Principle of Proof Writing

Jiaxin Hu

June 30, 2020

1 Math, Notation

1. Specify the variables/functions. Every time you use a variable/function, you should explain it, including its domain and meaning. Use $:=$ or $\triangleq$ for definition or assignment. The operator $=$ means a equal comparison.

   - Let $\mathcal{A} = (\mathcal{C}, \{\boldsymbol{M}_k\})$ denote the decision variables.
     $\rightarrow$ Let $\mathcal{A} = (\mathcal{C}, \{\boldsymbol{M}_k\}) \in \mathbb{R}^d$ denote the decision variables, where $d = \prod_k r_k + \sum_k r_k d_k$ is the number of parameters.
   - Let S denote the update mapping.
     $\rightarrow$ Let $S : \mathbb{R}^d \mapsto \mathbb{R}^d$ denote the update mapping.
   - The objective function is a function of tensor coefficient $\mathcal{B} = \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$
     $\rightarrow$ The objective function is a function of tensor coefficient $\mathcal{B} := \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$
   - $\left\| \mathcal{B}(\mathcal{A}^{(t)}) - \mathcal{B}(\mathcal{A}^*) \right\|_F \leq c \left\| \mathcal{A}^{(t)} - \mathcal{A}^* \right\|_F$ $\rightarrow \left\| \mathcal{B}(\mathcal{A}^{(t)}) - \mathcal{B}(\mathcal{A}^*) \right\|_F \leq c \left\| \mathcal{A}^{(t)} - \mathcal{A}^* \right\|_F, \forall t \in \mathbb{N}_+$.

2. Make the notation consistent. You should not change the variable/function you defined previously without any explanation. You also should not use the same notation for two different things.

   - *The notation $\mathcal{L}$ a shorthand of $\mathcal{L}_{\mathcal{Y}}(\cdot)$. You should make it clear before you use it.*
     Suppose $\mathcal{A}^*$ is a stationary point of $\mathcal{L}(\cdot)$.
     $\rightarrow$ For notational convenience, we drop the subscript $\mathcal{Y}$ from the objective $\mathcal{L}_{\mathcal{Y}}(\cdot)$. The objective function can be viewed either as a function of decision variables $\mathcal{A}$ or a function of coefficient tensor $\mathcal{B}$. With slight abuse of notation, we write both function as $\mathcal{L}(\cdot)$ ... Suppose $\mathcal{A}^*$ is a stationary point of $\mathcal{L}(\cdot)$.
   - *If you use ∇f to refer the derivative or gradient of a function, you should not use df or f' in the rest of the proof.*

$$\nabla^2 \mathcal{L}(\mathcal{A}^*) = \begin{bmatrix} d_{CC}^2\mathcal{L} & d_{CM_1}^2\mathcal{L} & \cdots & d_{CM_K}^2\mathcal{L} \\ d_{M_1 C}^2\mathcal{L} & d_{M_1 M_1}^2\mathcal{L} & \cdots & d_{M_1 M_K}^2\mathcal{L} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M_K C}^2\mathcal{L} & d_{M_K M_1}^2\mathcal{L} & \cdots & d_{M_K M_K}^2\mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla_{CC}^2\mathcal{L} & \nabla_{CM_1}^2\mathcal{L} & \cdots & \nabla_{CM_K}^2\mathcal{L} \\ \nabla_{M_1 C}^2\mathcal{L} & \nabla_{M_1 M_1}^2\mathcal{L} & \cdots & \nabla_{M_1 M_K}^2\mathcal{L} \\ \vdots & \vdots & \ddots & \vdots \\ \nabla_{M_K C}^2\mathcal{L} & \nabla_{M_K M_1}^2\mathcal{L} & \cdots & \nabla_{M_K M_K}^2\mathcal{L} \end{bmatrix}$$

- *You should not use ρ for spectral radius and contraction parameter at the same time.*
  Let ρ be the spectral radius of ∇S... Let $\rho = \rho + \epsilon$ be the contraction parameter.
  $\rightarrow$ Let ρ be the spectral radius of ∇S... Let $\rho_0 = \rho + \epsilon$ be the contraction parameter.
- *Use bold for matrices.*
  $(\mathcal{C}, M_1, \ldots, M_K) \quad \rightarrow (\mathcal{C}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_K)$.

3. Avoid unnecessary notation.

   - *You have already explained the domain of $\mathcal{A}$. The new notation Ω is unnecessary.*
     Let Ω denote the domain of $\mathcal{A}$ and Ω_O denote the equivalent class of $\mathcal{A}^*$ ... For $\mathcal{A} \in \Omega_O$,
     $\ldots$, For $\mathcal{A} \in \Omega/\Omega_O \ldots$
     $\rightarrow$ Let Ω_O denote the equivalent class of $\mathcal{A}^*$ ... For $\mathcal{A} \in \Omega_O$, $\ldots$, For $\mathcal{A} \in \mathbb{R}^d/\Omega_O$ ...

2 Language

1. **Grammar! Grammar! Grammar!**

   - some notations $\quad \rightarrow$ some notation
   - There exists a sub-sequences of iterate $\mathcal{A}$ ... $\quad \rightarrow$ There exist a sub-sequence of iterate $\mathcal{A}$
   - Combine the equation 7 and 8, we have ... $\quad \rightarrow$ Combining the equation 7 and 8, we have...
   - The set $\mathcal{E}$ only contains a finite number of different equivalent classes.
     $\rightarrow$ The set $\mathcal{E}$ contains only a finite number of equivalent classes.

2. Use sentences. The math notation or equation should be a noun or short clause in a sentence.

3. Be short and concise. Proof is also a part of academic writing.

   - The set $\mathcal{E}$ satisfies below two properties.
     $\rightarrow$ The set $\mathcal{E}$ satisfies two properties below.
   - *The statement to derive (1) is trivial whereas (2) needs explanation.*
     The set $\mathcal{E}_S$ satisfies two properties below: (1) is ... (2) is... (1) comes from ... (2) comes from...
     $\rightarrow$ The set $\mathcal{E}_S$ satisfies two properties below: (1) is ... (2) is... ,which is comes from...

4. Use formal expressions.

   - Trivially,... $\quad \rightarrow$ Therefore,...
   - In other words,... $\quad \rightarrow$ We conclude that,...

5. Avoid "can".

   - We can conclude that,... $\quad \rightarrow$ We conclude that,...

6. Pay attention to the comma",".
   Add comma "," before injections, such as "and, but, or, so, yet", when :

   - *we are connecting three or more items.* Our model handles three types of data including continuous, count, and binary data.

- *we are connecting two independent sentences.* The statistical convergence of our estimator is established, and we quantify the gain in accuracy compared to classical multivariate regression approach.

Do not add comma "," before injections, when:

- *we are connecting two verbs.*
  The paper presents the theory and discusses the results.
  Tom is not taking his ski helmet on the trip, but he is taking his new ski pants.
  Tom is not taking this ski helmet but is taking this new ski pants.

7. Always put an article before countable nouns.

   - There are elimination matrix and permutation matrix such that ...
     $\rightarrow$ There exist an elimination matrix and a permutation matrix such that ...

8. Use "that", if the object of "assume, show, imply" is a sentence.

   - Assume the matrix X is full rank. The equation implies X is invertible.
     $\rightarrow$ Assume that the matrix X is full rank. The equation implies that X is invertible.

9. Use `` to complie " in latex.

10. Keep a space before brackets.

11. Add a backslash after i.e. .

3   Logic

1. Use a clear proof structure. You can prove step by step from assumptions to the goal or you can use contradiction. Never mix these two structures in a single proof.

   - *Consider the proof of Uniqueness of tensor tucker decomposition.*

2. Avoid big leaps. Make every step concrete.

   - *Consider the first version of local convergence. I took the statement that ∇S is invariant to orthogonal transformation for granted. Then the whole proof went to a wrong direction.*

3. Reorganize. Check whether your proof logic is a "chain".

   - *The order of proof writing is not the same as the way you think. So, make it readable for reader.*

4. Summarize the cited or too detailed steps. Also be short and concise logically.

   - *I used implicit function theorem to show that each micro-step in update mapping S is continuously differentiable. However, it is unnecessary to put such detailed thing in the proof.*
   - *The way I showed $\nabla S = -(L+D)^{-1}L^T$ was exactly the same as reference. Just cite the reference.*