Basic Stochastic Analysis in Marine Engineering

Xiangjun Kong
xjunkong@gmail.com

Table of Contents

1 Letter Notations - Math Symbol Fraktur

$\mathfrak{A}$**A**,$\mathfrak{B}$B, $\mathfrak{C}$C, $\mathfrak{D}$D, $\mathfrak{E}$E, $\mathfrak{F}$F, $\mathfrak{G}$**G**, $\mathfrak{H}$**H**, $\mathfrak{I}$I, $\mathfrak{J}$J, $\mathfrak{K}$K, $\mathfrak{L}$L, $\mathfrak{M}$**M**,

$\mathfrak{N}$N, $\mathfrak{O}$O, $\mathfrak{P}$P, $\mathfrak{Q}$Q, $\mathfrak{R}$R, $\mathfrak{S}$S, $\mathfrak{T}$**T**,$\mathfrak{U}$U, $\mathfrak{V}$V, $\mathfrak{W}$W, $\mathfrak{X}$X, $\mathfrak{Y}$Y, $\mathfrak{Z}$Z

2 Gumbel Distribution

The Gumbel distribution is a particular case of the generalized extreme value distribution (also known as the Fisher-Tippett distribution). It is also known as the log-Weibull distribution and the double exponential distribution (Note that later name is sometimes confusingly used to refer to a totally different distribution, called Laplace distribution[1]). The cumulative distribution function of the Gumbel distribution is

$$F(x;\mu,\beta) = F_X(X \le x) = e^{-e^{-\frac{x-\mu}{\beta}}} = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t. \tag{1}$$

where μ and β are called location and shape parameters, respectively, and the probability density function $f_X(x)$ is defined as

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x-\mu}{\beta} - e^{-\frac{x-\mu}{\beta}}}. \tag{2}$$

The mode is μ, while the median is $\mu - \beta \ln(\ln 2)$, and the expectation[2] is given by

$$\mathrm{E}(X) = \bar{x} = \mu + \gamma\beta. \tag{3}$$

The standard deviation[3] σ is $\pi\beta/\sqrt{6}$, hence $\beta = \sigma\sqrt{6}/\pi \approx 0.78\sigma$. At the mode, where $x = \mu$, the value of $F(x;\mu,\beta)$ becomes $e^{-1} \approx 0.367879$, irrespective of the value of β. Euler constant $\gamma = 0.57721566490$.

Gumbel Distribution Modeling: One example for convergence of a distribution. Let (X_n) be a sequence of independent and identically distributed exponential random variables with parameter λ. Let M_n denote $\max\{X_1, \ldots, X_n\}$. Show there exists a random variable Z such that $M_n - \frac{1}{\lambda}\log(n)$ converges to the Gumbel distribution $F(x;0,1/\lambda)$ in (1) with parameters $\mu = 0$ and $\beta = 1/\lambda$.

[1] Laplace distribution has pdf

$$f(x;\mu,\beta) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}}.$$

Due to the absolute sign in the exponential, it becomes symmetric about the mean μ as two exponential distributions spliced together back-to-back. It is reminiscent of the normal distribution which is however expressed in terms of the squared difference from the mean μ, instead of the absolute difference, therefore the Laplace distribution has sharper peak and fatter tails than the normal distribution.

[2] Also called expected value, mean, average, or first moment and is a generalization of the weighted average of random variable X, $\bar{x} = \mathrm{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x$.

[3] The standard deviation is a measure of the variation or scattering of a sequence. The square is $\sigma^2 = \mathrm{E}\left[(X-\bar{x})^2\right] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\,\mathrm{d}x$.

Unfortunately we do not have convergence in probability, so that approach is not going to succeed. The basic trick is to note that $M_n \leq a$ if and only if $X_i \leq a$ for each $i = 1, \ldots, n$ and then use independence:

$$P(M_n - \tfrac{1}{\lambda}\log n \leq x) = P(M_n \leq \tfrac{1}{\lambda}\log n + x) = \prod_{i=1}^{n} P(X_i \leq \tfrac{1}{\lambda}\log n + x).$$

Now remembering that $P(X_i \leq a) = 1 - e^{-\lambda a}$, we find

$$P(M_n - \tfrac{1}{\lambda}\log n \leq x) = \left(1 - e^{-\lambda(\frac{1}{\lambda}\log n + x)}\right)^n = \left(1 - \frac{e^{-\lambda x}}{n}\right)^n \to e^{-e^{-\lambda x}}$$

as $n \to \infty$. Hence $M_n \to Z$ in distribution where $P(Z \leq x) = e^{-e^{-\lambda x}}$.

Extreme Design Value at certain probability p of non-exceedance can be found by solving the following equation

$$F(\hat{x}; \mu, \beta) = F_X(X \leq \hat{x}) = p \tag{4}$$

The extreme value is

$$\begin{aligned} \hat{x} &= -\beta \ln(-\ln p) + \mu \\ &= \bar{x} - \sigma\left[\gamma + \ln(-\ln p)\right]\frac{\sqrt{6}}{\pi} \end{aligned} \tag{5}$$

Note:

- Coefficient of variation (CoV) is defined by $\sigma/\bar{x}$
- The standard Gumbel distribution is the special case when $\mu = 0$ and $\beta = 1$ in (1), i.e., the pdf $f(x) = e^{-(x+e^{-x})}$ and its CDF $F(x) = e^{-e^{-x}}$.
- The Most Probable Maximum(**MPM**), as its name would suggest, is the extreme value at the mode probability where $x = \mu$, $p = F(x; \mu, \beta) = e^{-1} \approx 37\%$ percentile or 63% probability of exceedance. The extreme becomes $\hat{x}_{\text{MPM}} = \bar{x} - \sigma\gamma\frac{\sqrt{6}}{\pi} = \bar{x} - 0.4500536\sigma$.
- The Expected Maximum(**EM**) is simply the arithmetic mean of selected maxima, i.e., $\hat{x}_{\text{EM}} = \bar{x}$. In this case, the second term in (5) vanishes which results in the non-exceedance probability $p = e^{-e^{-\gamma}} = 0.570376$ or called 57% percentile.

3 Poisson Distribution

Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. We can obtain the Poisson distribution as a limit of binomial distributions if $n \to \infty$ and $p \to 0$ such that $\lambda := np$ (constant) is finite to represent the expected number of events per unit

time/area (or average rate of random events).
Suppose $X \sim B(n,p)$ with $\lambda = np$

$$\begin{aligned} P\{X=k\} &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} && (6)\\ &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} && (7)\\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} && (8) \end{aligned}$$

Then for $n \to \infty$ and moderate λ , we have

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \;\; \frac{n(n-1)\cdots(n-k+1)}{n^k} \to 1, \;\; \left(1-\frac{\lambda}{n}\right)^k \to 1 \tag{9}$$

Therefore,

$$P\{X=k\} \to e^{-\lambda} \frac{\lambda^k}{k!}. \tag{10}$$

The expected values of X and X^2 are

$$\begin{aligned} \mathrm{E}(X) &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!}\\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\\ &= \lambda e^{-\lambda} e^{\lambda} = \boldsymbol{\lambda} && (11)\\ \mathrm{E}(X^2) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{(k-1)!}\\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \frac{d}{d\lambda}(\lambda^k) = \lambda e^{-\lambda} \frac{d}{d\lambda} \left[\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right]\\ &= \lambda e^{-\lambda} \frac{d}{d\lambda} \left[\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right]\\ &= \lambda e^{-\lambda} \frac{d}{d\lambda}(\lambda e^{\lambda}) = \lambda e^{-\lambda}(e^{\lambda} + \lambda e^{\lambda}) = \boldsymbol{\lambda + \lambda^2}. && (12) \end{aligned}$$

So its variance $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \lambda$.

4 Generalized Mean

If p is a non-zero real number, $x_1, \ldots, x_n$ and $w_1, \ldots, w_n$ are positive real values and the corresponding positive weights (not necessarily normalized such that $\sum_{i=1}^n w_i = 1$), we define the weighted power mean or Generalized Mean as:

$$M_p(x_1,\dots,x_n)=\left(\frac{\sum_{i=1}^{n} w_i x_i^p}{\sum_{j=1}^{n} w_j}\right)^{\frac{1}{p}}=\exp\left(\frac{\ln\left(\frac{\sum w_i x_i{}^p}{\sum w_j}\right)}{p}\right) \tag{13}$$

$$M_0(x_1,\dots,x_n)=\left(\prod_{i=1}^{n} x_i^{w_i}\right)^{\frac{1}{\sum_{j=1}^{n} w_j}} \tag{14}$$

If we let $w_i=1/n$ for $i=1\dots n$, then we have

$$M_{-\infty}(\mathbf{x})=\lim_{p\to-\infty} M_p(x_1,\dots,x_n)=\min\{x_1,\dots,x_n\} \qquad \text{Minimum} \tag{15}$$

$$M_{-1}(\mathbf{x})=\frac{n}{\frac{1}{x_1}+\cdots+\frac{1}{x_n}} \qquad \text{Harmonic Mean} \tag{16}$$

$$M_0(\mathbf{x})=\lim_{p\to0} M_p(x_1,\dots,x_n)=\sqrt[n]{x_1\cdot\cdots\cdot x_n} \qquad \text{Geometric Mean} \tag{17}$$

$$M_1(\mathbf{x})=\frac{x_1+\cdots+x_n}{n} \qquad \text{Arithmetic Mean} \tag{18}$$

$$M_2(\mathbf{x})=\sqrt{\frac{x_1^2+\cdots+x_n^2}{n}} \qquad \text{RootMean Square} \tag{19}$$

$$M_3(\mathbf{x})=\sqrt[3]{\frac{x_1^3+\cdots+x_n^3}{n}} \qquad \text{Cubic Mean} \tag{20}$$

$$M_{+\infty}(\mathbf{x})=\lim_{p\to\infty} M_p(x_1,\dots,x_n)=\max\{x_1,\dots,x_n\} \qquad \text{Maximum} \tag{21}$$

Proof of $\lim_{p\to\infty} M_p=\max\{x_1,\dots,x_n\}$. Assume (possibly after relabeling and combining terms together) that $x_1\ge\cdots\ge x_n$. Then

$$\begin{aligned}
\lim_{p\to\infty} M_p(x_1,\dots,x_n)&=\lim_{p\to\infty}\left(\frac{\sum_{i=1}^{n} w_i x_i^p}{\sum_{j=1}^{n} w_j}\right)^{\frac{1}{p}}=x_1\lim_{p\to\infty}\left(\frac{\sum_{i=1}^{n} w_i\left(\frac{x_i}{x_1}\right)^p}{\sum_{j=1}^{n} w_j}\right)^{\frac{1}{p}}\\
&=x_1\exp\left(\lim_{p\to\infty}\frac{\ln\left(\frac{\sum w_i\left(\frac{x_i}{x_1}\right)^p}{\sum w_j}\right)}{p}\right) \qquad \text{(L'Hôpital's rule}\to\text{)}\\
&=x_1\exp\left(\lim_{p\to\infty}\frac{\left(\sum w_j\right)\sum w_i\left(\frac{x_i}{x_1}\right)^p\ln\frac{x_i}{x_1}}{\sum w_k\left(\frac{x_k}{x_1}\right)^p}\right)\\
&=x_1e^0=x_1=M_\infty(x_1,\dots,x_n)=\max\{X_1,\dots,X_n\},
\end{aligned} \tag{22}$$

where the last line comes from the fact that

$$\lim_{p\to\infty}\left(\frac{x_i}{x_1}\right)^p\ln\frac{x_i}{x_1}=0 \quad \text{for any } \frac{x_i}{x_1}\le1 \text{ and } i=1,\dots,n. \tag{23}$$

By L'Hôpital's rule, we prove M_0 (14) in the similar way for the limit $p \to 0$,

$$\lim_{p\to 0} \frac{\ln\left(\frac{\sum\limits_{i=1}^{n} w_i x_i^p}{\sum\limits_{j=1}^{n} w_j}\right)}{p} = \lim_{p\to 0} \frac{\ln\left(\sum\limits_{i=1}^{n} w_i x_i^p\right)' - \ln\left(\sum\limits_{j=1}^{n} w_j\right)'}{p'} \tag{24}$$

$$= \lim_{p\to 0} \frac{\sum\limits_{i=1}^{n} w_i x_i^p \ln x_i}{\sum_{k=1}^{n} w_k x_k^p} \tag{25}$$

$$= \lim_{p\to 0} \frac{\sum\limits_{i=1}^{n} w_i x_i^p \ln x_i}{\sum\limits_{k=1}^{n} w_k x_k^p} \tag{26}$$

$$= \sum_{i=1}^{n} \frac{w_i \ln x_i}{\lim_{p\to 0} \sum\limits_{k=1}^{n} w_k \left(\frac{x_k}{x_i}\right)^p} \tag{27}$$

$$= \sum_{i=1}^{n} \frac{w_i \ln x_i}{\sum\limits_{k=1}^{n} w_k} \tag{28}$$

$$= \frac{\sum\limits_{i=1}^{n} \ln x_i^{w_i}}{\sum\limits_{k=1}^{n} w_k} \tag{29}$$

$$= \ln\left(\prod_{i=1}^{n} x_i^{w_i}\right)^{\frac{1}{\sum\limits_{k=1}^{n} w_k}} \tag{30}$$

The proof of $\lim_{p\to -\infty} M_p = \min\{x_1, \dots, x_n\}$ follows from

$$M_{-\infty}(x_1, \dots, x_n) = \frac{1}{M_\infty(\frac{1}{x_1}, \dots, \frac{1}{x_n})} = \min\{X_1, \dots, X_n\}. \tag{31}$$

5 Weibull Distribution

We will introduce the general Weibull distribution with 3-parameter model and list a bunch of relevant formulas.

pdf	$f(x,\gamma,\beta) = \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\mu}{\beta}\right)^\gamma}$
CDF	$F(x) = 1 - e^{-\left(\frac{x-\mu}{\beta}\right)^\gamma}$
Reliability	$R(x) = e^{-\left(\frac{x-\mu}{\beta}\right)^\gamma}$
Hazard	$h(x) = \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1}$
Mean	$\mu + \beta\Gamma\left(1+\frac{1}{\gamma}\right)$
Median	$\mu + \beta(\ln 2)^{\frac{1}{\gamma}}$
Mode	$\mu + \beta\left(1-\frac{1}{\gamma}\right)^{\frac{1}{\gamma}}$
Variance	$\beta^2\left[\Gamma\left(1+\frac{2}{\gamma}\right) - \Gamma\left(1+\frac{1}{\gamma}\right)^2\right]$

(32)

with β the scale parameter (Characteristic Life), γ the shape parameter (Weibull Slope), μ the shift or location parameter (Waiting Time) and Γ is the Gamma function with $\Gamma(N) = (N-1)!$ for integer N.

Weibull distributions is widely used to model lifetimes that are not "memoryless" . For example, each of the following gives an application of the Weibull distribution, for example, the lifetime of a car battery or the survival/fatal probability of someone having the longevity over 100 years old. When $\gamma = 1$, the Weibull distribution becomes an exponential distribution (with rate parameter $\lambda = 1/\beta$), so the exponential distribution is a special case of both the Weibull distributions and the gamma distributions. The Weibull distribution interpolates between the exponential distribution ($\gamma = 1$) and the Rayleigh distribution ($\gamma = 2$ and $\beta = \sqrt{2}\sigma$, with σ being mode or the scale parameter of the Rayleigh distribution). CDF $F(x)$ represents the probability of Time-To-Failure (TTF) at most x and the survival function (aka reliability function) $R(t) = 1-F(t)$ is defined as the probability of survival for at least x units of time, in other words, the probability of non-exceedance or no failure before time x. The failure rate (or hazard function) can be defined as the ratio between the current probability of survival $f(X \leq x)$ and the reliability function $R(t)$

$$h(x) = \frac{f(x)}{R(x)} = \frac{f(x)}{1-F(x)} \tag{33}$$

The cumulative hazard function for the Weibull is the integral of the failure rate

$$H(x) = \int_{\mu}^{x} h(t)\,\mathrm{d}t = \int_{\mu}^{x} \frac{\gamma}{\beta}\left(\frac{t-\mu}{\beta}\right)^{\gamma-1} \mathrm{d}t = \left(\frac{x-\mu}{\beta}\right)^{\gamma}. \tag{34}$$

The n-th moment of a real-valued continuous function $f(x)$ of a real variable about a value c is

$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n f(x)\mathrm{d}x \tag{35}$$

For the second and higher moments ($n \geq 2$), the central moment (about the mean $c = \mu_1$) are usually used rather than the moments about zero, because they provide clearer information about the distribution's shape.

The n-th moment about the origin ($x = 0$) of a probability density function $f(x)$ is the expected value of X^n, denoted by $\mathrm{E}(X^n)$ and is called a raw moment, crude moment or non-central moment, denoted by μ'_n. The n-th moment about its mean μ_1 is usually called n-th central moment, denoted by μ^c_n. The non-central and central moments are related simply through (35)

$$\mu'_n = \sum_{i=0}^{n} \binom{n}{i} \mu_1^i \mu_1^{n-i} \mu^c_n; \quad \text{and } \mu^c_n = \sum_{i=0}^{n} \binom{n}{i} (-\mu_1)^i \mu_1^{n-i} \mu'_n \tag{36}$$

where the binary expansion and $c = \mu_1$ are used.

For Weibull distribution given in (32), the non-central moments can be shown as below

$$
\begin{aligned}
\mu'_n &= \int_{-\infty}^{\infty} x^n \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{\gamma}} \mathrm{d}x \\
&= \int_0^{\infty} \left(\mu + \beta t^{\frac{1}{\gamma}}\right)^n e^{-t} \mathrm{d}t && (37)\\
&= \sum_{i=0}^{n} \binom{n}{i} \mu^i \beta^{n-i} \int_0^{\infty} t^{\frac{n-i}{\gamma}} e^{-t} \mathrm{d}t && (38)\\
&= \sum_{i=0}^{n} \binom{n}{i} \mu^i \beta^{n-i} \Gamma\left(1 + \frac{n-i}{\gamma}\right) && (39)
\end{aligned}
$$

where in (37) used is $t = (\frac{x-\mu}{\beta})^{\gamma}$, i.e., $x = \mu + \beta t^{\frac{1}{\gamma}}$, $\mathrm{d}t = \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1}\mathrm{d}x$. The mean or expectation of general Weibull distribution can be immediately followed by evaluating (39) with $n = 1$, i.e.,

$$
\begin{aligned}
\mu'_1 &= \sum_{i=0}^{1} \binom{1}{i} \mu^i \beta^{1-i} \Gamma\left(1 + \frac{1-i}{\gamma}\right) && (40)\\
&= \binom{1}{0} \mu^0 \beta^1 \Gamma\left(1 + \frac{1}{\gamma}\right) + \binom{1}{1} \mu^1 \beta^0 \Gamma\left(1 + \frac{0}{\gamma}\right) && (41)\\
&= \mu + \beta\Gamma\left(1 + \frac{1}{\gamma}\right) && (42)
\end{aligned}
$$

This is actually the derivation of the formula given in (32). The other three non-central moments can be expanded from (39).

$$
\begin{aligned}
\mu'_2 &= \mu^2 + 2\mu\beta\Gamma\left(1 + \frac{1}{\gamma}\right) + \beta^2\Gamma\left(1 + \frac{2}{\gamma}\right) && (43)\\
\mu'_3 &= \mu^3 + 3\mu\beta^2\Gamma\left(1 + \frac{2}{\gamma}\right) + 3\mu^2\beta\Gamma\left(1 + \frac{1}{\gamma}\right) + \beta^3\Gamma\left(1 + \frac{3}{\gamma}\right) && (44)\\
\mu'_4 &= \mu^4 + 4\mu\beta^3\Gamma(1 + \frac{3}{\gamma}) + 6\mu^2\beta^2\Gamma(1 + \frac{2}{\gamma}) + 4\mu^3\beta\Gamma(1 + \frac{1}{\gamma}) + \beta^4\Gamma(1 + \frac{4}{\gamma}). && (45)
\end{aligned}
$$

One can see that the non-central moments present perfectly in a binomial pattern of $\mu'_n = (\mu'_1)^n = \left(\mu + \beta\Gamma(1 + \frac{1}{\gamma})\right)^n$ with $\Gamma(1 + \frac{1}{\gamma})^n$ replaced by $\Gamma(1 + \frac{n}{\gamma})$.

In numerical applications, the central moments about the mean $\mu_1 = \mu + \beta\Gamma\left(1 + \frac{1}{\gamma}\right)$ are utilized in the Weibull parameter identification. These central moments, denoted by $\mu_n^{c=\mu_1}$

can be similarly shown as

$$\begin{aligned}\mu_n^{c=\mu_1} &= \int_{-\infty}^{\infty} (x-\mu_1)^n \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{\gamma}} \mathrm{d}x \\ &= \int_0^{\infty} \beta^n \left[t^{\frac{1}{\gamma}} - \Gamma\left(1+\frac{1}{\gamma}\right)\right]^n e^{-t}\mathrm{d}t \qquad (46)\\ &= \beta^n \sum_{i=0}^{n} \binom{n}{i} (-1)^i \Gamma\left(1+\frac{1}{\gamma}\right)^i \int_0^{\infty} t^{\frac{n-i}{\gamma}} e^{-t}\mathrm{d}t \qquad (47)\\ &= \beta^n \sum_{i=0}^{n} \binom{n}{i} (-1)^i \Gamma\left(1+\frac{1}{\gamma}\right)^i \Gamma\left(1+\frac{n-i}{\gamma}\right) \qquad (48)\end{aligned}$$

As expected, the first central moment about the mean μ_1 is the expectation which vanishes $\mu_1^{c=\mu_1} = 0$ from (48). The variance $\mu_2^{c=\mu_1}$ (about the mean), the 3rd central moment and the 4th central moment can be found by the generating formula (48).

$$\begin{aligned}\mu_2^{c=\mu_1} &= \beta^2 \sum_{i=0}^{2} \binom{2}{i} (-1)^i \Gamma\left(1+\frac{1}{\gamma}\right)^i \Gamma\left(1+\frac{2-i}{\gamma}\right) \qquad (49)\\ &= \beta^2 \left[\Gamma\left(1+\frac{2-0}{\gamma}\right) + 2(-1)\Gamma\left(1+\frac{1}{\gamma}\right)^2 + \Gamma\left(1+\frac{1}{\gamma}\right)^2 \Gamma(1)\right] \qquad (50)\\ &= \beta^2 \left[\Gamma\left(1+\frac{2}{\gamma}\right) - \Gamma\left(1+\frac{1}{\gamma}\right)^2\right] \qquad (51)\end{aligned}$$

$$\begin{aligned}\mu_3^{c=\mu_1} &= \beta^3 \left[\sum_{i=0}^{3} \binom{3}{i} (-1)^i \Gamma\left(1+\frac{1}{\gamma}\right)^i \Gamma\left(1+\frac{3-i}{\gamma}\right)\right] \qquad (52)\\ &= \beta^3 \left[\Gamma\left(1+\frac{3}{\gamma}\right) - 3\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(1+\frac{2}{\gamma}\right) + 2\Gamma\left(1+\frac{1}{\gamma}\right)^3\right] \qquad (53)\end{aligned}$$

$$\begin{aligned}\mu_4^{c=\mu_1} =&\beta^4 \left[\sum_{i=0}^{4} \binom{4}{i} (-1)^i \Gamma\left(1+\frac{1}{\gamma}\right)^i \Gamma\left(1+\frac{4-i}{\gamma}\right)\right] \qquad (54)\\ =&\beta^4 \left[\Gamma\left(1+\frac{4}{\gamma}\right) - 4\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(1+\frac{3}{\gamma}\right) + 6\Gamma\left(1+\frac{1}{\gamma}\right)^2 \Gamma\left(1+\frac{2}{\gamma}\right)\right.\\ &\left. - 3\Gamma\left(1+\frac{1}{\gamma}\right)^4\right] \qquad (55)\end{aligned}$$

The variance (51) is the proof of the variance formulae given in (32). The n-th moment about the location parameter μ (actually degenerated or converted to the two-parameter Weibull model) can be readily expressed

$$\begin{aligned}\mu_n^{c=\mu} &= \int_{-\infty}^{\infty} (x-\mu)^n \frac{\gamma}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{\gamma}} \mathrm{d}x \\ &= \int_0^{\infty} \left(\beta t^{\frac{1}{\gamma}}\right)^n e^{-t}\mathrm{d}t \qquad (56)\\ &= \beta^n \Gamma\left(1+\frac{n}{\gamma}\right) \qquad (57)\end{aligned}$$

where $t = (\frac{x-\mu}{\beta})^\gamma$ is used in (56).

Let's end this section by one example of a product life-cycle estimation. If the mean Time To Failure for a product (for instance mobile phone) which follows a Weibull distribution is 5 years (42,750 hours) with a standard deviation of 10,000 hours, what is the probability that the phone will continuously work more than 6 years (51,300 hours)? Since we only have two parameters given, it will most probably fit a two-parameter Weibull distribution model with the location parameter $\mu = 0$. Using the two conditions, the mean TTF is

$$\bar{x} = \beta\Gamma(1 + \frac{1}{\gamma}) \tag{58}$$

and the standard deviation σ implies the variance

$$\sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma\left(1 + \frac{1}{\gamma}\right)^2\right]. \tag{59}$$

From (58), β is obtained in terms of $\bar{x}$ and γ and plugged into (59) to get

$$\frac{\sigma^2}{\bar{x}^2} + 1 = \frac{\Gamma\left(1 + \frac{2}{\gamma}\right)}{\Gamma\left(1 + \frac{1}{\gamma}\right)^2}. \tag{60}$$

A simple root-finding algorithm can be used to solve (60) for $\gamma \approx 4.917$ and from (58) β can be approximately 46605 (hours). So, the probability for that phone to work more than 6 years ($x = 51,300$) is $1 - F(x) = e^{-(x/\beta)^\gamma} = 20.1\%$.

6 Maximum Likelihood Estimation

Let x_1, ..., x_n be random sample of size n taken from a probability density function $f_X(x;\theta)$ with the unknown parameter θ. The likelihood function of this random sample is the joint density of the n random variables and is a function of the unknown parameter. In general, the likelihood function is defined as the joint probability of the random variables X_1, X_2,..., X_n with the one or more parameters, say θ_1, θ_2, ... as follows.

$$\begin{aligned}\mathcal{L}(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots) &= f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots) &(61)\\ &= f(x_1; \theta_1, \theta_2, \ldots)f(x_2; \theta_1, \theta_2, \ldots)\ldots f(x_n; \theta_1, \theta_2, \ldots)\\ &= P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n | \theta_1, \theta_2, \ldots) &(62)\end{aligned}$$

Note that the independent variables of the likelihood function are $\theta_1, \theta_2, \ldots$, not x_i which are the observed values.

Maximum Likelihood Estimation (MLE) of Weibull family can be derived by taking the derivative of the log–likelihood function. Since we are assuming that we have an independent sample of x_1, ..., x_n, we know

$$f(x_1, \ldots, x_n; \beta, \gamma, \mu) = \prod_{i=1}^{n} \left(\frac{\gamma}{\beta}\left(\frac{x_i - \mu}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x_i-\mu}{\beta}\right)^\gamma}\right). \tag{63}$$

Taking the natural logarithm to (63) so that all the products become summation and let L denote the resulting outcome

$$\mathrm{L}(\beta,\gamma,\mu) = n(\ln\gamma - \ln\beta) + (\gamma-1)\sum_{i=1}^{n}\ln\frac{x_i-\mu}{\beta} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\beta}\right)^{\gamma} \tag{64}$$

Now, the maximum likelihood estimator means that the solution parameters can maximize the likelihood function $\mathcal{L}$ by vanishing the derivatives of $\mathrm{L}(\beta,\gamma,\mu)$ which is less complicated since natural logarithmic function is monotonically increasing. Therefore, the maximum likelihood can be achieved by letting the derivatives of the log-likelihood equation to zeros and solving the non-linear set of the three equations, and subsequently the Weibull parameters β, γ and μ can be estimated.

$$\frac{\partial\,\mathrm{L}(\beta,\gamma,\mu)}{\partial\gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n}\ln\frac{x_i-\mu}{\beta}\left[1-\left(\frac{x_i-\mu}{\beta}\right)^{\gamma}\right] = 0 \tag{65}$$

$$\frac{\partial\,\mathrm{L}(\beta,\gamma,\mu)}{\partial\beta} = -\frac{n\gamma}{\beta} + \frac{\gamma}{\beta}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\beta}\right)^{\gamma} = 0 \tag{66}$$

$$\frac{\partial\,\mathrm{L}(\beta,\gamma,\mu)}{\partial\mu} = (1-\gamma)\sum_{i=1}^{n}\frac{1}{x_i-\mu} + \frac{\gamma}{\beta}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\beta}\right)^{\gamma-1} = 0 \tag{67}$$

The Weibull location parameter μ has very important physical significance. In many reliability applications, failures do not occur below a certain limit which is also known as a failure-free life (FFL) parameter in the engineering literature. The three-parameter Weibull model with this FFL parameter has been widely used to describe the reliability of product, see **citeChanseokPark.**

However, as one can see from (63), as the sample variable is approaching the location parameter μ, i.e., $x_i \to \mu$, $\ln(x_i-\mu) \to \pm\infty$ for $\gamma \neq 1$. For any set of solution, it is always possible to find another set of solution by making μ closer to x_{min} that can give a larger likelihood value $\mathcal{L}$. This makes the MLE method bod down and no straight-forward algorithm can be used to find the location parameter μ by MLE approach. One possible or widely accepted alternative is to set $\mu = x_{min} - 1/n$ and solve the rest two equations. It can be found from (66) that

$$\beta^{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^{\gamma} \quad \text{or } \beta = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^{\gamma}\right]^{\frac{1}{\gamma}}. \tag{68}$$

Insert (68) to (65) (or equivalently into (64) and take the derivative with respect to γ), rearrange the result and (65) becomes

$$\frac{\partial\,\mathrm{L}(\beta,\gamma,\mu)}{\partial\gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n}\ln(x_i-\mu) - \frac{n\sum_{i=1}^{n}(x_i-\mu)^{\gamma}\ln(x_i-\mu)}{\sum_{i=1}^{n}(x_i-\mu)^{\gamma}} = 0 \tag{69}$$

With the known μ, the equation has only one unknown γ, so the general root finding numerical method can be used to solve it. Extreme value distributions are the limiting distributions for the minimum or the maximum of a very large collection of random observations from the same arbitrary distribution. Gumbel (1958) showed that for any well-behaved initial distribution (i.e., $F(x)$ is continuous and has an inverse), only a few models are needed, depending on whether you are interested in the maximum or the minimum, and also if the observations are bounded above or below.

In the context of reliability modeling, extreme value distributions for the minimum are frequently encountered. For example, if a system consists of n identical components in series, and the system fails when the first of these components fails, then system failure times are the minimum of n random component failure times. Extreme value theory says that, independent of the choice of component model, the system model will approach a Weibull as n becomes large. The same reasoning can also be applied at a component level, if the component failure occurs when the first of many similar competing failure processes reaches a critical level.

The distribution often referred to as the Extreme Value Distribution (Type I) is the limiting distribution of the minimum of a large number of unbounded identically distributed random variables. The PDF and CDF are given by:

$$\begin{aligned} f(x) &= \tfrac{1}{\beta} e^{-\frac{x-\mu}{\beta} - e^{-\frac{x-\mu}{\beta}}}, \quad -\infty < x < \infty,\ \beta > 0 \\ F(x) &= 1 - e^{-e^{-\frac{x-\mu}{\beta}}}, \qquad -\infty < x < \infty,\ \beta > 0 \end{aligned} \tag{70}$$

The above Gumbel distribution (1) is to model the distribution of the maximum value. To model the minimum value, use the negative of the original values. If the x values are bounded below (as is the case with times of failure) then the limiting distribution is the Weibull.

The natural log-transformed Weibull distribution is the Extreme Value Distribution Model of Gumbel type.

In any modeling application for which the variable of interest is the minimum of many random factors, all of which can take positive or negative values, try the extreme value distribution as a likely candidate model. For lifetime distribution modeling, since failure times are bounded below by zero, the Weibull distribution is a better choice.

The Weibull distribution and the extreme value distribution have a useful mathematical relationship. If $t_1,\ t_3,\ \ldots,\ t_n > 0$ are a sample of random times of failure from a Weibull distribution, then $\ln(t_1)$, $\ln(t_2)$, $\ldots$, $\ln(t_n)$ are random observations from the extreme value distribution. In other words, the natural log of a Weibull random time is an extreme value random observation. If the Weibull distribution, i.e., $X \approx \mathrm{W}(\beta, \gamma)$, and $Y = \ln X$, has the shape parameter γ and the characteristic life β, then Gumbel the extreme value distribution (after taking natural logarithms) has the location $\mu = \ln(\beta)$, and the scale $\beta = \frac{1}{\gamma}$.

$$P(Y \le y) = P(\ln X \le y) = P(X \le e^y) \tag{71}$$

$$= 1 - e^{-\left(\frac{e^y}{\beta}\right)^\gamma} \tag{72}$$

$$= 1 - e^{-e^{(y-\ln\beta)\gamma}} \tag{73}$$

$$= 1 - e^{-e^{\frac{y-\ln\beta}{1/\gamma}}} \tag{74}$$

Because of this relationship, computer programs designed for the extreme value distribution can be used to analyze Weibull data. The situation exactly parallels using normal distribution programs to analyze lognormal data, after first taking natural logarithms of the data points.

7 Method of Moments Estimation

The method of moments method of estimation was introduced by Karl Pearson (1894, 1895). The procedure consists of equating as many population moments to sample moments as there are parameters to estimate. Mathematical support for this procedure comes from the principle of moments as discussed in detail in Kendall and Stuart (1969). In essence, this principle says that two distributions that have a finite number of lower moments in common will be approximations of one another. Thus, the distribution of the data is approximated by equating the moments of a distributional form to the data moments. To see how this could be done with the three-parameter Weibull distribution, and hence for the three-parameter Weibull, n-th moment about the mean for the population, also called the n-th central moment $\mu_n^{c=\mu_1}$, is denoted by (48). The corresponding sample moments are calculated in this way.

$$m_k = \sum_{i=1}^{n} \frac{x_i^k}{n} \tag{75}$$

and the central moments are

$$m_k^c = \sum_{i=1}^{n} \frac{(x_i - m_1)^k}{n} \tag{76}$$

8 Least Square Estimation

Performing rank regression on the sample distribution requires that a straight line mathematically be fitted to a set of data points such that the sum of the squares of the vertical deviations from the points to the line is minimized. This is in essence the same methodology as the probability plotting method, except that we use the principle of least squares to determine the line through the points, as opposed to just eyeballing it. The first step is to bring our function into a linear form. For a general Weibull distribution model, the cumulative density function is:

$$F(x) = 1 - e^{-\left(\frac{x-\mu}{\beta}\right)^{\gamma}} \tag{77}$$

8.1 Population

In statistics the term "population" has a slightly different meaning from the one given to it in ordinary speech. It need not refer only to people or to animate creatures –the population of Britain, for instance or the dog population of London. Statisticians also speak of a population of objects, or events, or procedures, or observations, including such things as the quantity of lead in urine, visits to the doctor, or surgical operations. A population is thus an aggregate of creatures, things, cases and so on. Population size: The total number of people in the group you are trying to study.

8.2 Sample

A sample is a set of individuals or objects collected or selected from a statistical population by a defined procedure. The elements of a sample are known as sample points, sampling units or observations. A population commonly contains too many individuals to study conveniently, so an investigation is often restricted to one or more samples drawn from it. A well chosen sample will contain most of the information about a particular population parameter but the relation between the sample and the population must be such as to allow true inferences to be made about a population from that sample. Sample size is the number of completed responses your survey receives. It' s called a sample because it only represents part of the group of people (or target population) whose opinions or behavior you care about.

For a series of sample points, it is simpler to use ordinal distribution function $F(x_i) = i/(n+1)$ (or $(i-1/2)/n$) instead of i/n to avoid boundary or numerical issues. Let's say you've sorted your data so $X_1, \ldots, X_n$ refers to the smallest up to the biggest values in the data. Then, X_i is a pretty good estimate of the (i/n)th quantile of the distribution. But i/n and $i/(n+1)$ are really close values and typically the true quantiles of i/n and $i/(n+1)$ are also really close. Thus, we can argue that X_i is a pretty good estimate of the $(i/(n+1))$th quantile of the distribution. So, why does all this matter? Consider the biggest value in the data set, X_n. Would this ever be a good estimate of quantile $n/n = 1$? Of course not. For a Weibull distribution, quantile 1 is always equal to ∞ yet X_n is always finite. That said, X_n is still a good approximation of quantile $n/(n+1)$ which will approach to 100% when $n \to \infty$. Therefore, we have a series of coordinate pairs $(x_i, F(x_i))$ which can be arranged into a paper plot.

Taking the natural logarithm of both sides of (77) and changing the variable notation from x to t (as x will be reserved for the line equation variable) yields:

$$\ln[1-F(t)] = -\left(\frac{t-\mu}{\beta}\right)^{\gamma} \tag{78}$$

After changing the signs and taking natural logarithm again to get

$$\ln[1-F(t)] = -\left(\frac{t-\mu}{\beta}\right)^{\gamma} \tag{79}$$

$$\ln\{-\ln[1-F(t)]\} = \underbrace{\ln\left(\ln\frac{1}{1-F(t)}\right)}_{y} = \gamma\underbrace{\ln(t-\mu)}_{x} + \underbrace{(-\gamma\ln\beta)}_{a} \tag{80}$$

In this form, the location parameter must be known beforehand or be determined by iteration to find the minimal error function. In order to determine whether or not a particular sample came from a population distribution that follows the determined Weibull distribution model, one may check the probability plot. A typical plot is like Fig.1 and if the plot shows the sample points (maxima or minima) fall close to a straight line, then the solved distribution parameters is a reasonable model. As the plot shows, the shape parameter γ represents the line slope which implies that the larger γ has steeper line relative to the horizontal (extreme value) axis and the tail becomes thicker.

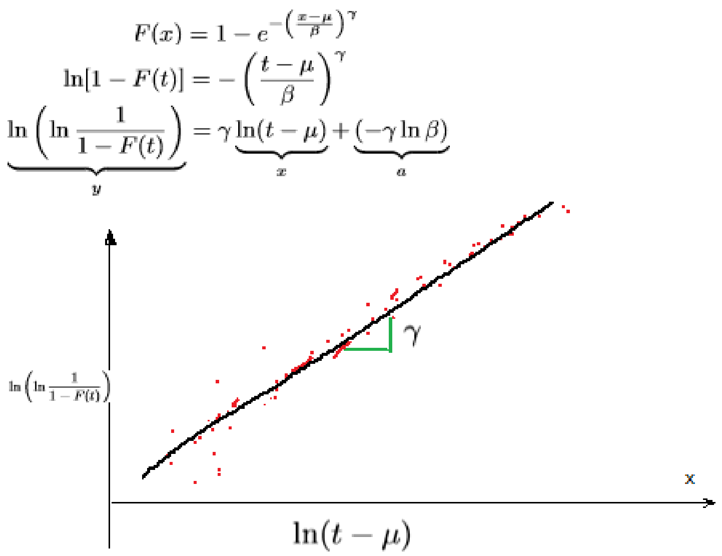

Fig. 1. Example of Least Square Fitting for Weibull Distribution

Alternatively, with any given shape parameter γ, the sample value can be extracted directly as the dependent variable and only one logarithmic operation is involved as shown below.

$$\underbrace{t}_{y} = \beta\underbrace{\{-\ln[1 - F(t)]\}^{\frac{1}{\gamma}}}_{x} + \underbrace{\mu}_{a} = \beta\left[\ln\frac{1}{1-F(t)}\right]^{\frac{1}{\gamma}} + \mu \tag{81}$$

For the first format in (80), if let $y = \ln\{-\ln[1-F(t)]\}$, $x = \ln(t-\mu)$, intercept $a = -\gamma\ln(\beta)$ and the slope $b = \gamma$, or $y = t$, $x = [\ln\frac{1}{1-F(t)}]^{1/\gamma}$ and $a = \mu$ for the second form (81), the linear equation $y = a + bx$ is readily formed. The least square estimation method (also known as regression analysis) is to find a line with the intercept and slope parameters such that the accumulated squared distances can reach a minimal. The well-known solutions are

$$\hat{a} = \frac{\sum_{i=1}^{n} y_i - \hat{b}\sum_{i=1}^{n} x_i}{n} = \bar{y} - \hat{b}\bar{x} \tag{82}$$

$$\hat{b} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{83}$$

where $\bar{x}$ and $\bar{y}$ are the arithmetic means for x and y respectively. The correlation coefficient is defined as follows:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{84}$$

σ_{xy} is the covariance of x and y, σ_x and σ_y the standard deviations of x and y, respectively. The estimator of ρ is the sample correlation coefficient $\hat{\rho}$, given by:

$$\hat{\rho} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}} \tag{85}$$

9 Design Limit States

In marine engineering applications, an extreme value of a response variable is always connected to a specific Return Period[4] (RP in short), with notation T_R. The most interested for the offshore owners or operators are how many years their production can be safely sustained. In normal practice, some operating limits as follows are defined as the warning boundary conditions beyond which a structure or a part of a structure exceeds a specified design requirement.

- Ultimate limit states (ULS) corresponding to the ultimate strength/load. For such ultimate condition, the representative value is corresponding to a load with an annual exceedance probability $q \leq 10^{-2}$ (or written as $q \leq 0.01$) i.e., once in every 100 years.

- Accidental limit states (ALS) corresponding to damage to components due to an accidental event or operational failure. The representative load is defined as an accidental load beyond the ULS which in general can have a probability corresponding to 10^{-4} annual exceedance, once every 10 000 years.

- Fatigue limit states (FLS) related to the possibility of failure due to the effect of cyclic loading. The representative value is defined as the expected load history.

- Serviceability limit states (SLS) corresponding to the criteria applicable to normal use or durability. The representative value is a specified value, dependent on operational requirements.

It is unrealistic to simulate an environments for 100-year real time duration. But it is possible to represent the environments in statistic way. To take account different scales of variation of for instance the wind speed, wave height, wave period, current speed and all directions etc, the whole variation history can be split into short term (say 3 hours, 1 minute or 1 second) and long term variability (like 1 year, 10 years, 100 years, 1000 years or 10 000 years). The short-term variation metocean data is usually described in a spectrum form representing a random 3-hour stationary sea state. A common procedure is to study

[4] The averaged length of time between any two events to occur. One plain and clear explanation is given by Ivan Haigh

a probability P that one event (typically the response amplitude X) of duration t does not exceed x within the RP T_R.

$$P(X \leq x) = 1 - \frac{t}{T_R} \tag{86}$$

For instance, a wave with the averaged period $t = 10s$ occur once over 100 year RP (see **?**), the exceedance probability Q can be defined as the inverse of the number of the response cycles N, i.e.,

$$Q = \frac{1}{N}. \tag{87}$$

where $N = (100 \cdot 365.25 \cdot 24 \cdot 3600)/10$. The corresponding non-exceedance probability is

$$P(X \leq x) = 1 - Q = 1 - \frac{1}{N} = 1 - 3.168808781 \cdot 10^{-9} \tag{88}$$

In the theory of maximum events, the non-exceedance probability of the maximum value of N, independent and identically distributed random variables X, can be written as
The most probable maximum for exceeding x once in N cycles can be found

$$P(X \leq x_{max}) = (P(X \leq x))^N \tag{89}$$

$$= \lim_{N\to\infty} \prod_{i=1}^{N} \left(1 - \frac{1}{N}\right) = \lim_{N\to\infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} \approx 0.37. \tag{90}$$

Physically, the time constant represents the elapsed time required for the system response to decay to zero if the system had continued to decay at the initial rate, because of the progressive change in the rate of decay the response will have actually decreased in value to $e^{-1} \approx 36.8\%$ in this time (say from a step decrease). In an increasing system, the time constant is the time for the system's step response to reach $1 - e^{-1} \approx 63.2\%$ of its final (asymptotic) value (say from a step increase). In radioactive decay the time constant is related to the decay constant λ, and it represents both the mean lifetime of a decaying system (such as an atom) before it decays, or the time it takes for all but 36.8% of the atoms to decay. For this reason, the time constant is longer than the half-life, which is the time for only 50% of the atoms to decay.

For short-term extremes, if the probability of exceedance during the 3-hour sea state is $F(x)$, then the total probability P (if most probable maximum, $P = e^{-1} \approx 0.37$) for any duration time t is

$$F(x)^N = P \tag{91}$$

$$F(x) = P^{1/N} \tag{92}$$

where P is the probability for x being exceeded once every N peaks during t seconds. The number of 3-hour sea states is $\frac{t}{N \cdot t_{3h}}$ and $t_{3h} = 10800s$ for 3-hour duration.
For long-term extreme, for instance 100yr ULS condition

$$F(x)^N = 1 - \frac{1}{100} \tag{93}$$

$$F(x) = (1 - 0.01)^{\frac{1}{N}} \tag{94}$$

Annual probability of non-exceedance for N (independent events per year), e.g. $N = 20/35$ (for 20 storms in 35 years). The exceedance probability $1/N$ of a single 3-hour maximum to exceed the value for the limit states are: $1/(100 \cdot 2922)$ and $1/(10000 \cdot 2922)$ for 100yr-ULS and 10000yr-ALS respectively.

10 Important Inequalities

The best well-known inequality is the so-called AM-GM inequality which is short for the two kinds of averaging methods, i.e., **A**rithmetic **M**ean and **G**eometric **M**ean. Every pupil knows that the arithmetic mean of any two numbers a and b is just

$$\mathfrak{A} = \frac{a+b}{2} \tag{95}$$

The simplest AM-GM inequality can be readily found by squaring the difference of two numbers.

$$(a-b)^2 = a^2 + b^2 - 2ab \geq 0. \tag{96}$$

From (96), it is obvious that we have proved that

$$a^2 + b^2 \geq 2ab. \tag{97}$$

If we add $2ab$ to both sides, the left will be just the square of $a + b$

$$a^2 + b^2 + 2ab = (a+b)^2 \geq 4ab. \tag{98}$$

Up to now, the inequalities are valid for arbitrary numbers of a and b. But for later convenience, we will assume all numbers to be positive reals otherwise explicitly stated.

$$\sqrt{ab} \leq \frac{a+b}{2} \tag{99}$$

If not known before, but now we define $\mathfrak{G} = \sqrt{ab}$ as the geometric mean for two positive real numbers a and b. The inequality (99) is what we called well-known AM-GM Inequality. If we square which becomes (99),

$$ab \leq \left(\frac{a+b}{2}\right)^2. \tag{100}$$

Since adding an equality equation to an inequality does not change the direction of the inequality, we add (97) and equality (98) with $2ab$ cancelled in the resulted inequality

$$\begin{align} a^2 + b^2 \qquad &\geq 2ab \tag{101} \\ a^2 + b^2 + 2ab &= (a+b)^2 \tag{102} \end{align}$$

and together with (100), we get

$$ab \leq \left(\frac{a+b}{2}\right)^2 \leq \frac{a^2+b^2}{2}. \tag{103}$$

The alternative form for arbitrary real a, b is

$$\sqrt{|ab|} \leq \frac{|a|+|b|}{2} \leq \sqrt{\frac{a^2+b^2}{2}}. \tag{104}$$

This will be widely utilized in proving various inequalities.

$$\mathfrak{M}_r(\mathbf{a}) = \Big(\mathfrak{A}(\mathbf{a}^r)\Big)^{\frac{1}{r}} = \left(\frac{1}{n}\sum_{i=1}^{n} a_i^r\right)^{\frac{1}{r}} \text{ or } = \left(\frac{\sum\limits_{i=1}^{n} p_i a_i^r}{\sum\limits_{i=1}^{n} p_i}\right)^{\frac{1}{r}} \tag{105}$$

$$\mathfrak{A}\ (\mathbf{a}) = \mathfrak{M}_1(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n} a_i \text{ or } = \frac{\sum\limits_{i=1}^{n} p_i a_i}{\sum\limits_{i=1}^{n} p_i} \tag{106}$$

$$\mathfrak{M}_{rs}(\mathbf{a}) = \Big(\mathfrak{M}_s(\mathbf{a}^r)\Big)^{\frac{1}{r}} = \Big(\mathfrak{A}(\mathbf{a}^{rs})\Big)^{\frac{1}{rs}} = \left(\frac{\sum\limits_{i=1}^{n} p_i a_i^{rs}}{\sum\limits_{i=1}^{n} p_i}\right)^{\frac{1}{rs}} \tag{107}$$

$$\mathfrak{H}(\mathbf{a}) = \mathfrak{M}_{-1}(\mathbf{a}) = \frac{\sum\limits_{i=1}^{n} p_i}{\sum\limits_{i=1}^{n} \frac{p_i}{a_i}} = \left(\frac{\sum\limits_{i=1}^{n} p_i a_i^{-1}}{\sum\limits_{i=1}^{n} p_i}\right)^{-1} \tag{108}$$

$$\mathfrak{G}(\mathbf{a}) = \mathfrak{M}_0(\mathbf{a}) = e^{\mathfrak{A}(\ln \mathbf{a})} = e^{\frac{1}{n}\sum\limits_{i=1}^{n} \ln a_i} \text{ or } = e^{\frac{1}{\sum p_i}\sum\limits_{i=1}^{n} p_i \ln a_i} = \left(\prod_{i=1}^{n} a_i^{p_i}\right)^{\frac{1}{\sum p_i}} \tag{109}$$

The geometric mean defined by (17) or (109). The homogeneity of an inequality in certain sets of variables often enables us to simplify our proofs by imposing an additional restriction (a normalisation) on them. Thus the means $M_r(\mathbf{a})$ of (105) are homogeneous, of degree 0, in the weights p_i, and we may always suppose, that $\sum p_i = 1$ prove that

$$(a_1^s + a_2^s + \cdots + a_n^s)^{\frac{1}{s}} \leq (a_1^r + a_2^r + \cdots + a_n^r)^{\frac{1}{r}} \tag{110}$$

when $0 < r < s$ (Theorem 19), we may suppose (since both sides are homogeneous in the $\mathbf{a}$ of degree 1) that $\sum_i \mathbf{a}_i^r = 1$. We have then

$$a_i^r \leq 1,\ a_i^s = (a_i^r)^{\frac{s}{r}} \leq a_i^r \tag{111}$$

and so $\sum_i a_i^s \leq \sum_i a_i^r = 1$. Without this preliminary normalisation, our proof would run

$$\frac{(\sum_i a_i^s)^{1/s}}{(\sum_i a_i^r)^{1/r}} = \left(\sum_i \frac{a_i^s}{(\sum_i a_i^r)^{s/r}}\right)^{1/s} = \left(\sum_i \left(\frac{a_i^r}{\sum_i a_i^r}\right)^{s/r}\right)^{1/s} \leq \left(\sum_i \frac{a_i^r}{\sum_i a_i^r}\right)^{1/s} = 1 \tag{112}$$

Therefore,

$$\left(\sum_i a_i^s\right)^{\frac{1}{s}} \leq \left(\sum_i a_i^r\right)^{\frac{1}{r}} \text{ for } 0 < r < s. \tag{113}$$

It might be instructing to know for $r > 0$, we can prove that

$$\mathfrak{M}_r(\mathbf{a}) = \left(\frac{\sum\limits_{i=1}^{n} p_i a_i^r}{\sum\limits_{i=1}^{n} p_i} \right)^{\frac{1}{r}} \le \left(\frac{\sum\limits_{i=1}^{n} p_i a_i^{2r}}{\sum\limits_{i=1}^{n} p_i} \right)^{\frac{1}{2r}} = \mathfrak{M}_{2r}(\mathbf{a}) \tag{114}$$

by clearing the power $1/(2r)$,

$$\left(\frac{\sum\limits_{i=1}^{n} p_i a_i^r}{\sum\limits_{i=1}^{n} p_i} \right)^{2} \le \frac{\sum\limits_{i=1}^{n} p_i a_i^{2r}}{\sum\limits_{i=1}^{n} p_i} \tag{115}$$

and continue to the denominators

$$\left(\sum_{i=1}^{n} p_i a_i^r \right)^2 \le \left(\sum_{i=1}^{n} p_i \right) \sum_{i=1}^{n} p_i a_i^{2r} \tag{116}$$

or in expanded form

$$(p_1 a_1^r + p_2 a_2^r + \cdots + p_n a_n^r)^2 \le (p_1 + p_2 + \cdots + p_n)\big(p_1 a_1^{2r} + p_2 a_2^{2r} + \cdots + p_n a_n^{2r}\big)\,. \tag{117}$$

If the weights are evenly distributed, i.e., $p_i = 1/n$ and $\sum p_i = 1$, then

$$\left(\sum_{i=1}^{n} \frac{1}{n} a_i^r \right)^2 \le \sum_{i=1}^{n} \frac{1}{n} a_i^{2r} \tag{118}$$

by clearing out $1/n$ on the left

$$\left(\sum_{i=1}^{n} a_i^r \right)^2 \le n \sum_{i=1}^{n} a_i^{2r}. \tag{119}$$

We can see (116) is a special form of C-S inequality by the inner product of sequences $\{\sqrt{p_i}\}$ and $\{\sqrt{p_i a_i^r}\}$.

In general, for $0 < r < s$, we have the **Power Mean Inequality** $M_r(\mathbf{a}) \le M_s(\mathbf{a})$, following the definition of Generalized mean in (105), also (13).

The next is to prove it.

Suppose first we write $r = s\alpha$, so that $0 < \alpha < 1$, and

$$p_i a_i^s = u_i,\ p_i = v_i, \quad \text{for } i = 1, \cdots, n \tag{120}$$

so that $v > 0$ and

$$p_i a_i^r = p_i a_i^{s\alpha} = (p_i a_i^s)^\alpha p_i^{1-\alpha} = u_i^\alpha v_i^{1-\alpha} \tag{121}$$

Then by Hölder's inequality (291) or (293), we have

$$\sum_{i=1}^{n} u_i^\alpha v_i^{1-\alpha} \le \left(\sum_{i=1}^{n} u_i \right)^\alpha \left(\sum_{i=1}^{n} v_i \right)^{1-\alpha} \tag{122}$$

and the equality applies only if u_i/v_i, is independent of i, i.e., a_i is independent of i. Hence

$$\sum_{i=1}^{n}(p_i a_i^s)^\alpha p_i^{1-\alpha} \leq \left(\sum_{i=1}^{n} p_i a_i^s\right)^\alpha \left(\sum_{i=1}^{n} p_i\right)^{1-\alpha} \tag{123}$$

$$\frac{\sum_{i=1}^{n}(p_i a_i^r)}{\sum_{i=1}^{n} p_i} \leq \left(\frac{\sum_{i=1}^{n} p_i a_i^s}{\sum_{i=1}^{n} p_i}\right)^\alpha \tag{124}$$

Finally the power of r-th root

$$\left(\frac{\sum_{i=1}^{n}(p_i a_i^r)}{\sum_{i=1}^{n} p_i}\right)^{\frac{1}{r}} \leq \left(\frac{\sum_{i=1}^{n} p_i a_i^s}{\sum_{i=1}^{n} p_i}\right)^{\frac{\alpha}{r}} = \left(\frac{\sum_{i=1}^{n} p_i a_i^s}{\sum_{i=1}^{n} p_i}\right)^{\frac{1}{s}} \tag{125}$$

This is exactly what we need to prove: $M_r(\mathbf{a}) \leq M_s(\mathbf{a})$ for $0 < r < s$. We shall call this the **Power Mean Inequality**. Another way to prove is to use the concave feature of the function $f(x) = x^\alpha$ for all $x > 0$ by Jensen's Inequality.

$$\frac{\sum_{i=1}^{n} p_i f\left(a_i^s\right)}{\sum_{i=1}^{n} p_i} = \frac{\sum_{i=1}^{n} p_i a_i^r}{\sum_{i=1}^{n} p_i} \leq f\left(\frac{\sum_{i=1}^{n} p_i a_i^s}{\sum_{i=1}^{n} p_i}\right) = \left(\frac{\sum_{i=1}^{n} p_i a_i^s}{\sum_{i=1}^{n} p_i}\right)^\alpha \blacksquare.$$

To arrive at (125), the same procedure can be followed as above, i.e., taking the power of r-th root of both sides.

If the weights are evenly distributed, i.e., $p_i = 1/n$ and $\sum p_i = 1$, then

$$\left(\frac{1}{n}\sum_{i=1}^{n} a_i^r\right)^{\frac{1}{r}} \leq \left(\frac{1}{n}\sum_{i=1}^{n} a_i^s\right)^{\frac{1}{s}} \tag{126}$$

Further, let $r = 1$ (implying $s \geq 1$) and define two means for two sequences $\mathbf{a}$ and $\mathbf{b}$

$$\mu_a = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{127}$$

$$\mu_b = \frac{1}{n}\sum_{i=1}^{n} b_i \tag{128}$$

Then it is found that

$$\sum_{i=1}^{n}\left(\frac{a_i}{\mu_a}\right)^s \geq n \tag{129}$$

and that

$$\sum_{i=1}^{n}\frac{\left(\frac{a_i}{\mu_a}\right)^s}{\frac{b_i}{\mu_b}} \geq n \tag{130}$$

where the later can be proved by Holder's inequality through a lot of manipulations.

and move factor n to the right

$$\left(\sum_{i=1}^{n} a_i^r\right)^{\frac{1}{r}} \leq n^{\left(\frac{1}{r}-\frac{1}{s}\right)}\left(\sum_{i=1}^{n} a_i^s\right)^{\frac{1}{s}} \tag{131}$$

or

$$\left(\sum_{i=1}^{n} a_i^r\right)^s \le n^{(s-r)} \left(\sum_{i=1}^{n} a_i^s\right)^r \tag{132}$$

The definition of geometric mean in (109) implies, that if further all $\mathbf{a}$ are positive and $r = 0 < s$, then

$$(\mathfrak{M}_0(\mathbf{a}))^s = (\mathfrak{G}(\mathbf{a}))^s = \mathfrak{G}(\mathbf{a}^s) \le \mathfrak{A}(\mathbf{a}^s) = (\mathfrak{M}_s(\mathbf{a}))^s . \tag{133}$$

The Power Mean L_r norm for positive a and b, can have weights introduced to the sum such that $p + q = 1$.

i.) If $1 < r$ implying $0 < 1 - 1/r < 1$ we have

$$pa + qb = (pa^r)^{1/r} p^{1-1/r} + (qb^r)^{1/r} q^{1-1/r} \tag{134}$$

$$\le (pa^r + qb^r)^{1/r} (p+q)^{1-1/r} = (pa^r + qb^r)^{1/r} \tag{135}$$

therefore,

$$(pa + qb)^r \le pa^r + qb^r. \tag{136}$$

ii.) If $0 < r < 1$,

$$pa^r + qb^r \le (pa + qb)^r \le (pa)^r + (qb)^r. \tag{137}$$

Lower bound for $0 < r < 1$ is just the reversed inequality direction of the above case for $1 < r$ and can be readily verified

$$\frac{pa^r + qb^r}{(pa + qb)^r} = p\left(\frac{a}{pa + qb}\right)^r + q\left(\frac{b}{pa + qb}\right)^r \le p\frac{a}{pa + qb} + q\frac{b}{pa + qb} = 1. \tag{138}$$

The upper bound can be proved by substituting $x = pa, y = qb$ for simple writing, then

$$\frac{x^r + y^r}{(x+y)^r} = \left(\frac{x}{x+y}\right)^r + \left(\frac{y}{x+y}\right)^r \ge \frac{x}{x+y} + \frac{y}{x+y} = 1. \tag{139}$$

Actually, the map $x \mapsto x^r$ for $x \ge 0$ and $r > 1$ is convex, since its second derivative is $r(r-1)x^{r-2} \ge 0$. We have

$$|pa + qb|^r \le (p|a| + q|b|)^r \le p|a|^r + q|b|^r, \quad \text{if } r > 1. \tag{140}$$

For evenly weighted case and $r > 1$, we have

$$\frac{a+b}{2} = \left(\frac{a^r}{2}\right)^{\frac{1}{r}} \left(\frac{1}{2}\right)^{1-\frac{1}{r}} + \left(\frac{b^r}{2}\right)^{\frac{1}{r}} \left(\frac{1}{2}\right)^{1-\frac{1}{r}} \tag{141}$$

$$\le \left(\frac{a^r + b^r}{2}\right)^{\frac{1}{r}} \left(\frac{1}{2} + \frac{1}{2}\right)^{1-\frac{1}{r}} = \left(\frac{a^r + b^r}{2}\right)^{\frac{1}{r}} \tag{142}$$

If we let $r = 2$ and $s = 3$, one important inequality can be obtained

$$\sqrt{\frac{a^2+b^2+c^2}{3}} \leq \sqrt[3]{\frac{a^3+b^3+c^3}{3}} \tag{143}$$

or

$$\left(a^2+b^2+c^2\right)^3 \leq 3\left(a^3+b^3+c^3\right)^2. \tag{144}$$

Products of Averages and Averages of Products

Suppose that $p_i \geq 0$ for all $i = 1, 2, \cdots, n$ and $\sum p_i = 1$. Show that if a_i and b_i are nonnegative real numbers that satisfy the termwise bound $1 \leq a_i b_i$ for all $i = 1, 2, \cdots, n$, then one also has the aggregate bound for the averages,

$$1 \leq \sum_{i=1}^{n} p_i a_i b_i \leq \sum_{i=1}^{n} p_i \sqrt{a_i b_i} \leq \left\{\sum_{i=1}^{n} p_i a_i\right\}\left\{\sum_{i=1}^{n} p_i b_i\right\}. \tag{145}$$

The cases in which $r = 0$ and an a_i is zero are trivial and we may ignore them. If every a_i is positive, and $r = 0 < s$, we have

10.1 Lagrangian Identity

$$\begin{aligned}
\left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right) - \left(\sum_{k=1}^{n} a_k b_k\right)^2 &= \sum_{1\leq i<j\leq n} \left(a_i b_j - a_j b_i\right)^2 && (146)\\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \left(a_i b_j - a_j b_i\right)^2 && (147)\\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \left(a_i b_j - a_j b_i\right)^2 && (148)
\end{aligned}$$

To put in the vector form,

$$\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - (\mathbf{a}\cdot\mathbf{b})^2 = |\mathbf{a}\times\mathbf{b}|^2 \tag{149}$$

Lagrangian identity might be easily obtained when we study the scalar product of two cross products $(\mathbf{a}\times\mathbf{b})$ and $(\mathbf{c}\times\mathbf{d})$ in vector analysis. It can be considered as the triple scalar product of $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{u}$ $(= \mathbf{c}\times\mathbf{d})$, which has the invariant property under any cyclic permutation of its three vectors.

$$\begin{aligned}
(\mathbf{a}\times\mathbf{b})\cdot\mathbf{u} &= \mathbf{a}\cdot(\mathbf{b}\times\mathbf{u})\\
&= \mathbf{a}\cdot[\mathbf{b}\times(\mathbf{c}\times\mathbf{d})]\\
&= \mathbf{a}\cdot[(\mathbf{b}\cdot\mathbf{d})c - (\mathbf{b}\cdot\mathbf{c})\mathbf{d})]. && (150)
\end{aligned}$$

Therefore, we have

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}). \tag{151}$$

In above, if we change $\mathbf{u} = \mathbf{a} \times \mathbf{b}$, (151) becomes the Lagrangian identity, i.e.,

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2. \tag{152}$$

10.2 Bernoulli's Inequalities

For any real number $x \geq -1$, we have

$$(1+x)^r \geq 1 + rx \text{ for } r \geq 1 \text{ or } r \leq 0 \text{ (strict if } x \neq 0 \ \&\ r \neq 0,\ 1.) \tag{153}$$

$$(1+x)^r \leq 1 + rx \text{ for } 0 \leq r \leq 1. \tag{154}$$

To generalize, if $a_1, a_2, \cdots, a_n$ are all reals such that either all are positive $a_i \geq 0$ or negative but within $a_i \in [-1, 0]$, then

$$(1+a_1)(1+a_2)\cdots(1+a_n) \geq 1 + a_1 + a_2 + \cdots + a_n \tag{155}$$

Bernoulli's inequality is a special case when $a_1 = a_2 = \cdots = a_n = x$. This generalized inequality can be proved by mathematical induction.

10.3 Schur Inequality

Let a, b, c be positive real numbers (in math notation $\forall a, b, c \in \mathbb{R}^+$), and n be integer. Then the following inequality holds:

$$a^n(a-b)(a-c) + b^n(b-c)(b-a) + c^n(c-a)(c-b) \geq 0 \tag{156}$$

with equality if and only if $a = b = c$ or $a = b$, $c = 0$ and permutations.

If $n = 0$, (156) is equivalent to the well-known

$$ab + bc + ca \leq a^2 + b^2 + c^2 \tag{157}$$

which actually holds $\forall a, b, c \in \mathbb{R}$.

If $n = 1$

$$ab(a+b) + bc(b+c) + ca(c+a) \leq a^3 + b^3 + c^3 + 3abc \quad \text{or} \tag{158}$$

$$(a+b-c)(b+c-a)(c+a-b) \leq abc \tag{159}$$

$$2a2b2c = 8abc \leq (a+b)(b+c)(c+a) \tag{160}$$

$$a^3(b+c) + b^3(c+a) + c^3(a+b) \leq \frac{2}{3}(a^2+b^2+c^2)^2 \tag{161}$$

where the following identities are often helpful.

$$\begin{aligned}(a+b+c)^3 &= a^3 + b^3 + c^3 + 3(a+b)(b+c)(c+a) \qquad (162)\\ &= a^3 + b^3 + c^3 + 3\big(ab(a+b)+bc(b+c)+ca(c+a)\big) + 6abc\end{aligned}$$

$$a^3 + b^3 + c^3 - 3abc = (a+b+c)\left(a^2 + b^2 + c^2 - ab - bc - ca\right) \tag{163}$$

$$(ab+bc+ca)^2 \leq 3\big(a^2b^2 + b^2c^2 + c^2a^2\big) \leq \big(a^2+b^2+c^2\big)^2 \leq 3\big(a^4+b^4+c^4\big) \tag{164}$$

10.4 Cauchy-Schwarz Inequality

The **C**auchy-**S**chwarz Inequality states that the product of sums of squares is no less than the square of a sum of products of a_i and $b_i \in \mathbb{R}$.

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2. \tag{165}$$

The equality holds if and only if $a_i = kb_i$, $i = 1 \ldots n$ for non-zero $k \in \mathbb{R}$.

In vector notations, if we define $\mathbf{u} = \{a_1, \ldots, a_n\}$ and $\mathbf{v} = \{b_1, \ldots, b_n\}$ then

$$|\langle \mathbf{u}, \mathbf{v}\rangle|^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle \cdot \langle \mathbf{v}, \mathbf{v}\rangle \tag{166}$$

The equality is taken if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

Another variant is obtained by substituting a_i with $\sqrt{a_i}$, b_i with $\sqrt{b_i}$ in (165) and both sides taken square roots.

$$\sum_{i=1}^{n} \sqrt{a_i b_i} \leq \sqrt{\sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i} \tag{167}$$

or explicitly

$$\sqrt{a_1 b_1} + \sqrt{a_2 b_2} + \ldots + \sqrt{a_n b_n} \leq \sqrt{a_1 + a_2 + \ldots + a_n}\sqrt{b_1 + b_2 + \ldots + b_n} \tag{168}$$

To directly prove (167), we can set back for first $a_i = c_i^2$ and $b_i = d_i^2$. By induction or the standard form of C-S in (165), assume we know that:

$$\left(\sum_{i=1}^{n} c_i d_i\right)^2 \leq \left(\sum_{i=1}^{n} c_i^2\right)\left(\sum_{i=1}^{n} d_i^2\right) \tag{169}$$

and we need to proceed $n+1$ step, i.e., to prove that:

$$\left(\sum_{i=1}^{n+1} c_i d_i\right)^2 \leq \left(\sum_{i=1}^{n+1} c_i^2\right)\left(\sum_{i=1}^{n+1} d_i^2\right). \tag{170}$$

We have:

$$\left(\sum_{i=1}^{n+1} c_i d_i\right)^2 = \left(\sum_{i=1}^{n} c_i d_i\right)^2 + (c_{n+1}d_{n+1})^2 + 2c_{n+1}d_{n+1}\left(\sum_{i=1}^{n} c_i d_i\right) \tag{171}$$

$$\sum_{i=1}^{n+1} c_i^2 \sum_{i=1}^{n+1} d_i^2 = \sum_{i=1}^{n} c_i^2 \sum_{i=1}^{n} d_i^2 + c_{n+1}^2 d_{n+1}^2 + c_{n+1}^2 \sum_{i=1}^{n} d_i^2 + d_{n+1}^2 \sum_{i=1}^{n} c_i^2 \tag{172}$$

hence in order to prove (171)≤(172), we just need to show that

$$2c_{n+1}d_{n+1}\left(\sum_{i=1}^{n} c_i d_i\right) \leq c_{n+1}^2 \sum_{i=1}^{n} d_i^2 + d_{n+1}^2 \sum_{i=1}^{n} c_i^2. \tag{173}$$

Since

$$c_{n+1}^2\sum_{i=1}^n d_i^2 - 2c_{n+1}d_{n+1}\sum_{i=1}^n c_i d_i + d_{n+1}^2\sum_{i=1}^n c_i^2 \tag{174}$$

$$= \left(c_{n+1}\sqrt{\sum_{i=1}^n d_i^2} - d_{n+1}\sqrt{\sum_{i=1}^n c_i^2}\right)^2 + 2c_{n+1}d_{n+1}\left(\sqrt{\sum_{i=1}^n d_i^2\sum_{i=1}^n c_i^2} - \sum_{i=1}^n c_i d_i\right)$$

$$\geq 0 \tag{175}$$

where (169) is used for the last brackets or (174) can be shown to have a non-positive discriminant of a bivariate polynomial[5] in c_{n+1} and d_{n+1}.

Remarks:
Mnemonic trick can be "**SSP** less than **PSS**", i.e., the **S**quare of the **S**um of **P**roducts $\leq$ the **P**roduct of the **S**um of **S**quares. Holder's inequality is the natural extension of this idea: for any positive integer n, the n-th power of the sum of products (of n non-negative sequences) is bounded by the product of the sum of the n-th powers.

10.4.1 Titu's inequality - Cauchy-Schwarz in Engel Form

The following inequality is known as Sedrakyan's inequality, Bergström's inequality, Engel's form or Titu's lemma, respectively, but due to its numerous applications won a designation of its own. It was referred to in the classic book Inequalities by E. F. Beckenbach and R. Bellman, (Springer, 1961). In recent years it became better known as Titu's Inequality in honor of Titu Andreescu. For any reals $a_1, a_2, a_3, \ldots, a_n$ and positive reals $b_1, b_2, b_3, \ldots, b_n$ we have

$$\frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} + \cdots + \frac{a_n^2}{b_n} \geq \frac{(a_1 + a_2 + \cdots + a_n)^2}{b_1 + b_2 + \cdots + b_n}. \tag{176}$$

with equality if and only if $\frac{a_1^2}{b_1} = \frac{a_2^2}{b_2} = \cdots = \frac{a_n^2}{b_n}$.
It is a direct consequence of the Cauchy–Schwarz inequality, obtained by substituting $a_i' = a_i/\sqrt{b_i}$ and $b_i' = \sqrt{b_i}$. This form is especially helpful when the inequality involves fractions where the numerator of the addend is a perfect square. Note that if the sequence of **b** is a

[5] Let $f(x, y) = ax^2 + bxy + cy^2$ be a bivariate polynomial or binary quadratic form and let $\Delta = b^2 - 4ac$ be the discriminant of f.

1. If Δ is a perfect square ($\Delta = 0, 1, 4, 9, \ldots$) then $f(x, y)$ can be factored as a product of two linear factors. In particular, if $\Delta = 0$ then $f(x, y)$ is the square of a linear factor and if $\Delta = 0, 1, 4, 9, \ldots$ then $f(x, y)$ is the product of two distinct linear factors.
2. If Δ is not a perfect square then $f(x, y)$ cannot be factored as a product of two linear factors with integer coefficients.
3. If Δ is not a perfect square then the only integer solution to $f(x, y) = 0$ is the solution $(x, y) = (0, 0)$.

permutation of **a**, then

$$\frac{a_1^2}{b_1}+\frac{a_2^2}{b_2}+\cdots+\frac{a_n^2}{b_n}\geq a_1+a_2+\cdots+a_n. \tag{177}$$

Ex 10.1 (The Simplest Example) . *Use C-S in Engel form to prove this.*

$$\frac{1}{a}+\frac{1}{b}\geq\frac{4}{a+b}. \tag{178}$$

Solution: Since the constants 1 in the nominators can be written in squared, i.e., 1 degree higher than the denominators a and b, Engel form C-S applies.

$$\frac{1^2}{a}+\frac{1^2}{b}\geq\frac{(1+1)^2}{a+b}. \tag{179}$$

Ex 10.2 (The Product of Square Root) . *Let a, b be positive real numbers, show that*

$$\sqrt{1+a^2}\sqrt{1+b^2}\geq a+b. \tag{180}$$

Ex 10.3 (The Sum of Square Example) . *Let a, b, c, d be positive real numbers such that $a^2+b^2+c^2+d^2=4$, show that*

$$\frac{a^2}{b}+\frac{b^2}{c}+\frac{c^2}{d}+\frac{d^2}{a}\geq 4. \tag{181}$$

Solution: From Cauchy-Schwarz:

$$\left(\frac{a^2}{b}+\frac{b^2}{c}+\frac{c^2}{d}+\frac{d^2}{a}\right)\left(a^2b+b^2c+c^2d+c^2a\right)\geq(a^2+b^2+c^2+d^2)^2=16 \tag{182}$$

therefore it suffices to show that

$$a^2b+b^2c+c^2d+c^2a\leq 4 \tag{183}$$

Using Cauchy-Schwarz again:

$$\begin{aligned} a^2b+b^2c+c^2d+c^2a &\leq\sqrt{(a^2+b^2+c^2+d^2)(a^2b^2+b^2c^2+c^2d^2+d^2a^2)} \\ &=2\sqrt{(a^2+c^2)(b^2+d^2)} \\ &\leq(a^2+c^2)+(b^2+d^2)=4 \end{aligned} \tag{184}$$

with AM-GM in the last step.

10.4.2 Weight-Power Summation Inequalities

The Titu's inequality (or in short Engel Form C-S, see 10.4.1) is actually the special case of the following generic Weight-Power Summation Inequalities for positive $a_i, b_i,\ i\in\{1,\cdots,n\}$ and $m=1$. The inequality below has been discovered by J. Radon in 1913 but is often referred to, along with Bergström's inequality (of which it is a generalization) as Titu's Lemma or Titu's inequality in honor of Titu Andreescu after the publication of Mathematical Olympiad Treasures in 2003.

i.) $m < -1$ or $m > 0$

$$\sum_{i=1}^{n} \frac{a_i^{m+1}}{b_i^m} \geq \frac{\left(\sum_{i=1}^{n} a_i\right)^{m+1}}{\left(\sum_{i=1}^{n} b_i\right)^m},\ m < -1 \text{ or } m > 0 \tag{185}$$

ii.) $-1 \leq m \leq 0$

$$\sum_{i=1}^{n} \frac{a_i^{m+1}}{b_i^m} \leq \frac{\left(\sum_{i=1}^{n} a_i\right)^{m+1}}{\left(\sum_{i=1}^{n} b_i\right)^m},\ -1 \leq m \leq 0 \tag{186}$$

Note that the inequalities have the feature where the nominators have one degree higher on the power indices than the denominators. For the second case $-1 \leq m \leq 0$, if we let $p = m + 1$, then since both $p, 1 - p = -m \in [0, 1]$, (186) becomes Hölder's Inequality

$$\sum_{i=1}^{n} a_i^p b_i^{1-p} \leq \left(\sum_{i=1}^{n} a_i\right)^p \left(\sum_{i=1}^{n} b_i\right)^{1-p},\ 0 \leq p \leq 1 \tag{187}$$

Let's give two alternative proofs of the above Weight-Power Summation inequalities (185) and (186).

First Proof: Let normalization parameters be

$$s = \frac{1}{a_1 + a_2 + \ldots + a_n},\ t = \frac{1}{b_1 + b_2 + \ldots + b_n} \tag{188}$$

such that $\sum\limits_{i=1}^{n} sa_i = 1$ and $\sum\limits_{i=1}^{n} tb_i = 1$. With the help of the Bernoulli's inequalities in (10.2), we might be able to prove our Weight-Power Summation inequalities. At least, we observed that our inequalities change the inequality directions exactly in the same way as the conditional Bernoulli's inequalities.

i.) $m < -1$ or $m > 0$

$$\sum_{i=1}^{n} \frac{(sa_i)^{m+1}}{(tb_i)^m} = \sum_{i=1}^{n} tb_i \left(\frac{sa_i}{tb_i}\right)^{m+1} \geq \sum_{i=1}^{n} tb_i \left[1 + (m+1)\left(\frac{sa_i}{tb_i} - 1\right)\right] \tag{189}$$

$$= \sum_{i=1}^{n} tb_i [(m+1)\frac{sa_i}{tb_i} - m] = \sum_{i=1}^{n} (m+1)sa_i - mtb_i = m+1-m = 1 \tag{190}$$

where (153) is used since the power $m + 1 > 1$ or $m < 0$, therefore,

$$\sum_{i=1}^{n} \frac{a_i^{m+1}}{b_i^m} \geq \frac{t^m}{s^{m+1}} = \frac{\left(\sum\limits_{i=1}^{n} a_i\right)^{m+1}}{\left(\sum\limits_{i=1}^{n} b_i\right)^m} \tag{191}$$

ii.) $-1 \le m \le 0$

$$\sum_{i=1}^{n}\frac{(sa_i)^{m+1}}{(tb_i)^m}=\sum_{i=1}^{n}tb_i\left(\frac{sa_i}{tb_i}\right)^{m+1}\le\sum_{i=1}^{n}tb_i\left[1+(m+1)\left(\frac{sa_i}{tb_i}-1\right)\right] \tag{192}$$

$$=\sum_{i=1}^{n}tb_i[(m+1)\frac{sa_i}{tb_i}-m]=\sum_{i=1}^{n}(m+1)sa_i-mtb_i=m+1-m=1 \tag{193}$$

where (154) is used since the power $0 \le m+1 \le 1$, so,

$$\sum_{i=1}^{n}\frac{a_i^{m+1}}{b_i^m}\le\frac{t^m}{s^{m+1}}=\frac{\left(\sum_{i=1}^{n}a_i\right)^{m+1}}{\left(\sum_{i=1}^{n}b_i\right)^m} \tag{194}$$

Second Proof: The second condition with $-1 \le m \le 0$ will be proved. Note that if the special cases $m=-1$ or $m=0$, the equality will plainly hold, so only $m \in (-1,0)$ shall be investigated. We still make use of (188), but instead of Bernoulli's inequality, we use Jensen's inequality (261) for a function $f(x)=x^{m+1}$ which is concave since $m+1 \in (0,1)$. Comparing with the formulation in (261) and observing that the weights $p_i = tb_i$ and independent variables $x_i = sa_i/(tb_i)$. If we sum up the products of weight p_i and the function's argument x_i, we will obtain 1, i.e., $\sum_{i=1}^{n} p_i x_i = 1$,

$$\sum_{i=1}^{n}\frac{(sa_i)^{m+1}}{(tb_i)^m}=\sum_{i=1}^{n}tb_i\left(\frac{sa_i}{tb_i}\right)^{m+1}\le\left(\sum_{i=1}^{n}tb_i\frac{sa_i}{tb_i}\right)^{m+1}=\left(\sum_{i=1}^{n}sa_i\right)^{m+1}=1 \tag{195}$$

Finally, the same result is obtained as (193). Similar proof for a convex power function x^{m+1} with $m<-1$ or $m>0$ can be used to prove (185).

Remarks: Above proofs used n-dimensional sequences for $i=1,\cdots,n$, actually more than two dimensions are unnecessary. Because if we can prove the case for $-1 \le m \le 0$ and $i=1,2$,

$$\frac{a_1^{m+1}}{b_1^m}+\frac{a_2^{m+1}}{b_2^m}\le\frac{(a_1+a_2)^{m+1}}{(b_1+b_2)^m}, \tag{196}$$

then it is straight-forward to repeat this for $i=3$ term, i.e.,

$$\frac{a_1^{m+1}}{b_1^m}+\frac{a_2^{m+1}}{b_2^m}+\frac{a_3^{m+1}}{b_3^m}\le\frac{(a_1+a_2)^{m+1}}{(b_1+b_2)^m}+\frac{a_3^{m+1}}{b_3^m}\le\frac{(a_1+a_2+a_3)^{m+1}}{(b_1+b_2+b_3)^m}. \tag{197}$$

This can just continue to complete the proof for all a_i, b_i until $i=n$. The advantage of this trick is that a traditional inductive method can be employed to only prove (196) if m is an integer.

In addition, the C-S inequality can be easily extended to

$$\sqrt{\sum_{k=1}^{n}a_kb_k}\le\sum_{k=1}^{n}\sqrt{a_kb_k}\le\sqrt{\left(\sum_{i=1}^{n}a_i\right)\left(\sum_{j=1}^{n}b_j\right)}. \tag{198}$$

Ex 10.4 (IMO 2001) *Let a, b, c be positive real numbers. Prove that*

$$\frac{a}{\sqrt{a^2+8bc}}+\frac{b}{\sqrt{b^2+8ac}}+\frac{c}{\sqrt{c^2+8ab}}\geq 1. \tag{199}$$

Solution: To use the Weight-Power Summation Inequality (185), the nominator needs to have one higher degree than the square root power which is 1/2. Clearly here, we need some manipulation for the nominators to have the degree 1+1/2 by multiplying both nominator and denominator of each term by $\sqrt{a}$, $\sqrt{b}$ and $\sqrt{c}$, respectively. So we have now $m = 1/2$ according to (185).

$$\frac{a^{\frac{3}{2}}}{\sqrt{a^3+8abc}}+\frac{b^{\frac{3}{2}}}{\sqrt{b^3+8cab}}+\frac{c^{\frac{3}{2}}}{\sqrt{c^3+8abc}}\geq\frac{(a+b+c)^{\frac{3}{2}}}{\sqrt{a^3+b^3+c^3+24abc}} \tag{200}$$

It is left to show $(a+b+c)^3 \geq a^3+b^3+c^3+24abc$ which is already given by (162) and (160). Also see (303) for a different approach to find the solution.

There is another upper bound of the inequality (199)

$$\frac{a^2+b^2+c^2}{ab+bc+ca}\geq\frac{a}{\sqrt{a^2+8bc}}+\frac{b}{\sqrt{b^2+8ac}}+\frac{c}{\sqrt{c^2+8ab}}\geq 1. \tag{201}$$

10.4.3 Generalized Weight-Power Summation Inequality

For non-negatives a_i and positive b_i, the following generalized Weight-Power Summation Inequality can be formulated.

$$\frac{\left(\sum\limits_{i=1}^{n}a_i\right)^{\alpha}}{\sum\limits_{i=1}^{n}b_i}\leq n^{\alpha-2}\sum_{i=1}^{n}\frac{a_i^{\alpha}}{b_i} \tag{202}$$

Starting from LHS, let's clear out the denominator (the summation on b_i) and the power index by multiplying $\sum_{i=1}^{n} b_i$ and taking the inverse power of α, respectively,

$$\sum_{i=1}^{n}a_i\leq n^{\left(1-\frac{2}{\alpha}\right)}\left(\sum_{i=1}^{n}\frac{a_i^{\alpha}}{b_i}\right)^{\frac{1}{\alpha}}\left(\sum_{i=1}^{n}b_i\right)^{\frac{1}{\alpha}} \tag{203}$$

On RHS, the indices of 3-term product sums up to 1 reminds us for the famous Holder's inequality, if rewriting n as a unit summation.

$$\sum_{i=1}^{n}1^{\left(1-\frac{2}{\alpha}\right)}\left(\frac{a_i^{\alpha}}{b_i}\right)^{\frac{1}{\alpha}}b_i^{\frac{1}{\alpha}}\leq\left(\sum_{i=1}^{n}1\right)^{\left(1-\frac{2}{\alpha}\right)}\left(\sum_{i=1}^{n}\frac{a_i^{\alpha}}{b_i}\right)^{\frac{1}{\alpha}}\left(\sum_{i=1}^{n}b_i\right)^{\frac{1}{\alpha}}\blacksquare \tag{204}$$

The other variant derivation can also be configured.

$$n^{\alpha-2}\sum_{i=1}^{n}\frac{a_i^{\alpha}}{b_i}=\frac{\left(\sum\limits_{i=1}^{n}1\right)^{\alpha-2}\sum\limits_{i=1}^{n}\frac{a_i^{\alpha}}{b_i}\sum\limits_{i=1}^{n}b_i}{\sum\limits_{i=1}^{n}b_i}\geq \tag{205}$$

$$\geq\frac{\left(\sum\limits_{i=1}^{n}\left(1^{\alpha-2}\cdot\frac{a_i^{\alpha}}{b_i}\cdot b_i\right)^{\frac{1}{\alpha-2+1+1}}\right)^{\alpha-2+1+1}}{\sum\limits_{i=1}^{n}b_i}=\frac{\left(\sum\limits_{i=1}^{n}a_i\right)^{\alpha}}{\sum\limits_{i=1}^{n}b_i}. \tag{206}$$

For example, for positives a, b and c we have:

$$\sum_{cyc}\frac{a}{b+c}=\sum_{cyc}\frac{a^3}{a^2(b+c)}\geq\frac{(a+b+c)^3}{3\sum_{cyc}(a^2b+a^2c)}. \tag{207}$$

Remarks:
If we define

$$\mu_a=\frac{1}{n}\sum_{i=1}^{n}a_i \tag{208}$$

$$\mu_b=\frac{1}{n}\sum_{i=1}^{n}b_i \tag{209}$$

$$\sum_{i=1}^{n}\frac{\left(\frac{a_i}{\mu_a}\right)^\alpha}{\frac{b_i}{\mu_b}}\geq n \tag{210}$$

or

$$\sum_{i=1}^{n}\left(\frac{a_i}{\mu_a}\right)^\alpha\geq n\frac{b_i}{\mu_b} \tag{211}$$

10.5 Rearrangement Inequality

Rearrangement Inequality might be the most intuitive and understandable tool when used to mathematic proofs. Many important inequalities can be stunningly proved by the rearrangement inequality, such as the Inequality of arithmetic and geometric means, the Cauchy–Schwarz inequality, and Chebyshev's sum inequality.

One particular consequence is that if $a_1\leq\cdots\leq a_n$ and $b_1\leq\cdots\leq b_n$ then by using $x_i:=a_i$ $y_i:=b_i$ and for all $i\in[1,n]$

$$x_1y_n+\cdots+x_ny_1\leq\ x_1y_{\sigma(1)}+\cdots+x_ny_{\sigma(n)}\leq\ x_1y_1+\cdots+x_ny_n \tag{212}$$

holds for every permutation σ of $1,\cdots,n$.

The proof is fairly handy. Consider a permutation $\sigma(1)$, $\sigma(2)$, $\cdots$, $\sigma(n)$ such that $S=x_1y_{\sigma(1)}+x_2y_{\sigma(2)}+\cdots+x_ny_{\sigma(n)}$ is maximized when the permutation becomes identity permutation, i.e., $\sigma(i)=i$ for all $i=1,\cdots,n$.

$$S=x_1\underset{1}{y_{\sigma(1)}}+x_2\underset{2}{y_{\sigma(2)}}+x_i\underset{i}{y_{\sigma(i)}}+\cdots+x_{n-1}\underset{n-1}{y_{\sigma(n-1)}}+x_n\underset{n}{y_{\sigma(n)}} \tag{213}$$

Suppose that there is at least one permutation appearing at the smallest positive index $\sigma(i)\neq i$ which means that $\sigma(i)=j>i$ (as the positions from 1 to $i-1$ are not changed from the identity permutation). Since i-th positive was taken by j-th. The next is to find who is taking the position where $\sigma(j)=j$ was. If it happens to be the interchanged case i.e.,

$$S\ =x_1\underset{1}{y_{\sigma(1)}}+x_2\underset{2}{y_{\sigma(2)}}+x_i\underset{i}{y_{\sigma(i)}}+\cdots+x_j\underset{j}{y_{\sigma(j)}}+\cdots+x_{n-1}\underset{n-1}{y_{\sigma(n-1)}}+x_n\underset{n}{y_{\sigma(n)}} \tag{214}$$

$$S'=x_1\underset{1}{y_{\sigma(1)}}+x_2\underset{2}{y_{\sigma(2)}}+x_i\underset{j}{y_{\sigma(i)}}+\cdots+x_j\underset{i}{y_{\sigma(j)}}+\cdots+x_{n-1}\underset{n-1}{y_{\sigma(n-1)}}+x_n\underset{n}{y_{\sigma(n)}} \tag{215}$$

now since $\sigma(j) = i < \sigma(i) = j$ in S' indicates that $y_i \le y_j$, the comparison with the identity permutation can result in

$$S' - S = x_i y_j + x_j y_i - (x_i y_i + x_j y_j) = (x_i - x_j)(y_j - y_i) \le 0. \tag{216}$$

(217)

So the interchanged product is no larger than one with the identity permutation. However if it is not i, find that $\sigma(k) = i < k$ as some number must be assigned to i.

Now, we have $i < j$ and $y_i \le y_j$. Similarly, $i < k$ and $x_i \le x_k$. Therefore,

$$0 \le (x_k - x_i)(y_j - y_i) \implies x_i y_j + x_k y_i \le x_k y_j + x_i y_i \tag{218}$$

so the sum $x_1 y_{\sigma(1)} + \cdots + x_n y_{\sigma(n)}$ is no larger than by changing $\sigma(i) = j, \sigma(k) = i$ to $\sigma(i) = i, \sigma(k) = j$.

Ex 10.5 (Nesbitt' s Inequality) *Let a, b, c be positive integers. Prove that,*

$$\frac{3}{2} \le \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \quad [< 2] \tag{219}$$

the lower bound inequality is called the Nesbitt' s Inequality and the upper bound is valid for a, b, c as triangle's sides.

First Proof: Prove this following lemma first:

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \ge \frac{2}{3}(ab+bc+ca)[\frac{1}{(a+b)^2} + \frac{1}{(b+c)^2} + \frac{1}{(c+a)^2} \tag{220}$$

But now, we know that

$$(ab+bc+ca)[\frac{1}{(a+b)^2} + \frac{1}{(b+c)^2} + \frac{1}{(c+a)^2}] \ge \frac{9}{4} \tag{221}$$

is from Iran TST 1996.

$$\begin{aligned}
(x+y+z)&\left(\frac{x}{(x+y)(z+x)} + \frac{y}{(y+z)(x+y)} + \frac{z}{(z+x)(y+z)}\right) \\
&= \frac{2(x+y+z)(xy+yz+zx)}{(x+y)(y+z)(z+x)} \\
&= 2\frac{(x+y)(y+z)(z+x) + xyz}{(x+y)(y+z)(z+x)} \\
&= 2\left(1 + \frac{+xyz}{(x+y)(y+z)(z+x)}\right) \\
&\le 2(1+\frac{1}{8}) = \frac{9}{4}
\end{aligned} \tag{222}$$

Second Proof: Multiply both sides by 2 first and then add 3 ones such that each denominator is summed to its nominator, i.e.,

$$3 \le \frac{2a+b+c}{b+c} + \frac{2b+c+a}{c+a} + \frac{2c+a+b}{a+b} \tag{223}$$

$$6 \le \frac{a+b}{b+c} + \frac{c+a}{b+c} + \frac{b+c}{c+a} + \frac{a+b}{c+a} + \frac{c+a}{a+b} + \frac{c+a}{b+c} + \frac{b+c}{a+b} \tag{224}$$

where the last line is direct from AM-GM rules $x/y + y/x + x/z + z/x + y/z + z/y \geq 6$ by $x = a+b, y = b+c, z = c+a$. Now let's assume that a, b, c are the triangle's sides. Let the half-perimeter $s = (a+b+c)/2$, and we have

$$s < a+b, \quad s < b+c, \quad s < c+a. \tag{225}$$

so

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} < \frac{a}{s} + \frac{b}{s} + \frac{c}{s} = 2. \tag{226}$$

$$\left(\frac{a}{a+b}\right)^2 + \left(\frac{b}{b+c}\right)^2 + \left(\frac{c}{c+d}\right)^2 + \left(\frac{d}{d+a}\right)^2 \geq 1 \tag{227}$$

$$\left(\frac{a}{a+b}\right)^3 + \left(\frac{b}{b+c}\right)^3 + \left(\frac{c}{c+d}\right)^3 + \left(\frac{d}{d+a}\right)^3 \geq \frac{1}{2} \tag{228}$$

$$\left(\frac{a}{a+b}\right)^5 + \left(\frac{b}{b+c}\right)^5 + \left(\frac{c}{c+d}\right)^5 + \left(\frac{d}{d+a}\right)^5 \geq \frac{1}{8} \tag{229}$$

Prove by power means.

Third Proof: Original inequality (219) which is equivalent to

$$\frac{a}{b+c}+1+\frac{b}{c+a}+1+\frac{c}{a+b}+1=(a+b+c)\left(\frac{1}{b+c}+\frac{1}{c+a}+\frac{1}{a+b}\right) \geq \frac{9}{2}. \tag{230}$$

However, we know from HM-AM for the three numbers $b+c$, $c+a$ and $a+b$ that

$$\frac{3}{\frac{1}{b+c}+\frac{1}{c+a}+\frac{1}{a+b}} \leq \frac{b+c+c+a+a+b}{3} = \frac{2}{3}(a+b+c). \tag{231}$$

This is exactly what we need to complete the proof.

Fourth Proof: We can also use rearrangement inequality to prove it.

$$\begin{aligned}
\frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b} &= \frac{1}{2}\left(\frac{a}{b+c}+\frac{b}{a+c}+\frac{c}{a+b}+\frac{a}{b+c}+\frac{b}{a+c}+\frac{c}{a+b}\right) \\
&\geq \frac{1}{2}\left(\frac{a}{a+c}+\frac{b}{a+b}+\frac{c}{b+c}+\frac{a}{a+b}+\frac{b}{b+c}+\frac{c}{a+c}\right) \\
&= \frac{3}{2}
\end{aligned} \tag{232}$$

Fifth Proof: If we denote $s := a+b+c$, we have to minimize

$$\frac{a}{s-a} + \frac{b}{s-b} + \frac{c}{s-c} \tag{233}$$

for $a, b, c > 0$.

Using new variables $x := a/s$, $y := b/s$, $z := c/s$ we can see, that this is equivalent to minimizing

$$\frac{x}{1-x} + \frac{y}{1-y} + \frac{z}{1-z} \tag{234}$$

for $x + y + z = 1$.

If we notice that the function

$$f(x) = \frac{x}{1-x} = -1 + \frac{1}{1-x} = -1 - \frac{1}{x-1} \tag{235}$$

is convex on the interval $(0, 1)$, we have

$$\frac{f(x) + f(y) + f(z)}{3} \geq f\left(\frac{x+y+z}{3}\right), \tag{236}$$

i.e.

$$f(x) + f(y) + f(z) \geq 3f(1/3). \tag{237}$$

By computing $f(1/3) = 1/2$ we see, that the last inequality is precisely

$$\frac{x}{1-x} + \frac{y}{1-y} + \frac{z}{1-z} \geq \frac{3}{2}. \tag{238}$$

The generalization is

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geq \frac{\sqrt{a}}{\sqrt{b}+\sqrt{c}} + \frac{\sqrt{b}}{\sqrt{a}+\sqrt{c}} + \frac{\sqrt{c}}{\sqrt{b}+\sqrt{a}} \tag{239}$$

or

$$\sum_{cyc} \frac{a^2}{b^2+c^2} \geq \sum_{cyc} \frac{a}{b+c}. \tag{240}$$

Ex 10.6 (a, b, c Permutation Inequality) *Let a, b, c be positive integers for which $abc = 1$. Prove that*

$$\frac{3}{2} \leq \sum \frac{1}{b(a+b)}. \tag{241}$$

OFFICIAL SOLUTION: LEMMA: Suppose that $x_1 \geq x_2 \geq x_3, y_1 \leq y_2 \leq y_3$, and $(y_1'; y_2'; y_3')$ is a permutation of $(y_1; y_2; y_3)$, then

$$x_1y_1' + x_2y_2' + x_3y_3' \geq x_1y_1 + x_2y_2 + x_3y_3. \tag{242}$$

Proof: Since $y_1 \leq y_2 \leq y_3$, we get $y_1' \geq y_1$ and $y_1' + y_2' \geq y_1 + y_2$; moreover,

$$y_1' + y_2' + y_3' = y_1 + y_2 + y_3 \tag{243}$$

$$x_1y_1'+x_2y_2'+x_3y_3' = (x_1 - x_2)y_1'+(x_2 - x_3)(y_1'+y_2')+x_3(y_1'+y_2'+y_3') \geq \tag{244}$$

$$(x_1 - x_2)y_1+(x_2 - x_3)(y_1+y_2)+x_3(y_1+y_2+y_3) = x_1y_1+x_2y_2+x_3y_3. \blacksquare \tag{245}$$

Denote by S the left-hand part of the desired inequality. Since S is invariant under the cyclical permutation of variables, we can assume $a \leq b \leq c$ or $a \geq b \geq c$. In both cases, by LEMMA, we get:

$$S = \frac{1}{(a+b)b} + \frac{1}{(b+c)c} + \frac{1}{(c+a)a} \geq \frac{1}{(a+b)c} + \frac{1}{(b+c)a} + \frac{1}{(c+a)b} = T. \tag{246}$$

Hence,

$$2S \geq S + T = \frac{b+c}{(a+b)bc} + \frac{c+a}{(b+c)ca} + \frac{a+b}{(c+a)ab} \geq 3(\frac{1}{abc})^{\frac{1}{3}} = 3. \blacksquare \tag{247}$$

Remarks: From WLOG: $0 \leq c \leq b \leq a$, we have

$$\begin{array}{ccccc} bc & \leq & ac & \leq & ab \\ b+c & \leq & a+c & \leq & a+b \\ \frac{1}{ab} & \leq & \frac{1}{ac} & \leq & \frac{1}{bc} \\ \frac{1}{a+b} & \leq & \frac{1}{a+c} & \leq & \frac{1}{b+c} \end{array} \tag{248}$$

Thus, by rearrangement inequalities, the following can be proved.

$$\frac{ab}{a+b} + \frac{ac}{a+c} + \frac{bc}{b+c} \leq \frac{ac}{a+b} + \frac{bc}{a+c} + \frac{ab}{b+c} \leq \frac{bc}{a+b} + \frac{ac}{a+c} + \frac{ab}{b+c} \tag{249}$$

$$\frac{ab}{a+b} + \frac{ac}{a+c} + \frac{bc}{b+c} \leq \frac{ab}{a+b} + \frac{bc}{a+c} + \frac{ac}{b+c} \leq \frac{bc}{a+b} + \frac{ac}{a+c} + \frac{ab}{b+c} \tag{250}$$

$$\frac{ab}{a+b} + \frac{ac}{a+c} + \frac{bc}{b+c} \leq \frac{ac}{a+b} + \frac{ab}{a+c} + \frac{bc}{b+c} \leq \frac{bc}{a+b} + \frac{ac}{a+c} + \frac{ab}{b+c} \tag{251}$$

$$\frac{ab}{a+b} + \frac{ac}{a+c} + \frac{bc}{b+c} \leq \frac{bc}{a+b} + \frac{ab}{a+c} + \frac{ac}{b+c} \leq \frac{bc}{a+b} + \frac{ac}{a+c} + \frac{ab}{b+c}. \tag{252}$$

Ex 10.7 (a, b, c Fraction Inequality with a parameter)

$$\frac{a}{b} + \frac{b}{c} + \frac{c}{a} + \frac{3n\sqrt[3]{abc}}{a+b+c} \geq 3 + n \tag{253}$$

For $n = 1$, We have:

$$\frac{a}{b} + \frac{a}{b} + \frac{b}{c} \geq 3\sqrt[3]{\frac{a^2b}{b^2c}} = \frac{3a}{\sqrt[3]{abc}} \tag{254}$$

$$\frac{b}{c} + \frac{b}{c} + \frac{c}{a} \geq \frac{3b}{\sqrt[3]{abc}} \tag{255}$$

$$\frac{c}{a} + \frac{c}{a} + \frac{a}{b} \geq \frac{3c}{\sqrt[3]{abc}} \tag{256}$$

Therefore:

$$\frac{a}{b} + \frac{b}{c} + \frac{c}{a} \geq \frac{a+b+c}{\sqrt[3]{abc}}. \tag{257}$$

So we just need to prove:

$$\frac{a+b+c}{\sqrt[3]{abc}} + \frac{3\sqrt[3]{abc}}{a+b+c} \geq 4 \tag{258}$$

If let $x = a + b + c$ and $y = \sqrt[3]{abc}$, then the above is equivalent to:

$$(a+b+c-\sqrt[3]{abc})(a+b+c-3\sqrt[3]{abc}) = (x-y)(x-3y) \geq 0 \tag{259}$$

which is true by AM-GM $x \geq 3y$ and equality holds when $a = b = c$.

It is readily shown for a factor $n \leq 3$ multiplied to $\sqrt[3]{abc}$, i.e.,

$$(a+b+c-3\sqrt[3]{abc})(a+b+c-n\sqrt[3]{abc}) = (x-3y)(x-ny) \geq 0. \tag{260}$$

and the condition $n \leq 3$ can be found from its determinant.

10.6 Jensen's Inequality

Suppose a real function $f(x)$ is convex for $x \in [a, b]$. Then, if $\sum_{i=1}^{n} p_i = 1$, we have the inequality

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i) \tag{261}$$

holds $\forall x_i \in [a, b]$ with $i = 1, \cdots n$. If x_i is a random variable with probability p_i, and f is convex, then we have $f(\mathrm{E}[X]) \leq \mathrm{E}[f(\mathbf{x})]$, provided the expectation[6] exist.

The convexity of the function can be checked by the sign of its second derivative, i.e., if $f''(x) \geq 0$ holds for all $x \in [a, b]$. A concave function can be changed by multiplying -1 to become convex. A good example of concave function is the logarithmic function $f(x) = \ln x$ which can be used to prove the generalized **AM-GM** inequality.

$$\mathfrak{G}(\mathbf{a}, \mathbf{p}) = \left(\prod_{i=1}^{n} a_i^{p_i}\right)^{\frac{1}{\sum\limits_{i=1}^{n} p_i}} \leq \frac{\sum\limits_{i=1}^{n} p_i a_i}{\sum\limits_{i=1}^{n} p_i} = \mathfrak{A}(\mathbf{a}, \mathbf{p}). \tag{263}$$

Further, convexity of a function is closely related to the logarithmic correspondences. Formally, a function $f(x)$ is called logarithmic concave (or log-concave) on the interval $[a, b]$ if $f > 0$ and $\ln f(x)$ is concave, i.e., $f(x)f''(x) \leq (f'(x))^2$ on $[a, b]$. Similar definition can be made to log-convex functions. We say $\ln f(x)$ (f must be a positive function) is convex, if $f(x)f''(x) \geq (f'(x))^2$ on $[a, b]$. To generalize for multivariables, suppose f is a twice differentiable function, so

$$\nabla^2 \ln f(\mathbf{x}) = \nabla \cdot \Big(\nabla \ln f(\mathbf{x})\Big) = \nabla \cdot \left(\frac{\nabla f(\mathbf{x})}{f(\mathbf{x})}\right) = \frac{f(\mathbf{x})\nabla^2 f(\mathbf{x}) - \nabla f(\mathbf{x}) \nabla f(\mathbf{x})}{f(\mathbf{x})^2}. \tag{264}$$

We[7] conclude that f is log-convex if and only if for all $\mathbf{x}$,

$$f(\mathbf{x})\nabla^2 f(\mathbf{x}) \geq \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^{'} \tag{265}$$

and log-concave if and only if for all $\mathbf{x}$,

$$f(\mathbf{x})\nabla^2 f(\mathbf{x}) \leq \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^{\cdot} \tag{266}$$

For example, the determinant of a positive definite[8] matrix $\mathbf{A}$ is log-concave.

[6] The expected value (or expectation) is the arithmetic mean of a large number of independently selected outcomes of a random variable $\mathrm{E}[X] = \sum\limits_{i=1}^{\infty} x_i p_i$ where p_i are the corresponding weight/probability functions. If the mathematical expectation $\mathrm{E}[X]$ exists, it satisfies the following property (with constants c_i):

$$\mathrm{E}\left[\sum_{i=1}^{k} c_i p_i(X)\right] = \sum_{i=1}^{k} c_i \, \mathrm{E}[p_i(X)] \tag{262}$$

[7] See, Convex Optimization, S.Boyd and L.Vandenberghe.

[8] Matrix $\mathbf{A}$ is positive definite then all of its eigenvalues are positive. Also for $\forall \mathbf{x}$, we have $\cos\langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle = \frac{\dot{\mathbf{x}} A\mathbf{x}}{\|\mathbf{x}\| \|\mathbf{A}\mathbf{x}\|} > 0 \iff \theta(\mathbf{x}, \mathbf{A}\mathbf{x}) \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

Ex 10.8 (Geometric Series Inequality) *Let a, b positive real numbers. Prove that*

$$\frac{1}{1+(ab)^{\frac{3}{2}}} \le \frac{1}{1+a+a^2+a^3} + \frac{1}{1+b+b^2+b^3} \tag{267}$$

Proof: Let's rewrite the right terms as multiplications

$$\frac{1}{1+(ab)^{\frac{3}{2}}} \le \frac{1}{(1+a)(1+a^2)} + \frac{1}{(1+b)(1+b^2)} \tag{268}$$

and observe that the function

$$f(x) = \frac{1}{(1+e^x)} \tag{269}$$

10.7 Young's Inequality

Let's start with a variant form of Young's inequality where the power indices are non-negative real numbers (excluding the case both are zeros).

$$x^\alpha y^\beta \le \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} y^{\alpha+\beta} \tag{270}$$

for all $x \ge 0$, $y \ge 0$ and $\alpha+\beta > 0$ (Note the implication here: either α or β can be 0, but not both!). One can be surprised to see $a^{2022}+b^{2022} \ge a^{2008}b^{14} + a^{14}b^{2008}$ (by applying (270) first, and keep the results, then, applying (270) again after switching α and β and summing up the results from the two times applications of (270)). The generalization to three factors can be

$$x^\alpha y^\beta z^\gamma + x^\beta y^\gamma z^\alpha + x^\gamma y^\alpha z^\beta \le x^{\alpha+\beta+\gamma} + y^{\alpha+\beta+\gamma} + z^{\alpha+\beta+\gamma} \tag{271}$$

If normalization is used i.e., $\alpha+\beta=1$, then (270) appears simply

$$x^\alpha y^\beta \le \alpha x + \beta y \tag{272}$$

and the case for three variables with $\alpha+\beta+\gamma=1$

$$x^\alpha y^\beta z^\gamma \le \alpha x + \beta y + \gamma z. \tag{273}$$

The variant form (270) is nothing newer other than the standard form

$$ab \le \frac{a^p}{p} + \frac{b^q}{q} \tag{274}$$

by setting $a = x^\alpha$, $b = y^\beta$, $p = (\alpha+\beta)/\alpha$ and $q = (\alpha+\beta)/\beta$.
Observing the different forms of Young's inequality, aren't they just direct generalization of the elementary AM-GM inequality from (103)? The generalized inequality (270) can be readily proved by exploring one of the most powerful but fundamental inequalities $x \le e^{x-1}$. Let the weighted arithmetic mean of the sequence $\{a_1,\dots,a_n\}$ is $A = \sum_j^n(\lambda_j a_j)$ with $\sum_j^n \lambda_j = 1$

$$\left(\frac{a_1}{A}\right)^{\lambda_1}\left(\frac{a_2}{A}\right)^{\lambda_2}\cdots\left(\frac{a_n}{A}\right)^{\lambda_n} = \frac{\prod_{k=1}^n a_k^{\lambda_k}}{A^{\sum_j^n \lambda_j}} \le \exp\left(\sum_k^n \frac{\lambda_k a_k}{A} - 1\right) = 1. \tag{275}$$

$$a_1^{\lambda_1} a_2^{\lambda_2} \ldots a_n{}^{\lambda_n} = \prod_{k=1}^{n} a_n^{\lambda_n} \leq \sum_{j}^{n} (\lambda_j a_j) = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n \tag{276}$$

If $n = 2$, $a_1 = x^{\alpha+\beta}$, $a_2 = y^{\alpha+\beta}$, $\lambda_1 = \frac{\alpha}{\alpha+\beta}$ and $\lambda_2 = \frac{\beta}{\alpha+\beta}$ then,

$$(x^{\alpha+\beta})^{\frac{\alpha}{\alpha+\beta}} (y^{\alpha+\beta})^{\frac{\beta}{\alpha+\beta}} = x^\alpha y^\beta \leq \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} y^{\alpha+\beta} \tag{277}$$

which is exactly (270). So there is no difficulty to arrive at (271) for $n = 3$, $a_1 = x^{\alpha+\beta+\gamma}$, $a_2 = y^{\alpha+\beta+\gamma}$, $a_3 = z^{\alpha+\beta+\gamma}$ and some trivial manipulations in (276).

Another variation form is characterized by putting the constants with the product in one side and keeping the bare sum of variables the other side, for example.

$$(px)^{\frac{1}{p}} (qy)^{\frac{1}{q}} \leq x + y \tag{278}$$

where $1/p + 1/q = 1$.

Ex 10.9 (AM-GM and Young's Inequality) *Prove the following inequalities. Which upper bound of them is larger or equal than the other? How about the condition of equality?*

$$abc \leq \left(\frac{a+b+c}{3}\right)^3 \tag{279}$$

$$abc \leq \frac{a^3+b^3+c^3}{3} \tag{280}$$

Solution: Both are direct results of AM-GM, and the upper bounds comparison arrives by the Power Mean Inequality (125). Together with (143) and for $n \geq 3$, we have

$$abc \leq \frac{a+b+c}{3} \leq \sqrt{\frac{a^2+b^2+c^2}{3}} \leq \sqrt[3]{\frac{a^3+b^3+c^3}{3}} \leq \sqrt[n]{\frac{a^n+b^n+c^n}{3}}. \tag{281}$$

It is obvious that the equality takes when all three a, b and c are identical.

10.7.1 Extension of Young' s inequality and Cauchy inequality

If $1 < p < \infty$ and q such that $p + q = pq$, or

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{282}$$

then

$$ab = \min_{0<t<\infty} \left(\frac{t^p a^p}{p} + \frac{t^{-q} b^q}{q}\right). \tag{283}$$

Let a real-valued function f of the positive real number $t \geq 0$

$$f(t) := \frac{t^p a^p}{p} + \frac{t^{-q} b^q}{q} \tag{284}$$

By setting the derivative over t equal to zero, it can be found that

$$t = (a^{-p}b^q)^{\frac{1}{p+q}} \tag{285}$$

Plugging it into (284), then the minimum of the function $f(t)$ can be found

$$\begin{aligned} f(t) &= \frac{t^p a^p}{p} + \frac{t^{-q}b^q}{q} && (286)\\ &\geq \frac{(a^{-p}b^q)^{\frac{p}{p+q}}a^p}{p} + \frac{(a^{-p}b^q)^{\frac{-q}{p+q}}b^q}{q} && (287)\\ &= \frac{a^{\frac{pq}{p+q}}b}{p} + \frac{ab^{\frac{pq}{p+q}}}{q} && (288)\\ &= ab(\frac{1}{p} + \frac{1}{q}) = ab && (289) \end{aligned}$$

where the condition of unity of harmonic sum of p and q (282) has been used. Similarity can be seen in the so-called the “Peter–Paul” inequality (one must “rob Peter to pay Paul”)

$$ab \leq \frac{a^2}{2\varepsilon} + \frac{\varepsilon b^2}{2} \tag{290}$$

10.8 Hölder's Inequality

Hölder's inequality states that, for non-negative sets $\{a_i\}$, $\{b_i\}$, $\{\ldots\}$, $\{z_i\}$ each with m-components, the inequality

$$\begin{aligned} &a_1^{\lambda_a}b_1^{\lambda_b}\cdots z_1^{\lambda_z} + a_2^{\lambda_a}b_2^{\lambda_b}\cdots z_2^{\lambda_z} + \cdots + a_m^{\lambda_a}b_m^{\lambda_b}\cdots z_m^{\lambda_z} \leq \\ &(a_1 + a_2 + \cdots + a_m)^{\lambda_a}(b_1 + b_2 + \cdots + b_m)^{\lambda_b}\cdots(z_1 + z_2 + \cdots + z_m)^{\lambda_z} \end{aligned} \tag{291}$$

holds for all $\lambda_a + \lambda_b + \cdots + \lambda_z = 1$. Note that the Hölder's inequality states the fact that the sum of geometric means of the corresponding components of the sets is smaller than the geometric mean of the sums of their individual components. The number of the addends in the smaller side equals the number of components of the set in the larger side. For instance, in the case of $\lambda_a = \lambda_b = \frac{1}{2}$, Hölder's inequality reduces to the square root form of C-S inequality (168).

It might be most elegant to cast Hölder's inequality in $m \times n$ matrix form

$$\begin{array}{cccc} \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ddots & a_{mn} \end{array}\right] & \overset{G}{\Longrightarrow} & \begin{array}{l} a_{11}^{\lambda_1}a_{12}^{\lambda_2}\cdots a_{1n}^{\lambda_n} = \prod_{j=1}^{n} a_{1j}^{\lambda_j} \\ a_{21}^{\lambda_1}a_{22}^{\lambda_2}\cdots a_{2n}^{\lambda_n} = \prod_{j=1}^{n} a_{2j}^{\lambda_j} \\ \cdots \quad \cdots \\ a_{m1}^{\lambda_1}a_{m2}^{\lambda_2}\cdots a_{mn}^{\lambda_n} = \prod_{j=1}^{n} a_{mj}^{\lambda_j} \end{array} \\ \begin{array}{cccc} S\Downarrow & S\Downarrow & S\Downarrow & S\Downarrow \\ \left(\sum_{i=1}^{m} a_{i1}\right)^{\lambda_1} & \left(\sum_{i=1}^{m} a_{i2}\right)^{\lambda_2} & \cdots & \left(\sum_{i=1}^{m} a_{in}\right)^{\lambda_n} \end{array} & & \end{array} \tag{292}$$

So Hölder's inequality becomes

$$\sum_{i=1}^{m}\prod_{j=1}^{n}a_{ij}^{\lambda_j} \leq \prod_{j=1}^{n}\left(\sum_{i=1}^{m}a_{ij}\right)^{\lambda_j} \tag{293}$$

or explicitly

$$\begin{aligned} &a_{11}^{\lambda_1}a_{12}^{\lambda_2}\cdots a_{1n}^{\lambda_n} + a_{21}^{\lambda_1}a_{22}^{\lambda_2}\cdots a_{2n}^{\lambda_n} + \cdots + a_{m1}^{\lambda_1}a_{m2}^{\lambda_2}\cdots a_{mn}^{\lambda_n} \leq \\ &(a_{11}+a_{21}+\cdots+a_{m1})^{\lambda_1}(a_{12}+a_{22}+\cdots+a_{m2})^{\lambda_2}\cdots(a_{1n}+a_{2n}+\cdots+a_{mn})^{\lambda_n} \end{aligned} \tag{294}$$

So (291) can be reformed in matrix form as

$$\begin{array}{cccc} \begin{bmatrix} a_1 & b_1 & \cdots & z_1 \\ a_2 & b_2 & \cdots & z_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_m & b_m & \ddots & z_n \end{bmatrix} & \overset{G}{\Longrightarrow} & \begin{array}{c} a_1^{\lambda_a}b_1^{\lambda_b}\cdots z_1^{\lambda_z} \\ a_2^{\lambda_a}b_2^{\lambda_b}\cdots z_2^{\lambda_z} \\ \cdots \\ a_m^{\lambda_a}b_m^{\lambda_b}\cdots z_m^{\lambda_z} \end{array} \\ \begin{array}{cccc} S\Downarrow & S\Downarrow & S\Downarrow & S\Downarrow \\ \left(\sum_{i=1}^{m}a_i\right)^{\lambda_a} & \left(\sum_{i=1}^{m}b_i\right)^{\lambda_b} & \cdots & \left(\sum_{i=1}^{m}z_i\right)^{\lambda_z} \end{array} & & \end{array} \tag{295}$$

The procedures in (292), (293) or (295) tell us that if we let **G**eomtric Averaging apply on rows and let **S**ummation apply on columns, then Hölder's inequality states that by the operation first with the geometric mean **G**, we get a smaller result than if we start with **S** operation first.

The standard Hölder's inequality in (291) requires to have normalized power indices, i.e., $\lambda_a + \lambda_b + \cdots + \lambda_z = 1$. Sometimes the power indices are expressed as similar to the standard form of Young's inequality (274).

$$|a_1b_1 + a_2b_2 + \ldots + a_nb_n| \leq (a_1^p + a_2^p + \ldots + a_n^p)^{\frac{1}{p}}(b_1^q + b_2^q + \ldots + b_n^q)^{\frac{1}{q}}. \tag{296}$$

where the numbers $p,\ q > 1$ are said to be Hölder conjugates of each other, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{297}$$

The special case $p = q = 2$ gives a form of the Cauchy–Schwarz inequality. One might encounter the inequality written in L_p-norm in other places, like linear algebra.

$$\|xy\|_1 \leq \|x\|_p\,\|y\|_q\,. \tag{298}$$

The L_p-norm is also called Hölder's norm which is defined by

$$\|x\|_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}. \tag{299}$$

Bound wave: product of 1st order waves:

$$(\omega_i-\omega_j)^2 \neq g(k_i-k_j)\tanh\left[(k_i-k_j)h\right]$$

While linear shallow dispersion is $(\omega_i-\omega_j)=\sqrt{gh}(k_i-k_j)$

The power indices can be in very different forms and are not necessarily summed to 1, but we can rearrange the inequality to become the standard form. Let's only take two sequences for a proof and the extension to multi sequences is straightforward. Assume that the sequences are $\{a_1,a_2,\ldots,a_n\}$ and $\{b_1,b_2,\ldots,b_n\}$. And α and β are any real positive power indices, the variant version of Hölder's inequality for two sequences can be written as

$$\begin{aligned}&\left(a_1^\alpha b_1^\beta\right)^{\frac{1}{\alpha+\beta}}+\left(a_2^\alpha b_2^\beta\right)^{\frac{1}{\alpha+\beta}}+\ldots+\left(a_n^\alpha b_n^\beta\right)^{\frac{1}{\alpha+\beta}}\\ &\leq\left((a_1+a_2+\ldots+a_n)^\alpha(b_1+b_2+\ldots+b_n)^\beta\right)^{\frac{1}{\alpha+\beta}}.\end{aligned} \tag{300}$$

When taking the power of $1/(\alpha+\beta)$ to both sides, the standard form is immediately retrieved with $\lambda_a=\alpha/(\alpha+\beta)$ and $\lambda_b=\beta/(\alpha+\beta)$. The power fractions in (300) help us to check homogeneity of terms, but they might complicate the application or computations, so sometimes we can also make it compact by clearing them off. Let $x_k=a_k^{\alpha+\beta}$, $y_k=b_k^{\alpha+\beta}$ and transfer the power to the left side, it becomes

$$\left(\sum_{i=1}^n x_i^\alpha\, y_i^\beta\right)^{\alpha+\beta}\leq\left(\sum_{i=1}^n x_i^{\alpha+\beta}\right)^\alpha\left(\sum_{i=1}^n y_i^{\alpha+\beta}\right)^\beta. \tag{301}$$

Ex 10.10 (IMO 2001) *As one application, look at Problem 2, IMO 2001:*

$$\left(\sum_{cyc}\frac{a}{\sqrt{a^2+8bc}}\right)^2\sum_{cyc}a(a^2+8bc) \tag{302}$$

$$\geq\left(\sum_{cyc}\sqrt[3]{\left(\frac{a}{\sqrt{a^2+8bc}}\right)^2a(a^2+8bc)}\right)^3=(a+b+c)^3. \tag{303}$$

In our case $n=3$, $\alpha=2$ and $\beta=1$, so

$$\sum_{cyc}\frac{a}{\sqrt{a^2+8bc}}\geq\frac{(a+b+c)^3}{a^3+b^3+c^3+24abc}\geq 1. \tag{304}$$

The last step comes from the fact that $(a+b)(b+c)(c+a)\geq 8abc$ or $a(b^2+c^2)+b(c^2+a^2)+c(a^2+b^2)\geq 6abc$.

Alternatively, we can also utilize Hölder's inequality in matrix form.

$$\begin{bmatrix}\frac{a}{\sqrt{a^2+8bc}} & \frac{a}{\sqrt{a^2+8bc}} & a(a^2+8bc)\\ \frac{b}{\sqrt{a^2+8ca}} & \frac{b}{\sqrt{a^2+8ca}} & b(b^2+8ca)\\ \frac{c}{\sqrt{a^2+8ab}} & \frac{c}{\sqrt{a^2+8ab}} & c(c^2+8ab)\end{bmatrix} \tag{305}$$

By swapping **G**ometric Meaning and **S**ummation on the rows and columns or applying (293), we will obtain

$$a+b+c\leq\sqrt[3]{\left(\sum_{cyc}\frac{a}{\sqrt{a^2+8bc}}\right)^2\sum_{cyc}a(a^2+8bc)}. \tag{306}$$

which is the same as (303). Also see (199) for other approach.

One might be encountered by this

$$(a_1 \cdots a_n)^{\frac{1}{n}} + (b_1 \cdots b_n)^{\frac{1}{n}} \leq \sqrt[n]{(a_1+b_1)\cdots(a_n+b_n)} \tag{307}$$

and be confused at first glance. But it is actually another variant form of Hölder's inequality just for two sequences $\{a_i\}$, $\{b_i\}$, $i = 1 \ldots n$.

$$a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_n^{\lambda_n} + b_1^{\lambda_1} b_2^{\lambda_2} \cdots b_n^{\lambda_n} \leq (a_1+b_1)^{\lambda_1}(a_2+b_2)^{\lambda_2}\cdots(a_n+b_n)^{\lambda_n} \tag{308}$$

and $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \frac{1}{n}$.

Ex 10.11 (Simple Sum Inequality) *Suppose that $a, b \in [0,1]$. Prove that*

$$\frac{1}{\sqrt{1+a^2}} + \frac{1}{\sqrt{1+b^2}} \leq \frac{2}{\sqrt{1+ab}}. \tag{309}$$

Solution 1: We have

$$\left(\frac{1}{\sqrt{1+a^2}} + \frac{1}{\sqrt{1+b^2}}\right)^2 = \frac{1}{1+a^2} + \frac{1}{1+b^2} + \frac{2}{\sqrt{(1+a^2)(1+b^2)}} \tag{310}$$

Using the AM-GM we have

$$\frac{2}{\sqrt{(1+a^2)(1+b^2)}} \leqslant \frac{1}{1+a^2} + \frac{1}{1+b^2}. \tag{311}$$

Therefore, we need to prove

$$\frac{1}{1+a^2} + \frac{1}{1+b^2} \leqslant \frac{2}{1+ab}, \tag{312}$$

which is equivalent to

$$\frac{(ab-1)(a-b)^2}{(1+a^2)(1+b^2+1)(1+ab)} \leqslant 0. \tag{313}$$

Solution 2

Let $a = \tan x$ and $b = \tan y$, with $0 \leq x, y \leq \frac{\pi}{4}$. The required inequality becomes

$$\cos x + \cos y \leq 2\frac{\sqrt{\cos x \cos y}}{\sqrt{\cos(x-y)}}, \tag{314}$$

or, equivalently,

$$(\cos x + \cos y)^2 \cos(x-y) \leq 4 \cos x \cos y. \tag{315}$$

Note that $\cos x + \cos y \le 2\cos\left(\frac{x+y}{2}\right)$. Thus, suffice it to prove that

$$\cos^2\left(\frac{x+y}{2}\right)\cos(x-y) \le \cos x \cos y. \tag{316}$$

Recollect that

$$\cos^2\left(\frac{x+y}{2}\right) = \frac{1+\cos(x+y)}{2} \tag{317}$$

and

$$\cos x \cos y = \frac{\cos(x-y)+\cos(x+y)}{2}, \tag{318}$$

so that we need to prove

$$\cos(x-y)+\cos(x-y)\cos(x+y) \le \cos(x-y)+\cos(x+y), \tag{319}$$

which is obvious since $\cos(x+y) \ge 0$ and $\cos(x-y) \le 1$. That completes the proof. Equality holds if $x = y$, i.e., if $a = b$.

Solution 3 By squaring both sides and then comparing the squared terms without square roots on the left with the half of the squared right. The cross term with square root is left to compare.

$$\frac{1}{1+a^2}+\frac{1}{1+b^2}+\frac{2}{\sqrt{(1+a^2)(1+b^2)}} \le \frac{4}{1+ab}. \tag{320}$$

Note that

$$\frac{1}{1+a^2}+\frac{1}{1+b^2}-\frac{2}{1+ab} = \frac{(ab-1)(a-b)^2}{(a^2+1)(b^2+1)(1+ab)} \le 0, \tag{321}$$

because $ab \le 1.$ and equality is attained for $a = b$. It follows that

$$\frac{1}{1+a^2}+\frac{1}{1+b^2} \le \frac{2}{1+ab}. \tag{322}$$

If we further apply AM-GM, then it is found that the larger part is already no larger than the half right one $2/(1+ab)$, i.e.,

$$\frac{2}{\sqrt{(1+a^2)(1+b^2)}} \le \frac{1}{1+a^2}+\frac{1}{1+b^2} \le \frac{2}{1+ab} \tag{323}$$

Therefore the proof can be done here. Of course if we would like to do a thorough check, we can apply the C-S inequality

$$\sqrt{(1+a^2)(1+b^2)} \ge 1+ab, \tag{324}$$

so that

$$\frac{2}{\sqrt{(1+a^2)(1+b^2)}} \le \frac{2}{1+ab}, \tag{325}$$

and, finally we add the proved (322) to both sides, the proof is then completely done.

$$\left(\frac{1}{\sqrt{1+a^2}}+\frac{1}{\sqrt{1+b^2}}\right)^2 \le \left(\frac{2}{\sqrt{1+ab}}\right)^2. \tag{326}$$

Equality is attained for $a=b=1$ or $a=b=0$.

Solution 4 Let

$$f(x)=\sqrt{1+e^{2x}},\ x\le 0. \tag{327}$$

$$f'(x)=e^{2x}\left(1+e^{2x}\right)^{-\frac{3}{2}}, \tag{328}$$

$$f''(x) = -2e^{2x}\left(1+e^{2x}\right)^{-\frac{3}{2}}+3e^{4x}\left(1+e^{2x}\right)^{-\frac{5}{2}} \tag{329}$$

$$= e^{2x}\left(1+e^{2x}\right)^{-\frac{5}{2}}\left[3e^{2x}-2(1+e^{2x})\right] \tag{330}$$

$$= e^{2x}\left(1+e^{2x}\right)^{-\frac{5}{2}}\left(e^{2x}-2\right)<0, \tag{331}$$

since $x<0$. Thus, $f(x)$ is concave and, according to Jensen's inequality,

$$f(x)+f(y)\le 2f\left(\frac{x+y}{2}\right) \tag{332}$$

or

$$(1+e^{2x})^{-\frac{1}{2}}+(1+e^{2y})^{-\frac{1}{2}}\le 2(1+e^{x+y})^{-\frac{1}{2}}. \tag{333}$$

Substituting $e^x=a$ and $a^y=b$ gives

$$(1+a^2)^{-\frac{1}{2}}+(1+b^2)^{-\frac{1}{2}}\le 2(1+ab)^{-\frac{1}{2}} \tag{334}$$

which is

$$\frac{1}{\sqrt{1+a^2}}+\frac{1}{\sqrt{1+b^2}}\le\frac{2}{\sqrt{1+ab}}. \tag{335}$$

Ex 10.12 (Simple Sum Inequality) *Suppose that $a,b\in[0,1]$. Prove that*

$$\frac{1}{\sqrt{1+a^2}}+\frac{1}{\sqrt{1+b^2}}\le\frac{2}{\sqrt{1+ab}}. \tag{336}$$

Ex 10.13 (2014 Czech & Slovak MO, Rnd. III, Cat. A) *For arbitrary non-negative numbers a and b prove*

$$\frac{a}{\sqrt{b^2+1}}+\frac{b}{\sqrt{a^2+1}}\ge\frac{a+b}{\sqrt{ab+1}} \tag{337}$$

Solution 1: The same trick as (199) can be applied. Multiply a and b to nominator and denominator on the left and use C-S in Engel form (176).Note we have to assume $ab > 0$ since if otherwise the equality already holds:

$$LHS = \frac{a^2}{a\sqrt{b^2+1}} + \frac{b^2}{b\sqrt{a^2+1}} \geq \frac{(a+b)^2}{a\sqrt{b^2+1}+b\sqrt{a^2+1}}. \tag{338}$$

Then we just need to show that

$$(a+b)\sqrt{ab+1} \geq a\sqrt{b^2+1} + b\sqrt{a^2+1}. \tag{339}$$

Keep clearing the square roots, we get $a^2+b^2+2 \geq 2\sqrt{(a^2+1)(b^2+1)}$ and then we can get $(a^2-b^2)^2 \geq 0$ which is true.

Solution 2: The Hölder's inequality can be used similarly as in (305). Write the power splitting formation in matrix.

$$\begin{bmatrix} \left(\frac{a}{\sqrt{b^2+1}}\right)^{\frac{1}{2}} & \left(a\sqrt{b^2+1}\right)^{\frac{1}{2}} \\ \left(\frac{b}{\sqrt{a^2+1}}\right)^{\frac{1}{2}} & \left(b\sqrt{a^2+1}\right)^{\frac{1}{2}} \end{bmatrix} \tag{340}$$

So we apply **G**-op and **S**-op on the rows and columns following (293) and we get

$$a+b \leq \left(\frac{a}{\sqrt{b^2+1}} + \frac{b}{\sqrt{a^2+1}}\right)^{\frac{1}{2}} \left(a\sqrt{b^2+1}+b\sqrt{a^2+1}\right)^{\frac{1}{2}}. \tag{341}$$

The next follows (338).

Solution 3: The Hölder's inequality in an alternative way to split the powers.

$$\begin{bmatrix} \left(\frac{a}{\sqrt{b^2+1}}\right)^{\frac{2}{3}} & \left(a(b^2+1)\right)^{\frac{1}{3}} \\ \left(\frac{b}{\sqrt{a^2+1}}\right)^{\frac{2}{3}} & \left(b(a^2+1)\right)^{\frac{1}{3}} \end{bmatrix} \tag{342}$$

After applying (293) and we get

$$a+b \leq \left(\frac{a}{\sqrt{b^2+1}} + \frac{b}{\sqrt{a^2+1}}\right)^{\frac{2}{3}} \left(a(b^2+1)+b(a^2+1)\right)^{\frac{1}{3}}. \tag{343}$$

Then we have

$$\left(\frac{a}{\sqrt{b^2+1}} + \frac{b}{\sqrt{a^2+1}}\right)^2 \geq \frac{(a+b)^3}{a(b^2+1)+b(a^2+1)} = \frac{(a+b)^2}{ab+1} \tag{344}$$

and it is done.

Solution 4: The inequality can be also solved by Jensen's inequality for function $f(x) = x^{-1/2}$:

$$\frac{a}{a+b}f(b^2+1) + \frac{b}{a+b}f(a^2+1) \geq f\left(\frac{a}{a+b}(b^2+1)+\frac{b}{a+b}(a^2+1)\right) = f(ab+1) \tag{345}$$

Ex 10.14 (Generic 2014 Czech & Slovak MO Problem) *For arbitrary non-negative numbers a, b and $n \geq 1$, prove the inequality*

$$\frac{a^n}{\sqrt{b^2+1}} + \frac{b^n}{\sqrt{a^2+1}} \geq \frac{a^n+b^n}{\sqrt{ab+1}} \tag{346}$$

Solution 1: WLOG: $a \geq b$ $(*)$. Firstly let's start from the case when $n = 2$. We can rewrite inequality as:

$$a^2\sqrt{a^2+1}(\sqrt{ab+1} - \sqrt{b^2+1}) \geq b^2\sqrt{b^2+1}(\sqrt{a^2+1} - \sqrt{ab+1}) \tag{347}$$

after multiplying both sides by $(\sqrt{ab+1} + \sqrt{b^2+1})(\sqrt{a^2+1} + \sqrt{ab+1})$ we get

$$a\sqrt{a^2+1}(\sqrt{ab+1} + \sqrt{a^2+1}) \geq b\sqrt{b^2+1}(\sqrt{b^2+1} + \sqrt{ab+1}) \tag{348}$$

which is true by WLOG. Alternative approach which is similar to (338) can be used.

$$\frac{a^2}{\sqrt{b^2+1}} + \frac{b^2}{\sqrt{a^2+1}} \geq \frac{(a^2+b^2)^2}{a^2\sqrt{b^2+1} + b^2\sqrt{a^2+1}} \tag{349}$$

$$\geq \frac{(a^2+b^2)^2}{\sqrt{a^4(b^2+1) + a^2b^2(a^2+1+b^2+1) + b^4(a^2+1)}} \tag{350}$$

$$= \frac{(a^2+b^2)^2}{\sqrt{(a^2+b^2)(a^2+b^2+2a^2b^2)}} \tag{351}$$

$$= \frac{a^2+b^2}{\sqrt{\frac{2a^2b^2}{a^2+b^2}+1}} \geq \frac{a^2+b^2}{\sqrt{ab+1}}. \tag{352}$$

Solution 2: Since $a, b > 0$, the Hölder's Inequality gives that

$$\left[a^n(b^2+1) + b^n(a^2+1)\right]\left(\frac{a^n}{\sqrt{b^2+1}} + \frac{b^n}{\sqrt{a^2+1}}\right)^2 \geq (a^n+b^n)^3 \tag{353}$$

$$\Longleftrightarrow \left(\frac{a^n}{\sqrt{b^2+1}} + \frac{b^n}{\sqrt{a^2+1}}\right)^2 \geq \frac{(a^n+b^n)^3}{ab\left(a^{n-1}b + b^{n-1}a\right) + a^n + b^n}. \tag{354}$$

It remains to show $\frac{(a^n+b^n)^3}{ab(a^{n-1}b+b^{n-1}a)+a^n+b^n} \geq \frac{(a^n+b^n)^2}{ab+1}$, which is reduced to $a^{n-1}b + b^{n-1}a \leq a^n + b^n$. Indeed, by rearrangement or factoring

$$a^n + b^n - a^{n-1}b - b^{n-1}a = (a-b)\left(a^{n-1} - b^{n-1}\right) \geq 0 \tag{355}$$

is clearly true for all positive real numbers a, b and $n \geq 1$. ■

Ex 10.15 (Fixed sum of four numbers) *Let a, b, c and d be positive numbers such that $a+b+c+d = 4$. Prove that*

- ***Subproblem I.***

$$\frac{a}{b} + \frac{b}{c} + \frac{c}{d} + \frac{d}{a} \leq \frac{4}{abcd}. \tag{356}$$

- ***Subproblem II.***

$$\left(\frac{1}{a^{12}} + \frac{1}{b^{12}} + \frac{1}{c^{12}} + \frac{1}{d^{12}}\right)(1 + 3abcd) \geq 16 \tag{357}$$

Solution 1 to *Subproblem I*: By clearing the denominator, the original is equivalent to

$$abcd\left(\frac{a}{b}+\frac{b}{c}+\frac{c}{d}+\frac{d}{a}\right)=a^2cd+ab^2d+abc^2+bcd^2\leq 4. \tag{358}$$

Let $\{a,b,c,d\}=\{x,y,z,t\}$, where $x\geq y\geq z\geq t$. Hence, look at the two sets of variables (x,y,z,t) and (xyz,xyt,xzt,yzt) which are the same ordered. By Rearrangement and two times AM-GM we obtain:

$$\begin{aligned}
&a^2cd+ab^2d+abc^2+bcd^2=a\cdot acd+b\cdot abd+c\cdot abc+d\cdot bcd\leq\\
\leq&x\cdot xyz+y\cdot xyt+z\cdot xzt+t\cdot yzt=xy(xz+yt)+zt(xz+yt)=\\
=&(xy+zt)(xz+yt)\leq\left(\frac{xy+xz+zt+yt}{2}\right)^2=\\
=&\left(\frac{(x+t)(y+z)}{2}\right)^2\leq\left(\frac{\left(\frac{x+y+z+t}{2}\right)^2}{2}\right)^2=4.
\end{aligned} \tag{359}$$

So $a^2cd+ab^2d+abc^2+bcd^2\leq 4$ is proved.

Solution 2 to *Subproblem I*: Let $S=a^2bc+b^2cd+c^2da+d^2ab$. We can easily find that:

$$\begin{aligned}
S-(ac+bd)(ab+cd)&=-bd(a-c)(b-d);\\
S-(bc+ad)(bd+ac)&=ac(a-c)(b-d)
\end{aligned}$$

which implies $S\leq\max\{(ac+bd)(ab+cd),(bc+ad)(bd+ac)\}$. By AG mean inequality:

$$\begin{aligned}
(ac+bd)(ab+cd)&\leq\left(\frac{(ac+bd)+(ab+cd)}{2}\right)^2\\
&=\frac{(a+d)^2(b+c)^2}{4}\\
&\leq\frac{1}{4}\left[\left(\frac{a+d+b+c}{2}\right)^2\right]^2=4
\end{aligned}$$

Similarly, we have

$$(bc+ad)(bd+ac)\leq 4$$

Thus we have $S\leq 4$.

Solution to *Subproblem II*:

$$\begin{aligned}
\left(\frac{1}{a^{12}}+\frac{1}{b^{12}}+\frac{1}{c^{12}}+\frac{1}{d^{12}}\right)(1+3abcd)&\geq\frac{4}{(abcd)^3}(1+3abcd)\\
&=\frac{4}{(abcd)^3}+\frac{12}{(abcd)^2}\\
&\geq 4+12=16
\end{aligned}$$

11 Some special series expansions

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots = -\sum_{k=1}^{\infty} \frac{x^k}{k}, \quad \forall\, |x| < 1 \tag{360}$$

$$\log(I_n - A) = -A - \frac{A^2}{2} - \frac{A^3}{3} - \ldots = -\sum_{k=1}^{\infty} \frac{A^k}{k}, \quad \forall\, \rho(A) < 1 \tag{361}$$

$$\log(x) = \int_0^1 \frac{(x-1)dt}{(x-1)t+1} \xrightarrow{\forall \lambda_i \notin \mathbb{R}^-} \log(A) = \int_0^1 \frac{(A-I_n)dt}{(A-I_n)t+1} \tag{362}$$

$$e^{\frac{2\pi x}{\sqrt{1-x^2}}} = 1 + 2\pi x + 2\pi^2 x^2 + (\pi + \frac{4\pi^3}{3})x^3 + \frac{2}{3}\pi^2(3+\pi^2)x^4 \tag{363}$$

$$+ (\frac{3}{4}\pi + 2\pi^3 + \frac{4}{15}\pi^5)x^5 + O(x^6) \tag{364}$$

$$e^{2\pi \tan \varepsilon} = e^{\frac{2\pi \sin \varepsilon}{\sqrt{1-\sin^2 \varepsilon}}} \tag{365}$$

$$= 1 + 2\pi\varepsilon + 2\pi^2\varepsilon^2 + \frac{2}{3}(\pi + 2\pi^3)\varepsilon^3 + \frac{2}{3}\pi^2(2+\pi^2)\varepsilon^4 \tag{366}$$

$$+ \frac{4}{15}(\pi + 5\pi^3 + \pi^5)\varepsilon^5 + O(\varepsilon^6) \tag{367}$$

References