Principle of Proof Writing

Jiaxin Hu

06/14/2020

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.
     $\to$ 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.
     $\to$ Let $S : \mathbb{R}^d \mapsto \mathbb{R}^d$ denote the update mapping.
   - The objective function is a function of tensor coefficient $\mathcal{B} = \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$
     $\to$ The objective function is a function of tensor coefficient $\mathcal{B} := \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$
   - $\left\|\mathcal{B}(\mathcal{A}^{(t)}) - \mathcal{B}(\mathcal{A}^*)\right\|_F \le c \left\|\mathcal{A}^{(t)} - \mathcal{A}^*\right\|_F \quad \to \left\|\mathcal{B}(\mathcal{A}^{(t)}) - \mathcal{B}(\mathcal{A}^*)\right\|_F \le c \left\|\mathcal{A}^{(t)} - \mathcal{A}^*\right\|_F, \ \forall t \in \mathbb{N}_+$.

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 $\mathcal{L}(\cdot)$.
     $\to$ 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.*

     $$\nabla^2 \mathcal{L}(\mathcal{A}^*) = \begin{pmatrix} d^2_{CC}\mathcal{L} & d^2_{CM_1}\mathcal{L} & \cdots & d^2_{CM_K}\mathcal{L} \\ d^2_{M_1C}\mathcal{L} & d^2_{M_1M_1}\mathcal{L} & \cdots & d^2_{M_1M_K}\mathcal{L} \\ \vdots & \vdots & \ddots & \vdots \\ d^2_{M_KC}\mathcal{L} & d^2_{M_KM_1}\mathcal{L} & \cdots & d^2_{M_KM_K}\mathcal{L} \end{pmatrix} \to \begin{pmatrix} \underset{\sim}{\nabla}^2_{CC}\mathcal{L} & \underset{\sim}{\nabla}^2_{CM_1}\mathcal{L} & \cdots & \underset{\sim}{\nabla}^2_{CM_K}\mathcal{L} \\ \underset{\sim}{\nabla}^2_{M_1C}\mathcal{L} & \underset{\sim}{\nabla}^2_{M_1M_1}\mathcal{L} & \cdots & \underset{\sim}{\nabla}^2_{M_1M_K}\mathcal{L} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{\sim}{\nabla}^2_{M_KC}\mathcal{L} & \underset{\sim}{\nabla}^2_{M_KM_1}\mathcal{L} & \cdots & \underset{\sim}{\nabla}^2_{M_KM_K}\mathcal{L} \end{pmatrix}$$

   - *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.
     $\to$ Let ρ be the spectral radius of ∇S... Let $\rho_0 = \rho + \epsilon$ be the contraction parameter.

- *Use bold for matrices.*
  $(\mathcal{C}, M_1, \ldots, M_K) \quad \to (\mathcal{C}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_K)$.

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 $\mathcal{A} \in \Omega/\Omega_O \ldots$
     $\to$ Let Ω_O denote the equivalent class of $\mathcal{A}^*$ ... For $\mathcal{A} \in \Omega_O$, $\ldots$, For $\mathcal{A} \in \mathbb{R}^d/\Omega_O \ldots$

2 LANGUAGE

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

   - some notations $\quad \to$ some notation
   - There exists a sub-sequences of iterate $\mathcal{A}$ ... $\quad \to$ There exist a sub-sequence of iterate $\mathcal{A}$
   - Combine the equation 7 and 8, we have ... $\quad \to$ Combining the equation 7 and 8, we have...

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 $\mathcal{E}$ only contains a finite number of different equivalent classes.
     $\to$ The set $\mathcal{E}$ contains only a finite number of equivalent classes.
   - The set $\mathcal{E}$ satisfies below two properties.
     $\to$ 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...
     $\to$ The set $\mathcal{E}_S$ satisfies two properties below: (1) is ... (2) is... ,which is comes from...

4. Use formal expressions.

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

5. Avoid "can".

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

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 $\nabla S = -(L+D)^{-1}L^T$ is exactly the same as reference. Just cite the reference.*