Random Matrices in Wireless Communications
Course 4: Inverse problems and random matrices: parameter estimation

Romain Couillet
ST-Ericsson, Supélec, FRANCE
romain.couillet@supelec.fr

Supélec

Outline

Outline

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving the eigenvalues (or in fact retrieving anything based on eigenvalues and eigenvectors) is called eigen-inference.
- If the distinct population eigenvalues, i.e. the distinct eigenvalues of $\mathbf{R}$, are small compared to N, much more can be done. This is the purpose of this course.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving the eigenvalues (or in fact retrieving anything based on eigenvalues and eigenvectors) is called eigen-inference.
- If the distinct population eigenvalues, i.e. the distinct eigenvalues of $\mathbf{R}$, are small compared to N, much more can be done. This is the purpose of this course.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \dots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving the eigenvalues (or in fact retrieving anything based on eigenvalues and eigenvectors) is called eigen-inference.
- If the distinct population eigenvalues, i.e. the distinct eigenvalues of $\mathbf{R}$, are small compared to N, much more can be done. This is the purpose of this course.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving the eigenvalues (or in fact retrieving anything based on eigenvalues and eigenvectors) is called eigen-inference.
- If the distinct population eigenvalues, i.e. the distinct eigenvalues of $\mathbf{R}$, are small compared to N, much more can be done. This is the purpose of this course.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving the eigenvalues (or in fact retrieving anything based on eigenvalues and eigenvectors) is called eigen-inference.
- If the distinct population eigenvalues, i.e. the distinct eigenvalues of $\mathbf{R}$, are small compared to N, much more can be done. This is the purpose of this course.

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called, after himself, G-estimators. Among those, we find
  - G_1*-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log\frac{n(n-1)^N}{(n-N)\prod_{k=1}^N(n-k)}\right]$$

with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

in probability.
  - G_3-estimator of the inverse covariance matrix,

$$G_3(\mathbf{R}_n) = \mathbf{R}_n^{-1}[1 - N/n]$$

  - and more than 50 others...
- However, Girko's proofs are rarely readable, if existent.
- As Bai puts it

"his proofs have puzzled many who attempt to understand, without success, Girko's arguments"

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called, after himself, G-estimators. Among those, we find
  - *G_1-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1} \left[\log\det(\mathbf{R}_n) + \log \frac{n(n-1)^N}{(n-N)\prod_{k=1}^N (n-k)} \right]$$

with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

in probability.

  - G_3-estimator of the inverse covariance matrix,

$$G_3(\mathbf{R}_n) = \mathbf{R}_n^{-1}[1 - N/n]$$

  - and more than 50 others...
- However, Girko's proofs are rarely readable, if existent.
- As Bai puts it

"his proofs have puzzled many who attempt to understand, without success, Girko's arguments"

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called, after himself, G-estimators. Among those, we find
  - *G_1-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log\frac{n(n-1)^N}{(n-N)\prod_{k=1}^N(n-k)}\right]$$

    with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

    in probability.
  - G_3-estimator of the inverse covariance matrix,

$$G_3(\mathbf{R}_n) = \mathbf{R}_n^{-1}[1 - N/n]$$

  - and more than 50 others...
- However, Girko's proofs are rarely readable, if existent.
- As Bai puts it

  "his proofs have puzzled many who attempt to understand, without success, Girko's arguments"

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis," http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called, after himself, G-estimators. Among those, we find
  - *G_1-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1} \left[\log\det(\mathbf{R}_n) + \log \frac{n(n-1)^N}{(n-N)\prod_{k=1}^N (n-k)} \right]$$

  with α_n any sequence such that $\alpha_n^{-2} \log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1} \log\det(\mathbf{R}) \to 0$$

  in probability.
  - G_3-estimator of the inverse covariance matrix,

$$G_3(\mathbf{R}_n) = \mathbf{R}_n^{-1}[1 - N/n]$$

  - and more than 50 others...
- However, Girko's proofs are rarely readable, if existent.
- As Bai puts it

"his proofs have puzzled many who attempt to understand, without success, Girko's arguments"

Outline

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^H\mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c\int \frac{t}{1+t\underline{m}_N(z)}dF^{\mathbf{T}_N}(t)\right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^H\mathbf{T}_N\mathbf{X}_N$.

- when $\mathbf{T}_N$ has eigenvalues $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{t_k}{1+t_k\underline{m}_N(z)}\right)^{-1}$$

- an N, n-consistent estimator for the t_k's was never found until recently...
- however, moment-based methods and free probability approaches provide simple solutions to estimate consistently all moments of $F^{\mathbf{T}_N}$.

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^H\mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c\int \frac{t}{1+t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t)\right)^{-1}$$

with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^H\mathbf{T}_N\mathbf{X}_N$.

- when $\mathbf{T}_N$ has eigenvalues $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{t_k}{1+t_k\underline{m}_N(z)}\right)^{-1}$$

- an N, n-consistent estimator for the t_k's was never found until recently...
- however, moment-based methods and free probability approaches provide simple solutions to estimate consistently all moments of $F^{\mathbf{T}_N}$.

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^H \mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$.
- when $\mathbf{T}_N$ has eigenvalues $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N} \sum_{k=1}^{K} n_k \frac{t_k}{1 + t_k \underline{m}_N(z)} \right)^{-1}$$

- an N, n-consistent estimator for the t_k's was never found until recently...
- however, moment-based methods and free probability approaches provide simple solutions to estimate consistently all moments of $F^{\mathbf{T}_N}$.

Reminder on moment-based approaches

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A}) C_{|V_2|}(\mathbf{B})$$

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)^{\mathsf{H}}$$

  whose e.s.d. converges weakly and almost surely to μ_B such that

$$\mu_B = \left(\left(\mu_\Gamma \boxslash \mu_c\right) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $\Gamma_N = \mathbf{R}_N \mathbf{R}_N^{\mathsf{H}}$.
- all basic matrix operations needed in wireless communications are accessible for convenient matrices (Gaussian, Vandermonde etc.)
- all operations are merely polynomial operations on the moments. As a consequence, for $\mathbf{B}_N = f(\mathbf{R}_N)$,

all moments of the l.s.d. of $\mathbf{B}_N$ are obtained as polynomials of those of the l.s.d. of $\mathbf{R}_N$

Reminder on moment-based approaches

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$
$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$$
$$\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)^{\mathsf{H}}$$

  whose e.s.d. converges weakly and almost surely to μ_B such that

$$\mu_B = \left(\left(\mu_\Gamma \boxslash \mu_c\right) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $\Gamma_N = \mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}$.
- all basic matrix operations needed in wireless communications are accessible for convenient matrices (Gaussian, Vandermonde etc.)
- all operations are merely polynomial operations on the moments. As a consequence, for $\mathbf{B}_N = f(\mathbf{R}_N)$,

all moments of the l.s.d. of $\mathbf{B}_N$ are obtained as polynomials of those of the l.s.d. of $\mathbf{R}_N$

Reminder on moment-based approaches

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)^{\mathsf{H}}$$

  whose e.s.d. converges weakly and almost surely to μ_B such that

$$\mu_B = \left(\left(\mu_\Gamma \boxslash \mu_c\right) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $\Gamma_N = \mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}$.
- all basic matrix operations needed in wireless communications are accessible for convenient matrices (Gaussian, Vandermonde etc.)
- all operations are merely polynomial operations on the moments. As a consequence, for $\mathbf{B}_N = f(\mathbf{R}_N)$,

all moments of the l.s.d. of $\mathbf{B}_N$ are obtained as polynomials of those of the l.s.d. of $\mathbf{R}_N$

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

the k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

- we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^H$

$$\mu_\Gamma = ((\mu_B \boxtimes \mu_c) \boxminus \delta_{\sigma^2}) \boxslash \mu_c$$

- for more involved models, the polynomial relations can be iterated and even automatically generated.

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

the k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

- we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)^{\mathrm{H}}$

$$\mu_{\Gamma} = \left(\left(\mu_B \boxtimes \mu_c\right) \boxminus \delta_{\sigma^2}\right) \boxslash \mu_c$$

- for more involved models, the polynomial relations can be iterated and even automatically generated.

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

the k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

  - we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

  and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^H$

$$\mu_{\Gamma} = ((\mu_B \boxtimes \mu_c) \boxminus \delta_{\sigma^2}) \boxslash \mu_c$$

- for more involved models, the polynomial relations can be iterated and even automatically generated.

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$, $\mathbf{D} \in \mathbb{C}^{N \times n}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n \to \infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N} \mathbf{Y}\mathbf{Y}^{\mathrm{H}})^k$$

$$D_k = \lim_{n \to \infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N} \mathbf{D}\mathbf{D}^{\mathrm{H}})^k$$

- For that model, we have the relations

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2 + 2c)D_1 + (1 + c)$$
$$M_3 = D_3 + (3 + 3c)D_2 + 3c{D_1}^2 + (1 + 3c + c^2)$$

hence

$$D_1 = M_1 - 1$$
$$D_2 = M_2 - (2 + 2c)M_1 + (1 + c)$$
$$D_3 = M_3 - (3 + 3c)M_2 - 3c{M_1}^2 + (6c^2 + 18c + 6)M_1 - (4c^2 + 12c + 4)$$

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N\times n}$, $\mathbf{D} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n\to\infty} \frac{1}{n}\operatorname{tr}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})^k$$

$$D_k = \lim_{n\to\infty} \frac{1}{n}\operatorname{tr}(\frac{1}{N}\mathbf{D}\mathbf{D}^{\mathrm{H}})^k$$

- For that model, we have the relations

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2+2c)D_1 + (1+c)$$
$$M_3 = D_3 + (3+3c)D_2 + 3cD_1{}^2 + (1+3c+c^2)$$

hence

$$D_1 = M_1 - 1$$
$$D_2 = M_2 - (2+2c)M_1 + (1+c)$$
$$D_3 = M_3 - (3+3c)M_2 - 3cM_1{}^2 + (6c^2+18c+6)M_1 - (4c^2+12c+4)$$

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N\times n}$, $\mathbf{D} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})^k$$

$$D_k = \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N}\mathbf{D}\mathbf{D}^{\mathsf{H}})^k$$

- For that model, we have the relations

$$\begin{aligned} M_1 &= D_1 + 1 \\ M_2 &= D_2 + (2+2c)D_1 + (1+c) \\ M_3 &= D_3 + (3+3c)D_2 + 3c{D_1}^2 + (1+3c+c^2) \end{aligned}$$

hence

$$\begin{aligned} D_1 &= M_1 - 1 \\ D_2 &= M_2 - (2+2c)M_1 + (1+c) \\ D_3 &= M_3 - (3+3c)M_2 - 3c{M_1}^2 + (6c^2+18c+6)M_1 - (4c^2+12c+4) \end{aligned}$$

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, “Finite Dimensional Statistical Inference,” *submitted to IEEE Trans. on Information Theory*.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]$$

  instead of their limits.
- we have further combinatorics theorems for all previous elementary models.
- *example:* the previous relations extend to

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2 + 2c)D_1 + (1 + c)$$
$$M_3 = D_3 + (3 + 3c)D_2 + 3cD_1^2 + (3 + 9c + 3c^2 + 3N^{-2})D_1 + (1 + 3c + c^2 + N^{-2})$$

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, “Finite Dimensional Statistical Inference,” *submitted to IEEE Trans. on Information Theory*.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]$$

  instead of their limits.
- we have further combinatorics theorems for all previous elementary models.
- *example:* the previous relations extend to

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2+2c)D_1 + (1+c)$$
$$M_3 = D_3 + (3+3c)D_2 + 3cD_1^2 + (3+9c+3c^2+3N^{-2})D_1 + (1+3c+c^2+N^{-2})$$

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, “Finite Dimensional Statistical Inference,” *submitted to IEEE Trans. on Information Theory*.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]$$

  instead of their limits.
- we have further combinatorics theorems for all previous elementary models.
- *example:* the previous relations extend to

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2+2c)D_1 + (1+c)$$
$$M_3 = D_3 + (3+3c)D_2 + 3cD_1^2 + (3+9c+3c^2+3N^{-2})D_1 + (1+3c+c^2+N^{-2})$$

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k - \mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k\right]\right]$$

- if the moments have Gaussian distributions (left to be proven for models other than sample covariance matrix), the full behaviour of the empirical moments is known.
- statistical maximum-likelihood/MMSE methods can then be used, see further Section 35.

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}-\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]\right]$$

- if the moments have Gaussian distributions (left to be proven for models other than sample covariance matrix), the full behaviour of the empirical moments is known.
- statistical maximum-likelihood/MMSE methods can then be used, see further Section 35.

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}-\mathrm{E}\left[\frac{1}{n}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]\right]$$

- if the moments have Gaussian distributions (left to be proven for models other than sample covariance matrix), the full behaviour of the empirical moments is known.
- statistical maximum-likelihood/MMSE methods can then be used, see further Section 35.

Related bibliography

- N. R. Rao, A. Edelman, “The polynomial method for random matrices,” Foundations of Computational Mathematics, *accepted for publication*.
- N. R. Rao, J. A. Mingo, R. Speicher, A. Edelman, “Statistical eigen-inference from large Wishart matrices,” Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- A. Masucci, Ø. Ryan, S. Yang, M. Debbah, “Finite Dimensional Statistical Inference,” *submitted to IEEE Trans. on Information Theory*.
- Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.
- Ø. Ryan, M. Debbah, “Free deconvolution for signal processing applications,” IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.
- Ø. Ryan, M. Debbah, “Asymptotic Behavior of Random Vandermonde Matrices With Entries on the Unit Circle,” IEEE Trans. on Information Theory, vol. 55, no. 7, pp. 3115-3147, 2009.
- I. M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” Annals of Statistics, vol. 99, no. 2, pp. 295-327, 2001.
- K. Johansson, “Shape Fluctuations and Random Matrices,” Comm. Math. Phys. vol. 209, pp. 437-476, 2000.
- Y. Q. Yin, Z. D. Bai, P. R. Krishnaiah, “On the limit of the largest eigenvalue of the large dimensional sample covariance matrix,” Probability Theory and Related Fields, vol. 78, no. 4, pp. 509-521, 1988.
- J. W. Silverstein, Z.D. Bai and Y.Q. Yin, “A note on the largest eigenvalue of a large dimensional sample covariance matrix,” Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168. 1988.
- C. A. Tracy, H. Widom, “On orthogonal and symplectic matrix ensembles,” Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
- P. Bianchi, M. Debbah, J. Najim, “Asymptotic independence in the spectrum of the Gaussian Unitary Ensemble,” arXiv preprint 0811.0979, 2008.

Outline

Advantages of using the Stieltjes transform

- for the same reasons as always, the Stieltjes transform carries all information on the underlying distribution, not only its moments.
- up to now, we obtained N, n-consistent estimators for every moment only
- if distribution is not compactly supported, moment approach is useless
- it would be more helpful to have N, n-consistent estimators of the powers themselves

Advantages of using the Stieltjes transform

- for the same reasons as always, the Stieltjes transform carries all information on the underlying distribution, not only its moments.
- up to now, we obtained N, n-consistent estimators for every moment only
- if distribution is not compactly supported, moment approach is useless
- it would be more helpful to have N, n-consistent estimators of the powers themselves

Advantages of using the Stieltjes transform

- for the same reasons as always, the Stieltjes transform carries all information on the underlying distribution, not only its moments.
- up to now, we obtained N, n-consistent estimators for every moment only
- if distribution is not compactly supported, moment approach is useless
- it would be more helpful to have N, n-consistent estimators of the powers themselves

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- it has long been thought the inverse problem of retrieving $t_1, \ldots, t_K$ from $\mathbf{B}_N$ was not possible.
- the problem was partially solved by Mestre in 2008!
- his technique uses elegant complex analysis tools. The description of this technique is the subject of this course.

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- it has long been thought the inverse problem of retrieving $t_1, \ldots, t_K$ from $\mathbf{B}_N$ was not possible.
- the problem was partially solved by Mestre in 2008!
- his technique uses elegant complex analysis tools. The description of this technique is the subject of this course.

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- it has long been thought the inverse problem of retrieving $t_1, \ldots, t_K$ from $\mathbf{B}_N$ was not possible.
- the problem was partially solved by Mestre in 2008!
- his technique uses elegant complex analysis tools. The description of this technique is the subject of this course.

Reminders of Part 3.

- consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$.
- in Part 3., we saw
  - that there is no eigenvalue outside the support with probability 1.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.
- these results are of crucial importance for the following.

Reminders of Part 3.

- consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$.
- in Part 3., we saw
  - that there is no eigenvalue outside the support with probability 1.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.
- these results are of crucial importance for the following.

Inverse problem for sample covariance matrix

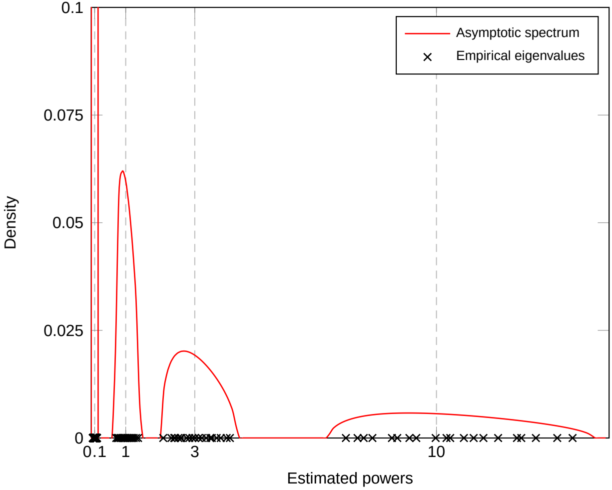

Figure: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

Eigen-inference for the sample covariance matrix model

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

Theorem

Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, *with* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$, *i.i.d. with entries of zero mean, variance* $1/n$, *and* $\mathbf{T}_N \in \mathbb{R}^{N\times N}$ *is diagonal with* K *distinct entries* $t_1,\dots,t_K$ *of multiplicity* $n_1,\dots,n_K$ *of same order as* n. *Let* $k \in \{1,\dots,K\}$. *Then, if the cluster associated to* t_k *is separated from the clusters associated to* $k-1$ *and* $k+1$, *as* $N,n\to\infty$, $N/n\to c$,

$$\hat{t}_k = \frac{n}{n_k}\sum_{m\in\mathcal{N}_k}(\lambda_m - \mu_m)$$

is an N,n*-consistent estimator of* t_k, *where* $\mathcal{N}_k = \{N - \sum_{i=k}^K n_i + 1,\dots,N-\sum_{i=k+1}^K n_i\}$, $\lambda_1,\dots,\lambda_N$ *are the eigenvalues of* $\mathbf{B}_N$ *and* $\mu_1,\dots,\mu_N$ *are the* N *solutions of*

$$m_{\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N}(\mu) = 0$$

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-inference for multi-source power estimation", *submitted to ISIT 2010*.

Lemma [Silverstein]

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

Taking $\mathbf{A} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and $y_i^2 = \frac{1}{N}\lambda_i$, the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathrm{H}}$ are the solutions of

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i - x} = 1$$

which is equivalent to

$$m_{\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N\mathbf{X}_N}(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - x} = 0$$

The μ_k's are then the eigenvalues of a matrix that is function of $\lambda_1, \ldots, \lambda_N$.

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-inference for multi-source power estimation", *submitted to ISIT 2010*.

Lemma [Silverstein]

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be diagonal with entries $\lambda_1, \dots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

Taking $\mathbf{A} = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ and $y_i^2 = \frac{1}{N}\lambda_i$, the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathsf{H}}$ are the solutions of

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i - x} = 1$$

which is equivalent to

$$m_{\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N}(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - x} = 0$$

The μ_k's are then the eigenvalues of a matrix that is function of $\lambda_1, \dots, \lambda_N$.

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-inference for multi-source power estimation", *submitted to ISIT 2010*.

Lemma [Silverstein]

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

Taking $\mathbf{A} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and $y_i^2 = \frac{1}{N}\lambda_i$, the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathsf{H}}$ are the solutions of

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i - x} = 1$$

which is equivalent to

$$m_{\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N}(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - x} = 0$$

The μ_k's are then the eigenvalues of a matrix that is function of $\lambda_1, \ldots, \lambda_N$.

Proof of the lemma

Let $\mathbf{A} \in \mathbb{C}^{N \times N}$ be Hermitian and $\mathbf{y} \in \mathbb{C}^N$. If μ is an eigenvalue of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ with eigenvector $\mathbf{x}$, we have

$$\begin{aligned}
(\mathbf{A} - \mathbf{y}\mathbf{y}^*)\mathbf{x} &= \mu\mathbf{x} \\
(\mathbf{A} - \mu I)x &= \mathbf{y}^* x \mathbf{y} \\
x &= \mathbf{y}^* x (\mathbf{A} - \mu I)^{-1}\mathbf{y} \\
\mathbf{y}^* x &= \mathbf{y}^* x \mathbf{y}^* (\mathbf{A} - \mu I)^{-1}\mathbf{y} \\
1 &= \mathbf{y}^* (\mathbf{A} - \mu I)^{-1}\mathbf{y}
\end{aligned}$$

Take $\mathbf{A}$ diagonal with entries $\lambda_1, \ldots, \lambda_N$, we then have

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - \mu} = 1 \tag{1}$$

Remarks on Mestre's result

- assuming cluster separation, the result consists in
  - taking the empirical *ordered* λ_i's inside the cluster (note that exact separation ensures there are n_k of these!)
  - getting the *ordered* eigenvalues $\mu_1, \ldots, \mu_N$ of

$$\text{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$$

    with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$. Keep only those of index inside $\mathcal{N}_k$.
  - take the difference and scale.

How to obtain this result?

- major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

$$\underline{m}_N(z) = \left(-z - c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.
- before going further, we need some reminders from complex analysis.

How to obtain this result?

- major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

$$\underline{m}_N(z) = \left(-z - c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.
- before going further, we need some reminders from complex analysis.

How to obtain this result?

- major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

$$\underline{m}_N(z) = \left(-z - c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.
- before going further, we need some reminders from complex analysis.

Reminders of complex analysis

W. Rudin, *Real and complex analysis*, McGraw-Hill, 2006.

- Cauchy integration formula

Theorem

Let $U \subset \mathbb{C}$ be an open set and $f : U \to \mathbb{C}$ be holomorphic on U. Let $\gamma \subset U$ be a continuous contour (i.e. closed path). Then, for a inside the surface formed by γ, we have

$$f(a) = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{z - a} dz$$

Reminders of complex analysis (2)

- Residue calculus

Theorem

Let γ be a contour on $\mathbb{C}$. For f holomorphic inside γ but on a discrete number of points, to compute the expression

$$\frac{1}{2\pi i}\oint_\gamma f(z)dz$$

one must

1. *determine the poles of f lying inside the surface formed by γ, i.e. those values a such that*

$$\lim_{z\to a}|f(z)| = \infty$$

2. *determine the order of each pole, i.e. the smallest k such that*

$$\lim_{z\to a}|(z-a)^k f(z)| < \infty$$

3. *compute the residues of f at the poles, i.e. evaluate the value*

$$\mathrm{Res}(f,a) \triangleq \lim_{z\to a}\frac{d^{k-1}}{dz^{k-1}}\left[(z-a)^k f(z)\right]$$

4. *the integral is then the sum of all residues.*

$$\frac{1}{2\pi i}\oint_\gamma f(z)dz = \sum_{a\in\{\text{ poles of } f\}} \mathrm{Res}(f,a)$$

Drawing a contour around t_k

- in the following, we make the cluster separability assumption for t_k, i.e. the cluster corresponding to t_k is separated from those corresponding to t_{k-1} and t_{k+1}.
- from the Cauchy integral formula, for a negatively oriented complex contour $\mathcal{C}_k$ enclosing t_k and only t_k,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} n_r \frac{\omega}{t_r - \omega} d\omega$$

- the idea is then
  1. to choose an appropriate integration contour featuring the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$.
  2. from the resulting expression, use the fact that, for N large, $m_F(z) \simeq m_{\mathbf{B}_N}(z)$, and replace m_F by $m_{\mathbf{B}_N}$
  3. $m_{\mathbf{B}_N}$ is a function of the empirical eigenvalues $\lambda_1, \ldots, \lambda_N$. By residue calculus, we obtain the estimate of t_k.

Drawing a contour around t_k

- in the following, we make the cluster separability assumption for t_k, i.e. the cluster corresponding to t_k is separated from those corresponding to t_{k-1} and t_{k+1}.
- from the Cauchy integral formula, for a negatively oriented complex contour $\mathcal{C}_k$ enclosing t_k and only t_k,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} n_r \frac{\omega}{t_r - \omega} d\omega$$

- the idea is then
  1. to choose an appropriate integration contour featuring the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$.
  2. from the resulting expression, use the fact that, for N large, $m_F(z) \simeq m_{\mathbf{B}_N}(z)$, and replace m_F by $m_{\mathbf{B}_N}$
  3. $m_{\mathbf{B}_N}$ is a function of the empirical eigenvalues $\lambda_1, \ldots, \lambda_N$. By residue calculus, we obtain the estimate of t_k.

Drawing a contour around t_k

- in the following, we make the cluster separability assumption for t_k, i.e. the cluster corresponding to t_k is separated from those corresponding to t_{k-1} and t_{k+1}.
- from the Cauchy integral formula, for a negatively oriented complex contour $\mathcal{C}_k$ enclosing t_k and only t_k,

$$t_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{t_r - \omega} d\omega$$

- the idea is then
  1. to choose an appropriate integration contour featuring the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$.
  2. from the resulting expression, use the fact that, for N large, $m_F(z) \simeq m_{\mathbf{B}_N}(z)$, and replace m_F by $m_{\mathbf{B}_N}$
  3. $m_{\mathbf{B}_N}$ is a function of the empirical eigenvalues $\lambda_1, \ldots, \lambda_N$. By residue calculus, we obtain the estimate of t_k.

Drawing a contour around t_k

- in the following, we make the cluster separability assumption for t_k, i.e. the cluster corresponding to t_k is separated from those corresponding to t_{k-1} and t_{k+1}.
- from the Cauchy integral formula, for a negatively oriented complex contour $\mathcal{C}_k$ enclosing t_k and only t_k,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} n_r \frac{\omega}{t_r - \omega} d\omega$$

- the idea is then
  1. to choose an appropriate integration contour featuring the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$.
  2. from the resulting expression, use the fact that, for N large, $m_F(z) \simeq m_{\mathbf{B}_N}(z)$, and replace m_F by $m_{\mathbf{B}_N}$
  3. $m_{\mathbf{B}_N}$ is a function of the empirical eigenvalues $\lambda_1, \ldots, \lambda_N$. By residue calculus, we obtain the estimate of t_k.

Drawing a contour around t_k

- in the following, we make the cluster separability assumption for t_k, i.e. the cluster corresponding to t_k is separated from those corresponding to t_{k-1} and t_{k+1}.
- from the Cauchy integral formula, for a negatively oriented complex contour $\mathcal{C}_k$ enclosing t_k and only t_k,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} n_r \frac{\omega}{t_r - \omega} d\omega$$

- the idea is then
  1. to choose an appropriate integration contour featuring the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$.
  2. from the resulting expression, use the fact that, for N large, $m_F(z) \simeq m_{\mathbf{B}_N}(z)$, and replace m_F by $m_{\mathbf{B}_N}$
  3. $m_{\mathbf{B}_N}$ is a function of the empirical eigenvalues $\lambda_1, \ldots, \lambda_N$. By residue calculus, we obtain the estimate of t_k.

Contour selection

- *very naive idea*: take any contour around t_k, as close as we want.
- However, the Stieltjes transform is ill-defined close to the real axis in the support!
- *naive idea*: take any contour around the cluster of t_k
- However, it is not true that t_k is inside its own cluster!
- *bright idea*: remember the inversion formula of the Stieltjes transform for the *conjugate* sample covariance matrix $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$,

$$z_N(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dF^{\mathbf{T}_N}(t)$$

and study again the graph of $x_N(m)$ its restriction to the real line...

Contour selection

- *very naive idea*: take any contour around t_k, as close as we want.
- However, the Stieltjes transform is ill-defined close to the real axis in the support!
- *naive idea*: take any contour around the cluster of t_k
- However, it is not true that t_k is inside its own cluster!
- *bright idea*: remember the inversion formula of the Stieltjes transform for the *conjugate* sample covariance matrix $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$,

$$z_N(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dF^{\mathbf{T}_N}(t)$$

and study again the graph of $x_N(m)$ its restriction to the real line...

Contour selection

- *very naive idea*: take any contour around t_k, as close as we want.
- However, the Stieltjes transform is ill-defined close to the real axis in the support!
- *naive idea*: take any contour around the cluster of t_k
- However, it is not true that t_k is inside its own cluster!
- *bright idea*: remember the inversion formula of the Stieltjes transform for the *conjugate* sample covariance matrix $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$,

$$z_N(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dF^{\mathbf{T}_N}(t)$$

and study again the graph of $x_N(m)$ its restriction to the real line...

Contour selection

- *very naive idea*: take any contour around t_k, as close as we want.
- However, the Stieltjes transform is ill-defined close to the real axis in the support!
- *naive idea*: take any contour around the cluster of t_k
- However, it is not true that t_k is inside its own cluster!
- *bright idea*: remember the inversion formula of the Stieltjes transform for the *conjugate* sample covariance matrix $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$,

$$z_N(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dF^{\mathbf{T}_N}(t)$$

and study again the graph of $x_N(m)$ its restriction to the real line...

Contour selection

- *very naive idea*: take any contour around t_k, as close as we want.
- However, the Stieltjes transform is ill-defined close to the real axis in the support!
- *naive idea*: take any contour around the cluster of t_k
- However, it is not true that t_k is inside its own cluster!
- *bright idea*: remember the inversion formula of the Stieltjes transform for the *conjugate* sample covariance matrix $\underline{\mathbf{B}}_N = \mathbf{X}_N^H \mathbf{T}_N \mathbf{X}_N$,

$$z_N(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dF^{\mathbf{T}_N}(t)$$

and study again the graph of $x_N(m)$ its restriction to the real line...

Inverse formula for the Stieltjes transform

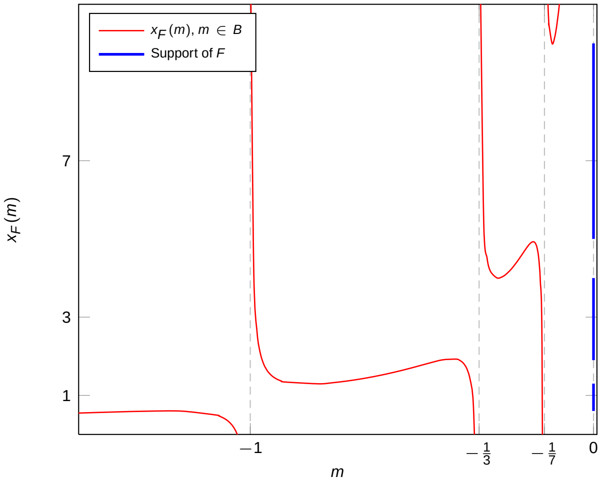

Figure: $x_{\underline{F}}(m)$, with $\underline{F}$ the l.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N\mathbf{X}_N$ with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, whenever $x_{\underline{F}}(m)$ is not increasing.

Playing with the asymptotes...

- remember that the clusters edges x_k^-, x_k^+ correspond to $x_k^- = x_N(m_k^-)$ and $x_k^+ = x_N(m_k^+)$ such that $x_N'(m_k^-) = x_N'(m_k^+) = 0$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour that crosses the real line (slightly on the left of) $-\frac{1}{m_k^-}$ and (slightly on the right of) $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Playing with the asymptotes. . .

- remember that the clusters edges x_k^-, x_k^+ correspond to $x_k^- = x_N(m_k^-)$ and $x_k^+ = x_N(m_k^+)$ such that $x_N'(m_k^-) = x_N'(m_k^+) = 0$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour that crosses the real line (slightly on the left of) $-\frac{1}{m_k^-}$ and (slightly on the right of) $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Playing with the asymptotes...

- remember that the clusters edges x_k^-, x_k^+ correspond to $x_k^- = x_N(m_k^-)$ and $x_k^+ = x_N(m_k^+)$ such that $x_N'(m_k^-) = x_N'(m_k^+) = 0$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour that crosses the real line (slightly on the left of) $-\frac{1}{m_k^-}$ and (slightly on the right of) $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Key step: change of variable

- consider:
  - two reals $\bar{x}_k^- = x_k^- - \varepsilon$ and $\bar{x}_k^+ = x_k^+ + \varepsilon$
  - any parametric curve $\bar{\Gamma}_k \subset \mathbb{C}$ such that

$$\bar{\Gamma}_k(0) = \bar{x}_k^-,\ \bar{\Gamma}_k(1) = \bar{x}_k^+,\ \bar{\Gamma}_k((0,1)) \subset \mathbb{C}^+$$

- with $m_N(z)$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}(z)$, define

$$\mathcal{C}_k = -1/m_N(\bar{\Gamma}_k) \cup -1/m_N(\bar{\Gamma}_k^*)$$

- denoting Γ_k the surface enclosed by $\bar{\Gamma}_k \cup \bar{\Gamma}_k^*$ properly oriented, we have,

$$t_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}w\underline{m}_N(w) + \frac{N-n}{n}\right)\frac{\underline{m}_N'(w)}{\underline{m}_N(w)^2}dw$$

- next figure presents the contour obtained by letting w move along a rectangle closely surrounding the real line.

Key step: change of variable

- consider:
  - two reals $\bar{x}_k^- = x_k^- - \varepsilon$ and $\bar{x}_k^+ = x_k^+ + \varepsilon$
  - any parametric curve $\bar{\Gamma}_k \subset \mathbb{C}$ such that

$$\bar{\Gamma}_k(0) = \bar{x}_k^-, \ \bar{\Gamma}_k(1) = \bar{x}_k^+, \ \bar{\Gamma}_k((0,1)) \subset \mathbb{C}^+$$

- with $m_N(z)$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}(z)$, define

$$\mathcal{C}_k = -1/m_N(\bar{\Gamma}_k) \cup -1/m_N(\bar{\Gamma}_k^*)$$

- denoting Γ_k the surface enclosed by $\bar{\Gamma}_k \cup \bar{\Gamma}_k^*$ properly oriented, we have,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\partial \Gamma_k} \left(\frac{N}{n} w \underline{m}_N(w) + \frac{N-n}{n} \right) \frac{\underline{m}_N'(w)}{\underline{m}_N(w)^2} dw$$

- next figure presents the contour obtained by letting w move along a rectangle closely surrounding the real line.

Key step: change of variable

- consider:
  - two reals $\bar{x}_k^- = x_k^- - \varepsilon$ and $\bar{x}_k^+ = x_k^+ + \varepsilon$
  - any parametric curve $\bar{\Gamma}_k \subset \mathbb{C}$ such that

$$\bar{\Gamma}_k(0) = \bar{x}_k^-,\ \bar{\Gamma}_k(1) = \bar{x}_k^+,\ \bar{\Gamma}_k((0,1)) \subset \mathbb{C}^+$$

- with $m_N(z)$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}(z)$, define

$$\mathcal{C}_k = -1/m_N(\bar{\Gamma}_k) \cup -1/m_N(\bar{\Gamma}_k^*)$$

- denoting Γ_k the surface enclosed by $\bar{\Gamma}_k \cup \bar{\Gamma}_k^*$ properly oriented, we have,

$$t_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\partial \Gamma_k} \left(\frac{N}{n} w \underline{m}_N(w) + \frac{N-n}{n} \right) \frac{\underline{m}_N'(w)}{\underline{m}_N(w)^2} dw$$

- next figure presents the contour obtained by letting w move along a rectangle closely surrounding the real line.

Key step: change of variable

- consider:
  - two reals $\bar{x}_k^- = x_k^- - \varepsilon$ and $\bar{x}_k^+ = x_k^+ + \varepsilon$
  - any parametric curve $\bar{\Gamma}_k \subset \mathbb{C}$ such that

$$\bar{\Gamma}_k(0) = \bar{x}_k^-,\ \bar{\Gamma}_k(1) = \bar{x}_k^+,\ \bar{\Gamma}_k((0,1)) \subset \mathbb{C}^+$$

- with $m_N(z)$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}(z)$, define

$$\mathcal{C}_k = -1/m_N(\bar{\Gamma}_k) \cup -1/m_N(\bar{\Gamma}_k^*)$$

- denoting Γ_k the surface enclosed by $\bar{\Gamma}_k \cup \bar{\Gamma}_k^*$ properly oriented, we have,

$$t_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}w\underline{m}_N(w) + \frac{N-n}{n}\right)\frac{\underline{m}_N'(w)}{\underline{m}_N(w)^2}dw$$

- next figure presents the contour obtained by letting w move along a rectangle closely surrounding the real line.

Selected contour

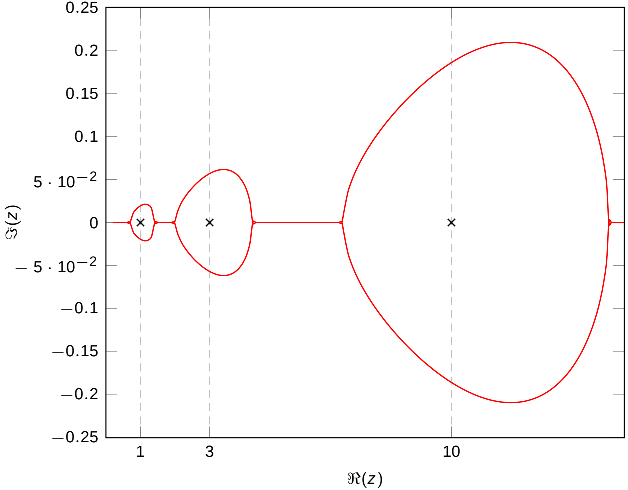

Figure: $z_N(m)$ and $z_N(m)^*$ as a function of m when m describes $(-\infty, \infty) + 10^{-8}i$. $\mathbf{T}_N$ is composed of three distinct entries, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 1/10$.

Random matrix theory in action

- as anticipated, we only need random matrix results once, as follows.
- we have that

$$m_{\underline{\mathbf{B}}_N}(z) - \underline{m}_N(z) \xrightarrow{\text{a.s.}} 0$$

for all z outside the support of $\underline{F}_N$, the distribution of Stieltjes transform $\underline{m}_N$, for all large N.
- on the integration contour, $\underline{m}_N(z)$ is moreover bounded and so, replacing $\underline{m}_N$ by $\hat{\underline{m}}_N$, and denoting

$$\hat{t}_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\partial \Gamma_k} \left(\frac{N}{n} w \hat{\underline{m}}_N(w) + \frac{N-n}{n} \right) \frac{\hat{\underline{m}}_N'(w)}{\hat{\underline{m}}_N(w)^2} dw$$

we have

$$\underline{m}_N(z) - \hat{\underline{m}}_N(z) \xrightarrow{\text{a.s.}} 0$$

- we then have our estimate, which we only need to compute.

Random matrix theory in action

- as anticipated, we only need random matrix results once, as follows.
- we have that

$$m_{\underline{\mathbf{B}}_N}(z) - \underline{m}_N(z) \xrightarrow{\text{a.s.}} 0$$

for all z outside the support of $\underline{F}_N$, the distribution of Stieltjes transform $\underline{m}_N$, for all large N.

- on the integration contour, $\underline{m}_N(z)$ is moreover bounded and so, replacing $\underline{m}_N$ by $\hat{\underline{m}}_N$, and denoting

$$\hat{t}_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}w\hat{\underline{m}}_N(w) + \frac{N-n}{n}\right)\frac{\hat{\underline{m}}'_N(w)}{\hat{\underline{m}}_N(w)^2}dw$$

we have

$$\underline{m}_N(z) - \hat{\underline{m}}_N(z) \xrightarrow{\text{a.s.}} 0$$

- we then have our estimate, which we only need to compute.

Random matrix theory in action

- as anticipated, we only need random matrix results once, as follows.
- we have that

$$m_{\underline{\mathbf{B}}_N}(z) - \underline{m}_N(z) \xrightarrow{\text{a.s.}} 0$$

for all z outside the support of $\underline{F}_N$, the distribution of Stieltjes transform $\underline{m}_N$, for all large N.
- on the integration contour, $\underline{m}_N(z)$ is moreover bounded and so, replacing $\underline{m}_N$ by $\hat{\underline{m}}_N$, and denoting

$$\hat{t}_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\partial \Gamma_k} \left(\frac{N}{n} w \hat{\underline{m}}_N(w) + \frac{N-n}{n} \right) \frac{\hat{\underline{m}}_N'(w)}{\hat{\underline{m}}_N(w)^2} dw$$

we have

$$\underline{m}_N(z) - \hat{\underline{m}}_N(z) \xrightarrow{\text{a.s.}} 0$$

- we then have our estimate, which we only need to compute.

Random matrix theory in action

- as anticipated, we only need random matrix results once, as follows.
- we have that

$$m_{\underline{\mathbf{B}}_N}(z) - \underline{m}_N(z) \xrightarrow{\text{a.s.}} 0$$

for all z outside the support of $\underline{F}_N$, the distribution of Stieltjes transform $\underline{m}_N$, for all large N.

- on the integration contour, $\underline{m}_N(z)$ is moreover bounded and so, replacing $\underline{m}_N$ by $\underline{\hat{m}}_N$, and denoting

$$\hat{t}_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\partial\Gamma_k} \left(\frac{N}{n} w \underline{\hat{m}}_N(w) + \frac{N-n}{n} \right) \frac{\underline{\hat{m}}_N'(w)}{\underline{\hat{m}}_N(w)^2} dw$$

we have

$$\underline{m}_N(z) - \underline{\hat{m}}_N(z) \xrightarrow{\text{a.s.}} 0$$

- we then have our estimate, which we only need to compute.

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- residue calculus, denote $f(w) = \left(\frac{N}{n} w \underline{\hat{m}}_N(w) + \frac{N-n}{n}\right) \frac{\underline{\hat{m}}'_N(w)}{\underline{\hat{m}}_N(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z \to \lambda_k} (z - \lambda_k) f(z) = -\frac{N}{n} \lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z \to \mu_k} (z - \lambda_k) f(z) = \lim_{z \to \mu_k} \frac{N}{n} \frac{(z - \mu_k) z m'(z)}{m(z)} = \frac{N}{n} \mu_k$$

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- residue calculus, denote $f(w) = \left(\frac{N}{n} w \underline{\hat{m}}_N(w) + \frac{N-n}{n}\right) \frac{\underline{\hat{m}}_N'(w)}{\underline{\hat{m}}_N(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z \to \lambda_k} (z - \lambda_k) f(z) = -\frac{N}{n} \lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z \to \mu_k} (z - \lambda_k) f(z) = \lim_{z \to \mu_k} \frac{N}{n} \frac{(z - \mu_k) z m'(z)}{m(z)} = \frac{N}{n} \mu_k$$

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- based on the asymptotes of

$$\underline{m}_N(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z}$$

we have

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

- what about μ_1? the trick is to use the fact that

$$\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{1}{z}dz = 0$$

which leads to

$$\frac{1}{2\pi i}\oint \partial\Gamma_k \frac{\underline{m}_N'(w)}{\underline{m}_N(w)}dw = 0$$

the empirical version of which is

$$\#\{i : \lambda_i \in \Gamma_k\} - \#\{i : \mu_i \in \Gamma_k\}$$

Since their difference tends to 0, there are as many λ_k's as μ_k's in the contour, hence μ_1 is asymptotically in the integration contour.

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- based on the asymptotes of

$$\underline{m}_N(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z}$$

we have

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

- what about μ_1? the trick is to use the fact that

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{z} dz = 0$$

which leads to

$$\frac{1}{2\pi i} \oint \partial\Gamma_k \frac{\underline{m}_N'(w)}{\underline{m}_N(w)} dw = 0$$

the empirical version of which is

$$\#\{i : \lambda_i \in \Gamma_k\} - \#\{i : \mu_i \in \Gamma_k\}$$

Since their difference tends to 0, there are as many λ_k's as μ_k's in the contour, hence μ_1 is asymptotically in the integration contour.

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- based on the asymptotes of

$$\underline{m}_N(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z}$$

we have

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

- what about μ_1? the trick is to use the fact that

$$\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{1}{z}dz = 0$$

which leads to

$$\frac{1}{2\pi i}\oint \partial\Gamma_k \frac{\underline{m}_N'(w)}{\underline{m}_N(w)}dw = 0$$

the empirical version of which is

$$\#\{i : \lambda_i \in \Gamma_k\} - \#\{i : \mu_i \in \Gamma_k\}$$

Since their difference tends to 0, there are as many λ_k's as μ_k's in the contour, hence μ_1 is asymptotically in the integration contour.

Outline

Situation

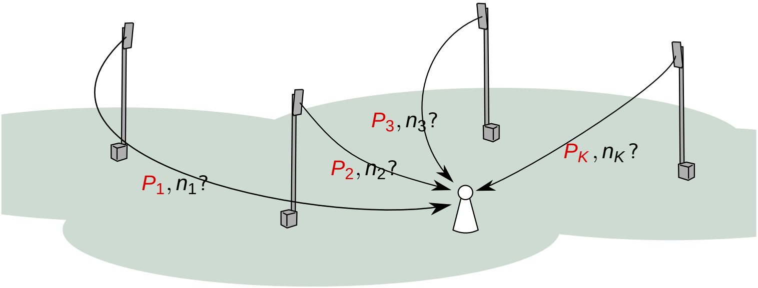

Figure: Power inference scenario

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{Y}\mathbf{Y}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{YY}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{YY}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{YY}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{Y}\mathbf{Y}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

Problem statement

- a device embedded with N antennas receive a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $F^{\mathbf{YY}^H}$, infer those of $F^{\mathbf{P}}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. extend Mestre's approach

System model

- at time t, source k transmit signal $\mathbf{x}_k^{(t)} \in \mathbb{C}^{n_k}$ with i.i.d. entries of zero mean and variance 1.
- we denote P_k the power emitted by user k
- the channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ from user k to the receiver has i.i.d. entries of zero mean and variance $1/N$.
- at time t, the additive noise is denoted $\sigma \mathbf{w}^{(t)}$, with $\mathbf{w}^{(t)} \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance 1.
- hence the receive signal $\mathbf{y}^{(t)}$ at time t,

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k} \mathbf{x}_k^{(t)} + \sigma \mathbf{w}_k^{(t)}$$

Gathering M time instant into $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_M] \in \mathbb{C}^{N \times M}$, this can be written

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma \mathbf{W}$$

with $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_K] \in \mathbb{C}^{N \times n}$, $n = \sum_{k=1}^{K} n_k$,
$\mathbf{P} = \mathrm{diag}(P_1, \dots, P_1, P_2, \dots, P_2, \dots, P_K, \dots, P_K)$ where P_k has multiplicity n_k on the diagonal, $\mathbf{X}^{\mathrm{H}} = [\mathbf{X}_1^{\mathrm{H}} \dots \mathbf{X}_K^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{C}^{n \times M}$, $\mathbf{X}_k = [\mathbf{x}_k^{(1)} \dots \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k \times M}$, $\mathbf{W}$ defined similarly.

System model

- at time t, source k transmit signal $\mathbf{x}_k^{(t)} \in \mathbb{C}^{n_k}$ with i.i.d. entries of zero mean and variance 1.
- we denote P_k the power emitted by user k
- the channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ from user k to the receiver has i.i.d. entries of zero mean and variance $1/N$.
- at time t, the additive noise is denoted $\sigma \mathbf{w}^{(t)}$, with $\mathbf{w}^{(t)} \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance 1.
- hence the receive signal $\mathbf{y}^{(t)}$ at time t,

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k} \mathbf{x}_k^{(t)} + \sigma \mathbf{w}_k^{(t)}$$

Gathering M time instant into $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_M] \in \mathbb{C}^{N \times M}$, this can be written

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma \mathbf{W}$$

with $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_K] \in \mathbb{C}^{N \times n}$, $n = \sum_{k=1}^{K} n_k$,
$\mathbf{P} = \mathrm{diag}(P_1, \dots, P_1, P_2, \dots, P_2, \dots, P_K, \dots, P_K)$ where P_k has multiplicity n_k on the diagonal, $\mathbf{X}^{\mathrm{H}} = [\mathbf{X}_1^{\mathrm{H}} \dots \mathbf{X}_K^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{C}^{n \times M}$, $\mathbf{X}_k = [\mathbf{x}_k^{(1)} \dots \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k \times M}$, $\mathbf{W}$ defined similarly.

Free deconvolution approach

- one can infer the moment of $F^{\mathbf{P}}$ from those of $F^{\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$.
- one can deconvolve $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with “deterministic matrix” $\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{Y}\mathbf{Y}^{\mathsf{H}} = (\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

    (the “deterministic” matrix can be taken random as long as it has a l.s.d.)
  - from $\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{H}\mathbf{H}^{\mathsf{H}}$

$$\mathbf{P}\mathbf{H}\mathbf{H}^{\mathsf{H}}$$

Free deconvolution approach

- one can infer the moment of $F^{\mathbf{P}}$ from those of $F^{\mathbf{YY}^{\mathsf{H}}}$.
- one can deconvolve $\mathbf{YY}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

    (the "deterministic" matrix can be taken random as long as it has a l.s.d.)
  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^{\mathsf{H}}$

$$\mathbf{PHH}^{\mathsf{H}}$$

Free deconvolution approach

- one can infer the moment of $F^{\mathbf{P}}$ from those of $F^{\mathbf{YY}^{\mathsf{H}}}$.
- one can deconvolve $\mathbf{YY}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

    (the "deterministic" matrix can be taken random as long as it has a l.s.d.)
  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^{\mathsf{H}}$

$$\mathbf{PHH}^{\mathsf{H}}$$

Free deconvolution approach

- one can infer the moment of $F^{\mathbf{P}}$ from those of $F^{\mathbf{YY}^{\mathsf{H}}}$.
- one can deconvolve $\mathbf{YY}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

    (the "deterministic" matrix can be taken random as long as it has a l.s.d.)
  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^{\mathsf{H}}$

$$\mathbf{PHH}^{\mathsf{H}}$$

Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \boxslash \mu_c \right) \boxminus \delta_{\sigma^2} \right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \boxslash \mu_c \right) \boxminus \delta_{\sigma^2} \right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}} \boxslash \mu_c\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathrm{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathsf{H}}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments m_k of $\mu_{\frac{1}{M}\mathbf{YY}^{\mathsf{H}}}$ as a function of the first moments d_k of $\mu_{\mathbf{P}}$ read

$$\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right)d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)d_2d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)d_1
\end{aligned}$$

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathsf{H}}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments m_k of $\mu_{\frac{1}{M}\mathbf{YY}^{\mathsf{H}}}$ as a function of the first moments d_k of $\mu_{\mathbf{P}}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right)d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)d_2d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)d_1
\end{aligned}
$$

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathsf{H}}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments m_k of $\mu_{\frac{1}{M}\mathbf{YY}^{\mathsf{H}}}$ as a function of the first moments d_k of $\mu_{\mathbf{P}}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right) d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right) d_1^2 \\
&+ \left(2N^{-1}n + 2M^{-1}n\right) d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right) d_3 \\
&+ \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right) d_2 d_1 \\
&+ \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right) d_1^3 \\
&+ \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right) d_2 \\
&+ \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right) d_1^2 \\
&+ \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right) d_1
\end{aligned}
$$

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathsf{H}}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments m_k of $\mu_{\frac{1}{M}\mathbf{YY}^{\mathsf{H}}}$ as a function of the first moments d_k of $\mu_{\mathbf{P}}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right) d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right) d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right) d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right) d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right) d_2 d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right) d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right) d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right) d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right) d_1
\end{aligned}
$$

Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the K solutions of the polynomial

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K$$

where the Π_m's (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the K solutions of the polynomial

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K$$

where the Π_m's (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1,\ldots,P_K$ are given by the K solutions of the polynomial

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K$$

where the Π_m's (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where X is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))$$

  with, for some p, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^{\circ}(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}$$

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where X is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

  with, for some p, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}$$

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where X is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

  with, for some p, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}$$

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

where X is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

with, for some p, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p} \det(\mathbf{C}^{-1}(\mathbf{p})) e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))} d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p})) e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))} d\mathbf{p}}$$

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where X is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))$$

  with, for some p, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^{\circ}(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}$$

Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

Stieltjes transform approach

- remember the matrix model

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W}$$

with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.

- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

and extend it into the matrix

$$\mathbf{Y}_{\rm ext} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

which is a sample covariance matrix model with random covariance matrix.

- since the covariance matrix clearly has an l.s.d., we have that the l.s.d. $\underline{m}(z)$ of $\mathbf{Y}_{\rm ext}^{\rm H}\mathbf{Y}_{\rm ext}$ is the unique solution, for $z \in \mathbb{C}^+$, of

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M}\int \frac{t}{1+t\underline{m}(z)}dH(t)$$

$$= -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int \frac{1}{t-(-\frac{1}{\underline{m}(z)})}dH(t)\right)$$

with H the l.s.d. of $\begin{pmatrix} \mathbf{H}\mathbf{P}\mathbf{H}^{\rm H}+\sigma^2\mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$.

Stieltjes transform approach

- remember the matrix model

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W}$$

  with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.
- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix} \in \mathbb{C}^{N\times M}$$

  and extend it into the matrix

$$\mathbf{Y}_{\rm ext} = \begin{bmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

  which is a sample covariance matrix model with random covariance matrix.
- since the covariance matrix clearly has an l.s.d., we have that the l.s.d. $\underline{m}(z)$ of $\mathbf{Y}_{\rm ext}^{\rm H}\mathbf{Y}_{\rm ext}$ is the unique solution, for $z \in \mathbb{C}^+$, of

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M}\int\frac{t}{1+t\underline{m}(z)}dH(t)$$

$$= -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int\frac{1}{t-(-\frac{1}{\underline{m}(z)})}dH(t)\right)$$

  with H the l.s.d. of $\begin{pmatrix}\mathbf{H}\mathbf{P}\mathbf{H}^{\rm H}+\sigma^2\mathbf{I}_N & 0\\ 0 & 0\end{pmatrix}$.

Stieltjes transform approach

- remember the matrix model

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W}$$

with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.

- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

and extend it into the matrix

$$\mathbf{Y}_{\mathrm{ext}} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

which is a sample covariance matrix model with random covariance matrix.

- since the covariance matrix clearly has an l.s.d., we have that the l.s.d. $\underline{m}(z)$ of $\mathbf{Y}_{\mathrm{ext}}^{\mathrm{H}}\mathbf{Y}_{\mathrm{ext}}$ is the unique solution, for $z \in \mathbb{C}^{+}$, of

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M}\int \frac{t}{1+t\underline{m}(z)}dH(t)$$

$$= -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int \frac{1}{t-(-\frac{1}{\underline{m}(z)})}dH(t)\right)$$

with H the l.s.d. of $\begin{pmatrix} \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{H}}+\sigma^2\mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$.

Second step

- Note now that $\mathbf{HPH}^{\mathrm{H}}$ is also a sample covariance matrix model, and therefore the l.s.d. of $\mathbf{HPH}^{\mathrm{H}}$ has Stieltjes transform $m_1(z)$, solution of the fixed-point equation in m_1

$$z(m_1) = -\frac{1}{m_1} + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{P_k}{1 + P_k m_1}$$

- Now, up to a shift of σ^2 and the addition of n zero eigenvalues, the l.s.d. of $\mathbf{HPH}^{\mathrm{H}}$ is H. More exactly,

$$\int \frac{1}{t-(z+\sigma^2)} dH(t) = \frac{N}{N+n} m_1(z) - \frac{n}{N+n}\frac{1}{z}$$

- reminding the previous equation

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int \frac{1}{t-(-\frac{1}{\underline{m}(z)})} dH(t)\right)$$

we then have the link between $\underline{m}$ and m_1,

$$z = -\frac{N}{M}\frac{1}{\underline{m}(z)^2} m_1(-1/\underline{m}(z) - \sigma^2) + \frac{N-M}{M}\frac{1}{\underline{m}(z)}$$

Second step

- Note now that $\mathbf{HPH}^{\mathsf{H}}$ is also a sample covariance matrix model, and therefore the l.s.d. of $\mathbf{HPH}^{\mathsf{H}}$ has Stieltjes transform $m_1(z)$, solution of the fixed-point equation in m_1

$$z(m_1) = -\frac{1}{m_1} + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{P_k}{1+P_k m_1}$$

- Now, up to a shift of σ^2 and the addition of n zero eigenvalues, the l.s.d. of $\mathbf{HPH}^{\mathsf{H}}$ is H. More exactly,

$$\int \frac{1}{t-(z+\sigma^2)}dH(t) = \frac{N}{N+n}m_1(z) - \frac{n}{N+n}\frac{1}{z}$$

- reminding the previous equation

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int \frac{1}{t-(-\frac{1}{\underline{m}(z)})}dH(t)\right)$$

we then have the link between $\underline{m}$ and m_1,

$$z = -\frac{N}{M}\frac{1}{\underline{m}(z)^2}m_1(-1/\underline{m}(z)-\sigma^2) + \frac{N-M}{M}\frac{1}{\underline{m}(z)}$$

Second step

- Note now that $\mathbf{HPH}^{\mathrm{H}}$ is also a sample covariance matrix model, and therefore the l.s.d. of $\mathbf{HPH}^{\mathrm{H}}$ has Stieltjes transform $m_1(z)$, solution of the fixed-point equation in m_1

$$z(m_1) = -\frac{1}{m_1} + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{P_k}{1 + P_k m_1}$$

- Now, up to a shift of σ^2 and the addition of n zero eigenvalues, the l.s.d. of $\mathbf{HPH}^{\mathrm{H}}$ is H. More exactly,

$$\int \frac{1}{t - (z + \sigma^2)} dH(t) = \frac{N}{N+n} m_1(z) - \frac{n}{N+n}\frac{1}{z}$$

- reminding the previous equation

$$z = -\frac{1}{\underline{m}(z)} + \frac{N+n}{M\underline{m}(z)}\left(1 - \frac{1}{\underline{m}(z)}\int \frac{1}{t - (-\frac{1}{\underline{m}(z)})} dH(t)\right)$$

we then have the link between $\underline{m}$ and m_1,

$$z = -\frac{N}{M}\frac{1}{\underline{m}(z)^2} m_1(-1/\underline{m}(z) - \sigma^2) + \frac{N-M}{M}\frac{1}{\underline{m}(z)}$$

Asymptotic spectrum

R. Couillet, J. W. Silverstein, M. Debbah, “Eigen-inference for multi-source power estimation”, *submitted to ISIT 2010*.

- all together, denoting $f(z) = m_1(-1/\underline{m}(z) - \sigma^2)$, the asymptotic spectrum of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ has Stietljes transform $m(z)$, $z \in \mathbb{C}^+$, such that

$$m(z) = \frac{M}{N}\underline{m}(z) + \frac{M-N}{N}\frac{1}{z}$$

where $\underline{m}(z)$ is the unique solution in $\mathbb{C}^+$ of

$$\frac{1}{\underline{m}(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{k=1}^{K}\frac{n_k P_k}{1 + P_k f(z)}$$

where $f(z)$ is given by

$$f(z) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N} z \underline{m}(z)^2$$

Asymptotic spectrum of $\frac{1}{M}\mathbf{YY}^H$

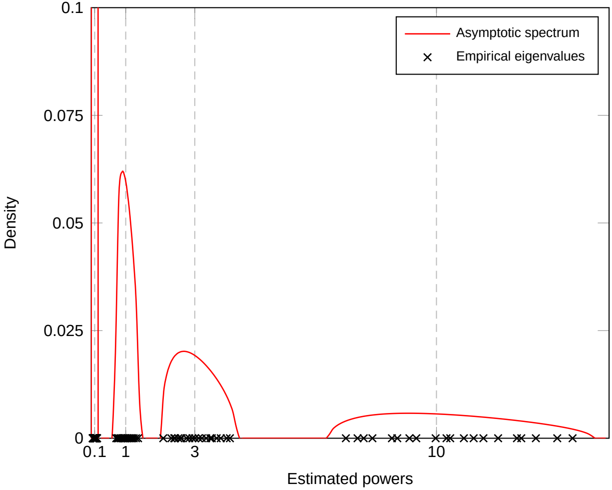

Figure: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{YY}^H$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

Contour definition

- the same approach as for the covariance matrix model can be followed
- assuming separation of cluster k, P_k is comprised between $-1/m_1(x_k^- \varepsilon)$ and $-1/m_1(x_k^+ + \varepsilon)$ for x_k^- and x_k^+ the edges of the k^{th} cluster of the support of $\frac{1}{M}\mathbf{YY}^{\mathrm{H}}$.
- reproducing the steps of Mestre's work, we have, for some contour $\partial\Gamma_k$,

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}wm_1(w) + \frac{N-n}{n}\right)\frac{m_1'(w)}{m_1(w)^2}dw$$

- the key here is to remember that

$$m_1(-1/\underline{m}(z) - \sigma^2) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N}z\underline{m}(z)^2$$

- we then make the variable change $w = -1/\underline{m}(z) - \sigma^2$ to get

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{m}(z)\right) + \frac{N-K}{K}\frac{1}{zm(z)}\right]\left[-\frac{1}{z\underline{m}(z)} - \frac{\underline{m}'(z)}{\underline{m}(z)^2} - \frac{m'(z)}{m(z)\underline{m}(z)}\right]dz$$

Contour definition

- the same approach as for the covariance matrix model can be followed
- assuming separation of cluster k, P_k is comprised between $-1/m_1(x_k^- \varepsilon)$ and $-1/m_1(x_k^+ + \varepsilon)$ for x_k^- and x_k^+ the edges of the k^{th} cluster of the support of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.
- reproducing the steps of Mestre's work, we have, for some contour $\partial\Gamma_k$,

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}wm_1(w) + \frac{N-n}{n}\right)\frac{m_1'(w)}{m_1(w)^2}dw$$

- the key here is to remember that

$$m_1(-1/\underline{m}(z) - \sigma^2) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N}z\underline{m}(z)^2$$

- we then make the variable change $w = -1/\underline{m}(z) - \sigma^2$ to get

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{m}(z)\right) + \frac{N-K}{K}\frac{1}{zm(z)}\right]\left[-\frac{1}{z\underline{m}(z)} - \frac{\underline{m}'(z)}{\underline{m}(z)^2} - \frac{m'(z)}{m(z)\underline{m}(z)}\right]dz$$

Contour definition

- the same approach as for the covariance matrix model can be followed
- assuming separation of cluster k, P_k is comprised between $-1/m_1(x_k^- \varepsilon)$ and $-1/m_1(x_k^+ + \varepsilon)$ for x_k^- and x_k^+ the edges of the k^{th} cluster of the support of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$.
- reproducing the steps of Mestre's work, we have, for some contour $\partial\Gamma_k$,

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}wm_1(w) + \frac{N-n}{n}\right)\frac{m_1'(w)}{m_1(w)^2}dw$$

- the key here is to remember that

$$m_1(-1/\underline{m}(z) - \sigma^2) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N}z\underline{m}(z)^2$$

- we then make the variable change $w = -1/\underline{m}(z) - \sigma^2$ to get

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{m}(z)\right) + \frac{N-K}{K}\frac{1}{z\underline{m}(z)}\right]\left[-\frac{1}{z\underline{m}(z)} - \frac{\underline{m}'(z)}{\underline{m}(z)^2} - \frac{m'(z)}{m(z)\underline{m}(z)}\right]dz$$

Contour definition

- the same approach as for the covariance matrix model can be followed
- assuming separation of cluster k, P_k is comprised between $-1/m_1(x_k^- \varepsilon)$ and $-1/m_1(x_k^+ + \varepsilon)$ for x_k^- and x_k^+ the edges of the k^{th} cluster of the support of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.
- reproducing the steps of Mestre's work, we have, for some contour $\partial\Gamma_k$,

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Gamma_k}\left(\frac{N}{n}wm_1(w) + \frac{N-n}{n}\right)\frac{m_1'(w)}{m_1(w)^2}dw$$

- the key here is to remember that

$$m_1(-1/\underline{m}(z) - \sigma^2) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N}z\underline{m}(z)^2$$

- we then make the variable change $w = -1/\underline{m}(z) - \sigma^2$ to get

$$P_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{m}(z)\right) + \frac{N-K}{K}\frac{1}{zm(z)}\right]\left[-\frac{1}{z\underline{m}(z)} - \frac{\underline{m}'(z)}{\underline{m}(z)^2} - \frac{m'(z)}{m(z)\underline{m}(z)}\right]dz$$

Using RMT and computing the residues

- we now know that $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, is close to its empirical counterpart

$$\hat{m}(z) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k - z}$$

- verifying that $m(z)$ is bounded along the integration contour, we can then replace limiting results by empirical ones, and get

$$\hat{P}_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{\hat{m}}(z)\right)+\frac{N-K}{K}\frac{1}{z\hat{m}(z)}\right]\left[-\frac{1}{z\underline{\hat{m}}(z)}-\frac{\underline{\hat{m}}'(z)}{\underline{\hat{m}}(z)^2}-\frac{\hat{m}'(z)}{\hat{m}(z)\underline{\hat{m}}(z)}\right]dz$$

- residue calculus then leads to 9 terms to be evaluated, the poles of which are at
  - $\eta_1, \ldots, \eta_N$, the solutions to
  $$\sum_{k=1}^{N}\frac{1}{\lambda_k - z} = 0$$
  - $\mu_1, \ldots, \mu_N$, the solutions to
  $$\sum_{k=1}^{M}\frac{1}{\lambda_k - z} = 0$$

  (denoting $\lambda_{N+1} = \ldots = \lambda_M = 0$.)
- proving the presence of the μ_k's and η_k's in the cluster under study is identical to the sample covariance matrix model approach.

Using RMT and computing the residues

- we now know that $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, is close to its empirical counterpart

$$\hat{m}(z) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k - z}$$

- verifying that $m(z)$ is bounded along the integration contour, we can then replace limiting results by empirical ones, and get

$$\hat{P}_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{\hat{m}}(z)\right)+\frac{N-K}{K}\frac{1}{z\hat{m}(z)}\right]\left[-\frac{1}{z\underline{\hat{m}}(z)}-\frac{\underline{\hat{m}}'(z)}{\underline{\hat{m}}(z)^2}-\frac{\hat{m}'(z)}{\hat{m}(z)\underline{\hat{m}}(z)}\right]dz$$

- residue calculus then leads to 9 terms to be evaluated, the poles of which are at
  - $\eta_1,\ldots,\eta_N$, the solutions to

$$\sum_{k=1}^{N}\frac{1}{\lambda_k - z} = 0$$

  - $\mu_1,\ldots,\mu_N$, the solutions to

$$\sum_{k=1}^{M}\frac{1}{\lambda_k - z} = 0$$

    (denoting $\lambda_{N+1} = \ldots = \lambda_M = 0$.)
- proving the presence of the μ_k's and η_k's in the cluster under study is identical to the sample covariance matrix model approach.

Using RMT and computing the residues

- we now know that $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, is close to its empirical counterpart

$$\hat{m}(z) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k - z}$$

- verifying that $m(z)$ is bounded along the integration contour, we can then replace limiting results by empirical ones, and get

$$\hat{P}_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{\hat{m}}(z)\right)+\frac{N-K}{K}\frac{1}{z\hat{m}(z)}\right]\left[-\frac{1}{z\underline{\hat{m}}(z)}-\frac{\underline{\hat{m}}'(z)}{\underline{\hat{m}}(z)^2}-\frac{\hat{m}'(z)}{\hat{m}(z)\underline{\hat{m}}(z)}\right]dz$$

- residue calculus then leads to 9 terms to be evaluated, the poles of which are at
  - $\eta_1,\ldots,\eta_N$, the solutions to

$$\sum_{k=1}^{N}\frac{1}{\lambda_k - z} = 0$$

  - $\mu_1,\ldots,\mu_N$, the solutions to

$$\sum_{k=1}^{M}\frac{1}{\lambda_k - z} = 0$$

  (denoting $\lambda_{N+1} = \ldots = \lambda_M = 0$.)

- proving the presence of the μ_k's and η_k's in the cluster under study is identical to the sample covariance matrix model approach.

Using RMT and computing the residues

- we now know that $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, is close to its empirical counterpart

$$\hat{m}(z) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k - z}$$

- verifying that $m(z)$ is bounded along the integration contour, we can then replace limiting results by empirical ones, and get

$$\hat{P}_k = \frac{n}{n_k}\frac{1}{2\pi i}\oint_{\partial\Omega_k}\left[\frac{N}{n}\left(1+\sigma^2\underline{\hat{m}}(z)\right) + \frac{N-K}{K}\frac{1}{z\hat{m}(z)}\right]\left[-\frac{1}{z\underline{\hat{m}}(z)} - \frac{\underline{\hat{m}}'(z)}{\underline{\hat{m}}(z)^2} - \frac{\hat{m}'(z)}{\hat{m}(z)\underline{\hat{m}}(z)}\right]dz$$

- residue calculus then leads to 9 terms to be evaluated, the poles of which are at
  - $\eta_1, \ldots, \eta_N$, the solutions to

$$\sum_{k=1}^{N}\frac{1}{\lambda_k - z} = 0$$

  - $\mu_1, \ldots, \mu_N$, the solutions to

$$\sum_{k=1}^{M}\frac{1}{\lambda_k - z} = 0$$

  (denoting $\lambda_{N+1} = \ldots = \lambda_M = 0$.)
- proving the presence of the μ_k's and η_k's in the cluster under study is identical to the sample covariance matrix model approach.

Stieltjes transform approach: final result

R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-inference for multi-source power estimation", *submitted to ISIT 2010*.

Theorem

Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$*, with* $\mathbf{Y}$ *defined as previously. Denote its ordered eigenvalues vector* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$*,* $\lambda_1 < \ldots, \lambda_N$*. Further assume asymptotic spectrum separability. Then, for* $k \in \{1, \ldots, K\}$*, as* N*,* n*,* M *grow large, we have*

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0$$

where the estimate $\hat{P}_k$ *is given by*

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i\in\mathcal{N}_k} (\eta_i - \mu_i)$$

with $\mathcal{N}_k = \{N - \sum_{i=k}^{K} n_i + 1, \ldots, N - \sum_{i=k+1}^{K} n_i\}$ *the set of indexes matching the cluster corresponding to* P_k*,* $(\eta_1, \ldots, \eta_N)$ *the ordered eigenvalues of* $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ *and* $(\mu_1, \ldots, \mu_N)$ *the ordered eigenvalues of* $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$*.*

Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed. *Note that this might be good enough!*.
- extension to the case when spectrum separation is not needed is being investigated at the moment.
- supposedly, it is possible to infer K, all n_K's and all P_K's using the Stieltjes transform method.

Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed. *Note that this might be good enough!*.
- extension to the case when spectrum separation is not needed is being investigated at the moment.
- supposedly, it is possible to infer K, all n_K's and all P_K's using the Stieltjes transform method.

Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed. *Note that this might be good enough!*.
- extension to the case when spectrum separation is not needed is being investigated at the moment.
- supposedly, it is possible to infer K, all n_k's and all P_k's using the Stieltjes transform method.

Performance comparison

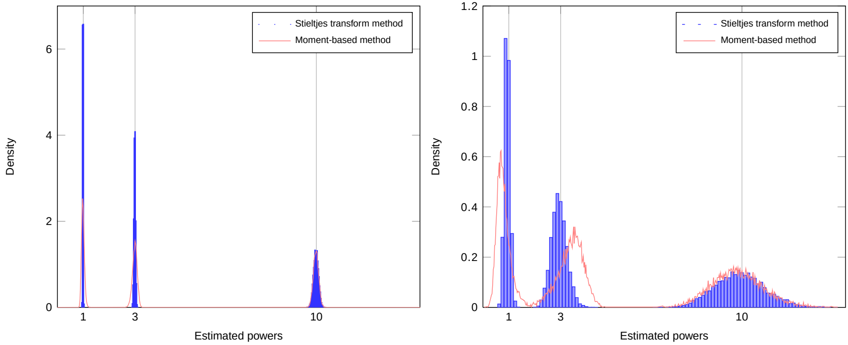

Figure: Multi-source power estimation, for $K = 3, P_1 = 1, P_2 = 3, P_3 = 10, n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, $\mathrm{SNR} = 10$ dB, for $10,000$ simulation runs; Top $n = 60$, bottom $n = 6$.

General comments and steps left to fulfill

- up to this day
  - the moment approach is much simpler to derive
  - it does not require any cluster separation
  - the finite size case is treated in the mean, which the Stieltjes transform approach cannot do.
  - however, the Stieltjes transform approach makes full use of the spectral knowledge, when the moment approach is limited to a few moments.
  - the results are more natural, and more "telling"
- in the future, it is expected that the cluster separation requirement can be overtaken.
- a natural general framework attached to the Stieltjes transform method could arise
- central limit results on the estimates is expected

General comments and steps left to fulfill

- up to this day
  - the moment approach is much simpler to derive
  - it does not require any cluster separation
  - the finite size case is treated in the mean, which the Stieltjes transform approach cannot do.
  - however, the Stieltjes transform approach makes full use of the spectral knowledge, when the moment approach is limited to a few moments.
  - the results are more natural, and more “telling”
- in the future, it is expected that the cluster separation requirement can be overtaken.
- a natural general framework attached to the Stieltjes transform method could arise
- central limit results on the estimates is expected

Related bibliography

- N. El Karoui, “Spectrum estimation for large dimensional covariance matrices using random matrix theory,” Annals of Statistics, vol. 36, no. 6, pp. 2757-2790, 2008.
- N. R. Rao, J. A. Mingo, R. Speicher, A. Edelman, “Statistical eigen-inference from large Wishart matrices,” Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- R. Couillet, M. Debbah, “Free deconvolution for OFDM multicell SNR detection”, PIMRC 2008, Cannes, France.
- X. Mestre, “Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates,” IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.
- R. Couillet, J. W. Silverstein, M. Debbah, “Eigen-inference for multi-source power estimation”, *submitted to ISIT 2010*.
- Z. D. Bai, J. W. Silverstein, “No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices,” The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.
- Z. D. Bai, J. W. Silverstein, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.
- J. Silverstein, Z. Bai, “Exact separation of eigenvalues of large dimensional sample covariance matrices” Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.
- Ø. Ryan, M. Debbah, “Free Deconvolution for Signal Processing Applications,” IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.

Outline

Living at the edge of the available mathematical tools

- the first applications of random matrix theory to wireless communications used to be
  - mere applications of existing theorems
  - limited to system analysis
- today,
  - the scope of applications has widened: multi-user, multi-antenna system analysis, concrete applications in self-organized networks, detection, estimation etc.
  - most of the new studies require new mathematical results

Living at the edge of the available mathematical tools

- the first applications of random matrix theory to wireless communications used to be
  - mere applications of existing theorems
  - limited to system analysis
- today,
  - the scope of applications has widened: multi-user, multi-antenna system analysis, concrete applications in self-organized networks, detection, estimation etc.
  - most of the new studies require new mathematical results

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?

Random matrix theory today and tomorrow

- today, a large number of applications linked to i.i.d. (Gaussian or not) models, with double correlation, variance profile, non-centered, Haar matrices has been treated.
- more structured matrices are more difficult to treat, especially on the analytic side
- results on eigenvectors are also less numerous
- finite size considerations yet limited to moment approaches
- eigen-inference methods need be developed to more involved models and gathered into a unified framework.
- the scalars appearing in fixed-point equations seem to compress the communication channel information to its simplest granularity:
  - is that true?
  - how and for which applications can this be used?