by 2154286 郑伟丞

Q1 : $inf\{sin(n) : n > 1\} = ?$, write down the answer and prove it.

answer: $inf\{sin(n) : n > 1\} = -1$

$inf\{sin(x) : x \in \mathbb{R}\} = -1$

So the problem is equivalent to proving $\forall \epsilon > 0, \exists k, n \in \mathbb{N}^*, |2k\pi - \frac{\pi}{2} - n| < \epsilon$

Let $\{x_t\} = -\frac{\pi}{2} + t, t \in \mathbb{N}$

$\exists p_t \in \mathbb{Z}, \exists r_t \in [0, 2\pi), -\frac{\pi}{2} + t = x_t = 2p_t\pi + r_t$

Obviously, for each $i \neq j$, $2\pi = \frac{(i-j)-(r_i-r_j)}{p_i-p_j}$ and $r_i \neq r_j$

Split $[0, 2\pi)$ into $\lfloor \frac{2\pi}{2\epsilon} \rfloor + 1$ pieces: $[0, \delta), [\delta, 2\delta), ..., [\theta - \delta, \theta), \delta = \frac{2\pi}{\lfloor \frac{2\pi}{2\epsilon} \rfloor + 1}$

Since $\{r_t\}$ is infinite, there exists r_i, r_j in the same one of the above intervals,and thus

$0 < r_i - r_j < \delta < 2\epsilon$

Since $x_i - x_j = i - j = 2(p_i - p_j)\pi + r_i - r_j$

Then $0 < (i - j) + 2(p_j - p_i)\pi = r_i - r_j < 2\epsilon$

Obviously, $\forall \nu \in \mathbb{R}, \exists \mu \in \mathbb{Z}, |\mu(r_i - r_j) - \nu| \leq \frac{1}{2}(r_i - r_j) < \epsilon$

Then have $\nu = \frac{\pi}{2} + 2w\pi, w \in \mathbb{Z}$ and $r_i - r_j = (i - j) + 2(p_j - p_i)\pi$,

it's satisfied that $|\mu(i - j) + 2[\mu(p_j - p_i) - w]\pi - \frac{\pi}{2}| < \epsilon$

Then have $k = \mu(p_j - p_i) - w, n = -\mu(i - j)$

when $i > j$,$\exists i, j \in \mathbb{N}^*, \exists w \in \mathbb{Z}, \exists \mu \in \mathbb{Z}^-$ to satisfy $k, n \in \mathbb{N}^*$

when $i > j$,$\exists i, j \in \mathbb{N}^*, \exists w \in \mathbb{Z}, \exists \mu \in \mathbb{N}^*$ to satisfy $k, n \in \mathbb{N}^*$

$\forall \epsilon > 0, \exists k, n \in \mathbb{N}^*, |2k\pi - \frac{\pi}{2} - n| < \epsilon$

Above all, the original statement is proved.

Q2 :Prove Cauchy-Schwartz inequality $\boldsymbol{x^T y \leq ||x||_2||y||_2}$ in two ways.

Method1:

Assume that vector y is a non-zero vector (if it is a zero vector, the equation holds)

Let $\boldsymbol{z} = \boldsymbol{x} - \frac{\boldsymbol{x}*\boldsymbol{y}}{\|\boldsymbol{y}\|_2^2} * \boldsymbol{y}$

Then $\boldsymbol{z} * \boldsymbol{z} = \|\boldsymbol{x}\|_2^2 - \frac{2\boldsymbol{x}*\boldsymbol{y}}{\|\boldsymbol{y}\|_2^2} \boldsymbol{x} * \boldsymbol{y} + \left(\frac{\boldsymbol{x}*\boldsymbol{y}}{\|\boldsymbol{y}\|_2^2}\right)^2 * \|\boldsymbol{y}\|_2^2 = \|\boldsymbol{z}\|_2^2 \geq 0$

$\therefore \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2 - \boldsymbol{x}(\boldsymbol{x} * \boldsymbol{y})^2 + (\boldsymbol{x} * \boldsymbol{y})^2 \geq 0$

$\therefore \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2 \geq (\boldsymbol{x} * \boldsymbol{y})^2$

So $\boldsymbol{x}^T \boldsymbol{y} \leq \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2$

Method2:

Let $\boldsymbol{z} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} - \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2^2}$

Then $\boldsymbol{z} * \boldsymbol{z} = \frac{\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} - \frac{2\boldsymbol{x}*\boldsymbol{y}}{\|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2} + \frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{y}\|_2^2}$

$\therefore \boldsymbol{2} - \frac{2\boldsymbol{x}*\boldsymbol{y}}{\|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2} \geq 0$

So $\boldsymbol{x}^T \boldsymbol{y} \leq \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2$

Q3 : Let $\mathbf{A}$ be a matrix of size $m \times n$. Denote the range space of $\mathbf{A}$ as $R(\mathbf{A})$ and the null space of $\mathbf{A}$ as $N(\mathbf{A})$,repectively. Prove $R(\mathbf{A}) = N^{\perp}(\mathbf{A}^T)$

$R(\boldsymbol{A}) = \{\boldsymbol{y} | \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n\}$

$N(\boldsymbol{A}^T) = \{\boldsymbol{x} | \boldsymbol{A}^T = 0, \boldsymbol{x} \in \mathbb{R}^m\}$

$N^{\perp}(\boldsymbol{A}^T) = \{\boldsymbol{y} | \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0, \forall \boldsymbol{x} \in N(\boldsymbol{A}^T)\}$

Let $\boldsymbol{A} = (\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{...} \boldsymbol{a_n})$, where $\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{...}, \boldsymbol{a_n} \in \mathbb{R}^m$

$\forall \boldsymbol{\mu} \in N(\boldsymbol{A}^T)$

$\boldsymbol{a_1^T \mu = a_2^T \mu = ... = a_n^T \mu = 0}$

So $x_1 \boldsymbol{a_1^T \mu} = x_2 \boldsymbol{a_2^T \mu} = ... = x_n \boldsymbol{a_n^T \mu} = 0$ where $x_1, x_2, ..., x_n \in \mathbb{R}$

So $(x_1 \boldsymbol{a_1^T} + x_2 \boldsymbol{a_2^T} + ... + x_n \boldsymbol{a_n^T})\boldsymbol{\mu} = 0$ where $x_1, x_2, ..., x_n \in \mathbb{R}$

Let $\boldsymbol{x} = (x_1, x_2, ..., x_n)^T$

So $x_1\boldsymbol{a_1^T}+x_2\boldsymbol{a_2^T}+...+x_n\boldsymbol{a_n^T}=\boldsymbol{x^T}\boldsymbol{A^T}$

$\boldsymbol{x^T}\boldsymbol{A^T}\boldsymbol{\mu}=(\boldsymbol{Ax})^{\boldsymbol{T}}\boldsymbol{\mu}=\langle\boldsymbol{Ax},\boldsymbol{\mu}\rangle=0$

So $N^{\perp}(\boldsymbol{A^T})=\{\boldsymbol{y}|\boldsymbol{y}=\boldsymbol{Ax},\boldsymbol{x}\in\mathbb{R}^{\boldsymbol{n}}\}=R(\boldsymbol{A})$

Q4 : For any two matrices, prove $trace(\boldsymbol{AB})=trace(\boldsymbol{BA})$

$(\boldsymbol{AB})_{ij}=\sum_k\boldsymbol{A}_{ik}\boldsymbol{B}_{kj}$

$trace(\boldsymbol{AB})=\sum_n\sum_k\boldsymbol{A}_{nk}\boldsymbol{B}_{kn}$

$(\boldsymbol{BA})_{ij}=\sum_k\boldsymbol{B}_{ik}\boldsymbol{A}_{kj}$

$trace(\boldsymbol{BA})$

$=\sum_n\sum_k\boldsymbol{B}_{nk}\boldsymbol{A}_{kn}$

$=\sum_n\sum_k\boldsymbol{A}_{kn}\boldsymbol{B}_{nk}$

$=\sum_k\sum_n\boldsymbol{A}_{nk}\boldsymbol{B}_{kn}$ (from n is equivalent to k)

$=\sum_n\sum_k\boldsymbol{A}_{nk}\boldsymbol{B}_{kn}$

$=trace(\boldsymbol{BA})$

Q5 : Prove $\boldsymbol{A}\succeq 0\Longleftrightarrow\langle\boldsymbol{A},\boldsymbol{B}\rangle$ for all $\boldsymbol{B}\succeq 0$

1.prove $\boldsymbol{A}\succeq 0\Longrightarrow\langle\boldsymbol{A},\boldsymbol{B}\rangle$ for all $\boldsymbol{B}\succeq 0$

$\because\boldsymbol{A}\succeq 0,\boldsymbol{B}\succeq 0$

$\therefore\boldsymbol{A}=\boldsymbol{P^T}\wedge\boldsymbol{P},\boldsymbol{B}=\boldsymbol{Q^T}\boldsymbol{DQ}$

where $\boldsymbol{P}$ is an orthogonal matrix, $\wedge$ is a diagonal matrix composed of the eigenvalues of $\boldsymbol{A}$, and the same applies to $\boldsymbol{B}$

$\langle\boldsymbol{A},\boldsymbol{B}\rangle$

$=trace(\boldsymbol{AB^T})$

$=trace(\boldsymbol{AB})$

$=trace(\boldsymbol{P^T}\wedge\boldsymbol{PQ^T}\boldsymbol{DQ})$

$= trace(\wedge \boldsymbol{PQ^TDQP^T})$ (from $trace(\boldsymbol{AB}) = trace(\boldsymbol{BA})$)

Let $\boldsymbol{C} = \boldsymbol{PQ^TDQP^T}$

$\boldsymbol{C}$ is a matrix with the same eigenvalues as $\boldsymbol{B}$, with non-negative diagonal elements, and $\wedge$ is a diagonal matrix with non-negative diagonal elements.

$\therefore trace(\wedge \boldsymbol{C}) \geq 0$

So $\langle \boldsymbol{A}, \boldsymbol{B} \rangle \geq 0$

2.prove $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ for all $\boldsymbol{B} \succeq 0 \Longrightarrow \boldsymbol{A} \succeq 0$

Suppose $\boldsymbol{A}$ is not a positive semidefinite matrix

Then $\exists \boldsymbol{x} \in \mathbb{R}$, $\boldsymbol{x^T A x} < 0$

Let $\boldsymbol{B} = \boldsymbol{xx^T}$

$\forall \boldsymbol{y} \in \mathbb{R}$, $\boldsymbol{y^T B y = y^T x x^T y = (x^T y)^T x^T y \geq 0}$

$\therefore \boldsymbol{B}$ must be a positive semidefinite matrix

$\langle \boldsymbol{A}, \boldsymbol{B} \rangle$

$= trace(\boldsymbol{AB}) = trace(\boldsymbol{Axx^T})$

Let $\boldsymbol{\mu} = \boldsymbol{Ax}$

$\boldsymbol{x^T \mu} = \sum_{i=1}^{n} x_i \mu_i$

$trace(\boldsymbol{\mu x^T}) = \sum_{i=1}^{n} \mu_i x_i$

$\therefore trace(\boldsymbol{\mu x^T}) = \boldsymbol{x^T \mu}$

$\therefore trace(\boldsymbol{Axx^T}) = \boldsymbol{x^T A x} < 0$

$\therefore \langle \boldsymbol{A}, \boldsymbol{B} \rangle < 0$

Contradictory to the premise, the assumption is not true

So $\boldsymbol{A} \succeq 0$

Above all, the original statement is proved.

Q6 : Define $f(\boldsymbol{x}) \triangleq ||\mathbf{A}\boldsymbol{x} - b||_2^2$. Compute $\nabla f(\boldsymbol{x}) and \nabla^2 f(\boldsymbol{x})$

$f(\boldsymbol{x}) \triangleq (\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^2$

$\frac{\partial f(\boldsymbol{x})}{\partial x_k}$ (1)

$= \frac{\partial \sum_i (\sum_j A_{ij}x_j - b_i)^2}{\partial x_k}$

$= 2\sum_i (\sum_j A_{ij}x_j - b_i)A_{ik}$

Let $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$

Then $D_i = \sum_j A_{ij}x_j - b_i$

So $(1) = 2\sum_i D_i A_{ik} = 2\sum_i A^T_{ki} D_i$ (2)

Let $\boldsymbol{B} = \boldsymbol{A^T}\boldsymbol{D}$

Then $B_k = \sum_i A^T_{ki} D_i$

So $(2) = 2B_k$

So $\nabla f(\boldsymbol{x}) = 2\boldsymbol{B} = x\boldsymbol{A^T(Ax-b)}$

$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_k \partial x_l} = \frac{2\partial \sum_i (\sum_j A_{ij}x_j - b_i)A_{ik}}{\partial x_l}$

$= 2\sum_i A_{ik}A_{il}$

$= 2\sum_i A^T_{ki}A_{il}$

$= 2(A^T A)_{kl}$

So $\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A^T}\boldsymbol{A}$

Q7 : Define $f(\boldsymbol{x}) \triangleq \|\boldsymbol{A} - \boldsymbol{x}\boldsymbol{x^T}\|_F^2$. Compute $\nabla f(\boldsymbol{x})$ and $\nabla^2 f(\boldsymbol{x})$.

$f(\boldsymbol{x}) \triangleq \sum_i \sum_j (A_{ij} - x_i x_j)^2$

$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = 2(A_{kk} - x_k^2)(-2x_k) - 2\sum_{i\neq k}(A_{ik} - x_i x_k)x_i - 2\sum_{j\neq k}(A_{ki} - x_k x_j)x_j$

$= -2(\sum_i (A_{ik} - x_i x_k)x_i + \sum_j (A_{kj} - x_k x_j)x_j)$ (1)

Let $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{x}\boldsymbol{x^T}$

Then $D_{ij} = A_{ij} - x_i x_j$

So $(1) = -2[(D^T x)_k + (Dx)_k]$

So $\nabla f(\boldsymbol{x}) = -2(\boldsymbol{D^T x} + \boldsymbol{Dx}) = -2[(\boldsymbol{A} - \boldsymbol{xx^T})^T + (\boldsymbol{A} - \boldsymbol{xx^T})]\boldsymbol{x}$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_k \partial x_l} = \frac{-2\partial(\sum_i (A_{ik} - x_i x_k)x_i + \sum_j (A_{kj} - x_k x_j)x_j}{\partial x_l} = \begin{cases} 4\sum_i x_i^2 - 4A_{kk} + 8x_k^2 \\ 8x_k x_l - 2A_{lk} - 2A_{kl} \end{cases}$$

So $\nabla^2 f(\boldsymbol{x}) = 8xx^T + 4x^T xI - 2A - 2A^T$

Q8 : For the logistic regression example in lecture notes, compute $\nabla E(\boldsymbol{x})$

$E(\boldsymbol{w}) = -\sum_{n=1}^{N}\{t_n ln y_n + (1-t_n)ln(1-y_n)\}$

$y_n = \sigma(\boldsymbol{w^T \phi_n}) = \frac{1}{1+e^{-\boldsymbol{w}^T\phi_n}}$

$\nabla_{(\boldsymbol{w^T\phi_n})} E(\boldsymbol{w}) = \nabla_{(\boldsymbol{w^T\phi_n})} y_n \nabla E(\boldsymbol{w}) = y_n(1-y_n)(\frac{1-t_n}{1-y_n} - \frac{t_n}{y_n}) = y_n - t_n$

$\boldsymbol{y} = (y_1, y_2, ..., y_N)^T$

$\boldsymbol{t} = (t_1, t_2, ..., t_N)^T$

$\boldsymbol{\phi} = (\phi_1^T, \phi_2^T, ..., \phi_N^T)^T$

So $\nabla E(\boldsymbol{w}) = \nabla_{\boldsymbol{w}}(\boldsymbol{\phi w})\nabla_{(\boldsymbol{\phi w})} E(\boldsymbol{w}) = \boldsymbol{\phi^T(y-t)}$

Q9 : Define $f(\boldsymbol{x}) \triangleq log\sum_{k=1}^{n} e^{x_k}$. Prove $\nabla^2 f(\boldsymbol{x}) \succeq 0$

$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial(\frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}})}{\partial x_j} = \begin{cases} \frac{e^{x_i}(\sum_{k=1}^{n} e^{x_k} - e^{x_i})}{(\sum_{k=1}^{n} e^{x_k})^2} & i = j \\ \frac{-e^{x_i + x_j}}{(\sum_{k=1}^{n} e^{x_k})^2} & i \neq j \end{cases}$$

$\forall \boldsymbol{x} \in \mathbb{R}$

$\boldsymbol{x^T}\nabla^2 f(\boldsymbol{x})\boldsymbol{x} = \sum_i \sum_j A_{ij}x_i x_j = \sum_i \frac{e^{x_i}(\sum_{k=1}^{n} e^{x_k} - e^{x_i})x_i^2}{(\sum_{k=1}^{n} e^{x_k})^2} - \sum_{i \neq j} \frac{e^{x_i + x_j} x_i x_j}{(\sum_{k=1}^{n} e^{x_k})^2}$

$= \frac{(\sum_{k=1}^{n} e^{x_k} x_k^2)(\sum_{k=1}^{n} e^{x_k}) - (\sum_{k=1}^{n} e^{x_k} x_k)^2}{(\sum_{k=1}^{n} e^{x_k})^2}$

From Cauchy's inequality we get

$(\sum_{k=1}^{n} e^{x_k} x_k)^2 \leq (\sum_{k=1}^{n} e^{x_k} x_k^2)(\sum_{k=1}^{n} e^{x_k})$

So $\boldsymbol{x^T}\nabla^2 f(\boldsymbol{x})\boldsymbol{x} \geq 0$

So $\nabla^2 f(\boldsymbol{x}) \succeq 0$

Q10 : Find at least one example in either of the following two fields that can be formulated as an optimization problem and show how to formulate it

1.EDA software

2.cluster scheduling for data centers

1.EDA software

在电子设计自动化（EDA）软件中，一个常见的优化问题是布局优化。布局优化的目标是巧妙地安排芯片上的电路元件位置和互连，以实现最佳性能，即尽量减小延迟，同时也考虑功耗和面积等因素。这个优化问题本质上是一个组合优化，通过给定一组元件和面积约束，旨在找到最佳的元件布局和连接方式，以最大化或最小化特定的目标函数（比如延迟、功耗、性能等）。优化任务包括确定每个元件的位置、方向以及每个连接的路径，同时需要充分考虑功耗和芯片面积等关键因素。

$min_{x,y,\theta,r}\ f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r})$

$s.t.$

$g(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r}) \leq 0$

$h(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r}) = 0$

其中, $\boldsymbol{x},\boldsymbol{y}$ 是组件的坐标, $\boldsymbol{\theta}$ 是组件的方向, $\boldsymbol{r}$ 是连接的路径. $f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r})$ 是目标函数，代表芯片的性能指标如功耗、延时。 $g(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r})$ 是不等式约束，表示芯片的面积限制，连接长度限制等。 $h(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{r})$ 是等式约束，表示组件之间的逻辑关系等。

2.cluster scheduling for data centers

数据中心集群中有多项任务需要执行。每个任务都需要一定的计算和存储资源，并需要在特定的服务器上运行。调度算法的目标是最大化数据中心资源的利用率，同时最小化任务完成时间。这个问题可以表示为一个优化问题，涉及任务的分配和执行顺序。优化变量是任务的分配和执行顺序，目标是在满足约束条件的前提下最大化资源利用率并最小化任务完成时间。

举例来说，考虑一个包含10台服务器的数据中心集群，每台服务器拥有8个CPU核心和16GB内存。共有20个任务需要执行，每个任务都有不同的计算和存储资源需求，以及特定的服务器限制。例如，任务A需要2个CPU核心和4GB内存，只能在服务器1、2、3上执行；而任务B需要4个CPU核心和8GB内存，可以在任何服务器上执行......

$min \max_i y_i$

$s.t.$

$\sum_{j=1}^{m} x_{ij} c_i \leq C_i y_i, i=1,...,n$

$\sum_{j=1}^{m} x_{ij} m_j \leq M_i y_i, i=1,...,n$

$\sum_{i=1}^{n} x_{ij} = 1, j=1,...,m$

$x_{ij} \in \{0,1\}, i=1,...,n, j=1,...,m$

$y_i = [0,1], i=1,...,n$

其中， n 是服务器的数量， m 是任务的数量, x_{ij} 是个01变量，表示 任务 j 是否分配给了服务器 i. y_i 是个连续变量，表示服务器 i 的资源利用率. c_j 和 m_j 分别表示任务所需的cpu核数和内存大小。 C_i和 M_i 分别表示服务器i的CPU核数和内存大小。

目标函数是最小化所有服务器之间的最大资源利用率，即均衡各个服务器的负载。 限制因素包括:

每台服务器的 CPU 和内存消耗不能超过其总量乘以资源利用率。

每个任务只能分配给一台服务器，并且必须满足其所需服务器的限制。

资源利用率必须介于 0 和 1 之间。