Optimization Theory and Algorithm	**Lecture 1 - XXX/XXXX/2021**

Lecture Title Here ...

Lecturer:Xiangyu Chang *Scribe: XXXXX*

Edited by: Junbo Hao

1 I am the section

In this lecture,

1.1 I am the subsection

- I am itemize.
- **I am bold**.
- *I am it*

2 Mathematics Formulation

This is the mathematics formulation $\sum_{i=1}^{n} \frac{1}{i^2}$.

We also can write it as

$$\sum_{i=1}^{n} \frac{1}{i^2}.$$

Let us do it for a complex form.

I need a number.

$$\sum_{i=1}^{n} \frac{1}{i^2}. \tag{1}$$

$$\begin{aligned} Y_t^{(i)} &= \frac{1}{K} \sum_{l \in \mathcal{S}_{\tau(t)}} M_l Z_{t-1}^{(l)} D_t^{(l)} + \frac{1}{K} \sum_{l \in \mathcal{S}_{\tau(t)}} N_{t-1}^{(l)} + N_t' \qquad (2) \\ &:= \frac{1}{K} \sum_{l \in \mathcal{S}_{\tau(t)}} M_l Z_{t-1}^{(l)} D_t^{(l)} + N_t + N_t', \end{aligned}$$

3 Algorithms and Code

Algorithm 1 I am the algorithm

1: **Input:** distributed dataset $\{A_i\}_{i=1}^m$, target rank k, iteration rank $r \geq k$, number of iterations T.
2: **Initialization:** orthonormal $Z_0^{(i)} = Z_0 \in \mathbb{R}^{d\times r}$ by QR decomposition on a random Gaussian matrix.
3: **for** $t = 1$ to T **do**
4: The i-th worker independently performs $Y_t^{(i)} = M_i Z_{t-1}^{(i)}$ for all $i \in [m]$, where $M_i = \frac{A_i^\intercal A_i}{s_i}$;
5: Each worker i sends $Y_t^{(i)}$ to the server and the server performs aggregation: $Y_t = \sum_{i=1}^m p_i Y_t^{(i)}$;
6: The server performs orthogonalization: $Z_t = \mathsf{orth}(Y_t)$ and broadcast Z_t to each worker such that $Z_t^{(i)} = Z_t$;
7: **end for**
8: **Output:** approximated eigen-space $Z_T \in \mathbb{R}^{d\times r}$ with orthonormal columns.

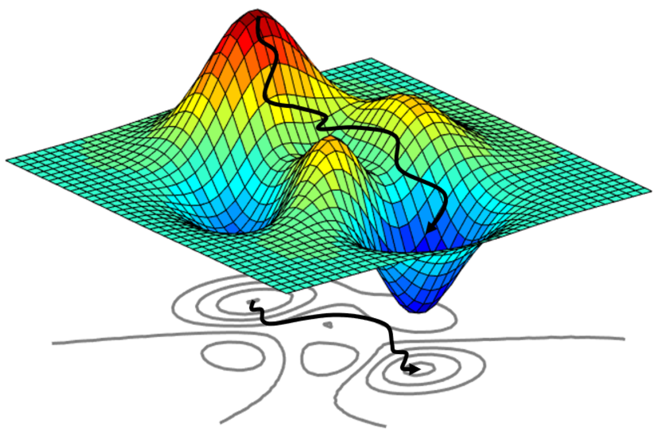

Figure 1: I am the Picture

4 Figures

5 Citation

You need to learn what a bib file is. How to make it, if you need.

This is the citation [Casella and Berger, 2002].

6 Theorem

Theorem 1 *I am a theorem.*

Example 1 *I am an example.*

Lemma 1 *I am a lemma.*

Definition 1 *I am a definition.*

7 Notations defined by ourselves

I am a scalar x or a.

I am a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$.

I am a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m \times n} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$, where $\mathbf{a}_i = (a_{1i}, \ldots, a_{im})^\top \in \mathbb{R}^m$ and $i = 1, \ldots, n$.

8 Coding

This is an example of coding.

```python
import numpy as np


def main():
    # test the numpy
    x = np.array([1, 2, 3, 4])
    print(x)

main()
```

References

[Casella and Berger, 2002] Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.