An L-Smooth Analysis of the Annealed Interval-Constrained Quantization-Aware Training Optimization Method `ASkewSGD`

Hok Fong Wong
Department of Computer Science and
Engineering, The Chinese University
of Hong Kong

Hoi-To Wai
Department of Systems Engineering and
Engineering Management, The Chinese
University of Hong Kong

Abstract

In this paper, we provide new insights for the algorithm `ASkewSGD` proposed by Leconte et al. (2023) for quantization-aware training in deep neural networks.

1 Introduction

Over the past decade, there is an increasing awareness that deep learning models have succeeded to achieve vast improvements for tasks such as computer vision, speech recognition and generation, and natural language processssing. While the accuracy of these models has significantly improved, the size of these models could be a potential problem when being deployed in resource-limited scenarios.

2 Related Works

3 Main Results

The camera-ready versions of the accepted papers are 9 pages, plus any additional pages needed for references and textual supplementary material.

Papers are in 2 columns with the overall line width of 6.75 inches (41 picas). Each column is 3.25 inches wide (19.5 picas). The space between the columns is .25 inches wide (1.5 picas). The left margin is 0.88 inches (5.28 picas). Use 10 point type with a vertical spacing of 11 points. Please use US Letter size paper instead of A4.

Paper title is 16 point, caps/lc, bold, centered between 2 horizontal rules. Top rule is 4 points thick and bottom rule is 1 point thick. Allow 1/4 inch space above and below title to rules.

Author descriptions are center-justified, initial caps. The lead author is to be listed first (left-most), and the Co-authors are set to follow. If up to three authors, use a single row of author descriptions, each one center-justified, and all set side by side; with more authors or unusually long names or institutions, use more rows.

Use one-half line space between paragraphs, with no indent.

4 FIRST LEVEL HEADINGS

First level headings are all caps, flush left, bold, and in point size 12. Use one line space before the first level heading and one-half line space after the first level heading.

4.1 Second Level Heading

Second level headings are initial caps, flush left, bold, and in point size 10. Use one line space before the second level heading and one-half line space after the second level heading.

4.1.1 Third Level Heading

Third level headings are flush left, initial caps, bold, and in point size 10. Use one line space before the third level heading and one-half line space after the third level heading.

Fourth Level Heading Fourth level headings must be flush left, initial caps, bold, and Roman type. Use one line space before the fourth level heading, and place the section text immediately after the heading with no line break, but an 11 point horizontal space.

4.2 Citations, Figure, References

4.2.1 Citations in Text

Citations within the text should include the author's last name and year, e.g., (Cheesman, 1985). Be sure that the

sentence reads correctly if the citation is deleted: e.g., instead of "As described by (Cheesman, 1985), we first frobulate the widgets," write "As described by Cheesman (1985), we first frobulate the widgets."

The references listed at the end of the paper can follow any style as long as it is used consistently.

4.2.2 Footnotes

Indicate footnotes with a number[1] in the text. Use 8 point type for footnotes. Place the footnotes at the bottom of the column in which their markers appear, continuing to the next column if required. Precede the footnote section of a column with a 0.5 point horizontal rule 1 inch (6 picas) long.[2]

4.2.3 Figures

All artwork must be centered, neat, clean, and legible. All lines should be very dark for purposes of reproduction, and art work should not be hand-drawn. Figures may appear at the top of a column, at the top of a page spanning multiple columns, inline within a column, or with text wrapped around them, but the figure number and caption always appear immediately below the figure. Leave 2 line spaces between the figure and the caption. The figure caption is initial caps and each figure should be numbered consecutively.

Make sure that the figure caption does not get separated from the figure. Leave extra white space at the bottom of the page rather than splitting the figure and figure caption.

This figure intentionally left non-blank

Figure 1: Sample Figure Caption

4.2.4 Tables

All tables must be centered, neat, clean, and legible. Do not use hand-drawn tables. Table number and title always appear above the table. See Table 1.

Use one line space before the table title, one line space after the table title, and one line space after the table. The table title must be initial caps and each table numbered consecutively.

---

[1]Sample of the first footnote.

[2]Sample of the second footnote.

Table 1: Sample Table Title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

5 SUPPLEMENTARY MATERIAL

If you need to include additional *textual* appendices (such as missing proofs, additional results) during submission, please include them in the same PDF file as the main paper. Non-textual supplementary material (such as code or videos) must be submitted in a separate ZIP file named 642-supp.zip (with 642 replaced with your paper ID on CMT).

The textual supplementary material must follow the same formatting instructions as in the main paper. The only difference is that it must be in a *single-column* format. Please see section A for further instructions regarding supplementary material.

6 SUBMISSION INSTRUCTIONS

To submit your paper to AISTATS 2023, please follow these instructions.

1. Download `aistats2023.sty`, `fancyhdr.sty`, and `sample_paper.tex` provided in our starter pack. Please, do not modify the style files as this might result in a formatting violation.

2. Use `sample_paper.tex` as a starting point.

3. Begin your document with

   ```
   \documentclass[twoside]{article}
   \usepackage{aistats2023}
   ```

   The `twoside` option for the class article allows the package `fancyhdr.sty` to include headings for even and odd numbered pages.

4. When you are ready to submit the manuscript, compile the latex file to obtain the pdf file.

5. Check that the content of your submission, *excluding* references and textual supplementary materials, is limited to **9 pages**. The number of pages containing references and supplementary materials alone is not limited.

6. Upload the PDF file along with other (non-textual) supplementary material files to the CMT website.

6.1 Camera-ready Papers

If your papers are accepted, you will need to submit the camera-ready version. Please make sure that you follow these instructions:

1. Change the beginning of your document to

   ```
   \documentclass[twoside]{article}
   \usepackage[accepted]{aistats2023}
   ```

   The option `accepted` for the package `aistats2023.sty` will write a copyright notice at the end of the first column of the first page. This option will also print headings for the paper. For the *even* pages, the title of the paper will be used as heading and for *odd* pages the author names will be used as heading. If the title of the paper is too long or the number of authors is too large, the style will print a warning message as heading. If this happens additional commands can be used to place as headings shorter versions of the title and the author names. This is explained in the next point.

2. If you get warning messages as described above, then immediately after `\begin{document}`, write

   ```
   \runningtitle{Provide here an
   alternative shorter version of the
   title of your paper}
   \runningauthor{Provide here the
   surnames of the authors of your
   paper, all separated by commas}
   ```

   Note that the text that appears as argument in `\runningtitle` will be printed as a heading in the *even* pages. The text that appears as argument in `\runningauthor` will be printed as a heading in the *odd* pages. If even the author surnames do not fit, it is acceptable to give a subset of author names followed by "et al."

3. The camera-ready versions of the accepted papers are 9 pages, plus any additional pages needed for references and textual supplementary material.

4. If you need to include additional appendices: The textual supplementary material must be included in a single PDF file with the main paper. Non-textual supplementary material (e.g. code) must be submitted in a separate supplementary material ZIP file.

5. Please, do not change the layout given by the above instructions and by the style file.

Acknowledgements

References

References follow the acknowledgements. Use an unnumbered third level heading for the references section. Please use the same font size for references as for the body of the paper—remember that references do not count against your page length total.

References

J. Alspector, B. Gupta, and R. B. Allen (1989). Performance of a stochastic learning microchip. In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems 1*, 748–760. San Mateo, Calif.: Morgan Kaufmann.

F. Rosenblatt (1962). *Principles of Neurodynamics.* Washington, D.C.: Spartan Books.

G. Tesauro (1989). Neurogammon wins computer Olympiad. *Neural Computation* **1**(3):321–323.

A   FORMATTING INSTRUCTIONS APPENDICES

The textual supplementary material should be included in the **same PDF file** as the main paper. The only difference is that the supplementary material must be in a *single-column* format.

B   MISSING PROOFS

The supplementary materials may contain detailed proofs of the results that are missing in the main paper.

B.1   Proof of Lemma 3

In this section, we present the detailed proof of Lemma 3 and then […]

C   ADDITIONAL EXPERIMENTS

If you have additional experimental results, you may include them in the supplementary materials.

C.1   The Effect of Regularization Parameter

Our algorithm depends on the regularization parameter λ. Figure 1 below illustrates the effect of this parameter on the performance of our algorithm. As we can see, […]

D   NON-TEXTUAL SUPPLEMENTARY MATERIAL

The (optional) non-textual supplementary material (e.g. code for the paper) should be submitted as a separate ZIP file.

- Your supplementary material file should be named 642-supp.zip (with 642 replaced with your paper ID on CMT).

- If you wish to include any code/datasets as part of the supplementary material, you can: (i) include it in the ZIP file of your supplementary material, and/or (ii) provide the public URL where the code can be found in the appropriate field of CMT submission form. If you promised to release code/datasets (either in the original submission PDF or during the rebuttal), the code/dataset must be released. In this case, the AISTATS Chairs will check the availability of the code/dataset and, if it isn't available by the camera-ready deadline, your submission may be excluded from the conference proceedings.

- If you wish to include any **videos** as part of the supplementary material, please do not include them in the ZIP file. Instead, simply provide the public URL where the video is available in the appropriate field of the CMT submission form. NOTE: This does not refer to the recorded presentation explaining your paper, which will be submitted/recorded separately, but to any other videos containing, e.g., experimental results.

- Do not include any textual supplementary material in the ZIP file. It must be included in the same PDF as the main paper.