

UNIVERSITÄT POTSDAM

MATHEMATISCHES VORTRAGEN UND SCHREIBEN

Annäherung an die Spektraldichte großer Matrizen

Autorin:

Carina SEIDEL

Betreuer:

Dr. Thomas MACH

29. September 2024



Zusammenfassung

Die Spektraldichte einer Matrix, auf Englisch auch Density of States (DOS) genannt, gibt die Wahrscheinlichkeit für Eigenwerte nahe eines Punktes $p \in \mathbb{R}$ an. Dies ist in vielen Bereichen der Physik, wie zum Beispiel der Quantenmechanik, von Interesse. Da Matrizen oft zu groß sind, um Eigenwerte effizient zu berechnen, gibt es verschiedene Methoden, sich der Spektraldichte anzunähern. Eine solche Methode ist die Kernel-Polynom-Methode, das Hauptthema dieses Berichts.

Der Ausgangspunkt für diese Arbeit ist das Paper APPROXIMATING SPECTRAL DENSITIES OF LARGE MATRICES von Lin Lin, Yousef Saad, and Chao Yang [1]. Meine eigene Arbeit war das Ergänzen und weitere Ausarbeiten einiger Definitionen, Erklärungen und Beweise zum besseren Verständnis sowie die Übersetzung ins Deutsche.

Inhaltsverzeichnis

1	Einleitung	2
1.1	Einführende Definitionen	2
1.2	Motivation und Problemstellung	3
2	Kernel-Polynom-Methode (KPM)	5
2.1	Überblick	5
2.2	Polynomiale Erweiterung durch Tschebyshev-Polynome . . .	6
2.3	Annäherung der Spektraldichte	7
3	Qualitätsanalyse	11
3.1	Der Begriff der Auflösung	11
3.2	Einschränkung des Schwartz-Raums	11
3.3	Regularisierung der δ -Distribution	13
3.4	Die Bedingung der Nicht-Negativität	14
4	Anhang	15

Einleitung

Um den Begriff der Spektraldichte zu definieren und zu verstehen, benötigen wir einige Definitionen vorab. Danach betrachten wir, warum die Berechnung der Spektraldichte nicht trivial ist.

1.1 Einführende Definitionen

Definition 1 (Funktional). Sei V ein \mathbb{R} -Vektorraum. Ein *Funktional* T ist eine Abbildung $T : V \rightarrow \mathbb{R}$

Definition 2 (Distribution). Sei $\emptyset \neq \Omega \subset \mathbb{R}^n$ offen. Sei \mathcal{E} der Raum der *Testfunktionen* über Ω . Eine *Distribution* T ist eine Abbildung $T : \mathcal{E} \rightarrow \mathbb{R}$ wobei für alle $g, g_1, g_2, \{g_n\}_{n \in \mathbb{N}} \in \mathcal{E}$ mit $\lim_{n \rightarrow \infty} g_n \rightarrow g$ gilt:

$$T(g_1 + \lambda g_2) = T(g_1) + \lambda T(g_2) \quad \text{und} \quad \lim_{n \rightarrow \infty} T(g_n) \rightarrow T(g)$$

Kurz: Eine Distribution T ist ein stetiges und lineares Funktional auf \mathcal{E}

Definition 3 (Delta-Distribution). Sei $\mathcal{E} = \mathcal{C}^\infty(\Omega)$ mit $0 \in \Omega \subset \mathbb{R}^n$. Dann ist

$$\delta : \mathcal{E} \rightarrow \mathbb{R}, f \mapsto f(0) \quad \text{mit} \quad \delta(f) = \langle \delta, f \rangle = f(0)$$

Bei dieser Definition ist die folgende wichtige Eigenschaft zu beachten:

$$\int_{-\infty}^{\infty} f(x) \delta(x - a) \, dx = \int_{-\infty}^{\infty} f(x) \delta(a - x) \, dx = f(a) \implies \int_{-\infty}^{\infty} \delta(x - a) \, dx = 1$$

Oftmals wird die Delta-Distribution auch als Delta-"Funktion" bezeichnet, obwohl es sich hierbei nicht um eine Funktion im eigentlichen Sinne handelt.

Dies führt nun zur zentralen Definition des Papers:

Definition 4 (Spektraldichte). Sei $A \in \mathbb{R}^{n \times n}$, $A^T = A$ und A spärlich besetzt. Dann ist die Spektraldichte definiert als

$$\phi(t) = \frac{1}{n} \sum_{j=1}^n \delta(t - \lambda_j)$$

wobei δ die Delta-Distribution und λ_j die Eigenwerte von A in nicht-absteigender Reihenfolge sind.

Die Anzahl der Eigenwerte in einem Intervall $[a, b]$ kann dann wie folgt ausgedrückt werden:

$$\nu_{[a,b]} = \int_a^b \sum_j \delta(t - \lambda_j) dt \equiv \int_a^b n\phi(t) dt \quad (1.1)$$

Definition 5 (Schwartz-Raum über \mathbb{R}). Der Schwartz-Raum über \mathbb{R} umfasst alle glatten Funktionen f , die schnell genug gegen 0 abfallen, wenn $|x|$ gegen unendlich geht. [3] In Formeln

$$\mathcal{S}(\mathbb{R}) := \left\{ f \in \mathcal{C}^\infty(\mathbb{R}) \mid \forall p, k \in \mathbb{N}_0 : \sup_{x \in \mathbb{R}} |x^p f^{(k)}(x)| < \infty \right\}$$

Im Weiteren werde ich das Symbol \mathcal{S} als Abkürzung für $\mathcal{S}(\mathbb{R})$ benutzen, da sich diese Arbeit allein mit dem reellen Kontext befasst.

1.2 Motivation und Problemstellung

Die Berechnung der Spektraldichte einer Matrix ist trivial, wenn die Eigenwerte bereits bekannt sind. Meistens ist das allerdings nicht der Fall, und die Berechnung der Eigenwerte bei sehr großen Matrizen ist zeit- und energieaufwendig. Gleichzeitig ist die DOS als eine Art Wahrscheinlichkeitsdichte über die Verteilung der Eigenwerte in vielen Anwendungen von großem Interesse. Es gibt also Bedarf an effizienten Methoden, die Spektraldichte möglichst kostengünstig abzuschätzen. Das Problem ist dabei, dass $\phi(t)$ aufgrund der Delta-Distribution keine "Funktion" im eigentlichen Sinne ist, die man punktweise auswerten kann.

Eine intuitive Idee wäre es zum Beispiel, ein Intervall $I \in \mathbb{R}$ zu wählen, sodass das Spektrum von A , $\sigma(A)$, eine Teilmenge von I ist. Wähle nun k Punkte t_i aus I , sodass diese das Intervall in Teilintervalle unterteilen:

$$\{t_i\}_{i=1}^k \subset I \quad \text{mit} \quad \bigcup_{i=1}^{k-1} [t_i, t_{i+1}] = I$$

Zähle nun die Eigenwerte in jedem Teilintervall. Berechne anschließend den Durchschnittswert von $\phi(t)$ in jedem Intervall mit $\nu_{[a,b]}$ aus Gleichung 1.1. Als Ergebnis erhält man Histogramme, die sich mit kleiner werdenden Teilintervallen, also größerem k und $(t_{i+1} - t_i) \rightarrow 0$, der Spektraldichte annähern. Hierbei gibt es leider ein Problem: Um die Eigenwerte in den Intervallen zu zählen, bedient man sich Hilfsmitteln wie zum Beispiel dem Sylvestreschen Trägheitssatz. Die Einzelheiten dieser Methode sind nicht Bestand dieser Arbeit, es wäre allerdings notwendig, eine Zerlegung $A - t_i I = LDL^T$ für alle t_i zu berechnen [2]. Dies sind teure und zeitaufwendige Berechnungen und damit keine gute Lösung. Nur eine Prozedur, bei der A mit Vektoren multipliziert wird, ist auch in größeren Dimensionen lohnenswert. Im Folgenden werden wir der Einfachheit halber immer annehmen, dass A symmetrisch und reell ist. Die Erweiterung auf hermitesche Matrizen ist eine vergleichsweise einfache Angelegenheit.

Kernel-Polynom-Methode (KPM)

2.1 Überblick

Bei der sogenannten Kernel-Polynom-Methode, kurz KPM, handelt es sich vielmehr um eine Klasse von Methoden, die mehrere Varianten umfasst. Wir werden im Folgenden die allgemeine Herangehensweise betrachten.

Wie der Name bereits nahelegt, ist die KPM eine polynomiale Erweiterung der Spektraldichte. Dabei werden die Koeffizienten der Polynome aus der Momentenmethode abgeleitet, um wie in der Statistik eine Schätzfunktion zu erhalten. Diese Methode beruht auf einem Resultat aus dem folgenden Theorem:

Theorem 1. Sei $A = A^T \in \mathbb{R}^{n \times n}$ mit Spektralzerlegung

$$A = U\Lambda U^T \quad \text{wobei} \quad UU^T = \mathbf{1}_n \text{ und } \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$$

Seien außerdem $\beta, v \in \mathbb{R}^n$ mit $v = U\beta$.

Gilt $v_i \sim_{\text{i.i.d.}} \mathcal{N}(0, 1)$ für die Komponenten $\{v_i\}_{i=1}^n$ von v , also

$$\mathbb{E}[v] = 0 \text{ und } \mathbb{E}[vv^T] = \mathbf{1}_n, \tag{2.1}$$

dann

$$\mathbb{E}[\beta\beta^T] = \mathbf{1}_n$$

Beweis Theorem 1. Es gilt

$$\mathbb{E}[v] = \mathbb{E}[U\beta] = U\mathbb{E}[\beta] = 0 \implies \mathbb{E}[\beta] = 0$$

Weiterhin gilt, dass

$$\mathbb{E}[vv^T] = \mathbb{E}[(U\beta)(U\beta)^T] = \mathbb{E}[U\beta\beta^T U^T] = U\mathbb{E}[\beta\beta^T]U^T = \mathbf{1}_n$$

Hieraus folgt, dass $\mathbb{E}[\beta\beta^T] = \mathbf{1}_n$ □

Dieses Theorem hat ein schönes Resultat, wenn man nun eine Matrixfunktion $f(A)$ betrachtet. Dann haben wir

$$\begin{aligned}
\mathbb{E} [v^T f(A) v] &= \mathbb{E} [(U\beta)^T f(U\Lambda U^T)(U\beta)] = \mathbb{E} [\beta^T U^T U f(\Lambda) U^T U \beta] \\
&= \mathbb{E} [\beta^T f(\Lambda) \beta] \\
&= \mathbb{E} \left[\sum_{j=1}^n \beta_j^2 f(\lambda_j) \right] \\
&= \sum_{j=1}^n f(\lambda_j) \mathbb{E} [\beta_j^2] \\
&= \sum_{j=1}^n f(\lambda_j)
\end{aligned}$$

also zusammengefasst

$$\mathbb{E} [v^T f(A) v] = \text{Spur}(f(A)) \quad (2.2)$$

2.2 Polynomiale Erweiterung durch Tschebyschev-Polynome

Aufgrund ihrer vielen einzigartigen Eigenschaften sind Tschebyschev-Polynome besonders gut zur polynomialen Erweiterung der Delta-Distribution geeignet. Mit Hilfe der trigonometrischen Funktionen können sie auch wie folgt ausgedrückt werden:

$$T_k(t) = \begin{cases} \cos(k \arccos(t)) & \text{für } k \in [-1, 1] \\ \cosh(k \operatorname{arcosh}(t)) & \text{für } k > 1 \\ (-1)^k \cosh(k \operatorname{arcosh}(-t)) & \text{für } k < -1 \end{cases}$$

Wir benutzen im Folgenden nur die Formel $T_k(t) = \cos(k \arccos(t))$. Daher müssen wir uns auf Matrizen beschränken, deren Eigenwerte im Intervall $[-1, 1]$ liegen. Sollte diese Voraussetzung nicht erfüllt sein, kann man die Eigenwerte entsprechend transformieren. Seien dazu λ_{us} und λ_{os} jeweils die untere bzw. obere Schranke für die Eigenwerte von A . Definiere

$$c := \frac{\lambda_{us} + \lambda_{os}}{2} \quad \text{und} \quad d := \frac{\lambda_{os} - \lambda_{us}}{2}$$

Dann ist $B = \frac{A - c \mathbf{1}_n}{d}$ eine Matrix mit Eigenwerten im Intervall $[-1, 1]$. Eine Veranschaulichung dazu ist im Anhang verlinkt.

Tschebyschev-Polynome können zudem mit der Rekursionsformel

$$T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$$

berechnet werden, wobei die Startbedingungen $T_0(t) = 1$ und $T_1(t) = t$ gelten.

Beachte auch, dass das Resultat in Gleichung 2.2 besagt, dass

$$\mathbb{E} [v^T T_k(A) v] = \sum_{j=1}^n T_k(\lambda_j) = \text{Spur}(T_k(A)) \quad (2.3)$$

gilt. Dies ist zentral im weiteren Vorgehen.

Sei nun

$$h(x) = \frac{1}{\sqrt{1-x^2}} \quad (2.4)$$

eine Gewichtsfunktion. Eine weitere Eigenschaft der Tschebyschev-Polynome ist, dass sie *orthogonal* bezüglich des mit h gewichteten Skalarproduktes

$$\langle f, g \rangle = \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} \cdot f(x) \cdot g(x) \, dx$$

sind. Das bedeutet, dass

$$\int_{-1}^1 \frac{1}{\sqrt{1-t^2}} \cdot T_k(t) \cdot T_l(t) \, dt = \begin{cases} 0 & \text{für } k \neq l \\ \pi & \text{für } k = l = 0 \\ \frac{\pi}{2} & \text{für } k = l \neq 0 \end{cases}$$

2.3 Annäherung der Spektraldichte

Multipliziere nun die Spektraldichte mit dem Inversen der Gewichtsfunktion 2.4:

$$\hat{\phi}(t) = \sqrt{1-t^2} \phi(t) = \sqrt{1-t^2} \times \frac{1}{n} \sum_{j=1}^n \delta(t - \lambda_j)$$

Sei nun $g \in \mathcal{S}$, dem in Definition 5 beschriebenen Schwartz-Raum, und $\mu_k \in \mathbb{R}$ Koeffizienten, die wir nachher berechnen, sodass die folgende Gleichung gilt:

$$\int_{-1}^1 \hat{\phi}(t) g(t) \, dt = \int_{-1}^1 \sum_{k=0}^{\infty} \mu_k T_k(t) g(t) \, dt \quad (2.5)$$

Gilt dies für beliebige $g \in \mathcal{S}$, so vereinfachen wir Gleichung 2.5 zu

$$\hat{\phi}(t) = \sum_{k=0}^{\infty} \mu_k T_k(t) \quad (2.6)$$

Nutze nun die Orthogonalität der Tschebyschev-Polynome aus, um einen bestimmten Koeffizienten μ_k zu berechnen:

$$\begin{aligned} \sum_{l=0}^{\infty} \mu_l T_l(t) = \hat{\phi}(t) &\implies \left(\sum_{l=0}^{\infty} \mu_l T_l(t) \right) \cdot T_k(t) = \hat{\phi}(t) \cdot T_k(t) \\ &\implies \int_{-1}^1 \frac{1}{\sqrt{1-t^2}} \cdot \left(\sum_{l=0}^{\infty} \mu_l T_l(t) \right) \cdot T_k(t) dt = \int_{-1}^1 \frac{1}{\sqrt{1-t^2}} \cdot \hat{\phi}(t) \cdot T_k(t) dt \\ &\implies \mu_k \cdot \frac{\pi}{2 - \delta_{k0}} = \int_{-1}^1 \frac{1}{\sqrt{1-t^2}} \cdot \sqrt{1-t^2} \cdot \phi(t) \cdot T_k(t) dt \\ &\implies \mu_k = \frac{2 - \delta_{k0}}{\pi} \cdot \int_{-1}^1 \phi(t) \cdot T_k(t) dt \end{aligned}$$

Durch Anwendung der Delta-Funktion erhält man:

$$\begin{aligned} \mu_k &= \frac{2 - \delta_{k0}}{\pi} \cdot \int_{-1}^1 \phi(t) \cdot T_k(t) dt = \frac{2 - \delta_{k0}}{\pi} \cdot \int_{-1}^1 \frac{1}{n} \sum_{j=1}^n \delta(t - \lambda_j) \cdot T_k(t) dt \\ &= \frac{2 - \delta_{k0}}{n\pi} \sum_{j=1}^n T_k(\lambda_j) \\ &= \frac{2 - \delta_{k0}}{n\pi} \text{Spur}(T_k(A)) \end{aligned}$$

Sei nun $n_{vec} \in \mathbb{R}$ und $v_0^{(1)}, v_0^{(2)}, \dots, v_0^{(n_{vec})}$ Vektoren, die die Bedingungen aus dem Theorem erfüllen, also $\mathbb{E}[v_0^{(k)}] = 0$ und $\mathbb{E}\left[v_0^{(k)} \left(v_0^{(k)}\right)^T\right] = \mathbb{1}_n$. Aus Gleichung 2.3 folgt, dass

$$\zeta_k = \frac{1}{n_{vec}} \sum_{l=1}^{n_{vec}} \left(v_0^{(l)}\right)^T T_k(A) v_0^{(l)}$$

ein guter Schätzer für $\text{Spur}(T_k(A))$ ist und damit

$$\mu_k \approx \frac{2 - \delta_{k0}}{n\pi} \zeta_k$$

Um die ζ_k zu bestimmen, sei im Folgenden $v_0 \equiv v_0^{(l)}$. Berechne nun mit Hilfe der Rekursionsformel für Tschebyschev-Polynome:

$$T_{k+1}(A)v_0 = 2AT_k(A)v_0 - T_{k-1}(A)v_0$$

Für $v_k \equiv T_k(A)v_0$ gilt also, dass

$$v_{k+1} = 2Av_k - v_{k-1}$$

Damit sind alle Bauteile zur Berechnung festgelegt und das Ziel der KPM erreicht: Anstatt rechenaufwendig Matrizen mit anderen Matrizen zu multiplizieren, müssen wir sie nur noch mit Vektoren multiplizieren. Nun können wir $\phi(t)$ beliebig nah annähern. Wie bereits erwähnt, ist eine unendlich genaue Annäherung nicht immer wünschenswert. Wegengilt

$$\lim_{k \rightarrow \infty} \mu_k \rightarrow 0$$

und wir interessieren uns nur für $T_k(t)$ mit $k \leq M$

Daher schätzen wir ϕ durch

$$\tilde{\phi}_M(t) = \frac{1}{\sqrt{1-t^2}} \sum_{k=0}^M \mu_k T_k(t) \quad (2.7)$$

Der folgende Pseudocode basiert auf [1, p. 10] und fasst die oben beschriebenen Schritte zusammen. Ich habe ihn selbst implementiert und im Anhang verlinkt.

Algorithm 1 Die Kernel-Polynom-Methode

Input: $A = A^T \in \mathbb{R}^{n \times n}$ mit Eigenwerten aus dem Intervall $[-1, 1]$

Output: Geschätzte Spektraldichte $\{\tilde{\phi}_M(t_i)\}$

```
    for  $k = 0 : M$  do
         $\zeta_k \leftarrow 0$ 
    end for
5: for  $l = 1 : n_{\text{vec}}$  do
    Wähle einen neuen zufälligen Vektor  $v_0^{(l)}$ ;  $\triangleright v_{0_i}^{(l)} \sim \text{i.i.d. } \mathcal{N}(0, 1)$ 
    for  $k = 0 : M$  do
        Berechne  $\zeta_k \leftarrow \zeta_k + \left(v_0^{(l)}\right)^T v_k(l)$ ;
        if  $k = 0$  then
10:          $v_1^{(l)} \leftarrow A v_0^{(l)}$ 
        else
             $v_{k+1}^{(l)} \leftarrow 2A v_k^{(l)} - v_{k-1}^{(l)}$   $\triangleright$  Drei-Term-Rekursion
        end if
    end for
15: end for
    for  $k = 0 : M$  do
         $\zeta_k \leftarrow \frac{\zeta_k}{n_{\text{vec}}}$ 
         $\mu_k \leftarrow \frac{2 - \delta_{k0}}{n\pi} \zeta_k$ 
    end for
20: Werte  $\tilde{\phi}_M(t_i)$  mit Gleichung 2.7 aus
```

Qualitätsanalyse

Sei $\tilde{\phi}(t)$ eine reguläre Funktion, die die Spektraldichte schätzt. Da $\phi(t)$ keine Funktion im eigentlichen Sinne ist, aber alle Annäherungen stetige Funktionen sind, ist

$$\left\| \phi(t) - \tilde{\phi}(t) \right\|_{L^p}$$

nicht definiert, um den Fehler abzuschätzen. Im Folgenden werden zwei Möglichkeiten vorgestellt, um dieses Problem zu umgehen. Die erste nutzt die Tatsache aus, dass $\delta(t)$ eine Distribution ist, die auf Testfunktionen aus dem Schwartz-Raum \mathcal{S} angewendet werden kann, während die zweite δ -Funktionen durch stetige und glatte Funktionen ersetzt und damit regularisiert. Diese zwei Vorgehensweisen sind eng miteinander verwandt, wie wir später noch sehen werden.

3.1 Der Begriff der Auflösung

In der Anwendung gibt es nur selten Bedarf für exakte Annäherung aller Eigenwerte von A . Diese wäre sogar höchst unstetig, wenn man sich an den intuitiven Ansatz mit den Histogrammen zurückerinnert. Die Exaktheit der Annäherung sollte sich daher an der gewünschten *Auflösung* orientieren. Hierbei ist gemeint, dass oft nur die Eigenwerte in einem Teilintervall $[a, b] \subset \sigma(A)$ interessant in einem gewissen Zusammenhang sind. Die Größe $b - a$ heißt dann Auflösung der Schätzung. Diese Erkenntnis gilt es als Parameter in unsere weiteren Betrachtungen mitzunehmen.

3.2 Einschränkung des Schwartz-Raums

In dieser ersten Methode fassen wir also $\delta(t)$ als Distribution auf. Sei nun $g \in \mathcal{C}^\infty(\mathbb{R})$ eine Testfunktion aus dem Schwartz-Raum \mathcal{S} definiert in Definition

5. Dann bedeutet dies gemäß Definition 2 der Distribution, dass

$$\langle \delta(\cdot - \lambda), g \rangle = \int_{-\infty}^{\infty} \delta(t - \lambda) g(t) \, dt = g(\lambda)$$

und für alle $p, k \in \mathbb{N}_0$

$$\sup_{t \in \mathbb{R}} |t^p g^{(k)}(t)| < \infty$$

Dann kann der Fehler wie folgt gemessen werden:

$$\epsilon_1 = \sup_{g \in \mathcal{S}} \left| \langle \phi, g \rangle - \langle \tilde{\phi}, g \rangle \right|$$

Um den Begriff der Auflösung auf diesen Fehler zu übertragen, machen wir zunächst eine Vorabüberlegung. Aus Gleichung 1.1 betrachte $\nu_{[a,b]}$ und definiere entsprechend

$$\tilde{\nu}_{[a,b]} = \int_a^b n \tilde{\phi}(t) \, dt$$

mit $\tilde{\phi}(t) \in \mathcal{C}^\infty(\mathbb{R})$. Angenommen, $n = 1$ und somit $\phi(t) = \delta(t)$. Eine unendliche Auflösung würde bedeuten, dass für beliebig kleine Intervalle $[a, b]$ der Term $|\nu_{[a,b]} - \tilde{\nu}_{[a,b]}|$ ebenfalls beliebig klein wird. Sei also $a = -\varepsilon, b = \varepsilon$. Aus der Definition der δ -Funktion folgt dann, dass

$$\lim_{\varepsilon \rightarrow 0+} \nu_{[-\varepsilon, \varepsilon]} = 1$$

während für glatte Funktionen $\tilde{\phi}$ selbstverständlich gilt, dass

$$\lim_{\varepsilon \rightarrow 0+} \tilde{\nu}_{[-\varepsilon, \varepsilon]} = 0$$

Dies zeigt, dass keine glatte Funktion unter stetiger Erhöhung der Auflösung zur Spektraldichte konvergiert. Eine sorgfältig abgewogene Annäherung wäre demnach genauso präzise wie eine konstante.

Wie im Vorherein festgestellt, ist eine endliche Auflösung aber oftmals genug. Wir können den Schwartz-Raum \mathcal{S} also einschränken. Zum Beispiel könnte man nur Gaussche Verteilungsfunktionen der Form

$$g_\sigma(t) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{t^2}{2\sigma^2}}$$

betrachten und \mathcal{S} dementsprechen auf den Unterraum

$$\mathcal{S}(\sigma; [\lambda_{us}, \lambda_{os}]) = \{g \mid g(t) \equiv g_\sigma(t - \lambda), \lambda \in [\lambda_{us}, \lambda_{os}]\}$$

einschränken. Hierbei sind λ_{us} und λ_{os} wie im vorherigen Abschnitt die jeweils obere und untere Schranke der Eigenwerte von A , während der Parameter σ die *Zielauflösung* ist. Wir können nun die folgende Metrik zur Qualitätsbewertung nutzen:

$$E \left[\tilde{\phi}; \mathcal{S}(\sigma; [\lambda_{lb}, \lambda_{ub}]) \right] = \sup_{g \in \mathcal{S}(\sigma; [\lambda_{lb}, \lambda_{ub}])} \left| \langle \phi, g \rangle - \langle \tilde{\phi}, g \rangle \right| \quad (3.1)$$

3.3 Regularisierung der δ -Distribution

In dieser zweiten Methode regularisieren wir die δ -Funktionen zum Beispiel mit der Gaußschen Normalverteilung mit angemessener Standardabweichung σ . Angemessene Standardabweichung bedeutet hier im Verhältnis zur gewünschten Auflösung: Je kleiner das σ ist, desto höher wird die Auflösung. Das σ sollte dabei so groß wie möglich für leichte Annäherung und so klein wie möglich für Genauigkeit sein. Die daraus entstandene Funktion $\phi_\sigma(t)$ ist dann wohldefiniert. Für $p = 1, 2$ und ∞ kann dann der folgende Fehler berechnet werden:

$$\epsilon_2 = \left\| \phi_\sigma(t) - \tilde{\phi}(t) \right\|_p$$

Sei also

$$\phi_\sigma(t) = \langle \phi(\cdot), g_\sigma(\cdot - t) \rangle = \sum_{j=1}^n g_\sigma(t - \lambda_j)$$

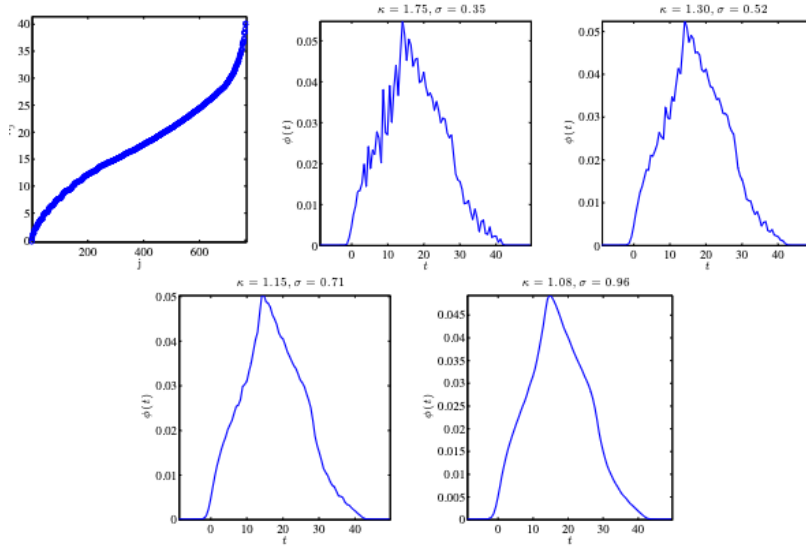
Dies ist dann nicht anderes als die "Weichzeichnung" der Spektraldichte durch Gauß-Funktionen der Breite σ . Genauso sei

$$\tilde{\phi}_\sigma(t) = \langle \tilde{\phi}(\cdot), g_\sigma(\cdot - t) \rangle$$

Dann ist

$$E \left[\tilde{\phi}; \mathcal{S}(\sigma; [\lambda_{lb}, \lambda_{ub}]) \right] = \sup_{g \in \mathcal{S}(\sigma; [\lambda_{lb}, \lambda_{ub}])} \left| \phi_\sigma(t) - \tilde{\phi}_\sigma(t) \right|$$

der L^∞ -Fehler zwischen zwei wohldefinierten Funktionen. Genau wie bei der anderen Methode, kontrolliert σ die Auflösung: Je größer das σ , desto glatter, aber auch ungenauer, ist die Annäherung. Bei kleineren Werten erhält man grobere Funktionen, bei denen man kleine Spitzen genau da erkennen kann, wo sich die Eigenwerte befinden. Die folgende Grafik entstammt aus [1, p. 6] und veranschaulicht diesen Effekt für vier unterschiedliche Werte von σ :



3.4 Die Bedingung der Nicht-Negativität

Als Wahrscheinlichkeitsverteilung ist die Spektraldichte nicht-negativ, also

$$\forall g \in \mathcal{S}, g \geq 0 : \langle \phi, g \rangle \geq 0$$

Einige numerische Annäherungen brechen mit dieser Eigenschaft, darunter auch die Kernel-Polynom-Methode. Das führt zu großen Fehlern und muss daher unbedingt beachtet werden. Methoden, um diese negativen Effekte einzudämmen, würden den Rahmen dieser Arbeit aber deutlich überschreiten.

Anhang

Hier ist der Link zu meinem GitHub Repository, in dem sich zusätzlich zu sämtlichen Latex Dateien für sowohl diese Arbeit als auch die Präsentation, ausführbarer, selbstgeschriebener Python-Code zur Kernel-Polynom-Methode befindet. <https://github.com/ClarinaSaxel/Spektraldichte>

Literaturverzeichnis

- [1] LIN LIN, YOUSSEF SAAD AND CHAO YANG, *Approximating Spectral Densities of Large Matrices*, <https://arxiv.org/pdf/1308.5467v2>
5. Oktober 2014.
- [2] G. H. GOLUB UND C. F. VAN LOAN, *Matrix Computations, 4th Edition*,
Johns Hopkins University Press, Baltimore, MD, 3rd ed., 2013
- [3] R. D. RICHTMYER, *Principles of advanced mathematical physics*, vol 1,
Springer Verlag, New York, 1981