Lecture Series on Math Fundamentals for MIMO Communications

Topic 2. Convergence of Deterministic/Random Scalar/Matrix and Applications in MIMO and Massive MIMO

Drs. Yindi Jing and Xinwei Yu
University of Alberta
April 15, 2020

Content

Part 1. Definitions and Results

1.1. Convergence for deterministic scalar/matrix sequence

Def. A scalar sequence $x_1, x_2, \cdots, x_M, \cdots$, or simply $\{x_M\}$, **converges** to a number x if for any $\epsilon > 0$, there exists a positive integer m such that for all $M \geq m$, $|x_M - x| < \epsilon$ holds.

This is also denoted as

$$\lim_{M\to\infty} x_M = x.$$

Remarks:

- x_M is deterministic.
- x_M can be either real-valued or complex-valued.
- x is a deterministic value independent of M.
- $|\cdot|$ is the absolute value or modulus, which is a size measure.

Examples:

- $\lim_{M\to\infty} \frac{(-1)^M}{M} = 0.$
- $\lim_{M\to\infty} (-1)^M$ does not exist.

Def. A $K \times N$ matrix sequence $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_M, \cdots$, or simply $\{\mathbf{X}_M\}$, **converges** to a matrix $\mathbf{X}$ if for any $\epsilon > 0$, there exists a positive integer m such that for all $M \geq m$, $\|\mathbf{X}_m - \mathbf{X}\| < \epsilon$ holds. This is also denoted as

$$\lim_{M \to \infty} \mathbf{X}_M = \mathbf{X}.$$

Remarks:

- $\mathbf{X}_M$ is deterministic meaning that all entries are deterministic.
- Elements of $\mathbf{X}_M$ can be either real-valued or complex-valued.
- $\mathbf{X}$ is a deterministic $K \times N$ matrix **independent of** M.
- K and N are **finite integers independent** of M.
- $\| \cdot \|$ denotes **matrix norm**, which is a size measure.

Review on matrix norm (for a $K \times N$ matrix $\mathbf{A}$)

- **Frobenius norm**: $\|\mathbf{A}\|_F = \left(\sum_{i=1}^{K}\sum_{j=1}^{N}|a_{ij}|^2\right)^{1/2}$, where a_{ij} is the (i,j)-th entry of $\mathbf{A}$.
- **p-norm induced from vector norm:**

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}, \text{ where } 1 \leq p \leq \infty.$$

  - 1-norm: $\|\mathbf{A}\|_1 = \max_{j=1,\cdots,N}\sum_{i=1}^{K}|a_{ij}|$.
  - ∞-norm: $\|\mathbf{A}\|_\infty = \max_{i=1,\cdots,K}\sum_{j=1}^{N}|a_{ij}|$.
  - 2-norm: $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$, the largest singular value of $\mathbf{A}$.
- **Entry-wise matrix p-norm:**

$$\|\mathbf{A}\|_{p,vec} = \|\text{vec}(\mathbf{A})\|_p = \left(\sum_{i=1}^{K}\sum_{j=1}^{N}|a_{ij}|^p\right)^{1/p}, \quad 1 \leq p < \infty.$$

  - **Max-norm** (for $p=\infty$): $\|\mathbf{A}\|_{\max} = \max_{i,j}|a_{i,j}|$.
- **All equivalent in the convergence definition**.

The convergence definition on Page 4 is also **equivalent to the entry-wise convergence given below.**

Def. A $K \times N$ (where K and N are fixed) matrix sequence $\mathbf{X}_M$ converges to a matrix $\mathbf{X}$ if every entry converges, i.e.,

$$\lim_{M\to\infty} [X_M]_{ij} = [X]_{ij}, \text{ for all } i = 1, \cdots, K, j = 1, \cdots, N.$$

Remarks:

- When **the dimensions are unlimited**, e.g., $K = M \to \infty$, $\lim_{M\to\infty} \mathbf{X}_M = \mathbf{X}$ is **not well-defined**.
- **Possibilities** for the case of unlimited dimensions.
  - **Approximate X_M by a simpler matrix sequence Y_M.** For example, if $\lim_{M\to\infty} \|\mathbf{X}_M - \mathbf{I}_M\| = 0$, we may say $\lim_{M\to\infty} X_M$ is the identity (map). Note that the convergence is in the sense of norm.
  - **Treat X_M as a sequence of operators**: More generally, one can define convergence in the sense of operators (strong/weak convergence etc.).

1.2. Convergence for random variable/matrix sequence [1]

- Multiple definitions: **convergence in different senses**
- Have slight difference

Def. A sequence of random variables $\{X_M\}$ **converges almost surely** or **almost everywhere** or **with probability 1** or **strongly** towards the random variable X if

$$\mathbb{P}\left(\lim_{M\to\infty} X_M = X\right) = 1.$$

- *Notation:* $X_M \xrightarrow{a.s.} X$ when $M \to \infty$ or $X_M \xrightarrow[M\to\infty]{a.s.} X$.
- $\mathbb{P}\left(\{w \in \Omega : \lim_{M\to\infty} X_M(\omega) = X(\omega)\}\right) = 1.$
- Meaning: Events for which X_M does not converge to X have zero-probability.

- The **strong law of large numbers** is concerned with almost sure convergence.

  If $Y_1, Y_2, \cdots, Y_M, \cdots$ is an infinite sequence of i.i.d. random variables with expected value μ, the sample average *converges almost surely* to the expected value, i.e.,

$$\frac{1}{M}\sum_{m=1}^{M} Y_m \xrightarrow[M\to\infty]{a.s.} \mu.$$

  – The limit μ in the above is a deterministic value. Therefor this is a special case of the almost sure convergence definition.

  – Widely used in research on massive MIMO.

- In the study of the SINR or capacity or a signal/interference component, it is important to know when a sequence of random variables becomes deterministic in the limit. Such behavior is often referred to as **asymptotic determinicity**.
- Many work interpret it mathematically as the almost sure convergence of the random variable sequence to a deterministic value.
- One disadvantage of such interpretation is the limited availability of mathematical tools for almost sure convergence, as it cannot be quantified through a norm or even a metric, resulting in great challenges in derivations.
- In fact, almost sure convergence is not topological. There is no topology on the space of random variables such that the almost surely convergent sequences are exactly the converging sequences with respect to that topology.

Def. A sequence of random variables $\{X_M\}$ is said to **converge in probability** towards the random variable X if for all $\epsilon > 0$,

$$\lim_{M\to\infty} \mathbb{P}\left(|X_M - X| > \epsilon\right) = 0.$$

- Notation: $X_M \xrightarrow{p} X$ when $M \to \infty$ or $X_M \underset{M\to\infty}{\xrightarrow{p}} X$.
- The $|\cdot|$ operation can be generalized to any distance measure.
- The **weak law of large numbers** is concerned with convergence in probability.

  If $Y_1, Y_2, \cdots, Y_M, \cdots$ is an infinite sequence of i.i.d. random variables with expected value μ, the sample average *converges in probability* to the expected value, i.e.,

$$\frac{1}{M}\sum_{m=1}^{M} Y_m \underset{M\to\infty}{\xrightarrow{p}} \mu.$$

- The i.i.d. assumption can be greatly relaxed, for example to $Y_1, \ldots, Y_M$ independent, and $\mathrm{Var}\left(Y_M\right) = \sigma^2$ for all M.

Def. A sequence of random variables $\{X_M\}$ is said to **converge in distribution**, or **converge weakly**, or **converge in law** to a random variable X if

$$\lim_{M\to\infty} F_M(x) = F(x)$$

at *all* x *where* $F(x)$ *is continuous*. Here $F_M(x)$ is the **cumulative distribution functions (CDF)** of X_M and $F(x)$ is the CDF of X.

- Notation: $X_M \xrightarrow{d} X$ when $M \to \infty$ or $X_M \xrightarrow[M\to\infty]{d} X$.
- In contrast to almost sure convergence, convergence in probability and convergence in distribution can be defined through the Lévy-Prokhorov and Ky Fan metrics, respectively.
- On the other hand, many seemingly innocent properties do not hold for convergence in distribution. For example, in general $X_M \xrightarrow{d} X$ and $Y_M \xrightarrow{d} Y$ do not necessarily imply $X_M + Y_M \xrightarrow{d} X + Y$ or $X_M Y_M \xrightarrow{d} XY$.

- **Central limit theorem** is concerned with convergence in distribution.

  If $Y_1, Y_2, \cdots, Y_M, \cdots$ is an infinite sequence of i.i.d. random variables with expected value μ and variance σ^2, as $M \to \infty$, the random variable sequence $\sqrt{M}\left(\frac{1}{M}\sum_{m=1}^{M} Y_m - \mu\right)$ *converges in distribution* to the normal distribution $\mathcal{N}(0, \sigma^2)$, that is,

$$\sqrt{M}\left(\frac{1}{M}\sum_{m=1}^{M} Y_m - \mu\right) \xrightarrow[M\to\infty]{d} \mathcal{N}(0, \sigma^2).$$

  – Application: For large M, the distribution of $\frac{1}{M}\sum_{m=1}^{M} Y_m$ is approximated as Gaussian whose mean is μ and variance is $\frac{\sigma^2}{M}$, i.e., $\frac{1}{M}\sum_{m=1}^{M} Y_m \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{M}\right)$.

Def. A sequence of random variables $\{X_M\}$ is said to **converge in mean square (m.s.)** to a random variable X if

$$\lim_{M \to \infty} \mathbb{E}\left(|X_M - X|^2\right) = 0.$$

- Notation: $X_M \xrightarrow{m.s.} X$ when $M \to \infty$ or $X_M \xrightarrow[M\to\infty]{m.s.} X$, or $X_M \xrightarrow[M\to\infty]{L^2} X$.
- Convergence in the rth-mean: changing the square to the rth-power.
- The definition is **based on *expectation*, not probability, leading to easier calculations in most situations**.

Discussions.

- Connections and differences.
  - a.s. convergence $\Rightarrow$ / m.s. convergence $\Rightarrow$ convergence in probability.
  - convergence in probability $\Rightarrow$ convergence in distribution.
  - a.s. convergence $\not\Rightarrow$ / $\not\Leftarrow$ m.s. convergence.
- Convenience in application.
  - m.s. convergence: **can be defined by a norm**; In fact, the norm can be defined through an inner product.
  - convergence in probability/distribution: **can be defined by a metric**;
  - a.s. convergence: **not topological**.
- The m.s. convergence is at the same time **the strongest and the most convenient to use**.

Properties.

Convergence	a.s.	p	d	m.s.
Linearity	✓	✓	×	✓
Continuous function	✓	✓	✓	×
Product	✓	✓	×	×

- **Linearity:**
  If $X_M \xrightarrow{a.s.} X$ and $Y_M \xrightarrow{a.s.} Y$, then for any a and b,
  $aX_M + bY_M \xrightarrow{a.s.} aX + bY$.
  Also holds for convergence in probability, convergence in m.s., but not for convergence in distribution.

- **Continuous function:**
  If $X_M \xrightarrow{a.s.} X$, for any *continuous function* h, $h(X_M) \xrightarrow{a.s.} h(X)$.
  Also holds for convergence in probability, convergence in distribution but not for convergence in m.s.

- For the **reciprocal function** $h(X_M) = 1/X_M$:
  If $X_M \xrightarrow{a.s./p} X$ and $\underline{\mathbb{P}(X=0)=0}$, we have $X_M \xrightarrow{a.s./p} X$.
  In particular, **if X_M converges to a deterministic value x**,
  $$X_M \xrightarrow{a.s./p} x \neq 0 \;\Rightarrow\; \frac{1}{X_M} \xrightarrow{a.s./p} \frac{1}{x}.$$
- **Product:**
  If $X_M \xrightarrow{a.s.} X$ and $Y_M \xrightarrow{a.s.} Y$, then $X_M Y_M \xrightarrow{a.s.} XY$.
  Also holds for convergence in probability, but not for convergence in distribution or convergence in m.s.
- **Ratio:**
  If $X_M \xrightarrow{a.s./p} X$, $Y_M \xrightarrow{a.s./p} Y$, and $\mathbb{P}(Y=0)=0$, then
  $X_M/Y_M \xrightarrow{a.s./p} X/Y$.
  Particularly, **if X_M, Y_M converge to deterministic values**,
  $$X_M \xrightarrow{a.s./p} x, \quad Y_M \xrightarrow{a.s./p} y \neq 0 \;\Rightarrow\; \frac{X_M}{Y_M} \xrightarrow{a.s./p} \frac{x}{y}.$$
  This is useful in massive MIMO analysis.

Def. The **squared coefficient of variance (SCV)** of a random variable is defined through:

$$\mathrm{SCV}(X) = \frac{\mathrm{Var}\left[X\right]}{\mathbb{E}\left[X\right]^2}.$$

- For a sequence of random variables $\{X_M\}$ with $\mathbb{E}\left[X_M\right] \to x$,
  - When $x \neq 0$, $\mathrm{SCV}(X_M) \to 0$ is equivalent to $X_M \xrightarrow[M\to\infty]{m.s.} x$.
  - When $x = 0$, $\mathrm{SCV}(X_M) \to 0$ implies $X_M \xrightarrow[M\to\infty]{m.s.} x$. But the converse does not hold.
- One can give random variables with positive (or negative) mean a fibre bundle structure and generalize coefficient of variance to a Riemannian metric on it.

- The concept of **asymptotic determinicity** of a random sequence $\{X_M\}$ can be defined through $\mathrm{SCV(X_M)} \rightarrow 0$ and $\mathbb{E}[X_M] \rightarrow x$.
  - Compared to the definition using a.s. convergence (Page 9), the definition through SCV is based on expectation, thus more convenient in applications and manipulations.
  - One could also define asymptotic determinicity through m.s. convergence. However, the SCV one is advantageous since it avoids, through normalization, mistakes caused by scaling. For example, X_M may be re-scaled to X_M/M^n, which converges in m.s. to 0, while $\mathrm{SCV}(X_M/M^n) = \mathrm{SCV}(X_M)$.

Generalization to convergence definitions of random matrix sequence

- The convergence definitions for a random *scalar* sequence can be naturally applied to a random *vector/matrix* sequence via combining with the entry-wise convergence on Page 6.

  Def. A sequence of $K \times N$ (where K and N are fixed) *random matrices* $\mathbf{X}_M$ **converges almost surely** to a random matrix $\mathbf{X}$ if for each entry, the sequence of random variables converges almost surely to the correspondence entry of $\mathbf{X}$, i.e.,

$$[\mathbf{X}_M]_{ij} \xrightarrow[M \to \infty]{a.s.} [\mathbf{X}]_{ij}, \quad \text{for all } i = 1, \cdots, K, \; j = 1, \cdots, N.$$

  Remarks:

  – Though usually not explicitly given, this definition has been widely used in massive MIMO research.

  – It is important that K **and** N **are fixed**, not increasing unboundedly with M.

The combination with the norm-based convergence on Page 4 is straightforward for the case of fixed K and N.

Def. A sequence of $K \times N$ (where K and N are fixed) *random matrices* $\mathbf{X}_M$ **converges almost surely** to a random matrix $\mathbf{X}$ if $\|\mathbf{X}_M - \mathbf{X}\|$ converges almost surely to 0, i.e.,

$$\|\mathbf{X}_M - \mathbf{X}\| \xrightarrow[M\to\infty]{a.s.} 0.$$

Remarks:

- $\|\cdot\|$ denotes any matrix norm. Note that all norms are equivalent as long as K, N are both fixed.
- When K and/or N increases with M, the limit can be studied, similarly to Page 6, when $\|\mathbf{X}_M - \mathbf{Y}_M\| \xrightarrow[M\to\infty]{a.s.} 0$ for some deterministic sequence $\mathbf{Y}_M$.

Part 2. Examples and Applications in MIMO and Massive MIMO Analysis

Simplest models to show the *usage* of the convergence definitions.

2.1. Channel hardening and distribution of MIMO capacity - convergence in distribution and more

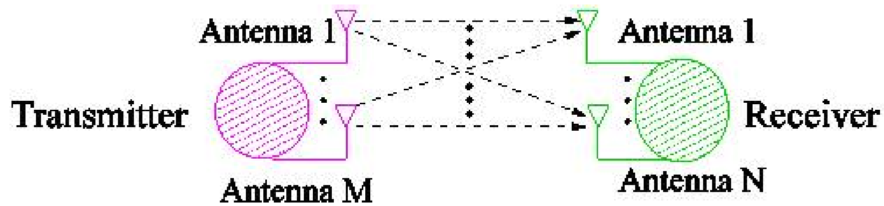
2.2. Achievable rate analysis for single-cell massive MIMO uplink under MRC - almost sure convergence

2.3. Outage probability analysis for massive MIMO downlink - convergence in distribution

2.4. Power scaling law for single-cell massive MIMO downlink - convergence in mean square and SCV determinicity

2.1. Channel hardening and distribution of MIMO capacity [4]

A point-to-point MIMO system



- M transmit antennas
- N receive antennas.
- Channel matrix $\mathbf{H}$ $(N \times M)$ with i.i.d. $\mathcal{CN}(0,1)$ entries [2]. Average transmit SNR: ρ.
- Gaussian signals; CSI at the receiver not the transmitter.
- **Mutual information (capacity):** $\mathcal{I} = \log(\mathbf{I} + \frac{\rho}{M}\mathbf{H}^H\mathbf{H})$.

Goals

- To compute the distribution of $\mathcal{I}$: use **central limit theorem** and more to approximate $\mathcal{I}$ as Gaussian.
- Asymptotic analysis for large M and/or N.

Review of results on <u>MIMO ergodic capacity</u> [3]: $\mathbb{E}\left[\mathcal{I}\right]$.

- If M is fixed and $N \to \infty$,

$$\mathbb{E}\left[\mathcal{I}\right] \approx M \log\left(1 + \frac{N\rho}{M}\right).$$

  – Cannot be put in simple limit form.

  – Approximation by **keeping the most dominant term**.

- If N is fixed and $M \to \infty$,

$$\lim_{M\to\infty} \mathbb{E}\left[\mathcal{I}\right] = N\log(1+\rho).$$

- If $M, N \to \infty$ and $\beta = N/M$ is fixed,

$$\mathbb{E}\left[\mathcal{I}\right] \approx \min\{M, N\} F(\beta, \rho),$$

  where

$$F(\beta,\rho) = \log\left[1 + \rho(\sqrt{\beta}+1)^2\right] + (\beta+1)\log\left(\frac{1+\sqrt{1-a}}{2}\right) - (\log e)\sqrt{\beta}\frac{1-\sqrt{1-a}}{1+\sqrt{1-a}}$$

$$+(\beta-1)\log\left(\frac{1+\alpha}{\alpha+\sqrt{1-\alpha}}\right), \quad a = \frac{2\rho\sqrt{\beta}}{1+\rho(\sqrt{\beta}+1)^2}, \quad \alpha = \frac{\sqrt{\beta}-1}{\sqrt{\beta}+1}.$$

Main results on the distribution of $\mathcal{I}$

- If M is fixed and $N \to \infty$,

$$\sqrt{N}\left[\mathcal{I} - M\log\left(1 + \frac{N\rho}{M}\right)\right] \xrightarrow{d} \mathcal{N}\left(0, M\log^2 e\right). \qquad (1)$$

Approximate distribution for large N:

$$\mathcal{I} \sim \mathcal{N}\left(M\log\left(1 + \frac{N\rho}{M}\right), \frac{M\log^2 e}{N}\right).$$

- If N fixed and $M \to \infty$,

$$\sqrt{M}\left[\mathcal{I} - N\log(1+\rho)\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{N\rho^2\log^2 e}{(1+\rho)^2}\right).$$

Approximate distribution:

$$\mathcal{I} \sim \mathcal{N}\left(N\log(1+\rho), \frac{N\rho^2\log^2 e}{M(1+\rho)^2}\right).$$

- If $M, N \to \infty$,
  - for low SNR,

$$\lim_{\rho \to 0} \frac{1}{\rho \log e} \sqrt{\frac{M}{N}} \left[\mathcal{I} - N\rho \log e \right] \xrightarrow{d} \mathcal{N}(0,1).$$

Approximate distribution: $\mathcal{I} \sim \mathcal{N}\left(N\rho \log e, \frac{N}{M}\rho^2 \log^2 e\right)$.
For the special case of fixed $\beta = N/M$, approximation distribution: $\mathcal{I} \sim \left(N\rho \log e, \beta\rho^2 \log^2 e\right)$.

  - for high SNR,

$$\lim_{\rho \to \infty} \frac{1}{\sigma_{MN}} \left[\mathcal{I} - \mu_{MN} \right] \xrightarrow{d} \mathcal{N}(0,1),$$

where $\mu_{MN} = k \log(\rho/M) + k \log 2 \left(\sum_{i=1}^{K-k} \frac{1}{i} - \gamma \right) + \log e \sum_{i=1}^{k-1} \frac{i}{K-i}$,

$$\sigma^2_{MN} = \log^2 e \left(\sum_{i=1}^{k-1} \frac{i}{(K-k+i)^2} + k \left[\frac{\pi^2}{6} - \sum_{i=1}^{K-1} \frac{1}{i^2} \right] \right), K = \max\{M, N\}, k = \min\{M, N\}.$$

Approximate distribution: $\mathcal{I} \sim \mathcal{N}\left(\mu_{MN}, \sigma^2_{MN}\right)$.
For the special case of fixed $\beta = N/M$, simplified approximations can be obtained.

Proof of (1). (for the first case: M is fixed and $N \to \infty$)

$$\mathcal{I} = \log(\mathbf{I} + \frac{\rho}{M}\mathbf{H}^H\mathbf{H}) = \sum_{m=1}^{M} \log\left(1 + \frac{\rho N}{M}\lambda_m\right), \qquad (2)$$

where $\lambda_1, \cdots, \lambda_M$ are the eigenvalues of $\frac{1}{N}\mathbf{H}^H\mathbf{H}$.

Define $\mathbf{W} \triangleq \mathbf{H}^H\mathbf{H}$, an $M \times M$ ***Wishart matrix***.

- From the law of large numbers,
  $\frac{1}{N}\mathbf{W} \xrightarrow{a.s.} \mathbf{I}$, following which $\tilde{\lambda}_m \triangleq \lambda_m - 1 \xrightarrow{a.s.} 0$.
- With Taylor expansion of (2),

$$\begin{aligned}
\mathcal{I} &= M \log\left(1 + \frac{\rho N}{M}\right) + \sum_{m=1}^{M} \log\left(1 + \frac{\frac{\rho N}{M}\tilde{\lambda}_m}{1 + \frac{\rho N}{M}}\right) \\
&= M \log\left(1 + \frac{\rho N}{M}\right) + \frac{\frac{\rho N}{M}\log e}{1 + \frac{\rho N}{M}} \sum_{m=1}^{M} \tilde{\lambda}_m + \mathcal{O}\left(\sum_{m=1}^{M} \tilde{\lambda}_m^2\right).
\end{aligned}$$

Denote the last term in the above as X, for the convenience.

- From properties of Wishart matrix, it can be shown that

$$\mathbb{E}\left[\sum_{m=1}^{M}\tilde{\lambda}_m\right] = \mathbb{E}\left[\frac{1}{N}\mathrm{tr}(\mathbf{H}^H\mathbf{H}) - M\right] = 0, \quad \mathrm{Var}\left[\sum_{m=1}^{M}\tilde{\lambda}_m\right] = \frac{M}{N},$$

$$\mathbb{E}\left[\sum_{m=1}^{M}\tilde{\lambda}_m^2\right] = \mathbb{E}\left[\sum_{m=1}^{M}(\lambda_m - 1)^2\right] = \frac{M^2}{N}, \quad \mathrm{Var}\left[\sum_{m=1}^{M}\tilde{\lambda}_m^2\right] = \mathcal{O}\left(\frac{1}{N^2}\right).$$

Thus, $\mathbb{E}[X] \sim \mathcal{O}\left(\frac{1}{N}\right), X - \mathbb{E}[X] \sim \mathcal{O}_p\left(\frac{1}{N^2}\right)$, where $\mathcal{O}_p$ represents the scaling in the probability sense (with probability 1).

As $\sum_{m=1}^{M}\tilde{\lambda}_m \sim \mathcal{O}_p\left(\frac{1}{N}\right)$, the effect of X diminishes in the sense of both probability and average when $N \to \infty$.

- Notice that

$$\frac{\sqrt{N}}{\sqrt{M}}\sum_{m=1}^{M}\tilde{\lambda}_m = \frac{1}{\sqrt{MN}}\left(N\sum_{m=1}^{M}\lambda_m - MN\right)$$
$$= \frac{1}{\sqrt{MN}}\left(\mathrm{tr}\mathbf{W} - MN\right) \xrightarrow{d} \mathcal{CN}(0,1),$$

where the last step is due to **central limit theorem**.

- By combining with the observation that

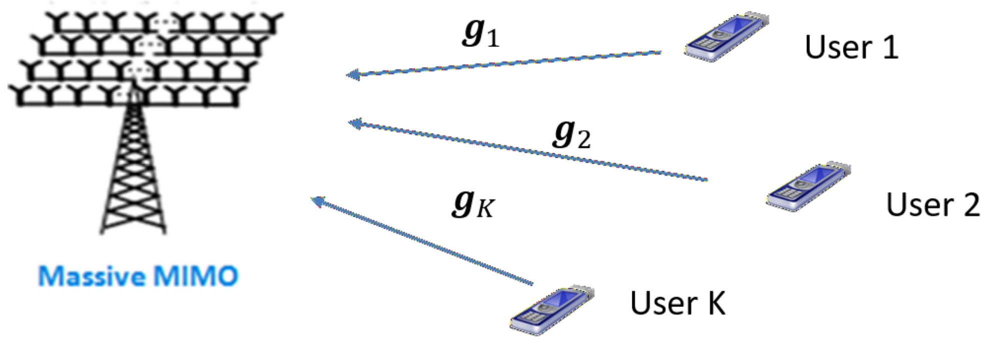
$$\frac{\frac{\rho N}{M}}{1+\frac{\rho N}{M}} = 1 + \mathcal{O}\left(\frac{1}{N}\right),$$

the capacity distribution result in (1) is obtained.

Discussions

- **Convergence in distribution**
- In addition, the mean of the capacity converges to a constant, the variance of the capacity diminishes to 0.
- The **SCV** of the capacity converges to 0 and the convergence rate is linear in $1/N$.
- The capacity distribution results have many applications, including the analysis of the outage probability, the user scheduling gain, required feedback rate, and the number of users needed for scheduling gain. More details are in [4].

2.2. Achievable rate analysis for single-cell multi-user massive MIMO Uplink with MRC



- M antennas at the base station (BS), K single-antenna users
- Channel matrix $(M \times K)$: $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$.
  - $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$: small-scale fading channel matrix. A *circularly symmetric complex Gaussian matrix* [2].
  - $\mathbf{D} = \mathrm{diag}\{[d_2, \cdots, d_K]\}$.
  - $\mathbf{g}_k = \sqrt{d_k}\mathbf{h}_k$: $M \times 1$ channel vector of User k.
  - $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$: the small-scale Rayleigh fading of User k.
  - $\mathbf{h}_i$'s are independent.
  - d_k: large-scale fading coefficient of User k.

- Estimated channel matrix: $\hat{\mathbf{G}} = \hat{\mathbf{H}}\mathbf{D}$.
  - **Model 1:** $\mathbf{H} = \hat{\mathbf{H}} + \Delta\mathbf{H}$, where $\hat{\mathbf{H}}$ and $\Delta\mathbf{H}$ are independent, $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$, $\hat{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, (1-\sigma_e^2)\mathbf{I})$.
    - * σ_e^2 represents the power of channel estimation error, depending on the training length and training power.
    - * Applies to MMSE channel estimation.
  - **Model 2:** $\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$, where $\mathbf{H}$ and $\Delta\mathbf{H}$ are independent, $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$, and <u>thus</u> $\hat{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, (1+\sigma_e^2)\mathbf{I})$.
    - * $1/\sigma_e^2$ represents the average SNR during training, depending on the training length and training power.
    - * Applies to channel estimation by scaling the received signals during training.
- <u>*Use Model 1 here.*</u> Similar analysis applies to Model 2.

Uplink with maximum-ratio combining (MRC)

- Users send information vector $\mathbf{s}$ with per-user power p_u. Entries of $\mathbf{s}$ are i.i.d. with zero-mean and unit-variance.
- Received signal vector at the BS:
$$\mathbf{x} = \sqrt{p_u}\mathbf{G}\mathbf{s} + \mathbf{w},$$
where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the noise vector at the BS.
- MRC at the BS to obtained processed vector $\mathbf{y}$ $(K \times 1)$:
$$\begin{aligned}\mathbf{y} = \hat{\mathbf{G}}^H\mathbf{x} &= \sqrt{p_u}\mathbf{D}^{\frac{1}{2}}\hat{\mathbf{H}}^H\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{s} + \mathbf{D}^{\frac{1}{2}}\hat{\mathbf{H}}^H\mathbf{w} \\ &= \sqrt{p_u}\mathbf{D}^{\frac{1}{2}}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{s} + \sqrt{p_u}\mathbf{D}^{\frac{1}{2}}\hat{\mathbf{H}}^H\Delta\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{s} + \mathbf{D}^{\frac{1}{2}}\hat{\mathbf{H}}^H\mathbf{w}.\end{aligned}$$
The kth element of $\mathbf{y}_k$ (to decode User k's information):
$$\begin{aligned}y_k = \sqrt{p_u}d_k\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k s_k + \sqrt{p_u}\sum_{j\neq k}\sqrt{d_k d_j}\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j s_j \\ +\sqrt{p_u}\sum_j\sqrt{d_k d_j}\hat{\mathbf{h}}_k^H\Delta\mathbf{h}_j s_j + \sqrt{d_k}\hat{\mathbf{h}}_k^H\mathbf{w}.\end{aligned}$$

Signal-to-interference-plus-noise-ratio (SINR) analysis

From the y_k formula on the previous slide, we have

$$\frac{\mathrm{SINR}_k}{M} = \frac{p_u d_k \left(\frac{1}{M} \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k \right)^2}{p_u \sum_{j \neq k} d_j \frac{1}{M} \left| \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_j \right|^2 + p_u \sigma_e^2 \sum_j d_j \left(\frac{1}{M} \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k \right) + \left(\frac{1}{M} \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k \right)} \tag{3}$$

The scalings with M is for each term to have *non-zero & bounded average or limit.*

- **a.s. convergence to deterministic values of some terms** in the SINR formula (3) from law of large numbers. **When $M \to \infty$, for fixed K**: As columns of $\hat{\mathbf{H}}$ are uncorrelated with zero-mean,

$$\frac{1}{M(1-\sigma_e^2)}\hat{\mathbf{H}}^H\hat{\mathbf{H}} \xrightarrow{a.s.} \mathbf{I}_K \Longleftrightarrow \frac{1}{M(1-\sigma_e^2)}\hat{\mathbf{h}}_i^H\hat{\mathbf{h}}_j \xrightarrow{a.s.} \delta_{ij}. \quad (4)$$

  - *Signal term:* From (4) and properties on Pages 15-16,

$$p_u d_k \left(\frac{1}{M}\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k\right)^2 \xrightarrow{a.s.} p_u d_k(1-\sigma_e^2)^2.$$

  - *Channel error and noise terms:* From (4) and Pages 15-16,

$$p_u\sigma_e^2\sum_j d_j\left(\frac{1}{M}\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k\right) \xrightarrow{a.s.} p_u\sigma_e^2(1-\sigma_e^2)\sum_j d_j.$$

$$\left(\frac{1}{M}\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k\right) \xrightarrow{a.s.} (1-\sigma_e^2).$$

- However, **for the interference term, we cannot claim a.s. convergence or m.s. convergence to a deterministic value.**
  - Via straightforward calculations.

$$\mathbb{E}\left[\frac{1}{M}\left|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j\right|^2\right] = 1, \quad \mathrm{Var}\left[\frac{1}{M}\left|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j\right|^2\right] = 1 - \frac{1}{M}.$$

    Notice that *the variance and the SCV do not converge to 0.* The term **does not converge to a deterministic value** in m.s. and is **not asymptotically deterministic** in SCV sense.
  - **a.s. convergence to a deterministic value is also untrue**, since it converge to a random variable with Gamma distribution. From central limit theorem, $\frac{1}{\sqrt{M}}\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j \xrightarrow{d} \mathcal{CN}(0, 1-\sigma_e^2)$. From the continuous function property on Page 15, $\frac{1}{M}\left|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j\right|^2 \xrightarrow{d} Gamma(1,1)$, the Gamma-distribution.

- **SINR result (an approximation) for large M:**

$$\frac{\mathrm{SINR}_k}{M} \approx \frac{d_k}{\sum_{j\neq k} d_j + \frac{\sigma_e^2}{1-\sigma_e^2}\sum_j d_j + \frac{1}{p_u(1-\sigma_e^2)}}.$$

By replacing the interference term with its mean and replace other terms with their deterministic limits.

- **Achievable rate result for large M:**

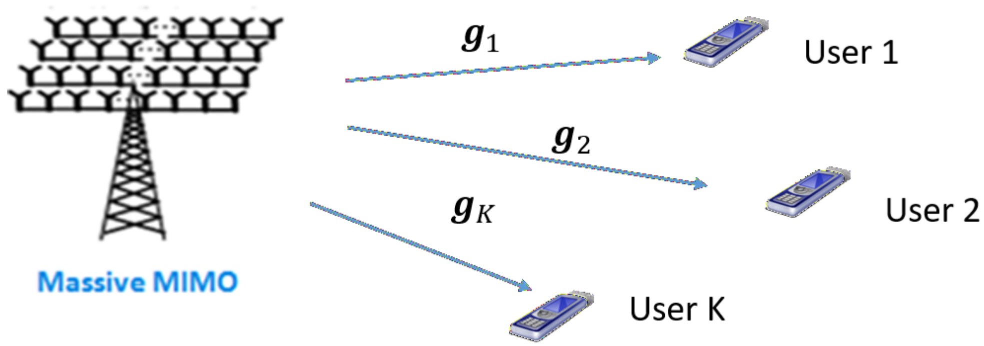
$$R_k \approx \log\left(1 + \frac{M d_k}{\sum_{j\neq k} d_j + \frac{\sigma_e^2}{1-\sigma_e^2}\sum_j d_j + \frac{1}{p_u(1-\sigma_e^2)}}\right).$$

From Jensen's inequality and convexity of $\log(1+1/x)$, the above rate-approximation is a lower bound.

Discussions

- The analysis is for the case of *fixed K* while $M \to \infty$.
- Similar analysis can be done for the downlink with maximum-ratio transmission (MRT).
- Many more general models and variations (e.g., multi-cell systems) have been considered, e.g., [5, 6].
- Can be extended to relay networks with MRT/MRC relaying.
- Analysis for the zero-forcing scheme is more complicated [10, 11].

2.3. Outage probability analysis for massive MIMO downlink [7]



- M antennas at the BS, K single-antenna users
- Channel matrix $(K \times M)$: $\mathbf{H} = [\mathbf{h}_1^t, \cdots, \mathbf{h}_K^t]^t$.
  - $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$: $1 \times M$ channel vector of User k.
  - Assume perfect CSI at the BS.
- BS sends s_k to User k. s_k's i.i.d. zero-mean and unit-variance.
- Signal vector $\mathbf{s} = [s_1, \cdots, s_K]^t$.
- BS transmit power: P_t.

- *maximum-ratio transmission (MRT) precoding* at the BS to produce the transmitted vector $\mathbf{x}$:

$$\mathbf{x} = \sqrt{\frac{P_t}{KM}}\mathbf{H}^H\mathbf{s}.$$

- The received signal vector:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}.$$

where $\mathbf{n}$ is the noise vector with i.i.d. $\mathcal{CN}(0,1)$ entries.

The received signal at the kth user:

$$y_k = \sqrt{\frac{P_t}{KM}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}x_k + \sqrt{\frac{P_t}{KM}}\sum_{j=1, j\neq k}^{K}\mathbf{h}_k\mathbf{h}_j^{\mathrm{H}}x_j + n_k,$$

where $n_k \sim \mathcal{CN}(0,1)$ is the noise.

- SINR:

$$\mathrm{SINR}_k = \frac{P_u M X_k^2}{1 + P_u Y_k},$$

where

$$P_u = \frac{P_t}{K}, \quad X_k \triangleq \frac{1}{M}\mathbf{h}_k\mathbf{h}_k^H, \quad Y_k \triangleq \sum_{j=1, j\neq k}^{K} \frac{1}{M}|\mathbf{h}_k\mathbf{h}_j^{\mathrm{H}}|^2.$$

  - P_u: transmit power per user.
  - Y_k: the scaled interference power.

    The scaling is for non-zero and bounded mean.

Claim: When $M \gg 1$, the probability density function (PDF) of Y_k has the following approximation:

$$f_{Y_k}(y) = (1-\eta)\sum_{i=0}^{\infty}\eta^i\phi\left(y; K+i-1, 1-\frac{1}{\sqrt{M}}\right), \tag{5}$$

where $\phi(y;\alpha,\theta) = \frac{y^{\alpha-1}e^{-y/\theta}}{\theta^{\alpha}(\alpha-1)!}$ for $y > 0$ is the PDF of Gamma distribution with shapa parameter α and scale parameter θ, and

$$\eta \triangleq \frac{K-1}{\sqrt{M}+K-2}.$$

With some rewriting, we have

$$f_{Y_k}(y) = \frac{\sqrt{M}}{\sqrt{M}+K-2}\eta^{-(K-2)}\left[e^{-\frac{\sqrt{M}}{\sqrt{M}+K-2}y} - e^{-\frac{\sqrt{M}}{\sqrt{M}-1}y}\sum_{n=0}^{K-3}\left(\frac{\sqrt{M}}{\sqrt{M}-1}\eta\right)^n\frac{y^n}{n!}\right]. \tag{6}$$

Proof.

- When $M \to \infty$, from **central limit theorem**, for $k \neq j$,

$$\frac{1}{\sqrt{M}}\mathbf{h}_k\mathbf{h}_j^H \xrightarrow{d} \mathcal{CN}(0,1)$$

In this claim, *the definition convergence in distribution is extended to complex random variable* [2] via joint CDF.

- From the continuous function property on Page 15,

$$\frac{1}{M}|\mathbf{h}_k\mathbf{h}_j^H|^2 \xrightarrow{d} Gamma(1,1),$$

where $Gamma(\alpha,\theta)$ represents the Gamma distribution.

- The correlation coefficient of $\frac{1}{M}|\mathbf{h}_k\mathbf{h}_j^H|^2$ and $\frac{1}{M}|\mathbf{h}_k\mathbf{h}_l^H|^2$ for $j \neq l$ can be calculated as $\rho_{jl} = 1/M$.

- So, Y_k can be approximated as sum of $K-1$ correlated Gamma random variables, whose PDF can be shown as (5). See [7] for details.

- With straightforward calculations,

$$\mathbb{E}\left[X_k^2\right] = 1 + \frac{1}{M}, \qquad \mathrm{Var}[X_k^2] = \frac{4}{M} + \mathcal{O}\left(\frac{1}{M}\right),$$
$$\mathbb{E}\left[Y_k\right] = K - 1, \qquad \mathrm{Var}[Y_k] = K - 1 + \frac{(K-1)(K-2)}{M}.$$

- Let γ_{th} be the SINR threshold and $\tilde{\gamma}_{th} \triangleq \gamma_{th}/M$.
- When $M \to \infty$ with $\tilde{\gamma}_{th}$ and K remain fixed/bounded,

$$\begin{aligned} P_{\text{out}} &= \mathbb{P}\left(\mathrm{SINR}_k \le \gamma_{th}\right) = \mathbb{P}\left(\frac{P_u M X_k^2}{1 + P_u Y_k} < \gamma_{th}\right) \\ &= \mathbb{P}\left(\frac{P_u(1+\tilde{X})}{1 + P_u Y_k} < \tilde{\gamma}_{th}\right) \to \mathbb{P}\left(\frac{P_u}{1 + P_u Y_k} < \tilde{\gamma}_{th}\right), \end{aligned} \quad (7)$$

where $\tilde{X} \triangleq X_k^2 - 1$ and the last step is because both the mean and variance of $\tilde{X}$ are $\mathcal{O}(1/M)$, while $\mathrm{Var}[Y_k] > K - 1$.

Outage probability result: By using the approximate PDF in (6) in (7), for $M \gg 1$,

$$P_{out} \approx \eta^{-(K-2)} e^{-\frac{\sqrt{M}}{\sqrt{M}+K-2}\left(\frac{M+1}{\gamma_{th}}-\frac{1}{P_u}\right)} - (1-\eta)\sum_{n=0}^{K-3}\frac{1}{n!}\eta^{n-K+2}\Gamma\left(n+1, \frac{\sqrt{M}}{\sqrt{M}-1}\left(\frac{M+1}{\gamma_{th}}-\frac{1}{P_u}\right)\right),$$

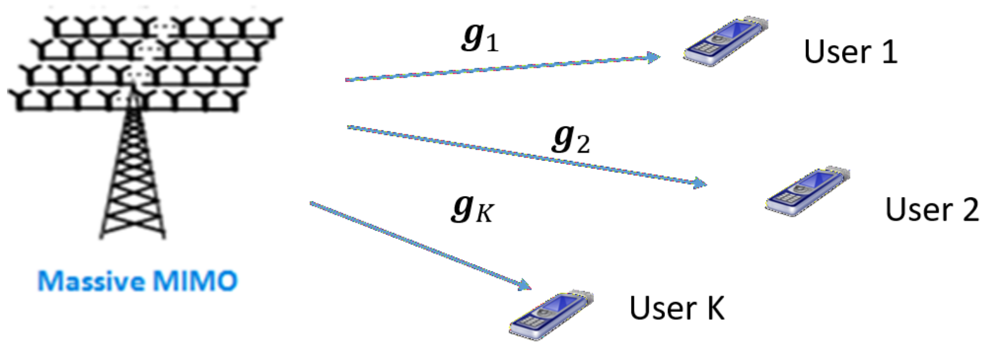
where $\Gamma(z, x)$ is the incomplete gamma function.

Discussions

- The result is an approximation, not a rigorous limiting result.
- For large M, the variance of X_k^2 is $\mathcal{O}(1/M)$, while the variance of Y_k is $\mathcal{O}(1)$. When $M \to \infty$, the signal power becomes deterministic, *the interference power does not* and its variance is significantly larger than that of the signal power.
- The same result can be obtained by approximating X_k^2 with its average, but keep Y_k as a random variable.
- [12] studied the outage probability for massive MIMO uplink with mixed ADCs and the ADC resolution profile optimization.

2.4. Power scaling law for single-cell massive MIMO downlink

The downlink of single-cell multi-user massive MIMO.



- M antennas at the BS, K single-antenna users
- Channel matrix $(K \times M)$ $\mathbf{H} = [\mathbf{h}_1^t, \cdots, \mathbf{h}_K^t]^t$.
  - $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$: $1 \times M$ channel vector of User k.

- Estimated CSI: $\hat{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, Q\mathbf{I})$.
  - $Q \triangleq 1/(1+E_t^{-1})$ representing the quality of the estimated channel.
  - E_t is the energy for the training phase, depending on training length and power.
  - Channel error: $\Delta\mathbf{h} = \mathbf{h} - \hat{\mathbf{h}}$, with i.i.d. $\mathcal{CN}(0, 1-Q)$ entries.
- BS transmit power: P_t.
- BS sends s_k to User k.

  s_k's are i.i.d., zero-mean and unit-variance.

  Signal vector: $\mathbf{s} = [s_2, \cdots, s_K]^t$.

SINR calculations

- *MRT precoding* at the BS to get the transmitted vector $\mathbf{x}$:

$$\mathbf{x} = \sqrt{\frac{P_t}{KMQ}} \hat{\mathbf{H}}^H \mathbf{s}.$$

- The received signal vector $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ is the noise vector with i.i.d. $\mathcal{CN}(0, 1)$ entries.

  The received signal at the kth user:

$$y_k = \sqrt{\frac{P_t}{KMQ}} \hat{\mathbf{h}}_k \mathbf{h}_k^{\mathrm{H}} s_k + \sqrt{\frac{P_t}{KMQ}} \sum_{j=1, j\neq k}^{K} \hat{\mathbf{h}}_k \mathbf{h}_j^{\mathrm{H}} s_j + n_k$$

$$= \sqrt{\frac{P_t}{KMQ}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} s_k + \sqrt{\frac{P_t}{KMQ}} \sum_{j\neq k} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_j^{\mathrm{H}} s_j + \sqrt{\frac{P_t}{KMQ}} \sum_{j=1}^{K} \hat{\mathbf{h}}_k \Delta \mathbf{h}_j^{\mathrm{H}} s_j + n_k,$$

  where $n_k \sim \mathcal{CN}(0, 1)$ is the noise.

- Define the powers of the *(scaled)* signal, interference, CSI error, and noise terms as follows:

$$P_s = \frac{1}{M^2}\left|\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H\right|^2,$$

$$P_{int} = \frac{1}{K-1}\frac{1}{M}\sum_{j\neq k}\left|\hat{\mathbf{h}}_k \hat{\mathbf{h}}_j^H\right|^2,$$

$$P_e = \frac{1-Q}{K}\frac{1}{M}\sum_{j=1}^{K}\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H.$$

  - The normalizations are to have *non-zero and bounded means*.
  - With calculations, we have

$$\mathrm{E}(P_s) = Q^2\left(1+\frac{1}{M}\right) \to Q^2, \qquad \mathrm{SCV}(P_s) = \mathcal{O}\left(\frac{1}{M}\right);$$

$$\mathrm{E}(P_{int}) = Q^2, \qquad \mathrm{SCV}(P_{int}) = \mathcal{O}\left(\frac{1}{M}+\frac{1}{K-1}\right);$$

$$\mathrm{E}(P_e) = Q\left(1-Q\right), \qquad \mathrm{SCV}\{P_e\} = \frac{1-Q}{KM}.$$

- The SINR of User k can be written as

$$\mathrm{SINR}_k = M \frac{P_s}{(K-1)\,P_{int} + KP_e + \frac{KQ}{P_t}}. \tag{8}$$

Scaling law analysis

- To understand the *effect and tradeoff* of system parameters and the *SINR behaviour* with respect to M.

- SINR simplification: From the SCVs on the previous slide, P_s and P_e are asymptotically deterministic in the SCV sense. We can replace then by there averages without changing the convergence behaviour of the SINR. Further, by replacing P_{int} with its averages, the following SINR approximation is obtained.

$$\mathrm{SINR}_k \approx \mathrm{SINR}_{k,L} = \frac{1}{\frac{K}{MQ} - \frac{1}{M} + \frac{K}{MQP_t}}.$$

- Define scaling exponents r_k, r_p, r_t for the **user number, transmit power, and training energy** as follows:

$$K = \mathcal{O}\left(M^{r_k}\right), \frac{1}{P_t} = \mathcal{O}\left(M^{r_p}\right), \frac{1}{E_t} = \mathcal{O}\left(M^{r_t}\right).$$

  – All exponents take values in $[0, 1]$, i.e., $0 \leq r_k, r_p, r_t \leq 1$ for practical massive MIMO systems.

  – They represent how the parameters scale with M, the BS antennas number.
  For example, when $r_k = 0$, the user number K is a constant; and when $r_k = 1$, K increases linearly with M.

- Similarly, define the scaling exponent of the SINR, r_s, as

$$\mathrm{SINR}_{k,L} = \mathcal{O}\left(M^{r_s}\right),$$

  which shows the asymptotic scaling of the SINR in M.

- **Scaling law result**

  For the single-cell massive MIMO downlink with MRT, the performance scaling law is

  $$r_s = 1 - r_t - r_k - r_p. \tag{9}$$

  Proof. Straightforward results from the SINR approximation and the definition of the exponents.

- **Remarks and discussion on typical scenarios.**

  1. Decreasing performance with increasing M (i.e., $r_s < 0$) contradicts the motivations of massive MIMO.

     The *necessary and sufficient condition* for the massive MIMO network to have *non-decreasing SINR* or *favourable SINR* is

     $$r_t + r_k + r_p \leq 1, \quad r_t, r_k, r_p \in [0, 1] \, .$$

     This provides many tradeoff laws. For example, if the training energy decreases linearly in M (i.e., $r_t = 1$), the transmit power and the user number must be constants.

2. For the case of $r_t = 0$, i.e., constant training energy, the scaling law is $r_s = 1 - (r_k + r_p)$.
   Notice that $1/M^{r_k + r_p} = P_t/K$, the *per-user transmission power*.
   The most power-saving design is: the per-user power decreases linearly with M, i.e., $r_k + r_p = 1$, which leads to constant SINR.
   With a larger M, the system can serve more users or consume less power, while maintaining certain SINR performance.
   Improvements in both aspects have limits: 1) $r_k = 1$, $r_p = 0$ and 2) $r_k = 0$, $r_p = 1$.
   Case 1) means: when K increases linearly with M, to achieve non-decreasing SINR, P_t must remain constant, and thus the goal of reducing P_t cannot be achieved.
   Case 2) means: when P_t is inversely proportional to M, the goal of serving more users cannot be achieved.
3. Other scenarios, e.g., decreasing training energy, constant or linearly increasing user number/transmit power, can be studied similarly from the scaling law [8, 9].

Asymptotic determinicity analysis

- Deterministic equivalence has been widely used in massive MIMO analysis. Many were based on a.s. convergence (sometime implicitly).
- **SCV-based definition for asymptotically determinicity** [8, 9].

  Definition. Let $\{X_M\}$ be a random variable sequence with *converging mean*. $\{X_M\}$ is said to be *asymptotically deterministic* if its **SCV** sequence *decreases at least linearly* with M.

  Another possible definition is: $\{X_M\}$ is said to be *asymptotically deterministic* if the **SCV** sequence *converges to 0*. (without condition on the convergence rate).

  For massive MIMO with very large but finite antennas, having linear or faster convergence rate helps the practicality of the results.

- **Result on asymptotically deterministic property**
  With the first definition on the previous slide, a *sufficient condition* for the SINR of the massive MIMO downlink to be **asymptotically deterministic** is

$$2r_t + r_k + 2r_p \geq 1. \tag{10}$$

  Proof. See Appendix B in [9].

- **Remarks and discussion on typical scenarios.**
  1. (10) combined with (9) implies $r_s \leq 1/2$, meaning that to have asymptotically deterministic SINR, the SINR scaling is no high than $\sqrt{M}$.
  2. Linearly increasing user number ($r_k = 1$) means asymptotically deterministic.
  3. To achieve both the highest SINR scaling $r_s = 1/2$ and asymptotically deterministic, the condition reduces to $r_k = 0$ (constant user number) and $r_t + r_p = 1/2$ (product of training energy and transmit power scales as $1/\sqrt{M}$).
  4. More scenarios can be discussed similarly.

Extensions

- The simple single-cell case is illustrated here to illustrate the use of SCV, the scaling law framework, and the asymptotically deterministic concept.
- [8] studied the power scaling law and asymptotic determinicity for multi-user massive MIMO relay networks.
  In addition, the scenario with linearly increasing SINR (not asymptotically deterministic) was studied, where the outage probability and error rate expressions were derived.
- [9] studied the power scaling law and asymptotic determinicity for the multi-cell case with pilot contamination and pilot contamination elimination, correlated channels, and zero-forcing precoding.

References

[1] H. Pishro-Nik, Introduction to Probability, Statistics, and Random Process, available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

[2] Y. Jing and X. Yu, *Some Math Definitions/Results and Applications in MIMO Communications - Topic 1. Complex Random Vector and Circularly Symmetric Complex Gaussian Matrix*, Available at http://www.ece.ualberta.ca/ yindi/MathBackground.html, Jan. 2020.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications,* vol. 10, pp. 585595, Nov. 1999.

[4] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, pp. 1893-1909, Sept. 2004.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[6] T. L. Marzetta, E. G. Larsson, H. Yang, H. Q. Ngo, *Fundamentals of Massive MIMO,* Cambridge University Press, 2016.

[7] C. Feng, Y. Jing, and S. Jin, "Interference and outage probability analysis for massive MIMO downlink with MF precoding," *IEEE Signal Processing Letters*, vol. 23, pp. 366-370, March 2016.

[8] Q. Wang and Y. Jing, "Performance analysis and scaling law of MRC/MRT relaying with CSI error in massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5882-5896, September 2017.

[9] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Performance scaling law for multicell multiuser massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 66, 9890-9903, November 2017.

[10] S. Rahimian, Y. Jing, and M. Ardakani, "Performance analysis of massive MIMO multi-way relay networks with low-resolution ADCs," Submitted to *IEEE Transactions on Wireless Communications*, https://arxiv.org/abs/1804.07254

[11] Q. Ding and Y. Jing, "SE analysis for mixed-ADC massive MIMO uplink with ZF receiver and imperfect CSI," *IEEE Transactions on Wireless Communications Letters*, Dec. 2019. Accepted. https://ieeexplore.ieee.org/document/8926490

[12] Q. Ding and Y. Jing, "Outage probability analysis for massive MIMO uplink with mixed-ADC," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 6293-6306, September 2018.