Multiple Constraints and Non-regular Solution in Deep Declarative Network

Suikei Wang

A thesis submitted for the degree of
Master of Machine Learning and Computer Vision
The Australian National University

September 2020

Except where otherwise indicated, this thesis is my own original work.

Suikei Wang
19 September 2020

to my parents, yyy (yyy is the people you want to dedicated this thesis to.)

Acknowledgments

The past two years at the Australian National University have been an invaluable experience for me. When I started my Master of Machine Learning and Computer Vision at the beginning of 2019, I could barely understand lectures, knew little about the country, and had never heard of the term "convex optimization". It is unbelievable that I have been doing a research project on this topic for a whole year. ANU has the top tier research group in this realm and how honored I am to be a postgraduate student here.

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Stephen Gould. I knew Stephen before the admission of the program and how privileged I am to work and learn with one of the most brilliant minds in our field. He always has an insightful and high-level view on this topic. More importantly, Stephen is extremely kind and patient as he is always willing to discuss and share ideas with us during the weekly meetings. Although sometimes I am stuck in some problems he used to give me constructive suggestions.

I would like to thank Dylan Campbell and Miaomiao Liu – another two giants of the Australian Centre for Robotic Vision – for offering my supervision and being my second examiner on my thesis. Dylan

Abstract

Put your abstract here.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Deep learning models composed with multiple parametrized processing layers can learn different levels of features and representations of data through the directed graph structure.

Put your introduction here. You could use \fix{ABCDEFG.} to leave your comments, see the box at the left side.

You have to rewrite your thesis!!!

1.2 Thesis Outline

How many chapters you have? You may have Chapter 2, Chapter **??**, Chapter **??**, Chapter 6, and Chapter 7.

1.3 Contribution

Part I

Deep Declarative Network: Multiple Constrained Declarative Nodes

Chapter 2

An Overview of Numerical Optimization

In this chapter, we aim to provide readers with an overview of numerical optimization. We begin with the theory of optimization (Section 2.1), from the existence of optimizers, to the optimality conditions for both unconstrained and constrained problems with duality. As the theoretical background of optimization, this field provides a solid solution for the algorithm.

We then formally define the optimization of unconstrained and constrained problems in Section 2.2 and describe the general regular solution for these problems based on the gradient calculation.

Next, we discuss briefly the bi-level optimization, which is a parameterized lower-level problem binds variables that appear in the objective of an upper-level problem (Section 2.3). Finally, we give a summary of the numerical optimization in constrained problems in Section 2.4.

2.1 Theory of Optimization

2.1.1 Existence of Optimizers

In optimization, a basic question is to determine the existence of a global minimizer for a given function f. There are several sufficient conditions on f to guarantee the existence, and the optimizer falls in the feasible set of solutions. For such feasible set, some definitions are following:

Definition 2.1. A subset $\Omega \in \mathbb{R}^n$ is called

- *bounded* if there is a constant $R > 0$ such that $\|x\| \leq R$ for all $x \in \Omega$
- *closed* if the limit point of any convergent sequence in Ω always lies in Ω
- *compact* if any sequence $\{x_k\}$ in Ω contains a subsequence that converges to a point in Ω

The following result gives a characterization of compact sets in $\mathbb{R}$.

Lemma 2.2 (Bolzano-Weierstrass theorem)**.** A subset Ω in $\mathbb{R}^n$ is *compact* if and only if it is bounded and closed.

2.1.2 Optimality Conditions for Unconstrained Problems

2.1.3 Optimality Conditions for Constrained Problems

2.1.4 Duality

2.2 Unconstrained and Constrained Optimization

2.2.1 Unconstrained Optimization

2.2.2 Equality Constrained Optimization

2.2.3 Inequality Constrained Optimization

2.3 Bi-level Optimization

2.4 Summary

Summary what you discussed in this chapter, and mention the story in next chapter. Readers should roughly understand what your thesis takes about by only reading words at the beginning and the end (Summary) of each chapter.

Chapter 3

Deep Declarative Network

Same as the last chapter, introduce the motivation and the high-level picture to readers, and introduce the sections in this chapter.

3.1 Notation

3.2 An Overview of Deep Declarative Network

3.2.1 Structure

3.2.2 Declarative Nodes

3.3 Learning

3.4 Back-propagation Through Declarative Nodes

3.4.1 Unconstrained

3.4.2 Equality Constrained

3.4.3 Inequatlity Constrained

3.5 Experiments

3.5.1 Platform

3.5.2 Implementation Details

3.5.3 Experimental Results

3.5.4 Analysis: Is the Gradient Always Converged?

3.6 Summary

Same as the last chapter, summary what you discussed in this chapter and be the bridge to next chapter. s

Chapter 4

The Future of Declarative Nodes

Same as the last chapter, introduce the motivation and the high-level picture to readers, and introduce the sections in this chapter.

4.1 Summary

Same as the last chapter, summary what you discussed in this chapter and be the bridge to the next chapter.

Part II

Deep Declarative Network: Non-regular Solution

Chapter 5

Solution of Deep Declarative Nodes

5.1 Software platform

5.2 Hardware platform

Table 5.1 shows how to include tables and Figure 5.1 shows how to include codes.

Architecture	Pentium 4	Atom D510	i7-2600
Model	P4D 820	Atom D510	Core i7-2600
Technology	90nm	45nm	32nm
Clock	2.8GHz	1.66GHz	3.4GHz
Cores × SMT	2 × 2	2 × 2	4 × 2
L2 Cache	1MB × 2	512KB × 2	256KB × 4
L3 Cache	none	none	8MB
Memory	1GB DDR2-400	2GB DDR2-800	4GB DDR3-1066

Table 5.1: Processors used in our evaluation.

```
int main(void)
{
  printf("Hello␣World\n");
  return 0;
}
```

(a)

```
void main(String[] args)
{
  System.out.println("Hello␣World");
}
```

(b)

Figure 5.1: Hello world in Java and C.

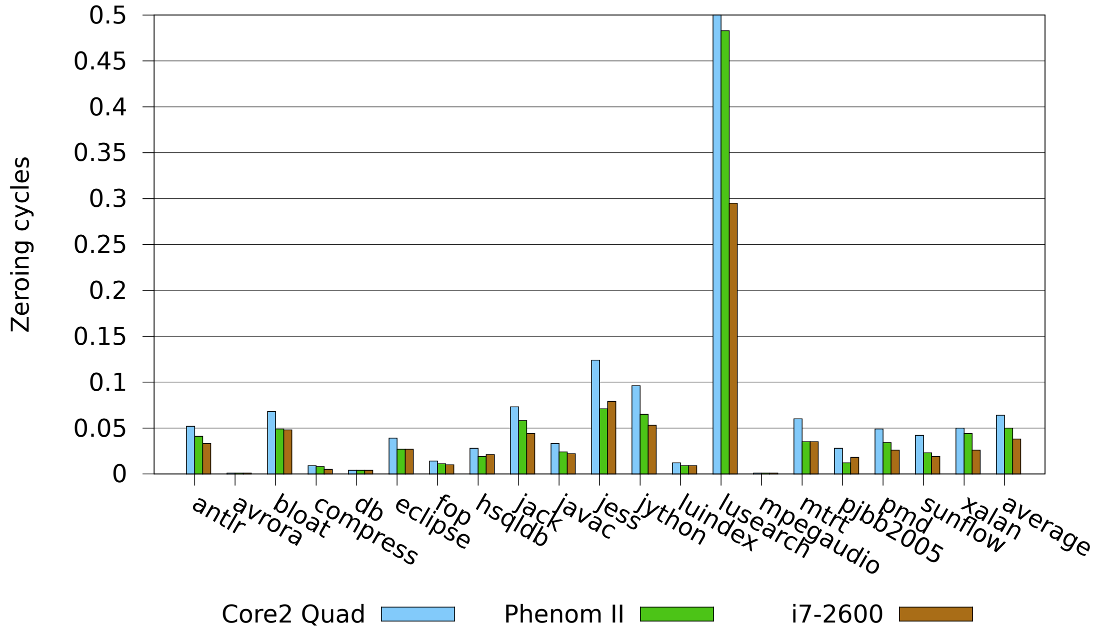
Chapter 6

Results

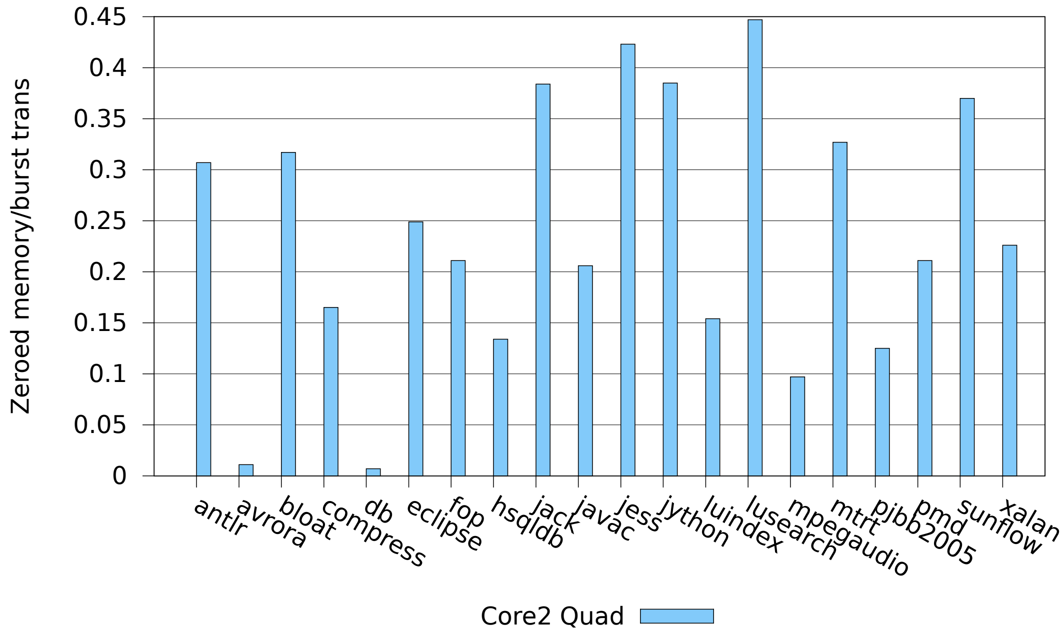
6.1 Direct Cost

Here is the example to show how to include a figure. Figure 6.1 includes two subfigures (Figure 6.1(a), and Figure 6.1(b));

6.2 Summary



(a) Fraction of cycles spent on zeroing



(b) BytesZeroed / BytesBurstTransactionsTransferred

Figure 6.1: The cost of zero initialization

Chapter 7

Conclusion

Summary your thesis and discuss what you are going to do in the future in Section 7.1.

7.1 Future Work

Good luck.

Bibliography

Appendix