MULTIVARIATE ISOTONIC REGRESSION AND ITS ALGORITHMS

A Thesis by

Linda Hoffmann

Bachelor of Science, Newman University, 2007

Submitted to the Department of Mathematics & Statistics
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

May 2009

MULTIVARIATE ISOTONIC REGRESSION
AND ITS ALGORITHMS

The following faculty members have examined the final copy of this thesis for form and content, and recommended that it be accepted in partial fulfillment of the requirement for the degree of Masters of Science with the major in Mathematics.

Xiaomi Hu, Committee Chair

Hari Mukerjee, Committee Member

Chu-Ping Vijverberg, Committee Member

ACKNOWLEDGMENTS

First of all, I want to thank my advisor Dr. Xiaomi Hu for all his time guiding me through the process of writing this Master's thesis paper. He patiently explained to me many questions I had, and I would go to his office every day if necessary. I also had the pleasure of taking classes with him, which helped me to gain a better understanding in the field of Statistics. Without the job offer as a GTA by Dr. Ken Miller, I would not have started a Master's program. Further, I can thank Lori Steiner and John Vogt, who supported me as an undergraduate at Newman University, and encouraged me to pursue a graduate level degree in Mathematics. Many other professors and teachers throughout my life made this thesis possible as well, either by teaching, mentoring, or supporting me.

Last but not least, I would like to thank my parents and family for giving me the opportunity to study abroad. Not only my family but friends encouraged me at times when I felt lost and wanted to give up. Although they do not necessarily teach me the subject of numbers, they are a big part of who I am today and the reason for the completion of this Master's thesis.

ABSTRACT

We use regression functions, which are the means of random variables, to interpret statistical inference. Often an order is imposed on the values of the regression function. Thus, we refer to the regression as an order restricted regression or an isotonic regression. In this paper we explain how to calculate multivariate isotonic regression. However, we investigate the case for a particular restriction on our elements. We impose relations between elements of the same row but not between rows. The technique is to decompose our multivariate model into univariate models so that prior knowledge about the simpler case can be used. Finally, we propose an algorithm to calculate multivariate isotonic regression. This algorithm could then be converted into a computer program.

TABLE OF CONTENTS

1 Introduction

One might wonder about the meaning and purpose of isotonic regression. Regression models predict occurrences in the world in which we live. Often, we have prior knowledge about unknown parameter spaces. In most cases there is some kind of ordering between the elements of a matrix, which is an element of those spaces. Therefore, the methods used to do statistical inference are different than those for the ordinary case. With those statistical constraints, statistical inference is known as order restricted statistical inference. We use regression functions, which are the means of random variables, to interpret our statistical inference. Since an order is imposed on the values of the regression function, we refer to the regression as an order restricted regression or an isotonic regression. Isotonic regression is used in many different fields, especially optimization. It also solves many problems in the industrial and medical field, where levels of treatments such as dose, temperature, and time are given restrictions.

Research involving isotonic regression go back to the 1950's. The first four statisticians to use this method were Dan Brunk, V.J. Chacko, David Bartholomew, and Constance van Eeden. The book *Statistical Inference under Order Restriction* by Richard Barlow, David Bartholomew, J.M. Bremner, and Dan Brunk was published in 1972 [1]. It is one of the first momographs in the field of order restricted regression and it serves as a basis in this field. A few years later, Tim Robertson, F.T. Wright, and R.L. Dykstra published *Order Restricted Statistical Inference* [6]. This book is often referred to as a follow-up of the first book because it displays research done in the 70's and 80's. The book *Constrained Statistical Inference* by M. Silvapulle and Pranab K. Sen [8] published in 2005 is one of the most recent commonly used books.

In 1955, Brunk Ayer and his co-workers invented one of the first algorithms for the simply ordered isotonic regression. Their algorithm is called pool adjacent violators algorithm or PAVA [6]. The problem with this algorithm is that it cannot be used for general orderings. Later, Brunk introduced an algorithm for solving isotonic regression with a partially ordered parameter space, called minimum lower set algorithm. Many other algorithms followed. These algorithms share the deficiency of a large number of iterations and other limited uses such as particular ordering. In 1983, Lee presented the min-max algorithm [5]. This algorithm allows for problems of larger size. Hu and Hansohm introduced the merge-and-chop algorithm in 2008 [4], which is mentioned later.

This thesis paper will introduce the concept of multivariate isotonic regression, including its algorithms. This branch of statistics is fairly new but many researchers have published their results. Most of the discoveries were found by Japanese researchers such as S. Sasabuchi, M. Inutsuka and D. D. S. Kulatunga. Multivariate isotonic regression allows statisticians to model the vast field of applications in a better and more complex way. The user has the option to integrate more variables of interest at once and integrate their relationship.

First, we consider the multivariate regression model $Y(x) \sim N(\mu(x), \Sigma)$ with $x \in \{x_1, \ldots, x_q\}$ as a predictor variable. The value of the regression function at x_i is $\mu(x_i) \in \mathbb{R}^p$. If we collect all those values, we can express $\mu(x) = (\mu(x_1), \ldots, \mu(x_q))$ as a matrix. Now, the restriction is put between the columns of $\mu(x)$. This restriction gives us a reflexive and transitive relation between the columns of $\mu(x)$. An example is given by S. Sasabuchi, M. Inutsuka, and D.D.S. Kulatunga in *A Multivariate Version of Isotonic Regression*, which was published in 1983. They studied the case $\mu(x_1) \le \mu(x_2) \le \cdots \le \mu(x_q)$, where $' \le'$ means the vectors are compared component wise. Later, Geng Zhi and Shi Ning-Zhong introduced a more specific ordering, called the umbrella ordering, in *Isotonic Regression in two Independent Variables under Umbrella Ordering* (1991).

Instead of a simple ordering method, we will relate our columns when the difference of two belong to a particular set. More precisely, each component of the difference will be compared to a given restriction. Then, there exists an ordering between elements of the same row. The rows themselves do not need to show any relation. We try to keep the statistical constraints as general as possible so that a variety of cases can be considered. Due to the prior knowledge of orderings on our target matrix, we will find more efficient results.

Over the years, many papers about algorithms for the univariate isotonic regression have been published. According to my knowledge, there is only one published algorithm for the multivariate case. In 1992, Syoichi Sasabuchi, Makoto Inutsuka, and D.D. Sarath Kulatunga introduced *An Algorithm for Computing Multivariate Isotonic Regression* [7]. Their algorithm works for partially ordered sets only. In the algorithm the given matrix is broken down into rows and columns so that the isotonic regression is found row by row. While one row is transformed, the others are fixed. Therefore, one can use the algorithms studied in the univariate case. We will also find the isotonic regression row by row so that prior univariate isotonic regression algorithms can be used. However, we will put less restriction on our order of the parameter space. The orderings will be reflexive and transitive. We will look at matrices for which the orderings are among elements of the same row. The goal is to establish an algorithm with lesser restrictions so that it can be used for many different cases of multivariate isotonic regression.

As mentioned above, this paper will introduce the concepts of multivariate isotonic regression, including its algorithms. Chapter 2 will introduce of some preliminary terminology, which is crucial for the study of isotonic regression. Then, the univariate model with its properties, algorithms, and applications is introduced. This chapter serves as a basis for the upcoming multivariate isotonic regression starting in chapter 3. The exploration of the orderings of vectors and the collection of isotonic matrices are presented in chapter 3. Chapter 4 deals with projections of matrices and defines multivariate isotonic regression. The purpose of multivariate isotonic regression can be found at the end of chapter 4, through a discussion of their applications. An algorithm for multivariate isotonic regression is introduced in chapter 5. It includes the sufficient conditions on the parameter space as well as the sample space. Conclusion remarks and acknowledgments for the completion of this paper are found in chapter 6 and 7 respectively.

2 Univariate isotonic regression

In this chapter, basic terminology of isotonic regression is reviewed. Further, results concerning with univariate isotonic regression found by different researcher are presented. This information serves as a foundation for later chapters.

2.1 Isotonic function

Let Ω be a set of elements called the index set with a binary relations denoted by $\ll$. We say that the relation $\ll$ is reflexive when for all $x \in \Omega$, $x \ll x$. We say that the relation $\ll$ is transitive when for all $x, y, z \in \Omega$, $x \ll y$ and $y \ll z$ imply $x \ll z$.

Definition 2.1. *Quasi order [6]*
A relation $\ll$ is a quasi-order (preorder) in Ω when it is reflexive and transitive.

An example of a quasi order is $\leq$ in $\mathbb{R}$. Once we have defined our quasi order, we can proceed with introducing a function f on Ω.

Let $\Omega = \{x_1, x_2, \ldots, x_q\}$ be our index space with the quasi order $\ll$ imposed between its elements and a function $f(x)$ defined on it. For the univariate case, let f be a vector such that

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix} \in \mathbb{R}^q \text{ where } f_i = f(x_i) \text{ for all } i = 1, \ldots, q.$$

From now on, $\ll$ refers to the quasi-order.

Definition 2.2. *Isotonic vector [6]*
A vector $f \in \mathbb{R}^q$ is isotonic or order restricted when for any $i \ll j$ we have $f_i \leq f_j$ with $i, j = 1, \ldots, q \in \Omega$.

A function can be considered as a vector so that Definition 2.2 applies to functions. Since f is a real valued function, we can refer to it as isotonic or univariate isotonic function if $\ll$ is defined on Ω and $i \ll j$ implies $f_i \leq f_j$.

2.2 Isotonic cone

Consider the linear space L and let A be a set of L. A is convex if for all $x, y \in L$ and for all $a \in [0, 1]$ imply $ax + (1 - a)y \in A$ [6]. We say A is cone if $ax \in A$ for all $a \geq 0$ [6].

Lemma 2.3. *Convex cone*
We say that $\mathcal{D}$ is a convex cone if and only if $f, g \in \mathcal{D}$ imply $af + bg \in \mathcal{D}$ for a, b being nonnegative.

Proof. Refer to [2, p 25] for verifications. □

Lemma 2.4. *Isotonic cone*
The collection of all isotonic functions defined on Ω is denoted by $\mathcal{C}$. $\mathcal{C}$ is a convex cone in $\mathbb{R}^q$.

Proof. Let f and $g \in \mathcal{C}$ and $a \in [0, 1]$. Define $z = af + (1 - a)g$ and let $x, y \in \Omega$ with $x \ll y$. We need to show that z is isotonic. We know that $x \ll y$ implies $f(x) \leq f(y)$ and $g(x) \leq g(y)$. Therefore, $x \ll y$ will also imply $af(x) + (1 - a)g(x) \leq af(y) + (1 - a)g(y)$ or, equivalently, $z(x) \leq z(y)$ since f and g are isotonic and real functions multiplied by a scalar. So, $z \in \mathcal{C}$ is an isotonic function and $\mathcal{C}$ is a convex set.

Let $f \in \mathcal{C}$ and a be a nonnegative scalar. We know that for any $x, y \in \Omega, x \ll y$ implies $f(x) \leq f(y)$. Now, if we multiply f with the scalar a, we do not change anything at this relation since we are dealing with real numbers and a is nonnegative. Therefore, $\mathcal{C}$ must be a cone. □

Before we move on, we will discuss the properties of closed sets. By definition, a set A is closed if it contains all of its limit points. In other words, all convergent sequences $\{x_i\}$ in A have a limit $p_i \in A$. So, if $\|x_n - p\| \to 0$ then $p \in A$ and A is closed.

Lemma 2.5. *The isotonic cone $\mathcal{C}$ is a closed set with respect to any norm induced by an inner product.*

A complete proof is given in [6, p 15].

2.3 Isotonic regression

Next, we will define our inner product. Then, we will be ready to introduce projections and isotonic regression.

Definition 2.6. *Inner product*

The inner product of $x, y \in \mathbb{R}^q$ is defined by

$$\langle x, y \rangle = \sum_{i=1}^{q} x_i y_i w_i \text{ with weight } w_i.$$

We note that $\{w_1, w_2, \ldots, w_q\}$ are usually given or chosen appropriately [8]. Further, it can be shown that an inner product induces a norm, $\langle x, x \rangle = \|x\|^2_{\langle \cdot,\cdot \rangle}$. Further,we note that a norm induces a distance, $\|x - y\|_{\langle \cdot,\cdot \rangle} = d(x, y)$.

Since the collection of all isotonic functions with the norm as defined in 2.6 is a closed convex cone $\mathcal{C} \in \mathbb{R}^q$, there exists a unique $f^* \in \mathcal{C}$ for which $\|f - f^*\| \leq \|f - g\|$ for all $g \in \mathcal{C}$ [6, p 15]. We refer to f^* as the projection of f onto $\mathcal{C}$ denoted by $P(f|\mathcal{C})$. This result is defined next.

Definition 2.7. *Projection [1]*

Let $\mathcal{C}$ be a closed isotonic cone in a space H. Then, the closest point of $\mathcal{C}$ to $f \in H$ is the $f^* = P(f|\mathcal{C})$.

Definition 2.8. *Three properties of $P(f|\mathcal{C})$*

A projection f^* onto an ordered restricted cone has following three properties [6]:

- $f^* \in \mathcal{C}$
- $\langle f - f^*, f^* \rangle = 0$
- $\langle f - f^*, g \rangle \leq 0$*, for all* $g \in \mathcal{C}$

In statistics, our goal is to find $\hat{\mu}(x)$ in a given model so that we receive the best estimation to a problem. In other words, we try to minimize the error terms. In order to do that, we first have to collect a random sample, $Y_{i1}, \ldots, Y_{in_i}$ for which we estimate the unknown parameter(s). Then, we decide on a fitted model. In most cases, we can make the assumption of normality so that the model is given by

$$Y(x) = \mu(x_i) + \epsilon \in \mathbb{R}, \tag{1}$$

where $\epsilon \sim N(0, \sigma^2)$ and $x_i \in \Omega = \{x_1, x_2, \ldots, x_q\}$.

If there are no restrictions on the parameter space, we can use the technique of Maximum Likelihood Estimation (MLE) to find the best estimator of our parameter(s). For instance, in the case of a normal distribution, where $\bar{Y}_i$ is the sample mean, the MLE of $\mu(x)$ is $\hat{\mu}(x)$, which is equal to $\bar{Y} = (\bar{Y}_1, \ldots, \bar{Y}_q)'$ [3].

If we put a quasi order on Ω and $\mu = (\mu(x_1) \quad \mu(x_2) \quad \cdots \quad \mu(x_p))'$ is isotonic as defined in 2.2, then we need to find the restricted MLE of μ. Using standard arguments, one can show that the projection of $\bar{Y}$ onto $\mathcal{C}$ is the restricted MLE of $\mu(x)$ denoted by $\mu^* = P(\hat{\mu}|\mathcal{C})$ with the inner product as defined in definition 2.6 [6] and with $w_i - n_i$. This projection is commonly known as isotonic regression.

2.4 Applications

In this section, a small portion of the wide field of applications in statistical inference and in optimization theories will be presented. We will take a look at some commonly used distributions.

2.4.1 The exponential family

The pdf of an exponential function is given by

$$f(x, \mu, \beta) = e^{p(\mu,\beta)+q(\mu)r(\beta)S(x,\beta)+T(x,\beta)}, \tag{2}$$

where we want to estimate μ while β is known. Further, we require

$$\frac{d}{d\mu}p(\mu, \beta) = -\mu\frac{d}{d\mu}q(\mu)r(\beta),\ \frac{d}{d\mu}q(\mu) > 0 \text{ and } r(\beta) > 0.$$

Then, $E[S(X, \beta)] = \mu$ and $Var[S(X, \beta)] = \frac{1}{q'(\mu)r(\beta)}$. We can show that the unbiased MLE of μ is $\bar{S} = \sum_{i=1}^{n} S(x_i, \beta)/n \sim (\mu, \frac{1}{nq'(\mu)r(\beta)})$. The restricted MLE of $\mu \in \mathcal{C}$ is given by $\mu^* = P(\bar{S}|\mathcal{C})$ with weights $w_i = n_i r(\beta_i)$ [6].

2.4.2 Discrete distributions

We can investigate the estimation of means of three different discrete distributions, the Bernoulli, the Poisson, and the geometric distribution. Note that all the pdf's can be written in exponential form so that we can use prior results. Further, the findings of the unbiased estimator using MLE methods can be found in *Intro to Mathematical Statistics* [3].

- For the Bernoulli distribution $f(x, \theta) = e^{ln(1-\theta)+(ln\frac{\theta}{1-\theta})x}$, $\bar{x}^* = P(\bar{x}|\mathcal{C})$ is the restricted MLE under $\theta \in \mathcal{C}$ with $w_i = n_i$.

- For the Poisson distribution $f(x, \lambda) = e^{-\lambda+(ln\lambda)x-lnx!}$, $\bar{x}^* = P(\bar{x}|\mathcal{C})$ is the restricted MLE under $\lambda \in \mathcal{C}$ where $w_i = n_i$.

- The geometric distribution with $f(x, \theta) = e^{-ln(1+\theta)+(ln\frac{\theta}{1+\theta})x}$ has $\bar{x}^* = P(\bar{x}|\mathcal{C})$ with $w_i = n_i$ as restricted MLE under $\theta \in \mathcal{C}$ [6].

2.4.3 Continuous distributions

Since the pdf of a normal model and a gamma distribution can be written as an exponential family, we can find their restricted MLE's, as well.

- The pdf of a normal distribution is $f(x, \mu, \sigma^2) = e^{-\frac{\mu^2}{2\sigma^2}-ln(\sqrt{2\pi\sigma^2})+\mu\frac{1}{\sigma^2}x-\frac{x^2}{2\sigma^2}}$. $\bar{x}^* = P(\bar{x}|\mathcal{C})$ is the restricted MLE of μ where $w_i = \frac{n_i}{\sigma_i^2}$.

- The pdf of a gamma distribution is given by $f(x, \alpha, \beta) = e^{-ln(\beta^\alpha\Gamma(\alpha))-\frac{1}{\beta}\alpha\frac{x}{\alpha}+lnx^{\alpha-1}}$. The restricted MLE of β is $\hat{\beta}^* = P(\hat{\beta}|\mathcal{C})$ with $w_i = \frac{n_i}{\alpha_i}$ [6].

2.5 Algorithms

We will now introduce an algorithm for computing isotonic regression when given a univariate model. Finding effective and efficient algorithms has been and still is an active area of the study of order restricted statistical inference. One of the famous and older ones is PAVA [6], [8]; but it is not included in this paper. Instead, we focus on a more recent algorithm.

2.5.1 Merge-&-chop procedure [4]

The Merge-&-chop procedure works for many different orderings on a finite set. The main idea behind the merge-and-chop procedure is to identify the largest and smallest value in a set after partitioning Ω into n finite sets. Then, we take out the largest or smallest valued set, which is called an upward or downward merge, and run the procedure again.

The algorithm stops when the space is exhausted. In other words, after n-iteration, the isotonic regression will be found with the help of partitioning. One advantage of this algorithm is that it can be easily implemented by a program because it consists of a finite number of iterations which identify the optimal values step-by-step by partitioning the given space.

3 Order of vectors and isotonic matrices

The main focus of this paper is to investigate multivariate isotonic regression. Before we start working with matrices, we will investigate vectors. Let $\Omega = \{j : j = 1, \ldots, q\}$ with quasi order $\ll$ defined on it. Since we are not in $\mathbb{R}^1$ anymore, we will use $\preceq$ instead of $\leq$ as our restriction on $\mathbb{R}^p$, where p is the dimension of our column vectors. For instance, if

$1 \ll 2 \ll 3$, where $\Omega = \{1, 2, 3\}$ for a population of three, then $x_1 \preceq x_2 \preceq x_3 \in \mathbb{R}^p$ is a vector valued function on Ω.

3.1 Order of vectors

In general, we know how to order real numbers. However, how do we order vectors? We use binary ordering with the properties of reflexive and transitive. Hence, a quasi order will be imposed between the vectors in $\mathbb{R}^p$, where p is the dimension.

3.1.1 Quasi order of vectors

Definition 3.1. *Let $\preceq$ be a binary relation of vectors in $\mathbb{R}^p$. This relation is a quasi order if it is reflexive, $x \preceq x$, and transitive, $x \preceq y$ and $y \preceq z$ implying $x \preceq z$, for all $x, y, z \in \mathbb{R}^p$.*

To get a better understanding what the previous definition indicates, we will consider three different examples.

Example 3.1 *Define $x \preceq y$ for $x = (x_1, x_2)'$ and $y = (y_1, y_2)' \in \mathbb{R}^2$ if $y_1 - x_1 + y_2 - x_2 \leq 0$ and $y_1 - x_1 - y_2 + x_2 \leq 0$ so that $\preceq$ is a quasi order in $\mathbb{R}^2$.*

Proof. Let $x, y, z \in \mathbb{R}^2$. First we will check if the relation is reflexive, i.e. $(x_1, x_2)' \preceq (x_1, x_2)'$. Since

$$x_1 - x_1 + x_2 - x_2 \leq 0 \text{ and } x_1 - x_1 - x_2 + x_2 \leq 0$$

are true, our relation is reflexive. The last question remains, if $(x_1, x_2)' \preceq (y_1, y_2)'$ and $(y_1, y_2)' \preceq (z_1, z_2)'$ imply $(x_1, x_2)' \preceq (z_1, z_2)'$. We know that

$$y_1 - x_1 + y_2 - x_2 \leq 0 \text{ and } z_1 - y_1 + z_2 - y_2 \leq 0$$

are true, since both statements are non-positive. If we add them, they are still non-positive. So,

$$y_1 - x_1 + y_2 - x_2 + z_1 - y_1 + z_2 - y_2 \leq 0$$

$$z_1 - x_1 + z_2 - x_2 \leq 0$$

implies that our relation is transitive. Therefore Example 3.1 defines a quasi order. □

Example 3.2 *Define $x \preceq y$ for $x = (x_1, x_2, x_3)'$ and $y = (y_1, y_2, y_3)' \in \mathbb{R}^3$ if $x_i \leq y_i$ for all $i = 1, 2, 3$ so that $\preceq$ is a quasi order in $\mathbb{R}^3$.*

Proof. Let $x, y, z \in \mathbb{R}^3$ and $i = 1, 2, 3$. First we will check if the relation is reflexive, i.e. $(x_1, x_2, x_3)' \preceq (x_1, x_2, x_3)'$. Since $x_i \leq x_i$, our relation is reflexive. For transitivity, we check if $(x_1, x_2, x_3)' \preceq (y_1, y_2, y_3)'$ and $(y_1, y_2, y_3)' \preceq (z_1, z_2, z_3)'$ imply $(x_1, x_2, x_3)' \preceq (z_1, z_2, z_3)'$. We know that $x_i \leq y_i$ and $y_i \leq z_i$ or, equivalently, $x_i \leq y_i \leq z_i$. Thus, $x_i \leq z_i$, and our relation is transitive. Therefore Example 3.2 defines a quasi order. □

Example 3.3 *Define $x \preceq y$ for $x = (x_1, x_2, x_3, x_4)'$ and $y = (y_1, y_2, y_3, y_4)' \in \mathbb{R}^4$ if $x_1 \leq y_1$, $x_3 = y_3$, and $x_4 \geq y_4$ so that $\preceq$ is a quasi order in $\mathbb{R}^4$.*

Proof. Let $x, y, z \in \mathbb{R}^4$ and $i = 1, 2, 3$. First we will check if the relation is reflexive, i.e. $(x_1, x_2, x_3, x_4)' \preceq (x_1, x_2, x_3, x_4)'$. Since $x_1 \leq x_1$, $x_3 = x_3$, and $x_4 \geq x_4$, our relation is reflexive. Next we check the reflexivity property. We know that $x_1 \leq y_1 \leq z_1$. Thus, $x_1 \leq z_1$. Further, $x_3 = y_3 = z_3$. Thus, $x_3 = z_3$. Additionally, $x_4 \geq y_4 \geq z_4$. Thus, $x_4 \geq z_4$, and our relation is transitive. Therefore Example 3.3 defines a quasi order. □

3.1.2 Quasi order of vectors defined by a convex cone

Another way to define a relation of vectors is by using a convex cone $\mathcal{D}$.

Lemma 3.2. *Let $\mathcal{D}$ be a convex cone in $\mathbb{R}^p$ as defined in the definition 2.3 and define $x \preceq y$ if $y - x \in \mathcal{D} \in \mathbb{R}^p$ for $x, y \in \mathbb{R}^p$. Then $\preceq$ defines a quasi order in $\mathbb{R}^p$.*

Proof. We know that the zero vector $0 \in \mathcal{D}$. Thus, for all $x \in \mathbb{R}^p$, we have $x - x \in \mathcal{D}$ implying $x \preceq x$ and $\preceq$ is reflexive. Let $x \preceq y$ and $y \preceq z$. Then $y - x \in \mathcal{D}$ and $z - y \in \mathcal{D}$. Further, the sum $(y - x) + (z - y) \in \mathcal{D}$ because $\mathcal{D}$ is a convex cone. This implies $z - x \in \mathcal{D}$ and $x \preceq z$; $\preceq$ is transitive. We have shown that $\preceq$ is a quasi order. □

Again, we will take a look at some examples. Each of them will be based on the previous three examples.

Example 3.4 *Let $\mathcal{D} = \{z = (z_1, z_2)' \in \mathbb{R}^2 : z_1 + z_2 \leq 0, z_1 - z_2 \leq 0\}$. Then $\mathcal{D}$ is a convex cone in $\mathbb{R}^2$ and the quasi order $\preceq$ defined with this $\mathcal{D}$ is the same order as in Example 3.1.*

Proof. Let $y, z \in \mathcal{D}$ and $a \geq 0$ and $b \geq 0$. We know $y_1 + y_2 \leq 0, y_1 - y_2 \leq 0, z_1 + z_2 \leq 0$, and $z_1 - z_2 \leq 0$. Then $ay + bz$ gives us $(ay_1 + bz_1) + (ay_2 + bz_2) = a(y_1 + y_2) + b(z_1 + z_2)$. Since a and b are nonnegative and each product is less or equal than zero, the entire sum is less or equal to zero. Secondly, we have $(ay_1 + bz_1) - (ay_2 + bz_2) = a(y_1 - y_2) + b(z_1 - z_2)$. Again we know that this expression is less than or equal to zero. Thus, $ay + bz \in \mathcal{D}$ and $\mathcal{D}$ is indeed a convex cone.

Recall the relation in Example 3.1: $y_1 - x_1 + y_2 - x_2 =\leq 0$ and $y_1 - x_1 - y_2 + x_2 = y_1 - x_1 - (y_2 - x_2) \leq 0$. Let $z_i = y_i - x_i$ then the two quasi orders defined are the same for Example 3.1 and 3.4. □

Example 3.5 *Let $\mathcal{D} = \{x = (x_1, x_2, x_3)' \in \mathbb{R}^3 : x_i \geq 0$ for all $i = 1, 2, 3\}$. Then $\mathcal{D}$ is a convex cone in $\mathbb{R}^3$ and the quasi order $\preceq$ defined with this $\mathcal{D}$ is the same order as in Example 3.2.*

Proof. Let $x, y \in \mathcal{D}$, $a \geq 0$ and $b \geq 0$, and $i = 1, 2, 3$. We know $x_i \geq 0$, and $y_i \geq 0$. Then $ax_i + by_i \geq 0$ and $ax + by \in \mathcal{D}$. In Example 3.2, $x_i \leq y_i$ this implies $y_i - x_i \geq 0$ and therefore an element in $\mathcal{D}$. The two quasi orders are the same. □

Example 3.6 *Let* $\mathcal{D} = \{x = (x_1, x_2, x_3, x_4)' \in \mathbb{R}^4 : x_1 \geq 0,\ x_3 = 0,\ \text{and}\ x_4 \leq 0\}$*. Then* $\mathcal{D}$ *is a convex cone in* $\mathbb{R}^4$ *and the quasi order* $\preceq$ *defined with this* $\mathcal{D}$ *is the same order as in Example 3.3.*

Proof. Let $x, y \in \mathcal{D}$ and $a \geq 0$ and $b \geq 0$. Then, $ax_1 + by_1 \geq 0$ since $x_1 \geq 0$ and $y_1 \geq 0$. Next, $ax_3 + by_3 = 0$ because $x_3 = 0$ and $y_3 = 0$. Also, $ax_4 + by_4 \leq 0$ due to the fact that $x_4 \leq 0$ and $y_4 \leq 0$. Thus, $ax + by \in \mathcal{D}$. In Example 3.3, $x_1 \leq y_1$ which implies $y_1 - x_1 \geq 0$ in $\mathcal{D}$. Also, $x_3 = y_3$ implies $y_3 - x_3 = 0$ and $x_4 \geq y_4$ implies $\rightarrow y_4 - x_4 \leq 0$. Thus the quasi orders defined in Example 3.3 and 3.6 are the same. □

3.1.3 A decomposable vector order

In Lemma 3.2, we declared that two vectors are related if their difference is in $\mathcal{D}$. Further, we may add that being in $\mathcal{D}$ means that each component of the difference of the two vectors is larger, smaller, equal to zero, or any real number. In other words, there are vector orders which can be decomposed into univariate orders for each of its components. However, for the univariate order there exists only four different cases, which the next lemma reflects.

Lemma 3.3. *Let* $r(\cdot)$ *be a mapping from* $\{1, \ldots, p\}$ *to*

$$\{" \geq ", " \leq ", " \leq " \text{ and } " \geq ", " \leq " \text{ or } " \geq "\}.$$

For $x, y \in \mathbb{R}^p$*, define* $x \preceq y$ *if* $x_i r(i) y_i$ *for all* $i = 1, \ldots, p$*. Then* $\preceq$ *is a quasi order in* $\mathbb{R}^p$*.*

The proof is omitted.

The next step is to investigate the above lemma and apply it to the examples we have seen.

Example 3.7 *If we let* $r(1) = r(2) = r(3) = " \leq "$*, then we get the same quasi order as defined in Example 3.2 and 3.5*

Example 3.8 *If we let* $r(1) = " \leq "$*,* $r(2) = " \leq "$ *or* $" \geq "$*,* $r(3) = " \leq "$ *and* $" \geq "$*, and* $r(4) = " \geq "$*, we obtain the same quasi order as defined in Example 3.3 and 3.6.*

The order defined in Lemma 3.3 is called decomposable because it can be decomposed into univariate quasi orders for each component. We have seen some examples for which this is true, such as Example 3.2, 3.3, 3.5, and 3.6. However, if different components of the same vector relate, we cannot decompose the quasi order on it. An example is given by Example 3.1 and 3.4; their orders are not decomposable.

We also notice that a decomposable order is an order defined by a convex cone as proven in Examples 3.5 and 3.6. Suppose $x = (x_1, \ldots, x_p)' \preceq (y_1, \ldots, y_p)' = y$ then $y_i - x_i$ is greater than or equal to zero, less than or equal to zero, or equal to zero since we are dealing with real numbers. So, each $y_i - x_i \in \mathcal{D}_i$ for all $i = 1, \ldots, p$.

3.2 Isotonic matrices

In this section, we will take a closer look how isotonic matrices are composed and what kind of properties they contain. In the previous section, we talked about orderings of vectors. Those orderings of vectors will lead to isotonic matrices. Different types of orderings for matrices will be considered but not all of them will be studied in more detail. Again, let $\ll$ be the quasi order in $\Omega = \{1, \ldots, q\}$ and $\preceq$ a quasi order on $\mathbb{R}^p$.

3.2.1 Isotonic matrices

Definition 3.4. *Isotonic matrix*
The matrix $X = (X_1, \ldots, X_q) \in \mathbb{R}^{p \times q}$ *is called an isotonic matrix if* $i \ll j$ *implies* $X_i \preceq X_j$.

To visualize the definition above, let us consider a few examples.

Example 3.9 *Let* $\Omega = \{1, 2, 3, 4, 5\}$ *with the quasi order* $1 \ll 2 \ll 3$ *and* $5 \ll 4 \ll 3$. *With respect to* $\preceq$ *in Example 3.2, where ordering was defined component wise, a matrix* $X = (x_{ij})_{3\times 5}$ *is isotonic if and only if* $x_{i1} \leq x_{i2} \leq x_{i3} \geq x_{i4} \geq x_{i5}$ *for all* $i = 1, 2, 3$.

Note that the ordering defined in Example 3.9 is actually called umbrella ordering and the rows of X are synchronized.

Example 1
Consider the case in which we collect data from different grades in high school and middle school. We are interested in the average amount of knowledge, represented by x_{ij}, of each subject such that we have following mean matrix:

$$\begin{pmatrix} & 6^{th}\text{ grade} & 7^{th}\text{ grade} & \cdots & 12^{th}\text{ grade} \\ \text{Math} & x_{11} & x_{12} & \cdots & x_{1q} \\ \text{English} & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & & & & \vdots \\ \text{Biology} & x_{p1} & x_{p2} & \cdots & x_{pq} \end{pmatrix}.$$

As we can imagine, we expect the number of the amount of knowledge to increase from the left to the right. In other words, the columns are related but not the rows. We can compare the columns component wise so that the rows are synchronized. For instance, if $x_{1j} \preceq x_{1k}$ then $x_{ij} \preceq x_{ik}$, for any $1 \leq i \leq p$. Further, the order is decomposable because it can be decomposed into univariate quasi orders for elements in rows as will be explained later.

Example 2
Imagine that we measure the average age of cats, dogs, women, and men after 3 month and after they have reached adulthood.

$$\begin{pmatrix} & \text{dogs} & \text{cats} & \text{women} & \text{men} \\ \text{Weight as an adult} & x_1 & y_1 & s_1 & t_1 \\ \text{Weight after 3 months} & x_2 & y_2 & s_2 & t_2 \end{pmatrix}.$$

In the above matrix, the first column relates to the second one, $X \ll Y$ and the third relates to the fourth one, $T \ll S$ but we would not compare the first two with the second two because they represent two completely different groups. Further, $y_1 - x_1 \leq 0$ and $y_2 - x_2 \leq 0$ so that $(y_1 - x_1) + (y_2 - x_2) \leq 0$. We can also assume that the weight difference is larger during adulthood, thus $(y_1 - x_1) - (y_2 - x_2) \leq 0$. We have the same results with S and T. We conclude that $\preceq$ is the same quasi order as defined in Example 3.4.

Previously, an isotonic matrix was defined by the ordering of its columns. However, there are many ways to obtain isotonic matrices. For instance, we could have ordering between the rows, any elements in the matrix, or between elements of the same row.

3.2.2 A convex cone

Lemma 3.5. *The collection of all possible isotonic $X's \in \mathbb{R}^{p \times q}$ will give us $\mathcal{C}$, a convex cone in $\mathbb{R}^{p \times q}$.*

Proof. Let X and $Y \in \mathcal{C}$, $a \in [0, 1]$, and β_i, γ_i be either any elements, columns, or rows of X and Y respectively. Define $Z = aX + (1 - a)Y$ and let $j, k \in \Omega$ with $j \ll k$. We need to show that Z is isotonic. We know that $j \ll k$ implies $\beta_j \leq \beta_k$ and $\gamma_j \leq \gamma_k$. Therefore, $j \ll k$ will also imply $a\beta_j + (1 - a)\gamma_j \leq a\beta_k + (1 - a)\gamma_k$ or equivalently $z_j \leq z_k$ since X and Y are isotonic only multiplied by a scalar. So, $Z \in \mathcal{C}$ is an isotonic matrix and $\mathcal{C}$ is a convex set.

Let $X \in \mathcal{C}$, a be a nonnegative scalar, and β_i be either any element, column, or row of X. We know that for any $j, k \in \Omega, j \ll k$ implies $\beta_j \leq \beta_k$. Now, if we multiply X with the scalar a, we do not change anything at this relation since we are dealing with real numbers as elements of X and a is nonnegative. Therefore, $\mathcal{C}$ must be a convex cone in $\mathbb{R}^{p \times q}$. □

3.2.3 A decomposable cone

The case in which two columns of X are related if the difference is an element of $\mathcal{D}$ is very specific. The rows of X are synchronized, and $\mathcal{D}$ and $\mathcal{C}$ are decomposed by their components. In this paper, we would rather investigate a more general case. We will look at matrices for which the rows are not synchronized but elements within one row may be related.

Before we start looking into more definitions, consider the problem of the age when a child starts to walk, eat, and talk. An experiment with different treatments can be conducted. The first group would be the control group, where the children or babies do not receive any help. In group #1, a child receives particular speaking exercises. The next group may receive help on how to eat and walk. So,

$$\begin{pmatrix} & \text{control group} & & \text{group \#1} & & \cdots & \text{group \#}q \\ \text{walk} & x_{11} & & x_{12} & \geq & \cdots & x_{1q} \\ \text{eat} & x_{21} & & x_{22} & \geq & \cdots & x_{2q} \\ \text{talk} & x_{31} & \geq & x_{32} & \leq & \cdots & x_q \end{pmatrix}.$$

As we see, we will have orderings between elements on the rows but not necessarily between different rows.

Definition 3.6. *Let $\mathcal{C}_{(i)}, i = 1, \ldots, p$ be the set of all rows of X, denoted by $X_{(i)}$, and $X_{(i)} \in \mathbb{R}^{p \times q}$ such that there exists a matrix $X \in \mathcal{C}$. In other words, $\mathcal{C}_{(i)} = \{X_{(i)} \in \mathbb{R}^q : \exists X = (X_{(1)}, \ldots, X_{(p)})' \in \mathcal{C}\}$.*

Lemma 3.7. *$\mathcal{C}_{(i)}$ as defined in Definition 3.6 is a convex cone. Therefore, we have p of $\mathcal{C}_{(i)}$ cones in $\mathcal{C}$.*

The proof is omitted because it is similar to the proof of Lemma 3.5.

Lemma 3.8. *Let $\mathcal{C}_{(i)}$ be the univariate isotonic cone with respect to a quasi order $\ll$ in Ω and the quasi order $r(\cdot)$ as defined in Lemma 3.3 in $\mathbb{R}^1$. Let $\mathcal{C}$ be the multivariate isotonic cone with respect to $\ll$ and $\preceq$ in $\mathbb{R}^q$. Then $X = (X_{(1)}, \ldots X_{(p)})' \in \mathcal{C}$ if and only if $X_{(i)} \in \mathcal{C}_{(i)}$ for all $i = 1, \ldots, p$ and we call $\mathcal{C}$ decomposable.*

Proof. ($\Rightarrow$) Assume $X \in \mathcal{C}$. Then for every $X_j \preceq X_k$ when $j \ll k$. By Lemma 3.3 this implies that for all $i = 1, \ldots, p$, $x_{ij} r(i) x_{ik}$. Thus $X_i \in \mathcal{C}_i$.
($\Leftarrow$) Assume $X_{(i)} \in \mathcal{C}_i$ for all $i = 1, \ldots, p$. Then $x_{ij} r(i) x_{ik}$. By Lemma 3.3, this implies that $X_j \preceq X_k$. Thus, $X \in \mathcal{C}$. □

Example 1, in which we measured the knowledge of students, represents a decomposable cone. Each row is a isotonic cone and the rows itself do not relate. However, example 2, in which we measured the weight of cats, dogs, men, and women, does not represent an isotonic cone.

Note that we consider a decomposable cone if each row is defined by a univariate isotonic order restriction. This order restriction of the i^{th} row in X depends on the relation in Ω. Further, each row may have a different restrictions. Thus, the rows may or may not be synchronized to each other. As we have proven in Lemma 3.8, the matrix looks like:

$$X = \begin{pmatrix} (x_{11} & x_{12} & \ldots & x_{1q}) : C_1 \\ (x_{21} & x_{22} & \ldots & x_{2q}) : C_2 \\ \vdots & & & \vdots \\ (x_{p1} & x_{p2} & \ldots & x_{pq}) : C_p \end{pmatrix}.$$

4 Multivariate isotonic regression

We have already seen in practice how multivariate isotonic regression may be useful. After a proposition, a summary of the purpose of this subject matter is given.

Proposition 4.1. *The projection of a matrix onto the isotonic cone is the multivariate isotonic regression and presents the best estimation for order restricted parameters.*

Our goal is to find an estimate of a multivariate regression function under order restrictions. This function can be displayed as a matrix $X \in \mathbb{R}^{p \times q}$. In the upcoming sections, we will show that $X^* = P_{\langle \cdot,\cdot \rangle}(X|\mathcal{C})$ is our multivariate isotonic regression or projection of X onto $\mathcal{C}$ induced by the inner product $\langle \cdot, \cdot \rangle$ as defined later. The three properties in Definition 2.8 apply here as well. Further, X^* exists and is unique in $\mathcal{C}$ such that $\|X - X^*\| = \inf\{\|X - Y\| : Y \in \mathcal{C}\}$. Therefore, we can compute $P_{\langle \cdot,\cdot \rangle}(f|\mathcal{C})$. An algorithm for the projections will be introduced in the next chapter.

4.1 Projection onto a multivariate isotonic cone

As mentioned in the last chapter, $\mathcal{C}$ is the collection of all isotonic matrices for which the elements of each row are related. This statement is defined next.

Definition 4.2. *Let $\mathcal{C}$ be the collection of all isotonic matrices, which have an ordering between the elements of the same row.*

Further, Lemma 3.8 stated that this particular convex cone is decomposable into $\mathcal{C}_1, \ldots, \mathcal{C}_p$. This result will be used in the upcoming sections.

4.1.1 A closed convex cone

Already for the univariate model, we defined inner product and norms. Let $X, Y \in \mathbb{R}^{p\times q}$ and $\langle X, Y\rangle$ be the inner product defined on $\mathbb{R}^{p\times q}$. Any inner product $\langle\cdot,\cdot\rangle$ is a mapping from $\mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$ to $\mathbb{R}^1$. It must satisfy

i. $\langle X, X\rangle \geq 0$ for all $X \in \mathbb{R}^{p\times q}$. The equality only holds if and only if $X = 0$.

ii. $\langle X, Y\rangle = \langle Y, X\rangle$ for all $X, Y \in \mathbb{R}^{p\times q}$.

iii. $\langle \alpha X + \beta Y, Z\rangle = \alpha\langle X, Z\rangle + \beta\langle Y, Z\rangle$ for all $X, Y, Z \in \mathbb{R}^{p\times q}$.

Further, the inner product induces a norm, i.e. $\|X\| = \sqrt{\langle X, X\rangle}$. Any norm $\|\cdot\|$ is a mapping from $\mathbb{R}^{p\times q}$ to $\mathbb{R}^0_+$, which is the collection of all non-negative real numbers. The norm must satisfy following properties:

i. $\|X\| \geq 0$ for all $X \in \mathbb{R}^{p\times q}$. The equality only holds if and only if $X = 0$.

ii. $\|X + Y\| \leq \|X\| + \|Y\|$ for all $X, Y \in \mathbb{R}^{p\times q}$.

iii. $\|\alpha X\| = |\alpha|\|X\|$, where α is a scalar in $\mathbb{R}^1$.

The relationship between norm and inner product is very useful. Since $\mathbb{R}^{p\times q}$ is a linear space with finite dimensions and a norm is induced by an inner product in that space, a sequence of matrices $\{X^{[n]}\} = \{(x_{ij}^{[n]})_{p\times q}\} \to B = (b_{ij})_{p\times q}$ if and only if every sequence of elements converges to that point, such that $\{x_{ij}^{[n]}\} \to b_{ij}$ for all $i = 1, \ldots, p$ and $j = 1, \ldots, q$. A proof can be found in [9].

Once we have defined an inner product, we can continue talking about $\mathcal{C}$.

Lemma 4.3. *Let $\|\cdot\|$ be a norm in $\mathbb{R}^{p\times q}$ induced from an inner product $\langle\cdot,\cdot\rangle$. The collection of all isotonic matrices $\mathcal{C}$ is a closed convex cone.*

Proof. We have already shown that $\mathcal{C}$ is convex cone (section 3.2.3). It is left to prove that $\mathcal{C}$ is also closed. We have to show that if $\{X^{[i]}\} \subseteq \mathcal{C}$ and $\{X^{[i]}\} \to P$ then $P \in \mathcal{C}$. We can write $X^{[i]} = (x_{jk}^{[i]})_{p\times q}$ and $P = (p_{jk})_{p\times q}$. Since $X^{[i]} \in \mathcal{C}$ let $x_{jk}^{[i]} \leq x_{jl}^{[i]}$ be one of the restrictions. If $\{X^{[i]}\} \to P$ then $x_{jk}^{[i]} \to p_{jk}$ and $x_{jl}^{[i]} \to p_{jl}$ [9]. Thus, $p_{jk} \leq p_{jl}$ and $P \in \mathcal{C}$. $\square$

4.1.2 Projection onto a closed convex cone

The properties mentioned earlier about projections apply for the projective matrix as well. However, we are dealing with a different space. Therefore, we will now study the properties of projections with matrices in more detail. We will show that there exists a matrix X^*, which produces the minimal distance to any matrix in H, our Hilbert space, and $\mathcal{C}$. Before introducing the three sufficient properties of projections, we will show that X^* is unique.

Definition 4.4. *Projection*

The projection of any matrix X in our entire space H onto the isotonic cone $\mathcal{C}$ is denoted by $X^* = P(X|\mathcal{C})$. The projection of X minimizes the distances between X and all matrices in $\mathcal{C}$, such that $\|X - X^*\| = \inf\{\|X - Y\| : Y \in \mathcal{C}\}$.

Lemma 4.5. *For all X in H, there exists a unique $X^* \in \mathcal{C}$ such that $\|X - X^*\| = \inf\{\|X - Y\| : Y \in \mathcal{C}\}$.*

Proof. First, we will prove existence. Let $v = \inf\{\|X - Y\| : Y \in \mathcal{C}\}$. Further, we define $\{Y_n\}$ be a sequence of matrices so that $\lim_{n\to\infty} \|Y_n - X\| = v$. Now, we will use the identity $\|Y + X\|^2 + \|Y - X\|^2 = 2(\|Y\|^2 + \|X\|^2)$, called the Parallelogram Law, to write:

$$\frac{1}{4}\|Y_m - Y_n\|^2 = \frac{1}{2}\|Y_m - X\|^2 + \frac{1}{2}\|Y_n - X\|^2 - \|\frac{1}{2}(Y_m + Y_n) - X\|^2.$$

We know that $\frac{1}{2}(Y_m + Y_n) \in \mathcal{C}$ since $\mathcal{C}$ is a convex set and $\|\frac{1}{2}(Y_m + Y_n) - X\|^2 \geq v^2$. Additionally, $\|Y_m - X\|^2$ and $\|Y_n - x\|^2$ equal v as m, n go to ∞. Therefore, with large n, m our above equation goes to zero and we can conclude that $\{Y\}$ is a Cauchy sequence because the elements become closer and closer. We will denote the limit of $\{Y\}$ as X^*. Since $\mathcal{C}$ is closed, $X^* \in \mathcal{C}$.

It is left to show uniqueness of X^*. Let $X_1^*, X_2^* \in \mathcal{C}$, such that $\|X - X_1^*\| = \inf\{\|X - Y\| : Y \in \mathcal{C}\} = v$ and $\|X - X_2^*\| = \inf\{\|X - Y\| : Y \in \mathcal{C}\} = v$. Then,

$$\frac{1}{4}\|X_1^* - X_2^*\|^2 = \frac{1}{2}\|X_1^* - X\|^2 + \frac{1}{2}\|X_2^* - X\|^2 - \|\frac{1}{2}(X_1^* + X_2^*) - X\|^2.$$

As mentioned previously, the above equation will go to zero. Thus, $\frac{1}{4}\|X_1^* - X_2^*\|^2 = 0$ implies $\|X_1^* - X_2^*\| = 0$ and $X_1^* = X_2^*$.

□

The three sufficient properties for a projective matrix onto a closed convex cone are presented next.

Lemma 4.6. *$X^* = P_{\mathcal{C}}(X)$ if and only if*

i. $X^* \in \mathcal{C}$,

ii. $\langle X - X^*, X^* \rangle = 0$ *and*

iii. $\langle X - X^*, Y \rangle \leq 0$ *for all* $Y \in \mathcal{C}$.

Proof. (i)

By Lemma 4.5, $X^* \in \mathcal{C}$.

□

Proof. (ii)

Let $X^* = P_A(x)$, then we know from above that $\|X - X^*\| \leq \|X - Y\|$ for all $Y \in \mathcal{C}$. Further, we can replace Y with $Y' = \alpha Y + (1-\alpha)X^*$ for $0 \leq \alpha \leq 1$ because $\mathcal{C}$ is convex. So,

$$
\begin{aligned}
&\|X - X^*\| \leq \|X - Y\| \\
&\|X - X^*\|^2 \leq \|X - Y\|^2 \\
&\|X - X^*\|^2 - \|X - Y\|^2 \\
&= \|X - X^*\|^2 - \|X - Y'\|^2 \\
&= \|X - X^*\|^2 - \|(X - X^*) + (X^* - Y')\|^2 \\
&= \|X - X^*\|^2 - \|(X - X^*) + (X^* - \alpha Y - (1-\alpha)X^*)\|^2 \\
&= \|X - X^*\|^2 - \|(X - X^*) + (\alpha X^* - \alpha Y)\|^2 \\
&= \langle X - X^*, X - X^*\rangle - \langle X - X^*, X - X^*\rangle - 2\langle X - X^*, \alpha(X^* - Y)\rangle - \alpha^2\|X^* - Y\|^2 \\
&-2\alpha\langle X - X^*, X^* - Y\rangle - \alpha^2\|X^* - Y\|^2 \\
&\leq 0.
\end{aligned}
$$

Now divide by -2α and let $\alpha \to 0$, we receive

$$\langle X - X^*, X^* - Y\rangle \geq 0 \text{ for all } Y \in \mathcal{C}.$$

Next, we replace Y with αX^*, which is still an element in $\mathcal{C}$, and we get the following inequality:

$$
\begin{aligned}
&\langle X - X^*, X^* - (\alpha X^*)\rangle \geq 0 \\
&\langle X - X^*, X^*\rangle - \alpha\langle X - X^*, X^*\rangle \geq 0 \\
&(1-\alpha)\langle x - x^*, x^*\rangle \geq 0.
\end{aligned}
$$

Depending on α, $\langle x - x^*, x^*\rangle$ will be positive (for $0 \leq \alpha \leq 1$) or negative (for $\alpha > 1$). This is a contradiction; therefore, $\langle X - X^*, X^*\rangle = 0$ since the inequality must be true for all α. So, we have shown that $\langle X - X^*, X^*\rangle = 0$ if $X^* = P_{\mathcal{C}}(X)$.

□

Proof. (iii)

From above results, we have $\langle X - X^*, X^*\rangle = 0$ and $\langle X - X^*, X^* - Y\rangle \geq 0$. So,

$$
\begin{aligned}
&\langle X - X^*, X^* - Y\rangle \geq 0 \\
&\langle X - X^*, X^*\rangle - \langle X - X^*, Y\rangle \geq 0 \\
&0 - \langle X - X^*, Y\rangle \geq 0 \\
&\Rightarrow \langle X - X^*, Y\rangle \leq 0.
\end{aligned}
$$

Next, we need to show the opposite direction is true, as well. Assume $\langle X - X^*, X^*\rangle = 0$ and $\langle X - X^*, Y\rangle \leq 0$ for all $Y \in \mathcal{C}$. By similar reasoning as above, we have

$$
\begin{aligned}
&\langle X - X^*, X^*\rangle - \langle X - X^*, Y\rangle \geq 0 \\
&\langle X - X^*, X^* - Y\rangle \geq 0.
\end{aligned}
$$

Since $\|X^* - Y\|^2 \geq 0$, we can write

$$\begin{aligned}
&\langle X - X^*, X^* - Y\rangle \geq -\tfrac{1}{2}\|X^* - Y\|^2 \\
&0 \geq -2\langle X - X^*, X^* - Y\rangle - \|X^* - Y\|^2 \\
&= \langle X - X^*, X - X^*\rangle - \langle X - X^*, X - X^*\rangle - 2\langle X - X^*, X^* - Y\rangle - \|X^* - Y\|^2 \\
&= \|X - X^*\|^2 - \|(X - X^*) + (X^* - Y)\|^2 \\
&= \|X - X^*\|^2 - \|X - Y\|^2 \\
&= \|X - X^*\|^2 - \|X - Y\|^2.
\end{aligned}$$

Then, $0 \geq \|X - X^*\|^2 - \|X - Y\|^2$ implies $\|X - X^*\| \leq \|X - Y\|$, which means $X^* = P_{\mathcal{C}}(X)$. □

4.2 Multivariate isotonic regression

Finally, we have the tools to introduce multivariate isotonic regression. Keep in mind that we are adding at least another observation to our univariate model; thus the population and sample model will get more complex.

4.2.1 A multivariate regression model

Our multivariate regression model with the random vector $Y(x)$ and the regression function $\mu(x)$ is given by

$$Y(x) = \mu(x) + \epsilon \in \mathbb{R}^p,\ \epsilon \sim N(0, \Sigma), \tag{3}$$

where $x \in \Omega = \{x_1, \ldots, x_q\}$. The regression function can be expressed as $\mu = (\mu_1, \ldots, \mu_q) \in \mathbb{R}^{p\times q}$, where $\mu_i = \mu(x_i)$. Further, we have $Y = \begin{pmatrix} Y_1 & Y_2 & \ldots & Y_q \end{pmatrix} \in \mathbb{R}^{p\times q}$, where $Y_i = Y(x_i)$. Suppose that we draw a random sample, which follows a normal model such that $Y_i \sim N(\mu_i, \Sigma)$ where Σ is the variance for all Y_i. We assume that Σ is known and every random vector $Y_i \in Y$ comes from a population with size n_i, e.g. $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_i}$ is a random sample from Y_i. Each component of that sample is a vector. Then, we can find the sample mean denoted by $\bar{Y}_i$. Again, we can express all of the sample means as $\bar{Y} = (\bar{Y}_1, \ldots, \bar{Y}_q) \in \mathbb{R}^{p\times q}$.

Our goal is to maximize the likelihood function. If there are no restriction on the parameter space, the solution is given by the MLE of $\mu_i \in \mathbb{R}^p$ denoted by $\hat{\mu}_i = \bar{y} \in \mathbb{R}^p$, which is stated and proved in the next Lemma.

Lemma 4.7. *Suppose the model and samples are described as above, $n = n_1 + \cdots + n_q$ and Σ is known.*
(a) The likelihood function μ is

$$L(\mu) = c \times e^{-\frac{1}{2}\sum_{i=1}^{q} n_i(\bar{Y}_i - \mu_i)'\Sigma^{-1}(\bar{Y}_i - \mu_i)},$$

where $c = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \times e^{-\frac{1}{2}\sum_{i=1}^{q} n_i \sum_{k=1}^{n_i}(Y_{i,k} - \bar{Y}_i)'\Sigma^{-1}(Y_{i,k} - \bar{Y}_i)}$.
(b) $\bar{Y}$ is the MLE for μ.

Proof. Part (a)

The likelihood function, which is joint pdf of the $Y_{11} \dots Y_{qn_q}$ transfered as a function of μ, can be expressed as

$$L(\mu, Y) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{q}\sum_{k=1}^{n_i}(Y_{i,k}-\mu_i)'\Sigma^{-1}(Y_{i,k}-\mu_i)}.$$

Note: $\sum_{k=1}^{n_i}(Y_{i,k} - \mu_i)'\Sigma^{-1}(Y_{i,k} - \mu_i)$

$$= \sum_{k=1}^{n_i}\left((Y_{i,k} - \bar{Y}_i)'\Sigma^{-1}(Y_{i,k} - \bar{Y}_i) + (\bar{Y}_i - \mu_i)'\Sigma^{-1}(\bar{Y}_i - \mu_i) + 2(\bar{Y}_i - \mu)'\Sigma^{-1}(Y_{i,k} - \bar{Y}_i)\right)$$

$$= \sum_{k=1}^{n_i}(Y_{i,k} - \bar{Y}_i)'\Sigma^{-1}(Y_{i,k} - \bar{Y}_i) + n_i(\bar{Y}_i - \mu_i)'\Sigma^{-1}(\bar{Y}_i - \mu_i)$$

since $2\sum_{k=1}^{n_i}(\bar{Y}_i - \mu)'\Sigma^{-1}(Y_{i,k} - \bar{Y}_i) = 0$.

So,

$$L(\mu, Y) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{q}\sum_{k=1}^{n_i}(Y_{i,k}-\bar{Y}_i)'\Sigma^{-1}(Y_{i,k}-\bar{Y}_i)} * e^{-\frac{1}{2}\sum_{i=1}^{q} n_i(\bar{Y}_i-\mu_i)'\Sigma^{-1}(\bar{Y}_i-\mu_i)}.$$

□

Proof. Part (b) $L(\mu, Y)$ is maximized when the part $n_i(\bar{Y}_i - \mu_i)'\Sigma^{-1}(\bar{Y}_i - \mu_i)$ of the exponent is minimized. Thus, if we replace μ_i by $\bar{Y}_i$ for all $i's$ then we maximize $L(\mu, Y)$. Therefore, the MLE of μ denoted by $\hat{\mu} = \bar{Y}$. □

4.2.2 Multivariate isotonic regression

We have seen how to find a good estimate for μ in a given model. However, often we have prior knowledge about the regression function μ such that $\mu \in \mathcal{C}$, where $\mathcal{C}$ is our decomposable isotonic cone. We have seen an example earlier. In section 3.2.3 an experiment was conducted in which babies were put randomly in different groups and the age at which they start walking, talking, and eating was measured. Keep in mind that for each entry in our sample matrix, we have n_i - observations, $i = 1, \dots, q$. Since this was an experiment, there was an ordering between groups but not between the different observations. Thus, it is an example of our special isotonic cone $\mathcal{C}$.

Before we can find the MLE of μ under the restriction that $\mu \in \mathcal{C}$, an inner product must be defined.

Lemma 4.8. *Let $X = (X_1, \dots, X_q)$, $Y = (Y_1, \dots, Y_q) \in \mathbb{R}^{p\times q}$. Then,*

$$\langle X, Y\rangle = \sum_{j=1}^{q} X_j' \left(\frac{\Sigma}{n_j}\right)^{-1} Y_j$$

defines an inner product for matrices.

Proof. First, we need to show that $\langle X, X\rangle \geq 0$ where the equality only holds if and only if $X = 0$ is true. Let $\left(\frac{\Sigma}{n_j}\right)^{-1} = B^{[j]}$. We have

$$\langle X, X\rangle = \sum_{j=1}^{q} X_j' B^{[j]} X_j \geq 0$$

since $X_j' B^{[j]} X_j$ for all j.

Further,

$$\begin{aligned}\langle X, X\rangle = 0 &\Leftrightarrow X_j' B^{[j]} X_j = 0 \text{ for all } j\\ &\Leftrightarrow X_j = 0\end{aligned}$$

and the equality only holds if and only if $X = 0$.

Next, we need to prove that $\langle X, Y\rangle = \langle Y, X\rangle$. So,

$$\langle X, Y\rangle = \sum_{j=1}^{q} X_j' B^{[j]} Y_j = \sum_{j=1}^{q} Y_j' B^{[j]} X_j = \langle Y, X\rangle.$$

It is left to show the linearity property. We have

$$\langle \alpha X + \beta Y, Z\rangle = \sum_{j=1}^{q} (\alpha X_j + \beta Y_j)' B^{[j]} Z_j$$

$$= \alpha \sum_{j=1}^{q} X_j' B^{[j]} Z_j + \beta \sum_{j=1}^{q} Y_j' B^{[j]} Z_j = \alpha\langle X, Z\rangle + \beta\langle Y, Z\rangle.$$

□

Once the inner product has been defined, we can proceed with the search of finding the restricted MLE. If we have an ordering on any elements on $\mu = (\mu_1 \ldots \mu_q)$ as described in earlier and where $\mu_i = (\mu_{ji})_{p\times 1}$ with $i = 1, \ldots, q$ and $j = 1, \ldots p$, then we can vectorize μ so that we have the following case:

$$vec(\mu) = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \\ \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \\ \vdots \\ \mu_{1q} \\ \vdots \\ \mu_{pq} \end{pmatrix} \in \mathbb{R}^{pq}.$$

Thus, our problem of finding an estimation to $\mu \in \mathcal{C}$ has become a one-dimensional problem. We can use algorithms introduced in the univariate case to solve for μ^*, the restricted MLE of μ. However, the problem of finding the restricted μ is more complicated if we want to impose any other order between the elements, rows, or columns of our μ. Remember, that we want μ to be in the specifically defined $\mathcal{C}$ by Definition 4.2.

In order to find the restricted MLE, we still have to work with the likelihood function. We have defined our inner product in Lemma 4.8 and we know that the inner product induces a norm $\|\cdot\|$. This means,

$$\sum_{j=1}^{q} n_j(\bar{Y}_j - \mu_i)'\Sigma^{-1}(\bar{Y}_j - \mu_i) = \langle(\bar{Y} - \mu), (\bar{Y} - \mu)\rangle = \|(\bar{Y} - \mu)\|^2$$

and $\|\bar{Y} - \mu\| = d(\bar{Y}, \mu)$.

Therefore, the likelihood function from Lemma 4.7 can be written as

$$L(\mu) = c * e^{-\frac{1}{2}\|\bar{Y}-\mu\|^2}$$

$$\text{with } c = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} * e^{-\frac{1}{2}\sum_{i=1}^{q} n_j \sum_{k=1}^{n_j}(Y_{i,k}-\bar{Y}_j)'\Sigma^{-1}(Y_{j,k}-\bar{Y}_j)}.$$

We can conclude that the likelihood function is maximized under the restriction $\mu \in \mathcal{C}$ if $\|\bar{Y} - \mu\|^2$ is minimized for $\mu \in \mathcal{C}$. This means, $\bar{Y}$ needs to be projected onto $\mathcal{C}$. This important result is listed in the next lemma.

Lemma 4.9. *Suppose the model, sample, and statistics are as described in section 4.2.1 and the restricted $\mu \in \mathcal{C}$ is as described in section 4.2.2. Then the MLE of μ under the restriction is $P(\bar{Y}|\mathcal{C})$.*

In conclusion, the projection onto the restricted isotonic cone gives us the restricted MLE of the multivariate regression function. Thus, this projection is often referred to as multivariate isotonic regression.

5 Algorithms for the multivariate isotonic regression

Currently, there exists only one algorithm for the multivariate isotonic regression [7]. In this paper, another algorithm, which is related to [7], is proposed. Recall that the goal of an algorithm is to find Y^* from an initial Y. Remember that the challenge is often that we have an order restriction not only on our columns but also on our rows. Recall that in this paper, we are interested in orderings between elements of the same row.

5.1 Assumptions

The proposed algorithm in this chapter needs three assumptions. The first assumption is made on the inner product $\langle \cdot, \cdot \rangle$, the next is made on the structure of $\mathcal{C}$, and the last assumption is made on the computation for univariate isotonic regression.

i. Inner product

   In Lemma 4.8, the inner product for the multivariate regression model was defined as

$$\langle X, Y \rangle = \sum_{j=1}^{q} X'_j \left(\frac{\Sigma}{n_j} \right)^{-1} Y_j, \text{ for } X, Y \in \mathbb{R}^{p \times q}.$$

   This is very specific. Instead, we will work with any positive definite matrix $A^{[j]} = \left(a_{ik}^{[j]} \right)_{p \times p}$ for $j = 1, \ldots, q$. Then the inner product of $X = (X_1, \ldots, X_q)$ and $Y = (Y_1, \ldots, Y_q)$ in $\mathbb{R}^{p \times q}$ is defined as

$$\langle X, Y \rangle = \sum_{j=1}^{q} X'_j A^{[j]} Y_j, \text{ for } X, Y \in \mathbb{R}^{p \times q}.$$

   Again, the norm $\| \cdot \|$ is induced by this inner product and we consider the projection of a matrix onto a decomposable cone with respect to this inner product.

   In $A^{[j]}$, the diagonal elements $a_{ii}^{[j]} > 0$ for all $i = 1, \ldots, p$ and $j = 1, \ldots, q$ because the matrix is a positive definite matrix. Therefore, the p inner products of $X, Y \in \mathbb{R}^q$ can be defined as $\langle x, y \rangle_{(i)} = x' diag \left(a_{ii}^{[1]}, \ldots, a_{ii}^{[q]} \right) y$ where $i = 1, \ldots, p$.

ii. Decomposable cone

   In Lemma 3.3 the mapping $r(\cdot)$ from $\{1, \ldots, p\}$ to $\{" \geq "," \leq "," \leq " \text{ and } " \geq "," \leq " \text{ or } " \geq "\}$ was defined. Consider $\ll_i$ as the quasi order on $\Omega = \{1, \ldots, q\}$, where $i = 1, \ldots, p$ indicates the row. Thus, $\ll_i$ and $r(i)$, which are quasi order in $\mathbb{R}^1$, determine a type of isotonic function on Ω and a closed convex cone $\mathcal{C}_{(i)}$ in $\mathbb{R}^q$. For all $i = 1, \ldots, p$, $\mathcal{C}_{(i)}$, which is the collection of all isotonic vectors of that particular type,

$$\mathcal{C} = \{X = (X_{(1)}, \ldots, X_{(p)})' \in \mathbb{R}^{p \times q} : X_{(i)} \in \mathcal{C}_{(i)} \text{ for all } i = 1, \ldots, p\}.$$

   As we recall from chapter 3, $\mathcal{C}$ is the decomposable cone, on which we will project the matrix of interest later. In Example 3.9, $\ll_1, \cdots, \ll_5$ are all identical. Since all the quasi orders were the same for each row, the rows of matrix X are actually synchronized. However $\ll_i$'s do not have to be the same. We want to consider the more general case, in which the rows do not necessarily relate.

iii. Computable univariate isotonic regressions

Chapter 2 came to the conclusion that with respect to an inner product, the projection of $f \in \mathbb{R}^q$ onto a closed convex cone exists. Thus, $P_{(i)}(f|\mathcal{C}_{(i)})$ exists and is unique. This projection is computable and we know that many algorithms exist. In the upcoming algorithm for the multivariate isotonic regression, we will include one of those algorithms and assume it works.

5.2 Rationale

Suppose $Y = (Y_1, \ldots, Y_q) = (Y_{(1)}, \ldots, Y_{(p)})' = (y_{ij})_{p\times q}$ is the matrix for which the projection needs to be computed. First, we initialize $Y = \mu^{[0]}$. Then we need to update $\mu^{[0]}$ to $\mu^{[1]}$ row by row; in general this algorithm updates $\mu^{[n]} \to \mu^{[n+1]}$. The limit of the sequence $\{\mu^{[n]}\}$ will give us $P(Y|\mathcal{C})$.

Since we are updating row by row, we reduced the multivariate isotonic regression to a univariate isotonic regression problem. This way, it is possible to use knowledge of univariate isotonic regression algorithms. After each row has been updated, we need to compute $d = \max\{|\mu_{ij}^{[n+1]} - \mu_{ij}^{[n]}| : i = 1, \ldots, p; j = 1, \ldots, q\}$. We will use d as a measurement tool for when to stop the algorithm. Thus, we compare our d with a predefined value δ_0 . If $d > \delta_0$, then we update each row of $\mu^{[n+1]}$ again; otherwise we stop and state that $Y^* = \mu^{[n+1]}$.

Next we will investigate what kind of procedure we can use for our algorithm. We only work with one row at a time while the other rows stay fixed. From section 4.2.2 we know that we need to find $\min \|Y - \mu\|^2$ for $\mu \in \mathcal{C}$. The question is how the i^{th} row of $\mu = (\mu_1, \ldots, \mu_q) = (\mu_{(1)}, \ldots, \mu_{(p)})' = (\mu_{ij})_{p\times q}$ can be updated so that we get the desired result.

We know that $\min \|Y - \mu\|^2 = \min \sum_{j=1}^{q} (Y_j - \mu_j)' A^{[j]} (Y_j - \mu_j)$. Further, all the rows except the i^{th} row, the one we are updating, are fixed. Therefore, we can divide up our vector.

$$Y_j - \mu_j = \begin{pmatrix} \begin{pmatrix} y_{1j} & - & \mu_{1j} \\ & \vdots & \\ y_{(i-1)j} & - & \mu_{(i-1)j} \end{pmatrix} \\ \begin{pmatrix} y_{ij} & - & \mu_{ij} \end{pmatrix} \\ \begin{pmatrix} y_{(i+1)j} & - & \mu_{(i+1)j} \\ & \vdots & \\ y_{pj} & - & \mu_{pj} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$$

Then, the problem of minimization can be written as

$$\begin{array}{l}(Y_j-\mu_j)'A^{[j]}(Y_j-\mu_j)\\ =(Z_1'Z_2Z_3')\times\left[\begin{array}{ccc}A_{11}^{(i-1)\times(i-1)} & A_{12}^{(i-1)\times i} & A_{13}^{(i-1)\times(p-(i+1))}\\ A_{21}^{i\times(i-1)} & A_{22}^{i\times i} & A_{23}^{i\times(p-(i+1))}\\ A_{31}^{(p-(i+1)\times(i-1)} & A_{32}^{(p-(i+1)\times i} & A_{33}^{(p-(i+1)\times(p-(i+1))}\end{array}\right]\times\left(\begin{array}{c}Z_1\\ Z_2\\ Z_3\end{array}\right)\\ \begin{array}{lclcl}=Z_1'A_{11}^{(i-1)\times(i-1)}Z_1 & + & Z_2A_{21}^{i\times(i-1)}Z_1 & + & Z_3'A_{31}^{(p-(i+1)\times(i-1)}Z_1\\ +Z_1'A_{12}^{(i-1)\times i}Z_2 & + & Z_2A_{22}^{i\times i}Z_2 & + & Z_3'A_{32}^{(p-(i+1)\times i}Z_2\\ +Z_1'A_{13}^{(i-1)\times(p-(i+1))}Z_3 & + & Z_2A_{23}^{i\times(p-(i+1))}Z_3 & + & Z_3'A_{33}^{(p-(i+1)\times(p-(i+1))}Z_3.\end{array}\end{array}$$

We note that we minimize the same term as when using the technique of finding the MLE described earlier. However, we are only interested in the parts which include y_{ij}. The other terms are constants and we will denote them by c_1. Thus, the minimization problem reduces to

$$c_1+Z_2A_{21}Z_1+Z_1'A_{21}Z_2+Z_2A_{22}Z_2+Z_3'A_{32}Z_2+Z_2A_{23}Z_3,\text{ where}$$

$$Z_2A_{21}Z_1=Z_1'A_{21}Z_2\text{ and }Z_3'A_{32}Z_2=Z_2A_{23}Z_3.$$

So, we have to minimize

$$c_1+Z_2A_{22}Z_2+2*(Z_2A_{21}Z_1+Z_2A_{23}Z_3).$$

If we substitute our variables back, we have

$$c_1+a_{ii}^{[j]}(y_{ij}-\mu_{ij})^2+2\times(y_{ij}-\mu_{ij})\left(\sum_{k=1}^{i-1}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}+\sum_{k=i+1}^{p}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}\right).$$

Then, we factor $a_{ii}^{[j]}$ and complete the square. Note that adding the term

$$\left(\sum_{k=1}^{i-1}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}+\sum_{k=i+1}^{p}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}\right)^2$$

does not affect our problem since it does not contain y_{ij}. We just change the constant to c_2. So, we get

$$c_2+a_{ii}^{[j]}\left(y_{ij}-\mu_{ij}+\frac{\sum_{k=1}^{i-1}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}+\sum_{k=i+1}^{p}[y_{kj}-\mu_{kj}]a_{kj}^{[j]}}{a_{ii}^{[j]}}\right)^2.$$

So, we actually have

$$c_2+\sum_{j=1}^{q}a_{ii}^{[j]}\left[y_{ij}-\mu_{ij}+\sum_{k=1,k\neq i}^{p}\frac{a_{ki}^{[j]}}{a_{ii}^{[j]}}(y_{kj}-\mu_{kj})\right]^2.$$

Let $f_{ij} = y_{ij} + \sum_{k=1,k\neq i}^{p} \frac{a_{ki}^{[j]}}{a_{ii}^{[j]}}(y_{kj} - \mu_{kj})$ and $f_{(i)} = (f_{i1}, \dots, f_{iq})' \in \mathbb{R}^q$, then

$$\|Y - \mu\|^2 = c_2 + \|f_{(i)} - \mu_{(i)}\|_{(i)}^2 \text{ for } j = 1, \dots q.$$

In other words, if we want to minimize $\|Y - \mu\|^2$, we should minimize $\|f_{(i)} - \mu_{(i)}\|_{(i)}^2$. This means $P_{(i)}(f_{(i)}|\mathcal{C}_{(i)})$ should and can be used to replace the i^{th} row of μ.

5.3 Algorithm

Based on the previous results, we can now introduce our algorithm. We note that updating the i^{th} row means that the other ones stay fixed, such that $\mu_{(k)} = \mu_{(k)}^{[n+1]}$ for $k = 1, \dots, i-1$ and $\mu_{(k)} = \mu_{(k)}^{[n]}$ for $k = i, \dots, p$.

Step 1 Initialize μ

Let n be the number of iteration such that $\mu \to \mu^{[n]}$. Before we start, $n = 0$; thus, $\mu = Y$ is actually $\mu^{[0]} = Y$.

Step 2 Update $\mu^{[n]}$ row by row

1. Set $i = 1$.
2. Update the i^{th} row of $\mu^{[n]}$ to that of $\mu^{[n+1]}$
   (a) Create $f_{(i)}^{[n]} = (f_{i1}^{[n]}, \dots, f_{iq}^{[n]})' \in \mathbb{R}^q$, where
   $$f_{ij}^{[n]} = y_{ij} + \frac{\sum_{k=1}^{i-1}(y_{kj} - \mu_{kj}^{[n+1]})a_{kj}^{[j]} + \sum_{k=i+1}^{p}(y_{kj} - \mu_{kj}^{[n]})a_{kj}^{[j]}}{a_{ii}^{[j]}}.$$
   (b) Compute $f_{(i)}^{*[n]} = P_{(i)}(f_{(i)}^{[n]}|\mathcal{C}_{(i)})$. The problem is a univariate one so that algorithms for the univariate isotonic regression can be used.
   (c) Update $\mu_{(i)}^{[n]}$ to $\mu_{(i)}^{[n+1]} = P_{(i)}(f_{(i)}^{[n]}|\mathcal{C}_{(i)})$.
3. If $i < p$, increase $i = i + 1$ and then go back to 2. of **Step 2**; otherwise set $n = n + 1$ and proceed to step **Step 3**.

Step 3 Check and output

If $d = \max\{|\mu_{ij}^{[n+1]} - \mu_{ij}^{[n]}| : i = 1, \dots, p; j = 1, \dots, q\} > \delta_0$, where δ_0 is a predefined small tolerance level, go back to **Step 2**; otherwise, display $\mu^{[n+1]}$ as $Y^* = P(Y|\mathcal{C})$.

For this algorithm, we assumed that the sequence $\{\mu^{[n+1]}\}$ converges. Those steps are very tedious. A similar proof as in [7] can be used, which established this result.

5.4 A theorem

We assume that $\{\mu^{[n]}\}$ converges. Next we will show that this sequence converges to $\lim_{n\to\infty}\mu^{[n]} = Y^*$, which means we need to prove that $\lim_{n\to\infty}\mu^{[n]}$ is a projection with $\lim_{n\to\infty}\mu^{[n]} \in \mathcal{C}$, where $\mathcal{C}$ is the isotonic cone as defined in Definition 4.2.

Theorem 5.1. *The $\lim_{n\to\infty}\mu^{[n]}$ is a projection of Y onto $\mathcal{C}$ denoted by $Y^* = P(Y|\mathcal{C})$.*

In order to prove the theorem above, we need to show that following three statements hold:

(a) $\lim_{n\to\infty}\mu^{[n]} \in \mathcal{C}$,

(b) $\langle Y - \lim_{n\to\infty}\mu^{[n]}, \lim_{n\to\infty}\mu^{[n]}\rangle = 0$ and

(c) $\langle Y - \lim_{n\to\infty}\mu^{[n]}, Z\rangle \leq 0$ for all $Z \in \mathcal{C}$.

Proof. (a)

After the first iteration, $\mu^{[1]} \in \mathcal{C}$ because each row $\mu^{[2]}_{(i)} \in \mathcal{C}_i$ and we know by Lemma 3.8 that $\mathcal{C}$ is a decomposable cone. Since $\mathcal{C}$ is closed, $\lim_{n\to\infty}\mu^{[n]} \in \mathcal{C}$.

□

Proof. (b)

Instead of projecting the entire matrix, we project row by row, which is done by step $2(b)$ of the algorithm. We define $f^{*[n]}_{(i)}$ as a projection in the univariate setting. By Lemma 4.6, $f^{*[n]}_{(i)} \in C_i$, $\langle f^{[n]}_{(i)} - f^{*[n]}_{(i)}, f^{*[n]}_{(i)}\rangle = 0$ and $\langle f^{[n]}_{(i)} - f^{*[n]}_{(i)}, g^{[n]}_{(i)}\rangle \leq 0$ for all $g^{[n]}_{(i)} \in C_i$. Now, if we let $n \to \infty$ then we can replace $f^{*[n]}_{ij} = \lim_{n\to\infty}\mu^{[n]}_{ij}$ and $\mu^{[n]}_{kj} = \lim_{n\to\infty}\mu^{[n]}_{kj}$ by the way the inner product for vectors is defined. So, we have

$$\langle f^{[n]}_{(i)} - f^{*[n]}_{(i)}, f^{*[n]}_{(i)}\rangle = 0$$

$$\sum_{j=1}^{q}(f^{[n]}_{ij} - f^{*[n]}_{ij})' * a^{[j]}_{ii} * f^{*[n]}_{ij} = 0$$

$$\sum_{j=1}^{q}\left(y_{ij} + \frac{\sum_{k=1}^{i-1}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj} + \sum_{k=i+1}^{p}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj}}{a^{[j]}_{ii}} - \lim_{n\to\infty}\mu^{[n]}_{ij}\right) *$$

$$a^{[j]}_{ii} * \lim_{n\to\infty}\mu^{[n]}_{ij} = 0$$

$$\sum_{j=1}^{q}\left(y_{ij}a^{[j]}_{ii} + \sum_{k=1}^{i-1}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj} + \sum_{k=i+1}^{p}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj} - \lim_{n\to\infty}\mu^{[n]}_{ij} * a^{[j]}_{ii}\right) *$$

$$\lim_{n\to\infty}\mu^{[n]}_{ij} = 0$$

$$\sum_{j=1}^{q}\left(y_{ij}a^{[j]}_{ii} - \lim_{n\to\infty}\mu^{[n]}_{ij}a^{[j]}_{ii} + \sum_{k=1}^{i-1}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj} + \sum_{k=i+1}^{p}(y_{kj} - \lim_{n\to\infty}\mu^{[n]}_{kj})a^{[j]}_{kj}\right) *$$

$$\lim_{n\to\infty} \mu_{ij}^{[n]} = 0$$

$$\sum_{j=1}^{q} \left(\sum_{k=1}^{p} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} \right) * \lim_{n\to\infty} \mu_{ij}^{[n]} = 0$$

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_i * \lim_{n\to\infty} \mu_{ij}^{[n]} = 0,$$

for all $i = 1, \ldots p$ and e_i is the i^{th} column of the identity matrix with dimension $p \times p$. By Lemma 4.8

$$\langle Y - \lim_{n\to\infty} \mu^{[n]}, \lim_{n\to\infty} \mu^{[n]} \rangle = \sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * \lim_{n\to\infty} \mu_j^{[n]}.$$

This statement is the same as

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * (e_1 \lim_{n\to\infty} \mu_{1j}^{[n]} + \cdots + e_p \lim_{n\to\infty} \mu_{pj}^{[n]}) =$$

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_1 \lim_{n\to\infty} \mu_{1j}^{[n]} + \cdots + \sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_p \lim_{n\to\infty} \mu_{pj}^{[n]}$$

$$= 0 + \cdots + 0$$

Therefore, $\langle Y - \lim_{n\to\infty} \mu^{[n]}, \lim_{n\to\infty} \mu^{[n]} \rangle = 0$.

□

Proof. (c)

We can use similar approach as in (b). Let $g = (g_1, \ldots, g_q) \in \mathcal{C}$ be a vector of $\mathbb{R}^q$ and $Z = (Z_1, \ldots, Z_q) \in \mathcal{C}$ be a matrix of dimension $p \times q$. If n goes to infinity then $g_j = z_{ij}$,$f_{ij}^{*[n]} = \lim_{n\to\infty} \mu_{ij}^{[n]}$ and $\mu_{kj}^{[n]} = \lim_{n\to\infty} \mu_{kj}^{[n]}$ so that

$$f_{(i)}^{[n]} - f_{(i)}^{*[n]}, g \rangle \leq 0$$

$$\sum_{j=1}^{q} (f_{ij}^{[n]} - f_{ij}^{*[n]})' * a_{ii}^{[j]} * g_j \leq 0$$

$$\sum_{j=1}^{q} \left(y_{ij} + \frac{\sum_{k=1}^{i-1} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} + \sum_{k=i+1}^{p} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]}}{a_{ii}^{[j]}} - \lim_{n\to\infty} \mu_{ij}^{[n]} \right) *$$

$$a_{ii}^{[j]} * z_{ij} \leq 0$$

$$\sum_{j=1}^{q} \left(y_{ij} a_{ii}^{[j]} + \sum_{k=1}^{i-1} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} + \sum_{k=i+1}^{p} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} - \lim_{n\to\infty} \mu_{ij}^{[n]} * a_{ii}^{[j]} \right) *$$

$$z_{ij} \leq 0$$

$$\sum_{j=1}^{q} \left(y_{ij} a_{ii}^{[j]} - \lim_{n\to\infty} \mu_{ij}^{[n]} a_{ii}^{[j]} + \sum_{k=1}^{i-1} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} + \sum_{k=i+1}^{p} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} \right) *$$

$$z_{ij} \leq 0$$

$$\sum_{j=1}^{q} \left(\sum_{k=1}^{p} (y_{kj} - \lim_{n\to\infty} \mu_{kj}^{[n]}) a_{kj}^{[j]} \right) * z_{ij} \leq 0$$

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_i * z_{ij} \leq 0,$$

for all $i = 1, \dots p$ and e_i is the i^{th} column of the identity matrix with dimension $p \times p$. By Lemma 4.8

$$\langle Y - \lim_{n\to\infty} \mu^{[n]}, Z \rangle = \sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * Z_j.$$

This statement is the same as

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * (e_1 z_{1j} + \cdots + e_p z_{pj}) =$$

$$\sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_1 z_{1j} + \cdots + \sum_{j=1}^{q} \left(Y_j - \lim_{n\to\infty} \mu_j^{[n]} \right) * A^{[j]} * e_p z_{pj}$$

$$\leq 0 + \cdots + 0$$

Therefore, $\langle Y - \lim_{n\to\infty} \mu^{[n]}, Z \rangle \leq 0$.

□

So, after a certain amount of finite iteration $Y - \mu^{[n]} \leq \delta_0$, for any δ_0.

6 Conclusion

In this thesis paper, basic concepts of univariate isotonic regression were reviewed. Then, the concepts were explained, an algorithm summarized, and applications presented. Afterwards, we extended the univariate model to the multivariate model. Multivariate isotonic regression was explained along with its concepts and applications. Further, we studied a particular convex cone $\mathcal{C}$, which under the general assumption was decomposable into its univariate isotonic cones. Finally, an algorithm was introduced and we proved that it converges to the multivariate isotonic regression.

Although an algorithm was introduced, we did not write a program for it. The reason for not programming the algorithm is that we left $\mathcal{C}$ very general. In order to develop a code, we would need to specify each $\mathcal{C}_{(i)}$. An example for the algorithm can be found in [7]. There, a very simple order was used and each row had the same ordering. Further, the proof of the conditions for which $\{\mu^{[n]}\}$ converges was omitted. We know if the sequence is bounded then there exists a subsequence which converges. One can show that the subsequence actually converges to multivariate isotonic regression. This implies that the sequence $\{\mu^{[n]}\}$ itself converges to the desired solution. So, finding the actual conditions under which $\{\mu^{[n]}\}$ converges should be studied further.

REFERENCES

References

[1] R. Barlow, D. Bartholomew, J. Bremner, and H. Brunk. *Statistical Inference under Order Restrictions*. John Wiley & Sons, New York, 1972.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, New York, 2004.

[3] R. V. Hogg, A. Craig, and J. W. McKean. *Intro to Mathematical Statistics*. Prentice Hall, Upper Saddle River, New Jersey, 2005.

[4] X. Hu and J. Hansohm. Merge and chop in the computation for isotonic regression. *Journal of Statistical planning and inference*, 138(10), 2008.

[5] C.-I. C. Lee. The min-max algorithm and isotonic regression. *The Annals of Statistics*, 11(2):467–477, 1983.

[6] T. Robertson, F. Wright, and R. Dykstra. *Order Restricted Statistical Inference*. John Wiley, New York, 1988.

[7] S. Sasabuchi, M. Inutsuka, and D. S. Kulatunga. An algorithm for computing multivariate isotonic regression. *Hiroshima Mathematical Journal*, 22(2):551–560, 1992.

[8] M. Silvapulle and P. K. Sen. *Constrained Statistical Inference*. John Wiley, New Jersey, 2005.

[9] R. S. Varga. *Matrix iterative Analysis*. Springer, New York, 2000.