Random Matrices and Machine Learning

Romain Couillet

romain.couillet@gipsa-lab.grenoble-inp.fr

GIPSA-lab, University Grenoble-Alps

March 22, 2021

Outline

Reminders: the sample covariance matrix model

Reminders: The sample covariance matrix model

The SCM model:

- $X_n = (X_{ij}) = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{C}^{N \times n}$ (or $\mathbb{R}^{N \times n}$) with i.i.d. zero mean unit variance entries and

$$\Sigma_n = \frac{1}{\sqrt{n}}[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n] = \frac{1}{\sqrt{n}}[R_N^{1/2}\boldsymbol{x}_1, \cdots, R_N^{1/2}\boldsymbol{x}_n]$$

with $R_N \in \mathbb{C}^{N \times N} \succeq 0$, so that

$$n\mathbb{E}\boldsymbol{s}_1\boldsymbol{s}_1^* = R_N^{1/2}\mathbb{E}\boldsymbol{x}_1\boldsymbol{x}_1^* R_N^{1/2} = R_N.$$

Reminders: The sample covariance matrix model

The SCM model:

- $X_n = (X_{ij}) = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{C}^{N \times n}$ (or $\mathbb{R}^{N \times n}$) with i.i.d. zero mean unit variance entries and

$$\Sigma_n = \frac{1}{\sqrt{n}}[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n] = \frac{1}{\sqrt{n}}[R_N^{1/2}\boldsymbol{x}_1, \cdots, R_N^{1/2}\boldsymbol{x}_n]$$

  with $R_N \in \mathbb{C}^{N \times N} \succeq 0$, so that

$$n\mathbb{E}\boldsymbol{s}_1\boldsymbol{s}_1^* = R_N^{1/2}\mathbb{E}\boldsymbol{x}_1\boldsymbol{x}_1^* R_N^{1/2} = R_N.$$

- R_N is the **population covariance** of the samples $(\sqrt{n}\boldsymbol{s}_1, \cdots, \sqrt{n}\boldsymbol{s}_n)$.

Reminders: The sample covariance matrix model

The SCM model:

- $X_n = (X_{ij}) = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{C}^{N \times n}$ (or $\mathbb{R}^{N \times n}$) with i.i.d. zero mean unit variance entries and

$$\Sigma_n = \frac{1}{\sqrt{n}}[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n] = \frac{1}{\sqrt{n}}[R_N^{1/2}\boldsymbol{x}_1, \cdots, R_N^{1/2}\boldsymbol{x}_n]$$

with $R_N \in \mathbb{C}^{N \times N} \succeq 0$, so that

$$n\mathbb{E}\boldsymbol{s}_1\boldsymbol{s}_1^* = R_N^{1/2}\mathbb{E}\boldsymbol{x}_1\boldsymbol{x}_1^* R_N^{1/2} = R_N.$$

- R_N is the **population covariance** of the samples $(\sqrt{n}\boldsymbol{s}_1, \cdots, \sqrt{n}\boldsymbol{s}_n)$.
- the matrix

$$\hat{R}_N = \frac{1}{n}R_N^{1/2} X_n X_n^* R_N^{1/2} = \Sigma_n \Sigma_n^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{s}_i \boldsymbol{s}_i^*$$

is the **sample covariance matrix**.

Reminders: Limiting eigenvalue distribution

Further notations:

- for $\lambda_1(A) \geq \ldots \lambda_N(A)$ the eigenvalues of Hermitian (symmetric) A:

$$L_N \equiv \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\hat{R}_N)} \quad \text{(random)}$$

$$L_N^R \equiv \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(R_N)} \quad \text{(deterministic)}$$

Reminders: Limiting eigenvalue distribution

Further notations:

- for $\lambda_1(A) \geq \ldots \lambda_N(A)$ the eigenvalues of Hermitian (symmetric) A:

$$L_N \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(\hat{R}_N)} \quad \text{(random)}$$

$$L_N^R \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(R_N)} \quad \text{(deterministic)}$$

- we further denote the Stieltjes transform

$$g_n(z) = \int (\lambda - z)^{-1} L_N(d\lambda) = \frac{1}{N}\mathrm{tr}\left(\hat{R}_N - zI_N\right)^{-1}.$$

Reminders: Limiting eigenvalue distribution

Further notations:

- for $\lambda_1(A) \geq \ldots \lambda_N(A)$ the eigenvalues of Hermitian (symmetric) A:

$$L_N \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(\hat{R}_N)} \quad \text{(random)}$$

$$L_N^R \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(R_N)} \quad \text{(deterministic)}$$

- we further denote the Stieltjes transform

$$g_n(z) = \int (\lambda - z)^{-1} L_N(d\lambda) = \frac{1}{N}\operatorname{tr}\left(\hat{R}_N - zI_N\right)^{-1}.$$

Asymptotics: We suppose that, as $n, N \to \infty$,

- $N/n \equiv c_n \to c > 0$

Reminders: Limiting eigenvalue distribution

Further notations:

- for $\lambda_1(A) \geq \ldots \lambda_N(A)$ the eigenvalues of Hermitian (symmetric) A:

$$L_N \equiv \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\hat{R}_N)} \quad \text{(random)}$$

$$L_N^R \equiv \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(R_N)} \quad \text{(deterministic)}$$

- we further denote the Stieltjes transform

$$g_n(z) = \int (\lambda - z)^{-1} L_N(d\lambda) = \frac{1}{N} \mathrm{tr} \left(\hat{R}_N - z I_N \right)^{-1}.$$

Asymptotics: We suppose that, as $n, N \to \infty$,

- $N/n \equiv c_n \to c > 0$
- $L_N^R \to L_\infty^R$ in distribution.

Reminders: Limiting spectral measure

Theorem (Limiting spectral distribution)

For $z \in \mathbb{C}^+ = \{w \in \mathbb{C},\ \Im[w] > 0\}$, as $n, N \to \infty$,

$$g_n(z) \xrightarrow{\text{a.s.}} t(z)$$

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

$$t(z) = \int \frac{L_\infty^R(du)}{-z(1 + uct(z)) + (1 - c)u}.$$

As a consequence,

$$L_N \xrightarrow{\text{a.s.}} \mathcal{F}$$

with $\mathcal{F}$ the unique probability measure such that $t(z) = \int (t - z)^{-1} \mathcal{F}(dt)$.

Reminders: Limiting spectral measure

Plotting $\mathcal{F}$: Numerically, we evaluate

$$t(x + \imath\varepsilon)$$

for some $\varepsilon \ll 1$ (e.g., $\varepsilon = 10^{-3}$)

Reminders: Limiting spectral measure

Plotting $\mathcal{F}$: Numerically, we evaluate

$$t(x + \imath\varepsilon)$$

for some $\varepsilon \ll 1$ $(\text{e.g., } \varepsilon = 10^{-3})$ and use the approximation

$$d\mathcal{F}(x) \simeq \frac{1}{\pi}\Im[t(x + \imath\varepsilon)]dx$$

Reminders: Limiting spectral measure

Plotting $\mathcal{F}$: Numerically, we evaluate

$$t(x + \imath\varepsilon)$$

for some $\varepsilon \ll 1$ (e.g., $\varepsilon = 10^{-3}$) and use the approximation

$$d\mathcal{F}(x) \simeq \frac{1}{\pi}\Im[t(x + \imath\varepsilon)]dx$$

Remark: This strongly assumes $\mathcal{F}$ has a density !!!

Reminders: Limiting spectral measure

Plotting $\mathcal{F}$: Numerically, we evaluate

$$t(x + \imath\varepsilon)$$

for some $\varepsilon \ll 1$ (e.g., $\varepsilon = 10^{-3}$) and use the approximation

$$d\mathcal{F}(x) \simeq \frac{1}{\pi}\Im[t(x + \imath\varepsilon)]dx$$

Remark: This strongly assumes $\mathcal{F}$ has a density !!!

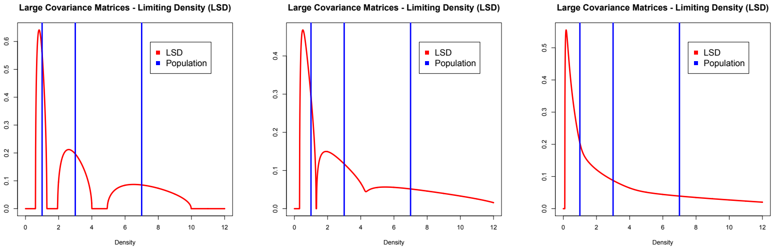


Figure: In red, limiting density for $c = .1$, $c = .3$, $c = .6$. In blue, 3 population eigenvalues of R_N, each of equal multiplicity.

Restating the theorem

The co-resolvent result: In the following it will be more convenient to work with $\Sigma_n^* \Sigma_n \in \mathbb{C}^{n \times n}$ than $\Sigma_n \Sigma_n^* \in \mathbb{C}^{N \times N}$.

Restating the theorem

The co-resolvent result: In the following it will be more convenient to work with $\Sigma_n^* \Sigma_n \in \mathbb{C}^{n\times n}$ than $\Sigma_n \Sigma_n^* \in \mathbb{C}^{N\times N}$.

- for $\tilde{L}_N = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(\Sigma_n^*\Sigma_n)}$,

$$\tilde{L}_N = \frac{n-N}{n}\delta_0 + \frac{N}{n}L_N$$

Restating the theorem

The co-resolvent result: In the following it will be more convenient to work with $\Sigma_n^* \Sigma_n \in \mathbb{C}^{n \times n}$ than $\Sigma_n \Sigma_n^* \in \mathbb{C}^{N \times N}$.

- for $\tilde{L}_N = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\Sigma_n^* \Sigma_n)}$,

$$\tilde{L}_N = \frac{n-N}{n} \delta_0 + \frac{N}{n} L_N$$

- in particular, as $L_N^R \to L_\infty^R$,

$$\tilde{L}_N \to \tilde{\mathcal{F}} = (1-c)\delta_0 + c\mathcal{F}.$$

Restating the theorem

The co-resolvent result: In the following it will be more convenient to work with $\Sigma_n^* \Sigma_n \in \mathbb{C}^{n \times n}$ than $\Sigma_n \Sigma_n^* \in \mathbb{C}^{N \times N}$.

- for $\tilde{L}_N = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\Sigma_n^* \Sigma_n)}$,

$$\tilde{L}_N = \frac{n-N}{n} \delta_0 + \frac{N}{n} L_N$$

- in particular, as $L_N^R \to L_\infty^R$,

$$\tilde{L}_N \to \tilde{\mathcal{F}} = (1-c)\delta_0 + c\mathcal{F}.$$

Theorem (Reexpressing the SCM limit)

For $z \in \mathbb{C}^+$, $\tilde{t}(z) = \int (\lambda - z)^{-1} \tilde{\mathcal{F}}(d\lambda)$ is the unique solution in $\mathbb{C}^+$ of

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}.$$

Restating the theorem

The co-resolvent result: In the following it will be more convenient to work with $\Sigma_n^*\Sigma_n \in \mathbb{C}^{n\times n}$ than $\Sigma_n\Sigma_n^* \in \mathbb{C}^{N\times N}$.

- for $\tilde{L}_N = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(\Sigma_n^*\Sigma_n)}$,

$$\tilde{L}_N = \frac{n-N}{n}\delta_0 + \frac{N}{n}L_N$$

- in particular, as $L_N^R \to L_\infty^R$,

$$\tilde{L}_N \to \tilde{\mathcal{F}} = (1-c)\delta_0 + c\mathcal{F}.$$

Theorem (Reexpressing the SCM limit)

For $z \in \mathbb{C}^+$, $\tilde{t}(z) = \int (\lambda - z)^{-1}\tilde{\mathcal{F}}(d\lambda)$ is the unique solution in $\mathbb{C}^+$ of

$$\tilde{t}(z) = \left(-z + c\int \frac{uL_\infty^R(du)}{1+u\tilde{t}(z)}\right)^{-1}.$$

Remark. $t(z), \tilde{t}(z)$ are linked by $czt(z) = 1 - c + z\tilde{t}(z)$ (from $\tilde{\mathcal{F}} = (1-c)\delta_0 + c\mathcal{F}$).

Restating the theorem

Proof. Starting from the result (inverted on both sides):

$$\frac{1}{\tilde{t}(z)} = -z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)}$$

multiply by $\tilde{t}(z)$ ($\neq 0$) to get

$$1 = -z\tilde{t}(z) + c \left(1 - \int \frac{L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)$$

so that, using $czt(z) = 1 - c + z\tilde{t}(z)$,

$$t(z) = -\frac{1}{z} \int \frac{L_\infty^R(du)}{1 + u\tilde{t}(z)} = \int \frac{L_\infty^R(du)}{-z - zu\tilde{t}(z)}$$

and finally, again with $czt(z) = 1 - c + z\tilde{t}(z)$,

$$t(z) = \int \frac{L_\infty^R(du)}{-z(1 + uct(z)) + (1-c)u}.$$

Both results are thus equivalent.

Outline

Motivation

Results so far: $L_N \xrightarrow{\mathrm{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?
- most importantly in practice: how to estimate R_N from $\Sigma_n \Sigma_n^*$?

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?
- most importantly in practice: how to estimate R_N from $\Sigma_n \Sigma_n^*$?
- we only have the link

$$R_N \leftrightarrow L_N^R \leftrightarrow g_n(z) \rightarrow t(z) \leftrightarrow \mathcal{F} \leftrightarrow L_N$$

how to go backwards?

$$L_N \xrightarrow{?} \dots \xrightarrow{?} R_N$$

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?
- most importantly in practice: how to estimate R_N from $\Sigma_n \Sigma_n^*$?
- we only have the link

$$R_N \leftrightarrow L_N^R \leftrightarrow g_n(z) \to t(z) \leftrightarrow \mathcal{F} \leftrightarrow L_N$$

how to go backwards?

$$L_N \overset{?}{\to} \ldots \overset{?}{\to} R_N$$

Follow-up of the class: this is the objective of what follows!

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?
- most importantly in practice: how to estimate R_N from $\Sigma_n \Sigma_n^*$?
- we only have the link

$$R_N \leftrightarrow L_N^R \leftrightarrow g_n(z) \to t(z) \leftrightarrow \mathcal{F} \leftrightarrow L_N$$

  how to go backwards?

$$L_N \stackrel{?}{\to} \ldots \stackrel{?}{\to} R_N$$

Follow-up of the class: this is the objective of what follows!

- from $\Sigma_n \Sigma_n^*$, estimate eigenvalues $\lambda_1(R_N), \ldots, \lambda_N(R_N)$?

Motivation

Results so far: $L_N \xrightarrow{\text{a.s.}} \mathcal{F}$ defined by Stieltjes transform $t(z)$ for all $z \in \mathbb{C}^+$:

- how to plot $\mathcal{F}$? is it a continuous measure?
- most importantly in practice: how to estimate R_N from $\Sigma_n \Sigma_n^*$?
- we only have the link

$$R_N \leftrightarrow L_N^R \leftrightarrow g_n(z) \to t(z) \leftrightarrow \mathcal{F} \leftrightarrow L_N$$

how to go backwards?

$$L_N \overset{?}{\to} \dots \overset{?}{\to} R_N$$

Follow-up of the class: this is the objective of what follows!

- from $\Sigma_n \Sigma_n^*$, estimate eigenvalues $\lambda_1(R_N), \dots, \lambda_N(R_N)$?
- from $\Sigma_n \Sigma_n^*$, estimate eigenvectors $u_1(R_N), \dots, u_N(R_N)$?

Outline

From Stieltjes transform to measure

Proposition (Condition on density measure)

For g_μ Stieltjes transform of μ with real support and finite mass. Assume

$$\frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[g_\mu(x_0 + \imath y) \right] \equiv I(x_0) \text{ exists}$$

for all $x \in \mathcal{V}(x_0)$. Then μ has a density in x_0 equal to $I(x_0)$.

Proof. See course notes.

Application to the SCM model

Objective: The goal here is to show:

Application to the SCM model

Objective: The goal here is to show:

- that the limit law $\mathcal{F}$ has a density

Application to the SCM model

Objective: The goal here is to show:

- that the limit law $\mathcal{F}$ has a density
- that $\mathcal{F}$ can be plotted exactly (not only approximately through $\frac{1}{\pi}\Im[t(x+\imath\varepsilon)]$)

Application to the SCM model

Objective: The goal here is to show:

- that the limit law $\mathcal{F}$ has a density
- that $\mathcal{F}$ can be plotted exactly (not only approximately through $\frac{1}{\pi}\Im[t(x+\imath\varepsilon)]$)
- that the support $\mathrm{supp}(\mathcal{F})$ is well identified

Application to the SCM model

Objective: The goal here is to show:

- that the limit law $\mathcal{F}$ has a density
- that $\mathcal{F}$ can be plotted exactly (not only approximately through $\frac{1}{\pi}\Im[t(x+\imath\varepsilon)]$)
- that the support $\mathrm{supp}(\mathcal{F})$ is well identified

Beyond the spectrum: Most importantly, this is a required first step to:

- create a **strong link** between R_N and $\Sigma_n\Sigma_n^*$

Application to the SCM model

Objective: The goal here is to show:

- that the limit law $\mathcal{F}$ has a density
- that $\mathcal{F}$ can be plotted exactly (not only approximately through $\frac{1}{\pi}\Im[t(x+\imath\varepsilon)]$)
- that the support $\mathrm{supp}(\mathcal{F})$ is well identified

Beyond the spectrum: Most importantly, this is a required first step to:

- create a **strong link** between R_N and $\Sigma_n\Sigma_n^*$
- provide **new statistical inference** tools on R_N (eigenvalues and eigenvectors).

Application to the SCM model

Starting point:

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}.$$

Application to the SCM model

Starting point:

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}.$$

Two fundamental identities: for all $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$,

$$(\star) \quad (\tilde{t}(z_1) - \tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))} \right) = (z_1 - z_2)\tilde{t}(z_1)\tilde{t}(z_2)$$

Application to the SCM model

Starting point:

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}.$$

Two fundamental identities: for all $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$,

$$(\star) \quad (\tilde{t}(z_1) - \tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))} \right) = (z_1 - z_2)\tilde{t}(z_1)\tilde{t}(z_2)$$

$$(\star\star) \quad (z_1\tilde{t}(z_1) - z_2\tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))} \right)$$
$$= (z_1 - z_2) c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2) u L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))}.$$

Application to the SCM model

Starting point:

$$\tilde{t}(z) = \left(-z + c \int \frac{uL_\infty^R(du)}{1 + u\tilde{t}(z)}\right)^{-1}.$$

Two fundamental identities: for all $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$,

$$(\star) \quad (\tilde{t}(z_1) - \tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2)u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))}\right) = (z_1 - z_2)\tilde{t}(z_1)\tilde{t}(z_2)$$

$$(\star\star) \quad (z_1\tilde{t}(z_1) - z_2\tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2)u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))}\right)$$
$$= (z_1 - z_2)c \int \frac{\tilde{t}(z_1)\tilde{t}(z_2)uL_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))}.$$

Elements of proof.

- follows immediately from the "scalar resolvent identity": $a^{-1} - b^{-1} = a^{-1}b^{-1}(b - a)$ (remember the matrix form $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$).
- applied to the scalar inverse $\tilde{t}(z) = (-z + c\int u/(1 + u\tilde{t}(z))L_\infty^R(du))^{-1}$.

Application to the SCM model

Important corollary: taking $z_1 = z \in \mathbb{C}^+$ and $z_2 = \bar{z}$ in $(\star)$,

$$2\imath\Im[\tilde{t}(z)]\left(1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}(z)|^2}\right) = 2\imath\Im[z]|\tilde{t}(z)|^2$$

Application to the SCM model

Important corollary: taking $z_1 = z \in \mathbb{C}^+$ and $z_2 = \bar{z}$ in $(\star)$,

$$2\imath\Im[\tilde{t}(z)]\left(1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}(z)|^2}\right) = 2\imath\Im[z]|\tilde{t}(z)|^2$$

so that, since $\Im[z] > 0$ and $\Im[\tilde{t}(z)] > 0$ (Stieltjes transform of real supported measure),

$$\boxed{(\star\star\star) \quad \forall z \in \mathbb{C}^+, \quad 1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}(z)|^2} > 0.}$$

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$,

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^\circ(x) \quad \textit{exists.}$$

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$*,*

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^\circ(x) \quad \textit{exists.}$$

This implies that $\tilde{\mathcal{F}}$ *has a density* f *in all* $\mathbb{R} \setminus \{0\}$ *equal to*

$$f(x) \equiv \frac{1}{\pi} \Im[\tilde{t}^\circ(x)].$$

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$,

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^\circ(x) \quad \textit{exists.}$$

This implies that $\tilde{\mathcal{F}}$ *has a density* f *in all* $\mathbb{R} \setminus \{0\}$ *equal to*

$$f(x) \equiv \frac{1}{\pi} \Im[\tilde{t}^\circ(x)].$$

Proof. Two steps:

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$,

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^\circ(x) \quad \textit{exists.}$$

This implies that $\tilde{\mathcal{F}}$ has a density f in all $\mathbb{R} \setminus \{0\}$ equal to

$$f(x) \equiv \frac{1}{\pi} \Im[\tilde{t}^\circ(x)].$$

Proof. Two steps:

Step 1. Show that $\tilde{t}(z)$ **bounded** in a neighborhood $B(x) \subset \mathbb{C}^+$ of $x \in \mathbb{R}$.

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$,

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^{\circ}(x) \quad \textit{exists.}$$

This implies that $\tilde{\mathcal{F}}$ *has a density* f *in all* $\mathbb{R} \setminus \{0\}$ *equal to*

$$f(x) \equiv \frac{1}{\pi} \Im[\tilde{t}^{\circ}(x)].$$

Proof. Two steps:

Step 1. Show that $\tilde{t}(z)$ **bounded** in a neighborhood $B(x) \subset \mathbb{C}^+$ of $x \in \mathbb{R}$.

Step 2. Extract **converging subsequences** $(z_n, \tilde{t}(z_n)) \to (x, \tilde{t})$ and show that $\tilde{t}$ is unique.

Application to the SCM model

Theorem (Existence of a density)

For all $x \in \mathbb{R} \setminus \{0\}$,

$$\lim_{z \in \mathbb{C}^+ \to x} \tilde{t}(z) = \tilde{t}^\circ(x) \quad \textit{exists.}$$

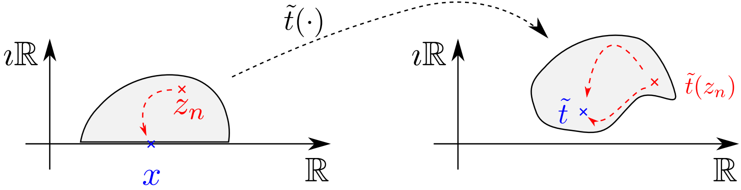
This implies that $\tilde{\mathcal{F}}$ *has a density* f *in all* $\mathbb{R} \setminus \{0\}$ *equal to*

$$f(x) \equiv \frac{1}{\pi} \Im[\tilde{t}^\circ(x)].$$

Proof. Two steps:

Step 1. Show that $\tilde{t}(z)$ **bounded** in a neighborhood $B(x) \subset \mathbb{C}^+$ of $x \in \mathbb{R}$.

Step 2. Extract **converging subsequences** $(z_n, \tilde{t}(z_n)) \to (x, \tilde{t})$ and show that $\tilde{t}$ is unique.



Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

For $z_0 \in \mathbb{C}^+$ **fixed**, from $(\star\star)$,

$$\left(z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)\right) \left(1 - c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right)$$

$$= (z_n - z_0) c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}.$$

Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

For $z_0 \in \mathbb{C}^+$ **fixed**, from $(\star\star)$,

$$\left(z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)\right) \left(1 - c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right)$$

$$= (z_n - z_0) c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}.$$

Goal: show that $|\tilde{t}(z_n)| \to \infty$ induces **contradiction!**

Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

For $z_0 \in \mathbb{C}^+$ **fixed**, from $(\star\star)$,

$$\left(z_n\tilde{t}(z_n) - z_0\tilde{t}(z_0)\right)\left(1 - c\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right)$$

$$= (z_n - z_0)c\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}.$$

Goal: show that $|\tilde{t}(z_n)| \to \infty$ induces **contradiction!**

- By Cauchy-Swcharz's inequality,

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| \leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}}\sqrt{\int \frac{|\tilde{t}(z_0)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}}$$

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| \leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}}\sqrt{\int \frac{|\tilde{t}(z_0)|^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}}.$$

Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

For $z_0 \in \mathbb{C}^+$ **fixed**, from $(\star\star)$,

$$\left(z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)\right) \left(1 - c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right)$$

$$= (z_n - z_0) c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}.$$

Goal: show that $|\tilde{t}(z_n)| \to \infty$ induces **contradiction!**

- By Cauchy-Swcharz's inequality,

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right| \leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z_n)|^2}} \sqrt{\int \frac{|\tilde{t}(z_0)|^2 u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z_0)|^2}}$$

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right| \leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z_n)|^2}} \sqrt{\int \frac{|\tilde{t}(z_0)|^2 L_\infty^R(du)}{|1 + u\tilde{t}(z_0)|^2}}.$$

so that, from $(\star\star\star)$, as $n \to \infty$,

$$\limsup_n c \left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right| < 1, \quad \limsup_n c \left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}\right| < \infty.$$

Application to the SCM model

Step 1. By contradiction: assume $|\tilde{t}(z_n)| \to \infty$ for $z_n \in B(x) \to x \in \mathbb{R} \setminus \{0\}$, B **bounded**.

For $z_0 \in \mathbb{C}^+$ **fixed**, from $(\star\star)$,

$$\left(z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)\right)\left(1 - c\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right)$$
$$= (z_n - z_0)c\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}.$$

Goal: show that $|\tilde{t}(z_n)| \to \infty$ induces **contradiction!**

- By Cauchy-Swcharz's inequality,

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| \le \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}}\sqrt{\int \frac{|\tilde{t}(z_0)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}}$$

$$\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| \le \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}}\sqrt{\int \frac{|\tilde{t}(z_0)|^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}}.$$

so that, from $(\star\star\star)$, as $n \to \infty$,

$$\limsup_n c\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| < 1, \quad \limsup_n c\left|\int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))}\right| < \infty.$$

- **Careful!** left inequality is strict because $\int \frac{|\tilde{t}(z_0)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2} < 1$ but $\lim_n \int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2} \le 1$ only!

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_{n} |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n\tilde{t}(z_n) - z_0\tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n\tilde{t}(z_n) - z_0\tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By $(\star)$,

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})$$

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By $(\star)$,

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n}) \tilde{t}(z_{1,n}) \tilde{t}(z_{2,n}) \overset{n\to\infty}{\longrightarrow} 0.$$

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By ($\star$),

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c\int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) \overset{n\to\infty}{\longrightarrow} 0.$$

To show that $\tilde{t}_1 = \tilde{t}_2$, it suffices to show that

$$\liminf_n \left| 1 - c\int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))} \right| > 0.$$

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By $(\star)$,

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) \overset{n\to\infty}{\longrightarrow} 0.$$

To show that $\tilde{t}_1 = \tilde{t}_2$, it suffices to show that

$$\liminf_n \left| 1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))} \right| > 0.$$

This is again "in spirit" the **strict inequality** case of cauchy-Schwarz, but $z_{1,n}, z_{2,n}$ are not fixed!

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n\tilde{t}(z_n) - z_0\tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n\tilde{t}(z_n) - z_0\tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By ($\star$),

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c\int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) \overset{n\to\infty}{\longrightarrow} 0.$$

To show that $\tilde{t}_1 = \tilde{t}_2$, it suffices to show that

$$\liminf_n \left|1 - c\int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))}\right| > 0.$$

This is again "in spirit" the **strict inequality** case of cauchy-Schwarz, but $z_{1,n}, z_{2,n}$ are not fixed!

An option is to use a polarization method on $\Re\left[c\int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))}\right]$.

(see details in course notes)

Application to the SCM model

Step 1. As a consequence, we find

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty.$$

But we assumed

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \to \infty$$

(recall $z_n \not\to 0$). This is a contradiction!

Step 2. Since $\tilde{t}(z_n)$ bounded for $z_n \in B(x)$, let $z_{1,n}, z_{2,n} \to x$ with $\tilde{t}(z_{1,n}) \to \tilde{t}_1$ and $\tilde{t}(z_{2,n}) \to \tilde{t}_2$. By ($\star$),

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))}\right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) \overset{n\to\infty}{\longrightarrow} 0.$$

To show that $\tilde{t}_1 = \tilde{t}_2$, it suffices to show that

$$\liminf_n \left| 1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))} \right| > 0.$$

This is again "in spirit" the **strict inequality** case of cauchy-Schwarz, but $z_{1,n}, z_{2,n}$ are not fixed!

An option is to use a polarization method on $\Re\left[c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}) u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))}\right]$.

(see details in course notes)

This concludes the proof. $\square$

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c\int \frac{uL_\infty^R(du)}{1 + u\tilde{t}(z)}\right)^{-1}$$

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c \int \frac{uL_{\infty}^{R}(du)}{1 + u\tilde{t}(z)} \right)^{-1} \Leftrightarrow z = -\frac{1}{\tilde{t}(z)} + c \int \frac{uL_{\infty}^{R}(du)}{1 + u\tilde{t}(z)}$$

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1} \Leftrightarrow z = -\frac{1}{\tilde{t}(z)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)}$$

with

- $\tilde{t}(z)$ **defined on** $\mathbb{C}^+$ with **image** $\tilde{t}(\mathbb{C}^+)$,

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c\int \frac{uL_\infty^R(du)}{1+u\tilde{t}(z)}\right)^{-1} \Leftrightarrow z = -\frac{1}{\tilde{t}(z)} + c\int \frac{uL_\infty^R(du)}{1+u\tilde{t}(z)}$$

with

- $\tilde{t}(z)$ **defined on** $\mathbb{C}^+$ with **image** $\tilde{t}(\mathbb{C}^+)$, and so with **explicit inverse**

$$\begin{aligned} z(\cdot) : \tilde{t}(\mathbb{C}^+) &\to \mathbb{C}^+ \\ \tilde{t} &\mapsto -\frac{1}{\tilde{t}} + c\int \frac{uL_\infty^R(du)}{1+u\tilde{t}}. \end{aligned}$$

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1} \Leftrightarrow z = -\frac{1}{\tilde{t}(z)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)}$$

with

- $\tilde{t}(z)$ **defined on** $\mathbb{C}^+$ with **image** $\tilde{t}(\mathbb{C}^+)$, and so with **explicit inverse**

$$\begin{aligned} z(\cdot) : \tilde{t}(\mathbb{C}^+) &\to \mathbb{C}^+ \\ \tilde{t} &\mapsto -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}. \end{aligned}$$

- tempting to **extend** $z(\cdot)$ to $\mathbb{C} \setminus \{\tilde{t} \in \mathbb{R}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$, where $z(\cdot)$ is "valid".

Application to the SCM model

Important Remark. Recall that

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1} \Leftrightarrow z = -\frac{1}{\tilde{t}(z)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)}$$

with

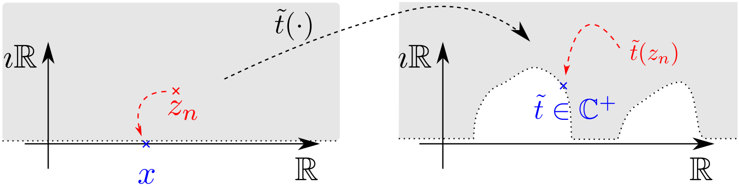
- $\tilde{t}(z)$ **defined on** $\mathbb{C}^+$ with **image** $\tilde{t}(\mathbb{C}^+)$, and so with **explicit inverse**

$$\begin{aligned} z(\cdot) : \tilde{t}(\mathbb{C}^+) &\to \mathbb{C}^+ \\ \tilde{t} &\mapsto -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}. \end{aligned}$$

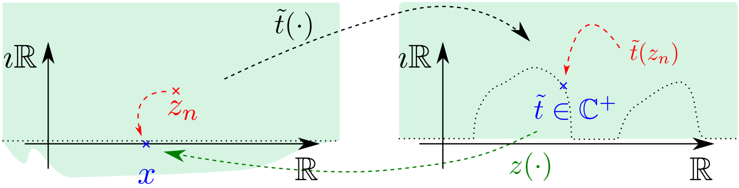
- tempting to **extend** $z(\cdot)$ to $\mathbb{C} \setminus \{\tilde{t} \in \mathbb{R}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$, where $z(\cdot)$ is "valid".
- in particular, the restriction to $\mathbb{R}^* \setminus \{\tilde{t} \in \mathbb{R}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$ of this "extension" defines

$$\begin{aligned} x(\cdot) : \mathbb{R}^* \setminus \{\tilde{t} \in \mathbb{R}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\} &\to \mathbb{R} \\ \tilde{t} &\mapsto -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}. \end{aligned}$$

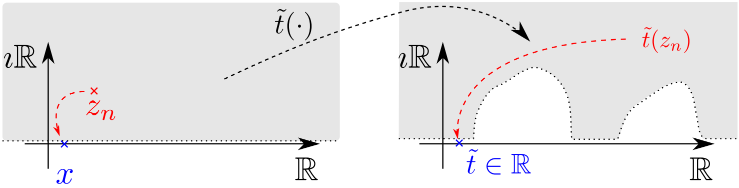
Application to the SCM model



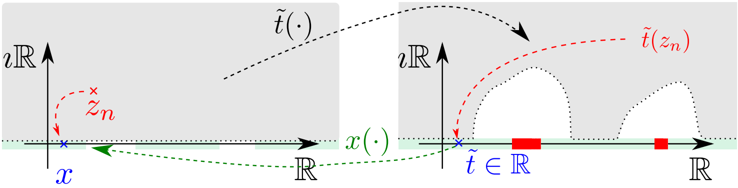
Application to the SCM model



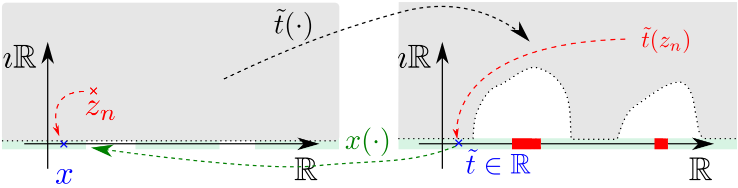
Application to the SCM model



Application to the SCM model

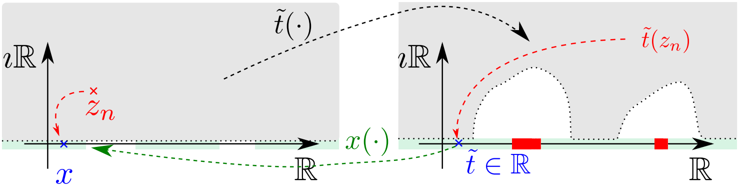


Application to the SCM model



The function $x(\cdot)$:

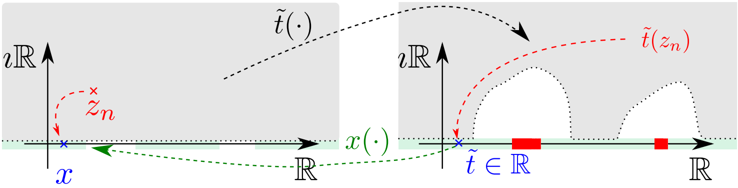
Application to the SCM model



The function $x(\cdot)$:

- coincides on $\tilde{t}(\mathbb{R}^* \setminus \mathrm{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \mathrm{supp}(\mathcal{F})$

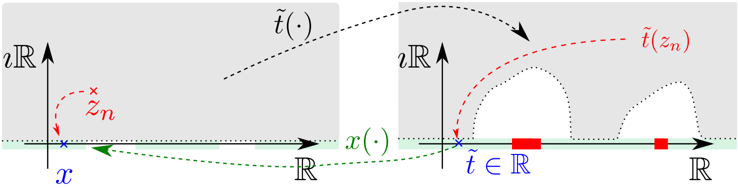
Application to the SCM model



The function $x(\cdot)$:

- coincides on $\tilde{t}(\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F})$
- on the rest, is an extension of $z(\cdot)$ with **no physical interpretation**.

Application to the SCM model

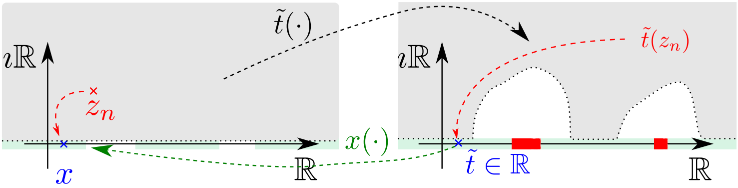


The function $x(\cdot)$:

- coincides on $\tilde{t}(\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F})$
- on the rest, is an extension of $z(\cdot)$ with **no physical interpretation**.

The intuition: recall that $\tilde{t}(x) = \int (v-x)^{-1} \tilde{\mathcal{F}}(dv)$ is **increasing on all connected components of** $\mathbb{R} \setminus \operatorname{supp}(\tilde{\mathcal{F}})$.

Application to the SCM model



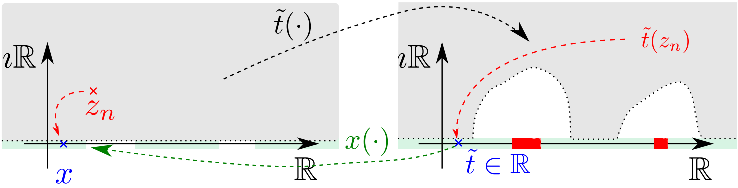
The function $x(\cdot)$:

- coincides on $\tilde{t}(\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F})$
- on the rest, is an extension of $z(\cdot)$ with **no physical interpretation**.

The intuition: recall that $\tilde{t}(x) = \int (v - x)^{-1} \tilde{\mathcal{F}}(dv)$ is **increasing on all connected components of** $\mathbb{R} \setminus \operatorname{supp}(\tilde{\mathcal{F}})$.

As such, **we expect** that:

Application to the SCM model



The function $x(\cdot)$:

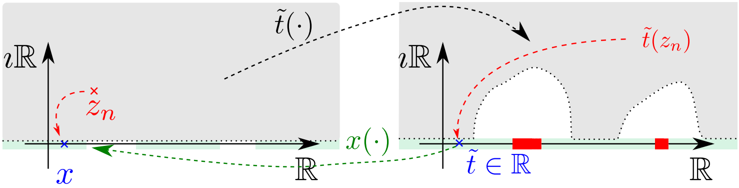
- coincides on $\tilde{t}(\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F})$
- on the rest, is an extension of $z(\cdot)$ with **no physical interpretation**.

The intuition: recall that $\tilde{t}(x) = \int (v - x)^{-1} \tilde{\mathcal{F}}(dv)$ is **increasing on all connected components of** $\mathbb{R} \setminus \operatorname{supp}(\tilde{\mathcal{F}})$.

As such, **we expect** that:

- on $\tilde{t}(\mathbb{R}^* \setminus \operatorname{supp}(\mathcal{F}))$, $x(\tilde{t})$ is (well defined and) **increasing** (local inverse of increasing function!)

Application to the SCM model



The function $x(\cdot)$:

- coincides on $\tilde{t}(\mathbb{R}^* \setminus \mathrm{supp}(\mathcal{F}))$ with the restriction of (the extension of) $\tilde{t}$ on $\mathbb{R}^* \setminus \mathrm{supp}(\mathcal{F})$
- on the rest, is an extension of $z(\cdot)$ with **no physical interpretation**.

The intuition: recall that $\tilde{t}(x) = \int (v-x)^{-1} \tilde{\mathcal{F}}(dv)$ is **increasing on all connected components of $\mathbb{R} \setminus \mathrm{supp}(\tilde{\mathcal{F}})$**.

As such, **we expect** that:

- on $\tilde{t}(\mathbb{R}^* \setminus \mathrm{supp}(\mathcal{F}))$, $x(\tilde{t})$ is (well defined and) **increasing** (local inverse of increasing function!)
- elsewhere, $x(\tilde{t})$ **may not** be increasing (otherwise, it would have an increasing local inverse satisfying $\tilde{t}(x)$ equation: but this is not a proof!)

Application to the SCM model

Theorem (Exact limiting spectrum description)

Let $z(\cdot)$ *and* $x(\cdot)$ *as above. Then,*

Application to the SCM model

Theorem (Exact limiting spectrum description)

Let $z(\cdot)$ and $x(\cdot)$ as above. Then,

1. *if $x_0 \neq 0 \in \operatorname{supp}(\mathcal{F})$, a* **positive density point** *of $\mathcal{F}$, then $\tilde{t}^\circ(x_0)$ is unique solution $\tilde{t}$* ***with*** *$\Im[\tilde{t}] > 0$ of*
$$z(\tilde{t}) = x_0.$$

Application to the SCM model

Theorem (Exact limiting spectrum description)

Let $z(\cdot)$ and $x(\cdot)$ as above. Then,

1. *if $x_0 \neq 0 \in \operatorname{supp}(\mathcal{F})$, a* **positive density point** *of $\mathcal{F}$, then $\tilde{t}^\circ(x_0)$ is unique solution $\tilde{t}$* **with** *$\Im[\tilde{t}] > 0$ of*
$$z(\tilde{t}) = x_0.$$
2. *if $x_0 \neq 0 \notin \operatorname{supp}(\mathcal{F})$, then*
$$\tilde{t}^\circ(x_0) \in \mathbb{R}^* \setminus \{\tilde{t}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$$

Application to the SCM model

Theorem (Exact limiting spectrum description)

Let $z(\cdot)$ and $x(\cdot)$ as above. Then,

1. *if $x_0 \neq 0 \in \operatorname{supp}(\mathcal{F})$, a* **positive density point** *of $\mathcal{F}$, then $\tilde{t}^\circ(x_0)$ is unique solution $\tilde{t}$* **with** *$\Im[\tilde{t}] > 0$ of*

$$z(\tilde{t}) = x_0.$$

2. *if $x_0 \neq 0 \notin \operatorname{supp}(\mathcal{F})$, then*

$$\tilde{t}^\circ(x_0) \in \mathbb{R}^* \setminus \{\tilde{t}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$$

and is the unique real solution to $x(\tilde{t}) = x_0$ **with** *$x'(\tilde{t}) > 0$.*

Application to the SCM model

Theorem (Exact limiting spectrum description)

Let $z(\cdot)$ and $x(\cdot)$ as above. Then,

1. *if $x_0 \neq 0 \in \operatorname{supp}(\mathcal{F})$, a* **positive density point** *of $\mathcal{F}$, then $\tilde{t}^\circ(x_0)$ is unique solution $\tilde{t}$* **with** *$\Im[\tilde{t}] > 0$ of*
$$z(\tilde{t}) = x_0.$$

2. *if $x_0 \neq 0 \notin \operatorname{supp}(\mathcal{F})$, then*
$$\tilde{t}^\circ(x_0) \in \mathbb{R}^* \setminus \{\tilde{t}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$$
and is the unique real solution to $x(\tilde{t}) = x_0$ **with** *$x'(\tilde{t}) > 0$.*

   Conversely, for $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}, -1/\tilde{t} \in \operatorname{supp}(L_\infty^R)\}$ such that $x'(\tilde{t}_0) > 0$,
$$x(\tilde{t}_0) \notin \operatorname{supp}(\mathcal{F}).$$

Application to the SCM model

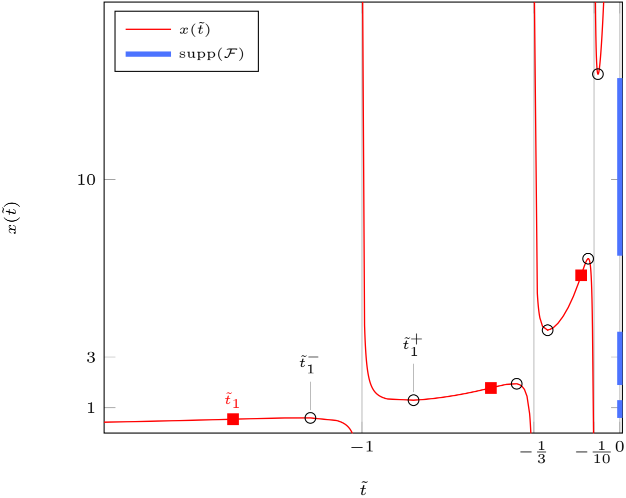


Figure: $x(\tilde{t})$ for $\tilde{t} \in \mathbb{R}$, R_N diagonal with 3 masses in 1, 3, 10, $c = 1/10$. Support $\mathrm{supp}(\mathcal{F})$ underlined on y-axis.

Application to the SCM model

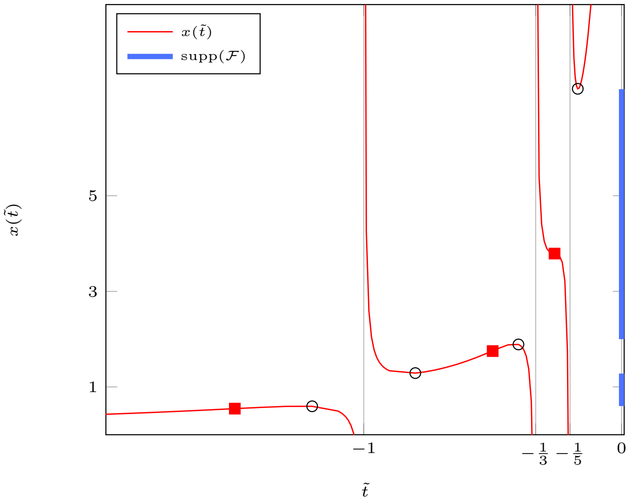


Figure: $x(\tilde{t})$ for $\tilde{t} \in \mathbb{R}$, R_N diagonal with 3 masses in 1, 3, 5, $c = 1/10$. Support $\mathrm{supp}(\mathcal{F})$ underlined on y-axis.

Application to the SCM model

Proof. [1. Existence] Of course $0 \neq \tilde{t}^\circ(x_0) = \lim_{y \downarrow 0} \tilde{t}(x_0 + \imath y)$ with

$$x_0 + \imath y = -\frac{1}{\tilde{t}(x_0 + \imath y)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + \imath y)}.$$

Application to the SCM model

Proof. [1. Existence] Of course $0 \neq \tilde{t}^\circ(x_0) = \lim_{y \downarrow 0} \tilde{t}(x_0 + \imath y)$ with

$$x_0 + \imath y = -\frac{1}{\tilde{t}(x_0+\imath y)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0+\imath y)}.$$

Besides,

$$\left| \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0+\imath y)} \right|^2 \leq \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(x_0+\imath y)|^2} \leq \frac{1}{\Im[\tilde{t}(x_0+\imath y)]^2}$$

since $|1 + u\tilde{t}(x_0+\imath y)|^2 \geq t^2 \Im[\tilde{t}(x_0+\imath y)]^2$.

Proof. [1. Existence] Of course $0 \neq \tilde{t}^\circ(x_0) = \lim_{y \downarrow 0} \tilde{t}(x_0 + \imath y)$ with

$$x_0 + \imath y = -\frac{1}{\tilde{t}(x_0 + \imath y)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + \imath y)}.$$

Besides,

$$\left| \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + \imath y)} \right|^2 \leq \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(x_0 + \imath y)|^2} \leq \frac{1}{\Im[\tilde{t}(x_0 + \imath y)]^2}$$

since $|1 + u\tilde{t}(x_0 + \imath y)|^2 \geq t^2 \Im[\tilde{t}(x_0 + \imath y)]^2$.

But $\Im[\tilde{t}(x_0 + \imath y)]$ converges to non-zero value, so $\frac{1}{\Im[\tilde{t}(x_0 + \imath y)]^2}$ bounded for all small y.

Application to the SCM model

Proof. [1. Existence] Of course $0 \neq \tilde{t}^\circ(x_0) = \lim_{y \downarrow 0} \tilde{t}(x_0 + \imath y)$ with

$$x_0 + \imath y = -\frac{1}{\tilde{t}(x_0 + \imath y)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + \imath y)}.$$

Besides,

$$\left| \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + \imath y)} \right|^2 \leq \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(x_0 + \imath y)|^2} \leq \frac{1}{\Im[\tilde{t}(x_0 + \imath y)]^2}$$

since $|1 + u\tilde{t}(x_0 + \imath y)|^2 \geq t^2 \Im[\tilde{t}(x_0 + \imath y)]^2$.

But $\Im[\tilde{t}(x_0 + \imath y)]$ converges to non-zero value, so $\frac{1}{\Im[\tilde{t}(x_0 + \imath y)]^2}$ bounded for all small y.

By dominated convergence, as $y \downarrow 0$,

$$x_0 = -\frac{1}{\tilde{t}^\circ(x_0)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0)}$$

proving **existence**.

Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$

Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$
- $B_1 = B(\tilde{t}_1, \varepsilon)$ and $B_2 = B(\tilde{t}_2, \varepsilon)$ with $B(\tilde{t}_1, \varepsilon) \cap B(\tilde{t}_2, \varepsilon) = \emptyset$.

Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$
- $B_1 = B(\tilde{t}_1, \varepsilon)$ and $B_2 = B(\tilde{t}_2, \varepsilon)$ with $B(\tilde{t}_1, \varepsilon) \cap B(\tilde{t}_2, \varepsilon) = \emptyset$.

Then

1. $x_0 \in z(B_1) \cap z(B_2)$ which is open by the open mapping theorem ($z(\cdot)$ being analytic there)

Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$
- $B_1 = B(\tilde{t}_1, \varepsilon)$ and $B_2 = B(\tilde{t}_2, \varepsilon)$ with $B(\tilde{t}_1, \varepsilon) \cap B(\tilde{t}_2, \varepsilon) = \emptyset$.

Then

1. $x_0 \in z(B_1) \cap z(B_2)$ which is open by the open mapping theorem ($z(\cdot)$ being analytic there)
2. so there exists $z' \in \mathbb{C}^+ \cap (z(B_1) \cap z(B_2))$ with $z' = z(\tilde{t}'_1) = z(\tilde{t}'_2)$ for some $t'_1 \in B_1$, $t'_2 \in B_2$

Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$
- $B_1 = B(\tilde{t}_1, \varepsilon)$ and $B_2 = B(\tilde{t}_2, \varepsilon)$ with $B(\tilde{t}_1, \varepsilon) \cap B(\tilde{t}_2, \varepsilon) = \emptyset$.

Then

1. $x_0 \in z(B_1) \cap z(B_2)$ which is open by the open mapping theorem ($z(\cdot)$ being analytic there)
2. so there exists $z' \in \mathbb{C}^+ \cap (z(B_1) \cap z(B_2))$ with $z' = z(\tilde{t}'_1) = z(\tilde{t}'_2)$ for some $t'_1 \in B_1$, $t'_2 \in B_2$
3. but this is impossible since then $t'_1 = \tilde{t}(z')$, $t'_2 = \tilde{t}(z')$, which breaks uniqueness of $\tilde{t}$ in $\mathbb{C}^+$.

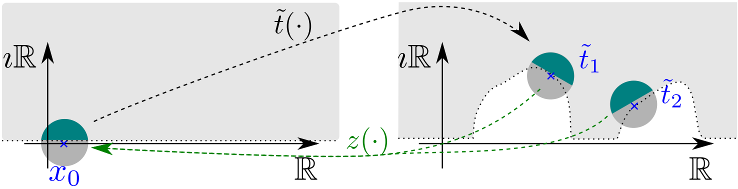
Application to the SCM model

[1. Uniqueness] Let

- $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ with $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$
- $B_1 = B(\tilde{t}_1, \varepsilon)$ and $B_2 = B(\tilde{t}_2, \varepsilon)$ with $B(\tilde{t}_1, \varepsilon) \cap B(\tilde{t}_2, \varepsilon) = \emptyset$.

Then

1. $x_0 \in z(B_1) \cap z(B_2)$ which is open by the open mapping theorem ($z(\cdot)$ being analytic there)
2. so there exists $z' \in \mathbb{C}^+ \cap (z(B_1) \cap z(B_2))$ with $z' = z(\tilde{t}'_1) = z(\tilde{t}'_2)$ for some $t'_1 \in B_1$, $t'_2 \in B_2$
3. but this is impossible since then $t'_1 = \tilde{t}(z')$, $t'_2 = \tilde{t}(z')$, which breaks uniqueness of $\tilde{t}$ in $\mathbb{C}^+$.



Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^{\circ}(x_0) \neq 0$ (see later).

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^{\circ}(x_0) \neq 0$ (see later). As $x_0 \notin \operatorname{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^{\circ}(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^\circ(x_0) \neq 0$ (see later). As $x_0 \notin \mathrm{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^\circ(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

But, for all $\tilde{t} \in B(\tilde{t}(x_0)) \cap \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}$$

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^\circ(x_0) \neq 0$ (see later). As $x_0 \notin \operatorname{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^\circ(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

But, for all $\tilde{t} \in B(\tilde{t}(x_0)) \cap \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}} = -\frac{1}{\tilde{t}} + \frac{c}{\tilde{t}} \left(1 - \frac{1}{\tilde{t}} r\left(-\frac{1}{\tilde{t}}\right)\right)$$

where

$$r(z) = \int (u - z)^{-1} L_\infty^R(du)$$

is the Stieltjes transform of L_∞^R.

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^\circ(x_0) \neq 0$ (see later). As $x_0 \notin \operatorname{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^\circ(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

But, for all $\tilde{t} \in B(\tilde{t}(x_0)) \cap \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}} = -\frac{1}{\tilde{t}} + \frac{c}{\tilde{t}} \left(1 - \frac{1}{\tilde{t}} r\left(-\frac{1}{\tilde{t}} \right) \right)$$

where

$$r(z) = \int (u - z)^{-1} L_\infty^R(du)$$

is the Stieltjes transform of L_∞^R.

As a consequence, $r(-\frac{1}{\tilde{t}}) \to \ell \in \mathbb{R}$ as $\tilde{t} \to \tilde{t}(x_0)$ (and this is also true for all x in a neighborhood of x_0).

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^\circ(x_0) \neq 0$ (see later). As $x_0 \notin \operatorname{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^\circ(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

But, for all $\tilde{t} \in B(\tilde{t}(x_0)) \cap \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}} = -\frac{1}{\tilde{t}} + \frac{c}{\tilde{t}} \left(1 - \frac{1}{\tilde{t}} r\left(-\frac{1}{\tilde{t}}\right)\right)$$

where

$$r(z) = \int (u - z)^{-1} L_\infty^R(du)$$

is the Stieltjes transform of L_∞^R.

As a consequence, $r(-\frac{1}{\tilde{t}}) \to \ell \in \mathbb{R}$ as $\tilde{t} \to \tilde{t}(x_0)$ (and this is also true for all x in a neighborhood of x_0). This implies that

$$-\frac{1}{\tilde{t}(x_0)} \notin \operatorname{supp}(L_\infty^R)$$

(since $\Im[r(-\frac{1}{\tilde{t}})] \to 0$ for all $\tilde{t} \to \tilde{t}(x_0)$).

Application to the SCM model

[2. Preliminary] We can show that $\tilde{t}^{\circ}(x_0) \neq 0$ (see later). As $x_0 \notin \mathrm{supp}(\mathcal{F})$, $\tilde{t}(\cdot)$ defined in x_0 as

$$\tilde{t}(x_0) = \tilde{t}^{\circ}(x_0) = \int \frac{1}{u - x_0} \tilde{\mathcal{F}}(du)$$

for which, of course, $\tilde{t}'(x_0) > 0$ **(on $\mathbb{R}$)**. Hence, $\tilde{t}$ complex analytic around $(x_0, \tilde{t}(x_0))$.

But, for all $\tilde{t} \in B(\tilde{t}(x_0)) \cap \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_{\infty}^{R}(du)}{1 + u\tilde{t}} = -\frac{1}{\tilde{t}} + \frac{c}{\tilde{t}} \left(1 - \frac{1}{\tilde{t}} r \left(-\frac{1}{\tilde{t}} \right) \right)$$

where

$$r(z) = \int (u - z)^{-1} L_{\infty}^{R}(du)$$

is the Stieltjes transform of L_{∞}^{R}.

As a consequence, $r(-\frac{1}{\tilde{t}}) \to \ell \in \mathbb{R}$ as $\tilde{t} \to \tilde{t}(x_0)$ (and this is also true for all x in a neighborhood of x_0). This implies that

$$-\frac{1}{\tilde{t}(x_0)} \notin \mathrm{supp}(L_{\infty}^{R})$$

(since $\Im[r(-\frac{1}{\tilde{t}})] \to 0$ for all $\tilde{t} \to \tilde{t}(x_0)$). Equivalently,

$$\tilde{t}(x_0) \notin \left\{ \tilde{t}; -\frac{1}{\tilde{t}} \in \mathrm{supp}(L_{\infty}^{R}) \right\}.$$

Application to the SCM model

[2. Existence] This implies that

- $x(\tilde{t}(x_0))$ well defined

Application to the SCM model

[2. Existence] This implies that

- $x(\tilde{t}(x_0))$ well defined
- and $\tilde{t}(x_0)$ is a solution to

$$x(\tilde{t}(x_0)) = -\frac{1}{\tilde{t}(x_0)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0)}$$

(analytic extension of $z(\tilde{t})$ possible around $(\tilde{t}(x_0), x_0)$).

Application to the SCM model

[2. Existence] This implies that

- $x(\tilde{t}(x_0))$ well defined
- and $\tilde{t}(x_0)$ is a solution to

$$x(\tilde{t}(x_0)) = -\frac{1}{\tilde{t}(x_0)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0)}$$

(analytic extension of $z(\tilde{t})$ possible around $(\tilde{t}(x_0), x_0)$).

- As $\tilde{t}'(x_0) > 0$, by local inverse, $x'(\tilde{t}(x_0)) > 0$.

Application to the SCM model

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$.

Application to the SCM model

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$. Then,

$$(\tilde{t}_1 - \tilde{t}_2)\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)}\right) = 0.$$

Application to the SCM model

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$. Then,

$$(\tilde{t}_1 - \tilde{t}_2)\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)}\right) = 0.$$

Since $\tilde{t}_1 \neq \tilde{t}_2$, necessarily

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} = 1.$$

Application to the SCM model

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$. Then,

$$(\tilde{t}_1 - \tilde{t}_2)\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)}\right) = 0.$$

Since $\tilde{t}_1 \neq \tilde{t}_2$, necessarily

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} = 1.$$

But, $x'(\tilde{t}_i) > 0$ is equivalent to $c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_i|^2} < 1$, and by Cauchy–Schwarz,

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} \leq \sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_1|^2}}\sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_2|^2}} < 1.$$

Application to the SCM model

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$. Then,

$$(\tilde{t}_1 - \tilde{t}_2)\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)}\right) = 0.$$

Since $\tilde{t}_1 \neq \tilde{t}_2$, necessarily

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} = 1.$$

But, $x'(\tilde{t}_i) > 0$ is equivalent to $c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_i|^2} < 1$, and by Cauchy–Schwarz,

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} \leq \sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_1|^2}}\sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_2|^2}} < 1.$$

[2. Uniqueness] Let $\tilde{t}_1 \neq \tilde{t}_2$ two real solutions of $x(\tilde{t}_i) = x_0$ with $x'(\tilde{t}_i) > 0$. Then,

$$(\tilde{t}_1 - \tilde{t}_2)\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)}\right) = 0.$$

Since $\tilde{t}_1 \neq \tilde{t}_2$, necessarily

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} = 1.$$

But, $x'(\tilde{t}_i) > 0$ is equivalent to $c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_i|^2} < 1$, and by Cauchy–Schwarz,

$$c\int \frac{u^2 L_\infty^R(du)}{(1+u\tilde{t}_1)(1+u\tilde{t}_2)} \leq \sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_1|^2}}\sqrt{c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}_2|^2}} < 1.$$

This is a contradiction, and thus:

$$\tilde{t}_1 = \tilde{t}_2.$$

Application to the SCM model

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$.

Application to the SCM model

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c \int \frac{u \imath\varepsilon_n L_\infty^R(du)}{1 + u\imath\varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$.

Application to the SCM model

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c \int \frac{u\imath\varepsilon_n L_\infty^R(du)}{1 + u\imath\varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$. By dominated convergence, in the limit, $0 = -1!$ so $\tilde{t}^\circ(x_0) \neq 0$.

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c \int \frac{u \imath \varepsilon_n L_\infty^R(du)}{1 + u \imath \varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$. By dominated convergence, in the limit, $0 = -1!$ so $\tilde{t}^\circ(x_0) \neq 0$.

[2. Converse] If $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \operatorname{supp}(L_\infty^R)\}$, then for $z \in \mathbb{C}^+$ and $\tilde{t} \in B(\tilde{t}_0)$,

$$(\tilde{t}(z) - \tilde{t}) \left(1 - c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z))(1 + u\tilde{t})} \right) = (z - x(\tilde{t}))\tilde{t}(z)\tilde{t}.$$

Application to the SCM model

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c\int \frac{u\imath\varepsilon_n L_\infty^R(du)}{1 + u\imath\varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$. By dominated convergence, in the limit, $0 = -1!$ so $\tilde{t}^\circ(x_0) \neq 0$.

[2. Converse] If $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \operatorname{supp}(L_\infty^R)\}$, then for $z \in \mathbb{C}^+$ and $\tilde{t} \in B(\tilde{t}_0)$,

$$(\tilde{t}(z) - \tilde{t})\left(1 - c\int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z))(1 + u\tilde{t})}\right) = (z - x(\tilde{t}))\tilde{t}(z)\tilde{t}.$$

For $z \in \mathbb{C}^+ \to x(\tilde{t})$, by Cauchy–Schwarz and $x'(\tilde{t}) > 0$ for $\tilde{t} \in B(\tilde{t}_0)$ (so that $1 - c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}|^2} > 0$),

Application to the SCM model

[Why $\tilde{t}^{\circ}(x_0) \neq 0$?] Assume $\tilde{t}^{\circ}(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c \int \frac{u \imath\varepsilon_n L_{\infty}^{R}(du)}{1 + u \imath\varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$. By dominated convergence, in the limit, $0 = -1!$ so $\tilde{t}^{\circ}(x_0) \neq 0$.

[2. Converse] If $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \operatorname{supp}(L_{\infty}^{R})\}$, then for $z \in \mathbb{C}^+$ and $\tilde{t} \in B(\tilde{t}_0)$,

$$(\tilde{t}(z) - \tilde{t}) \left(1 - c \int \frac{u^2 L_{\infty}^{R}(du)}{(1 + u\tilde{t}(z))(1 + u\tilde{t})} \right) = (z - x(\tilde{t}))\tilde{t}(z)\tilde{t}.$$

For $z \in \mathbb{C}^+ \to x(\tilde{t})$, by Cauchy–Schwarz and $x'(\tilde{t}) > 0$ for $\tilde{t} \in B(\tilde{t}_0)$ (so that $1 - c \int \frac{u^2 L_{\infty}^{R}(du)}{|1+u\tilde{t}|^2} > 0$),

$$\tilde{t}(z) \to \tilde{t} \in \mathbb{R}.$$

Application to the SCM model

[Why $\tilde{t}^\circ(x_0) \neq 0$?] Assume $\tilde{t}^\circ(x_0) = 0$. By analyticity of $\tilde{t}$ in $B(x_0)$,

$$z_n \imath\varepsilon_n = -1 + c \int \frac{u\imath\varepsilon_n L_\infty^R(du)}{1 + u\imath\varepsilon_n}$$

for some $z_n \to x_0$ and $\imath\varepsilon_n = \tilde{t}(z_n)$ for $\varepsilon_n \to 0$. By dominated convergence, in the limit, $0 = -1!$ so $\tilde{t}^\circ(x_0) \neq 0$.

[2. Converse] If $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \operatorname{supp}(L_\infty^R)\}$, then for $z \in \mathbb{C}^+$ and $\tilde{t} \in B(\tilde{t}_0)$,

$$(\tilde{t}(z) - \tilde{t}) \left(1 - c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z))(1 + u\tilde{t})}\right) = (z - x(\tilde{t}))\tilde{t}(z)\tilde{t}.$$

For $z \in \mathbb{C}^+ \to x(\tilde{t})$, by Cauchy–Schwarz and $x'(\tilde{t}) > 0$ for $\tilde{t} \in B(\tilde{t}_0)$ (so that $1 - c\int \frac{u^2 L_\infty^R(du)}{|1+u\tilde{t}|^2} > 0$),

$$\tilde{t}(z) \to \tilde{t} \in \mathbb{R}.$$

Hence $x(\tilde{t}_0) \notin \operatorname{supp}(\mathcal{F})$. □

Outline

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right)$$

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u - w}.$$

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u-w}.$$

- creates link between Stieltjes transforms $\tilde{t}(z)$ and $r(z)$

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u - w}.$$

- creates link between Stieltjes transforms $\tilde{t}(z)$ and $r(z)$
- so indirectly between observable $\hat{R}_N$ and hidden R_N.

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u-w}.$$

- creates link between Stieltjes transforms $\tilde{t}(z)$ and $r(z)$
- so indirectly between observable $\hat{R}_N$ and hidden R_N.

Question: can we estimate R_N from $\hat{R}_N$ with this link? In particular,

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u-w}.$$

- creates link between Stieltjes transforms $\tilde{t}(z)$ and $r(z)$
- so indirectly between observable $\hat{R}_N$ and hidden R_N.

Question: can we estimate R_N from $\hat{R}_N$ with this link? In particular,

- the spectral measure of R_N?

Motivation

Important formula: We already saw that

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right), \quad r(w) = \int \frac{L_\infty^R(du)}{u-w}.$$

- creates link between Stieltjes transforms $\tilde{t}(z)$ and $r(z)$
- so indirectly between observable $\hat{R}_N$ and hidden R_N.

Question: can we estimate R_N from $\hat{R}_N$ with this link? In particular,

- the spectral measure of R_N?
- the subspaces (eigenvectors) of R_N?

Outline

A few reminders of complex analysis

Proposition (Cauchy Integral)

For $\mathcal{C}$ a closed positively oriented contour in $\mathbb{C}^+$,

A few reminders of complex analysis

Proposition (Cauchy Integral)

For $\mathcal{C}$ a closed positively oriented contour in $\mathbb{C}^+$,

- *if z is inside the surface described by $\mathcal{C}$ and f holomorphic on $U \supset \mathcal{C}$,*

$$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} \frac{f(w)}{w-z} dw = f(z).$$

A few reminders of complex analysis

Proposition (Cauchy Integral)

For $\mathcal{C}$ a closed positively oriented contour in $\mathbb{C}^+$,

- *if z is inside the surface described by $\mathcal{C}$ and f holomorphic on $U \supset \mathcal{C}$,*

$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-z} dw = f(z).$$

- *if z is outside,*

$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-z} dw = 0.$$

Proposition (Cauchy Integral)

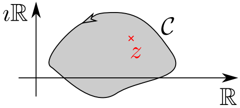
For $\mathcal{C}$ a closed positively oriented contour in $\mathbb{C}^+$,

- *if z is inside the surface described by $\mathcal{C}$ and f holomorphic on $U \supset \mathcal{C}$,*

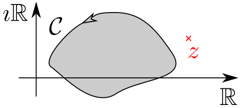
$$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} \frac{f(w)}{w-z}dw = f(z).$$

- *if z is outside,*

$$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} \frac{f(w)}{w-z}dw = 0.$$



$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} \frac{f(w)}{w-z}dw = f(z)$



$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} \frac{f(w)}{w-z}dw = 0$

A few reminders of complex analysis

Proposition (Residue calculus)

Let $\mathcal{C}$ as above and f holomorphic on an open set including $\mathcal{C}$. Let $a_1, \ldots, a_L$ be the singularities of f within the surface described by $\mathcal{C}$.

A few reminders of complex analysis

Proposition (Residue calculus)

Let $\mathcal{C}$ as above and f holomorphic on an open set including $\mathcal{C}$. Let $a_1, \ldots, a_L$ be the singularities of f within the surface described by $\mathcal{C}$. Then,

$$\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(z)dz = \sum_{i=1}^{L} \mathrm{Res}(f, a_i)$$

A few reminders of complex analysis

Proposition (Residue calculus)

Let $\mathcal{C}$ as above and f holomorphic on an open set including $\mathcal{C}$. Let $a_1, \ldots, a_L$ be the singularities of f within the surface described by $\mathcal{C}$. Then,

$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(z) dz = \sum_{i=1}^{L} \mathrm{Res}(f, a_i)$$

where

$$\mathrm{Res}(f, a) = \frac{1}{(n-1)!} \lim_{z \to a} \frac{d^{n-1}}{dz^{n-1}} ((z-a)^n f(z))$$

for $n \in \{1, 2, \ldots\}$ the smallest index such that the limit is finite.

A few reminders of complex analysis

Proposition (Residue calculus)

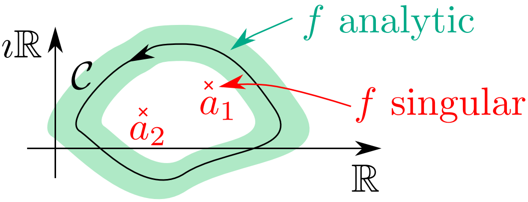
Let $\mathcal{C}$ as above and f holomorphic on an open set including $\mathcal{C}$. Let $a_1, \ldots, a_L$ be the singularities of f within the surface described by $\mathcal{C}$. Then,

$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(z)dz = \sum_{i=1}^{L} \mathrm{Res}(f, a_i)$$

where

$$\mathrm{Res}(f, a) = \frac{1}{(n-1)!} \lim_{z \to a} \frac{d^{n-1}}{dz^{n-1}}((z-a)^n f(z))$$

for $n \in \{1, 2, \ldots\}$ the smallest index such that the limit is finite.



$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w)dw = \mathrm{Res}(f, a_1) + \mathrm{Res}(f, a_2)$$

Proposition (Vitali's convergence theorem)

Let $(f_n)_{n\in\mathbb{N}}$ *a series of analytic functions in* $D \subset \mathbb{C}$ *such that* $|f_n(z)| \leq M$ *for all* $z \in D$.

Proposition (Vitali's convergence theorem)

Let $(f_n)_{n\in\mathbb{N}}$ a series of analytic functions in $D \subset \mathbb{C}$ such that $|f_n(z)| \leq M$ for all $z \in D$.
Assume that $f_n(z_k) \to f(z_k)$ for all $z_1, z_2, \ldots$ in a set with a limit point in D.

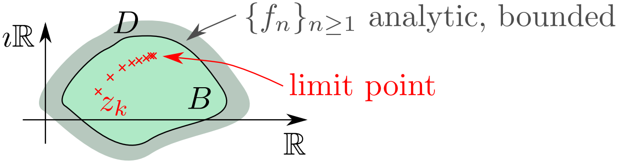
A few reminders of complex analysis

Proposition (Vitali's convergence theorem)

Let $(f_n)_{n\in\mathbb{N}}$ a series of analytic functions in $D \subset \mathbb{C}$ such that $|f_n(z)| \leq M$ for all $z \in D$.
Assume that $f_n(z_k) \to f(z_k)$ for all $z_1, z_2, \ldots$ in a set with a limit point in D.
Then f_n converges uniformly on all $B \subset D$ strictly and f is analytic.

Proposition (Vitali's convergence theorem)

Let $(f_n)_{n\in\mathbb{N}}$ a series of analytic functions in $D \subset \mathbb{C}$ such that $|f_n(z)| \leq M$ for all $z \in D$.
Assume that $f_n(z_k) \to f(z_k)$ for all $z_1, z_2, \ldots$ in a set with a limit point in D.
Then f_n converges uniformly on all $B \subset D$ strictly and f is analytic.



Outline

Objective

Linear functional of the $\lambda_i(R_N)$:

Objective

Linear functional of the $\lambda_i(R_N)$: We want to estimate

$$G(f) = \int f(t) L_\infty^R(dt)$$

Objective

Linear functional of the $\lambda_i(R_N)$: We want to estimate

$$G(f) = \int f(t) L_\infty^R(dt) = \frac{1}{N}\sum_{i=1}^{n} f(\lambda_i(R_N))$$

for extensible in a complex analytic function near $\mathrm{supp}(L_\infty^R)$.

Objective

Linear functional of the $\lambda_i(R_N)$: We want to estimate

$$G(f) = \int f(t) L_\infty^R(dt) = \frac{1}{N} \sum_{i=1}^{n} f(\lambda_i(R_N))$$

for extensible in a complex analytic function near $\mathrm{supp}(L_\infty^R)$.

Examples:

- $f(t) = t$ gives $\frac{1}{N}\mathrm{tr}(R_N)$

Objective

Linear functional of the $\lambda_i(R_N)$: We want to estimate

$$G(f) = \int f(t) L_\infty^R(dt) = \frac{1}{N} \sum_{i=1}^n f(\lambda_i(R_N))$$

for extensible in a complex analytic function near $\operatorname{supp}(L_\infty^R)$.

Examples:

- $f(t) = t$ gives $\frac{1}{N}\operatorname{tr}(R_N)$
- $f(t) = t^k$ gives $\frac{1}{N}\operatorname{tr}(R_N^k)$

Objective

Linear functional of the $\lambda_i(R_N)$: We want to estimate

$$G(f) = \int f(t) L_\infty^R(dt) = \frac{1}{N} \sum_{i=1}^{n} f(\lambda_i(R_N))$$

for extensible in a complex analytic function near $\mathrm{supp}(L_\infty^R)$.

Examples:

- $f(t) = t$ gives $\frac{1}{N}\mathrm{tr}(R_N)$
- $f(t) = t^k$ gives $\frac{1}{N}\mathrm{tr}(R_N^k)$
- $f(t) = \log(1 + st)$ gives $\det(I_N + sR_N)$ (up to taking $\exp(N\cdot)$)

Central idea

Central idea: By Cauchy's theorem,

$$G(f) = \int \left(\frac{1}{2\pi \imath} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_\infty^R(dt)$$

Central idea

Central idea: By Cauchy's theorem,

$$G(f) = \int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_\infty^R(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w) r(w) dw$$

with $r(w) = \int (t-w)^{-1} L_\infty^R(dt)$ Stieltjes transform of L_∞^R.

Central idea

Central idea: By Cauchy's theorem,

$$G(f) = \int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_\infty^R(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w) r(w) dw$$

with $r(w) = \int (t-w)^{-1} L_\infty^R(dt)$ Stieltjes transform of L_∞^R.

- we then link $r(z)$ to $\tilde{t}(z)$ to relate $G(f)$ to $\tilde{t}(z)$:

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r \left(-\frac{1}{\tilde{t}(z)} \right)$$

Central idea

Central idea: By Cauchy's theorem,

$$G(f) = \int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_\infty^R(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w) r(w) dw$$

with $r(w) = \int (t-w)^{-1} L_\infty^R(dt)$ Stieltjes transform of L_∞^R.

- we then link $r(z)$ to $\tilde{t}(z)$ to relate $G(f)$ to $\tilde{t}(z)$:

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)} \right)$$

- or even better:

$$r\left(-\frac{1}{\tilde{t}}(z) \right) = -z\tilde{t}(z)t(z)$$

with $-czt(z) = c - 1 - z\tilde{t}(z)$.

Central idea

Central idea: By Cauchy's theorem,

$$G(f) = \int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_\infty^R(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w) r(w) dw$$

with $r(w) = \int (t-w)^{-1} L_\infty^R(dt)$ Stieltjes transform of L_∞^R.

- we then link $r(z)$ to $\tilde{t}(z)$ to relate $G(f)$ to $\tilde{t}(z)$:

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right)$$

- or even better:

$$r\left(-\frac{1}{\tilde{t}}(z)\right) = -z\tilde{t}(z)t(z)$$

with $-czt(z) = c - 1 - z\tilde{t}(z)$.

- then, variable change!!!

$$w = -\frac{1}{\tilde{t}(z)}$$

Central idea

The problem . . . :

Central idea

The problem . . . :

✓ $\tilde{t}$, t and r analytic on $\mathbb{C}^+$ and $\mathbb{C}^-$,

Central idea

The problem . . . :

✓ $\tilde{t}$, t and r analytic on $\mathbb{C}^+$ and $\mathbb{C}^-$, $\Rightarrow$ variable change

$$w = -\frac{1}{\tilde{t}(z)}$$

always possible on $\mathbb{C}^+$ and $\mathbb{C}^-$

Central idea

The problem . . . :

✓ $\tilde{t}$, t and r analytic on $\mathbb{C}^+$ and $\mathbb{C}^-$, $\Rightarrow$ variable change

$$w = -\frac{1}{\tilde{t}(z)}$$

always possible on $\mathbb{C}^+$ and $\mathbb{C}^-$

✗ but $\mathcal{C}$ crosses the real axis!

Central idea

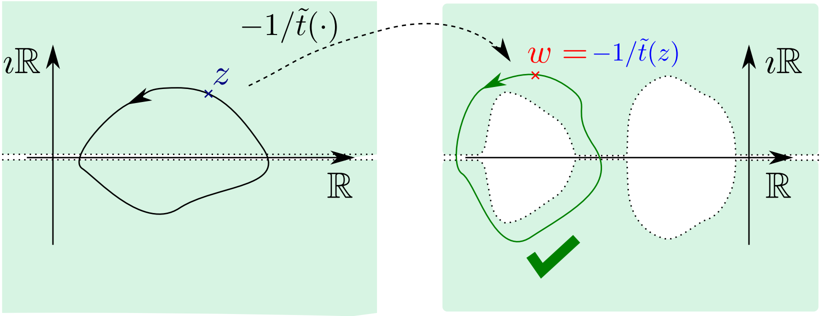
The problem . . . :

✓ $\tilde{t}$, t and r analytic on $\mathbb{C}^+$ and $\mathbb{C}^-$, $\Rightarrow$ variable change

$$w = -\frac{1}{\tilde{t}(z)}$$

always possible on $\mathbb{C}^+$ and $\mathbb{C}^-$

✗ but $\mathcal{C}$ crosses the real axis! What happens to $w = -\frac{1}{\tilde{t}(z)}$ when crossing $\mathbb{R}$?



Central idea

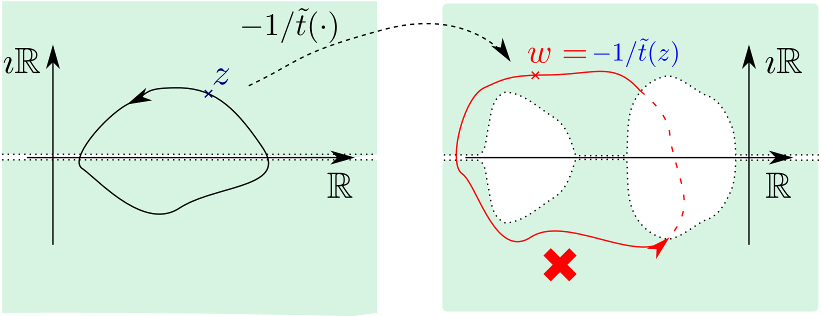
The problem . . .:

✓ $\tilde{t}$, t and r analytic on $\mathbb{C}^+$ and $\mathbb{C}^-$, $\Rightarrow$ variable change

$$w = -\frac{1}{\tilde{t}(z)}$$

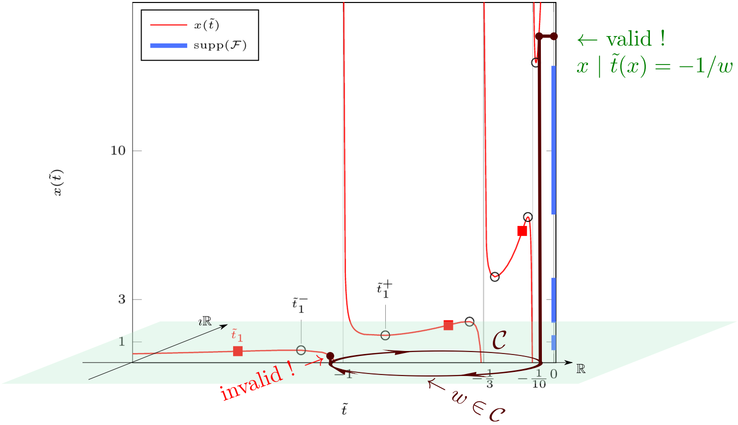
always possible on $\mathbb{C}^+$ and $\mathbb{C}^-$

✗ but $\mathcal{C}$ crosses the real axis! What happens to $w = -\frac{1}{\tilde{t}(z)}$ when crossing $\mathbb{R}$?



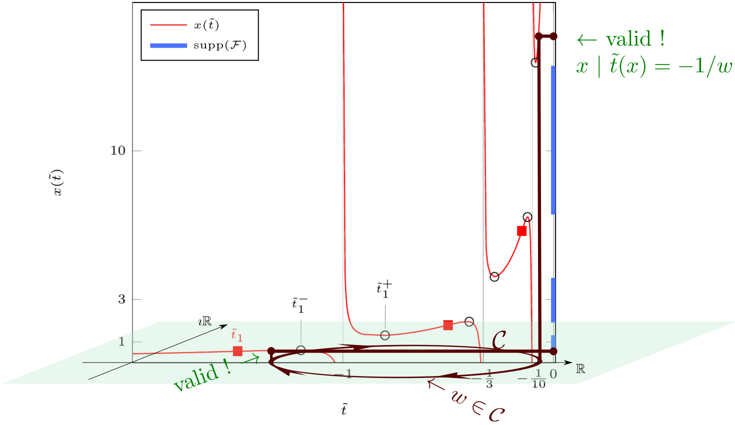
Central idea

Back to $x(\tilde{t})$!: Recall $L_\infty^R = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_3 + \frac{1}{3}\delta_{10}$



Central idea

Back to $x(\tilde{t})$!: Recall $L_\infty^R = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_3 + \frac{1}{3}\delta_{10}$



Main result

Consequence: not every contour $\mathcal{C}$ is valid!

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\operatorname{supp}(\mathcal{F})$

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\mathrm{supp}(\mathcal{F})$
   $\Rightarrow$ So crossing $\mathbb{R}$ outside the support!

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\operatorname{supp}(\mathcal{F})$
   $\Rightarrow$ So crossing $\mathbb{R}$ outside the support!
2. define $\mathcal{C} = -1/\tilde{t}(\mathcal{C}_{\mathcal{F}})$, which is always defined!

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\operatorname{supp}(\mathcal{F})$
   $\Rightarrow$ So crossing $\mathbb{R}$ outside the support!
2. define $\mathcal{C} = -1/\tilde{t}(\mathcal{C}_{\mathcal{F}})$, which is always defined!
3. from the drawing of $x(\cdot) \leftrightarrow \tilde{t}(\cdot)$, $\mathcal{C}$ "contains" $\operatorname{supp}(L_{\infty}^{R})$

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\operatorname{supp}(\mathcal{F})$
   $\Rightarrow$ So crossing $\mathbb{R}$ outside the support!
2. define $\mathcal{C} = -1/\tilde{t}(\mathcal{C}_{\mathcal{F}})$, which is always defined!
3. from the drawing of $x(\cdot) \leftrightarrow \tilde{t}(\cdot)$, $\mathcal{C}$ "contains" $\operatorname{supp}(L_\infty^R)$ $\Rightarrow$ It is valid!

Main result

Consequence: not every contour $\mathcal{C}$ is valid!

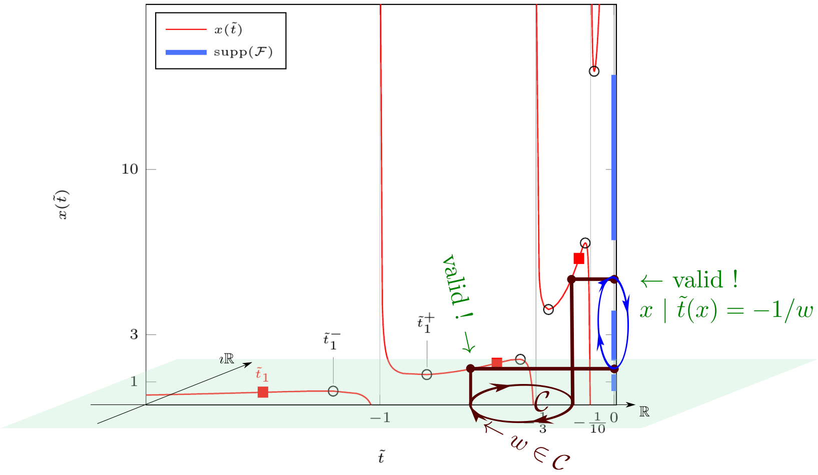
The method: one must go backwards:

1. design $\mathcal{C}_{\mathcal{F}}$ **any contour** circling around $\mathrm{supp}(\mathcal{F})$
   $\Rightarrow$ So crossing $\mathbb{R}$ outside the support!
2. define $\mathcal{C} = -1/\tilde{t}(\mathcal{C}_{\mathcal{F}})$, which is always defined!
3. from the drawing of $x(\cdot) \leftrightarrow \tilde{t}(\cdot)$, $\mathcal{C}$ "contains" $\mathrm{supp}(L_{\infty}^{R})$ $\Rightarrow$ It is valid!

Remark: In fact, not always! e.g., estimate $\frac{1}{N}\mathrm{tr}(R_N^{-1})$ from $\hat{R}_N$ when $n < N$?

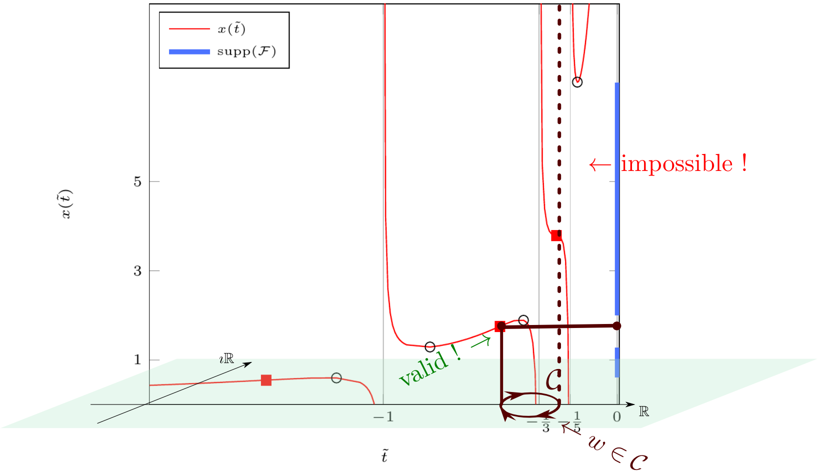
Going further...

What if we want to estimate $f(1)$, $f(3)$? : in $L_\infty^R = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_3 + \frac{1}{3}\delta_5$



Going further...

What if we want to estimate $f(1)$, $f(3)$? : in $L_\infty^R = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_3 + \frac{1}{3}\delta_5$



Going further...

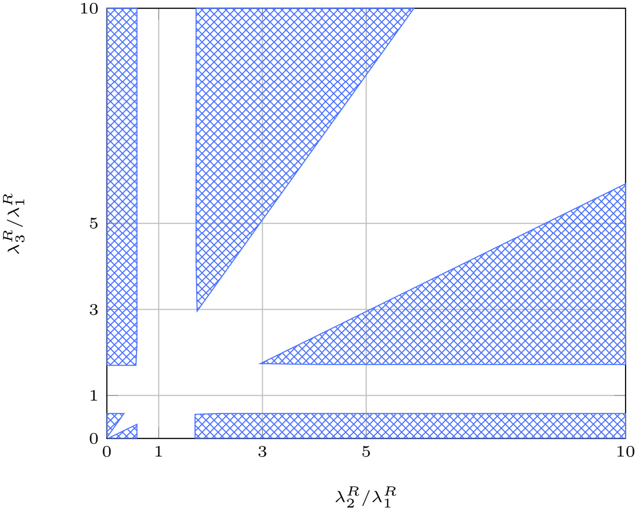


Figure: Subsets of $\lambda_1^R \leq \lambda_2^R \leq \lambda_3^R$ (hatched region) for which detectability condition over $L_\infty^R = \frac{1}{3}(\delta_{\lambda_1^R} + \delta_{\lambda_2^R} + \delta_{\lambda_3^R})$ is satisfied, $c = 1/10$.

Outline

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_{\infty}^{R}(du)$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_{\infty}^{R}(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_{\mathcal{F}})}$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_\infty^R(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_\mathcal{F})}$$

for $\mathcal{C}_\mathcal{F}$ any contour circling around $\mathrm{supp}(\mathcal{F})$.

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_\infty^R(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_\mathcal{F})}$$

for $\mathcal{C}_\mathcal{F}$ any contour circling around $\operatorname{supp}(\mathcal{F})$.

The strategy: for $w = -\frac{1}{\tilde{t}(z)}$,

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{t}(z)}\right) r\left(-\frac{1}{\tilde{t}(z)}\right)\frac{\tilde{t}'(z)}{\tilde{t}(z)^2} dz.$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_\infty^R(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_{\mathcal{F}})}$$

for $\mathcal{C}_{\mathcal{F}}$ any contour circling around $\operatorname{supp}(\mathcal{F})$.

The strategy: for $w = -\frac{1}{\tilde{t}(z)}$,

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_{\mathcal{F}}} f\left(-\frac{1}{\tilde{t}(z)}\right) r\left(-\frac{1}{\tilde{t}(z)}\right)\frac{\tilde{t}'(z)}{\tilde{t}(z)^2} dz.$$

- but we know that

$$r\left(-\frac{1}{\tilde{t}(z)}\right) = -z\tilde{t}(z)t(z)$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w)r(w)dw, \quad r(w) = \int (u-w)^{-1} L_\infty^R(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_\mathcal{F})}$$

for $\mathcal{C}_\mathcal{F}$ any contour circling around $\mathrm{supp}(\mathcal{F})$.

The strategy: for $w = -\frac{1}{\tilde{t}(z)}$,

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{t}(z)}\right) r\left(-\frac{1}{\tilde{t}(z)}\right)\frac{\tilde{t}'(z)}{\tilde{t}(z)^2}dz.$$

- but we know that

$$r\left(-\frac{1}{\tilde{t}(z)}\right) = -z\tilde{t}(z)t(z)$$

- so that

$$G(f) = \frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{t}(z)}\right) zt(z)\frac{\tilde{t}'(z)}{\tilde{t}(z)}dz.$$

Statistical inference

Back to the original problem:

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}} f(w) r(w) dw, \quad r(w) = \int (u-w)^{-1} L_\infty^R(du)$$

now with

$$\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}_\mathcal{F})}$$

for $\mathcal{C}_\mathcal{F}$ any contour circling around $\operatorname{supp}(\mathcal{F})$.

The strategy: for $w = -\frac{1}{\tilde{t}(z)}$,

$$G(f) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{t}(z)}\right) r\left(-\frac{1}{\tilde{t}(z)}\right)\frac{\tilde{t}'(z)}{\tilde{t}(z)^2} dz.$$

- but we know that

$$r\left(-\frac{1}{\tilde{t}(z)}\right) = -z\tilde{t}(z)t(z)$$

- so that

$$G(f) = \frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{t}(z)}\right) zt(z)\frac{\tilde{t}'(z)}{\tilde{t}(z)} dz.$$

and we are fully in the domain of **limiting** observables!

Statistical inference

From limit to observations: recall that

$$\tilde{g}_n(z) \equiv \frac{1}{n}\operatorname{tr}(\Sigma_n^*\Sigma_n - zI_n)^{-1} \xrightarrow{\text{a.s.}} \tilde{t}(z)$$

for all $z \in \mathbb{C}^+$ (equivalent to $g_n(z) = \frac{1}{N}\operatorname{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1}$).

Statistical inference

From limit to observations: recall that

$$\tilde{g}_n(z) \equiv \frac{1}{n}\mathrm{tr}(\Sigma_n^*\Sigma_n - zI_n)^{-1} \xrightarrow{\text{a.s.}} \tilde{t}(z)$$

for all $z \in \mathbb{C}^+$ (equivalent to $g_n(z) = \frac{1}{N}\mathrm{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1}$).

- tempting to replace $\tilde{t}(z)$ by $\tilde{g}_n(z)$ in expression of $G(f)$!

Statistical inference

From limit to observations: recall that

$$\tilde{g}_n(z) \equiv \frac{1}{n}\mathrm{tr}(\Sigma_n^*\Sigma_n - zI_n)^{-1} \xrightarrow{\text{a.s.}} \tilde{t}(z)$$

for all $z \in \mathbb{C}^+$ (equivalent to $g_n(z) = \frac{1}{N}\mathrm{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1}$).

- tempting to replace $\tilde{t}(z)$ by $\tilde{g}_n(z)$ in expression of $G(f)$!
- but, for this, we need a uniform convergence: (almost surely! **Beware of quantifiers!**)

$$\max_{z\in\mathcal{C}_{\mathcal{F}}} |\tilde{g}_n(z) - \tilde{t}(z)| \xrightarrow{\text{a.s.}} 0.$$

Statistical inference

From limit to observations: recall that

$$\tilde{g}_n(z) \equiv \frac{1}{n}\operatorname{tr}(\Sigma_n^*\Sigma_n - zI_n)^{-1} \xrightarrow{\text{a.s.}} \tilde{t}(z)$$

for all $z \in \mathbb{C}^+$ (equivalent to $g_n(z) = \frac{1}{N}\operatorname{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1}$).

- tempting to replace $\tilde{t}(z)$ by $\tilde{g}_n(z)$ in expression of $G(f)$!
- but, for this, we need a uniform convergence: (almost surely! **Beware of quantifiers!**)

$$\max_{z\in\mathcal{C}_\mathcal{F}} |\tilde{g}_n(z) - \tilde{t}(z)| \xrightarrow{\text{a.s.}} 0.$$

- two problems:
  1. $\mathcal{C}_\mathcal{F}$ crosses $\mathbb{R}$!

Statistical inference

From limit to observations: recall that

$$\tilde{g}_n(z) \equiv \frac{1}{n}\mathrm{tr}(\Sigma_n^* \Sigma_n - zI_n)^{-1} \xrightarrow{\text{a.s.}} \tilde{t}(z)$$

for all $z \in \mathbb{C}^+$ (equivalent to $g_n(z) = \frac{1}{N}\mathrm{tr}(\Sigma_n \Sigma_n^* - zI_N)^{-1}$).

- tempting to replace $\tilde{t}(z)$ by $\tilde{g}_n(z)$ in expression of $G(f)$!
- but, for this, we need a uniform convergence: (almost surely! **Beware of quantifiers!**)

$$\max_{z \in \mathcal{C}_\mathcal{F}} |\tilde{g}_n(z) - \tilde{t}(z)| \xrightarrow{\text{a.s.}} 0.$$

- two problems:
  1. $\mathcal{C}_\mathcal{F}$ crosses $\mathbb{R}$!
  2. does point-wise imply uniform convergence?

Statistical inference

Intermediary (but fundamental) result:

Theorem ("No eigenvalue outside the support")

Assume that $\mathbb{E}[|X_{ij}|^4] < \infty$ *and*

$$\max_{1 \le i \le N} \left\{ \operatorname{dist}(\lambda_i(R_N), \operatorname{supp}(L_\infty^R)) \right\} \to 0. \quad \textit{(no spike condition)}$$

Statistical inference

Intermediary (but fundamental) result:

Theorem ("No eigenvalue outside the support")

Assume that $\mathbb{E}[|X_{ij}|^4] < \infty$ *and*

$$\max_{1 \leq i \leq N} \left\{ \operatorname{dist}(\lambda_i(R_N), \operatorname{supp}(L_\infty^R)) \right\} \to 0. \quad \textit{(no spike condition)}$$

For $\varepsilon > 0$, *let* $A \supset \operatorname{supp}(\mathcal{F})$ *the* ε*-opening of* $\operatorname{supp}(\mathcal{F})$.

Intermediary (**but fundamental**) **result**:

Theorem ("No eigenvalue outside the support")

Assume that $\mathbb{E}[|X_{ij}|^4] < \infty$ *and*

$$\max_{1 \le i \le N} \left\{ \mathrm{dist}(\lambda_i(R_N), \mathrm{supp}(L_\infty^R)) \right\} \to 0. \quad \textit{(no spike condition)}$$

For $\varepsilon > 0$, *let* $A \supset \mathrm{supp}(\mathcal{F})$ *the* ε*-opening of* $\mathrm{supp}(\mathcal{F})$. *Then, for all large* n *almost surely,* $\hat{R}_N = \Sigma_n \Sigma_n^*$ *has no eigenvalue in* A.

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

Statistical inference

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_{\mathcal{F}}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_{\mathcal{F}}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence:

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_{\mathcal{F}}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence: let

$$\Omega = \bigcap_{z_i} \Omega_{z_i}$$

for $\{z_i\}_{i \geq 1}$ having a limit point in $\mathcal{C}_{\mathcal{F}}$.

Statistical inference

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_\mathcal{F}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence: let

$$\Omega = \bigcap_{z_i} \Omega_{z_i}$$

for $\{z_i\}_{i\geq 1}$ having a limit point in $\mathcal{C}_\mathcal{F}$.

$\Rightarrow$ This is a probability 1 set!

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_\mathcal{F}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large,
(problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence:
let

$$\Omega = \bigcap_{z_i} \Omega_{z_i}$$

for $\{z_i\}_{i\geq 1}$ having a limit point in $\mathcal{C}_\mathcal{F}$.

$\Rightarrow$ This is a probability 1 set!
- On Ω, apply Vitali:

$$\sup_{z\in\mathcal{C}_\mathcal{F}} |\tilde{g}_n(z) - \tilde{t}(z)| \to 0$$

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_\mathcal{F}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence: let
$$\Omega = \bigcap_{z_i} \Omega_{z_i}$$
for $\{z_i\}_{i\geq 1}$ having a limit point in $\mathcal{C}_\mathcal{F}$.

  $\Rightarrow$ This is a probability 1 set!
- On Ω, apply Vitali:
$$\sup_{z\in\mathcal{C}_\mathcal{F}} |\tilde{g}_n(z) - \tilde{t}(z)| \to 0$$
and thus, the convergence is almost sure!

Completing the calculus: let us then suppose R_N has no isolated eigenvalue,

- by the theorem, for $z \in \mathcal{C}_\mathcal{F}$, on Ω_z of proba 1, $\tilde{g}_n(z)$ bounded for all n large, (problem of crossing $\mathbb{R}$ solved!)
- we now apply Vitali's theorem to move from point-wise to uniform convergence: let
$$\Omega = \bigcap_{z_i} \Omega_{z_i}$$
for $\{z_i\}_{i\geq 1}$ having a limit point in $\mathcal{C}_\mathcal{F}$.

  $\Rightarrow$ This is a probability 1 set!
- On Ω, apply Vitali:
$$\sup_{z\in\mathcal{C}_\mathcal{F}} |\tilde{g}_n(z) - \tilde{t}(z)| \to 0$$
and thus, the convergence is almost sure!
- we conclude:
$$G(f) - \frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mathcal{F}} f\left(-\frac{1}{\tilde{g}_n(z)}\right) z g_n(z)\frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)} dz \xrightarrow{\text{a.s.}} 0.$$

Statistical inference

From complex contour to exact calculus: We already have a consistent estimator of $G(f)$, but not intuitive, numerical evaluation!

Statistical inference

From complex contour to exact calculus: We already have a consistent estimator of $G(f)$, but not intuitive, numerical evaluation!

- we can go further by noticing (with $\lambda_1 \geq \ldots \geq \lambda_n$ vaps of $\Sigma_n^* \Sigma_n$),

$$\tilde{g}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z}$$

$$g_n(z) = \frac{n}{N} \tilde{g}_n(z) + \frac{N-n}{Nz}$$

$$\tilde{g}_n'(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(\lambda_i - z)^2}$$

all rational functions!

From complex contour to exact calculus: We already have a consistent estimator of $G(f)$, but not intuitive, numerical evaluation!

- we can go further by noticing (with $\lambda_1 \geq \ldots \geq \lambda_n$ vaps of $\Sigma_n^* \Sigma_n$),

$$\tilde{g}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z}$$
$$g_n(z) = \frac{n}{N} \tilde{g}_n(z) + \frac{N-n}{Nz}$$
$$\tilde{g}_n'(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(\lambda_i - z)^2}$$

  all rational functions!

- if $f(w)$ "simple", we can use residue calculus!

Example ($f(x) = x^2$)

For $f(x) = x^2$, we know, with probability 1,

Example ($f(x) = x^2$)

For $f(x) = x^2$, we know, with probability 1,

$$G(f) \quad = \quad \frac{1}{2\pi\imath}\oint_{\mathcal{C}_{\mathcal{F}}} z g_n(z)\frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^3}dz + o(1)$$

Example $(f(x) = x^2)$

For $f(x) = x^2$, we know, with probability 1,

$$\begin{aligned} G(f) &= \frac{1}{2\pi\imath}\oint_{\mathcal{C}_{\mathcal{F}}} z g_n(z)\frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^3}dz + o(1) \\ &= \frac{1}{2\pi\imath}\oint_{\mathcal{C}_{\mathcal{F}}} \left(\frac{n}{N}\frac{z\tilde{g}_n'(z)}{\tilde{g}_n(z)^2} + \underbrace{\frac{n-N}{N}\frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^2}}_{0} \right) dz + o(1) \end{aligned}$$

Example ($f(x) = x^2$)

For $f(x) = x^2$, we know, with probability 1,

$$\begin{aligned} G(f) &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} z g_n(z) \frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^3} dz + o(1) \\ &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} \left(\frac{n}{N} \frac{z\tilde{g}_n'(z)}{\tilde{g}_n(z)^2} + \underbrace{\frac{n-N}{N} \frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^2}}_{0} \right) dz + o(1) \\ &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} \frac{n}{N} \frac{z\tilde{g}_n'(z)}{\tilde{g}_n(z)^2} dz + o(1). \end{aligned}$$

Example ($f(x) = x^2$)

For $f(x) = x^2$, we know, with probability 1,

$$
\begin{aligned}
G(f) &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} z g_n(z) \frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^3} dz + o(1) \\
&= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} \left(\frac{n}{N} \frac{z\tilde{g}_n'(z)}{\tilde{g}_n(z)^2} + \underbrace{\frac{n-N}{N} \frac{\tilde{g}_n'(z)}{\tilde{g}_n(z)^2}}_{0} \right) dz + o(1) \\
&= \frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} \frac{n}{N} \frac{z\tilde{g}_n'(z)}{\tilde{g}_n(z)^2} dz + o(1).
\end{aligned}
$$

By integration by parts, this is, with probability 1:

$$
G(f) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_{\mathcal{F}}} \frac{n}{N} \frac{1}{\tilde{g}_n(z)} dz + o(1).
$$

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_{\mathcal{F}}$ are such that $\tilde{g}_n(z) = 0$.

Statistical inference

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_{\mathcal{F}}$ are such that $\tilde{g}_n(z) = 0$.

Useful lemma:

Lemma

Let $\Lambda \in \mathbb{R}^{n\times n}$ diagonal and $a \in \mathbb{R}^n$. Then, the eigenvalues of $\Lambda - aa^*$ are either eigenvalues of Λ or the roots of $1 = a^*(\Lambda - xI_n)^{-1}a$.

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_{\mathcal{F}}$ are such that $\tilde{g}_n(z) = 0$.

Useful lemma:

Lemma

Let $\Lambda \in \mathbb{R}^{n\times n}$ diagonal and $a \in \mathbb{R}^n$. Then, the eigenvalues of $\Lambda - aa^*$ are either eigenvalues of Λ or the roots of $1 = a^*(\Lambda - xI_n)^{-1}a$.

Proof.

Let x not an eigenvalue of Λ, then by Sylverster's identity,

$$\det(\Lambda - aa^* - xI_n) = \underbrace{\det(\Lambda - xI_n)}_{\neq 0}\det(I_n - aa^*(\Lambda - xI_n)^{-1}) \propto 1 - a^*(\Lambda - xI_n)^{-1}a.$$

□

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_{\mathcal{F}}$ are such that $\tilde{g}_n(z) = 0$.

Useful lemma:

Lemma

Let $\Lambda \in \mathbb{R}^{n\times n}$ diagonal and $a \in \mathbb{R}^n$. Then, the eigenvalues of $\Lambda - aa^*$ are either eigenvalues of Λ or the roots of $1 = a^*(\Lambda - xI_n)^{-1}a$.

Proof.

Let x not an eigenvalue of Λ, then by Sylverster's identity,

$$\det(\Lambda - aa^* - xI_n) = \underbrace{\det(\Lambda - xI_n)}_{\neq 0} \det(I_n - aa^*(\Lambda - xI_n)^{-1}) \propto 1 - a^*(\Lambda - xI_n)^{-1}a.$$

$\square$

Calculus continuation: apply the lemma for $\Lambda = \operatorname{diag}(\lambda_i)$ and $a = \sqrt{\lambda/n}$ with λ the vector of λ_i's.

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_{\mathcal{F}}$ are such that $\tilde{g}_n(z) = 0$.

Useful lemma:

Lemma

Let $\Lambda \in \mathbb{R}^{n\times n}$ diagonal and $a \in \mathbb{R}^n$. Then, the eigenvalues of $\Lambda - aa^*$ are either eigenvalues of Λ or the roots of $1 = a^*(\Lambda - xI_n)^{-1}a$.

Proof.

Let x not an eigenvalue of Λ, then by Sylverster's identity,

$$\det(\Lambda - aa^* - xI_n) = \underbrace{\det(\Lambda - xI_n)}_{\neq 0}\det(I_n - aa^*(\Lambda - xI_n)^{-1}) \propto 1 - a^*(\Lambda - xI_n)^{-1}a.$$

$\square$

Calculus continuation: apply the lemma for $\Lambda = \operatorname{diag}(\lambda_i)$ and $a = \sqrt{\lambda/n}$ with λ the vector of λ_i's.
$\Rightarrow$ The eigenvalues of

$$\Lambda - \frac{1}{n}\sqrt{\lambda}\sqrt{\lambda}$$

are the roots x_j of $1 = \frac{1}{n}\sum_{i=1}^n \lambda_i(\lambda_i - x)^{-1}$,

Example ($f(x) = x^2$ (continued))

Expand $\tilde{g}_n(z)$ as a rational function: the poles inside $\mathcal{C}_\mathcal{F}$ are such that $\tilde{g}_n(z) = 0$.

Useful lemma:

Lemma

Let $\Lambda \in \mathbb{R}^{n \times n}$ diagonal and $a \in \mathbb{R}^n$. Then, the eigenvalues of $\Lambda - aa^*$ are either eigenvalues of Λ or the roots of $1 = a^*(\Lambda - xI_n)^{-1}a$.

Proof.

Let x not an eigenvalue of Λ, then by Sylverster's identity,

$$\det(\Lambda - aa^* - xI_n) = \underbrace{\det(\Lambda - xI_n)}_{\neq 0} \det(I_n - aa^*(\Lambda - xI_n)^{-1}) \propto 1 - a^*(\Lambda - xI_n)^{-1}a.$$

$\square$

Calculus continuation: apply the lemma for $\Lambda = \operatorname{diag}(\lambda_i)$ and $a = \sqrt{\lambda/n}$ with λ the vector of λ_i's.
$\Rightarrow$ The eigenvalues of

$$\Lambda - \frac{1}{n}\sqrt{\lambda}\sqrt{\lambda}$$

are the roots x_j of $1 = \frac{1}{n}\sum_{i=1}^n \lambda_i(\lambda_i - x)^{-1}$, or equivalently of
$0 = \frac{1}{n}\sum_{i=1}^n (\lambda_i - x)^{-1} = \tilde{g}_n(x)$.

Example ($f(x) = x^2$ (continued))

Position of the roots? By Weyl's interlacing lemma, the λ_i's are interlaced with the roots x_i of $0 = g_n(x)$.

Example ($f(x) = x^2$ (continued))

Position of the roots? By Weyl's interlacing lemma, the λ_i's are interlaced with the roots x_i of $0 = g_n(x)$.
$\Rightarrow$ They are all inside the contour $\mathcal{C}_{\mathcal{F}}$!

Example ($f(x) = x^2$ (continued))

Position of the roots? By Weyl's interlacing lemma, the λ_i's are interlaced with the roots x_i of $0 = g_n(x)$.
$\Rightarrow$ They are all inside the contour $\mathcal{C}_{\mathcal{F}}$!

Final step, the residue calculus:

$$\mathrm{Res}(x_i) = \lim_{z \to x_i} \frac{n}{N} \frac{z - x_i}{\tilde{g}_n(z)}.$$

Example ($f(x) = x^2$ (continued))

Position of the roots? By Weyl's interlacing lemma, the λ_i's are interlaced with the roots x_i of $0 = g_n(x)$.
$\Rightarrow$ They are all inside the contour $\mathcal{C}_{\mathcal{F}}$!

Final step, the residue calculus:

$$\mathrm{Res}(x_i) = \lim_{z \to x_i} \frac{n}{N} \frac{z - x_i}{\tilde{g}_n(z)}.$$

By Taylor (or l'Hospital rule),

$$\mathrm{Res}(x_i) = \lim_{z \to x_i} \frac{n}{N} \frac{1}{\tilde{g}'_n(z)} = \frac{n}{N} \frac{1}{\tilde{g}'_n(x_i)}.$$

Example ($f(x) = x^2$ (continued))

Position of the roots? By Weyl's interlacing lemma, the λ_i's are interlaced with the roots x_i of $0 = g_n(x)$.
$\Rightarrow$ They are all inside the contour $\mathcal{C}_{\mathcal{F}}$!

Final step, the residue calculus:

$$\operatorname{Res}(x_i) = \lim_{z \to x_i} \frac{n}{N} \frac{z - x_i}{\tilde{g}_n(z)}.$$

By Taylor (or l'Hospital rule),

$$\operatorname{Res}(x_i) = \lim_{z \to x_i} \frac{n}{N} \frac{1}{\tilde{g}_n'(z)} = \frac{n}{N} \frac{1}{\tilde{g}_n'(x_i)}.$$

And we conclude, with probability 1,

$$G(f) = \sum_{i \in S} \frac{n}{N} \frac{1}{\tilde{g}_n'(x_i)} + o(1)$$

with $S = \{i \mid \lambda_i \text{ inside } \mathcal{C}_{\mathcal{F}}\}$.

Some consequences: evaluating the eigenvalues of $R_N = \mathrm{diag}(1, \ldots, 1, 3, \ldots, 3, 7, \ldots, 7)$?

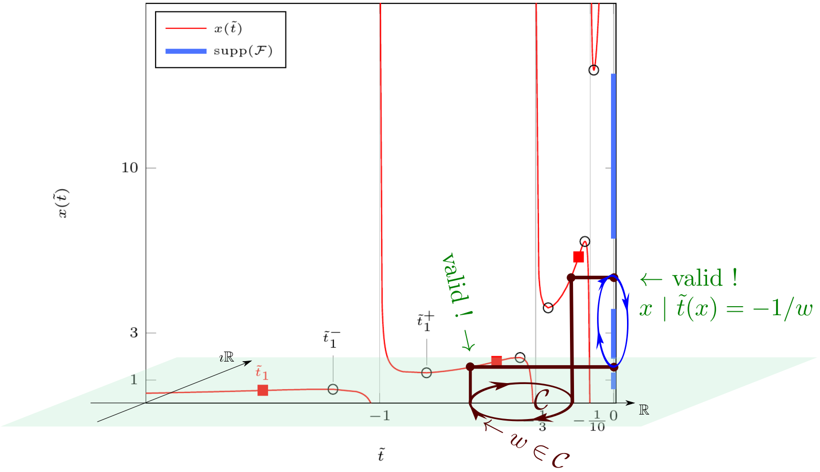
Some consequences: evaluating the eigenvalues of
$R_N = \mathrm{diag}(1, \ldots, 1, 3, \ldots, 3, 7, \ldots, 7)$?

- same idea but contour $\mathcal{C}_{\mathcal{F}}$ **only** around corresponding "hump" in spectrum of $\hat{R}_N$.

Statistical inference

Some consequences: evaluating the eigenvalues of $R_N = \mathrm{diag}(1, \dots, 1, 3, \dots, 3, 7, \dots, 7)$?

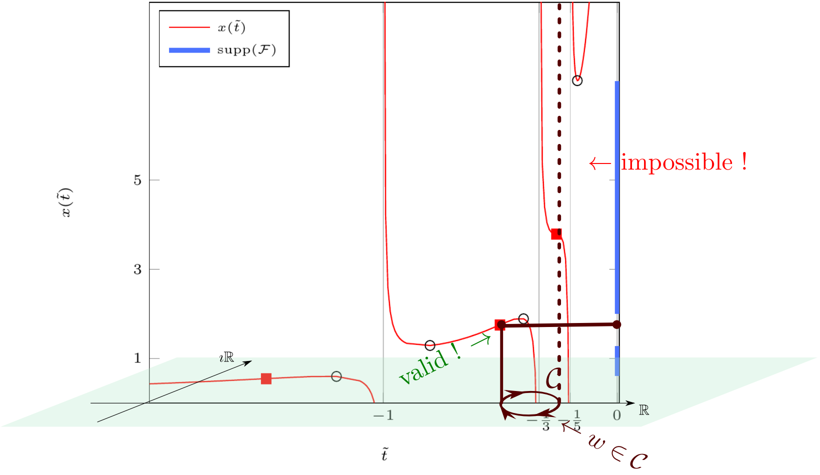
- same idea but contour $\mathcal{C}_{\mathcal{F}}$ **only** around corresponding "hump" in spectrum of $\hat{R}_N$.



Statistical inference

Some consequences: evaluating the eigenvalues of
$R_N = \mathrm{diag}(1, \ldots, 1, 3, \ldots, 3, 7, \ldots, 7)$?

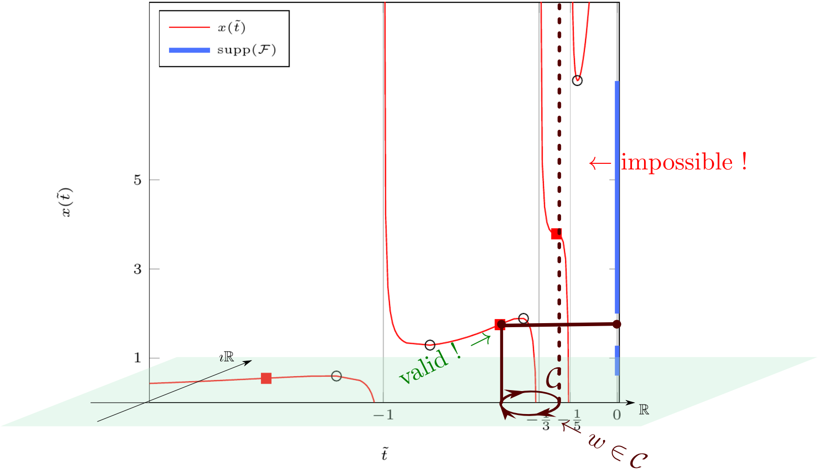
- same idea but contour $\mathcal{C}_\mathcal{F}$ **only** around corresponding "hump" in spectrum of $\hat{R}_N$.



Statistical inference

Some consequences: evaluating the eigenvalues of
$R_N = \mathrm{diag}(1, \ldots, 1, 3, \ldots, 3, 7, \ldots, 7)$?

- same idea but contour $\mathcal{C}_\mathcal{F}$ **only** around corresponding "hump" in spectrum of $\hat{R}_N$.



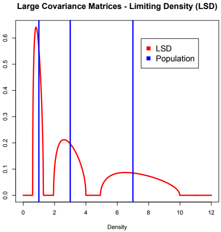
- only works if spectrum is "disjoint"!

Statistical inference

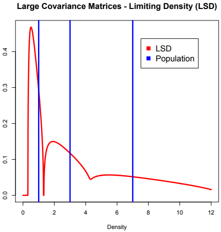
- This all depends on c! Remember...

Statistical inference

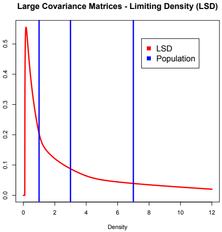
- This all depends on c! Remember...
  Limiting density for $c = .1$, $c = .3$, $c = .6$



can evaluate 1, 3, 7



can evaluate 1



cannot evaluate any.

- or can't we ??? (see lab session)

Outline

Outline

Position of the problem

Setup:

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\left\{ \begin{array}{c} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{array} \right.$$

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x, y) \in \mathbb{R}^+$$

Position of the problem

Setup:

- data $x_1,\ldots,x_n \in \mathbb{R}^p$ (for us, $n,p \gg 1$)
- k classes (for us, $k \ll n,p$):

$$\left\{ \begin{array}{c} x_1,\ldots,x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1},\ldots,x_n \in \mathcal{C}_k. \end{array} \right.$$

- affinity metric: for $x,y \in \mathbb{R}^p$,

$$\kappa(x,y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x,y) = f\left(\frac{1}{p}\|x-y\|^2\right)$$

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x,y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x,y) = f\left(\frac{1}{p}\|x-y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

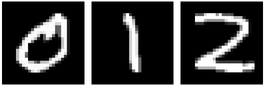
$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x, y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x, y) = f\left(\frac{1}{p}\|x - y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

Example: MNIST data



Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

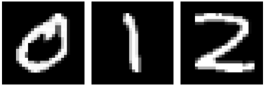
$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x,y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x,y) = f\left(\frac{1}{p}\|x-y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

Example: MNIST data



- $x_i =$ "pixels of images" ?

Position of the problem

Setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

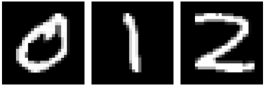
$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x, y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x, y) = f\left(\frac{1}{p}\|x - y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

Example: MNIST data



- x_i = "pixels of images" ?
- x_i = "smart features" ? (HOG, SURF, neural-net type [VGG, ResNet, etc.])

Position of the problem

Objective: perform clustering (i.e., unsupervised classification)!

Position of the problem

Objective: perform clustering (i.e., unsupervised classification)!

- metric to minimize:

$$\min_{\substack{\hat{\mathcal{C}}_1 \dots \hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i = \{1,\dots,n\} \\ \hat{\mathcal{C}}_i \cap \hat{\mathcal{C}}_j = \emptyset}} \sum_{i=1}^{k} \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \notin \hat{\mathcal{C}}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

Position of the problem

Objective: perform clustering (i.e., unsupervised classification)!

- metric to minimize:

$$\min_{\substack{\hat{\mathcal{C}}_1 \ldots \hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i = \{1,\ldots,n\} \\ \hat{\mathcal{C}}_i \cap \hat{\mathcal{C}}_j = \emptyset}} \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \notin \hat{\mathcal{C}}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

- can be rewritten

$$\min_{\substack{\hat{\mathcal{C}}_1 \ldots \hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i = \{1,\ldots,n\} \\ \hat{\mathcal{C}}_i \cap \hat{\mathcal{C}}_j = \emptyset}} \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \in \{1,\ldots,n\}}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|} - \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \in \hat{\mathcal{C}}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

Position of the problem

Objective: perform clustering (i.e., unsupervised classification)!

- metric to minimize:

$$\min_{\substack{\hat{\mathcal{C}}_1 \ldots \hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i = \{1,\ldots,n\} \\ \hat{\mathcal{C}}_i \cap \hat{\mathcal{C}}_j = \emptyset}} \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \notin \hat{\mathcal{C}}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

- can be rewritten

$$\min_{\substack{\hat{\mathcal{C}}_1 \ldots \hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i = \{1,\ldots,n\} \\ \hat{\mathcal{C}}_i \cap \hat{\mathcal{C}}_j = \emptyset}} \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \in \{1,\ldots,n\}}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|} - \sum_{i=1}^k \sum_{\substack{j \in \hat{\mathcal{C}}_i \\ \bar{j} \in \hat{\mathcal{C}}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

- or, denoting $F_{ji} = \frac{1}{\sqrt{|\hat{\mathcal{C}}_i|}} \delta_{x_j \in \hat{\mathcal{C}}_i}$ and $\mathcal{F} \subset \mathbb{R}^{n \times k}$ the **discrete** set of valid F's:

Position of the problem

Objective: perform clustering (i.e., unsupervised classification)!

- metric to minimize:

$$\min_{\substack{\hat{\mathcal{C}}_1\ldots\hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i=\{1,\ldots,n\} \\ \hat{\mathcal{C}}_i\cap\hat{\mathcal{C}}_j=\emptyset}} \sum_{i=1}^k \sum_{\substack{j\in\hat{\mathcal{C}}_i \\ \bar{j}\notin\hat{\mathcal{C}}_i}} \frac{\kappa(x_j,x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

- can be rewritten

$$\min_{\substack{\hat{\mathcal{C}}_1\ldots\hat{\mathcal{C}}_k \\ \cup_{i=1}^k \hat{\mathcal{C}}_i=\{1,\ldots,n\} \\ \hat{\mathcal{C}}_i\cap\hat{\mathcal{C}}_j=\emptyset}} \sum_{i=1}^k \sum_{\substack{j\in\hat{\mathcal{C}}_i \\ \bar{j}\in\{1,\ldots,n\}}} \frac{\kappa(x_j,x_{\bar{j}})}{|\hat{\mathcal{C}}_i|} - \sum_{i=1}^k \sum_{\substack{j\in\hat{\mathcal{C}}_i \\ \bar{j}\in\hat{\mathcal{C}}_i}} \frac{\kappa(x_j,x_{\bar{j}})}{|\hat{\mathcal{C}}_i|}$$

- or, denoting $F_{ji} = \frac{1}{\sqrt{|\hat{\mathcal{C}}_i|}}\delta_{x_j\in\hat{\mathcal{C}}_i}$ and $\mathcal{F}\subset\mathbb{R}^{n\times k}$ the **discrete** set of valid F's:

$$\min_{F\in\mathcal{F}} \sum_{i=1}^k F_{\cdot i}^{\mathsf{T}}(D-K)F_{\cdot i} = \min_{F\in\mathcal{F}} \operatorname{tr} F^{\mathsf{T}}(D-K)F$$

where

$K = \{\kappa(x_i,x_j)\}_{i,j=1}^n$ and $D = \operatorname{diag}(\{\sum_{j=1}^n K_{ij}\}_{i=1}^n)$ (degrees of the "graph" K).

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^{\mathsf{T}}F = I_k\}$

$$\min_{F,\ F^{\mathsf{T}}F=I_k} \mathrm{tr} F^{\mathsf{T}}(D-K)F$$

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \mathrm{tr} F^\mathsf{T} (D - K) F$$

- this is an eigenvector problem!

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n \times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \operatorname{tr} F^\mathsf{T} (D - K) F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k “smallest” eigenvectors of $D - K$.

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^{\mathsf{T}}F = I_k\}$

$$\min_{F,\ F^{\mathsf{T}}F=I_k} \operatorname{tr} F^{\mathsf{T}}(D-K)F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k "smallest" eigenvectors of $D-K$.
- gives spectral clustering algorithm:
  1. retrieve k smallest eigenvectors of $D-K$ into $V=[v_1,\ldots,v_k]\in\mathbb{R}^{n\times k}$
  2. cluster the (small dimensional!) rows $V_{i\cdot}\in\mathbb{R}^n$ with k-means
  3. the k classes obtained are the estimates of $\mathcal{C}_1,\ldots,\mathcal{C}_k$.

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \operatorname{tr} F^\mathsf{T}(D-K)F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k "smallest" eigenvectors of $D-K$.
- gives spectral clustering algorithm:
  1. retrieve k smallest eigenvectors of $D-K$ into $V=[v_1,\ldots,v_k]\in\mathbb{R}^{n\times k}$
  2. cluster the (small dimensional!) rows $V_{i\cdot}\in\mathbb{R}^n$ with k-means
  3. the k classes obtained are the estimates of $\mathcal{C}_1,\ldots,\mathcal{C}_k$.

Important remark: geometric interpretation

$$\operatorname{tr} F^\mathsf{T}(D-K)F = \frac{1}{2}\sum_{i=1}^{k}\sum_{j\bar{j}} K_{j\bar{j}}(F_{ji}-F_{\bar{j}i})^2 \quad \text{(prove it!)}$$

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \operatorname{tr} F^\mathsf{T}(D-K)F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k "smallest" eigenvectors of $D-K$.
- gives spectral clustering algorithm:
  1. retrieve k smallest eigenvectors of $D-K$ into $V=[v_1,\ldots,v_k]\in\mathbb{R}^{n\times k}$
  2. cluster the (small dimensional!) rows $V_{i\cdot}\in\mathbb{R}^n$ with k-means
  3. the k classes obtained are the estimates of $\mathcal{C}_1,\ldots,\mathcal{C}_k$.

Important remark: geometric interpretation

$$\operatorname{tr} F^\mathsf{T}(D-K)F = \frac{1}{2}\sum_{i=1}^{k}\sum_{j\bar{j}} K_{j\bar{j}}(F_{ji}-F_{\bar{j}i})^2 \quad \text{(prove it!)}$$

- we must have $F_{ji}\simeq F_{\bar{j}i}$ if $K_{j\bar{j}}\gg 1$

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \operatorname{tr} F^\mathsf{T} (D-K) F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k "smallest" eigenvectors of $D-K$.
- gives spectral clustering algorithm:
  1. retrieve k smallest eigenvectors of $D-K$ into $V = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$
  2. cluster the (small dimensional!) rows $V_{i\cdot} \in \mathbb{R}^n$ with k-means
  3. the k classes obtained are the estimates of $\mathcal{C}_1, \ldots, \mathcal{C}_k$.

Important remark: geometric interpretation

$$\operatorname{tr} F^\mathsf{T} (D-K) F = \frac{1}{2} \sum_{i=1}^{k} \sum_{j\bar{j}} K_{j\bar{j}} (F_{ji} - F_{\bar{j}i})^2 \quad \text{(prove it!)}$$

- we must have $F_{ji} \simeq F_{\bar{j}i}$ if $K_{j\bar{j}} \gg 1$
- F_{ji} could be distinct from $F_{\bar{j}i}$ if $K_{j\bar{j}} \simeq 0$

Position of the problem

Relaxation into spectral clustering:

- problem is discrete, NP-complete.
- solution: relaxing $\mathcal{F}$ into $\{F \in \mathbb{R}^{n\times k},\ F^\mathsf{T} F = I_k\}$

$$\min_{F,\ F^\mathsf{T} F = I_k} \operatorname{tr} F^\mathsf{T}(D-K)F$$

- this is an eigenvector problem!
  $\Rightarrow$ Solution F is concatenation of k "smallest" eigenvectors of $D-K$.
- gives spectral clustering algorithm:
  1. retrieve k smallest eigenvectors of $D-K$ into $V=[v_1,\ldots,v_k]\in\mathbb{R}^{n\times k}$
  2. cluster the (small dimensional!) rows $V_{i\cdot}\in\mathbb{R}^n$ with k-means
  3. the k classes obtained are the estimates of $\mathcal{C}_1,\ldots,\mathcal{C}_k$.

Important remark: geometric interpretation

$$\operatorname{tr} F^\mathsf{T}(D-K)F = \frac{1}{2}\sum_{i=1}^{k}\sum_{j\bar{j}} K_{j\bar{j}}(F_{ji}-F_{\bar{j}i})^2 \quad \text{(prove it!)}$$

- we must have $F_{ji}\simeq F_{\bar{j}i}$ if $K_{j\bar{j}}\gg 1$
- F_{ji} could be distinct from $F_{\bar{j}i}$ if $K_{j\bar{j}}\simeq 0$

We will see this will be a problem in large dimensions!

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

- hence

$$K \simeq \begin{pmatrix} [K_1]_{n_1 \times n_1} & [\varepsilon]_{n_1 \times n_2} & \cdots \\ [\varepsilon]_{n_2 \times n_1} & [K_2]_{n_2 \times n_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

- hence

$$K \simeq \begin{pmatrix} [K_1]_{n_1 \times n_1} & [\varepsilon]_{n_1 \times n_2} & \cdots \\ [\varepsilon]_{n_2 \times n_1} & [K_2]_{n_2 \times n_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- in particular, with $j_a = (0_{n_1}, \ldots, 1_{n_a}, \ldots, 0_{n_k})^\mathsf{T}$,

$$K j_a \simeq D j_a$$

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

- hence

$$K \simeq \begin{pmatrix} [K_1]_{n_1 \times n_1} & [\varepsilon]_{n_1 \times n_2} & \cdots \\ [\varepsilon]_{n_2 \times n_1} & [K_2]_{n_2 \times n_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- in particular, with $j_a = (0_{n_1}, \ldots, 1_{n_a}, \ldots, 0_{n_k})^\mathsf{T}$,

$$K j_a \simeq D j_a \Rightarrow (D - K) j_a \simeq 0$$

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

- hence

$$K \simeq \begin{pmatrix} [K_1]_{n_1 \times n_1} & [\varepsilon]_{n_1 \times n_2} & \cdots \\ [\varepsilon]_{n_2 \times n_1} & [K_2]_{n_2 \times n_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- in particular, with $j_a = (0_{n_1}, \ldots, 1_{n_a}, \ldots, 0_{n_k})^\mathsf{T}$,

$$K j_a \simeq D j_a \Rightarrow (D - K) j_a \simeq 0$$

so that j_a, canonical vector of $\mathcal{C}_a$ eigenvector of $D - K$!

Position of the problem

The Ng-Weiss-Jordan (wrong) intuition:

- say $f(t) = \exp(-t/2)$: the Gaussian (or heat, or RBF) kernel.
- *intuition*: for easy clustering tasks,

$$\|x_i - x_j\| \begin{cases} \gg 1, & \text{if } x_i, x_j \text{ in different classes} \\ \ll 1, & \text{if } x_i, x_j \text{ in same class} \end{cases}$$

- hence

$$K \simeq \begin{pmatrix} [K_1]_{n_1 \times n_1} & [\varepsilon]_{n_1 \times n_2} & \cdots \\ [\varepsilon]_{n_2 \times n_1} & [K_2]_{n_2 \times n_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- in particular, with $j_a = (0_{n_1}, \ldots, 1_{n_a}, \ldots, 0_{n_k})^{\mathsf{T}}$,

$$K j_a \simeq D j_a \Rightarrow (D - K) j_a \simeq 0$$

  so that j_a, canonical vector of $\mathcal{C}_a$ eigenvector of $D - K$!

Refinement: implies also

$$D^{-\frac{1}{2}} K D^{-\frac{1}{2}} (D^{\frac{1}{2}} j_a) = D^{\frac{1}{2}} j_a$$

more stable in practice.

Position of the problem

From theory to practice: **not at all what was expected!!**

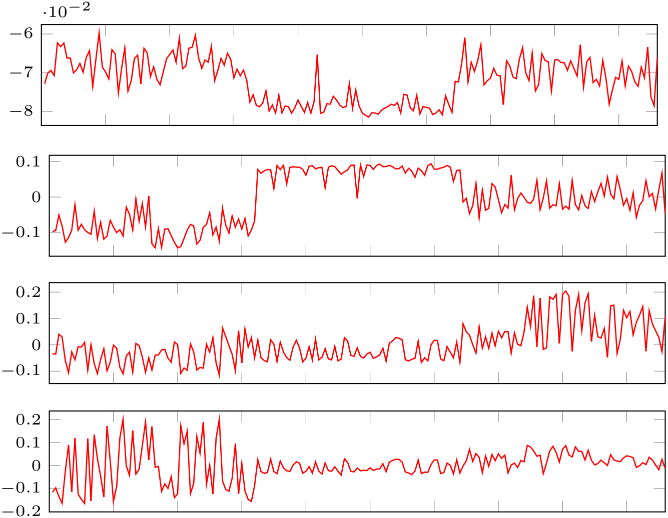


Figure: 4 dominant eigenvectors of $L = D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ (**red**); MNIST data (0, 1, 2).

Position of the problem

From theory to practice: not at all what was expected!!

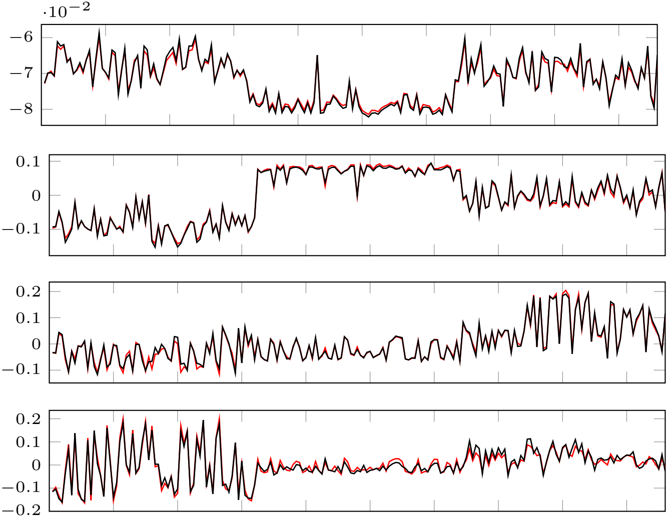


Figure: 4 dominant eigenvectors of $L = D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ (**red**), asymptotic approximation $\hat{L}$ (**black**); MNIST data (0, 1, 2).

Position of the problem

From theory to practice: **not at all what was expected!!**

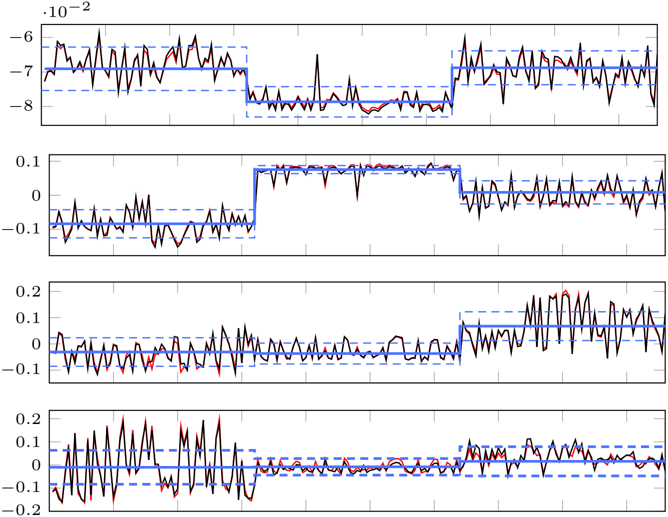


Figure: 4 dominant eigenvectors of $L = D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ (**red**), asymptotic approximation $\hat{L}$ (**black**) vs. Gaussian theory (1σ et 2σ) (**blue**); MNIST data (0, 1, 2).

Position of the problem

What's wrong with Ng-Weiss-Jordan?

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution?

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$.

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2)$$

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right) \right)$$

or equivalently

$$\mathbb{P}\left(x^\mathsf{T} \mu > x^\mathsf{T}(-\mu) \mid x \in \mathcal{C}_2 \right)$$

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(x^\mathsf{T}\mu > x^\mathsf{T}(-\mu) \mid x \in \mathcal{C}_2\right) \Leftrightarrow \mathbb{P}\left(x^\mathsf{T}\mu > 0 \mid x \in \mathcal{C}_2\right).$$

Position of the problem

What's wrong with Ng-Weiss-Jordan? their small dimensional intuition!

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(x^\mathsf{T}\mu > x^\mathsf{T}(-\mu) \mid x \in \mathcal{C}_2\right) \Leftrightarrow \mathbb{P}\left(x^\mathsf{T}\mu > 0 \mid x \in \mathcal{C}_2\right).$$

But $x \in \mathcal{C}_2 \Leftrightarrow x \sim \mathcal{N}(-\mu, I_p)$, so $x^\mathsf{T}\mu \sim \mathcal{N}(-\|\mu\|^2, \|\mu\|^2)$

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(x^{\mathsf{T}}\mu > x^{\mathsf{T}}(-\mu) \mid x \in \mathcal{C}_2\right) \Leftrightarrow \mathbb{P}\left(x^{\mathsf{T}}\mu > 0 \mid x \in \mathcal{C}_2\right).$$

But $x \in \mathcal{C}_2 \Leftrightarrow x \sim \mathcal{N}(-\mu, I_p)$, so $x^{\mathsf{T}}\mu \sim \mathcal{N}(-\|\mu\|^2, \|\mu\|^2)$, i.e.,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(-(t + \|\mu\|^2)^2/(2\|\mu\|^2))dt$$

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(x^\mathsf{T}\mu > x^\mathsf{T}(-\mu) \mid x \in \mathcal{C}_2\right) \Leftrightarrow \mathbb{P}\left(x^\mathsf{T}\mu > 0 \mid x \in \mathcal{C}_2\right).$$

But $x \in \mathcal{C}_2 \Leftrightarrow x \sim \mathcal{N}(-\mu, I_p)$, so $x^\mathsf{T}\mu \sim \mathcal{N}(-\|\mu\|^2, \|\mu\|^2)$, i.e.,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(-(t + \|\mu\|^2)^2/(2\|\mu\|^2))dt = \|\mu\| Q(\|\mu\|)$$

with $Q(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)du$.

Position of the problem

What's wrong with Ng-Weiss-Jordan? **their small dimensional intuition!**

A detour to Bayes' optimality: assume $\mathbb{R}^p \ni x_i \in \mathcal{C}_a \Leftrightarrow x_i \sim \mathcal{N}(\mu_a, C_a)$ (Gaussian mixture model)

- as $n, p \to \infty$, we need to control growth rate of $\mu_1, \ldots, \mu_k$ and $C_1, \ldots, C_k$ to have a non-trivial/degenerate problem!
- how to control μ_a, C_a evolution? $\Rightarrow$ find worst case scenario when μ_i's, C_i's perfectly known!

Neyman-Pearson test: assume (temporarily) $k = 2$, μ_1, μ_2, C_1, C_2 known

- μ_i *evolution*: say $\mu_1 = -\mu_2 = \mu$ and $C_1 = C_2 = I_p$. Then by Neyman-Pearson,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \mathbb{P}\left((2\pi)^{-p/2} \exp\left(-\|x - \mu_1\|^2/2\right) > (2\pi)^{-p/2} \exp\left(-\|x - \mu_2\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(x^{\mathsf{T}}\mu > x^{\mathsf{T}}(-\mu) \mid x \in \mathcal{C}_2\right) \Leftrightarrow \mathbb{P}\left(x^{\mathsf{T}}\mu > 0 \mid x \in \mathcal{C}_2\right).$$

But $x \in \mathcal{C}_2 \Leftrightarrow x \sim \mathcal{N}(-\mu, I_p)$, so $x^{\mathsf{T}}\mu \sim \mathcal{N}(-\|\mu\|^2, \|\mu\|^2)$, i.e.,

$$\mathbb{P}(x \to \mathcal{C}_1 \mid x \in \mathcal{C}_2) = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(-(t + \|\mu\|^2)^2/(2\|\mu\|^2))dt = \|\mu\| Q(\|\mu\|)$$

with $Q(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)du$.

Only non-trivial if $\|\mu_1 - \mu_2\| = 2\|\mu\| = O(1)$! (with respect to p)

Position of the problem

Neyman-Pearson test:

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr} C_1 - \mathrm{tr} C_2 = \varepsilon p$)

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr}C_1 - \mathrm{tr}C_2 = \varepsilon p$)

$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1)$

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr}C_1 - \mathrm{tr}C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr}C_1 - \mathrm{tr}C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr} C_1 - \mathrm{tr} C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$)

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr}C_1 - \mathrm{tr}C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$) so, for $p \gg 1$,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}_{\mathcal{N}(0,1)}\left(w > \sqrt{p/3}(\log(1+\varepsilon)(1+\varepsilon)/\varepsilon - 1)\right)$$

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr} C_1 - \mathrm{tr} C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$) so, for $p \gg 1$,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}_{\mathcal{N}(0,1)}\left(w > \sqrt{p/3}(\log(1+\varepsilon)(1+\varepsilon)/\varepsilon - 1)\right)$$

Only non-trivial if $\varepsilon \to 0$!

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\operatorname{tr}C_1 - \operatorname{tr}C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2}\exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$) so, for $p \gg 1$,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}_{\mathcal{N}(0,1)}\left(w > \sqrt{p/3}(\log(1+\varepsilon)(1+\varepsilon)/\varepsilon - 1)\right)$$

Only non-trivial if $\varepsilon \to 0$!, and then, by Taylor,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}(w > \sqrt{p/3}\varepsilon)$$

Position of the problem

Neyman-Pearson test:

- C_i *evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr} C_1 - \mathrm{tr} C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2}\exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$) so, for $p \gg 1$,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}_{\mathcal{N}(0,1)}\left(w > \sqrt{p/3}(\log(1+\varepsilon)(1+\varepsilon)/\varepsilon - 1)\right)$$

Only non-trivial if $\varepsilon \to 0!$, and then, by Taylor,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}(w > \sqrt{p/3}\varepsilon)$$

Only non-trivial if $\varepsilon = O(p^{-\frac{1}{2}})$!

Position of the problem

Neyman-Pearson test:

- *C_i evolution*: assume now $\mu_1 = \mu_2 = 0$ and $C_1 = I_p$, $C_2 = (1+\varepsilon)I_p$. (so that $\mathrm{tr} C_1 - \mathrm{tr} C_2 = \varepsilon p$)

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) = \mathbb{P}\left((1+\varepsilon)^{-p/2} \exp\left(-\|x\|^2/2(1+\varepsilon)\right) > \exp\left(-\|x\|^2/2\right)\right)$$

or equivalently

$$\mathbb{P}\left(\|x\|^2/p > \log(1+\varepsilon)(1+\varepsilon)/\varepsilon \mid x \in \mathcal{C}_1\right)$$

By the CLT, $\|x\|^2/p \to \mathcal{N}(1, 3/p)$ (we used $\mathbb{E}[|x_j|^4] = 3$ for $x \in \mathbb{R}^p$) so, for $p \gg 1$,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}_{\mathcal{N}(0,1)}\left(w > \sqrt{p/3}(\log(1+\varepsilon)(1+\varepsilon)/\varepsilon - 1)\right)$$

Only non-trivial if $\varepsilon \to 0$!, and then, by Taylor,

$$\mathbb{P}(x \to \mathcal{C}_2 \mid x \in \mathcal{C}_1) \simeq \mathbb{P}(w > \sqrt{p/3}\varepsilon)$$

Only non-trivial if $\varepsilon = O(p^{-\frac{1}{2}})$!

- *Conclusion*: non-trivial conditions:

$$\|\mu_1 - \mu_2\| = O(1), \quad \mathrm{tr}(C_1 - C_2) = O(\sqrt{p})$$

Position of the problem

Neyman-Pearson test: the question of $\mathrm{tr}((C_a - C_b)^2)$

Neyman-Pearson test: the question of $\mathrm{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\mathrm{tr}((C_a - C_b)^2) = O(1)$$

Position of the problem

Neyman-Pearson test: the question of $\mathrm{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\mathrm{tr}((C_a - C_b)^2) = O(1)$$

- for spectral clustering (so with C_i's **un**known), we will see:

Position of the problem

Neyman-Pearson test: the question of $\mathrm{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\mathrm{tr}((C_a - C_b)^2) = O(1)$$

- for spectral clustering (so with C_i's **un**known), we will see:

$\mathrm{tr}((C_a - C_b)^2) = O(p)$ for standard methods! (e.g., Gaussian kernel)

Position of the problem

Neyman-Pearson test: the question of $\operatorname{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\operatorname{tr}((C_a - C_b)^2) = O(1)$$

- for spectral clustering (so with C_i's **un**known), we will see:

$$\operatorname{tr}((C_a - C_b)^2) = O(p) \quad \text{for standard methods! (e.g., Gaussian kernel)}$$

- **but** can be dramatically improved with **random matrix-tuned kernel**

Position of the problem

Neyman-Pearson test: the question of $\mathrm{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\mathrm{tr}((C_a - C_b)^2) = O(1)$$

- for spectral clustering (so with C_i's **un**known), we will see:

$$\mathrm{tr}((C_a - C_b)^2) = O(p) \quad \text{for standard methods! (e.g., Gaussian kernel)}$$

- **but** can be dramatically improved with **random matrix-tuned kernel** into

$$\mathrm{tr}((C_a - C_b)^2) = O(\sqrt{p})$$

Position of the problem

Neyman-Pearson test: the question of $\operatorname{tr}((C_a - C_b)^2)$

- from Neyman-Pearson (so with C_i's known), non-triviality when

$$\operatorname{tr}((C_a - C_b)^2) = O(1)$$

- for spectral clustering (so with C_i's **un**known), we will see:

$$\operatorname{tr}((C_a - C_b)^2) = O(p) \quad \text{for standard methods! (e.g., Gaussian kernel)}$$

- **but** can be dramatically improved with **random matrix-tuned kernel** into

$$\operatorname{tr}((C_a - C_b)^2) = O(\sqrt{p})$$

Can we do better???

Position of the problem

The (non-trivial) setup:

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x, y) \in \mathbb{R}^+$$

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x,y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x,y) = f\left(\frac{1}{p}\|x-y\|^2\right)$$

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x,y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x,y) = f\left(\frac{1}{p}\|x-y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

Position of the problem

The (non-trivial) setup:

- data $x_1, \ldots, x_n \in \mathbb{R}^p$ (for us, $n, p \gg 1$)
- k classes (for us, $k \ll n, p$):

$$\begin{cases} x_1, \ldots, x_{n_1} \in \mathcal{C}_1 \\ \vdots \\ x_{n-n_k+1}, \ldots, x_n \in \mathcal{C}_k. \end{cases}$$

- affinity metric: for $x, y \in \mathbb{R}^p$,

$$\kappa(x, y) \in \mathbb{R}^+ \quad \text{e.g., } \kappa(x, y) = f\left(\frac{1}{p}\|x - y\|^2\right)$$

(we will see later why $\frac{1}{p}$ useful)

- **asymptotic non-triviality conditions:** for $\mu^\circ = \sum_{a=1}^k \frac{n_a}{n}\mu_a$ and $C^\circ = \sum_{a=1}^k \frac{n_a}{n}C_a$

$$\|\mu_a - \mu^\circ\| = O(1) \quad \operatorname{tr}(C_a - C_a) = O(\sqrt{p}) \quad \|C_a\| = O(1)$$

Position of the problem

Fundamental result:

Position of the problem

Fundamental result: (really, really fundamental!!!)

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n} \left| \frac{1}{p}\|x_i - x_j\|^2 - \tau_p \right| \xrightarrow{\text{a.s.}} 0$$

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n}\left|\frac{1}{p}\|x_i-x_j\|^2-\tau_p\right| \xrightarrow{\text{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i-x_j\|^2=\tau_p+O_p(p^{-\frac{1}{2}})$$

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n} \left| \frac{1}{p}\|x_i - x_j\|^2 - \tau_p \right| \xrightarrow{\mathrm{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i - x_j\|^2 = \tau_p + O_p(p^{-\frac{1}{2}})$$

and in particular

$$\frac{1}{p}x_i^\mathsf{T} x_j = 0 + O_p(p^{-\frac{1}{2}})$$

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n}\left|\frac{1}{p}\|x_i - x_j\|^2 - \tau_p\right| \xrightarrow{\text{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i - x_j\|^2 = \tau_p + O_p(p^{-\frac{1}{2}})$$

and in particular

$$\frac{1}{p}x_i^\mathsf{T} x_j = 0 + O_p(p^{-\frac{1}{2}})$$

- does this mean $K \simeq f(\tau_p)1_n 1_n^\mathsf{T}$?, clustering impossible ?

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n} \left| \frac{1}{p}\|x_i - x_j\|^2 - \tau_p \right| \xrightarrow{\text{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i - x_j\|^2 = \tau_p + O_p(p^{-\frac{1}{2}})$$

and in particular

$$\frac{1}{p}x_i^\mathsf{T} x_j = 0 + O_p(p^{-\frac{1}{2}})$$

- does this mean $K \simeq f(\tau_p)1_n 1_n^\mathsf{T}$?, clustering impossible ?
- **yes and no**!!

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n} \left| \frac{1}{p}\|x_i - x_j\|^2 - \tau_p \right| \xrightarrow{\text{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i - x_j\|^2 = \tau_p + O_p(p^{-\frac{1}{2}})$$

and in particular

$$\frac{1}{p}x_i^\mathsf{T} x_j = 0 + O_p(p^{-\frac{1}{2}})$$

- does this mean $K \simeq f(\tau_p)1_n 1_n^\mathsf{T}$?, clustering impossible ?
- **yes and no**!!
- **Careful**: same error as Marčenko-Pastur theorem:

$$\max_{i,j} \left| [\frac{1}{p}XX^\mathsf{T}]_{ij} - [I_p]_{ij} \right| \xrightarrow{\text{a.s.}} 0 \quad \textbf{but} \quad \frac{1}{p}XX^\mathsf{T} \not\to I_p!!$$

Position of the problem

Fundamental result: (really, really fundamental!!!)
Under the above conditions, for $\tau_p \equiv (2/p)\mathrm{tr}C^\circ$,

$$\max_{1\leq i,j\leq n} \left| \frac{1}{p}\|x_i - x_j\|^2 - \tau_p \right| \xrightarrow{\text{a.s.}} 0$$

In fact, precisely,

$$\frac{1}{p}\|x_i - x_j\|^2 = \tau_p + O_p(p^{-\frac{1}{2}})$$

and in particular

$$\frac{1}{p}x_i^\mathsf{T} x_j = 0 + O_p(p^{-\frac{1}{2}})$$

- does this mean $K \simeq f(\tau_p)1_n 1_n^\mathsf{T}$?, clustering impossible ?
- **yes and no**!!
- **Careful**: same error as Marčenko-Pastur theorem:

$$\max_{i,j} \left| [\frac{1}{p}XX^\mathsf{T}]_{ij} - [I_p]_{ij} \right| \xrightarrow{\text{a.s.}} 0 \quad \textbf{but} \quad \frac{1}{p}XX^\mathsf{T} \not\to I_p!!$$

Key idea: Taylor expansion of K_{ij} around $f(\tau_p)$!

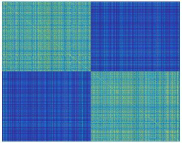
Position of the problem

In image: Kernel $K_{ij} = \exp(-\frac{1}{2p}\|x_i - x_j\|^2)$ and second eigenvector v_2 ($x_i \sim \mathcal{N}(\pm\mu, I_p)$, $\mu = (2, 0, \ldots, 0)^\mathsf{T} \in \mathbb{R}^p$).

Position of the problem

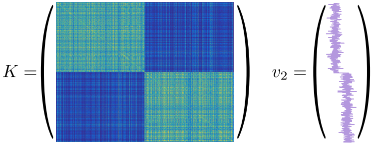
In image: Kernel $K_{ij} = \exp(-\frac{1}{2p}\|x_i - x_j\|^2)$ and second eigenvector v_2 ($x_i \sim \mathcal{N}(\pm\mu, I_p)$, $\mu = (2, 0, \ldots, 0)^\mathsf{T} \in \mathbb{R}^p$).

$p = 4$, $n = 1000$

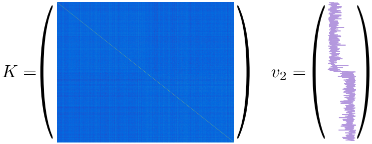
$K =$  $v_2 =$ 

In image: Kernel $K_{ij} = \exp(-\frac{1}{2p}\|x_i - x_j\|^2)$ and second eigenvector v_2 $(x_i \sim \mathcal{N}(\pm\mu, I_p),\ \mu = (2, 0, \ldots, 0)^\mathsf{T} \in \mathbb{R}^p)$.

$p = 4,\ n = 1000$



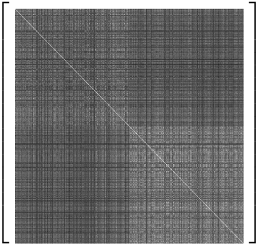
$p = \mathbf{400},\ n = 1000$

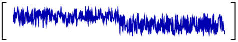


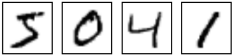
Position of the problem

And for real data?

MNIST
raw
$p = 784,\ n = 500$
$\downarrow$

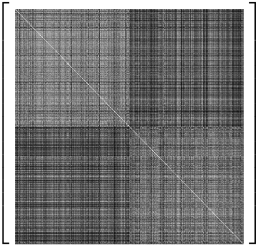
$\mathbf{K} =$ 

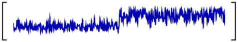
$\mathbf{v}_2 =$ 



(ici, classes "5" et "0")

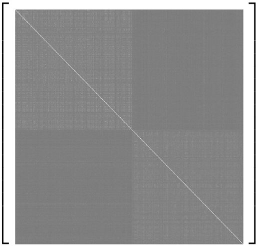
ImageNet
VGG-features
$p = 3084,\ n = 500$
$\downarrow$

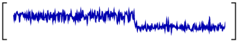
$\mathbf{K} =$ 

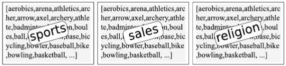
$\mathbf{v}_2 =$ 

(ici, classes "bird" et "plane")

20NewsGroup
BERT embedding
$p = 300,\ n = 500$
$\downarrow$

$\mathbf{K} =$ 

$\mathbf{v}_2 =$ 



(ici, classes "sports" et "sales")

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \quad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^{n} \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0) \frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0) \frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0) \frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p\right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion!

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0)\frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0)\left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0)\frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p \right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion!
  See here why:

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0) \frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0) \frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p \right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion!
  See here why: what are the spectral norms of?

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} \sim \mathcal{N}(0,1) = O(1)?$$

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \qquad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0) \frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0) \frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p \right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion!
  See here why: what are the spectral norms of?

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} \sim \mathcal{N}(0,1) = O(1)? \quad \Rightarrow O(1)! \quad \text{(semi-circle law)}$$

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \quad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0)\frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0) \frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p \right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion! See here why: what are the spectral norms of?

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} \sim \mathcal{N}(0,1) = O(1)? \quad \Rightarrow O(1)! \quad \text{(semi-circle law)}$$

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} = 1?$$

Taylor expansion

The Taylor expansion: For simplicity, consider the simpler case (with $\|\mu_i\| = O(1)$)

$$K = \left\{ \frac{1}{p} x_i^\mathsf{T} x_j \right\}_{i,j=1}^n \quad \text{(so we forget the terms } \|x_i\|^2/p \text{ and } \|x_j\|^2/p\text{)}$$

Taylor expansion now around $f(0)$:

- just develop (for, say, $x_i \sim \mathcal{N}(\mu_a, C_a)$ and $x_j \sim \mathcal{N}(\mu_b, C_b)$):

$$f\left(\frac{1}{p} x_i^\mathsf{T} x_j\right) = f(0) + f'(0)\frac{1}{p} x_i^\mathsf{T} x_j + \frac{1}{2} f''(0) \left(\frac{1}{p} x_i^\mathsf{T} x_j\right)^2 + \ldots$$

  Hence:

$$K = f(0) 1_n 1_n^\mathsf{T} + f'(0) \frac{1}{p} X^\mathsf{T} X + \frac{1}{2} f''(0) \left\{ \left(x_i^\mathsf{T} x_j / p \right)^2 \right\}_{i,j=1}^n + \ldots$$

- **key difficulty**: moving from entries Taylor expansion to matrix Taylor expansion!
  See here why: what are the spectral norms of?

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} \sim \mathcal{N}(0,1) = O(1)? \quad \Rightarrow O(1)! \quad \text{(semi-circle law)}$$

$$\frac{1}{\sqrt{n}} X \in \mathbb{R}^{n \times n} \quad X_{ij} = X_{ji} = 1? \quad \Rightarrow O(\sqrt{p})! \quad \text{(rank-1 matrix)}$$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^\mathsf{T}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^{\mathsf{T}}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$\|\frac{1}{p}W^{\mathsf{T}}W\| = O(1)$ (SCM model!)

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^\mathsf{T}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$\|\frac{1}{p}W^\mathsf{T}W\| = O(1)$ (SCM model!) and $\|\frac{1}{p}(MJ^\mathsf{T})^\mathsf{T}(MJ^\mathsf{T})\| = \|\frac{1}{p}J\underbrace{M^\mathsf{T}M}_{=O(1)}J^\mathsf{T}\| = O(1)$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^{\mathsf{T}}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$$\|\frac{1}{p}W^{\mathsf{T}}W\| = O(1) \quad \text{(SCM model!)} \quad \text{and} \quad \|\frac{1}{p}(MJ^{\mathsf{T}})^{\mathsf{T}}(MJ^{\mathsf{T}})\| = \|\frac{1}{p}J \underbrace{M^{\mathsf{T}}M}_{=O(1)} J^{\mathsf{T}}\| = O(1)$$

so finally

$$\frac{1}{p}X^{\mathsf{T}}X = O_{\|\cdot\|}(1)$$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^\mathsf{T}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$$\|\frac{1}{p}W^\mathsf{T}W\| = O(1) \quad \text{(SCM model!)} \quad \text{and} \quad \|\frac{1}{p}(MJ^\mathsf{T})^\mathsf{T}(MJ^\mathsf{T})\| = \|\frac{1}{p}J\underbrace{M^\mathsf{T}M}_{=O(1)}J^\mathsf{T}\| = O(1)$$

so finally

$$\frac{1}{p}X^\mathsf{T}X = O_{\|\cdot\|}(1)$$

- in addition, writing $(x_i^\mathsf{T}x_j)^2 = \mathbb{E}[(x_i^\mathsf{T}x_j)^2] + Z_{ij}$, and with $\operatorname{tr}(C_aC_b) = O(p)$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^\mathsf{T}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$$\|\frac{1}{p}W^\mathsf{T}W\| = O(1) \quad \text{(SCM model!)} \quad \text{and} \quad \|\frac{1}{p}(MJ^\mathsf{T})^\mathsf{T}(MJ^\mathsf{T})\| = \|\frac{1}{p}J\underbrace{M^\mathsf{T}M}_{=O(1)}J^\mathsf{T}\| = O(1)$$

so finally

$$\frac{1}{p}X^\mathsf{T}X = O_{\|\cdot\|}(1)$$

- in addition, writing $(x_i^\mathsf{T}x_j)^2 = \mathbb{E}[(x_i^\mathsf{T}x_j)^2] + Z_{ij}$, and with $\operatorname{tr}(C_aC_b) = O(p)$

$$\frac{1}{p^2}\left\{(x_i^\mathsf{T}x_j)^2\right\}_{i,j=1}^p = \frac{1}{p^2}\left\{\begin{bmatrix} \operatorname{tr}C_1^2 1_{n_1}1_{n_1}^\mathsf{T} & \operatorname{tr}C_1C_2 1_{n_1}1_{n_2}^\mathsf{T} & \cdots \\ \operatorname{tr}C_1C_2 1_{n_2}1_{n_1}^\mathsf{T} & \operatorname{tr}C_2^2 1_{n_2}1_{n_2}^\mathsf{T} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\right\}_{i,j=1}^p + \underbrace{Z}_{\substack{\mathbb{E}Z_{ij}=0 \\ \operatorname{Var}Z_{ij}=O(p^{-2})}} + \ldots$$

Taylor expansion

Continuing the expansion: with $\mathcal{C}_a \ni x_i = \mu_a + w_i$,

- while $\|\mu_a\| = O(1)$ and $\|w_i\| = O(\sqrt{p})$...,

$$X = W + MJ^\mathsf{T}, \quad M = [\mu_1, \ldots, \mu_k], \quad J = [j_1, \ldots, j_k]$$

in which

$$\|\frac{1}{p} W^\mathsf{T} W\| = O(1) \quad \text{(SCM model!)} \quad \text{and} \quad \|\frac{1}{p}(MJ^\mathsf{T})^\mathsf{T}(MJ^\mathsf{T})\| = \|\frac{1}{p} J \underbrace{M^\mathsf{T} M}_{=O(1)} J^\mathsf{T}\| = O(1)$$

so finally

$$\frac{1}{p} X^\mathsf{T} X = O_{\|\cdot\|}(1)$$

- in addition, writing $(x_i^\mathsf{T} x_j)^2 = \mathbb{E}[(x_i^\mathsf{T} x_j)^2] + Z_{ij}$, and with $\mathrm{tr}(C_a C_b) = O(p)$

$$\frac{1}{p^2}\left\{(x_i^\mathsf{T} x_j)^2\right\}_{i,j=1}^p = \frac{1}{p^2}\left\{\begin{bmatrix} \mathrm{tr} C_1^2 1_{n_1} 1_{n_1}^\mathsf{T} & \mathrm{tr} C_1 C_2 1_{n_1} 1_{n_2}^\mathsf{T} & \cdots \\ \mathrm{tr} C_1 C_2 1_{n_2} 1_{n_1}^\mathsf{T} & \mathrm{tr} C_2^2 1_{n_2} 1_{n_2}^\mathsf{T} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\right\}_{i,j=1}^p + \underbrace{Z}_{\substack{\mathbb{E} Z_{ij} = 0 \\ \mathrm{Var} Z_{ij} = O(p^{-2})}} + \ldots$$

$$= O_{\|\cdot\|}(1) + O_{\|\cdot\|}(p^{-\frac{1}{2}})$$

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^{\mathsf{T}}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^{\mathsf{T}})^{\mathsf{T}}(W + MJ^{\mathsf{T}})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^{\mathsf{T}}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ...

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**
- as a consequence, we can anticipate:

Main result

Summing up:

$$K = f(0)\underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0)\underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2}\underbrace{\frac{1}{p}JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**
- as a consequence, we can anticipate:
  1. **phase transition phenomena**: below some threshold $g(M,T)$, **clustering impossible!**

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**
- as a consequence, we can anticipate:
  1. **phase transition phenomena**: below some threshold $g(M,T)$, **clustering impossible!**
  2. beyond the threshold, **eigenvector alignment known**

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**
- as a consequence, we can anticipate:
  1. **phase transition phenomena**: below some threshold $g(M,T)$, **clustering impossible!**
  2. beyond the threshold, **eigenvector alignment known**

$\Rightarrow$ Exact performance !

Main result

Summing up:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

for $T = \{\frac{1}{p}\mathrm{tr} C_a C_b\}_{a,b=1}^k$.

- this is ... **a spiked model!!**
- as a consequence, we can anticipate:
  1. **phase transition phenomena**: below some threshold $g(M,T)$, **clustering impossible!**
  2. beyond the threshold, **eigenvector alignment known**

$\Rightarrow$ Exact performance !

Back to $D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$: can be studied similarly, just more painful!

Main result: comparison to simulations

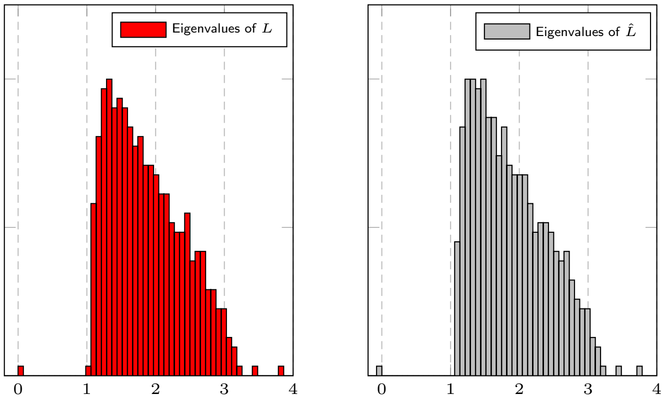


Figure: Eigenvalues of $L = D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ versus **Taylor expansion** $\hat{L}$, $k = 3$, $p = 2048$, $n = 512$, $c_1 = c_2 = 1/4$, $c_3 = 1/2$, $[\mu_a]_j = 4\boldsymbol{\delta}_{aj}$, $C_a = (1 + 2(a-1)/\sqrt{p})I_p$, $f(x) = \exp(-x/2)$ **(Gaussian kernel)**.

Main result: comparison to **real data** simulations

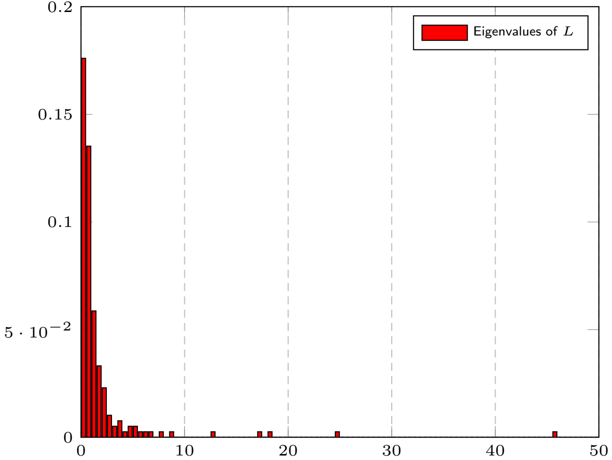


Figure: Eigenvalues of L (red) and **(equivalent Gaussian model)** $\hat{L}$ (white), MNIST data, $p = 784$, $n = 192$.

Main result: comparison to **real data** simulations

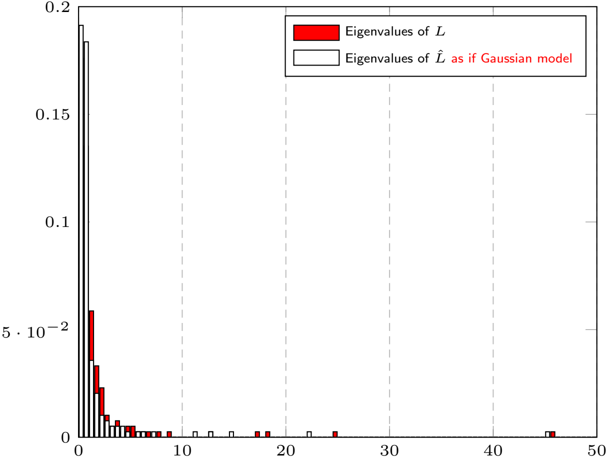


Figure: Eigenvalues of L (red) and **(equivalent Gaussian model)** $\hat{L}$ (white), MNIST data, $p = 784$, $n = 192$.

Main result: comparison to **real data** simulations

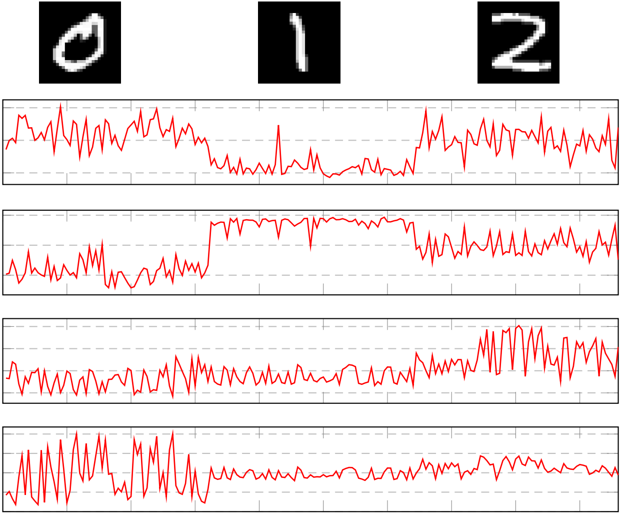


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST data (**red**) and theoretical findings (**blue**).

Main result: comparison to **real data** simulations

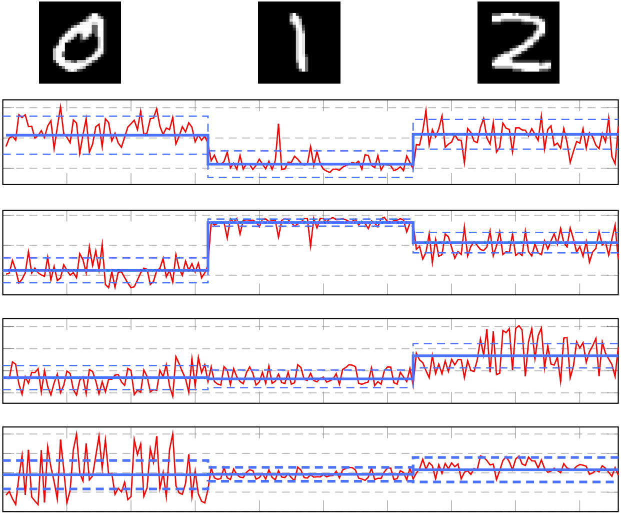


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ for MNIST data (**red**) and theoretical findings (**blue**).

Main result: comparison to **real data** simulations

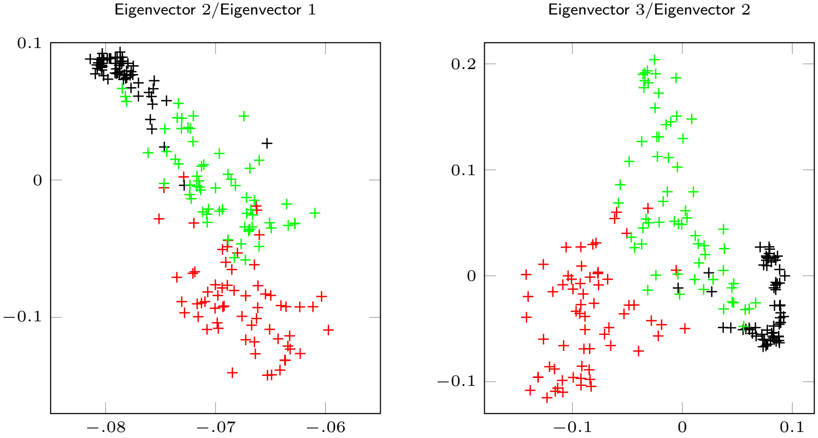


Figure: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Main result: comparison to **real data** simulations

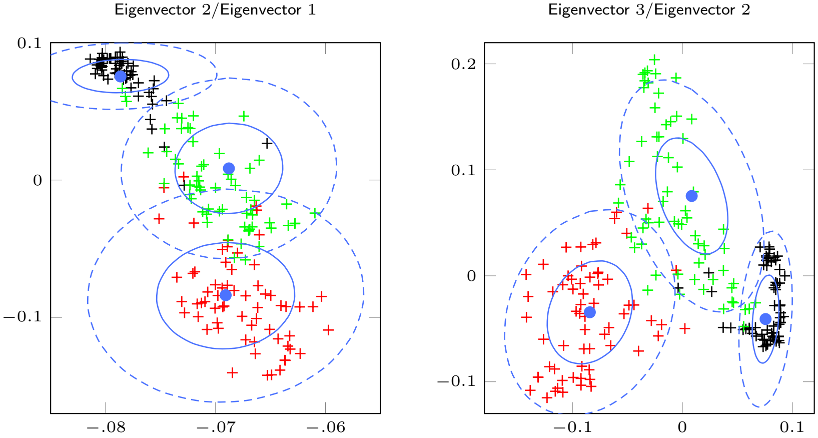


Figure: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p}JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$?

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} J T J^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears!

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow K$ **deterministic !!**

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow$ K **deterministic !!**

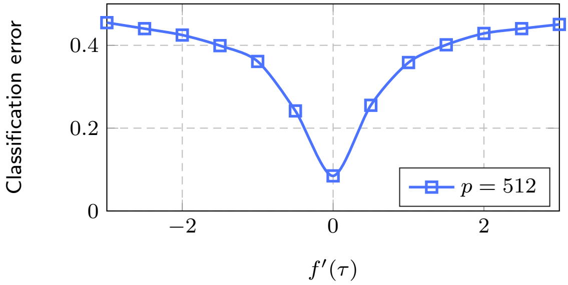


Figure: Kernel with $f(\tau) = 4$, $f''(\tau) = 2$, $x_i \sim \mathcal{N}(0, C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow$ K **deterministic !!**

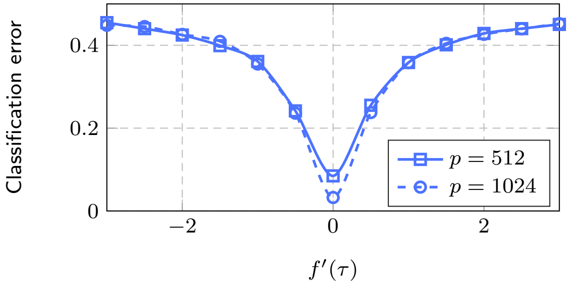


Figure: Kernel with $f(\tau) = 4$, $f''(\tau) = 2$, $x_i \sim \mathcal{N}(0, C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^{\mathsf{T}}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^{\mathsf{T}})^{\mathsf{T}}(W + MJ^{\mathsf{T}})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^{\mathsf{T}}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow K$ **deterministic !!**

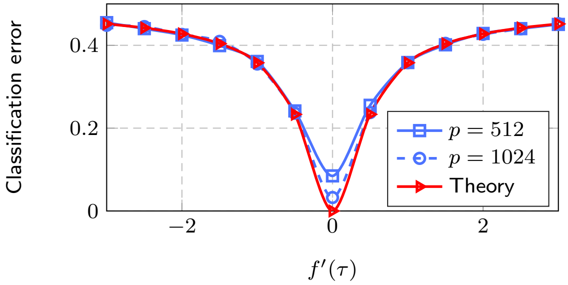


Figure: Kernel with $f(\tau) = 4$, $f''(\tau) = 2$, $x_i \sim \mathcal{N}(0, C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow K$ **deterministic !!**

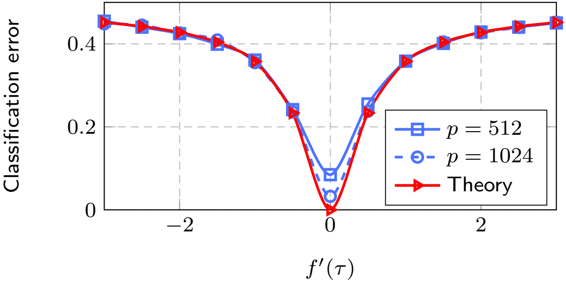


Figure: Kernel with $f(\tau) = 4$, $f''(\tau) = 2$, $x_i \sim \mathcal{N}(0, C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

- **Trivial classification** when $M = 0$ and $\|T\| = O(1)$.

Going further: the surprising $f'(\tau)=0$ case

Remember:

$$K = f(0)\underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + \underbrace{f'(0)\frac{1}{p}(W+MJ^\mathsf{T})^\mathsf{T}(W+MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2}\underbrace{\frac{1}{p}JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0)=0$? $\Rightarrow$ noise W disappears! $\Rightarrow K$ **deterministic !!**

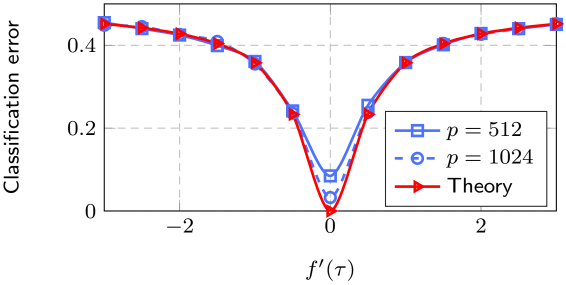


Figure: Kernel with $f(\tau)=4$, $f''(\tau)=2$, $x_i \sim \mathcal{N}(0,C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

- **Trivial classification** when $M=0$ and $\|T\| = O(1)$.
- this means we can **reduce** $\frac{1}{p}\operatorname{tr}((C_a - C_b)^2)$ **at least by** $1/\sqrt{p}$!

Going further: the surprising $f'(\tau) = 0$ case

Remember:

$$K = f(0) \underbrace{1_n 1_n^\mathsf{T}}_{=O_{\|\cdot\|}(n)} + f'(0) \underbrace{\frac{1}{p}(W + MJ^\mathsf{T})^\mathsf{T}(W + MJ^\mathsf{T})}_{=O_{\|\cdot\|}(1)} + \frac{f''(0)}{2} \underbrace{\frac{1}{p} JTJ^\mathsf{T}}_{=O_{\|\cdot\|}(1)} + o_{\|\cdot\|}(1)$$

- what if $f'(0) = 0$? $\Rightarrow$ noise W disappears! $\Rightarrow$ K **deterministic !!**

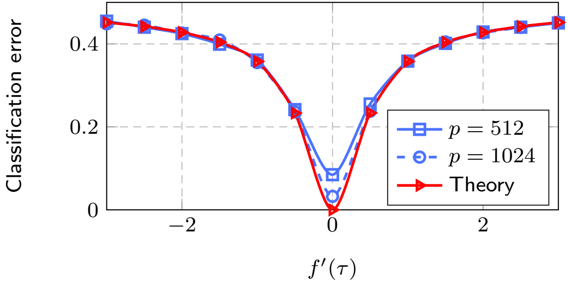


Figure: Kernel with $f(\tau) = 4$, $f''(\tau) = 2$, $x_i \sim \mathcal{N}(0, C_a)$, $C_1 = I_p$, $[C_2]_{i,j} = .4^{|i-j|}$, $c_0 = \frac{1}{4}$.

- **Trivial classification** when $M = 0$ and $\|T\| = O(1)$.
- this means we can **reduce** $\frac{1}{p}\mathrm{tr}((C_a - C_b)^2)$ **at least by** $1/\sqrt{p}$!
- **Key remark**: what is the shape of this "optimal" $f(t)$???

Outline

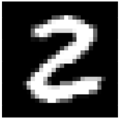
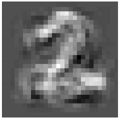
Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x, y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

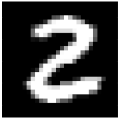
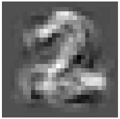
- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x, y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors

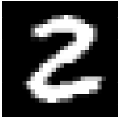
 

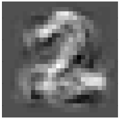
- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x,y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors





- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

Definition (Concentrated random vectors)

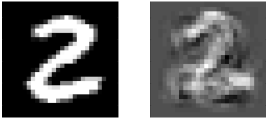
Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $z \in E$ is q-exponentially **concentrated** if for any 1-**Lipschitz function**[1] $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

$$\mathbb{P}\left\{|\mathcal{F}(z) - \mathbb{E}\mathcal{F}(z)| > t\right\} \leq Ce^{-c\,t^q}$$

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x,y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors



- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

Definition (Concentrated random vectors)

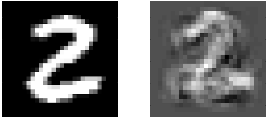
Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $z \in E$ is q-exponentially **concentrated** if for any 1-**Lipschitz function**[1] $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

$$\mathbb{P}\left\{|\mathcal{F}(z) - \mathbb{E}\mathcal{F}(z)| > t\right\} \leq C e^{-c\, t^q} \quad \text{denoted} \quad z \in \mathcal{O}(e^{-\cdot^q}).$$

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x, y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors



- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

Definition (Concentrated random vectors)

Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $z \in E$ is q-exponentially **concentrated** if for any 1-**Lipschitz function**[1] $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

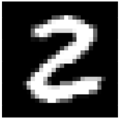
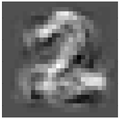
$$\mathbb{P}\left\{|\mathcal{F}(z) - \mathbb{E}\mathcal{F}(z)| > t\right\} \leq Ce^{-c\,t^q} \quad \text{denoted} \quad z \in \mathcal{O}(e^{-\cdot^q}).$$

(P1) $X \sim \mathcal{N}(0, I_p)$ is 2-exponentially **concentrated**.

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x, y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors

- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

Definition (Concentrated random vectors)

Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $z \in E$ is q-exponentially **concentrated** if for any 1-**Lipschitz function**[1] $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

$$\mathbb{P}\left\{|\mathcal{F}(z) - \mathbb{E}\mathcal{F}(z)| > t\right\} \leq Ce^{-c\,t^q} \quad \text{denoted} \quad z \in \mathcal{O}(e^{-\cdot^q}).$$

(P1) $X \sim \mathcal{N}(0, I_p)$ is 2-exponentially **concentrated**.

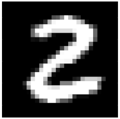
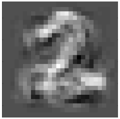
(P2) If $X \in \mathcal{O}(e^{-\cdot^q})$ and $\mathcal{G}$ is $\|\mathcal{G}\|_{lip}$-**Lipschitz**, then

$$\mathcal{G}(X) \in \mathcal{O}\left(e^{-(\cdot/\|\mathcal{G}\|_{lip})^q}\right).$$

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x,y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip}\,\|x - y\|_E$.

Notion of Concentrated Vectors

- **Observation**: RMT seems to predict ML performances for **real data** even under **Gaussian assumptions!**
- **But** real data badly modelled by Gaussian vectors

- **Fundamental remark**: Gaussian vectors are special instances of **concentrated random vectors**!

Definition (Concentrated random vectors)

Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $z \in E$ is q-exponentially **concentrated** if for any 1-**Lipschitz function**[1] $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

$$\mathbb{P}\left\{|\mathcal{F}(z) - \mathbb{E}\mathcal{F}(z)| > t\right\} \leq Ce^{-c\,t^q} \quad \text{denoted} \quad z \in \mathcal{O}(e^{-\cdot^q}).$$

(P1) $X \sim \mathcal{N}(0, I_p)$ is 2-exponentially **concentrated**.

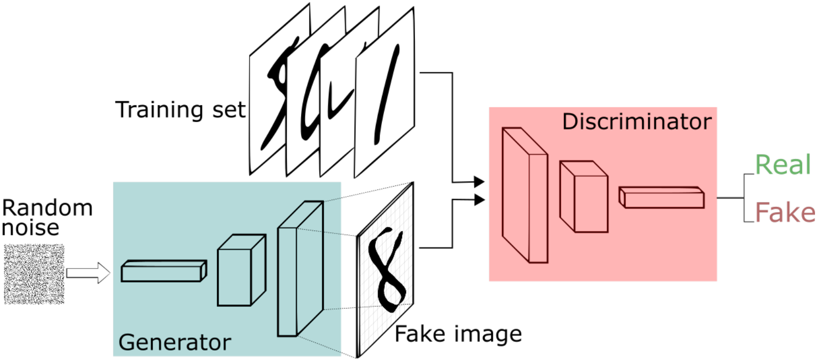
(P2) If $X \in \mathcal{O}(e^{-\cdot^q})$ and $\mathcal{G}$ is $\|\mathcal{G}\|_{lip}$-**Lipschitz**, then

$$\mathcal{G}(X) \in \mathcal{O}\left(e^{-(\cdot/\|\mathcal{G}\|_{lip})^q}\right).$$

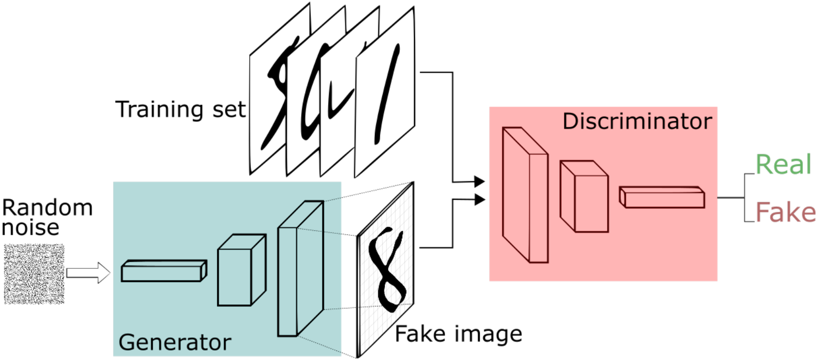
"Concentrated vectors are stable through Lipschitz maps."

[1] **Reminder:** $\mathcal{F} : E \to F$ is $\|\mathcal{F}\|_{lip}$-Lipschitz if $\forall (x, y) \in E^2 : \|\mathcal{F}(x) - \mathcal{F}(y)\|_F \leq \|\mathcal{F}\|_{lip} \|x - y\|_E$.

GAN data: An Example of Concentrated Vectors



GAN data: An Example of Concentrated Vectors



Generated image $= \mathcal{G}$(Gaussian) (with $\mathcal{G}$ Lipschitz!)

GAN data: An Example of Concentrated Vectors

Figure: Images generated by the BigGAN model [Brock *et al*, ICLR'19].

Mixture of Concentrated Vectors

New assumption: k **concentrated random vector classes**,

$$X = [\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}] \in \mathbb{R}^{p \times n}$$

Mixture of Concentrated Vectors

New assumption: k **concentrated random vector classes**,

$$X = [\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}] \in \mathbb{R}^{p\times n}$$

with

$$\mu_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i], \quad C_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i x_i^\intercal]$$

Mixture of Concentrated Vectors

New assumption: k concentrated random vector classes,

$$X = [\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}] \in \mathbb{R}^{p \times n}$$

with

$$\mu_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i], \;\; C_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i x_i^\mathsf{T}]$$

Examples:

- Gaussian mixture $x_i \sim \mathcal{N}(\mu_a, C_a)$

Mixture of Concentrated Vectors

New assumption: k **concentrated random vector classes**,

$$X = [\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}] \in \mathbb{R}^{p \times n}$$

with

$$\mu_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i], \;\; C_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i x_i^\intercal]$$

Examples:

- Gaussian mixture $x_i \sim \mathcal{N}(\mu_a, C_a)$
- mixture of Lipschitz functionals of Gaussians $x_i = \phi_a(z_i)$, $z_i \sim \mathcal{N}(0, I_p)$

Mixture of Concentrated Vectors

New assumption: k **concentrated random vector classes**,

$$X = [\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}] \in \mathbb{R}^{p \times n}$$

with

$$\mu_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i], \;\; C_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i x_i^\intercal]$$

Examples:

- Gaussian mixture $x_i \sim \mathcal{N}(\mu_a, C_a)$
- mixture of Lipschitz functionals of Gaussians $x_i = \phi_a(z_i)$, $z_i \sim \mathcal{N}(0, I_p)$
- **GAN images!**

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

Behavior of Gram Matrices for Concentrated Vectors

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \textit{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly**,

$$\left\|\mathbb{E}[Q(z)] - \bar{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right)$$

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly**,

$$\left\|\mathbb{E}[Q(z)] - \bar{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right)$$

where $\bar{Q}(z)$ **only depends on** $\mu_1, \ldots, \mu_k$, $C_1, \ldots, C_k$.

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly**,

$$\left\|\mathbb{E}[Q(z)] - \bar{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right)$$

where $\bar{Q}(z)$ **only depends on** $\mu_1, \ldots, \mu_k$, $C_1, \ldots, C_k$.

- **Exactly** the same result as for Gaussian mixture!

Behavior of Gram Matrices for Concentrated Vectors

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly***,*

$$\left\| \mathbb{E}[Q(z)] - \bar{Q}(z) \right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}} \right)$$

where $\bar{Q}(z)$ **only depends on** $\mu_1, \ldots, \mu_k$, $C_1, \ldots, C_k$.

- **Exactly** the same result as for Gaussian mixture!

$\Rightarrow$ universality

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly**,

$$\left\|\mathbb{E}[Q(z)] - \bar{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right)$$

where $\bar{Q}(z)$ **only depends on** $\mu_1, \ldots, \mu_k$, $C_1, \ldots, C_k$.

- **Exactly** the same result as for Gaussian mixture!

  $\Rightarrow$ universality

- **Careful** **(again!)**: does NOT mean that Gaussian mixture models real data!

Behavior of Gram Matrices for Concentrated Vectors

Theorem (Universality result)

Under the non-trivial growth rate assumptions,

$$Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q}) \text{ in } (\mathbb{R}^{n\times n}, \|\cdot\|)$$

and, **most importantly**,

$$\left\|\mathbb{E}[Q(z)] - \bar{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right)$$

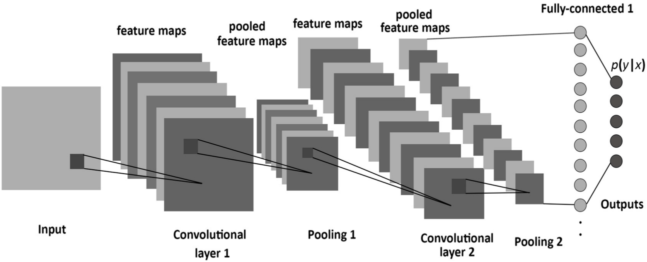
where $\bar{Q}(z)$ **only depends on** $\mu_1, \ldots, \mu_k$, $C_1, \ldots, C_k$.

- **Exactly** the same result as for Gaussian mixture!

  $\Rightarrow$ universality

- **Careful** (again!): does NOT mean that Gaussian mixture models real data!
- only means (but this is huge!) that **ML algorithms treat GAN data as if Gaussian**

Application to CNN Representations of GAN Images



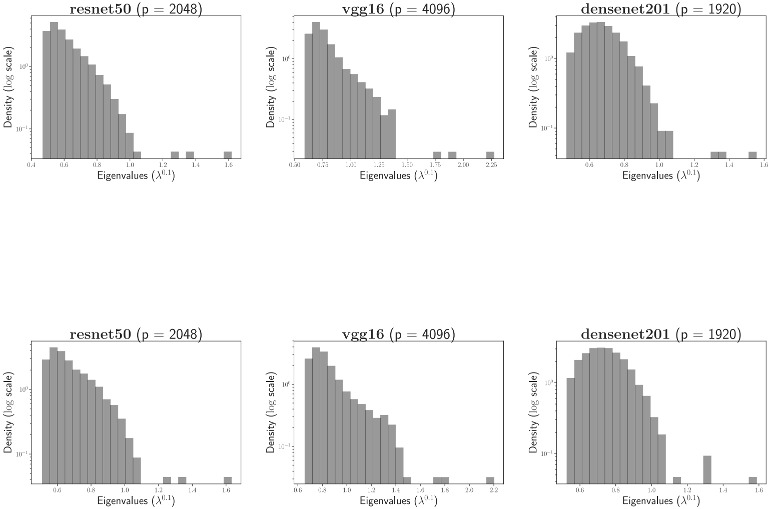
- CNN representations correspond to the **one before last** layer.

Application to CNN representations of GAN Images

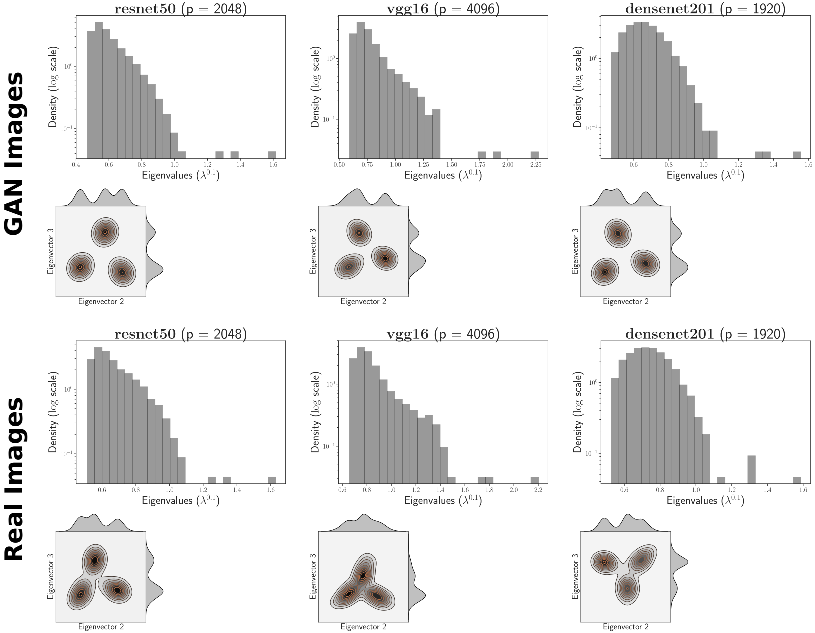
Application to CNN representations of GAN Images

GAN Images

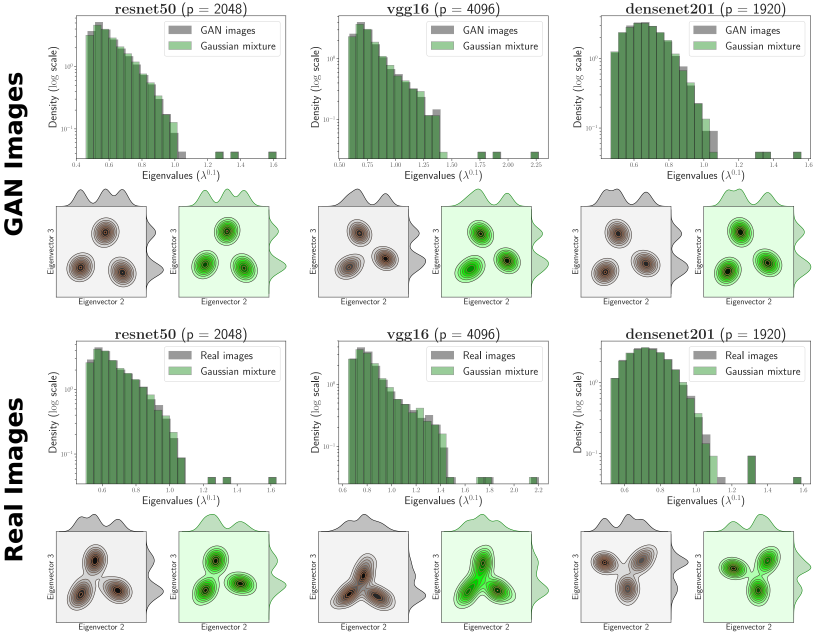
Real Images



Application to CNN representations of GAN Images



Application to CNN representations of GAN Images



Takeaway message

Random matrix theory **explains the inner working of practical ML algorithms**

Takeaway message

Random matrix theory **explains the inner working of practical ML algorithms**

and this is **provably valid for real data!**