) - 2\right) \\
&= -r^{-1} \cdot \frac{\sec\left(\theta\right)\ln\left(\cot(\theta)\right) + \sec\left(\theta\right)\tan(\theta)\ln\left(\cot(\theta)\right) + 2\sec\left(\theta\right)\tan(\theta) - 2\sec\left(\theta\right)}{\ln\left(\cot\left(\theta\right)\right)^3} \\
&= f^{[xx]}(r) \cdot \frac{g^{[xx]}(\theta)}{g(\theta)} \\
l''^{[r,\theta]}_{xy}(r,\theta) &= l''_{xy}\left(r\cos(\theta), r\sin(\theta)\right) \\
&= r^{-1}\csc\left(\theta\right)\ln\left(\cot\left(\theta\right)\right)^{-3}\left(\ln\left(\cot\left(\theta\right)\right) + \tan(\theta)\ln\left(\cot\left(\theta\right)\right) + 2\tan(\theta) - 2\right) \\
&= r^{-1}\ln\left(\cot\left(\theta\right)\right)^{-3}\left(\csc\left(\theta\right)\ln\left(\cot\left(\theta\right)\right) + \sec(\theta)\ln\left(\cot\left(\theta\right)\right) + 2\sec(\theta) - 2\csc\left(\theta\right)\right) \\
&= r^{-1} \cdot \frac{\csc\left(\theta\right)\ln\left(\cot\left(\theta\right)\right) + \sec(\theta)\ln\left(\cot\left(\theta\right)\right) + 2\sec(\theta) - 2\csc\left(\theta\right)}{\ln\left(\cot\left(\theta\right)\right)^3} \\
&= f^{[xy]}(r) \cdot \frac{g^{[xy]}(\theta)}{g(\theta)} \\
l^{[r,\theta]}_{yy}(r,\theta) &= l''_{xy}\left(r\cos(\theta), r\sin(\theta)\right) \\
&= -r^{-1}\cot(\theta)\csc(\theta)\ln\left(\cot(\theta)\right)^{-3}\left(\ln\left(\cot(\theta)\right) + \tan(\theta)\ln\left(\cot(\theta)\right) + 2\tan(\theta) - 2\right) \\
&= -r^{-1} \cdot \frac{\csc(\theta)\ln\left(\cot(\theta)\right) + \csc(\theta)\cot(\theta)\ln\left(\cot(\theta)\right) + 2\csc(\theta) - 2\csc(\theta)\cot(\theta)}{\ln\left(\cot(\theta)\right)^3}
\end{aligned}$$

$$= f^{[yy]}(r) \cdot \frac{g^{[yy]}(\theta)}{g(\theta)}$$

In the above derivations, we see that the resulting polar representations give a set of separable functions parametrised by independent variables, r and θ.

These functions are well defined except possibly for

$$\theta = \frac{\pi}{4}$$

On inspection we find that the θ parametrised parts of l at $\frac{\pi}{4}$ evaluates to a fraction of the form $\frac{0}{0}$. For limits of this form, we can use l'Hôpital's rule, repeatedly if need be.

Evaluation of the Derivatives

On evaluation we find that we require more than simply the first derivatives, the derivation of these can be found in appendix H.

We begin our analysis with $g(\theta)$ as this indicates (for all three functions) how many derivatives we require if a limit is to exist.

The derivatives of $g(\theta)$ evaluate as follows:

$$\begin{aligned}
g({}^{\pi}/_{4}) =& \ln\left(\cot({}^{\pi}/_{4})\right)^3 \\
=& \,0 \\
g'({}^{\pi}/_{4}) =& -3\csc({}^{\pi}/_{4})\sec({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right)^2 \\
=& \,0 \\
g''({}^{\pi}/_{4}) =& \,3\csc^2({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right)^2 - 3\sec^2({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right)^2 \\
&+ 6\csc^2({}^{\pi}/_{4})\sec^2({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right) \\
=& \,0 \\
g^{(3)}({}^{\pi}/_{4}) =& \,18\sec^3({}^{\pi}/_{4})\csc({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right) - 18\csc^3({}^{\pi}/_{4})\sec({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right) \\
&- 6\csc^2({}^{\pi}/_{4})\cot({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right)^2 - 6\sec^2({}^{\pi}/_{4})\tan({}^{\pi}/_{4})\ln\left(\cot({}^{\pi}/_{4})\right)^2 \\
&- 6\csc^3({}^{\pi}/_{4})\sec^3({}^{\pi}/_{4}) \\
=& -48
\end{aligned}$$

Here we see that the following are necessary requirements in order for a limit to be proven to exist by l'Hôpital's rule.

$$\begin{aligned}
&g^{[xx]}({}^{\pi}/_{4}) = 0, \qquad && g^{[xx]\prime}({}^{\pi}/_{4}) = 0, \qquad && g^{[xx]\prime\prime}({}^{\pi}/_{4}) = 0 \\
&g^{[xy]}({}^{\pi}/_{4}) = 0, && g^{[xy]\prime}({}^{\pi}/_{4}) = 0, && g^{[xy]\prime\prime}({}^{\pi}/_{4}) = 0 \\
&g^{[yy]}({}^{\pi}/_{4}) = 0, && g^{[yy]\prime}({}^{\pi}/_{4}) = 0, && g^{[yy]\prime\prime}({}^{\pi}/_{4}) = 0
\end{aligned}$$

On evaluation we see that these properties holds.

Evaluation of derivatives of $g^{[xx]}(\theta)$ at $\frac{\pi}{4}$

$$\begin{aligned}
g^{[xx]}(^{\pi}/_{4}) =&\ \sec(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\tan(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ 2\sec(^{\pi}/_{4})\tan(^{\pi}/_{4}) - 2\sec(^{\pi}/_{4})\\
=&\ 0\\
g^{[xx]\prime}(^{\pi}/_{4}) =&\ \sec^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ \sec(^{\pi}/_{4})\tan(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) - \sec^2(^{\pi}/_{4})\csc(^{\pi}/_{4}) + \sec^3(^{\pi}/_{4})\\
&+ 2\sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4}) - 2\sec(^{\pi}/_{4})\tan(^{\pi}/_{4})\\
=&\ 0\\
g^{[xx]\prime\prime}(^{\pi}/_{4}) =&\ 5\sec^3(^{\pi}/_{4})\tan(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\tan^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ \sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&- \sec^4(^{\pi}/_{4})\csc(^{\pi}/_{4}) - 5\sec^3(^{\pi}/_{4}) + \sec(^{\pi}/_{4})\csc^2(^{\pi}/_{4})\\
&+ 6\sec^3(^{\pi}/_{4})\tan(^{\pi}/_{4}) + 2\sec(^{\pi}/_{4})\tan^3(^{\pi}/_{4}) - 2\sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4})\\
=&\ 0\\
g^{[xx](3)}(^{\pi}/_{4}) =&\ 18\sec^3(^{\pi}/_{4})\tan^2(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + 5\sec^5(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ \sec(^{\pi}/_{4})\tan^4(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\tan^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ 5\sec^3(^{\pi}/_{4})\tan(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) - 3\sec^5(^{\pi}/_{4}) - \sec^4(^{\pi}/_{4})\csc(^{\pi}/_{4})\\
&+ \sec^3(^{\pi}/_{4})\csc^2(^{\pi}/_{4})\\
&- 20\sec^3(^{\pi}/_{4})\tan(^{\pi}/_{4}) + \sec^2(^{\pi}/_{4})\csc(^{\pi}/_{4}) - 2\csc^3(^{\pi}/_{4})\\
&+ 23\sec^3(^{\pi}/_{4})\tan^2(^{\pi}/_{4}) + 2\sec(^{\pi}/_{4})\tan^4(^{\pi}/_{4}) - 2\sec(^{\pi}/_{4})\tan^3(^{\pi}/_{4})\\
=&\ -8\sqrt{2}
\end{aligned}$$

Evaluation of derivatives of $g^{[xy]}(\theta)$ at $\frac{\pi}{4}$

$$\begin{aligned}
g^{[xy]}(^{\pi}/_{4}) =&\ \csc(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + 2\sec(^{\pi}/_{4}) - 2\csc(^{\pi}/_{4})\\
=&\ 0\\
g^{[xy]\prime}(^{\pi}/_{4}) =&\ \sec(^{\pi}/_{4})\tan(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) - \csc(^{\pi}/_{4})\cot(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&- \csc^2(^{\pi}/_{4})\sec(^{\pi}/_{4}) - \sec^2(^{\pi}/_{4})\csc(^{\pi}/_{4})\\
&+ 2\sec(^{\pi}/_{4})\tan(^{\pi}/_{4}) + 2\csc(^{\pi}/_{4})\cot(^{\pi}/_{4})\\
=&\ 0\\
g^{[xy]\prime\prime}(^{\pi}/_{4}) =&\ \csc^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec^3(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) - \sec^3(^{\pi}/_{4}) + \csc^3(^{\pi}/_{4})\\
&+ \csc(^{\pi}/_{4})\cot^2(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4})) + \sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4})\ln(\cot(^{\pi}/_{4}))\\
&+ 2\sec(^{\pi}/_{4})\tan^2(^{\pi}/_{4}) - 2\csc(^{\pi}/_{4})\cot^2(^{\pi}/_{4})\\
&+ \sec(^{\pi}/_{4})\csc^2(^{\pi}/_{4}) - \csc(^{\pi}/_{4})\sec^2(^{\pi}/_{4})\\
=&\ 0
\end{aligned}$$

$$\begin{aligned} g^{[xy](3)}(\pi/4) =&\ 5\sec^3(\pi/4)\tan(\pi/4)\ln(\cot(\pi/4)) - 5\csc^3(\pi/4)\cot(\pi/4)\ln(\cot(\pi/4)) \\ &+ \sec(\pi/4)\tan^3(\pi/4)\ln(\cot(\pi/4)) - \csc(\pi/4)\cot^3(\pi/4)\ln(\cot(\pi/4)) \\ &- \sec^4(\pi/4)\csc(\pi/4) - \csc^4(\pi/4)\sec(\pi/4) + 2\sec(\pi/4)\tan^3(\pi/4) \\ &+ 2\csc(\pi/4)\cot^3(\pi/4) + \sec^2(\pi/4)\csc(\pi/4) - 2\csc^3(\pi/4) \\ &+ \csc^2(\pi/4)\sec(\pi/4) - 2\sec^3(\pi/4) \\ =&\ -8\sqrt{2} \end{aligned}$$

Evaluation of derivatives of $g^{[yy]}(\theta)$ at $\frac{\pi}{4}$

$$\begin{aligned} g^{[yy]}(\pi/4) =&\ \csc(\pi/4)\ln(\cot(\pi/4)) + \csc(\pi/4)\cot(\pi/4)\ln(\cot(\pi/4)) + 2\csc(\pi/4) \\ &- 2\csc(\pi/4)\cot(\pi/4) \\ =&\ 2\sqrt{2} - 2\sqrt{2} \\ =&\ 0 \\ g^{[yy]\prime}(\pi/4) =&\ -\csc(\pi/4)\cot(\pi/4)\ln(\cot(\pi/4)) - \csc(\pi/4)\cot^2(\pi/4)\ln(\cot(\pi/4)) \\ &- \csc^3(\pi/4)\ln(\cot(\pi/4)) - \csc^2(\pi/4)\sec(\pi/4) + \csc^3(\pi/4) \\ &+ 2\csc(\pi/4)\cot^2(\pi/4) - 2\csc(\pi/4)\cot(\pi/4) \\ =&\ 0 \\ g^{[yy]\prime\prime}(\pi/4) =&\ 5\csc^3(\pi/4)\cot(\pi/4)\ln(\cot(\pi/4)) + \csc(\pi/4)\cot^3(\pi/4)\ln(\cot(\pi/4)) \\ &+ \csc(\pi/4)\cot^2(\pi/4)\ln(\cot(\pi/4)) + \csc^3(\pi/4)\ln(\cot(\pi/4)) \\ &+ \csc^4(\pi/4)\sec(\pi/4) + 5\csc^3(\pi/4) - \csc(\pi/4)\sec^2(\pi/4) \\ &- 6\csc^3(\pi/4)\cot(\pi/4) - 2\csc(\pi/4)\cot^3(\pi/4) + 2\csc(\pi/4)\cot^2(\pi/4) \\ =&\ 0 \\ g^{[yy](3)}(\pi/4) =&\ -18\csc^3(\pi/4)\cot^2(\pi/4)\ln(\cot(\pi/4)) - 5\csc^5(\pi/4)\ln(\cot(\pi/4)) \\ &- \csc(\pi/4)\cot^4(\pi/4)\ln(\cot(\pi/4)) - \csc(\pi/4)\cot^3(\pi/4)\ln(\cot(\pi/4)) \\ &- 5\csc^3(\pi/4)\cot(\pi/4)\ln(\cot(\pi/4)) - 3\csc^5(\pi/4) - \csc^4(\pi/4)\sec(\pi/4) \\ &+ \csc^3(\pi/4)\sec^2(\pi/4) - 20\csc^3(\pi/4)\cot(\pi/4) + \csc^2(\pi/4)\sec(\pi/4) - 2\sec^3(\pi/4) \\ &+ 23\csc^3(\pi/4)\cot^2(\pi/4) + 2\csc(\pi/4)\cot^4(\pi/4) - 2\csc(\pi/4)\cot^3(\pi/4) \\ =&\ -8\sqrt{2} \end{aligned}$$

The Limit

By l'Hôpital's rule we get the following results:

$$\frac{g^{[xx]}(\pi/4)}{g(\pi/4)} = \frac{g^{[xx](3)}(\pi/4)}{g^{(3)}(\pi/4)} = \frac{-8\sqrt{2}}{-48} = \frac{\sqrt{2}}{6}$$
$$\frac{g^{[xy]}(\pi/4)}{g(\pi/4)} = \frac{g^{[xy](3)}(\pi/4)}{g^{(3)}(\pi/4)} = \frac{-8\sqrt{2}}{-48} = \frac{\sqrt{2}}{6}$$
$$\frac{g^{[yy]}(\pi/4)}{g(\pi/4)} = \frac{g^{[yy](3)}(\pi/4)}{g^{(3)}(\pi/4)} = \frac{-8\sqrt{2}}{-48} = \frac{\sqrt{2}}{6}$$

We can now evaluate the limit of $\nabla^2 l$

First we calculate the limit of the second order derivatives:

$$\begin{aligned}
\lim_{(x,y)\to(c,c)} l''_{xx}(x,y) &= \lim_{r\to r_c}\left[f^{[xx]}(r)\cdot\right]\cdot \lim_{\theta\to\pi/4}\left[\frac{g^{[xx]}(\theta)}{g(\theta)}\right] \\
&= -r_c^{-1}\cdot\frac{\sqrt{2}}{6} \\
&= -\frac{1}{6c} \\
\lim_{(x,y)\to(c,c)} l''_{xy}(x,y) &= \lim_{r\to r_c}\left[f^{[xy]}(r)\cdot\right]\cdot \lim_{\theta\to\pi/4}\left[\frac{g^{[xy]}(\theta)}{g(\theta)}\right] \\
&= r_c^{-1}\cdot\frac{\sqrt{2}}{6} \\
&= \frac{1}{6c} \\
\lim_{(x,y)\to(c,c)} l''_{yy}(x,y) &= \lim_{r\to r_c}\left[f^{[yy]}(r)\cdot\right]\cdot \lim_{\theta\to\pi/4}\left[\frac{g^{[yy]}(\theta)}{g(\theta)}\right] \\
&= -r_c^{-1}\cdot\frac{\sqrt{2}}{6} \\
&= -\frac{1}{6c}
\end{aligned}$$

But c was simply the value of x (and y) hence we get the following result

Let

$$S = \left\{ (x,y)^T \;\middle|\; x,y>0 \right\}$$

The Hessian of the log mean temperature difference, $\nabla^2 l : S \to \mathbb{R}^{2\times 2}$, is defined as

$$\nabla^2 l(x,y) = \begin{cases} \dfrac{1}{6x}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, & x = y \\ y^{-1}\ln(w)^{-2}\, k(w) \begin{pmatrix} -w^{-1} & 1 \\ 1 & -w \end{pmatrix}, & x \neq y \end{cases}$$

where

$$k(w) = 1 + w^{-1} + 2w^{-1}\ln(w)^{-1} - 2\ln(w)^{-1}$$

E Limits of m Gradient and Hessian

The proofs presented here are similar to the proof given in section 3.4 and the limit will be proven via polar coordinates, where the limit being proven is:

$$(r, \theta)^T \to \left(r_c, \frac{\pi}{4}\right)^T$$

where $r_c = c\sqrt{2} > 0$

E.1 Limit of ∇m

The gradient of m is defined as

$$\begin{aligned}
\nabla m\,(x, y) &= \frac{1}{(x-y)^2}\begin{pmatrix} 1 - \ln(w) - w^{-1} \\ 1 + \ln(w) - w \end{pmatrix} \\
&= \begin{pmatrix} m'_x\,(x, y) \\ m'_y\,(x, y) \end{pmatrix}
\end{aligned}$$

Polar Representation

We derive the polar representations of $m'_x\,(x, y)$ and $m'_y\,(x, y)$ separately by making the transformations defined by eqs. (3.2) to (3.4)

$$\begin{aligned}
m'^{[r,\theta]}_x(r, \theta) &= m'_x\,(r\cos(\theta), r\sin(\theta)) \\
&= \frac{1 - \ln(\cot(\theta)) - \tan(\theta)}{(r\cos(\theta) - r\sin(\theta))^2} \\
&= r^{-2} \cdot \frac{1 - \ln(\cot(\theta)) - \tan(\theta)}{(\cos(\theta) - \sin(\theta))^2} \\
&= p(r) \cdot \frac{q^{[x]}(\theta)}{q(\theta)} \\
m'^{[r,\theta]}_y(r, \theta) &= m'_y\,(r\cos(\theta), r\sin(\theta)) \\
&= \frac{1 + \ln(\cot(\theta)) - \cot(\theta)}{(r\cos(\theta) - r\sin(\theta))^2} \\
&= r^{-2} \cdot \frac{1 + \ln(\cot(\theta)) - \cot(\theta)}{(\cos(\theta) - \sin(\theta))^2} \\
&= p(r) \cdot \frac{q^{[y]}(\theta)}{q(\theta)}
\end{aligned}$$

These functions are well defined except possibly for

$$\theta = \frac{\pi}{4}$$

Evaluation of the Derivatives

The derivation of the following derivatives can be found in appendix I

We begin our analysis with $q(\theta)$ as this indicates (for both functions) how many derivatives we require if a limit is to exist.

$$
\begin{aligned}
q(\pi/4) &= (\cos(\pi/4) - \sin(\pi/4))^2 = \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\right)^2 &&= 0 \\
q'(\pi/4) &= -2\cos\left(2\cdot\frac{\pi}{4}\right) = -2\cos\left(\frac{\pi}{2}\right) &&= 0 \\
q''(\pi/4) &= 4\sin\left(2\cdot\frac{\pi}{4}\right) = 4\sin\left(\frac{\pi}{2}\right) &&= 4
\end{aligned}
$$

Here we see that the following are necessary requirements in order for a limit to be proven to exist by l'Hôpital's rule.

$$
\begin{aligned}
q^{[x]}(\pi/4) &= 0, & q^{[x]\prime}(\pi/4) &= 0 \\
q^{[y]}(\pi/4) &= 0, & q^{[y]\prime}(\pi/4) &= 0
\end{aligned}
$$

On evaluation we see that these properties hold.

Evaluation of derivatives of $q^{[x]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
q^{[x]}(\pi/4) &= 1 - \ln\left(\cot(\pi/4)\right) - \tan(\pi/4) &&= 0 \\
q^{[x]\prime}(\pi/4) &= \csc(\pi/4)\sec(\pi/4) - \sec^2(\pi/4) &&= 0 \\
q^{[x]\prime\prime}(\pi/4) &= \sec^2(\pi/4) - \csc^2(\pi/4) - 2\sec^2(\pi/4)\tan(\pi/4) &&= -4
\end{aligned}
$$

Evaluation of derivatives of $q^{[y]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
g^{[y]}(\pi/4) &= 1 + \ln\left(\cot(\pi/4)\right) - \cot(\pi/4) &&= 0 \\
g^{[y]\prime}(\pi/4) &= -\csc(\pi/4)\sec(\pi/4) + \csc^2(\pi/4) &&= 0 \\
g^{[y]\prime\prime}(\pi/4) &= \csc^2(\pi/4) - \sec^2(\pi/4) - 2\csc^2(\pi/4)\cot(\pi/4) &&= -4
\end{aligned}
$$

The Limit

By l'Hôpital's rule we get the following results:

$$
\begin{aligned}
\frac{q^{[x]}(\pi/4)}{q(\pi/4)} &= \frac{q^{[x]\prime\prime}(\pi/4)}{q''(\pi/4)} = \frac{-4}{4} = -1 \\
\frac{q^{[y]}(\pi/4)}{q(\pi/4)} &= \frac{q^{[y]\prime\prime}(\pi/4)}{q''(\pi/4)} = \frac{-4}{4} = -1
\end{aligned}
$$

We can now evaluate the limit of ∇m

First we calculate the limit of the first order derivatives:

$$\begin{aligned}
\lim_{(x,y)\to(c,c)} m'_x(x,y) &= \left[\lim_{r\to r_c} p(r)\right]\cdot\left[\lim_{\theta\to\pi/4}\frac{q^{[x]}(\theta)}{q(\theta)}\right] \\
&= r_c^{-2}\cdot(-1) \\
&= -\frac{1}{2c^2} \\
\lim_{(x,y)\to(c,c)} m'_y(x,y) &= \left[\lim_{r\to r_c} p(r)\right]\cdot\left[\lim_{\theta\to\pi/4}\frac{q^{[y]}(\theta)}{q(\theta)}\right] \\
&= r_c^{-2}\cdot(-1) \\
&= -\frac{1}{2c^2}
\end{aligned}$$

But c was simply the value of x (and y) hence we get the following result

Let

$$S = \left\{ (x,y)^T \;\middle|\; x,y>0 \right\}$$

The gradient of the reciprocal of the log mean temperature difference, $\nabla m : S \to \mathbb{R}^2$, is defined as

$$\nabla m(x,y) = \begin{cases} \begin{pmatrix} -1/2x^2 \\ -1/2x^2 \end{pmatrix}, & x = y \\ \dfrac{1}{(x-y)^2}\begin{pmatrix} 1-\ln(w)-w^{-1} \\ 1+\ln(w)-w \end{pmatrix}, & x \neq y \end{cases}$$

E.2 Limit of $\nabla^2 m$

The Hessian of m is defined as

$$\nabla^2 m(x,y) = \frac{1}{(x-y)^3}\begin{pmatrix} 2\ln(w)+4w^{-1}-w^{-2}-3 & w-w^{-1}-2\ln(w) \\ w-w^{-1}-2\ln(w) & 2\ln(w)-4w+w^2+3 \end{pmatrix}$$

Polar Representation

We begin by deriving the polar representation m''_{xx}, m''_{xy} and m''_{yy} separately by making the transformations defined by eqs. (3.2) to (3.4) of the second order derivatives.

$$\begin{aligned}
m''^{[r,\theta]}_{xx}(r,\theta) &= m''_{xx}(r\cos(\theta), r\sin(\theta)) \\
&= \frac{1}{(r\cos(\theta)-r\sin(\theta))^3}\Big(2\ln(\cot(\theta))+4\tan(\theta)-\tan^2(\theta)-3\Big) \\
&= r^{-3}\cdot\frac{2\ln(\cot(\theta))+4\tan(\theta)-\tan^2(\theta)-3}{(\cos(\theta)-\sin(\theta))^3}
\end{aligned}$$

$$
\begin{aligned}
&= p^{[xx]}(r) \cdot \frac{q^{[xx]}(\theta)}{q(\theta)} \\
m''^{[r,\theta]}_{xy}(r,\theta) &= m''_{xy}(r\cos(\theta), r\sin(\theta)) \\
&= \frac{1}{(r\cos(\theta) - r\sin(\theta))^3}\left(\cot(\theta) - \tan(\theta) - 2\ln\left(\cot(\theta)\right)\right) \\
&= r^{-3} \cdot \frac{\cot(\theta) - \tan(\theta) - 2\ln\left(\cot(\theta)\right)}{(\cos(\theta) - \sin(\theta))^3} \\
&= p^{[xy]}(r) \cdot \frac{q^{[xy]}(\theta)}{q(\theta)} \\
m''^{[r,\theta]}_{yy}(r,\theta) &= m''_{yy}(x,y) \\
&= \frac{1}{(r\cos(\theta) - r\sin(\theta))^3}\Big(2\ln\left(\cot(\theta)\right) - 4\cot(\theta) + \cot^2(\theta) + 3\Big) \\
&= r^{-3} \cdot \frac{2\ln\left(\cot(\theta)\right) - 4\cot(\theta) + \cot^2(\theta) + 3}{(\cos(\theta) - \sin(\theta))^3} \\
&= p^{[yy]}(r) \cdot \frac{q^{[yy]}(\theta)}{q(\theta)}
\end{aligned}
$$

In the above derivations, we see that the resulting polar representations give a set of separable functions parametrised by independent variables, r and θ.

These functions are well defined except possibly for

$$\theta = \frac{\pi}{4}$$

On inspection we find that the θ parametrised parts of m at $\frac{\pi}{4}$ evaluates to a fraction of the form $\frac{0}{0}$. For limits of this form, we can use l'Hôpital's rule, repeatedly if need be.

Evaluation of the Derivatives

On evaluation we find that we require more than simply the first derivatives, the derivation of these can be found in appendix I.

We begin our analysis with $q(\theta)$ as this indicates how many derivatives we require if a limit is to exist. i.e. We continue to take derivatives until the first non-zero is found.

The derivatives of $q(\theta)$ evaluate as follows:

$$
\begin{aligned}
q\left(\pi/4\right) &= \left(\cos\left(\pi/4\right) - \sin\left(\pi/4\right)\right)^3 \\
&= 0 \\
q'\left(\pi/4\right) &= -3\cos\left(2\left(\pi/4\right)\right)\left(\cos\left(\pi/4\right) - \sin\left(\pi/4\right)\right) \\
&= 0 \\
q''\left(\pi/4\right) &= 6\sin\left(2\left(\pi/4\right)\right)\left(\cos\left(\pi/4\right) - \sin\left(\pi/4\right)\right) + 3\cos\left(2\left(\pi/4\right)\right)\left(\sin\left(\pi/4\right) + \cos\left(\pi/4\right)\right) \\
&= 0 \\
q^{(3)}\left(\pi/4\right) &= 15\cos\left(2\left(\pi/4\right)\right)\left(\cos\left(\pi/4\right) - \sin\left(\pi/4\right)\right) - 12\sin\left(2\left(\pi/4\right)\right)\left(\sin\left(\pi/4\right) + \cos\left(\pi/4\right)\right) \\
&= -12\sqrt{2}
\end{aligned}
$$

Here we see that the following are necessary requirements in order for a limit to be proven to exist by l'Hôpital's rule.

$$
\begin{aligned}
&q^{[xx]}(\pi/4) = 0, \qquad && q^{[xx]\prime}(\pi/4) = 0, \qquad && q^{[xx]\prime\prime}(\pi/4) = 0 \\
&q^{[xy]}(\pi/4) = 0, && q^{[xy]\prime}(\pi/4) = 0, && q^{[xy]\prime\prime}(\pi/4) = 0 \\
&q^{[yy]}(\pi/4) = 0, && q^{[yy]\prime}(\pi/4) = 0, && q^{[yy]\prime\prime}(\pi/4) = 0
\end{aligned}
$$

On evaluation we see that these properties holds.

Evaluation of derivatives of $q^{[xx]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
q^{[xx]}\left(\pi/4\right) &= 2\ln\left(\cot\left(\pi/4\right)\right) + 4\tan\left(\pi/4\right) - \tan^2\left(\pi/4\right) - 3 \\
&= 0 \\
q^{[xx]\prime}\left(\pi/4\right) &= -2\csc\left(\pi/4\right)\sec\left(\pi/4\right) + 4\sec^2\left(\pi/4\right) - 2\sec^2\left(\pi/4\right)\tan\left(\pi/4\right) \\
&= 0 \\
q^{[xx]\prime\prime}\left(\pi/4\right) &= 2\csc^2\left(\pi/4\right) - 2\sec^2\left(\pi/4\right) + 8\sec^2\left(\pi/4\right)\tan\left(\pi/4\right) \\
&\quad - 4\sec^2\left(\pi/4\right)\tan^2\left(\pi/4\right) - 2\sec^4\left(\pi/4\right) \\
&= 0 \\
q^{[xx](3)}\left(\pi/4\right) &= 16\sec^2\left(\pi/4\right)\tan^2\left(\pi/4\right) - 4\csc^2\left(\pi/4\right)\cot\left(\pi/4\right) - 4\sec^2\left(\pi/4\right)\tan\left(\pi/4\right) \\
&\quad + 8\sec^4\left(\pi/4\right) - 8\sec^2\left(\pi/4\right)\tan^3\left(\pi/4\right) - 16\sec^4\left(\pi/4\right)\tan\left(\pi/4\right) \\
&= -32
\end{aligned}
$$

Evaluation of derivatives of $q^{[xy]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
q^{[xy]}(\pi/4) &= \cot(\pi/4) - \tan(\pi/4) - 2\ln\left(\cot(\pi/4)\right) \\
&= 0 \\
q^{[xy]\prime}(\pi/4) &= -\csc^2(\pi/4) - \sec^2(\pi/4) + 2\csc(\pi/4)\sec(\pi/4) \\
&= 0 \\
q^{[xy]\prime\prime}(\pi/4) &= 2\csc^2(\pi/4)\cot(\pi/4) - 2\sec^2(\pi/4)\tan(\pi/4) - 2\csc^2(\pi/4) + 2\sec^2(\pi/4) \\
&= 0 \\
q^{[xy](3)}(\pi/4) &= -4\csc^2(\pi/4)\cot^2(\pi/4) - 2\csc^4(\pi/4) - 4\sec^2(\pi/4)\tan^2(\pi/4) \\
&\quad - 2\sec^4(\pi/4) + 4\csc^2(\pi/4)\cot(\pi/4) + 4\sec^2(\pi/4)\tan(\pi/4) \\
&= -16
\end{aligned}
$$

Evaluation of derivatives of $q^{[yy]}(\theta)$ at $\frac{\pi}{4}$

$$
\begin{aligned}
q^{[yy]}(\pi/4) &= 2\ln\left(\cot(\pi/4)\right) - 4\cot(\pi/4) + \cot^2(\pi/4) + 3 \\
&= 0 \\
q^{[yy]\prime}(\pi/4) &= -2\csc(\pi/4)\sec(\pi/4) + 4\csc^2(\pi/4) - 2\csc^2(\pi/4)\cot(\pi/4) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
q^{[yy]\prime\prime}(\pi/4) &= 2\csc^2(\pi/4) - 2\sec^2(\pi/4) - 8\csc^2(\pi/4)\cot(\pi/4) \\
&\quad + 4\csc^2(\pi/4)\cot^2(\pi/4) + 2\csc^4(\pi/4) \\
&= 0 \\
q^{[yy](3)}(\pi/4) &= -4\csc^2(\pi/4)\cot(\pi/4) - 4\sec^2(\pi/4)\tan(\pi/4) + 16\csc^2(\pi/4)\cot^2(\pi/4) \\
&\quad + 8\csc^4(\pi/4) - 8\csc^2(\pi/4)\cot^3(\pi/4) - 16\csc^4(\pi/4)\cot(\pi/4) \\
&= -32
\end{aligned}
$$

The Limit

By l'Hôpital's rule we get the following results:

$$
\begin{aligned}
\frac{q^{[xx]}(\pi/4)}{q(\pi/4)} &= \frac{q^{[xx](3)}(\pi/4)}{q^{(3)}(\pi/4)} = \frac{-32}{-12\sqrt{2}} = \frac{8}{3\sqrt{2}} \\
\frac{q^{[xy]}(\pi/4)}{q(\pi/4)} &= \frac{q^{[xy](3)}(\pi/4)}{q^{(3)}(\pi/4)} = \frac{-16}{-12\sqrt{2}} = \frac{4}{3\sqrt{2}} \\
\frac{q^{[yy]}(\pi/4)}{q(\pi/4)} &= \frac{q^{[yy](3)}(\pi/4)}{q^{(3)}(\pi/4)} = \frac{-32}{-12\sqrt{2}} = \frac{8}{3\sqrt{2}}
\end{aligned}
$$

We can now evaluate the limit of $\nabla^2 m$

First we calculate the limit of the second order derivatives:

$$
\begin{aligned}
\lim_{(x,y)\to(c,c)} m''_{xx}(x,y) &= \lim_{r\to r_c}\left[p^{[xx]}(r)\right]\cdot\lim_{\theta\to\pi/4}\left[\frac{q^{[xx]}(\theta)}{q(\theta)}\right] = r_c^{-3}\cdot\frac{8}{3\sqrt{2}} = \frac{2}{3c^3} \\
\lim_{(x,y)\to(c,c)} m''_{xy}(x,y) &= \lim_{r\to r_c}\left[p^{[xy]}(r)\right]\cdot\lim_{\theta\to\pi/4}\left[\frac{q^{[xy]}(\theta)}{q(\theta)}\right] = r_c^{-3}\cdot\frac{4}{3\sqrt{2}} = \frac{1}{3c^3} \\
\lim_{(x,y)\to(c,c)} m''_{yy}(x,y) &= \lim_{r\to r_c}\left[p^{[yy]}(r)\right]\cdot\lim_{\theta\to\pi/4}\left[\frac{q^{[yy]}(\theta)}{q(\theta)}\right] = r_c^{-3}\cdot\frac{8}{3\sqrt{2}} = \frac{2}{3c^3}
\end{aligned}
$$

But c was simply the value of x (and y) hence we get the following result

Let

$$
S = \left\{\, (x,y)^T \;\middle|\; x, y > 0 \,\right\}
$$

The Hessian of the reciprocal of the log mean temperature difference, $\nabla^2 m : S \to \mathbb{R}^{2\times 2}$, is defined as

$$
\nabla^2 m(x,y) = \begin{cases} \dfrac{1}{3x^3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, & x = y \\ \dfrac{1}{(x-y)^3}\begin{pmatrix} 2\ln(w) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln(w) \\ w - w^{-1} - 2\ln(w) & 2\ln(w) - 4w + w^2 + 3 \end{pmatrix}, & x \neq y \end{cases}
$$

F Derivation of l Gradient and Hessian

Throughout this derivation we make use of theorem 2.2.
Function definition:

$$l\left(x,y\right)=\frac{x-y}{\ln\left(^{x}/_{y}\right)}$$

F.1 Gradient Derivation

Partial derivative of l w.r.t x

$$u=x-y,\quad u'_x=1,\quad v=\ln\left(\frac{x}{y}\right),\quad v'_x=\frac{1}{x}$$

$$\begin{aligned}
l'_x\left(x,y\right)&=\frac{1}{\ln\left(^{x}/_{y}\right)^2}\left(\ln\left(\frac{x}{y}\right)-\frac{1}{x}\left(x-y\right)\right)\\
&=\frac{1}{\ln\left(^{x}/_{y}\right)^2}\left(\ln\left(\frac{x}{y}\right)+\frac{y}{x}-1\right)\\
&=\ln\left(w\right)^{-2}\left(w^{-1}+\ln\left(w\right)-1\right)
\end{aligned}$$

Partial derivative of l w.r.t y

$$u=x-y,\quad u'_y=-1,\quad v=\ln\left(\frac{x}{y}\right),\quad v'_y=-\frac{1}{y}$$

$$\begin{aligned}
l'_y\left(x,y\right)&=\frac{1}{\ln\left(^{x}/_{y}\right)^2}\left(-\ln\left(\frac{x}{y}\right)+\frac{1}{y}\left(x-y\right)\right)\\
&=\frac{1}{\ln\left(^{x}/_{y}\right)^2}\left(\frac{x}{y}-1-\ln\left(\frac{x}{y}\right)\right)\\
&=\ln\left(w\right)^{-2}\left(w-\ln\left(w\right)-1\right)
\end{aligned}$$

F.1.1 Gradient of l

$$\nabla l\left(x,y\right)=\ln\left(w\right)^{-2}\begin{pmatrix}w^{-1}+\ln\left(w\right)-1\\ w-\ln\left(w\right)-1\end{pmatrix}$$

F.2 Hessian Derivation

Second partial derivative of l w.r.t x

The partial derivative of l w.r.t x is

$$l'_x(x,y) = \frac{{}^{y}/_{x} + \ln\left({}^{x}/_{y}\right) - 1}{\ln\left({}^{x}/_{y}\right)^2}$$

we will differentiate this again w.r.t. x.

We split the sums and differentiate them separately, we will then form the second partial derivative.

- For $\ln\left({}^{x}/_{y}\right)^{-1}$

$$\begin{aligned}\left(\ln\left(\frac{x}{y}\right)^{-1}\right)'_x &= -\frac{1}{x}\ln\left(\frac{x}{y}\right)^{-2} \\ &= -x^{-1}\ln(w)^{-2}\end{aligned}$$

- For $\ln\left({}^{x}/_{y}\right)^{-2}$

$$\begin{aligned}\left(\ln\left(\frac{x}{y}\right)^{-2}\right)'_x &= -2\cdot\frac{1}{x}\ln\left(\frac{x}{y}\right)^{-3} \\ &= -2x^{-1}\ln(w)^{-3}\end{aligned}$$

- For $\frac{{}^{y}/_{x}}{\ln\left({}^{x}/_{y}\right)^2}$

$$\begin{aligned} u &= \frac{y}{x} = w^{-1}, & u'_x &= -\frac{y}{x^2} = -w^{-1}x^{-1} \\ v &= \ln\left(\frac{x}{y}\right)^2 = \ln(w)^2, & v'_x &= \frac{2}{x}\ln\left(\frac{x}{y}\right) = 2x^{-1}\ln(w)\end{aligned}$$

$$\begin{aligned}\left(\frac{{}^{y}/_{x}}{\ln\left({}^{x}/_{y}\right)^2}\right)'_x &= \frac{-x^{-1}w^{-1}\ln(w)^2 - 2x^{-1}w^{-1}\ln(w)}{\ln(w)^4} \\ &= \frac{-x^{-1}w^{-1}\ln(w) - 2x^{-1}w^{-1}}{\ln(w)^3} \\ &= -x^{-1}w^{-1}\ln(w)^{-2} - 2x^{-1}w^{-1}\ln(w)^{-3}\end{aligned}$$

Hence the second partial derivative of l w.r.t x is

$$\begin{aligned} l''_{xx}(x,y) &= -x^{-1}w^{-1}\ln(w)^{-2} - 2x^{-1}w^{-1}\ln(w)^{-3} + \left(-x^{-1}\ln(w)^{-2}\right) - \left(-2x^{-1}\ln(w)^{-3}\right) \\ &= -x^{-1}\ln(w)^{-2}\left(1 + w^{-1} + 2w^{-1}\ln(w)^{-1} - 2\ln(w)^{-1}\right) \\ &= -y^{-1}\ln(w)^{-2}\,w^{-1}\left(1 + w^{-1} + 2w^{-1}\ln(w)^{-1} - 2\ln(w)^{-1}\right)\end{aligned}$$

Second partial derivative of l w.r.t y

The partial derivative of l w.r.t y is

$$l'_y(x,y) = \frac{{}^x/_y - \ln({}^x/_y) - 1}{\ln({}^x/_y)^2}$$

we will differentiate this again w.r.t. y.

We split the sums and differentiate them separately, we will then form the second partial derivative.

- For $\frac{{}^x/_y}{\ln({}^x/_y)^2}$

$$u = \frac{x}{y} = w, \qquad u'_y = -\frac{x}{y^2} = -wy^{-1}$$
$$v = \ln\left(\frac{x}{y}\right)^2 = \ln(w)^2, \qquad v'_y = -\frac{2}{y}\ln\left(\frac{x}{y}\right) = -2y^{-1}\ln(w)$$

$$\begin{aligned}\left(\frac{{}^x/_y}{\ln({}^x/_y)^2}\right)'_y &= \frac{-wy^{-1}\ln(w)^2 + 2wy^{-1}\ln(w)}{\ln(w)^4}\\ &= \frac{2wy^{-1} - wy^{-1}\ln(w)}{\ln(w)^3}\\ &= 2wy^{-1}\ln(w)^{-3} - wy^{-1}\ln(w)^{-2}\end{aligned}$$

- For $\ln({}^x/_y)^{-1}$

$$\begin{aligned}\left(\ln\left(\frac{x}{y}\right)^{-1}\right)'_y &= y^{-1}\ln\left(\frac{x}{y}\right)^{-2}\\ &= y^{-1}\ln(w)^{-2}\end{aligned}$$

- For $\ln({}^x/_y)^{-2}$

$$\begin{aligned}\left(\ln\left(\frac{x}{y}\right)^{-2}\right)'_y &= 2y^{-1}\ln\left(\frac{x}{y}\right)^{-3}\\ &= 2y^{-1}\ln(w)^{-3}\end{aligned}$$

Hence the second partial derivative of l w.r.t y is

$$\begin{aligned}l''_{yy}(x,y) &= 2wy^{-1}\ln(w)^{-3} - wy^{-1}\ln(w)^{-2} - y^{-1}\ln(w)^{-2} - 2y^{-1}\ln(w)^{-3}\\ &= 2wy^{-1}\ln(w)^{-3} - wy^{-1}\ln(w)^{-2} - y^{-1}\ln(w)^{-2} - 2y^{-1}\ln(w)^{-3}\\ &= -y^{-1}\ln(w)^{-2}\,w\left(1 + w^{-1} + 2w^{-1}\ln(w)^{-1} - 2\ln(w)^{-1}\right)\end{aligned}$$

Mixed partial derivative of l w.r.t x and y

The partial derivative of l w.r.t x is

$$l'_x(x,y) = \frac{{}^y/_x + \ln\left({}^x/_y\right) - 1}{\ln\left({}^x/_y\right)^2}$$

we will differentiate this again w.r.t. y.

We split the sums and differentiate them separately, we will then form the mixed partial derivative.

- For $\frac{{}^y/_x}{\ln({}^x/_y)^2}$

$$\begin{aligned} u &= \frac{y}{x} = w^{-1}, & u'_y &= \frac{1}{x} = x^{-1} \\ v &= \ln\left(\frac{x}{y}\right)^2 = \ln(w)^2, & v'_y &= -\frac{2}{y}\ln\left(\frac{x}{y}\right) = -2y^{-1}\ln(w) \end{aligned}$$

$$\begin{aligned} \left(\frac{{}^y/_x}{\ln({}^x/_y)^2}\right)'_y &= \frac{x^{-1}\ln(w)^2 + 2w^{-1}y^{-1}\ln(w)}{\ln(w)^4} \\ &= \frac{x^{-1}\ln(w) + 2w^{-1}y^{-1}}{\ln(w)^3} \\ &= x^{-1}\ln(w)^{-2} + 2w^{-1}y^{-1}\ln(w)^{-3} \end{aligned}$$

- For $\ln\left({}^x/_y\right)^{-1}$

$$\begin{aligned} \left(\ln\left(\frac{x}{y}\right)^{-1}\right)'_y &= \frac{1}{y}\ln\left(\frac{x}{y}\right)^{-2} \\ &= y^{-1}\ln(w)^{-2} \end{aligned}$$

- For $\ln\left({}^x/_y\right)^{-2}$

$$\begin{aligned} \left(\ln\left(\frac{x}{y}\right)^{-2}\right)'_y &= \frac{2}{y}\ln\left(\frac{x}{y}\right)^{-3} \\ &= 2y^{-1}\ln(w)^{-3} \end{aligned}$$

Hence the mixed partial derivative of l w.r.t x and y is

$$\begin{aligned} l''_{xy}(x,y) &= x^{-1}\ln(w)^{-2} + 2w^{-1}y^{-1}\ln(w)^{-3} + y^{-1}\ln(w)^{-2} - 2y^{-1}\ln(w)^{-3} \\ &= y^{-1}\ln(w)^{-2}\left(1 + w^{-1} + 2w^{-1}\ln(w)^{-1} - 2\ln(w)^{-1}\right) \end{aligned}$$

F.2.1 Hessian of l

We have:

$$\begin{aligned}
l''_{xx}(x,y) &= -y^{-1}\ln(w)^{-2}w^{-1}\left(1+w^{-1}+2w^{-1}\ln(w)^{-1}-2\ln(w)^{-1}\right)\\
l''_{yy}(x,y) &= -y^{-1}\ln(w)^{-2}w\left(1+w^{-1}+2w^{-1}\ln(w)^{-1}-2\ln(w)^{-1}\right)\\
l''_{xy}(x,y) &= y^{-1}\ln(w)^{-2}\left(1+w^{-1}+2w^{-1}\ln(w)^{-1}-2\ln(w)^{-1}\right)\\
l''_{yx}(x,y) &= l''_{xy}(x,y)
\end{aligned}$$

If we let

$$k(w) = 1+w^{-1}+2w^{-1}\ln(w)^{-1}-2\ln(w)^{-1}$$

we get

$$\begin{aligned}
l''_{xx}(x,y) &= -y^{-1}\ln(w)^{-2}k(w)\,w^{-1}\\
l''_{yy}(x,y) &= -y^{-1}\ln(w)^{-2}k(w)\,w\\
l''_{xy}(x,y) &= y^{-1}\ln(w)^{-2}k(w)\\
l''_{yx}(x,y) &= y^{-1}\ln(w)^{-2}k(w)
\end{aligned}$$

This gives a Hessian of

$$\begin{aligned}
\nabla^2 l(x,y) &= \begin{pmatrix} l''_{xx}(x,y) & l''_{xy}(x,y)\\ l''_{yx}(x,y) & l''_{yy}(x,y)\end{pmatrix}\\
&= y^{-1}\ln(w)^{-2}k(w)\begin{pmatrix} -w^{-1} & 1\\ 1 & -w\end{pmatrix}
\end{aligned}$$

where

$$k(w) = 1+w^{-1}+2w^{-1}\ln(w)^{-1}-2\ln(w)^{-1}$$

G Derivation of m Gradient and Hessian

Throughout this derivation we make use of theorem 2.2.
Function:

$$m\left(x,y\right)=\frac{\ln\left({}^{x}\!/_{y}\right)}{x-y}$$

Defining $w=\frac{x}{y}$, we get:

$$m\left(x,y\right)=\frac{\ln\left(w\right)}{y\left(w-1\right)}$$

G.1 Gradient Derivation

Partial derivative of m w.r.t x

$$u=\ln\left(\frac{x}{y}\right),\quad u'_x=\frac{1}{x},\quad v=x-y,\quad v'_x=1$$

$$\begin{aligned}m'_x(x,y)&=\frac{\frac{1}{x}\left(x-y\right)-\ln\left({}^{x}\!/_{y}\right)}{\left(x-y\right)^2}\\&=\frac{1-{}^{y}\!/_{x}-\ln({}^{x}\!/_{y})}{y^2\left({}^{x}\!/_{y}-1\right)^2}\\&=\frac{1-w^{-1}-\ln\left(w\right)}{y^2\left(w-1\right)^2}\end{aligned}$$

Partial derivative of m w.r.t y

$$u=\ln\left(\frac{x}{y}\right),\quad u'_y=-\frac{1}{y},\quad v=x-y,\quad v'_y=-1$$

$$\begin{aligned}m'_y\left(x,y\right)&=\frac{-\frac{1}{y}\left(x-y\right)+\ln\left({}^{x}\!/_{y}\right)}{\left(x-y\right)^2}\\&=\frac{1+\ln\left({}^{x}\!/_{y}\right)-{}^{x}\!/_{y}}{\left(x-y\right)^2}\\&=\frac{1+\ln\left(w\right)-w}{y^2\left(w-1\right)^2}\end{aligned}$$

G.1.1 Gradient of m

$$\nabla m\left(x,y\right)=\frac{1}{y^{2}\left(w-1\right)^{2}}\begin{pmatrix}1-\ln\left(w\right)-w^{-1}\\ 1+\ln\left(w\right)-w\end{pmatrix}$$
$$=\frac{1}{\left(x-y\right)^{2}}\begin{pmatrix}1-\ln\left(w\right)-w^{-1}\\ 1+\ln\left(w\right)-w\end{pmatrix}$$

G.2 Hessian Derivation

Second partial derivative of m w.r.t x

The partial derivative of m w.r.t x is

$$m'_{x}\left(x,y\right)=\frac{1-{}^{y}\!/_{x}-\ln({}^{x}\!/_{y})}{\left(x-y\right)^{2}}$$

we will differentiate this again w.r.t. x.

We split the sums and differentiate them separately, we will then form the second partial derivative.

- For $\left(x-y\right)^{-2}$

$$\left(\left(x-y\right)^{-2}\right)'_{x}=-2\left(x-y\right)^{-3}$$

- For $\frac{{}^{y}\!/_{x}}{\left(x-y\right)^{2}}$

$$u=\frac{y}{x},\quad u'_{x}=-\frac{y}{x^{2}},\quad v=\left(x-y\right)^{2},\quad v'_{x}=2\left(x-y\right)$$

$$\left(\frac{{}^{y}\!/_{x}}{\left(x-y\right)^{2}}\right)'_{x}=\frac{1}{\left(x-y\right)^{4}}\left(-\frac{y}{x^{2}}\left(x-y\right)^{2}-2\cdot\frac{y}{x}\left(x-y\right)\right)$$
$$=\frac{1}{\left(x-y\right)^{3}}\left(-\frac{y}{x^{2}}\left(x-y\right)-2\cdot\frac{y}{x}\right)$$
$$=\frac{1}{\left(x-y\right)^{3}}\left(\frac{y^{2}}{x^{2}}-3\cdot\frac{y}{x}\right)$$

- For $\frac{\ln({}^{x}\!/_{y})}{\left(x-y\right)^{2}}$

$$u=\ln\left(\frac{x}{y}\right),\quad u'_{x}=\frac{1}{x},\quad v=\left(x-y\right)^{2},\quad v'_{x}=2\left(x-y\right)$$

$$\left(\frac{\ln\left({}^{x}\!/_{y}\right)}{(x-y)^2}\right)'_x = \frac{1}{(x-y)^4}\left(\frac{1}{x}(x-y)^2 - 2\ln\left(\frac{x}{y}\right)(x-y)\right)$$
$$= \frac{1}{(x-y)^3}\left(\frac{1}{x}(x-y) - 2\ln\left(\frac{x}{y}\right)\right)$$
$$= \frac{1}{(x-y)^3}\left(1 - \frac{y}{x} - 2\ln\left(\frac{x}{y}\right)\right)$$

Hence the second partial derivative of m w.r.t x is

$$m''_{xx}(x,y) = \frac{1}{(x-y)^3}\left[-2 - \left(\frac{y^2}{x^2} - 3\cdot\frac{y}{x}\right) - \left(1 - \frac{y}{x} - 2\ln\left(\frac{y}{x}\right)\right)\right]$$
$$= \frac{1}{(x-y)^3}\left[4\cdot\frac{y}{x} - \frac{y^2}{x^2} - 3 + 2\ln\left(\frac{y}{x}\right)\right]$$
$$= \frac{1}{(x-y)^3}\left[4w^{-1} - w^{-2} - 3 + 2\ln(w)\right]$$

Second partial derivative of m w.r.t y

The partial derivative of m w.r.t y is

$$m'_y(x,y) = \frac{1 + \ln\left({}^{x}\!/_{y}\right) - {}^{x}\!/_{y}}{(x-y)^2}$$

we will differentiate this again w.r.t. y.

We split the sums and differentiate them separately, we will then form the second partial derivative.

- For $(x-y)^{-2}$

$$\left((x-y)^{-2}\right)'_y = 2(x-y)^{-3}$$

- For $\frac{{}^{x}\!/_{y}}{(x-y)^2}$

$$u = \frac{x}{y},\quad u'_y = -\frac{x}{y^2},\quad v = (x-y)^2,\quad v'_y = -2(x-y)$$

$$\left(\frac{{}^{x}\!/_{y}}{(x-y)^2}\right)'_y = \frac{1}{(x-y)^4}\left(-\frac{x}{y^2}(x-y)^2 + 2\cdot\frac{x}{y}(x-y)\right)$$
$$= \frac{1}{(x-y)^3}\left(2\cdot\frac{x}{y} - \frac{x}{y^2}(x-y)\right)$$
$$= \frac{1}{(x-y)^3}\left(3\cdot\frac{x}{y} - \frac{x^2}{y^2}\right)$$

- For $\frac{\ln(x/y)}{(x-y)^2}$

$$u = \ln\left(\frac{x}{y}\right), \quad u'_y = -\frac{1}{y}, \quad v = (x-y)^2, \quad v'_y = -2(x-y)$$

$$\begin{aligned}\left(\frac{\ln(x/y)}{(x-y)^2}\right)'_y &= \frac{1}{(x-y)^4}\left(-\frac{1}{y}(x-y)^2 + 2\ln\left(\frac{x}{y}\right)(x-y)\right)\\ &= \frac{1}{(x-y)^3}\left(2\ln\left(\frac{x}{y}\right) - \frac{1}{y}(x-y)\right)\\ &= \frac{1}{(x-y)^3}\left(2\ln\left(\frac{x}{y}\right) - \frac{x}{y} + 1\right)\end{aligned}$$

Hence the second partial derivative of m w.r.t y is

$$\begin{aligned}m''_{yy}(x,y) = \frac{1}{(x-y)^3}\left[2 + \left(2\ln\left(\frac{x}{y}\right) - \frac{x}{y} + 1\right) - \left(3\cdot\frac{x}{y} - \frac{x^2}{y^2}\right)\right]\\ = \frac{1}{(x-y)^3}\left[3 + 2\ln\left(\frac{x}{y}\right) - 4\cdot\frac{x}{y} + \frac{x^2}{y^2}\right]\\ = \frac{1}{(x-y)^3}\left[3 + 2\ln(w) - 4w + w^2\right]\end{aligned}$$

Mixed partial derivative of m w.r.t x and y

The partial derivative of m w.r.t x is

$$m'_x(x,y) = \frac{1 - y/x - \ln(x/y)}{(x-y)^2}$$

we will differentiate this again w.r.t. y.

We split the sums and differentiate them separately, we will then form the mixed partial derivative.

- For $(x-y)^{-2}$

$$\left((x-y)^{-2}\right)'_y = 2(x-y)^{-3}$$

- For $\frac{y/x}{(x-y)^2}$

$$u = \frac{y}{x}, \quad u'_y = \frac{1}{x}, \quad v = (x-y)^2, \quad v'_y = -2(x-y)$$

$$\left(\frac{y/x}{(x-y)^2}\right)'_y = \frac{1}{(x-y)^4}\left(\frac{1}{x}(x-y)^2 + 2\cdot\frac{y}{x}(x-y)\right)$$

$$= \frac{1}{(x-y)^3}\left(\frac{1}{x}(x-y) + 2 \cdot \frac{y}{x}\right)$$
$$= \frac{1}{(x-y)^3}\left(1 + \frac{y}{x}\right)$$

- For $\frac{\ln({}^{x}/_{y})}{(x-y)^2}$

$$u = \ln\left(\frac{x}{y}\right), \quad u'_y = -\frac{1}{y}, \quad v = (x-y)^2, \quad v'_y = -2(x-y)$$

$$\left(\frac{\ln({}^{x}/_{y})}{(x-y)^2}\right)'_y = \frac{1}{(x-y)^4}\left(-\frac{1}{y}(x-y)^2 + 2\ln\left(\frac{x}{y}\right)(x-y)\right)$$
$$= \frac{1}{(x-y)^3}\left(2\ln\left(\frac{x}{y}\right) - \frac{1}{y}(x-y)\right)$$
$$= \frac{1}{(x-y)^3}\left(2\ln\left(\frac{x}{y}\right) - \frac{x}{y} + 1\right)$$

Hence the mixed partial derivative of m w.r.t x and y is

$$m''_{xy}(x,y) = \frac{1}{(x-y)^3}\left[2 - \left(1 + \frac{y}{x}\right) - \left(2\ln\left(\frac{x}{y}\right) - \frac{x}{y} + 1\right)\right]$$
$$= \frac{1}{(x-y)^3}\left[\frac{x}{y} - \frac{y}{x} - 2\ln\left(\frac{x}{y}\right)\right]$$
$$= \frac{1}{(x-y)^3}\left[w - w^{-1} - 2\ln(w)\right]$$

G.2.1 Hessian of m

We have:

$$m''_{xx}(x,y) = \frac{1}{(x-y)^3}\left(2\ln(w) + 4w^{-1} - w^{-2} - 3\right)$$
$$m''_{yy}(x,y) = \frac{1}{(x-y)^3}\left(2\ln(w) - 4w + w^2 + 3\right)$$
$$m''_{xy}(x,y) = \frac{1}{(x-y)^3}\left(w - w^{-1} - 2\ln(w)\right)$$
$$m''_{yx}(x,y) = m''_{xy}(x,y)$$

Hence the Hessian is:

$$\nabla^2 m(x,y) = \begin{pmatrix} m''_{xx}(x,y) & m''_{xy}(x,y) \\ m''_{yx}(x,y) & m''_{yy}(x,y) \end{pmatrix}$$
$$= \frac{1}{(x-y)^3}\begin{pmatrix} 2\ln(w) + 4w^{-1} - w^{-2} - 3 & w - w^{-1} - 2\ln(w) \\ w - w^{-1} - 2\ln(w) & 2\ln(w) - 4w + w^2 + 3 \end{pmatrix}$$

H l Polar Representation Derivatives of θ functions

Through these derivations we make use of theorem 2.2.

Derivatives of Polar Representation of l - θ functions

We have that the polar representation of l is

$$l_{r,\theta}(r,\theta) = f(r) \cdot \frac{g_1(\theta)}{g_2(\theta)}$$

where

$$\begin{aligned} g_1(\theta) &= \cos(\theta) - \sin(\theta) \\ g_2(\theta) &= \ln(\cot(\theta)) \end{aligned}$$

The derivatives of these functions are:

$$\begin{aligned} g_1'(\theta) &= -\sin(\theta) - \cos(\theta) \\ g_2'(\theta) &= -\csc(\theta)\sec(\theta) \end{aligned}$$

Derivatives of Polar Representation of ∇l - θ functions

We begin with the derivative of $g(\theta)$ as it is the denominator of both functions.

$$\begin{aligned} g(\theta) =&\ \ln(\cot(\theta))^2 \\ g'(\theta) =&\ -2\csc(\theta)\sec(\theta)\ln(\cot(\theta)) \\ g''(\theta) =&\ 2\csc^2(\theta)\ln(\cot(\theta)) - 2\sec^2(\theta)\ln(\cot(\theta)) + 2\csc^2(\theta)\sec^2(\theta) \end{aligned}$$

We now calculate the derivatives of $g^{[x]}(\theta)$ and $g^{[y]}(\theta)$, the numerators of the limits that we are trying to prove.

$$\begin{aligned} g^{[x]}(\theta) =&\ \tan(\theta) + \ln(\cot(\theta)) - 1 \\ g^{[x]\prime}(\theta) =&\ \sec^2(\theta) - \csc(\theta)\sec(\theta) \\ g^{[x]\prime\prime}(\theta) =&\ 2\sec^2(\theta)\tan(\theta) + \csc^2(\theta) - \sec^2(\theta) \end{aligned}$$

$$\begin{aligned} g^{[y]}(\theta) =&\ \cot(\theta) - \ln(\cot(\theta)) - 1 \\ g^{[y]\prime}(\theta) =&\ -\csc^2(\theta) + \csc(\theta)\sec(\theta) \\ g^{[y]\prime\prime}(\theta) =&\ 2\csc^2(\theta)\cot(\theta) - \csc^2(\theta) + \sec^2(\theta) \end{aligned}$$

Derivatives of Polar Representation of $\nabla^2 l$ - θ functions

We begin with the derivatives of $g(\theta)$ as it is the denominator of all three functions.

$$g(\theta) = \ln(\cot(\theta))^3$$

$$
\begin{aligned}
g'(\theta) =& -3\csc(\theta)\sec(\theta)\ln(\cot(\theta))^2 \\
g''(\theta) =& 3\csc^2(\theta)\ln(\cot(\theta))^2 - 3\sec^2(\theta)\ln(\cot(\theta))^2 + 6\csc^2(\theta)\sec^2(\theta)\ln(\cot(\theta)) \\
g^{(3)}(\theta) =& 18\sec^3(\theta)\csc(\theta)\ln(\cot(\theta)) - 18\csc^3(\theta)\sec(\theta)\ln(\cot(\theta)) - 6\csc^2(\theta)\cot(\theta)\ln(\cot(\theta))^2 \\
& - 6\sec^2(\theta)\tan(\theta)\ln(\cot(\theta))^2 - 6\csc^3(\theta)\sec^3(\theta)
\end{aligned}
$$

We now calculate the derivatives of $g^{[xx]}(\theta)$, $g^{[xy]}(\theta)$ and $g^{[yy]}(\theta)$, the numerators of the limits that we are trying to prove.

$$
\begin{aligned}
g^{[xx]}(\theta) =& \sec(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan(\theta)\ln(\cot(\theta)) + 2\sec(\theta)\tan(\theta) - 2\sec(\theta) \\
g^{[xx]\prime}(\theta) =& \sec^3(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan^2(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan(\theta)\ln(\cot(\theta)) \\
& - \sec^2(\theta)\csc(\theta) + \sec^3(\theta) + 2\sec(\theta)\tan^2(\theta) - 2\sec(\theta)\tan(\theta) \\
g^{[xx]\prime\prime}(\theta) =& 5\sec^3(\theta)\tan(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan^3(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan^2(\theta)\ln(\cot(\theta)) \\
& + \sec^3(\theta)\ln(\cot(\theta)) - \sec^4(\theta)\csc(\theta) - 5\sec^3(\theta) + \sec(\theta)\csc^2(\theta) \\
& + 6\sec^3(\theta)\tan(\theta) + 2\sec(\theta)\tan^3(\theta) - 2\sec(\theta)\tan^2(\theta) \\
g^{[xx](3)}(\theta) =& 18\sec^3(\theta)\tan^2(\theta)\ln(\cot(\theta)) + 5\sec^5(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan^4(\theta)\ln(\cot(\theta)) \\
& + \sec(\theta)\tan^3(\theta)\ln(\cot(\theta)) + 5\sec^3(\theta)\tan(\theta)\ln(\cot(\theta)) - 3\sec^5(\theta) \\
& - \sec^4(\theta)\csc(\theta) + \sec^3(\theta)\csc^2(\theta) - 20\sec^3(\theta)\tan(\theta) + \sec^2(\theta)\csc(\theta) \\
& - 2\csc^3(\theta) + 23\sec^3(\theta)\tan^2(\theta) + 2\sec(\theta)\tan^4(\theta) - 2\sec(\theta)\tan^3(\theta)
\end{aligned}
$$

$$
\begin{aligned}
g^{[xy]}(\theta) =& \csc(\theta)\ln(\cot(\theta)) + \sec(\theta)\ln(\cot(\theta)) + 2\sec(\theta) - 2\csc(\theta) \\
g^{[xy]\prime}(\theta) =& \sec(\theta)\tan(\theta)\ln(\cot(\theta)) - \csc(\theta)\cot(\theta)\ln(\cot(\theta)) - \csc^2(\theta)\sec(\theta) \\
& - \sec^2(\theta)\csc(\theta) + 2\sec(\theta)\tan(\theta) + 2\csc(\theta)\cot(\theta) \\
g^{[xy]\prime\prime}(\theta) =& \csc^3(\theta)\ln(\cot(\theta)) + \sec^3(\theta)\ln(\cot(\theta)) - \sec^3(\theta) + \csc^3(\theta) \\
& + \csc(\theta)\cot^2(\theta)\ln(\cot(\theta)) + \sec(\theta)\tan^2(\theta)\ln(\cot(\theta)) + 2\sec(\theta)\tan^2(\theta) \\
& - 2\csc(\theta)\cot^2(\theta) + \sec(\theta)\csc^2(\theta) - \csc(\theta)\sec^2(\theta) \\
g^{[xy](3)}(\theta) =& 5\sec^3(\theta)\tan(\theta)\ln(\cot(\theta)) - 5\csc^3(\theta)\cot(\theta)\ln(\cot(\theta)) \\
& + \sec(\theta)\tan^3(\theta)\ln(\cot(\theta)) - \csc(\theta)\cot^3(\theta)\ln(\cot(\theta)) \\
& - \sec^4(\theta)\csc(\theta) - \csc^4(\theta)\sec(\theta) + 2\sec(\theta)\tan^3(\theta) + 2\csc(\theta)\cot^3(\theta) \\
& + \sec^2(\theta)\csc(\theta) - 2\csc^3(\theta) + \csc^2(\theta)\sec(\theta) - 2\sec^3(\theta)
\end{aligned}
$$

$$
\begin{aligned}
g^{[yy]}(\theta) =& \csc(\theta)\ln(\cot(\theta)) + \csc(\theta)\cot(\theta)\ln(\cot(\theta)) + 2\csc(\theta) - 2\csc(\theta)\cot(\theta) \\
g^{[yy]\prime}(\theta) =& -\csc(\theta)\cot(\theta)\ln(\cot(\theta)) - \csc(\theta)\cot^2(\theta)\ln(\cot(\theta)) - \csc^3(\theta)\ln(\cot(\theta)) \\
& - \csc^2(\theta)\sec(\theta) + \csc^3(\theta) + 2\csc(\theta)\cot^2(\theta) - 2\csc(\theta)\cot(\theta) \\
g^{[yy]\prime\prime}(\theta) =& 5\csc^3(\theta)\cot(\theta)\ln(\cot(\theta)) + \csc(\theta)\cot^3(\theta)\ln(\cot(\theta)) + \csc(\theta)\cot^2(\theta)\ln(\cot(\theta)) \\
& + \csc^3(\theta)\ln(\cot(\theta)) + \csc^4(\theta)\sec(\theta) + 5\csc^3(\theta) - \csc(\theta)\sec^2(\theta) \\
& - 6\csc^3(\theta)\cot(\theta) - 2\csc(\theta)\cot^3(\theta) + 2\csc(\theta)\cot^2(\theta) \\
g^{[yy](3)}(\theta) =& -18\csc^3(\theta)\cot^2(\theta)\ln(\cot(\theta)) - 5\csc^5(\theta)\ln(\cot(\theta)) - \csc(\theta)\cot^4(\theta)\ln(\cot(\theta)) \\
& - \csc(\theta)\cot^3(\theta)\ln(\cot(\theta)) - 5\csc^3(\theta)\cot(\theta)\ln(\cot(\theta)) - 3\csc^5(\theta)
\end{aligned}
$$

$$- \csc^4(\theta)\sec(\theta) + \csc^3(\theta)\sec^2(\theta) - 20\csc^3(\theta)\cot(\theta) + \csc^2(\theta)\sec(\theta)$$
$$- 2\sec^3(\theta) + 23\csc^3(\theta)\cot^2(\theta) + 2\csc(\theta)\cot^4(\theta) - 2\csc(\theta)\cot^3(\theta)$$

I m Polar Representation Derivatives of θ functions

Through these derivations we make use of theorem 2.2.

Derivatives of Polar Representation of m - θ functions

We have that the polar representation of m is

$$m^{[r,\theta]}(r,\theta) = p(r)\cdot\frac{q_1(\theta)}{q_2(\theta)}$$

where

$$\begin{aligned} q_1(\theta) &= \ln(\cot(\theta)) \\ q_2(\theta) &= \cos(\theta) - \sin(\theta) \end{aligned}$$

The derivatives of these functions are:

$$\begin{aligned} q_1'(\theta) &= -\csc(\theta)\sec(\theta) \\ q_2'(\theta) &= -\sin(\theta) - \cos(\theta) \end{aligned}$$

Derivatives of Polar Representation of ∇m - θ functions

We begin with the derivative of $q(\theta)$ as it is the denominator of both functions.

$$\begin{aligned} q(\theta) &= (\cos(\theta) - \sin(\theta))^2 \\ q'(\theta) &= 2(-\sin(\theta) - \cos(\theta))(\cos(\theta) - \sin(\theta)) \\ &= -2\left(\cos^2(\theta) - \sin^2(\theta)\right) \\ &= -2\cos(2\theta) \\ q''(\theta) &= 4\sin(2\theta) \end{aligned}$$

We now calculate the derivatives of $q^{[x]}(\theta)$ and $q^{[y]}(\theta)$, the numerators of the limits that we are trying to prove.

$$\begin{aligned} q^{[x]}(\theta) &= 1 - \ln(\cot(\theta)) - \tan(\theta) \\ q^{[x]\prime}(\theta) &= \csc(\theta)\sec(\theta) - \sec^2(\theta) \\ q^{[x]\prime\prime}(\theta) &= -\csc(\theta)\cot(\theta)\sec(\theta) + \csc(\theta)\sec(\theta)\tan(\theta) - 2\sec(\theta)\tan(\theta)\sec(\theta) \\ &= \sec^2(\theta) - \csc^2(\theta) - 2\sec^2(\theta)\tan(\theta) \end{aligned}$$

$$\begin{aligned} q^{[y]}(\theta) &= 1 + \ln(\cot(\theta)) - \cot(\theta) \\ q^{[y]\prime}(\theta) &= -\csc(\theta)\sec(\theta) + \csc^2(\theta) \\ q^{[y]\prime\prime}(\theta) &= \csc(\theta)\cot(\theta)\sec(\theta) - \csc(\theta)\sec(\theta)\tan(\theta) - 2\csc(\theta)\cot(\theta)\csc(\theta) \\ &= \csc^2(\theta) - \sec^2(\theta) - 2\csc^2(\theta)\cot(\theta) \end{aligned}$$

Derivatives of Polar Representation of $\nabla^2 m$ - θ functions

We begin with the derivatives of $q(\theta)$ as it is the denominator of all three functions.

$$\begin{aligned}
q(\theta) &= \left(\cos(\theta) - \sin(\theta)\right)^3 \\
q'(\theta) &= 3\left(-\sin(\theta) - \cos(\theta)\right)\left(\cos(\theta) - \sin(\theta)\right)^2 \\
&= -3\left(\sin(\theta) + \cos(\theta)\right)\left(\cos(\theta) - \sin(\theta)\right)\left(\cos(\theta) - \sin(\theta)\right) \\
&= -3\left(\cos^2(\theta) - \sin^2(\theta)\right)\left(\cos(\theta) - \sin(\theta)\right) \\
&= -3\cos(2\theta)\left(\cos(\theta) - \sin(\theta)\right) \\
q''(\theta) &= 6\sin(2\theta)\left(\cos(\theta) - \sin(\theta)\right) - 3\cos(2\theta)\left(-\sin(\theta) - \cos(\theta)\right) \\
&= 6\sin(2\theta)\left(\cos(\theta) - \sin(\theta)\right) + 3\cos(2\theta)\left(\sin(\theta) + \cos(\theta)\right) \\
q^{(3)}(\theta) &= 12\cos(2\theta)\left(\cos(\theta) - \sin(\theta)\right) + 6\sin(2\theta)\left(-\sin(\theta) - \cos(\theta)\right) \\
&\quad - 6\sin(2\theta)\left(\sin(\theta) + \cos(\theta)\right) + 3\cos(2\theta)\left(\cos(\theta) - \sin(\theta)\right) \\
&= 15\cos(2\theta)\left(\cos(\theta) - \sin(\theta)\right) - 6\sin(2\theta)\left(\sin(\theta) + \cos(\theta)\right) - 6\sin(2\theta)\left(\sin(\theta) + \cos(\theta)\right) \\
&= 15\cos(2\theta)\left(\cos(\theta) - \sin(\theta)\right) - 12\sin(2\theta)\left(\sin(\theta) + \cos(\theta)\right)
\end{aligned}$$

We now calculate the derivatives of $q^{[xx]}(\theta)$, $q^{[xy]}(\theta)$ and $q^{[yy]}(\theta)$, the numerators of the limits that we are trying to prove.

$$\begin{aligned}
q^{[xx]}(\theta) &= 2\ln\left(\cot(\theta)\right) + 4\tan(\theta) - \tan^2(\theta) - 3 \\
q^{[xx]\prime}(\theta) &= -2\csc(\theta)\sec(\theta) + 4\sec^2(\theta) - 2\sec^2(\theta)\tan(\theta) \\
q^{[xx]\prime\prime}(\theta) &= -2\left(-\csc(\theta)\cot(\theta)\sec(\theta) + \csc(\theta)\sec(\theta)\tan(\theta)\right) \\
&\quad + 4\left(2\sec(\theta)\tan(\theta)\sec(\theta)\right) \\
&\quad - 2\left(2\sec(\theta)\tan(\theta)\sec(\theta)\tan(\theta) + \sec^2(\theta)\sec^2(\theta)\right) \\
&= 2\csc^2(\theta) - 2\sec^2(\theta) + 8\sec^2(\theta)\tan(\theta) - 4\sec^2(\theta)\tan^2(\theta) - 2\sec^4(\theta) \\
q^{[xx](3)}(\theta) &= 2\left(-2\csc(\theta)\cot(\theta)\csc(\theta)\right) \\
&\quad - 2\left(2\sec(\theta)\tan(\theta)\sec(\theta)\right) \\
&\quad + 8\left(2\sec(\theta)\tan(\theta)\sec(\theta)\tan(\theta) + \sec^2(\theta)\sec^2(\theta)\right) \\
&\quad - 4\left(2\sec(\theta)\tan(\theta)\sec(\theta)\tan^2(\theta) + 2\sec^2(\theta)\sec^2(\theta)\tan(\theta)\right) \\
&\quad - 2\left(4\sec(\theta)\tan(\theta)\sec^3(\theta)\right) \\
&= -4\csc^2(\theta)\cot(\theta) - 4\sec^2(\theta)\tan(\theta) + 16\sec^2(\theta)\tan^2(\theta) + 8\sec^4(\theta) \\
&\quad - 8\sec^2(\theta)\tan^3(\theta) - 8\sec^4(\theta)\tan(\theta) - 8\sec^4(\theta)\tan(\theta) \\
&= 16\sec^2(\theta)\tan^2(\theta) - 4\csc^2(\theta)\cot(\theta) - 4\sec^2(\theta)\tan(\theta) + 8\sec^4(\theta) \\
&\quad - 8\sec^2(\theta)\tan^3(\theta) - 16\sec^4(\theta)\tan(\theta)
\end{aligned}$$

$$q^{[xy]}(\theta) = \cot(\theta) - \tan(\theta) - 2\ln\left(\cot(\theta)\right.$$

$$
\begin{aligned}
q^{[xy]\prime}(\theta) &= -\csc^2(\theta) - \sec^2(\theta) + 2\csc(\theta)\sec(\theta) \\
q^{[xy]\prime\prime}(\theta) &= 2\csc(\theta)\cot(\theta)\csc(\theta) - 2\sec(\theta)\tan(\theta)\sec(\theta) \\
&\quad - 2\csc(\theta)\cot(\theta)\sec(\theta) + 2\csc(\theta)\sec(\theta)\tan(\theta) \\
&= 2\csc^2(\theta)\cot(\theta) - 2\sec^2(\theta)\tan(\theta) - 2\csc^2(\theta) + 2\sec^2(\theta) \\
q^{[xy](3)}(\theta) &= 2\left(-2\csc(\theta)\cot(\theta)\csc(\theta)\cot(\theta) - \csc^2(\theta)\csc^2(\theta)\right) \\
&\quad - 2\left(2\sec(\theta)\tan(\theta)\sec(\theta)\tan(\theta) + \sec^2(\theta)\sec^2(\theta)\right) \\
&\quad + 4\csc(\theta)\cot(\theta)\csc(\theta) + 4\sec(\theta)\tan(\theta)\sec(\theta) \\
&= -4\csc^2(\theta)\cot^2(\theta) - 2\csc^4(\theta) - 4\sec^2(\theta)\tan^2(\theta) \\
&\quad - 2\sec^4(\theta) + 4\csc^2(\theta)\cot(\theta) + 4\sec^2(\theta)\tan(\theta)
\end{aligned}
$$

$$
\begin{aligned}
q^{[yy]}(\theta) &= 2\ln\left(\cot(\theta)\right) - 4\cot(\theta) + \cot^2(\theta) + 3 \\
q^{[yy]\prime}(\theta) &= -2\csc(\theta)\sec(\theta) + 4\csc^2(\theta) - 2\csc^2(\theta)\cot(\theta) \\
q^{[yy]\prime\prime}(\theta) &= 2\csc(\theta)\cot(\theta)\sec(\theta) - 2\csc(\theta)\sec(\theta)\tan(\theta) - 8\csc(\theta)\cot(\theta)\csc(\theta) \\
&\quad + 4\csc(\theta)\cot(\theta)\csc(\theta)\cot(\theta) - 2\csc^2(\theta)\csc^2(\theta) \\
&= 2\csc^2(\theta) - 2\sec^2(\theta) - 8\csc^2(\theta)\cot(\theta) \\
&\quad + 4\csc^2(\theta)\cot^2(\theta) + 2\csc^4(\theta) \\
q^{[yy](3)}(\theta) &= -4\csc(\theta)\cot(\theta)\csc(\theta) - 4\sec(\theta)\tan(\theta)\sec(\theta) + 16\csc(\theta)\cot(\theta)\csc(\theta)\cot(\theta) \\
&\quad + 8\csc^2(\theta)\csc^2(\theta) - 8\csc(\theta)\cot(\theta)\csc(\theta)\cot^2(\theta) - 8\csc^2(\theta)\csc^2(\theta)\cot(\theta) \\
&\quad - 8\csc(\theta)\cot(\theta)\csc^3(\theta) \\
&= -4\csc^2(\theta)\cot(\theta) - 4\sec^2(\theta)\tan(\theta) + 16\csc^2(\theta)\cot^2(\theta) \\
&\quad + 8\csc^4(\theta) - 8\csc^2(\theta)\cot^3(\theta) - 8\csc^4(\theta)\cot(\theta) - 8\csc^4(\theta)\cot(\theta) \\
&= -4\csc^2(\theta)\cot(\theta) - 4\sec^2(\theta)\tan(\theta) + 16\csc^2(\theta)\cot^2(\theta) \\
&\quad + 8\csc^4(\theta) - 8\csc^2(\theta)\cot^3(\theta) - 16\csc^4(\theta)\cot(\theta)
\end{aligned}
$$

References

[1] I. P. Androulakis, C. D. Maranas, and C. A. Floudas. “αBB: A global optimization method for general constrained nonconvex problems”. English. In: *Journal of Global Optimization* 7.4 (1995), pp. 337–363. ISSN: 0925-5001. DOI: 10.1007/BF01099647. URL: http://dx.doi.org/10.1007/BF01099647.

[2] P. Belotti et al. “Mixed-integer nonlinear optimization”. In: *Acta Numerica* 22 (May 2013), pp. 1–131. ISSN: 1474-0508. DOI: 10.1017/S0962492913000032. URL: http://journals.cambridge.org/article_S0962492913000032.

[3] H. T. Broeck. “Economic Selection of Exchanger Sizes”. In: *Industrial & Engineering Chemistry* 36.1 (1944), pp. 64–67. DOI: 10.1021/ie50409a013. eprint: http://dx.doi.org/10.1021/ie50409a013. URL: http://dx.doi.org/10.1021/ie50409a013.

[4] H. Chang and J. Guo. “Heat Exchange Network Design for an Ethylene Process Using Dual Temperature Approach”. In: *Tamkang Journal of Science and Engineering* 8.4 (2005), p. 283.

[5] J. J. Chen. “Comments on improvements on a replacement for the logarithmic mean”. In: *Chem. Eng. Sci.* 42.10 (1987), pp. 2488–2489.

[6] E. Chong and S. H. Zak. *An Introduction to Optimization*. Vol. 76. John Wiley & Sons, 2013.

[7] A.R. Ciric and C.A. Floudas. “Heat exchanger network synthesis without decomposition”. In: *Computers & Chemical Engineering* 15.6 (1991), pp. 385–396. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/0098-1354(91)87017-4. URL: http://www.sciencedirect.com/science/article/pii/0098135491870174.

[8] *Climate Change Act 2008*. 2008. URL: http://www.legislation.gov.uk/ukpga/2008/27/contents (visited on 05/15/2015).

[9] *Coin-or branch and cut (CBC)*. URL: https://projects.coin-or.org/Cbc (visited on 06/02/2015).

[10] *Digest of UK Energy Consumption*. 2011.

[11] Element Energy et al. *The potential for recovering and using surplus heat from industry*. 2014. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/294900/element_energy_et_al_potential_for_recovering_and_using_surplus_heat_from_industry.pdf (visited on 04/23/2015).

[12] M. Escobar and I. E. Grossmann. *Mixed-Integer Nonlinear Programming Models for Optimal Simultaneous Synthesis of Heat Exchangers Network*. Available from CyberInfrastructure for MINLP [www.minlp.org, a collaboration of Carnegie Mellon University and IBM Research] at: www.minlp.org/library/problem/index.php?i=93. Modification of: 16:45:40, April 30 2010.

[13] M. Escobar and J. O. Trierweiler. “Optimal heat exchanger network synthesis: A case study comparison”. In: *Applied Thermal Engineering* 51.1-2 (2013), pp. 801–826. ISSN: 1359-4311. DOI: http://dx.doi.org/10.1016/j.applthermaleng.2012.10.022. URL: http://www.sciencedirect.com/science/article/pii/S1359431112006783.

[14] C. A. Floudas. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Nonlinear and Mixed-integer Optimization: Fundamentals and Applications. Oxford University Press, USA, 1995. ISBN: 9780195356557. URL: https://books.google.co.uk/books?id=OhTfOjSkq18C.

[15] C. A. Floudas, A. R. Ciric, and I. E. Grossmann. “Automatic synthesis of optimum heat exchanger network configurations”. In: *AIChE Journal* 32.2 (1986), pp. 276–290. ISSN: 1547-5905. DOI: 10.1002/aic.690320215. URL: http://dx.doi.org/10.1002/aic.690320215.

[16] K. C. Furman and N. V. Sahinidis. “A Critical Review and Annotated Bibliography for Heat Exchanger Network Synthesis in the 20th Century”. In: *Industrial & Engineering Chemistry Research* 41.10 (2002), pp. 2335–2370. DOI: 10.1021/ie010389e. eprint: http://dx.doi.org/10.1021/ie010389e. URL: http://dx.doi.org/10.1021/ie010389e.

[17] B. Geißler et al. “Using piecewise linear functions for solving MINLPs”. In: *Mixed Integer Nonlinear Programming*. Springer, 2012, pp. 287–314.

[18] A. M Geoffrion. “Elements of Large-Scale Mathematical Programming Part I: Concepts”. In: *Management Science* 16.11 (1970), pp. 652–675. DOI: 10.1287/mnsc.16.11.652. eprint: http://dx.doi.org/10.1287/mnsc.16.11.652. URL: http://dx.doi.org/10.1287/mnsc.16.11.652.

[19] R. E. Gomory. “Outline of an algorithm for integer solutions to linear programs”. In: *Bull. Amer. Math. Soc.* 64.5 (Sept. 1958), pp. 275–278. URL: http://projecteuclid.org/euclid.bams/1183522679.

[20] C. E. Gounaris, R. Misener, and C. A. Floudas. “Computational Comparison of Piecewise Linear Relaxations for Pooling Problems”. In: *Industrial Engineering & Chemistry Research* 48.12 (2009), pp. 5742–5766. DOI: 10.1021/ie8016048.

[21] *Gurobi Optimization, Inc.* URL: http://www.gurobi.com/ (visited on 06/02/2015).

[22] W. E. Hart, J. Watson, and D. L. Woodruff. “Pyomo: modeling and solving mathematical programs in Python”. English. In: *Mathematical Programming Computation* 3.3 (2011), pp. 219–260. ISSN: 1867-2949. DOI: 10.1007/s12532-011-0026-8. URL: http://dx.doi.org/10.1007/s12532-011-0026-8.

[23] K. F. Huang, E. M. Al-mutairi, and I. A. Karimi. “Heat exchanger network synthesis using a stagewise superstructure with non-isothermal mixing”. In: *Chemical Engineering Science* 73 (2012), pp. 30–43. ISSN: 0009-2509. DOI: http://dx.doi.org/10.1016/j.ces.2012.01.032. URL: http://www.sciencedirect.com/science/article/pii/S0009250912000413.

[24] C. S. Hwa. “Mathematical Formulation and Optimization of Heat Exchanger Networks Using Separable Programming”. In: *AIChE- IChemE Symposium Series 4*. AIChE, 1965, pp. 101–106.

[25] *IBM ILOG CPLEX Optimization Studio.* URL: http://www-03.ibm.com/software/products/en/ibmilogcpleoptistud (visited on 06/02/2015).

[26] F. P. Incropera and D. P. De Witt. *Fundamentals of Heat and Mass Transfer.* Wiley, 1995. ISBN: 0471457280.

[27] F. A. Al-Khayyal and J. E. Falk. "Jointly Constrained Biconvex Programming". In: *Mathematics of Operations Research* 8.2 (1983), pp. 273–286. DOI: 10.1287/moor.8.2.273. eprint: http://dx.doi.org/10.1287/moor.8.2.273. URL: http://dx.doi.org/10.1287/moor.8.2.273.

[28] T. Koch et al. "MIPLIB 2010". In: *Mathematical Programming Computation* 3.2 (2011), pp. 103–163. DOI: 10.1007/s12532-011-0025-9. URL: http://mpc.zib.de/index.php/MPC/article/view/56/28.

[29] S. R. Lay. *Analysis: with an introduction to proof.* Pearson, 2014. ISBN: 032174747x.

[30] *LLNL Flow Charts.* URL: https://flowcharts.llnl.gov (visited on 06/14/2015).

[31] G. P. McCormick. "Computability of global solutions to factorable nonconvex programs: Part I -Convex underestimating problems". In: *Mathematical Programming* 10.1 (Dec. 1976), pp. 147–175.

[32] R. Misener, J. P. Thompson, and C. A. Floudas. "APOGEE: Global optimization of standard, generalized, and extended pooling problems via linear and logarithmic partitioning schemes". In: *Computers & Chemical Engineering* 35.5 (2011), pp. 876–892. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/j.compchemeng.2011.01.026. URL: http://www.sciencedirect.com/science/article/pii/S0098135411000366.

[33] H. Mittelmann. *Benchmarks for Optimization Software.* URL: http://plato.asu.edu/bench.html (visited on 06/02/2015).

[34] S. A. Papoulias and I. E. Grossmann. "A structural optimization approach in process synthesis-II: Heat recovery networks". In: *Computers & Chemical Engineering* 7.6 (1983), pp. 707–721. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/0098-1354(83)85023-6. URL: http://www.sciencedirect.com/science/article/pii/0098135483850236.

[35] W. R. Paterson. "A replacement for the logarithmic mean". In: *Chemical Engineering Science* 39.11 (1984), pp. 1635–1636. ISSN: 0009-2509. DOI: http://dx.doi.org/10.1016/0009-2509(84)80090-1. URL: http://www.sciencedirect.com/science/article/pii/0009250984800901.

[36] M. Spivak. *Calculus.* Cambridge University Press, 2006. ISBN: 9780521867443.

[37] G. Strang. *Introduction to Linear Algebra.* Wellesley-Cambridge, 2009. ISBN: 9780980232714.

[38] D. S. Wicaksono and I. A. Karimi. "Piecewise MILP under-and overestimators for global optimization of bilinear programs". In: *AIChE Journal* 54.4 (2008), pp. 991–1008.

[39] *XPRESS Solver Engine.* URL: http://www.solver.com/xpress-solver-engine (visited on 06/02/2015).

[40] T.F. Yee and I.E. Grossmann. "Simultaneous optimization models for heat integration-II. Heat exchanger network synthesis". In: *Computers & Chemical Engineering* 14.10 (1990), pp. 1165–1184. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/0098-1354(90)85010-8. URL: http://www.sciencedirect.com/science/article/pii/0098135490850108.

[41] A. Zavala-Río, R. Femat, and R. Santiesteban-Cos. "An analytical study of the logarithmic mean temperature difference". In: *Revista Mexicana de Ingeniería Química* 4 (2005), pp. 201–212.