POLITECNICO DI MILANO

Facoltà di Ingegneria
Scuola di Ingegneria Industriale e dell'Informazione
Dipartimento di Elettronica, Informazione e Bioingegneria

Master of Science in
Computer Science and Engineering
Ingegneria Informatica

Stochastic Variance Reduced Policy Gradient

Supervisor:

MARCELLO RESTELLI

Assistant Supervisor:

MATTEO PAPINI

MATTEO PIROTTA

Master Graduation Thesis by:

DAMIANO BINAGHI
Student Id n. 858458

CANONACO GIUSEPPE
Student Id n. 222222

Academic Year 2017-2018

Dediche

ACKNOWLEDGMENTS

Here you can put acknowledgements to people that helped you during the thesis. Remember that helping students to write thesis is part of the job of some of them, and they're also paid for that. Please make sure to thank them for what they weren't supposed to do.

Remember also that this page is part of your thesis. I know that your boyfriend/girlfriend is very important to you and you cannot live without her/him, as it is for me. But there's no need to put her/his name here unless she/he gave a proper contribution to this work. Same goes for friends, parents, drinking buddies and so on.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

EMODPS Evolutionary Multi-Objective Direct Policy Search

OS operating system

XML eXtensible Markup Language

ABSTRACT

An abstract is a brief of a research article, thesis, review, conference proceeding or any in-depth analysis of a particular subject or discipline, and is often used to help the reader quickly ascertain the paper's purpose. When used, an abstract always appears at the beginning of a manuscript or typescript, acting as the point-of-entry for any given academic paper or patent application. Abstracting and indexing services for various academic disciplines are aimed at compiling a body of literature for that particular subject.

Max 2200 characters, spaces included.

SOMMARIO

Per abstract si intende il sommario di un documento, senza l'aggiunta di interpretazioni e valutazioni. L'abstract si limita a riassumere, in un determinato numero di parole, gli aspetti fondamentali del documento esaminato. Solitamente ha forma "indicativo-schematica"; presenta cioé notizie sulla struttura del testo e sul percorso elaborativo dell'autore.

Max 2200 caratteri compresi gli spazi.

ESTRATTO

"... il testo delle tesi redatte in lingua straniera dovrà essere introdotto da un ampio estratto in lingua italiana, che andrà collocato dopo l'abstract."

PREFACE

A preface is an introduction to a book or other literary work written by the work's author. A preface generally covers the story of how the book came into being, or how the idea for the book was developed.

MOTIVATION

Graduating is not the motivation that one expects here.

1 AN INTRODUCTION TO THE WRITING OF SCIENTIFIC TEXTS

Science, my boy, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth.

— Verne *Journey to the Center of the Earth* 1957

1.1 GENERAL RULES

A scientific manuscript should be short, complete, clear, and logical. It should convey the messages in the easiest, but still correct way. Before writing, it may be useful to identify who the text is addressed to or the potential readers. The type of readers should then lead the writing style. For example, the following type of text are all different, both for what concern the contents and the way the manuscript is organized:

A BOOK it is usually written for students;

A SCIENTIFIC PAPER it is written for peers: from scientists to scientists;

A USER MANUAL it is written for learning how to use a software, for example.

When writing the thesis, keep in mind that the reader is probably a person who knows about the specific research field, but does not know what you did, so do not take things for granted.

It may be useful to develop the text in different stages. First, define the structure of the text (see also Section 1.2). Second, make a list of the main points in each section. Then, further develop the list by sketching a couple of sentences for each point. Finally, write the entire text. This may be helpful to focus on the main contents of the text, the messages you would like to convey, and, not less importantly, to share and discuss the organization of the manuscript with your supervisor.

In general, each paragraph should convey one message. Try to keep the sentences short, because this will ease the reading and understanding process. Avoid using phrases that are just a repetition, do not add any new message, or do not further contribute to understanding. Avoid using synonyms: in scientific writing they may confuse the readers. For example:

> We use a modeling framework composed of an optimization step followed by a simulation step. The modeling

~~procedure~~ allows to analyze the system evolution.
We use a modeling framework composed of an optimization step followed by a simulation step. The modeling framework allows to analyze the system evolution.

Use the active form if you are referring to what you did. For example:

~~The HBV model is chosen because it is a parsimonious model.~~
We chose the HBV model because it is a parsimonious model.

The passive form, instead, is preferred to indicate what is done in the literature. For example: "The HBV model is used because it is a parsimonious model."

Be careful in using pronouns such as "it" or "they" to refer to a concept in the previous sentence. It is fine, if the previous sentence convey just one concept, otherwise the pronoun may be confusing. In these cases, it is better to repeat the concept in the second sentence. For example:

We assigned the same parameters to the different land use classes in the catchment. ~~They~~ are taken from Wilson et al. (2014).
We assigned the same parameters to the different land use classes in the catchment. The land use classes are taken from Wilson et al. (2014).

Bullet lists are welcome. The items should be ranked following a rationale, for example they can be ranked by importance, preference, priority, etc.

1.2 THE STRUCTURE OF A SCIENTIFIC TEXT

The typical structure is described in the following. Nevertheless each thesis may have a different structure which can be customized to better highlight the work done. Before writing the entire thesis, you should identify a temptative structure and discuss it with your supervisor (see Section 1.1).

INTRODUCTION The Introduction should introduce the reader to the argument developed in the thesis. It should set the context and the reason why the topic is interesting. State the research questions addressed in the thesis and the objectives of the work. The novelties of the work should also be highlighted. You can also briefly mention the methods used.

LITERATURE REVIEW It contains the description of previous works and what has been already done in the literature. The text should be complemented with appropriate references (see Section 1.3).

MATERIALS AND METHODS It should describe the methods and techniques employed to address the research questions, the workflows, the hypothesis, the experimental setting, etc. If you are using a specific dataset, you should introduce it in this section. In principle, any reader should be able to reproduce the work and experiments you did by reading this section.

CASE STUDY If you are focusing on one or more case studies, this section should include the description of the study sites. Do not describe anything about the study site, but only include information that are relevant for your work, to understand what follows or why the application is relevant.

RESULTS This section should describe the results of the analysis. The text should be complemented with figures and tables, which are intended to support the evidences described in the text and to better comunicate the results (see Section 1.5).

DISCUSSION This section contains the interpretations of the results and state the conclusions you can draw from the analysis. You should also discuss the limitations of your approach and identify future research paths. This section may be merged with the "Results" section.

CONCLUSION You should make a short summary of the objectives of the work, the main results, and what you can conclude from your analysis. If you listed one or more research questions, you may want to formulate the reply which should be based, of course, on the evidences produced by your work. You can also briefly mention limitations and future works.

1.3 IN-TEXT CITATIONS

Citations usually goes just after the sentence you are quoting or paraphasing using brackets. The most commonly used type of citation is: author, year. When the reference has two or three authors of the source, cite all of them. If there are more than three authors, you can cite the first author followed by "et al." If the authors are subject of a sentence, only the year of publication is included in the brackets. For example:

> Air passing above a forested area is more humid than the one passing over a not forested area (Spracklen et al., 2012).

> Spracklen et al. (2012) demonstrated an increase of moisture in the air that passes above a forested area than a not forested one.

Do not copy and paste form other sources unless you are directly quoting from a work. For example:

> The approach is summarized by Philbrick and Kitanidis (1997): "Together modelling and optimization are sometimes called Systems Analysis."

1.4 ACRONYMS

Acronyms are welcome, but it is better not to over-use them. A text full of acronyms is difficult to read and may confuse the readers or force them to remember what they mean or to go back to the point where they are defined. For example:

> In the US, the notion of an NWO became popular after the terrorist attacks on the WTC. However, officials in NATO and the WTO rarely refer to an NWO in proceedings relating to the GATT, and it can be said that the MVTO, the MFN clause, and SROs have little to do with an NWO.[1].

Acronyms should be always defined the first time they are used (this may not apply to very popular acronyms, e.g.,, NASA, IPCC, etc., which may remain undefined). If you use the suggested LaTeX environment, `acronyms`, as explained in subsection 2.3.2 you can avoid to remember which is the first time you used the acronym. It does that for you. For example:

> In this thesis we adopt Evolutionary Multi-Objective Direct Policy Search (EMODPS). EMODPS is an approximate dynamic programming method that combines direct policy search and multi-objective evolutionary algorithms.

Be consistent: once you defined an acronym, readers will expect that you use the acronym instead of the full description. It is better to summarize all the acronyms in a dedicated table. This is done automatically if you use write every acronym in the dedicated environment `acronyms`. You find the list after the table of contents. The command transform the word into a link to the list for people for forget the meaning while reading.

1 Example taken from http://www.scribendi.com/advice/the_correct_use_of_acronyms.en.html

1.5 FIGURES AND TABLES

Figures and tables can be used to present a large amount of information to the reader and can be used to assist communication. Still, try to keep the minimum number of figures/tables. Figures and tables are not self-explanatory. Use the text to focus on the important point the reader should draw from them.

Tables/figures should be sequentially numbered as they appear in the text (in other words, Figure 2 can not be cited before Figure 1). They should appear close to the text where they are cited first. All the figures/tables included in the thesis must be cited somewhere in the text.[2]

Each figure/table should be understood without reference to the text. Captions should, then, contain a brief description of the figure's contents, for instance an explanation of the legend (see Figure 1.1).

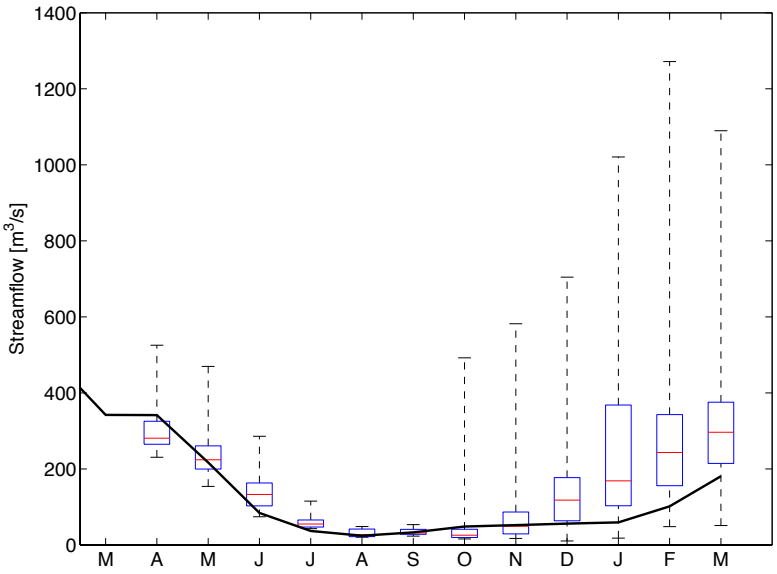

Figure 1.1: Streamflow forecast ensemble on April 1st 2000. The boxes show the 25%, 50%, and 75% percentiles, while the whiskers represent the minimum and maximum ensemble members, thus indicating the full range of uncertainty. The solid black line represents the observations.

Figures should always be good quality figures. Axis tag, labels, and legend should have an appropriate font size. Keep always the original figure (for example, `.fig` if you created the figure with matlab): this will allow you or who is working with/after you to quickly modify it or recover the original data, if needed. When applicable, keep the

2 LaTeX is of great help here: for instance, you should not worry about number ordering. Specific code for figures as well as for tables is given in subsection 2.3.2.

same axis ranges to the figures: it will ease the comparison between different figures. When applicable/possible, use subfigures to ease the comparison between figures.

Finally, if you are using figures, tables and/or data from other sources, for example from other publications, cite always the reference. For example:
"[...], taken/modified from Jones et al. (1950)"

1.6 MATHEMATICS

Inline vs full line vs numbered equation. Punctuation within equations.

Contemporary writers label the mathematical results in their papers as lemmas, theorems, propositions and so on. Typically the major results of the paper are called theorems, the lesser results are called propositions. The propositions are typically ingredients in the proofs of the theorems that are stand-alone statements of may be of independent interest), and the small technical results are called lemmas. This probably varies quite a bit from writer to writer (and perhaps also from field to field).

1.7 MARGIN NOTES

Margin notes are used to study a new text or paper. To read carefully a given text, it is essential that you write "all over" what you read, usually in the margins or between the lines?wherever there is space. By annotating, you'll be creating a shorthand version of what you're reading to which you can return later for reference. It's always much easier to navigate something you read a few days ago if you have taken detailed marginal notes. We recommend to use margin notes when reviewing you own thesis. Marginal notes help you to conceptualize the piece as you read.

Here are some general tips for annotating non-fiction articles. Right next to each paragraph, write a brief phrase summarizing what happens in it. Underline passages that seem crucial to the point of the paragraph or to the larger thesis of the piece. Note sections you don't fully understand with a question, like "What does this mean?" or "Doesn't this contradict my earlier argument?". Outline points being made and the examples given to support them. If you lists a number of reasons or factors that cause something or says something can be divided up into a certain number of parts, list the parts or factors in the margin.

1.8 SUBMISSION TO UNIVERSITY: FOLLOW THE INSTRUCTIONS

Visit this link for the updated information about the content of the thesis.

"Alcune Scuole forniscono linee guida specifiche cui i laureandi devono attenersi per la redazione della tesi. Per ulteriori informazioni: www.tedoc.polimi.it/..."

1.8.1 *Archiving electronic documents: PDF/A*

PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for the digital preservation of electronic documents. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding). The ISO requirements for PDF/A file viewers include color management guidelines, support for embedded fonts, and a user interface for reading embedded annotations.

Universities usually requires this standard but they're also not aware that common programs like MS Word, OpenOffice and so on aren't really able to produce compliant PDFs. In Latex, there's some development going on but at the time of writing, the available commands are still too obscure and buggy. So in the end, forget the PDF/A for now.[3]

3 Or DIY and then make a pull request on github :D.

2 STATE OF THE ART

In this section, we provide the essential background on policy gradient methods and stochastic variance-reduced gradient methods for finite-sum optimization.

2.1 POLICY GRADIENT

A Reinforcement Learning task (**sutton1998reinforcement**) can be modelled with a discrete-time continuous Markov Decision Process (MDP) $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \rho\}$, where $\mathcal{S}$ is a continuous state space; $\mathcal{A}$ is a continuous action space; $\mathcal{P}$ is a Markovian transition model, where $\mathcal{P}(s'|s, a)$ defines the transition density from state s to s' under action a; $\mathcal{R}$ is the reward function, where $\mathcal{R}(s, a) \in [-R, R]$ is the expected reward for state-action pair (s, a); $\gamma \in [0, 1)$ is the discount factor; and ρ is the initial state distribution. The agent's behaviour is modelled as a policy π, where $\pi(\cdot|s)$ is the density distribution over $\mathcal{A}$ in state s. We consider episodic MDPs with effective horizon H.[1] In this setting, we can limit our attention to trajectories of length H. A trajectory τ is a sequence of states and actions $(s_0, a_0, s_1, a_1, \ldots, s_{H-1}, a_{H-1})$ observed by following a stationary policy, where $s_0 \sim \rho$. We denote with $p(\tau|\pi)$ the density distribution induced by policy π on the set $\mathcal{T}$ of all possible trajectories (see Appendix **??** for the definition), and with $\mathcal{R}(\tau)$ the total discounted reward provided by trajectory τ: $\mathcal{R}(\tau) = \sum_{t=0}^{H-1} \gamma^t \mathcal{R}(s_t, a_t)$. Policies can be ranked based on their expected total reward: $J(\pi) = \mathbb{E}_{\tau \sim p(\cdot|\pi)} [\mathcal{R}(\tau)|M]$. Solving an MDP M means finding $\pi^* \in \arg\max_\pi \{J(\pi)\}$.

Policy gradient methods restrict the search for the best performing policy over a class of parametrized policies $\Pi_\theta = \{\pi_\theta : \theta \in \mathbb{R}^d\}$, with the only constraint that π_θ is differentiable w.r.t. θ. For sake of brevity, we will denote the performance of a parametric policy with $J(\theta)$ and the probability of a trajectory τ with $p(\tau|\theta)$ (in some occasions, $p(\tau|\theta)$ will be replaced by $p_\theta(\tau)$ for the sake of readability). The search for a locally optimal policy is performed through gradient ascent, where the policy gradient is **sutton2000policy**; **Peters2008reinf**:

$$\nabla J(\theta) = \mathop{\mathbb{E}}_{\tau \sim p(\cdot|\theta)} [\nabla \log p_\theta(\tau) \mathcal{R}(\tau)] . \tag{2.1}$$

Notice that the distribution defining the gradient is induced by the current policy. This aspect introduces a nonstationarity in the sampling process. Since the underlying distribution changes over time, it

1 The episode duration is a random variable, but the optimal policy can reach the target state (i.e., absorbing state) in at most H steps. This has not to be confused with a finite horizon problem where the optimal policy is non-stationary.

is necessary to resample at each update or use weighting techniques such as importance sampling. Here, we consider the *online learning scenario*, where trajectories are sampled by interacting with the environment at each policy change. In this setting, stochastic gradient ascent is typically employed. At each iteration $k > 0$, a batch $\mathcal{D}_N^k = \{\tau_i\}_{i=0}^N$ of $N > 0$ trajectories is collected using policy π_{θ_k}. The policy is then updated as $\theta_{k+1} = \theta_k + \alpha \widehat{\nabla}_N J(\theta_k)$, where α is a step size and $\widehat{\nabla}_N J(\theta)$ is an estimate of Eq. (2.1) using $\mathcal{D}_N^k$. The most common policy gradient estimators (e.g., REINFORCE (**williams1992simple**) and G(PO)MDP (**baxter2001infinite**)) can be expressed as follows

$$\widehat{\nabla}_N J(\theta) = \frac{1}{N} \sum_{n=1}^{N} g(\tau_i|\theta), \quad \tau_i \in \mathcal{D}_N^k, \tag{2.2}$$

where $g(\tau_i|\theta)$ is an estimate of $\nabla \log p_\theta(\tau_i) \mathcal{R}(\tau_i)$. Although the REINFORCE definition is simpler than the G(PO)MDP one, the latter is usually preferred due to its lower variance. We refer the reader to Appendix **??** for details and a formal definition of g.

The main limitation of plain policy gradient is the high variance of these estimators. The naïve approach of increasing the batch size is not an option in RL due to the high cost of collecting samples, i.e., by interacting with the environment. For this reason, literature has focused on the introduction of baselines (i.e., functions $b : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$) aiming to reduce the variance (**williams1992simple**; **Peters2008reinf**; **Thomas2017actionbaseline**; **wu2018variance**), see Appendix **??** for a formal definition of b. These baselines are usually designed to minimize the variance of the gradient estimate, but even them need to be estimated from data, partially reducing their effectiveness. On the other hand, there has been a surge of recent interest in variance reduction techniques for gradient optimization in supervised learning (SL). Although these techniques have been mainly derived for finite-sum problems, we will show in Section **??** how they can be used in RL. In particular, we will show that the proposed SVRPG algorithm can take the best of both worlds (i.e., SL and RL) since it can be plugged into a policy gradient estimate using baselines. The next section has the aim to describe variance reduction techniques for finite-sum problems. In particular, we will present the SVRG algorithm that is at the core of this work.

2.2 INSTALL LaTeX

If you don't have already a LaTeX system installed, this section will explain everything you need. The easiest way to get LaTeX is to install TeXLive, which works on all operating systems (OSs). In `https://www.tug.org/texlive/` you find the instructions and the files needed - and also get in touch with minimalism of TeXusers.

Then you will need an editor: I strongly recommend TeXworks because it's very simple and available on all the platforms. Also you don't need to install it, it's already included in TeXLive. The official documentation of TeXworks is available here;[2] I strongly recommend the reading of chapter 3. Alternatevely you can read an italian manual: profs.sci.univr.it/ gregorio/introtexworks.pdf (just 13 pages, read it!).[3]

After opening TeXworks, I strongly suggest to set these two additional things:

- open Preferences, then go the Composition tab: in the second box there, the "Process instruments", push the plus button. In the window just opened, write `Biber` in the "Name" field, `biber` in the "Program" field (lowercase!) and then press the plus button to add the argument `$basename`;
- again in the same window, set "Hide console output" to "never".

Then just test the installation of the template:

A. go into the template home folder;
B. open the file `ClassicThesis_DEIB.tex`;
C. select `pdfLaTeX` from the dropdown menu in the top right of the TeXworks window;
D. press the rounded green button: it compiles the `.tex` file for the first time and open the resulting `.pdf`;
E. select `Biber` from the same dropdown menu and press again the green button: this compiles the bibliography, a thing you need to repeat only when you change the file `Bibliography.bib`;
F. select `pdfLaTeX` again and recompile: this is needed to build indices and crossreferences;

The above compilation procedure is the standard way to translate the LaTeX code into pdfs.

2.2.1 *Online editor*

If the above procedure seems too difficult to you and you have an internet connection always available, you might think to use an online editor. The best choice at the time of writing is `http:\\sharelatex.com` where you can even find this template after registration to the site by looking for "Classic Thesis At DEIB". Your project will be saved

2 https://docs.google.com/file/d/0B5iVT8Q7W44pMk1WSFRKcDRlMU0/preview
3 If you already have a preferred editor, just keep using yours.

on their server but you can also download them. The platform allows up to two authors for free accounts.

There is no need to provide instructions for its use since the website has them. They also have an online LaTeXguide.

2.3 USE LaTeX WITHIN THIS TEMPLATE

2.3.1 *File structure*

The template is organized in multiple file and folders:

A. `ClassicThesis_DEIB.tex` is the main file to be compiled, found in the root folder. You should just add the source filenames you want to include and any hyphenation you need to explictly specify.

B. `classicthesis-config.tex` contains options that can be chosen for this template, like the `draft` one that prints date and time at the bottom of every page. It contains also the definition for the title, the author and others stuff displayed in the titlepage. Comments within the file should guide you.[4] Take a look at it!

C. `Bibliography.bib` is the *Bibtex* database: it is a normal textfile where you should put books and articles read;

D. `Chapters` contains the files for the main chapters of your thesis; this is where you will add the chapters text, as well these very words in line 41 of the file `Conclusion.tex`;

E. `CodeFiles` contains any code snippet you want to include in your thesis with the environment `listings`; it might be some relevant Matlab or C code, as well as long bash scripts;

F. `FrontBackmatter` contains various files that are included in the main one to produce abstract, titlepages, acknowledgements, ... Follow the instructions below to modify them in order to suits your needs;

G. `Images` contains the `.pdf` or `.png` versions of the images of the thesis organized in subfolder per chapter.

To modify abstract, preface, acknowledgements snd acronyms, you need to go into the folder `FrontBackmatter` where you will find the following:

ABSTRACT.TEX contains the text displayed as "abstract" and "sommario" just after the list of figures, tables, etc. Modify the text and leave the rest.

4 comments are the rows starting with %.

ACKNOWLEDGMENTS.TEX contains the text put just before the table of contents. Modify the text to suit your needs.

ACRONYMS.TEX contains the environment `acronym` with the definition of all the acronyms that will be used within the text. Add your own to the list and put the longest as parameter of the environment.

AUTOPARTS folder contains things that should work without your intervention. Forget them.

DEDICATION.TEX same usage and structure as `Acknowledgements.tex`.

ESTRATTO.TEX Politecnico di Milano requires an italian long excerpt of theses written in foreign languages.

FRONTESPIZIO.TEX and `FrontespizioIT.tex` are the cover page in english and italian, respectively. Politecnico di Milano requires the italian version of the english cover, so there it is. Both should work perfectly if you modify section 2 of the file `classicthesis-config.tex`, but you may not like the style so modify them as you prefer.

PREFACE.TEX same usage and structure as `Acknowledgements.tex`.

PUBLICATION.TEX same usage and structure as `Acknowledgements.tex`, but not included by default. Activate it by uncommenting the relevant line in `ClassicThesis_DEIB.tex`.

RETROFRONTESPIZIO.TEX contains the colophon. In most cases is fine as it already is.

2.3.2 *Special environments*

In addition to common LaTeX environments, this thesis is set to use:

Use these environments: they make the thesis less bland and readable.

- the command `\graffito{}` is used to create margin notes. The limits in number of words or length of word must be seen as a motivation to keep the notes short and simple;
- `\begin{aenumerate}` to produce an `\enumerate` with letters instead of numbers, as in the file list above;
- footnotes are useful to provide extra information. Usually they are not required to understand a paragraph but provide interesting details. This keeps the main body of text concise. You can create them with `\footnote{text}`.[5]

5 They should be placed after the punctuation mark and preferably at the end of the paragraph. In fact, they should not interrupt the reading flow. If you need to put a footnote in the middle of a paragraph, or of a sentence then the note should be part of the main text.

- `\ac{}` and its variations, defined by package `acronyms`, provide nice handling for acronyms, like eXtensible Markup Language (XML), produced with the code `\ac{XML}`. List them within the environment `acronym` in the file `FrontBackmatter/Acronyms.tex`.

2.3.3 *Citing, quoting and referencing*

References to bibliography are produced in the usual way with `\cite{bib_key}` (Bringhurst, 2008); don't forget the brackets which have to be added by hand. There also variations of the command, like `\citeauthor{bib_key}`, `\citetitle{bib_key}` and others that you can find in the `bibtex` manual.

`\blockcquote[][]{}{}` "produce a quotation with reference to author and page" see Bringhurst, 2008, p. 111. If the quotation is longer than two rows is indented. This behavior is provided by the package `csquotes`, which settings are in `classicthesis-config.tex`. The package also provides `\enquote{the citation}` that produces "correct quotation style" according to the language in use.

There is a set of commands to refer to chapters, sections, subsections, appendices, figures, tables and equations, like `\myChap{label_key}` to produce chapter 1. There are also capital versions of the commands (`\MyChap{}` produces Chapter 1). They need a `\label{name}` anchor next to the referred thing.

- `\myChap` for chapters;
- `\mySec` for sections;
- `\mySubsec` for subsections;
- `\myAppendix` for appendices;
- `\myFig` for figures;
- `\myTab` for tables;
- `\myEq` for equations;

2.3.4 *Figures and tables*

Figures are handled usually with the code

```
\begin{figure}
\centering
\includegraphics[width=\columnwidth]{Images/your_image_name.pdf}
\caption[Short description]{Long description.}
\label{fig:a_name}
\end{figure}
```

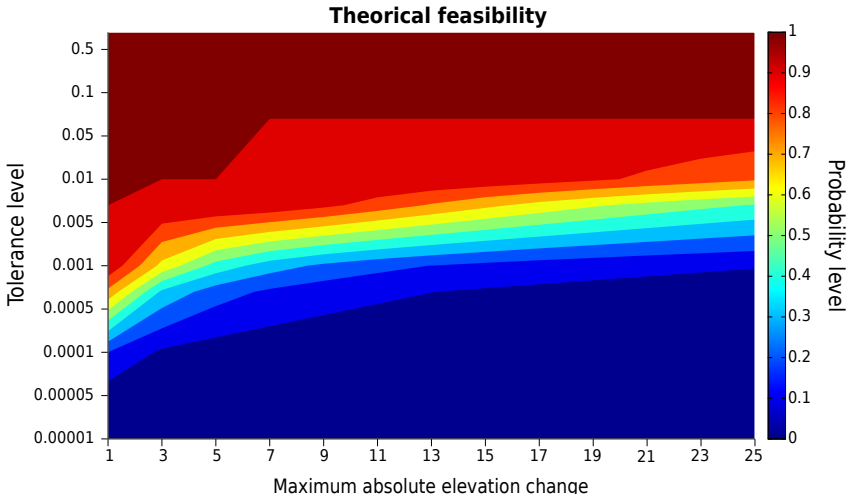

Figure 2.1: Thing taken from our master thesis whose meaning have been completely forgotten.

ALGORITHM	PARAMETER	SUGGESTED VALUES
Any	NFE	10 000 ÷ 200 000
	Population Size	10 ÷ 1000
GDE3	DE step size	0.0 ÷ 1.0
	Crossover rate	0.0 ÷ 1.0

Table 2.1: Parameters needed for things that are not needed anymore themselves.

which produces things like figure 2.1. Take care of the short description: it appears in the list of figures and should be just a reference, not a exhaustive description. Of course, you need to put the image file `your_image_name.pdf` in folder `Images/`. We suggest to keep things organized: create a folder for each chapter and keep the original source/working file in the same place.

Tables are produced with the code

```
\begin{table}[tb]
\footnotesize
\centering
\begin{tabularx}{0.8\textwidth}{llrcl}
\toprule
\tableheadline{l}{Algorithm} &
\tableheadline{l}{Parameter} &
\tableheadlineMore{3}{c}{Suggested Values} \\
\midrule
\tablefirstcol{l}{Any}
& \acs{NFE}  & $10\,000 $ & $ \div $ & $ 200\,000$ \\
& Population Size  &  $10 $ & $ \div  $ & $ 1000$ \\
```

```
\midrule
\tablefirstcol{l}{\ac{GDE3}}
& \ac{DE} step size & $0.0 $ & $\div $ & $ 1.0$ \\
& Crossover rate  & $0.0$ & $ \div $ & $ 1.0$ \\
\bottomrule
\end{tabularx}
\caption[Short description]{Long description.}
\label{tab:MOEAandParameters}
\end{table}
```

which produces table 2.1. `\myfloatalign`, `\tableheadline{}{}` and its variation `\tableheadlineMore{}{}{}` and `\tablefirstcol{}{}` are used to give a common style to all tables in the document. Use them! They are defined in `classicthesis-config.tex`.

2.3.5 *Math*

You can produce an equation like $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k^2} = \frac{\pi^2}{6}$ by embedding this code in the line:

```
$\lim_{n \to \infty}\sum_{k=1}^n \frac{1}{k^2}= \frac{\pi^2}{6}$.
```

Equation that spans the full line like:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k^2} = \frac{\pi^2}{6}$$

are produced with something like this:

```
\[
\lim_{n \to \infty}\sum_{k=1}^n \frac{1}{k^2}= \frac{\pi^2}{6}.
\]
```

If you need to refer to the equation later on, you need to number and label it. It is done via

```
\begin{equation}
\label{eq:euler}
e^{i\pi}+1=0.
\end{equation}
```

$$e^{i\pi} + 1 = 0. \tag{2.3}$$

From equation (2.3) you can see how `\myEq{eq:euler}` should be used.

Numeric sets requires specific font as $\forall x \in \mathbb{R}$ which is produced with `$\forall x \in \mathbb{R}$`. Matrices like

$$A = \begin{bmatrix} x_{11} & x_{12} & \dots \\ x_{21} & x_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

requires

```
A=
\begin{bmatrix}
x_{11} & x_{12} & \dots \\
x_{21} & x_{22} & \dots \\
\vdots & \vdots & \ddots
\end{bmatrix}.
```

Multiline equation can be produced with different environments like `split` and `cases`.

$$\begin{aligned} a &= b+c-d \\ &= e-f \\ &= g+h \\ &= i. \end{aligned}$$

comes from

```
\begin{split}
a &= b+c-d \\
  &= e-f \\
&= g+h \\
&= i.
\end{split}.
```

$$f(n) := \begin{cases} 2n+1, & \text{con } n \text{ dispari,} \\ n/2, & \text{con } n \text{ pari.} \end{cases}$$

comes from

```
\begin{split}
a &= b+c-d \\
  &= e-f \\
&= g+h \\
&= i.
\end{split}.
```

Definition like

Definition 2.1 (Gauss)**.** The math guy find obvious that $\int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}$.

are produced with the code

```
\begin{definition}[Gauss]
The math guy find obvious that
$\int_{-\infty}^{+\infty}
e^{-x^2}\,dx=\sqrt{\pi}$.
\end{definition}
```

There also a number of other similar environments, like `observation`, `theorem` with or without name, `corollary` and `lemma`.

Observation 2.2. *But many people like me don't find it obvious.*

Theorem 2.1. Mathematicians are very rare, if any.

Theorem 2.2 (Pythagorean)**.** The square of the hypotenuse of a triangle is equal to the sum of the squares of the other two sides.

Demonstration is left for exercise.

Corollary 2.3. *A line segment whose length is incommensurable so the ratio of which is not a rational number, can be constructed using a straightedge and compass.*

Lemma 2.4. *Pythagoras's theorem enables construction of incommensurable lengths because the hypotenuse of a triangle is related to the sides by the square root operation.*

You can also proof your theorem with the environment `proof`.

Theorem 2.5 (Surprise)**.** We have $\log(-1)^2 = 2\log(-1)$.

Proof. We have $\log(1)^2 = 2\log(1)$. But also we have $\log(-1)^2 = \log(1) = 0$. So $2\log(-1) = 0$. ■

There's also the cute little square at the end.

2.4 CONTRIBUTING TO THIS TEMPLATE

Suggestion and improvements are welcome at `https://github.com/Lordmzn/ClassicThesis-at-DEIB` or via email at `emanuele.mason@polimi.it`, `andrea.cominola@polimi.it` or `daniela.anghileri@polimi.it`.

BIBLIOGRAPHY

Bringhurst, R. (2008). *The Elements of Typographic Style*. Version 3.2. Point Roberts, WA, USA: Hartley & Marks Publishers (cit. on p. 14).

Knuth, D. E. (1974). "Computer Programming as an Art." In: *Communications of the ACM* 17.12, pp. 667–673 (cit. on p. 21).

Verne, J. (1957). *Journey to the Center of the Earth*. Classics illustrated. Huge Print Press. ISBN: 9780758311993 (cit. on p. 1).

APPENDIX EXAMPLE: CODE LISTINGS

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— Knuth, "Computer Programming as an Art," 1974

A.1 THE `listings` PACKAGE TO INCLUDE SOURCE CODE

Source code is usually not part of the text of a thesis, but if it is an original contribution it makes sense to le the code speak by itself instead of describing it. The package `listings` provide the proper layout tools. Refer to its manual if you need to use it, an example is given in listing A.1.

Listing A.1: Code snippet with the recursive function to evaluate the pdf of the sum Z_N of N random variables equal to X.

```
std::vector<int> values_of_x(number_of_values_of_x,
  min_value_of_x);
for (unsigned int i = 1; i < number_of_values_of_x; i++) {
  values_of_x[i] = values_of_x[i - 1] + 1;
}
prob_x = 1.0 / number_of_values_of_x;
std::vector<std::vector<double> > p_z;
for (unsigned int idx = 0; idx < p_z.size(); idx++) {
  p_z[idx] = std::vector<double>(
    (max_value_of_x * (idx + 1) - min_value_of_x
      * (idx + 1)) + 1, INIT_VALUE);
}

double prob(int Z, int value_of_z) {
  if (value_of_z < min_value_of_x * Z ||
    value_of_z > max_value_of_x * Z) {
      return 0.0;
  }
  if (value_of_z < min_value_of_z ||
    value_of_z > max_value_of_z) {
      return 0.0;
  }
  int idx_value_of_z = -(min_value_of_z - value_of_z);
  int idx_N = Z - 1;
  if (p_z[idx_N][idx_value_of_z] == -2.0) {
    if (Z > 1) {
      double pp = 0.0;
      for (unsigned int i = 0; i < number_of_values_of_x; i++) {
        pp += prob(Z - 1, value_of_z - values_of_x[i], p);
      }
      p_z[idx_N][idx_value_of_z] = prob_x * pp;
    } else {
      if (Z == 1) {
        for (unsigned int j = 0; j < number_of_values_of_x; j++)
            {
          if (value_of_z == values_of_x[j]) {
            p_z[idx_N][idx_value_of_z] = prob_x;
            break;
          }
        }
      }
      if (p_z[idx_N][idx_value_of_z] == INIT_VALUE) {
        p_z[idx_N][idx_value_of_z] = 0.0;
      }
    }
  }
  return p_z[idx_N][idx_value_of_z];
}
```