RWTH AACHEN UNIVERSITY

MASTERS THESIS

Fast Projection on Birkhoff polytope

Author:
Mohammad SAIFULLAH

Supervisor:
Dr. Rolf BARDELI

A thesis submitted in fulfilment of the requirements
for the degree of Masters of Science

in the

Research Group Name
Fraunhofer IAIS

February 4, 2016

Declaration of Authorship

I, Mohammad SAIFULLAH, declare that this thesis titled, "Fast Projection on Birkhoff polytope" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

RWTH AACHEN UNIVERSITY

Abstract

Faculty Name
Fraunhofer IAIS

Masters of Science

Fast Projection on Birkhoff polytope

by Mohammad SAIFULLAH

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too…

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor…

Contents

List of Figures

List of Tables

List of Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

Physical Constants

Speed of Light $c = 2.997\,924\,58 \times 10^{8}\,\mathrm{m\,s^{-1}}$ (exact)

List of Symbols

a	distance	m
P	power	W ($\mathrm{J\,s^{-1}}$)
ω	angular frequency	rad

For/Dedicated to/To my…

Chapter 1

Chapter Title Here

1.1 Welcome and Thank You

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

1.2 Learning LaTeX

LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for LaTeX is actually a simple, plain text file that contains *no formatting*. You tell LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the `\emph{text}` command and put the text I want in italics in between the curly braces. This means that LaTeX is a "mark-up" language, very much like HTML.

1.2.1 A (not so short) Introduction to LaTeX

If you are new to LaTeX, there is a very good eBook – freely available online as a PDF file – called, "The Not So Short Introduction to LaTeX". The book's title is typically shortened to just *lshort*. You can download the latest version (as it is occasionally updated) from here: http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf

It is also available in several other languages. Find yours from the list on this page: http://www.ctan.org/tex-archive/info/lshort/

It is recommended to take a little time out to learn how to use LaTeX by creating several, small 'test' documents, or having a close look at several templates on: http://www.LaTeXTemplates.com

Making the effort now means you're not stuck learning the system when what you *really* need to be doing is writing your thesis.

1.2.2 A Short Math Guide for LaTeX

If you are writing a technical or mathematical thesis, then you may want to read the document by the AMS (American Mathematical Society) called, "A Short Math Guide for LaTeX". It can be found online here: `http://www.ams.org/tex/amslatex.html` under the "Additional Documentation" section towards the bottom of the page.

1.2.3 Common LaTeX Math Symbols

There are a multitude of mathematical symbols available for LaTeX and it would take a great effort to learn the commands for them all. The most common ones you are likely to use are shown on this page: `http://www.sunilpatel.co.uk/latex-type/latex-math-symbols/`

You can use this page as a reference or crib sheet, the symbols are rendered as large, high quality images so you can quickly find the LaTeX command for the symbol you need.

1.2.4 LaTeX on a Mac

The LaTeX distribution is available for many systems including Windows, Linux and Mac OS X. The package for OS X is called MacTeX and it contains all the applications you need – bundled together and pre-customised – for a fully working LaTeX environment and workflow.

MacTeX includes a custom dedicated LaTeX editor called TeXShop for writing your '**`.tex`**' files and BibDesk: a program to manage your references and create your bibliography section just as easily as managing songs and creating playlists in iTunes.

1.3 Getting Started with this Template

If you are familiar with LaTeX, then you should explore the directory structure of the template and then proceed to place your own information into the *THESIS INFORMATION* block of the **`main.tex`** file. You can then modify the rest of this file to your unique specifications based on your degree/university. Section 1.5 on page 4 will help you do this. Make sure you also read section 1.7 about thesis conventions to get the most out of this template.

If you are new to LaTeX it is recommended that you carry on reading through the rest of the information in this document.

Before you begin using this template you should ensure that its style complies with the thesis style guidelines imposed by your institution. In most cases this template style and layout will be suitable. If it is not, it may only require a small change to bring the template in line with your institution's recommendations. These modifications will need to be done on the **`MastersDoctoralThesis.cls`** file.

1.3.1 About this Template

This LaTeX Thesis Template is originally based and created around a LaTeX style file created by Steve R. Gunn from the University of Southampton (UK), department of

Electronics and Computer Science. You can find his original thesis style file at his site, here: http://www.ecs.soton.ac.uk/~srg/softwaretools/document/templates/

Steve's **ecsthesis.cls** was then taken by Sunil Patel who modified it by creating a skeleton framework and folder structure to place the thesis files in. The resulting template can be found on Sunil's site here: http://www.sunilpatel.co.uk/thesis-template

Sunil's template was made available through http://www.LaTeXTemplates.com where it was modified many times based on user requests and questions. Version 2.0 and onwards of this template represents a major modification to Sunil's template and is, in fact, hardly recognisable. The work to make version 2.0 possible was carried out by Vel and Johannes Böttcher.

1.4 What this Template Includes

1.4.1 Folders

This template comes as a single zip file that expands out to several files and folders. The folder names are mostly self-explanatory:

Appendices – this is the folder where you put the appendices. Each appendix should go into its own separate **.tex** file. An example and template are included in the directory.

Chapters – this is the folder where you put the thesis chapters. A thesis usually has about six chapters, though there is no hard rule on this. Each chapter should go in its own separate **.tex** file and they can be split as:

- Chapter 1: Introduction to the thesis topic
- Chapter 2: Background information and theory
- Chapter 3: (Laboratory) experimental setup
- Chapter 4: Details of experiment 1
- Chapter 5: Details of experiment 2
- Chapter 6: Discussion of the experimental results
- Chapter 7: Conclusion and future directions

This chapter layout is specialised for the experimental sciences.

Figures – this folder contains all figures for the thesis. These are the final images that will go into the thesis document.

1.4.2 Files

Included are also several files, most of them are plain text and you can see their contents in a text editor. After initial compilation, you will see that more auxiliary files are created by LaTeX or BibTeX and which you don't need to delete or worry about:

example.bib – this is an important file that contains all the bibliographic information and references that you will be citing in the thesis for use with BibTeX. You can write it manually, but there are reference manager programs available that will create and manage it for you. Bibliographies in LaTeX are a large subject and you may need to read about BibTeX before starting with this. Many modern reference managers will allow you to export your references in BibTeX format which greatly eases the amount of work you have to do.

MastersDoctoralThesis.cls – this is an important file. It is the class file that tells LaTeX how to format the thesis. If you need to change the layout or structure of the thesis, you will likely need to open this file and find the part relevant to what you are trying to do.

main.pdf – this is your beautifully typeset thesis (in the PDF file format) created by LaTeX. It is supplied in the PDF with the template and after you compile the template you should get an identical version.

main.tex – this is an important file. This is the file that you tell LaTeX to compile to produce your thesis as a PDF file. It contains the framework and constructs that tell LaTeX how to layout the thesis. It is heavily commented so you can read exactly what each line of code does and why it is there. After you put your own information into the *THESIS INFORMATION* block – you have now started your thesis!

Files that are *not* included, but are created by LaTeX as auxiliary files include:

main.aux – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file.

main.bbl – this is an auxiliary file generated by BibTeX, if it is deleted, BibTeX simply regenerates it when you run the 'main' file. Whereas the `.bib` file contains all the references you have, this `.bbl` file contains the references you have actually cited in the thesis and is used to build the bibliography section of the thesis.

main.blg – this is an auxiliary file generated by BibTeX, if it is deleted BibTeX simply regenerates it when you run the main `.tex` file.

main.lof – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It tells LaTeX how to build the *List of Figures* section.

main.log – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It contains messages from LaTeX, if you receive errors and warnings from LaTeX, they will be in this `.log` file.

main.lot – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It tells LaTeX how to build the *List of Tables* section.

main.out – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file.

So from this long list, only the files with the `.bib`, `.cls` and `.tex` extensions are the most important ones. The other auxiliary files can be ignored or deleted as LaTeX and BibTeX will regenerate them.

1.5 Filling in Your Information in the `main.tex` File

You will need to personalise the thesis template and make it your own by filling in your own information. This is done by editing the `main.tex` file in a text editor.

Open the file and scroll down to the second large block titled *THESIS INFORMATION* where you can see the entries for *University Name*, *Department Name*, etc ...

Fill out the information about yourself, your group and institution. You can also insert web links, if you do, make sure you use the full URL, including the `http://` for this. If you don't want these to be linked, simply remove the `\href{url}{name}` and only leave the name.

When you have done this, save the file and recompile `main.tex`. All the information you filled in should now be in the PDF, complete with web links. You can now begin your thesis proper!

1.6 The `main.tex` File Explained

The **`main.tex`** file contains the structure of the thesis. There are plenty of written comments that explain what pages, sections and formatting the LaTeX code is creating. Each major document element is divided into commented blocks with titles in all capitals to make it obvious what the following bit of code is doing. Initially there seems to be a lot of LaTeX code, but this is all formatting, and it has all been taken care of so you don't have to do it.

Begin by checking that your information on the title page is correct. For the thesis declaration, your institution may insist on something different than the text given. If this is the case, just replace what you see with what is required in the *DECLARATION PAGE* block.

Then comes a page which contains a funny quote. You can put your own, or quote your favourite scientist, author, person, and so on. Make sure to put the name of the person who you took the quote from.

Following this is the abstract page which summaries your work in a condensed way and can almost be used as a standalone document to describe what you have done. The text you write will cause the heading to move up so don't worry about running out of space.

Next come the acknowledgements. On this page, write about all the people who you wish to thank (not forgetting parents, partners and your advisor/supervisor).

The contents pages, list of figures and tables are all taken care of for you and do not need to be manually created or edited. The next set of pages are more likely to be optional and can be deleted since they are for a more technical thesis: insert a list of abbreviations you have used in the thesis, then a list of the physical constants and numbers you refer to and finally, a list of mathematical symbols used in any formulae. Making the effort to fill these tables means the reader has a one-stop place to refer to instead of searching the internet and references to try and find out what you meant by certain abbreviations or symbols.

The list of symbols is split into the Roman and Greek alphabets. Whereas the abbreviations and symbols ought to be listed in alphabetical order (and this is *not* done automatically for you) the list of physical constants should be grouped into similar themes.

The next page contains a one line dedication. Who will you dedicate your thesis to?

Finally, there is the block where the chapters are included. Uncomment the lines (delete the `%` character) as you write the chapters. Each chapter should be written in its own file and put into the *Chapters* folder and named **`Chapter1`**, **`Chapter2`**, etc... Similarly for the appendices, uncomment the lines as you need them. Each appendix should go into its own file and placed in the *Appendices* folder.

After the preamble, chapters and appendices finally comes the bibliography. The bibliography style (called *`authoryear`*) is used for the bibliography and is a fully featured style that will even include links to where the referenced paper can be found online. Do not underestimate how grateful your reader will be to find that a reference to a paper is just a click away. Of course, this relies on you putting the URL information into the BibTeX file in the first place.

1.7 Thesis Features and Conventions

To get the best out of this template, there are a few conventions that you may want to follow.

One of the most important (and most difficult) things to keep track of in such a long document as a thesis is consistency. Using certain conventions and ways of doing things (such as using a Todo list) makes the job easier. Of course, all of these are optional and you can adopt your own method.

1.7.1 Printing Format

This thesis template is designed for double sided printing (i.e. content on the front and back of pages) as most theses are printed and bound this way. This means that the inner margin is always wider than the outer for binding. Four out of five people will now judge the margins by eye and think, “I never noticed that before”. Switching to one sided printing is as simple as uncommenting the `oneside` option of the `documentclass` command at the top of the **`main.tex`** file. You may then wish to adjust the margins to suit specifications from your institution.

The headers for the pages contain the page number on the outer side (so it is easy to flick through to the page you want) and the chapter name on the inner side.

The text is set to 11 point by default with single line spacing, again, you can tune the text size and spacing should you want or need to using the options at the very start of **`main.tex`**. The spacing can be changed similarly by replacing the `singlespacing` with `onehalfspacing` or `doublespacing`.

1.7.2 Using US Letter Paper

The paper size used in the template is A4, which is the standard size in Europe. If you are using this thesis template elsewhere and particularly in the United States, then you may have to change the A4 paper size to the US Letter size. To do this, you will need to open the **`MastersDoctoralThesis.cls`** file and navigate to the `MARGINS` block where you can change `a4paper` to `letterpaper`.

Due to the differences in the paper size, the resulting margins may be different to what you like or require (as it is common for institutions to dictate certain margin sizes). If this is the case, then the margin sizes can be tweaked by modifying the values in the same block as where you set the paper size. Now your document should be set up for US Letter paper size with suitable margins.

1.7.3 References

The `biblatex` package is used to format the bibliography and inserts references such as this one (Pak, 1999). The options used in the **`main.tex`** file mean that the in-text citations of references are formatted with the author(s) listed with the date of the publication. Multiple references are separated by semicolons (e.g. (Paffenholz, 2013; Pak, 1999)) and references with more than three authors are only show the first author with *et al.* indicating there are more authors (e.g. (Jesus A. De Loera, 2007)). This is done automatically for you. To see how you use references, have a look at the **`Chapter1.tex`** source file. Many reference managers allow you to simply drag the reference into the document as you type.

Scientific references should come *before* the punctuation mark if there is one (such as a comma or period). The same goes for footnotes[1]. You can change this but the most important thing is to keep the convention consistent throughout the thesis. Footnotes themselves should be full, descriptive sentences (beginning with a capital letter and ending with a full stop). The APA6 states: "Footnote numbers should be superscripted, [...], following any punctuation mark except a dash." The Chicago manual of style states: "A note number should be placed at the end of a sentence or clause. The number follows any punctuation mark except the dash, which it precedes. It follows a closing parenthesis."

The bibliography is typeset with references listed in alphabetical order by the first author's last name. This is similar to the APA referencing style. To see how LaTeX typesets the bibliography, have a look at the very end of this document (or just click on the reference number links in in-text citations).

A Note on bibtex

The bibtex backend used in the template by default does not correctly handle unicode character encoding (i.e. "international" characters). You may see a warning about this in the compilation log and, if your references contain unicode characters, they may not show up correctly or at all. The solution to this is to use the biber backend instead of the outdated bibtex backend. This is done by finding this in **`main.tex`**: *`backend=bibtex`* and changing it to *`backend=biber`*. You will then need to delete all auxiliary BibTeX files and navigate to the template directory in your terminal (command prompt). Once there, simply type `biber main` and biber will compile your bibliography. You can then compile **`main.tex`** as normal and your bibliography will be updated. An alternative is to set up your LaTeX editor to compile with biber instead of bibtex, see here for how to do this for various editors.

1.7.4 Tables

Tables are an important way of displaying your results, below is an example table which was generated with this code:

```
\begin{table}
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\tabhead{Groups} & \tabhead{Treatment X} & \tabhead{Treatment Y} \\
\midrule
1 & 0.2 & 0.8\\
2 & 0.17 & 0.7\\
3 & 0.24 & 0.75\\
4 & 0.68 & 0.3\\
\bottomrule\\
\end{tabular}
\end{table}
```

You can reference tables with `\ref{<label>}` where the label is defined within the table environment. See **`Chapter1.tex`** for an example of the label and citation (e.g. Table 1.1).

[1] Such as this footnote, here down at the bottom of the page.

TABLE 1.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

1.7.5 Figures

There will hopefully be many figures in your thesis (that should be placed in the *Figures* folder). The way to insert figures into your thesis is to use a code template like this:

```
\begin{figure}
\centering
\includegraphics{Figures/Electron}
\decoRule
\caption[An Electron]{An electron (artist's impression).}
\label{fig:Electron}
\end{figure}
```

Also look in the source file. Putting this code into the source file produces the picture of the electron that you can see in the figure below.

FIGURE 1.1: An electron (artist's impression).

Sometimes figures don't always appear where you write them in the source. The placement depends on how much space there is on the page for the figure. Sometimes there is not enough room to fit a figure directly where it should go (in relation to the text) and so LaTeX puts it at the top of the next page. Positioning figures is the job of LaTeX and so you should only worry about making them look good!

Figures usually should have captions just in case you need to refer to them (such as in Figure 1.1). The `\caption` command contains two parts, the first part, inside the square brackets is the title that will appear in the *List of Figures*, and so should be short. The second part in the curly brackets should contain the longer and more descriptive caption text.

The `\decoRule` command is optional and simply puts an aesthetic horizontal line below the image. If you do this for one image, do it for all of them.

LaTeX is capable of using images in many formats such as PDF, JPEG, PNG and more.

1.7.6 Typesetting mathematics

If your thesis is going to contain heavy mathematical content, be sure that LaTeX will make it look beautiful, even though it won't be able to solve the equations for you.

The "Not So Short Introduction to LaTeX" (available on CTAN) should tell you everything you need to know for most cases of typesetting mathematics. If you need more information, a much more thorough mathematical guide is available from the AMS called, "A Short Math Guide to LaTeX" and can be downloaded from: `ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf`

There are many different LaTeX symbols to remember, luckily you can find the most common symbols here. You can use the web page as a quick reference or crib sheet and because the symbols are grouped and rendered as high quality images (each with a downloadable PDF), finding the symbol you need is quick and easy.

You can write an equation, which is automatically given an equation number by LaTeX like this:

```
\begin{equation}
E = mc^{2}
\label{eqn:Einstein}
\end{equation}
```

This will produce Einstein's famous energy-matter equivalence equation:

$$E = mc^2 \tag{1.1}$$

All equations you write (which are not in the middle of paragraph text) are automatically given equation numbers by LaTeX. If you don't want a particular equation numbered, use the unnumbered form:

```
\[ a^{2}=4 \]
```

1.8 Sectioning and Subsectioning

You should break your thesis up into nice, bite-sized sections and subsections. LaTeX automatically builds a table of Contents by looking at all the `\chapter{}`, `\section{}` and `\subsection{}` commands you write in the source.

The Table of Contents should only list the sections to three (3) levels. A `chapter{}` is level zero (0). A `\section{}` is level one (1) and so a `\subsection{}` is level two (2). In your thesis it is likely that you will even use a `subsubsection{}`, which is level three (3). The depth to which the Table of Contents is formatted is set within **`MastersDoctoralThesis.cls`**.

1.9 In Closing

You have reached the end of this mini-guide. You can now rename or overwrite this pdf file and begin writing your own **`Chapter1.tex`**' and the rest of your thesis. The easy work of setting up the structure and framework has been taken care of for you. It's now your job to fill it out!

Good luck and have lots of fun!

Guide written by —
Sunil Patel: www.sunilpatel.co.uk
Vel: LaTeXTemplates.com

Chapter 2

The Birkhoff Polytope

In this Chapter I am going to describe about Birkhoff polytope B_n which is sometimes considered to be one of the most important polytopes in many sphere. Before that I will define some basic types.

2.1 Hypermatrix

A generalisation of matrix to an $n_1 \times n_1 \times$ ... array of numbers.
Example: 3-dimensional hypermatrix(cube). It is a $2 \times 2 \times 2$ matrix over $\mathbb{R}$.

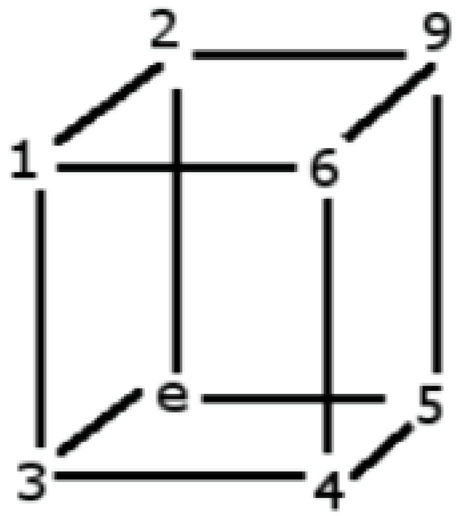

FIGURE 2.1: 3-dimensional hypermatrix(cube).

2.2 Doubly-Stochastic Matrix

Doubly Stochastic Matrix is a square matrix of non-negative real numbers, each of whose rows and columns sum to 1. If rows and columns of some matrices sum to n, it can also be considered as doubly stochastic matrix because we can just device all the entries by n, which results to true Doubly Stochastic.

Examples:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1/6 & 5/6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1/6 & 0 & 5/6 \\ 5/6 & 0 & 0 & 1/6 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

2.3 Permutation Matrix

A matrix obtained by permuting the rows of an $n \times n$ identity matrix according to some permutation of the numbers 1 to n. So the number of $n \times n$ permutation matrics is $n!$.

The permutation matrics of order 3 are:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

2.4 Polytopes

In elementary geometry, a polytope is a geometric object with flat sides, and may exist in any general number of dimensions n as an n-dimensional polytope or n-polytope. For example a two-dimensional polygon is a 2-polytope and a three-dimensional polyhedron is a 3-polytope

Mathematically, A polytope $P \subseteq \mathbb{R}^d$ is the convex hull $P = conv(v_1, ..., v_k)$ of a finite set of points $v_1, ..., v_k \in \mathbb{R}^d$. Dually any polytope can be written as the bounded intersection of a finite number of affine half-spaces in the form $P = \{x \mid Ax \leq b\}$

2.4.1 Elements of polytope

A proper face of F of a polytope P is the intersection of P with and affine hyperplane H such that P is completely contained in one of the closed half spaces defined by H. The empty set and the polytope P a face of P. Any face F is itself polytope. The dimension of a polytope $P \subseteq \mathbb{R}^d$ is the dimension of the minimum affine space containin it. It is full dimensional if its dimension is d.

0-dimensional faces of P are called $vertices$, 1-dimensional faces are edges. Proper faces of maximal dimensional are called facets. P is the convex hull of its vertices, and the vertices of any face are subset of the vertices of P. Thus, a polytope has only a finite number of faces. Let f_i be the number of i-dimensional faces of $P, 0 \leq i \leq dimP - 1$. The f-vector of a d-dimensional polytope P is the non-negative integral vector $f(P) = (f_0, ..., f_(d-1)$.

The face lattice or combinatorial type $\mathcal{L}(P)$ of a polytope P is the partially ordered set of all faces of P (including the empty face and P itself). This defines Eulerian lattice. Figure shows 2.1 this lattice.

It contains all combinatorial information of the polytope. Two polytopes P,P' are combinatorially isomorphic or have the same combinatorial type if their face lattices are isomorphic as posets.

An r-dimensional simplex (or r-simplex) is the convex hull of r+1 affinely independent points in $\mathbb{R}^d$. A polytope is called simplicial if all facets are simplices. It is simple if the dual is simplicial. Equally, a d-dimensilanl polytope P is simple if each vertex is incident to precisely d edges. The d-dimensional 0/1-cube C^d is the convex hull of all d-dimensinal 0/1-vectors. This is a simple d-polytope with 2^d vertices and 2d facets. More generally we denote by a d-cube any d-dimensional polytope that is combinatorially isomorphic to the 0/1-cube (it need not be full dimensional).

2.4.2 Properties of Polytope

Let $P_1 \subset \mathbb{R}^{d_1}$ and $P_2 \subset \mathbb{R}^{d_2}$ be two (geometrically realised) polytopes with vertex sets $V(P_1) = \{v_1, ..., v_k$ and $V(P_2) = \{w_1, ..., w_l\}$. With $0^{(d)}$ we denote he d-dimensional zero vector.

The (geometric) product of P_1 and P_2 is the polytope

$$P_1 \times P_2 = conv((v_i, w_i) \in \mathbb{R}^{d_1+d_2} \mid 1 \leq i \leq k, i \leq j \leq l) \tag{2.1}$$

This is the same as the set of all points (v, w) for $v \in P_1$ and $w \in P_2$. The (geometric) join of P_1 and P_2 is the polytope

$$P_1 \star P_2 := conv(P_1 \times \{0^{d_2}\} \times \{0\} \cup \{0^{d_1}\} \times P_2 \times \{1\}) \subseteq \mathbb{R}^{d_1+d_2+1} \tag{2.2}$$

More generally we say that a polytope P is a product or join of two polytopes P_1 and P_2, if P is combinatorially isomorphic to the geometric prouct or geometric join of some realisations of the face lattices P_1, or P_2.

If F is face of a polytope $P = \{x \mid Ax \leq b\} \subseteq \mathbb{R}^d$ and $\langle c, x\rangle \leq d$ a linear funtional defining F, then the $wedgewedge_F(P)$ of P over F is defined to be the polytope.

$$wedge_F(P) = \{(x, x_0) \in \mathbb{R}^{d+1} \mid Ax \leq b, 0 \leq x_0 \leq d - \langle c, x\rangle\} \tag{2.3}$$

Again, we say more generally that P is wedge of a polytope Q over some face F of Q if P is combinatorially equivalently to $wedge_F(Q)$

2.5 Birkhoff Polytope

Birkhoff polytope B_n is sometimes considered to be one of the most important polytopes in many sphere. Birkhoff polytope is also called assignment polytope, the polytope of doubly stochastic matrices, or the perfect matching polytope of complete bipartite graph $K_(n, n)$,transportation polytope. It surprisingly appears in various branches of mathematics from geometry to enumerative combinatorics to optimisation theory to Statistics.

The Birkhoff polytope B_n is the convex hull of all $(n \times n)$ permutation metrices, i.e.matrices which consists precisely one 1 in every row and column, and zeros at all places. Equivalently, B_n is the set of all non-negative $(n \times n)$-metrices, whose rows and columns all sum to 1, or the perfect matching polytope of the complete bipartite graph $K_(n, n)$. The Birkhoff polytope B_n has dimension $(n-1)^2$ with $n!$ vertices and n_2 facets. The Birkhoff-von Neumann Theoram illustrated, B_n can be understood as the intersection of the positive orthant with a family of of hyperplanes.

Birkhoff polytopes are widely studied as a class of polytopes in the area of optimisation, statistics, enumerative combinatorics or representation theory. Despite, Combinatorial and Geometrical structure of Birkhoff polytope and its algorithmic treatment are still open to disover.

A Birkhoff polytope B_n is a polytope defined by the following equations and inequalities:

$$a_{i,j} \geqslant 0, \sum_{i=1}^{n} a_{i,j} = 1, \sum_{j=1}^{n} a_{i,j} = 1 \text{ for all } 1 \leqslant i, j \leqslant n. \tag{2.4}$$

($a_{i,j}$ can be thought of as $n \times n$ doubly stochastic matrices. It can be realise that B_n has dimension $(n-1)^2$ as values of $a_{i,j}, 1 \leqslant i, j \leqslant n$ determine the rest. Different way helps to realise that vertices of P_n are permutation matrices.

2.6 Properties of B_n

Here I am going to describe some of the properties of Birkhoff polytope.

2.6.1 Vetices

The Birkhoff polytope has $n!$ vertices. This was derived from the Birkhoff-von Neumann theorem.

2.6.2 Edges

The edges of the Birkhoff polytope corresponds to pairs of permutations differing by a cycle:

(σ, ω) such that $\sigma^{-1}\omega$ is a cycle.

This implies that the graph of B_n is a Caley graph of the symmetric group S_n. This ialso implies that the graph of B_3 is a complete graph K_6, and thus B_3 is a neighbourly prototype.

2.6.3 Facets

The Birkhoff polytope lies within and $(n^2 - 2n + 1)$-dimensional affine subspace of the n^2-dimensional space of all $n \times n$ matrices: this subspace is determined by the linear equality constraints that the sum of each row and each column be one. Within this subspace, it is defined by n^2 linear inequalities, one for each coordinate of the matrix, specifying that the coordinate be non-negative. Therefore, it has exactly n^2 facets.

2.6.4 Symmetries

The Birkhoff polytop B_n is both vertex-transitive and facet-transitive. This is not regular for $n > 2$.

2.6.5 Volume

One of the hardest open problem is to find the volume of a Birkhoff polytopes. Volume calculation was possible for $n \leqslant 10$. It is known to be equal to the volume of polytope associated with standard Young tabuleaux. The following asymptotic formula was founded by Rodney Canifeld and Brenden McKay:

$$vol(B_n) = \exp(-(n-1)^2 \ln n + n^2 - (n - \frac{1}{2}) \ln(2\pi) + \frac{1}{3} + \mathcal{O}(1)) \tag{2.5}$$

Chapter 3

Quadratic Programming

An optimisation problem with a quadratic objective function and linear constraints is called a quadratic program. Problem of this type are important in their own right, and they also arise as subproblems in methods for general constrained optimisations.

3.1 Definition of QP

The general quadratic program(QP) can be stated as

$$\begin{array}{lll} \min\limits_{x} & q(x) = \frac{1}{2}x^T G x + x^T c & \\ \text{subject to} & a_i^T x = b_i & i \in \mathbb{E} \\ & a_i^T x \geqslant b_i & i \in \mathbb{I} \end{array} \tag{3.1}$$

where G is a symmetric $n \times n$ matrix, E and I are finite sets of indices, and c,x, and $a_i, i \in E \cup I$, are vectors in $\mathbb{R}^n$. Quadratic programs can always be solved (or shown to be infeasible) in a finite amount of computation, but the effort required to find a solution depends strongly on the characteristics of the objective function and the number of inequality constraints. If the Hessian matrix G is positive semidefinite, we say that 3.1 is a convex QP, and in this case the problem is often similar in difficulty to a linear program. (Strictly convex QPs are those in which G is positive definite.) Nonconvex QPs, in which G is an indefinite matrix, can be more challenging because they can have several stationary points and local minima.

In this chapter we will try to show different kinds of quadratic programs but we will mainly focus on convex quadratic program and different algorithms of convex quadratic programs.

3.2 Classification of QP's

Some classification of quadratic program's:

- Unconstriained QP
- Box constrained QP
- Equality constrained QP
- Inequality constrained QP.

Important Definitions

Logarithmic barrier

Consider inequalities $Ax \leqslant b$ with A of size $m \times n$ and with rows a_i^T. Define

$$P = \{x \mid Ax \leqslant b\} \text{ and } P^0 = \{x \mid Ax < b\}$$

logarithmic barrier for the inequalities $Ax \leqslant b$:

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x) \quad \text{with domain } P^0$$

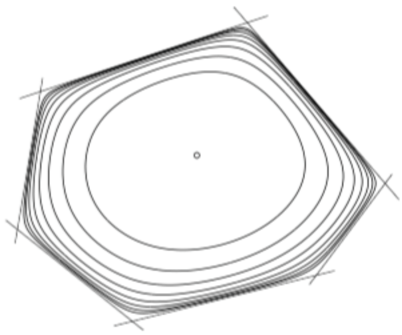

FIGURE 3.1: Logarithmic barrier function.

Gradient

Gradient $\nabla\phi(x)$ is the n-vector with $\nabla\phi(x)_i = \frac{\delta\phi(x)}{\delta x_i}$

$$\nabla\phi(x) = \sum_{k=1}^{m} \frac{1}{b_k - a_k^T x} a_k = A^T d_x$$

d_x denotes the positive $m-$vector

$$d_x = \left(\frac{1}{b_1 - a_1^T x}, ..., \frac{1}{b_m - a_m^T x}\right)$$

KKT System

KKT System is

Hessian Matrix

Hessian Matrix Gradient $\nabla^2\phi(x)$ is the $n \times n$-matrix with $\nabla\phi(x)_{ij} = \frac{\delta^2\phi(x)}{\delta x_i \delta x_j}$

$$\nabla^2\phi(x) = \sum_{k=1}^{m} \frac{1}{(b_k - a_k^T x)^2} a_k a_k^T = A^T diag(d_x)^2 A$$

Central Path

Consider the linear programming problem in standard form:

$$\min c^T x, \quad \text{subject to} Ax = b, \geqslant 0$$

where c and x are vectors in $\mathbb{R}^n$, b is a vector in $\mathbb{R}^m$, and A is an $m \times n$ matrix with full row rank. The dual problem of the above problem is

$$\max b^T \lambda, \quad \text{subject to} A^T \lambda + s = c, x \geqslant 0$$

where λ is a vector in $\mathbb{R}^m$ and s is a vector in $\mathbb{R}^n$ The primal-dual feasible set $\mathcal{F}$ and strictly feasible set $\mathcal{F}^0$ are defined as

$$\mathcal{F} = \{(x, \lambda, s) \mid Ax = b, A^T \lambda + s = c, (x, s) \geqslant 0\}$$
$$\mathcal{F}^0 = \{(x, \lambda, s) \mid Ax = b, A^T \lambda + s = c, (x, s) > 0\}$$

The central path C is an arc of strictly feasible points. It is parameterized by a scalar $\eta > 0$, and each point $(x_\tau, \lambda_\tau, s_\tau) \in \mathcal{C}$ satisfies the following equations:

$$\begin{aligned} A^T \lambda + s = c, & \\ Ax = b, x_i, s_i & = \tau \quad i = 1, 2, ...n. \\ (x, s) > 0. & \end{aligned}$$

So the central path is defined as:

$$\mathcal{C} = \{(x_\tau, \lambda_\tau, s_\tau) \mid \tau > 0\}$$

Another way of defining $\mathcal{C}$ is:

$$F(x_\tau, \lambda_\tau, s_\tau) = \begin{bmatrix} 0 \\ 0 \\ \tau e \end{bmatrix}, (x_\tau, s_\tau) > 0$$

the conditions approximate more and more colsely as τ goes to zero. If $\mathcal{C}$ converges to anything as $\tau \downarrow 0$, it must converge to a primal-dual solution of the linear program.

Active set

The Active set $\mathbb{A}(x)$ at any feasible x consists of the equality constraint indices from ε together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

$$\mathbb{A}(x) = \varepsilon \cup i \in \mathbb{I} \mid c_i(x) = 0$$

At a feasible point x, the inequality constraint $i \in \mathbb{I}$ is said to be active if $c_i(x) = 0$ and inactive if the strict inequality $c_i(x) > 0$ is satisfied.

3.3 Equality-Constrained Quadratic Programs

We start this section with of algorithms for quadratic programming by considering the case of equality constrained

Properties of Equality-constrained QPs

To make it simple, we write the equality constraint in a form of matrix and define it as follows:

$$\begin{aligned} \min_{x} \quad & q(x) = \frac{1}{2}x^T G x + x^T c \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{3.2}$$

where A is the $m \times n$ Jacobian of constraints (with $m \leqslant n$) whose rows are $a_i^T, i \in \mathbb{E}$ and b is the vecotor in $\mathbb{R}^n$ whose componants are $b_i, i \in \mathbb{E}$. Currently, we consider that A has a full row rank (rank m) so that the constraints are consistent.

The first-order necessary conditions for x^* to be a solution of 3.2 state that there is a vector λ^* such that the following system of equations is satisfied:

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix} \tag{3.3}$$

These conditions are a consequence of the general result for first-order optimality conditions. λ^* is called the vector of Lagrange multipliers. In 3.3 we can write x^* as $x^* = x + p$ which makes it useful for computation, where x is some estimate of the solution and p is the desired step. By introducing this and rearranging the equations, we obtain

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda* \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix} \tag{3.4}$$

where

$$h = Ax - b, g = c + Gx, \quad p = x^* - x. \tag{3.5}$$

The matrix 3.4 is called the Karush-Kuhn-Tucker (KKT) matrix, and the following result gives conditions under which it is nonsingular. We will use Z to denote the $n \times (n-m)$ matrix whose columns are basis for the null space of A. That is, z has full rank and satisfies $AZ = 0$.

Lemma

Lemma 3.3.1. *Let A have full row rank, and assume that the reduce Hessian matrix $Z^T G Z$ is positive definite. Then the KKT matrix*

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \tag{3.6}$$

is nonsingular, and hence there is a unique vector pair (x^*, λ^*) satisfying 3.3.

♣

So, when the conditions of the 3.3.1 is satisfied, there is a unique vector pair (x^*, λ^*) that satiesfies the first-order necessary condition for 3.2. In fact, the second order sufficient conditions are also satisfied at (x^*, λ^*), so x^* is a strict local minimizer of 3.2. In fact we can use a direct argument to show that x^* is a global solution of 3.2.

Theorem

Theorem 3.3.2. *Let A have full row rank and assume that the reduced-Hessian matrix $Z^T G Z$ is positive definite. Then the vector x^* satisfying 3.3 is the unique global solution of 3.2.*

♣

Proof. Let x be any other feasible point (satisfying $Ax = b$), and as before, let p denote the difference $x^* - x$. Since $Ax^* = Ax = b$, we have that $Ap = 0$. By substituting into the objective function 3.2, we get

$$\begin{aligned} q(x) &= \frac{1}{2}(x^* - p)^T G(x^* - p) + C^T(x^* - p) \\ &= \frac{1}{2}p^T Gp - p^T Gx^* - C^T p + q(x^*) \end{aligned} \tag{3.7}$$

From 3.3 we have that $Gx^* = -c + A^T\lambda^*$, so from $Ap = 0$ we have that

$$p^T Gx^* = p^T(-c) + A^T\lambda^* = -p^T c. \tag{3.8}$$

By substituting this relation into 3.7, we obtain

$$q(x) = \frac{1}{2}p^T Gp + q(x^*). \tag{3.9}$$

Since p lies in the null space of A, we can write $p = Zu$ for some vector $u \in \mathbb{R}^{n-m}$, so that

$$q(x) = \frac{1}{2}u^T Z^T GZu + q(x^*). \tag{3.10}$$

By positive definiteness of $Z^T GZ$, we conclude that $q(x) > q(x^*)$ except when $u = 0$, that is, when $x = x*$. Therefore, x^* is the unique global solution of 3.2 □

When the reduced Hessian matrix $Z^T GZ$ is positive semidefinite with zero eigenvalues, the vector x^* satisfying 3.4 is a local minimizer but not a strict local minimizer. If the reduced Hessian has negative eigenvalues, then x^* is only a stationary point, not a local minimizer.

3.4 Direct Solution of the KKT System

In this section we discuss the efficient methods of solving KKT system. The KKT system is always indefinite if $m \geqslant 1$. We can apply direct techniques to solve indefinite KKT system.

Factoring the Full Scale System

One option for solving KKT system is to perform triangular factorization on the full KKT matrix and then perform backward and forward substitution with the triangular factors. It is not possible to apply Cholesky factorization as it is indefinite.Another option can be Gaussian Elimination to obtain the L and U factors, but this method does not consider the symmetry.

So, the most effective strategy is to use symmetric indefinite factorization which has the form of:

$$P^T KP = LDL^T \tag{3.11}$$

where P is an appropriately chosen permutation matrix. L is lower triangular with $diag(L) = I$ and D is block diagonal. Based on 3.11, the KKT system 3.4 is solved as follows:

$$\begin{aligned} solve \quad Ly &= P^T \begin{bmatrix} g \\ h \end{bmatrix} \\ solve \quad D\hat{y} &= y \\ solve \quad L^T \bar{y} &= \hat{y} \\ set \quad \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} &= P\bar{y} \end{aligned} \tag{3.12}$$

This approach of factoring the full $(n+m) \times (n+m)$ KKt matrix is quite effective on many problems. It can be expensive when the permutation matrix P are not able to maintain sparsity in the L factor.

Range-space approach

The range-space approach is useful when $G \in \mathbb{R}^{n\times n}$ is symmetric positive definite. We can multiply the first part of the equation 3.4 by $AG^{-}1$ and then subtract the second part to obtain a linear system in the vector λ^* alone:

$$(AG^{-1}A^T\lambda^*) = (AG^{-1}g - h) \tag{3.13}$$

We solve this symmetric semidifinite system for λ^* and then recover p from the first equation in 3.4 by solving

$$Gp = A^T\lambda^* - g \tag{3.14}$$

This approach requires us to perform operation with G^{-1}, as well as to compute the factorization of the $m \times n$ matrix $AG^{-1}A^T$. That is why it is useful when:

- G is well conditioned and easily invertible (e.g., G is diagonal or block-diagonal),
- B^{-1} is known explicitely (e.g., by means of a quasi-Newton updating formula),
- the number m of equality constraints is small.

Null-space approach

The null-space approach does not require regularity of G and thus has a wider range of applicability than the range-space approach.

We assume that $A \in \mathbb{R}^{m\times n}$ has full row rank m and that Z^TGZ is positive definite, where $Z \in \mathbb{R}^{n\times(n-m)}$ is the matrix whose columns span Ker A which can be computed by QR factorization.

We partition the vector x^* according to

$$x^* = Yw_y + Zw_z \tag{3.15}$$

where $Y \in \mathbf{R}^{n\times m}$ is such that $[Y \quad Z] \in \mathbb{R}^{n\times n}$ is nonsingular and $w_y \in \mathbb{R}^m, w_Z \in \mathbb{R}^{n-m}$.

Substituting 3.15 into the 3.4, we get

$$\begin{aligned} Ax^* = AYw_Y + AZw_Z &= c \\ AZ &= 0 \end{aligned} \tag{3.16}$$

i.e., Yw_Y is a particular solution of $Ax = c$.

Since $A \in \mathbb{R}^{m\times n}$ has rank m and $[Y \quad Z] \in \mathbb{R}^{n\times n}$ is nonsingular, the product matrix $[Y \quad Z] = [AY \quad 0] \in \mathbb{R}^{m\times m}$ is non singular. Hence, w_Y is well determined by 3.16.

On the otherhand, substituting 3.15 into the first equation of 3.4, we get

$$GYw_Y + GZw_Z + A^T\lambda^* = b. \tag{3.17}$$

Multiplying by Z^T and observing $Z^TA^T = (AZ)^T = 0$ yields

$$Z^TGZw_Z = Z^Tb - Z^TGYw_Y. \tag{3.18}$$

The reduced KKT system 3.18 can be solved by a Cholesky factorization of the reduced Hessian $Z^TBZ \in \mathbb{R}^{(n-m)\times(n-m)}$. Once w_Y and w_Z have been computed as the solution of 3.17 and 3.17, x^* is obtained according to 3.15.

Finally, the Lagrange multiplier turns out to be the solution of the linear system arising from multiplication of the equation 3.15 by Y^T:

$$(AY)^T\lambda^* = Y^Tb - Y^TGx^* \tag{3.19}$$

3.5 Iterative solution of the KKT system

Direct solution of the KKT system can be sometimes expensive, the possible alternative is iterative method to solve KKT system. one of the famous iterative method is Conjugate Gradient method. Although it is not recommended for solving the full system using the CG mathod because it can be unstable on systems that are not positive definite. An iterative solver can be applied either to the entire KKT system or, as in the null-space and range-space approach, use the special structure of the KKT matrix.

3.6 Inequality-Constrained Problems

Inequality-constrained quadratic programs are QPs which consists of inequality constraints and may or may not consist equality constraints. Several classes of algorithms for solving convex quadratic programs containing inequality and equality constraints are available. Active-set method and Interior-point are most famous for their accuracy and capability to solve large problems.

3.6.1 Properties of Inequality-constrained problems

Optimality condition for inequality-constrained problems

Lagrangian for the problem 3.1 is:

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^TGx + x^Tc - \sum_{i\in\mathcal{I}\cup\varepsilon}\lambda_i(a_i^Tx - b_i) \tag{3.20}$$

The active set $\mathcal{A}(x^*)$ consists of indices of the constraints for which equality holds at x^*:

$$\mathcal{A}(x^*) = \{i \in \varepsilon \cup \mathcal{I} \mid a_i^Tx^* = b_i\} \tag{3.21}$$

According to the KKT condition for this problem, it is found that any solution x^* for some Lagrange multipliers $\lambda_i^*, i \in \mathcal{A}(x^*)$ of 3.1 satisfies the following conditions:

$$\begin{aligned} Gx^* + c - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* a_i = 0 & \\ a_i^T x^* = b_i \quad & \text{for all } i \in \mathcal{A}(x^*), \\ a_i^T x^* \geqslant b_i \quad & \text{for all } i \in \mathcal{I}/\mathcal{A}(x^*), \\ \lambda_i^* \geqslant 0, \quad & \text{for all } i \in \mathcal{I} \cap \mathcal{A}(x^*), \end{aligned} \tag{3.22}$$

Incase of convex QP, that is when G is positive semidefinite, 3.22 are sufficient for x^* to be a global solution.

Theorem

Theorem 3.6.1. *If x^* satisfies the conditions 3.22 for some $\lambda_i^*, i \in \mathcal{A}(x^*)$, and G is positive semidefinite, then x^* is a global solution of 3.1.*

Degeneracy

Degeneracy initiates difficulties for some optimisation algorithm. It appears in the following situation:

- the active constraint gradients $a_i, i \in \mathcal{A}(x^*)$, are linearly dependent at the solution x^*, and/or
- there is some index $i \in \mathcal{A}(x^*)$ such that all Lagrange multipliers satifying KKT conditions have $\lambda_i^* = 0$

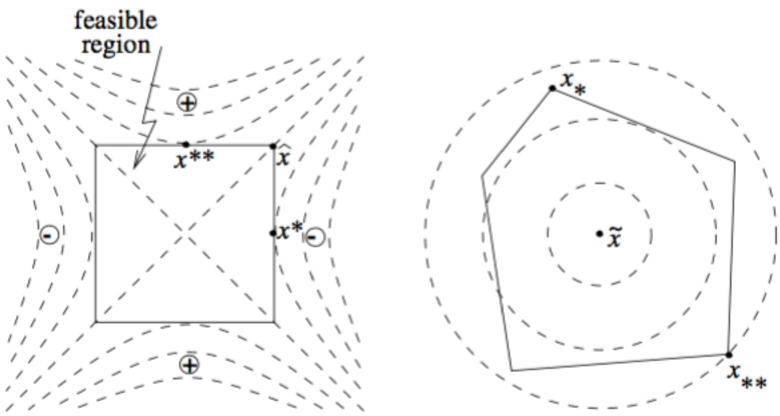

FIGURE 3.2: Nonconvex quadratic programs.

In the figure 3.3 two instances are visualized. There is only a single active constraint at the solution x^* in the left picture, that is lso an unconstrained minimizer of the objective function. According to KKT condition, $Gx^* + c = 0$, such that the lone Lagrange multiplier should be zero. 3 constraints are active at the solution x^* in the right-side image. As each of the three constraint gradients is a a vector in $\mathbb{R}^2$, they should be linearly independent.

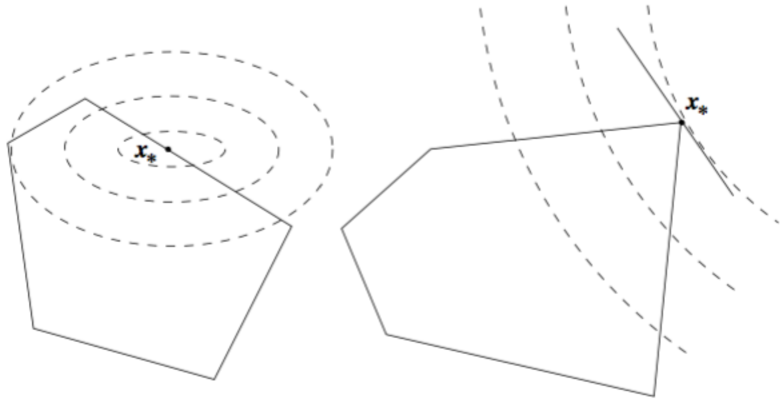

FIGURE 3.3: Degenerate solutions of quadratic programs.

Specifically for two reason Degeneracy can cause problems for optimization algorithms,

- First linear dependence of the active constraint gradients can cause numerical difficulties as several matrices which required to factor become rank deficient.
- Second when the problem contains weakly active constraints, it is difficult for the algorithm to determine whether these constraints are active at the solution.

3.7 Active-set Methods for Convex QPs

Active-set methods are important method for solving convex quadratic programs which consists equality and inequality constraints. Active set method starts by finding a feasible point during initial phase and then search for a solution along the edges and faces of the feasible set by solving a sequence of equality-constrained QPs.

If it was possible to know the contents of the active set earlier, it would be straight forward to get the solution by solving an equality-constrained QP of the form:

$$\begin{aligned} \min_x \quad & q(x) = \frac{1}{2}x^T Gx + x^T c \\ \text{subject to} \quad & a_i^T x = b_i \qquad \forall i \in \mathcal{A}(x^*) \end{aligned}$$

But usually it is not possible to know $\mathcal{A}(x^*)$ and termination of this set is a major challnage for algorithms.

Primal active-set method follows step from one iteration to another by solving a quadratic subproblem in which some of the equality constraints and all the inequality constraints are imposed as equalities which is referred to as the working set. Working set at kth iterate x_k is denoted by $\mathcal{W}_k$

Consider an iterate x_k and the working set $\mathcal{W}_k$, it is necessary to check if x_k minimizes the quadratic q in the subspace defined by working set. Otherwise step p is computed by solving an equality-constrained subproblem in which the constraints corresponding to the working set $\mathcal{W}_k$ are regarded as equalities and all other constraints are temporarily disregarded. We define this subproblem in terms of the step p:

$$p = x - x_k, \quad g_k = Gx_k + c.$$

By substituting x with $(x_k + p)$ into 3.1,

$$q(x) = q(x_k + p) = \frac{1}{2}p^T G p + g_k^T p + \rho_k$$

ρ_k is independent of p. So without considering ρ_k, we can write the QP subproblem to be solved at kth iteration as:

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}p^T G p + g_k^T p \\ \text{subject to} \quad & a_i^T p = 0 \qquad \forall i \in \mathcal{W}_k \end{aligned} \tag{3.23}$$

Solution of this subproblem is denoted by p_k. Consider optimal p_k is nonzero for the moment. We need to calculate displacement along this direction. If $x_k + p_k$ is feasible with respect to all the constraints, then we can set:

$$x_{k+1} = x_k + p_k$$

Otherwise, we set:

$$x_{k+1} = x_k + \alpha_k p_k \tag{3.24}$$

Where α_k is the step length and is chosen possible largest value in the range $[0, 1]$ for which all constraints are satisfied.

Selecting α_k

If $a_i^T p_k \geqslant 0$ for some $i \notin \mathcal{W}_k$, then for all $\alpha_k \geqslant 0$ we have $a_i^T(x_k + \alpha_k p_k) \geqslant a_i^T x_k \geqslant b_i$. So, constraint i will be satisfied for all nonnegative choices of the step-length parameter. Whenever $a_i^T p_k < 0$ for some $i \notin \mathcal{W}_k$, we have $a_i^T(x_k + \alpha_k p_k) \geqslant b_i$ only if

$$\alpha_k \leqslant \frac{b_i - a_i^T x_k}{a_i^T p_k}$$

So, to maximize the decrement of q, α_k should be as large as possible in $[0, 1]$ subject to retaining feasibility, so we get the following equation:

$$\alpha_k = \min\left(1, \min_{i \notin \mathcal{W}_k, a_i^T p_k < 0} \frac{b_i - a_i^T x_k}{a_i^T p_k}\right) \tag{3.25}$$

The constraints i for which the minimum is achieved called blocking constraints.

If $\alpha_k = 1$ and no new constraints are active at $x_k + \alpha_k p_k$, then there is no blocking constraints on this iteration.

If $\alpha_k < 1$, step along p_k was blocked by some constraints not in $\mathcal{W}_k$, $\mathcal{W}_{k+1}$ is built by adding one of the blocking constraints to $\mathcal{W}_k$

This method is repeated untill a point $\hat{x}$ has been achieved that minimizes the quadratic objective function over its current working set $\hat{\mathcal{W}}$. Identifying this point is not hard because the subproblem as solution $p = 0$. Since $p = 0$ satisfies the optimality condition 3.4 for 3.23, it is found that:

$$\sum_{i \in \hat{\mathcal{W}}} a_i \hat{\lambda}_i = g = G\hat{x} + c \tag{3.26}$$

for some Lagrange multipliers $\hat{\lambda}_i, i \in \hat{\mathcal{W}}$. It follows that x^* and λ^* satisfy the first KKT condition, if the multipliers are defined corresponding to the inequality constraints that are not in the working set to be zero. As there are some control imposed on the step length, x^* is also feasible with respect to all the constraints, so the second and third KKT conditions are satisfied at this point.

Considering the signs of the multipliers corresponding to the inequality constraints in the working set, that is, the indices $i \in \hat{\mathcal{W}} \cap \mathcal{I}$. The fourth KKT condition is also satisfied if these multipliers are all nonnegative. So it can be concluded that, $\hat{x}$ is a KKT point for the original problem 3.1. In fact, since G is positive semidefinite, we have from Theorem *** that $\hat{x}$ is a global solution of 3.1.

If on the other hand, if there exists some $j \in \hat{\mathcal{W}} \cap \mathcal{I}$, such that

$$\lambda^* < 0$$

That constraints has to be removed from the active set and solve new subproblem. This will decreases the objective function. The following theorem states that this strategy produces a direction p at the next iteration that is feasible with respect to the removed constraint.

Theorem

Theorem 3.7.1. *Suppose that the point $\hat{x}$ satisfies first-order conditions for the equality-constrained subproblem with working set $\hat{W}$; that is, equation 3.25 is satisfied along with $a_i^T \hat{x} = b_i$ for all $i \in \hat{W}$. Suppose, too, that the constraint gradients $a_i, i \in \hat{W}$, are linearly independent and that there is an index $j \in \hat{W}$ such that $l\hat{ambda}_j < 0$.*
Let p be the solution obtained by dropping the constraint j and solving the following subproblem:

$$\begin{aligned} \min_p \quad & \frac{1}{2} p^T G p + (G\hat{x} + c)^T p, \\ \text{subject to} \quad & a_i^T p = 0 \qquad \forall i \in \hat{\mathcal{W}} \text{ with } i \neq j \end{aligned} \tag{3.27}$$

Then p is a feasible direction for constraint j,that is, $a_j^T p \geqslant 0$. Moreover,if p satisfies second- order sufficient conditions for above equation, then we have that $a_j^T > 0$, and that p is a descent direction for $q(.)$.

Whenever p_k obtained from 3.23 is nonzero and satisfies second-order sufficient optimality conditions for the current working set, it is a direction of strict descent for $q(.)$

Theorem

Theorem 3.7.2. *Suppose that the solution p_k of 3.23 is nonzero and satisfies the second order sufficient conditions for optimality for that problem. Then the function $q(.)$ is strictly decreasing along the direction p_k .*

So it can be concluded that, When G is positive definite-the second order sufficient conditions are satisfied for all feasible subproblems of the form 3.23. Hence, it follows from the result that a strict decrease in $q(.)$ can be obtained whenever $p_k \neq 0$.

Specification of the Active-set method for convex QP

The whole Active-set algorithm can be specified as following:

To implement active-set method efficiently an important key is reuse of nformation from solving the equality-constrained subproblem at the next iteration. The only difference between two consecutive subproblems is that the working set grows or shrinks by a single component. Efficient codes perform updates of the matrix factorizations obtained at the previous iteration, rather than calculating them from scratch each time.

3.8 Interior-Point Method

Interior-point methods follow iterative approach which is a good candidate as alternative of active-set method. This method is also know as trajectory-following, path-following method. Here only convex quadratic programming problem with inequality constraints will be focused. It is easier to the problem as follows:

$$\begin{aligned} \min_{x} \quad & q(x) = \frac{1}{2}x^T G x + x^T c \\ \text{subject to} \quad & Ax \geqslant b \end{aligned} \tag{3.28}$$

where $G \in \mathbb{R}^{n \times n}$ is symmetric, positive semidefinite, $A \in \mathbb{R}^{m\times n}$.

$$A = [a_i]_{i\in\mathcal{I}} \quad b = [b_i]_{i\in\mathcal{I}}, \quad \mathcal{I} = \{1, 2, 3..., m\}$$

KKT conditions for this notation can be written as follows:

$$\begin{aligned} Gx - A^T\lambda + c = 0, \\ Ax - b \geqslant 0, \\ (Ax - b)_i\lambda_i = 0, \quad i = 1, 2, ..., m \\ \lambda \geqslant 0. \end{aligned} \tag{3.29}$$

By introducing the slack vector $y \geqslant 0$, conditions can be rewritten as:

$$
\begin{aligned}
Gx - A^T\lambda + c = 0, \\
Ax - y - b = 0, \\
y_i\lambda_i = 0, \quad i = 1, 2, ..., m \\
(y, \lambda) \geqslant 0,
\end{aligned}
\tag{3.30}
$$

As G is positive semidefinite, above KKT conditions are necessary and sufficient to solve convex quadratic program.

Let, current iterate (x, y, z) that

satisfies $(y, \lambda) > 0$, a complementary measure μ can be defined as:

$$\mu = \frac{y^T \lambda}{m} \tag{3.31}$$

Derived path-following for the KKT conditions by considering the above KKT conditions:

$$F(x, y, \lambda : \sigma, \mu) = \begin{bmatrix} Gx - A^T\lambda + c \\ Ax - y - b \\ \mathcal{Y}\Lambda e - \sigma\mu e \end{bmatrix} = 0 \tag{3.32}$$

Where

$$\mathcal{Y} = diag(y_1, y_2, ..., y_m), \quad \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_1), \quad e = (1, 1, ..., 1)^T$$

and $\sigma \in [0, 1]$. The solution of 3.27 for all positive values of σ and μ define the central path, which is a trajectory that leads to the solution to the QP as $\sigma\mu$ tends to zero.

After

selecting μ and applying Newton's method to 3.27 the following linear system is achieved:

$$\begin{bmatrix} G & 0 & -A^T \\ A & -I & 0 \\ 0 & \Lambda & \mathcal{Y} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -r_d \\ -r_p \\ -\Lambda \mathcal{Y} e + \sigma \mu e \end{bmatrix} \tag{3.33}$$

where

$$r_d = Gx - A^T\lambda + c \quad r_p = Ax - y - b$$

The new iterate (x^+, y^+, λ^+) are obtained by means of

$$(x^+, y^+, \lambda^+) = (x, y, \lambda) + \alpha(\Delta x, \Delta y, \Delta \lambda)$$

α is chosen

such that $(x^+, \lambda^+) > 0$ and possibly to satisfy various other conditions.

Chapter 4

Cholesky Factorization

Cholesky factorization is an important term in the feld of linear algebra. It is the process of decomposing a Hermitian, positive-definite matrix into product of a lower and upper triangular matrix and its conjygate transpose. Which is useful for numerical optimization and Monte Carlo

sim-
u-
la-
tion.
It
was
named
af-
ter
the
name
the
name
of
its
dis-
cov-
erer
Andre-
Louis
Cholesky.
It
is
be-
lieved
that
cholesky
de-
com-
po-
si-
tion
is
al-
most
twice
as
ef-
fi-
cient
as
the
LU
de-
com-
po-
si-
tion.
for

solving system of linear equation.

4.1 Definition and Existance

The Cholesky factorization is only defined for symmetric or Hermitian positive definite matrices.

Definition 4.1.1. A matrix

$A \in \mathbb{R}^{m\times m}$ is symmetric positive definite (SPD) if and only if it is symmetric ($A^T = A$) and for all nonzero vectors $x \in \mathbb{R}^m$ it is the case that $x^T Ax > 0$.

Theorem

Theorem 4.1.1. *Given a SPD matrix A there exist a lower triangular matrix L such that $A = LL^T$.*

The lower triangular

matrix L is known as the Cholesky factor and LL^T is known as the Cholesky factorization of A. It is unique if the diagonal elements of L are restricted to be positive.

The converse holds

trivially: if A can be written as LL^* for some invertible L, lower triangular or otherwise, then A is Hermitian and positive definite.

4.2 LDL decomposition

Although the

focus of this chapter is Cholesky factorization, it is worth to make an understanding about a closely related variant of the Cholesky factorization, the LDL decomposition. It is represented as,

$A = LDL^*$

where L is a lower triangular matrix and D is a diagonal matrix.LDL decomposition relates to Cholesky decomposition by the following way

$$A = LDL^* = LD^{\frac{1}{2}} D^{\frac{1}{2}*} L^* = LD^{\frac{1}{2}} (LD^{\frac{1}{2}})^*$$

When LDL

is ef- fi- ciently im- ple- mented, it takes the same space and com- plex- ity to build and use. LDL method method is used for those cases where no Cholesky de- com- po- si- tion pos- si- ble.

4.3 Application of Choleskey Factorization

The Cholesky decomposition is mostly used when numerical solution of linear equations $Ax = b$ is required. When A is symmetric and positive definite, it is possible to solve $Ax = b$ by first computing the Cholesky decomposition $A = LL^*$, then solving $Ly = b$ for y by forward substitution, and finally solving $L* x = y$ for x

by back sub-sti-tu-tion.

The Cholesky de-com-po-si-tion helps to achieve su-pe-rior ef-fi-ciency and nu-mer-i-cal sta-bil-ity. Com-pared to the *LU* de-com-po-si-tion, this can per-form al-most two times efficiently.

Linear least squares

Systems of the form $Ax = b$ with A symmetric and positive definite arise quite often in applications. For instance, the normal equations in linear least squares problems are of this form. It may also happen that matrix A comes from an energy functional which must be positive from physical considerations; this happens frequently in the numerical solution of partial differential equations.

Nonlinear optimization

Nonlinear multivariate functions may be minimized over their parameters using variants of Newton's method called quasi-Newton methods. At each iteration, the search takes a step s defined by solving $Hs = -g$ for s, where s is the step, g is the gradient vector of the function's partial first derivatives with respect to the parameters, and H is an approximation to the Hessian matrix of partial second derivatives formed by repeated rank 1 updates at each iteration. Two well-known update formulae are called Davidon–Fletcher–Powell (DFP) and

Broyden–Fletcher–Goldfarb–Shanno (BFGS). Loss of the positive-definite condition through round-off error is avoided if rather than updating an approximation to the inverse of the Hessian, one updates the Cholesky decomposition of an approximation of the Hessian matrix itself.

Monte Carlo simulation

The Cholesky decomposition is commonly used in the Monte Carlo method for simulating systems with multiple correlated variables: The correlation matrix is decomposed, to give the lower-triangular L. Applying this to a vector of uncorrelated samples, u, produces a sample vector Lu with the covariance properties of the system being modeled.

For a simplified example that shows the economy one gets from Cholesky's decomposition, say one needs to generate two correlated normal variables x_1 and x_2. All one needs to do is to generate two uncorrelated Gaussian random variables z_1 and z_2. We set $x_1 = z_1$ and $x_2 = \rho z_1 + \sqrt{1-\rho^2} z_2$.

Matrix inversion

The explicit inverse of a

Hermitian matrix can be computed via Cholesky decomposition, in a manner similar to solving linear systems, using n^3 operations ($\frac{1}{2}n^3$ multiplications). The entire inversion can even be efficiently performed in-place.

A non-Hermitian matrix B can also be inverted using the following identity, where BB^* will always be Hermitian:

$$\mathbf{B}^{-1} = \mathbf{B}^*(\mathbf{B}\mathbf{B}^*)^{-1}.$$

4.4 The Cholesky algorithm

Most usual

al- go- rithm for Cholesky fac- tor- iza- tion $chol(A)$ can be de- rived as: (*Greek low- er- case let- ters refers to scalars, lower case let- ter ref- eres to vec- tor and up- per- case let- ters ref- eres to ma- tri- ces*).

Partition

$$A = \left(\begin{array}{c|c} \alpha_{11} & \star \\ \hline a_{21} & A_{22} \end{array} \right) \quad \text{and } L = \left(\begin{array}{c|c} \lambda_{11} & 0 \\ \hline \lambda_{21} & L_{22} \end{array} \right).$$

By substituting these partitioned matrices into $A = LL^T$, it is found that:

$$\left(\begin{array}{c|c} \alpha_{11} & \star \\ \hline a_{21} & A_{22} \end{array} \right) = \left(\begin{array}{c|c} \lambda_{11} & 0 \\ \hline \lambda_{21} & L_{22} \end{array} \right) \left(\begin{array}{c|c} \lambda_{11} & 0 \\ \hline \lambda_{21} & L_{22} \end{array} \right)^T = \left(\begin{array}{c|c} \lambda_{11}^2 & \star \\ \hline \lambda_{11} l_{21} & l_{21} l_{21}^T + L_{22} L_{22}^T \end{array} \right)$$

so that

$$\left(\begin{array}{c|c} \alpha_{11} = \lambda_{11}^2 & \star \\ \hline a_{21} = \lambda_{11} l_{21} & A_{22} = l_{21} l_{21}^T + L_{22} L_{22}^T \end{array} \right)$$

and hence

$$\left(\begin{array}{c|c} \lambda_{11} = \sqrt{\alpha_{11}} & \star \\ \hline l_{21} = a_{21}/\lambda_{11} & L_{22} = chol(A_{22} - l_{21} l_{21}^T) \end{array} \right)$$

These equalities directs to the

al-
go-
rithm

1. Partition $A \leftarrow \left(\begin{array}{c|c} \alpha_{11} & \star \\ \hline a_{21} & A_{22} \end{array}\right)$
2. Overwrite $\alpha_{11} := \lambda_{11} = \sqrt{\alpha_{11}}$
3. Overwrite $a_{21} := l_2 1 = a_{21}/\lambda_1 1$.
4. Overwrite $A_{22} := A_{22} - l_{21} l_{21}^T$ (updating lower triangular part of $A_2 2$)
5. Continue with $A = A_{22}$

Result: Write here the result
initialization **while** *While condition* **do**
instructions **if** *condition* **then**
instructions1 instructions2
else
instructions3
algorithm[How to write algorithms]How to write algorithms

Appendix A

Appendix Title Here

Write your Appendix content here.

Bibliography

Jesus A. De Loera Fu Liu, Ruriko Yoshida (2007). “A Generating Function for all Semi-Magic Squares and the Volume of the Birkhoff Polytope”. In: *Cornel University Library*, pp. 1–24. URL: http://arxiv.org/abs/math/0701866.

Paffenholz, Andreas (2013). “FACES OF BIRKHOFF POLYTOPES”. In: *Cornel University Library*, pp. 1–29. URL: http://arxiv.org/abs/1304.3948.

Pak, Igor (1999). “Four Questions on Birkhoff Polytope”. In: *Annals of combinatorics*, pp. 83–90. URL: http://www.math.ucla.edu/~pak/papers/bir.pdf.