Towards Practical Private Data Release

Michael Fine Harvard University
Cambridge MA, 02138
mfine@college.harvard.edu

1 Towards Practical Private Data Release

1.1 Introduction

In the standard model of differential privacy, analysts submit queries to a trusted curator, who returns a noisy answer to each in an online fashion. While simple, this paradigm is suboptimal in a number of ways. It requires analysts to profoundly change their workflow -- instead of being able to inspect, query, and manipulate data they have access to, they must now submit queries to a curator without ever being able to look at the data. Moreover, as [TODO cite] Dwork notes, this model gives up *analyst privacy* -- the curator is necessarily aware of each query the analyst makes of the data.

Ideally, the curator would release a *differentially private synthetic dataset* -- a data structure that, while differential private, "looks like" the true dataset. More formally, given a sensitive dataset $\mathcal{X} \in \mathbb{R}^{m \times n}$, we are looking for a differentially private sanitizer M such that $\tilde{\mathcal{X}} = M(\mathcal{X})$ approximates $\mathcal{X}$ with respect to a class of queries Q:

$$\forall q \in Q: \;\; |q(\mathcal{X}) - q(\tilde{\mathcal{X}})| \leq \alpha$$

for some constant α.

[TODO lit review]

1.2 DualQuery

DualQuery views the problem of private data release as a zero-sum game between the query player and the data player. The value of the game is the difference between the query run on the data players move and the true dataset. In DualQuery's formulation, the query player uses a no-regret learning algorithm, while the data player finds a best response using an optimization algorithm.

1.3 OracleQuery

OracleQuery generalizes the no-regret adversarial approach of DualQuery to support any heuristic optimization oracle, while reducing the oracle-runtime dependence to $\log |Q|$ rather than $|Q|$. Unfortunately, it has a few drawbacks

- Runtime depends linearly on $1/\delta$, which is infeasible when δ is cryptographically $(< 10^{-100})$ small.
- Probabilistic optimization oracle must be able to certify that the solution it returns is optimal (though it is allowed to simply not return a result with small probability). While this works for certain optimization procedures, it is infeasible/impossible for procedures like gradient descent.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

1.4 Generative Adversarial Networks

In a parallel line of research, Generative Adversarial Networks (Goodfellow 2016) have shown great promise in generative realistic looking, high dimensionsional images. Quite similar to the DualQuery approach, a GAN is trained by pairing two deep neural networks, a generator and a discriminator. The generator aims to generate realistic samples, while the discriminator tries to distinguish between fake and real samples.

Beyond standard problems with training GAN's (mode collapse etc), the primary issue with the GAN formulation is that so far we've not been able to make any theoretical guarantees about the worst case error (or even average case) of a query over GAN-generated dataset. This is extremely important for our scenario, where analysts would like to know with high probability any result they obtain on the synthetic dataset would approximately hold on the true dataset. This lack of theoretical guarantees actually stems from two related problems:

1. No guarantee that the converged solution is a global optimum (or global nash equilibrium)
2. No guarantee that even a global nash equilibria in the *parameter space* is really optimal in the *distribution space*.

1.4.1 How combine theoretical guarantees with distribution learning capabilities of GANs

2 Submission of papers to NeurIPS 2019

NeurIPS requires electronic submissions. The electronic submission site is

`https://cmt.research.microsoft.com/NeurIPS2019/`

Please read the instructions below carefully and follow them faithfully.

2.1 Style

Papers to be submitted to NeurIPS 2019 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Additional pages *containing only acknowledgments and/or cited references* are allowed. Papers that exceed eight pages of content (ignoring references) will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2019 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper compared to earlier years.

Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

2.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the World Wide Web at

`http://www.neurips.cc/`

The file `neurips_2019.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2019 is `neurips_2019.sty`, rewritten for LaTeX 2_{ε}. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

Preprint option If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text “Preprint. Work in progress.” in the footer. This version may be distributed as you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2019.tex` may be used as a “shell” for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in

3 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors’ names are set in boldface, and each name is centered above the corresponding address. The lead author’s name is to be listed first (left-most), and the co-authors’ names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

regarding figures, tables, acknowledgments, and references.

4 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

4.1 Headings: second level

Second-level headings should be in 10-point type.

4.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

5 Citations, figures, tables, references

These instructions apply to everyone.

5.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

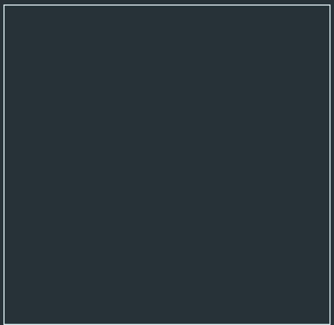

Figure 1: Sample figure caption.

```
http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf
```

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2019` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2019}
```

As submission is double blind, refer to your own published work in the third person. That is, use “In the previous work of Jones et al. [4],” not “In our previous work [4].” If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form “A. Anonymous.”

5.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

5.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

[1] Sample of the first footnote.

[2] As in this example.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

5.4 Tables

All tables must be centered, neat, clean and legible. The table number and

Place 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 lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

6 Exercise 4

IN PROGRESS

$$\sum_{t=1}^{T}\langle w^{(t)} - w^*, v_t\rangle = \sum_{t=1}^{T}\left[\frac{1}{2\eta_t}\left(-||w^{(t+1)} - w^*||^2 + ||w^{(t)} - w^*||^2\right)\right] + \sum_{t=1}^{T}\frac{\eta_t}{2}||v_t||^2$$

Factoring the first sum on the right side, pulling out the boundary terms

$$\sum_{t=2}^{T}\left[\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right)||w^{(t)} - w^*||^2\right] + \frac{1}{\eta_1}||w^{(1)} - w^*||^2 - \frac{1}{\eta_T}||w^{(T+1)} - w^*||^2 + \sum_{t=1}^{T}\frac{\eta_t}{2}||v_t||^2$$

We lower bound $||v_t||^2$ by ρ, $||w^*||$ by B, remove negative terms, and use the boundedness of $||w^{(t)} - w^*|| < 2B$ to get

$$\sum_{t=1}^{T}\langle w^{(t)} - w^*, v_t\rangle \le 2\sum_{t=2}^{T}\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right)B - \frac{1}{\eta_1}B^2 + \frac{\rho^2}{2}\sum_{t=1}^{T}\eta_t$$

$$\le 2\sum_{t=2}^{T}\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right)B - \frac{B}{\rho} + \frac{B\rho}{2}\sum_{t=1}^{T}1/\sqrt{T}$$

$$\le \sum_{t=2}^{T}\left[||w^{(t)} - w^*||^2\right] + T^{3/2}B\rho$$

Where the last step follows from the inequality $\sum_{i=1}^{T}\frac{1}{\sqrt{t}} \le 2\sqrt{T} - 1$.

7 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

8 Preparing PDF files

Please prepare submission files with paper size “US Letter,” and not, for example, “A4.”

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

8.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.