Universität Leipzig
Fakultät für Mathematik und Informatik
Mathematisches Institut

Diplomarbeit

Expandergraphen und Derandomisierung

Marburg, den 7. August 2014

Vorgelegt von: Christian Fetsch
Studiengang: Mathematik, Diplom
Betreuer: Prof. Dr. Andreas Thom

Danken möchte ich Professor Andreas Thom für seine freundliche geduldige Betreuung; außerdem meinen Eltern für ihre Unterstützung; und meiner Frau Karin, die immer an mich geglaubt hat.

Inhaltsverzeichnis

1 Einführung

In der vorliegenden Arbeit wird der Artikel „Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications“ von Wigderson und Xiao [16] detailliert nachvollzogen. Es wird aufgezeigt, wie der Beweis der matrixwertigen Chernoff-Ungleichung von Ahlswede und Winter [1] verläuft. Mit diesen Ergebnissen und der Methode der pessimistischen Schätzer wird schließlich der Beweis des Alon-Roichman-Theorems [3] entrandomisiert.

Zunächst wird ein kurzer Überblick über das Gebiet der Expandergraphen gegeben. Es folgt die Vorarbeit zum Haupttheorem in Form nützlicher Sätze und Lemmata. Dieses wird im Abschnitt „Matrixwertige Chernoff-Abschätzung“ behandelt. Im Abschnitt „Probabilistische Beweise und Derandomisierung“ geht es um die Methode der pessimistischen Schätzer, mit der einige probabilistische Beweise verbessert werden können. Mit der Derandomisierung des Haupttheorems in diesem Abschnitt ergeben sich Anwendungen in zahlreichen Bereichen der Mathematik. Eine davon ist die Entrandomisierung des Alon-Roichman-Theorems, welche die deterministische Konstruktion eines bestimmten Expandergraphen erlaubt. Diese wird im letzten Abschnitt behandelt.

Verwendete Symbole

$\mathrm{Sym}(n)$	Menge der symmetrischen $n \times n$-Matrizen
$\ell^2(V)$	Menge der ℓ^2-Funktionen auf der Knotenmenge V
$\mathrm{P}[A]$	Wahrscheinlichkeit des Ereignisses A
$\mathrm{d}f$	Simpliziale Korandabbildung, angewandt auf Funktion f
$\mathrm{Tr}(A)$	Spur der Matrix A
$\mathrm{Cay}(H; S)$	Cayleygraph der Gruppe H erzeugt von S
$\mathrm{supp}(X)$	Trägermenge der Zufallsvariable X
e^x	reelle Exponentialfunktion von x
$\exp(A)$	Matrix-Exponentialfunktion von A
$\|\vec{v}\|$	Euklidische Norm des Vektors $\vec{v}$
$\|f\|_2$	ℓ^2-Norm der Funktion f auf einer Knotenmenge
$\|A\|$	Operatornorm der Matrix A
$\|A\|_2$	2-Schattennorm der Matrix A
$\|A\|_p$	p-Schattennorm der Matrix A
∂F	Kantenrand der Knotenmenge F
$\langle f \mid g \rangle$	ℓ^2-Skalarprodukt der Funktionen f und g
J	Matrix, die in allen Einträgen Einsen hat
$V \setminus F$	Menge V ohne die Elemente der Menge F

2 Expandergraphen

Wigderson und Xiao derandomisieren den Beweis des Alon-Roichman-Theorems, welches besagt, dass sich zu einer beliebigen Gruppe ein gewisser Expandergraph konstruieren lässt, indem man zufällige Elemente auswählt. Den Expandergraphen gilt heute großes Forschungsinteresse. Sie haben viele Anwendungen, sowohl in der Informatik als auch in der Mathematik. Es handelt sich um Graphen, die wenige Kanten haben, aber trotzdem eine hohe Konnektivität besitzen. Sie tauchen bei der Konstruktion fehlerkorrigierender Codes [14] auf. Durch Irrfahrten auf Expandergraphen lassen sich Pseudozufallszahlen erzeugen [8] oder Zufälligkeitsextraktoren konstruieren [13] [11]. Solche Irrfahrten spielen auch eine Rolle beim Entrandomisieren von Produktgraphen [2]. In diesem Abschnitt wird eine Definition von Expandergraphen angegeben und es werden einige zentrale Sätze behandelt, die zum Verständnis des Themas beitragen.

2.1 Definitionen

Die Graphen, die in dieser Arbeit behandelt werden, sind ungerichtet und dürfen mehrfache Kanten aufweisen.

Definition 2.1. Ein *Graph* $G = (V, E)$ besteht aus einer Menge V von *Knoten* und einer Menge E von *Kanten*. Kanten können zwei verschiedene Knoten verbinden oder einen Knoten mit sich selbst. Eine solche Kante nennt man *Schleife*. Hat V endlich viele Elemente, nennt man G einen endlichen Graphen.

Im Folgenden werden wir uns nur mit endlichen Graphen beschäftigen.

Definition 2.2. Für einen Graphen G mit Knotenmenge V und Kantenmultimenge E definieren wir die *Adjazenzmatrix* A folgendermaßen. Die Einträge von A sind mit Paaren von Knoten $v, w \in V$ indiziert, und es soll

$$A_{vw} = \text{Anzahl der Kanten zwischen } v \text{ und } w \text{ für } v \neq w \text{ und}$$

$$A_{vv} = \text{Anzahl der Schleifen an } v$$

gelten.

Das heißt, die Einträge der Adjazenzmatrix geben die Anzahl der Kanten zwischen je zwei Knoten an. Die Einträge auf der Diagonalen zählen die Schleifen an dem entsprechenden Knoten. A ist symmetrisch und legt den Graphen G eindeutig fest.

Definition 2.3. Sei $d \geq 2$ eine natürliche Zahl. Wir nennen den Graphen G genau dann *d-regulär*, wenn für alle Knoten $w \in V$ gilt, dass $\sum_{v \in V} A_{vw} = d$.

Anschaulich bedeutet *d-regulär*, dass jeder Knoten d ausgehende Kanten hat, wobei Schleifen als eine Kante gezählt werden.

Sei G ein endlicher Graph mit n Knoten. Dann handelt es sich bei A um eine symmetrische $n \times n$-Matrix. A besitzt also n reelle Eigenwerte, wenn man deren Vielfachheiten berücksichtigt. Wir bezeichnen diese Eigenwerte als λ_i und betrachten sie der Größe nach geordnet:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Die Menge der Eigenwerte von A nennen wir das *Spektrum* des Graphen G.

Für einen beliebigen Graphen $G = (V, E)$ betrachten wir nun Funktionen $f\colon V \to \mathbb{C}$ auf der Knotenmenge von G. Wir definieren

$$\ell^2(V) := \left\{ f\colon V \to \mathbb{C} \;\middle|\; \sum_{v \in V} |f(v)|^2 < +\infty \right\}$$

zusammen mit der Norm

$$\|f\|_2 := \left(\sum_{v \in V} \overline{f(v)} f(v) \right)^{1/2}.$$

Solange V endlich ist, etwa $|V| = n$, sind offensichtlich alle Funktionen $f\colon V \to \mathbb{C}$ in $\ell^2(V)$. Wir können uns jede dieser Funktionen als Vektor im $\mathbb{C}^n$ vorstellen, und die Adjazenzmatrix auf sie anwenden. Zur besseren Übersicht sind hier die Knoten von G als $v_1, \ldots, v_n$ notiert:

$$Af = \begin{pmatrix} A_{v_1 v_1} & A_{v_1 v_2} & \cdots & A_{v_1 v_n} \\ \vdots & \vdots & & \vdots \\ A_{v_n v_1} & A_{v_n v_2} & \cdots & A_{v_n v_n} \end{pmatrix} \begin{pmatrix} f(v_1) \\ f(v_2) \\ \vdots \\ f(v_n) \end{pmatrix}$$

$$= \begin{pmatrix} A_{v_1 v_1} f(v_1) + A_{v_1 v_2} f(v_2) + \cdots + A_{v_1 v_n} f(v_n) \\ \vdots \\ A_{v_n v_1} f(v_1) + A_{v_n v_2} f(v_2) + \cdots + A_{v_n v_n} f(v_n) \end{pmatrix}.$$

Für die Kombination von Adjazenzmatrix und Funktion gilt also $(Af)(v_i) = \sum_{j=1}^{n} A_{v_i v_j} f(v_j)$, oder einfach

$$(Af)(v) = \sum_{w \in V} A_{vw} f(w).$$

2.2 Eigenschaften regulärer Graphen

Nun können wir einen ersten Satz über reguläre Graphen beweisen. Der Beweis stammt von Davidoff, Sarnak und Valette [6]:

Proposition 2.4. *Sei G ein endlicher d-regulärer Graph mit n Knoten. Dann gilt*

1. $\lambda_1 = d$;

2. $|\lambda_i| \leq d$ *für* $2 \leq i \leq n$;

3. *Die Vielfachheit des Eigenwerts* λ_1 *beträgt genau dann* 1, *wenn* G *zusammenhängend ist.*

Beweis. Wir beweisen zunächst (1.) und (2.). Dazu zeigen wir, dass d ein Eigenwert von G ist und für einen beliebigen Eigenwert λ gilt, dass $|\lambda| \leq d$ ist. Wir betrachten die konstante Funktion $f \equiv 1$ auf V. Es gilt

$$Af(v) = \sum_{w \in V} A_{vw} f(w) = \sum_{w \in V} A_{vw} = d = d \cdot f(v)$$

für alle $v \in V$. f ist also eine Eigenfunktion von A mit zugehörigem Eigenwert d. Sei nun λ ein beliebiger Eigenwert, mit zugehöriger Eigenfunktion f. Wir betrachten einen Knoten $v \in V$ mit

$$|f(v)| = \max_{w \in V} |f(w)|.$$

Wir gehen davon aus, dass $f(v) > 0$ ist. Ansonsten ersetzen wir f durch $-f$. Nun sehen wir, dass

$$f(v)|\lambda| = |f(v)\lambda| = |\sum_{w \in V} A_{vw} f(w)|$$

$$\leq \sum_{w \in V} A_{vw} |f(w)| \leq f(v) \sum_{w \in V} A_{vw} = f(v)d.$$

Das Ergebnis erhalten wir, indem wir beide Seiten der Gleichung durch $f(v)$ teilen.

Um (3.) zu beweisen, nehmen wir zunächst an, G sei zusammenhängend. Es sei f eine reellwertige Eigenfunktion zum Eigenwert d. Wir wollen nun zeigen, dass f konstant ist. Dazu wählen wir wie oben einen Knoten $v \in V$ mit $|f(v)| = \max_{w \in V} |f(w)|$. Wir schreiben $f(v)$ als

$$f(v) = \frac{(Af)(v)}{d} = \sum_{w \in V} \frac{A_{vw}}{d} f(w).$$

Dies ist eine Konvexkombination der $f(w)$, und diese sind betragsmäßig kleiner oder gleich $f(v)$. Daher muss $f(v) = f(w)$ für jedes $w \in V$ mit $A_{vw} \neq 0$ gelten, also für jeden Knoten, der zu v benachbart ist. Mit dem selben Argument zeigen wir, dass f den gleichen Wert auf den Nachbarn

dieser Knoten hat, und so weiter. Da G zusammenhängend ist, ist f also auf ganz G konstant.

Nun nehmen wir an, G sei nicht zusammenhängend. Sei v ein Knoten in G. Sei V_1 die Menge aller Knoten w, von denen ein Pfad nach v in G existiert. Sei $V_2 = G \setminus V_1$. Wir haben nun G in zwei d-reguläre Graphen mit den Knotenmengen V_1 und V_2 aufgeteilt, denn sobald ein Knoten $u \in V$ benachbart zu einem Knoten in V_1 ist, gilt $u \in V_1$. Dies gilt auch für V_2. Wir definieren nun die Funktionen

$$f_1(v) = \begin{cases} 1, & \text{falls } v \in V_1 \\ 0, & \text{falls } v \in V_2 \end{cases} \quad \text{und} \quad f_2(v) = \begin{cases} 0, & \text{falls } v \in V_1 \\ 1, & \text{falls } v \in V_2. \end{cases}$$

Bei f_1 und f_2 handelt es sich offensichtlich um linear unabhängige Eigenfunktionen von A zum Eigenwert d. Die Dimension des zugehörigen Eigenraumes ist also größer als 1. Damit ist auch die Vielfachheit des Eigenwerts d größer als 1, und wir haben Aussage (3.) gezeigt. □

2.3 Isoperimetrische Konstante und Eigenwerte

Sei G ein d-regulärer Graph mit Knotenmenge V und Kantenmenge E. Für eine Menge $F \subseteq V$ von Knoten definieren wir den *Rand* ∂F als die Menge der Kanten, die F mit ihrem Komplement $V \setminus F$ verbinden.

Anmerkung. *Es gilt offensichtlich, dass* $\partial F = \partial(V \setminus F)$ *ist.*

Definition 2.5. Die *isoperimetrische Konstante*, *Kantenexpansions-Konstante* oder *Cheeger-Konstante* $h(G)$ ist definiert als

$$h(G) = \inf \left\{ \frac{|\partial F|}{\min\{|F|, |V \setminus F|\}} \;\middle|\; F \subseteq V \text{ und } 0 < |F| < \infty \right\}.$$

Falls G nur endlich viele Knoten besitzt, lautet eine äquivalente Definition

$$h(G) = \min \left\{ \frac{|\partial F|}{|F|} \;\middle|\; F \subseteq V, 0 < |F| \leq \frac{|V|}{2} \right\}.$$

Die isoperimetrische Konstante gibt ein gewisses Maß dafür, wie gut die Knoten eines Graphen verbunden sind. Bei einem nicht zusammenhängenden Graphen beträgt sie Null, da die Zusammenhangskomponenten keine ausgehenden Kanten haben.

Definition 2.6. Einen Graphen G nennen wir (d, ε)-*Expander*, falls er d-regulär ist und $h(G) \geq \varepsilon$ gilt. Man findet auch die Bezeichnung (n, d, ε)-*Expander* für einen Expander mit n Knoten.

Definition 2.7. Sei d eine natürliche Zahl, (G_n) eine Folge d-regulärer Graphen mit $|G_n| \to \infty$ für $n \to \infty$. Wir nennen (G_n) eine *Folge von Expandern*, wenn ein $\varepsilon > 0$ existiert, so dass $h(G_i) \geq \varepsilon$ für alle i gilt. Man spricht auch oft von einer *Familie von Expandern*, auch wenn diese Bezeichnung ungenau ist.

Die Graphen einer Folge von Expandern müssen keine aufsteigende Folge bilden, das heißt G_i muss nicht in G_{i+1} enthalten sein.

Die isoperimetrische Konstante hängt eng mit dem betragsmäßig zweitgrößten Eigenwert des Graphen zusammen. Der folgende Satz zeigt diese wichtige Verbindung. Der Beweis stammt, bis auf leichte Änderungen, von Davidoff, Sarnak und Valette [6].

Satz 2.8. *Sei $G = (V, E)$ ein endlicher zusammenhängender d-regulärer Graph ohne Schleifen. Sei $\lambda = \lambda_2$ der betragsmäßig zweitgrößte Eigenwert von G. Dann gilt*

$$\frac{k - \lambda}{2} \leq h(G) \leq \sqrt{2k(k - \lambda)}.$$

Beweis. a) Wir fangen mit der ersten Ungleichung an. Dazu geben wir der Kantenmenge E eine beliebige Ausrichtung. Jede Kante $e \in E$ besitzt nun einen Ursprungsknoten e^- und einen Zielknoten e^+. Wir können nun die *simpliziale Korandabbildung* $\mathrm{d}\colon \ell^2(V) \to \ell^2(E)$ definieren, mit

$$\mathrm{d}f(e) = f(e^+) - f(e^-),$$

für $f \in \ell^2(V)$ und $e \in E$. Wir betrachten nun $\ell^2(V)$ und $\ell^2(E)$ zusammen mit dem hermiteschen Skalarprodukt

$$\langle f \mid g \rangle = \sum_{v \in V} \overline{f(v)} g(v)$$

bzw. dem entsprechend definierten Skalarprodukt für $\ell^2(E)$. Wir definieren den *adjungierten* (oder *konjungiert-transponierten*) Operator $\mathrm{d}^*\colon \ell^2(E) \to \ell^2(V)$, festgelegt durch

$$\langle \mathrm{d}f \mid g \rangle = \langle f \mid \mathrm{d}^* g \rangle$$

für alle $f \in \ell^2(V)$, $g \in \ell^2(E)$. Außerdem definieren wir eine Funktion $\delta\colon V \times E \to \{-1, 0, 1\}$ durch

$$\delta(v, e) = \begin{cases} 1, & \text{falls } v = e^+ \\ -1, & \text{falls } v = e^- \\ 0, & \text{sonst.} \end{cases}$$

Nun können wir d umformulieren als

$$\mathrm{d}f(e) = \sum_{v \in V} \delta(v, e) f(v)$$

und d^* als

$$\mathrm{d}^* g(v) = \sum_{e \in E} \delta(v, e) g(e),$$

wobei $v \in V$ und $g \in \ell^2(E)$. Diese Formulierungen sind korrekt, denn es gilt

$$\langle f \mid \mathrm{d}^* g \rangle = \sum_{v \in V} \overline{f(v)} \Big(\sum_{e \in E} \delta(v, e) g(e) \Big).$$

Vergleichen wir hier die Koeffizienten, erhalten wir für jede Kante $e \in E$ die Terme $\overline{f(e^+)} g(e)$ und $-\overline{f(e^-)} g(e)$. Umordnen ergibt

$$= \sum_{e \in E} \overline{(f(e^+) - f(e^-))} g(e) = \langle \mathrm{d}f \mid g \rangle .$$

Wir definieren nun den *kombinatorischen Laplace-Operator*

$$\Delta = \mathrm{d}^* \mathrm{d}\colon \ell^2(V) \to \ell^2(V).$$

Es gilt

$$\Delta = d \cdot \mathrm{Id} - A;$$

dies können wir uns folgendermaßen deutlich machen: Wir betrachten $\mathrm{d}^* \mathrm{d} f$ auf einem Knoten v, für eine beliebige Funktion $f \in \ell^2(V)$. An der Gleichung

$$\Delta f(v) = \mathrm{d}^* \mathrm{d} f(v) = \sum_{e \in E} \delta(v, e) \mathrm{d} f(e)$$

sehen wir, dass Δ nur auf den zu v benachbarten Kanten wirkt. Für jede ausgehende Kante e wird $f(e^+) - f(e^-)$ aufsummiert, hier gilt $e^+ = v$. Für

jede eingehende Kante wird $-(f(e^+) - f(e^-))$ summiert, mit $e^- = v$. Da G d-regulär ist, besitzt v insgesamt d benachbarte Kanten, wir erhalten also $d \cdot f(v)$ von den positiven Summanden. Die negativen Summanden haben jeweils den Funktionswert von f an den mit v verbundenen Knoten. Das Aufsummieren dieser Werte ist aber genau die Wirkung der Adjazenzmatrix, daher der Teil mit $-Af(v)$.

Wir sehen hier, dass Δ unabhängig von der Wahl der Ausrichtung des Graphen ist. Auf einer Basis aus Eigenfunktionen von A hat Δ die Form

$$\Delta = \begin{pmatrix} 0 & & & 0 \\ & d-\lambda_2 & & \\ & & \ddots & \\ 0 & & & d-\lambda_n \end{pmatrix}.$$

Der Eigenwert 0 gehört hier zum Raum der konstanten Funktionen auf V. Für eine Funktion f mit $\sum_{v\in V} f(v) = 0$ (d.h. f steht orthogonal auf den konstanten Funktionen in $\ell^2(V)$), gilt also

$$\|\mathrm{d}f\|_2^2 = \langle \mathrm{d}f \mid \mathrm{d}f \rangle = \langle \Delta f \mid f \rangle \geq (d-\lambda_2)\|f\|_2^2. \tag{1}$$

Um dies zu zeigen, zerlegen wir f in Eigenfunktionen $f_1, \ldots, f_n$ von A. Hierbei soll $f_1 \equiv 0$ der konstante Anteil von f sein. Damit erhalten wir

$$\langle \Delta f \mid f \rangle = (d-\lambda_2)f_2^2 + \cdots + (d-\lambda_n)f_n^2 \geq (d-\lambda_2)(f_2^2 + \cdots + f_n^2) = (k-\lambda_2)\|f\|_2^2.$$

Diese Ungleichung werden wir auf eine bestimmte Funktion f anwenden. Wir halten eine Teilmenge F von V fest und bestimmen

$$f(v) = \begin{cases} |V \setminus F|, & \text{falls } v \in F \\ -|F|, & \text{falls } v \in V \setminus F. \end{cases}$$

Es gilt $\sum_{v\in V} f(v) = 0$ und $\|f\|_2^2 = |F||V \setminus F|^2 + |V \setminus F||F|^2 = |F||V \setminus F||V|$. Außerdem gilt:

$$\mathrm{d}f(e) = \begin{cases} 0, & \text{falls } e^- \text{ und } e^+ \text{ beide in } F \text{ oder beide in } V \setminus F \text{ liegen;} \\ \pm|V|, & \text{falls } e \text{ Knoten in } F \text{ und } V \setminus F \text{ verbindet.} \end{cases}$$

Somit erhalten wir $\|\mathrm{d}f\|_2^2 = |V|^2|\partial F|$. Anwenden von Ungleichung (1) ergibt

$$|V|^2|\partial F| \leq (d - \lambda_2)|F||V \setminus F||V|$$

und Teilen durch $|F||V|^2$ ergibt

$$\frac{|\partial F|}{|F|} \leq (d - \lambda_2)\frac{|V \setminus F|}{|V|}.$$

Gehen wir nun von einer Teilmenge $F \subset V$ mit $|F| \leq \frac{|V|}{2}$ aus, erhalten wir $\frac{|\partial F|}{|F|} \geq \frac{k-\lambda_2}{2}$, und damit $h(G) \geq \frac{k-\lambda_2}{2}$.

b) Nun wenden wir uns der zweiten Ungleichung zu. Für eine nichtnegative Funktion f auf V sei

$$B_f = \sum_{e \in E} |f(e^+)^2 - f(e^-)^2|.$$

Dieser Beweis nutzt drei Eigenschaften der Zahl B_f, die wir im Folgenden aufzeigen werden. Es ist wichtig zu bemerken, dass durch den Absolutbetrag in der Definition von B_f die Orientierung des Graphen irrelevant ist. Die Werte, die f annimmt, bezeichnen wir mit $\beta_r > \beta_{r-1} > \cdots > \beta_1 > \beta_0$. Wir legen die Mengen

$$L_i = \{\, v \in V \mid f(v) \geq \beta_i \,\} \quad (i = 0, 1, \ldots, r)$$

fest. Die L_i stellen anschaulich eine Einteilung des Graphen mittels Höhenlinien dar.

1. Eigenschaft. $B_f = \sum_{i=1}^{r} |\partial L_i|(\beta_i^2 - \beta_{i-1}^2)$.

Wir betrachten zunächst die Menge E_f der Kanten $e \in E$ mit der Eigenschaft $f(e^+) \neq f(e^-)$. Es gilt offensichtlich $B_f = \sum_{e \in E_f} |f(e^+)^2 - f(e^-)^2|$. Wir werden die Betragsstriche leicht los, indem wir diese Summe als

$$B_f = \sum_{e \in E_f} (\beta_{i(e)}^2 - \beta_{j(e)}^2)$$

schreiben, mit $i(e) > j(e)$. Durch Einfügen einer Nullsumme erhalten wir

$$= \sum_{e \in E_f} (\beta_{i(e)}^2 - \beta_{i(e)-1}^2 + \beta_{i(e)-1}^2 - \cdots - \beta_{i(e)+1}^2 + \beta_{i(e)+1}^2 - \beta_{j(e)}^2)$$

$$= \sum_{e \in E_f} \sum_{l=j(e)+1}^{i(e)} (\beta_l^2 - \beta_{l-1}^2).$$

Wie oft tritt nun der Summand $(\beta_i^2 - \beta_{i-1}^2)$ auf? Er entsteht für jede Kante, die einen Knoten v mit $f(v) \geq \beta_i$ und einen Knoten w mit $f(w) < \beta_i$ verbindet, ein Mal. Diese Kanten bilden aber genau den Rand von L_i, was unsere erste Aussage beweist.

2. Eigenschaft. $B_f \leq \sqrt{2d} \cdot \|\mathrm{d}f\|_2 \|f\|_2$.

Es gilt

$$\begin{aligned} B_f &= \sum_{e \in E} |f(e^+) + f(e^-)| \cdot |f(e^+) - f(e^-)| \\ &\leq \left[\sum_{e \in E} (f(e^+) + f(e^-))^2\right]^{1/2} \left[\sum_{e \in E} (f(e^+) - f(e^-))^2\right]^{1/2} \end{aligned}$$

(Cauchy-Schwarz-Ungleichung)

$$\leq \sqrt{2}\left[\sum_{e \in E} (f(e^+)^2 + f(e^-)^2)\right]^{1/2} \|\mathrm{d}f\|_2$$

und, da $(a+b)^2 \leq 2(a^2+b^2)$

$$= \sqrt{2d} \cdot \left[\sum_{v \in V} f(v)^2\right]^{1/2} \|\mathrm{d}f\|_2 = \sqrt{2d} \|f\|_2 \|\mathrm{d}f\|_2.$$

3. Eigenschaft. Für ein f mit $|\operatorname{supp} f| \leq \frac{|V|}{2}$ gilt $B_f \geq h(G)\|f\|_2^2$.

Wegen $|\operatorname{supp} f| \leq \frac{|V|}{2}$ gilt, dass $\beta_0 = 0$ und $|L_i| \leq \frac{|V|}{2}$ für $i = 1, \ldots, r$. Daraus folgt $|\partial L_i| \geq h(G)|L_i|$. Aus Eigenschaft **1.** folgt nun

$$\begin{aligned} B_f &\geq h(G) \sum_{i=1}^{r} |L_i|(\beta_i^2 - \beta_{i-1}^2) \\ &= h(G)\left[|L_r|\beta_r^2 + (|L_{r-1}| - |L_r|)\beta_{r-1}^2 + \cdots + (|L_1| - |L_2|)\beta_1^2\right] \\ &= h(G)\left[|L_r|\beta_r^2 + \sum_{i=1}^{r-1} |L_i - L_{i+1}|\beta_i^2\right]. \end{aligned}$$

Nun ist $L_i \setminus L_{i+1}$ aber genau die Menge der Knoten, auf denen f den Wert β_i annimmt, das heißt, der Term in der eckigen Klammer beträgt $\|f\|_2^2$.

Schluss. Wir wenden nun die gefundenen Eigenschaften auf eine sorgfältig ausgewählte Funktion f an. Sei g eine reellwertige Eigenfunktion des Operators Δ, zum Eigenwert $d - \lambda_2$. Wir definieren die Menge $V^+ := \{ v \in V \mid g(v) > 0 \}$. Es gilt $V^+ \neq \emptyset$, da $\sum_{v \in V} g(v) = 0$ und $g \neq 0$. Wir dürfen $|V^+| \leq \frac{|V|}{2}$ annehmen, dafür ersetzen wir wenn nötig g durch $-g$. Wir definieren die Funktion f als $f = \max\{ g, 0 \}$. Für ein $v \in V^+$ gilt, da $g \leq 0$ auf $V \setminus V^+$ ist,

$$\begin{aligned}(\Delta f)(v) &= d \cdot f(v) - \sum_{w \in V} A_{vw} f(w) = d \cdot g(v) - \sum_{w \in V^+} A_{vw} g(w) \\ &\leq d \cdot g(v) - \sum_{w \in V} A_{vw} g(w) = (\Delta g)(v) = (d - \lambda_2) g(v).\end{aligned}$$

Mit Hilfe dieser Abschätzung erhalten wir

$$\|\mathrm{d}f\|_2^2 = \langle \Delta f \mid f \rangle = \sum_{v \in V^+} (\Delta f)(v) g(v) \leq (d - \lambda_2) \sum_{v \in V^+} g(v)^2 \leq (d - \lambda_2) \|f\|_2^2.$$

Mit den Eigenschaften **2.** und **3.** folgt

$$h(G) \|f\|_2^2 \leq B_f \leq \sqrt{2d} \cdot \|\mathrm{d}f\|_2 \|f\|_2 \leq \sqrt{2d \cdot (d - \lambda_2)} \|f\|_2^2,$$

die Behauptung folgt durch Herauskürzen von $\|f\|_2^2$. □

Aus den Sätzen 2.4 und 2.8 folgt sofort:

Korollar 2.9. *Sei $(G_m)_{m \geq 0}$ eine Folge endlicher, zusammenhängender, d-regulärer Graphen ohne Schleifen, so dass $|V_m| \to +\infty$ für $m \to +\infty$. Die Folge $(G_m)_{m \geq 1}$ ist genau dann eine Folge von Expandern, wenn ein $\varepsilon > 0$ existiert mit $d - \lambda_2(G_m) \geq \varepsilon$ für jedes $m \geq 1$.*

3 Matrixwertige Chernoffabschätzung

Dieses Kapitel befasst sich mit der Hinleitung zu einer Abschätzung, die es erlauben wird, Aussagen über Summen matrixwertiger Zufallsvariablen zu machen. Die Wahrscheinlichkeit, dass eine solche Summe eine vorgegebene Schranke übersteigt, werden wir am Ende des Kapitels nach oben hin abschätzen können.

3.1 Definitionen

Dieser Abschnitt enthält grundlegende Definitionen der in den Kapiteln 3 und 4 verwendeten Begriffe.

Im Folgenden soll $\mathrm{Sym}(d)$ die Menge der reellwertigen symmetrischen $d \times d$-Matrizen bezeichnen. Wir schreiben I oder I_d für die Identitätsmatrix in $\mathrm{Sym}(d)$. Für $A \in \mathrm{Sym}(d)$ seien $\lambda_1(A) \geq \cdots \geq \lambda_d(A)$ die nach Größe geordneten Eigenwerte von A. Wir nutzen die *Operatornorm* für Matrizen

$$\|A\| = \max_v \frac{\|Av\|}{\|v\|} = \max_i |\lambda_i(A)|$$

und die *Spur* einer Matrix,

$$\mathrm{Tr}(A) = \sum_{i=1}^{d} \lambda_i(A).$$

Außerdem nutzen wir die Tatsache, dass mittels einer Orthonormalbasis $(v_1, \ldots, v_d)$ des $\mathbb{R}^d$ die Spur als

$$\mathrm{Tr}(A) = \sum_{i=1}^{d} \langle v_i, Av_i \rangle$$

ausgedrückt werden kann, wobei $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt im $\mathbb{R}^d$ bezeichnet.

Eine Matrix $A \in \mathrm{Sym}(d)$ bezeichnen wir als *positiv semidefinit* (p.s.d.), wenn alle ihre Eigenwerte größer oder gleich null sind. A ist genau dann p.s.d., wenn $\langle v, Av \rangle \geq 0$ für alle $v \in \mathbb{R}^d$ gilt. Im Folgenden soll die Schreibweise $A \geq 0$ bedeuten, dass A p.s.d. ist. Dies erlaubt uns, auf den symmetrischen Matrizen eine Ordnung zu definieren. Wir sagen, dass $A \leq B$ genau dann, wenn $B - A \geq 0$. Für zwei Matrizen $A \leq B$ soll das Intervall $[A, B]$ die Menge der symmetrischen Matrizen C mit $A \leq C$ und $C \leq B$ bezeichnen.

Die *Exponentialfunktion für Matrizen*, definiert durch die Reihe

$$\exp(A) = \sum_{l=0}^{\infty} \frac{A^l}{l!},$$

konvergiert für alle Matrizen und es gilt $\exp(A) \geq 0$, wenn A symmetrisch ist. Eine Basis aus Eigenvektoren von $A \in \mathrm{Sym}(d)$ ist ebenfalls eine Basis aus

Eigenvektoren von $\exp(A)$ und es gilt $\lambda_i(\exp(A)) = \mathrm{e}^{\lambda_i(A)}$ für alle $1 \leq i \leq d$. Die matrixwertigen Zufallsvariablen, mit denen wir uns beschäftigen werden, sehen folgendermaßen aus:

Sei $f\colon [n] \to [-I_d, I_d]$, wobei $[n] = \{1, \ldots, n\}$. Sei X eine beliebige Verteilung über $[n]$. Die Variable $f(X)$ verhält sich nun ähnlich einer begrenzten diskreten Zufallsvariablen mit Werten in den rellen Zahlen, die wir uns als Funktionen $f\colon [n] \to [-1, 1]$ vorstellen können. Auf natürliche Art und Weise erhalten wir auch den Erwartungswert $\mathbb{E}[f(X)] = \sum_{i=1}^{n} \mathrm{P}[X = i] \cdot f(i)$. Es ist leicht zu sehen, dass Spur und Erwartungswert kommutieren. Es gilt also $\mathbb{E}[\mathrm{Tr}(f(X))] = \mathrm{Tr}(\mathbb{E}[f(X)])$. Als *Träger* von X, $\mathrm{supp}(X)$, bezeichnen wir alle Werte, die X mit positiver Wahrscheinlichkeit annehmen kann. Wir sagen, dass eine Aussage über eine Zufallsvariable X *immer* gilt, wenn sie für jedes Element von $\mathrm{supp}(X)$ gilt.

Wir werden an einigen Stellen die Landausche $\mathcal{O}$-Notation verwenden. Für zwei Funktionen $f(x)$ und $g(x)$ und eine vorgegebene Stelle a bedeutet $f(x) = \mathcal{O}(g(x))$, dass

$$\limsup_{x \to a} \frac{f(x)}{g(x)} < \infty$$

ist. Häufig werden mit der $\mathcal{O}$-Schreibweise statt Funktionen Folgen miteinander verglichen, oder sie dient zur Angabe der Größenordnung eines auftauchenden Wertes.

3.2 Unterstützende Sätze

Die Sätze, die in diesem Abschnitt aufgeführt sind, dienen dem Beweis des Haupttheorems im nächsten Abschnitt.

Lemma 3.1. *Seien A und B Matrizen in* $\mathrm{Sym}(d)$ *und gelte $B \geq 0$, dann ist* $\mathrm{Tr}(AB) \leq \|A\| \cdot \mathrm{Tr}(B)$.

Beweis. Seien $v_1, \ldots, v_d$ Eigenvektoren von A, die eine orthonormale Basis bilden. Ihre Eigenwerte seien $\lambda_i = \lambda_i(A)$. Es gilt

$$\mathrm{Tr}(AB) = \sum_{i=1}^{d} \langle v_i, ABv_i \rangle = \sum_{i=1}^{d} \lambda_i \langle v_i, Bv_i \rangle.$$

Da B positiv semidefinit ist, gilt $\langle v_i, Bv_i\rangle \geq 0$, also erhalten wir

$$\mathrm{Tr}(AB) \leq \sum_{i=1}^{d} \max_j \lambda_j \langle v_i, Bv_i\rangle \leq \|A\| \cdot \mathrm{Tr}(B).$$

□

Satz 3.2 (Golden-Thompson-Ungleichung)**.** *Seien A und B hermitesche $n \times n$ Matrizen. Dann gilt*

$$\mathrm{Tr}(\exp(A+B)) \leq \mathrm{Tr}(\exp(A) \cdot \exp(B)) \ .$$

Dieser Beweis stammt aus Terence Taos Blog [15]. Er nutzt eine nichtkommutative Hölder-Ungleichung für Matrizen, diese beweisen wir im folgenden Lemma.

Definition 3.3 (p-Schattennorm)**.** Für eine gerade natürliche Zahl $p = 2, 4, 6, \ldots$ definieren wir die *p-Schattennorm* $\|A\|_p$ einer beliebigen $n \times n$-Matrix A durch

$$\|A\|_p := ((\mathrm{Tr}(AA^*))^{p/2})^{1/p}.$$

Lemma 3.4. *Seien A_1, A_2 beliebige $n \times n$ Matrizen und $p = 2, 4, 8, \ldots$ eine gerade Zweierpotenz. Dann gilt*

$$|\mathrm{Tr}(A_1 A_2 \ldots A_p)| \leq \|A_1\|_p \|A_2\|_p \cdots \|A_p\|_p. \tag{2}$$

Beweis. Die 2-Schattennorm

$$\|A\|_2 = (\mathrm{Tr}(AA^*))^{1/2}$$

ist die Hilbertraumnorm, die von dem Frobenius-Skalarprodukt

$$\langle A, B\rangle := \mathrm{Tr}(AB^*)$$

induziert wird. Hierbei handelt es sich um eine positiv definite hermitesche Sesquilinearform, deshalb können wir die Cauchy-Schwarz-Ungleichung anwenden und erhalten

$$|\mathrm{Tr}(A_1 A_2^*)| \leq \|A_1\|_2 \|A_2\|_2.$$

Da $\|A_2\|_2 = \|A_2^*\|_2$ gilt, folgt insbesondere

$$|\mathrm{Tr}(A_1 A_2)| \leq \|A_1\|_2 \|A_2\|_2.$$

Wir werden dies nun durch Induktion über p verallgemeinern, um (2) zu erhalten. Den Fall $p = 2$ haben wir oben bereits gezeigt. Sei $p \geq 4$ eine Zweierpotenz. Wir teilen $A_1 \dots A_p$ in $p/2$ Paare auf und erhalten mit Hilfe der Induktionsbehauptung

$$|\mathrm{Tr}(A_1 A_2 \cdots A_p)| \leq \|A_1 A_2\|_{p/2} \|A_3 A_4\|_{p/2} \cdots \|A_{p-1} A_p\|_{p/2}. \tag{3}$$

Weiterhin betrachten wir

$$\|A_1 A_2\|_{p/2}^{p/2} = \mathrm{Tr}(\underbrace{A_1 A_2 A_2^* A_1^* \cdots A_1 A_2 A_2^* A_1^*}_{p/4\text{-mal}}).$$

Wir nutzen die zyklische Vertauschungseigenschaft der Spur, um den Faktor A_1^* von ganz rechts nach links zu bewegen. Nun können wir wieder die Induktionshypothese anwenden und erhalten

$$\|A_1 A_2\|_{p/2}^{p/2} \leq \|A_1^* A_1\|_{p/2} \|A_2 A_2^*\|_{p/2} \cdots \|A_1^* A_1\|_{p/2} \|A_2 A_2^*\|_{p/2}.$$

Es gilt aber $\|A_1^* A_1\|_{p/2} = \|A_1\|_p^2$ und $\|A_2 A_2^*\|_{p/2} = \|A_2\|_p^2$, wieder wegen der Eigenschaft der Spur. Wir erhalten also

$$\|A_1 A_2\|_{p/2} \leq \|A_1\|_p \|A_2\|_p$$

sowie ähnliche Aussagen für $\|A_3 A_4\|_{p/2}$ usw. Diese setzen wir in (3) ein und erhalten (2). □

Beweis der Golden-Thompson-Ungleichung. Wir gehen hier von (2) aus und setzen $A_1 = \cdots = A_p = AB$ für hermitesche Matrizen A und B. Damit erhalten wir

$$\mathrm{Tr}((AB)^p) \leq \|AB\|_p^p.$$

Da $A = A^*$ gilt, erhalten wir mit zyklischer Vertauschung

$$\mathrm{Tr}((AB)^p) \leq \mathrm{Tr}((A^2 B^2)^{p/2})$$

für $p = 2, 4, 8, \dots$ Wir wenden diese beiden Schritte weiter an (als nächstes

würden wir $\mathrm{Tr}((A^2B^2)^{p/2}) \leq \|A^2B^2\|_{p/2}^{p/2}$ abschätzen) und erhalten schließlich

$$\mathrm{Tr}((AB)^p) \leq \mathrm{Tr}(A^pB^p). \tag{4}$$

Hier ersetzen wir nun A und B durch $\exp(A/p)$ bzw. $\exp(B/p)$. Dies ergibt

$$\mathrm{Tr}((\exp(A/p)\exp(B/p))^p) \leq \mathrm{Tr}(\exp(A)\exp(B)). \tag{5}$$

Wir betrachten diese Ungleichung genauer beim Übergang $p \to \infty$: Es gilt $\exp(A/p) = I + A/p + \mathcal{O}(1/p^2)$ und $\exp(B/p) = I + B/p + \mathcal{O}(1/p^2)$, also ist $\exp(A/p)\exp(B/p) = \exp((A+B)/p + \mathcal{O}(1/p^2))$. Die linke Seite der Ungleichung (5) beträgt also $\mathrm{Tr}(\exp(A+B+\mathcal{O}(1/p)))$. Wenn wir nun den Limes $p \to \infty$ bilden, erhalten wir die Golden-Thompson-Ungleichung.

□

Lemma 3.5. *Sei A eine Matrix in $\mathrm{Sym}(d)$, $f\colon [n] \to \mathrm{Sym}(d)$, X eine Zufallsvariable mit Werten in $[n]$. Dann gilt*

$$\mathrm{Tr}(\mathbb{E}_X[\exp(A + f(X))]) \leq \|\mathbb{E}[\exp(f(X))]\| \cdot \mathrm{Tr}(\exp(A)).$$

Beweis. Wir beginnen mit der linken Seite der Ungleichung und vertauschen Spur und Erwartungswert. Dann gilt

$$\mathrm{Tr}(\mathbb{E}[\exp(A + f(X))]) = \mathbb{E}[\mathrm{Tr}(\exp(A + f(X)))].$$

Mit der Golden-Thompson-Ungleichung erhalten wir

$$\leq \mathbb{E}[\mathrm{Tr}(\exp(f(X))\exp(A))],$$

nach Vertauschen von Spur und Erwartungswert

$$\leq \mathrm{Tr}(\mathbb{E}[\exp(f(X))]\exp(A)).$$

Hierauf wenden wir Lemma 3.1 an und erhalten

$$\leq \|\mathbb{E}[\exp(f(X))]\| \cdot \mathrm{Tr}(\exp(A)).$$

□

Wir nutzen hier eine abgewandelte Version der gewöhnlichen Markov-Ungleichung. Zur Erinnerung:

Satz 3.6 (Markov-Ungleichung)**.** *Sei X eine reellwertige Zufallsvariable, a eine positive reelle Konstante und $h\colon \mathbb{R} \to [0, \infty[$ eine monoton wachsende Funktion. Dann gilt*

$$\mathrm{P}\left[X \geq a\right] \leq \frac{\mathbb{E}[h(X)]}{h(a)}.$$

Der Beweis ist wohlbekannt. Die matrixwertige Markov-Ungleichung ähnelt der skalaren sichtlich; für unsere Zwecke reicht es, wenn X nur abzählbar viele Werte annimmt:

Satz 3.7 (Markov-Ungleichung für Matrizen)**.** *Für $\gamma > 0$, eine Abbildung $g\colon [n] \to \mathrm{Sym}(d)$ mit $g(x) \geq 0$ für jedes $x \in [n]$ und eine Zufallsvariable X mit Werten in $[n]$ gilt*

$$\mathrm{P}[g(X) \not\leq \gamma I] \leq \frac{1}{\gamma} \operatorname{Tr}(\mathbb{E}[g(X)]).$$

Beweis. Die Aussage $g(X) \not\leq \gamma I$ bedeutet, dass mindestens ein Eigenwert von g(X) größer als γ sein muss. Also gilt

$$\mathrm{P}[g(X) \not\leq \gamma I] = \mathrm{P}[\|g(X)\| > \gamma].$$

Hierauf wenden wir die gewöhnliche Markov-Ungleichung an und erhalten

$$\mathrm{P}[\|g(X)\| > \gamma] \leq \frac{1}{\gamma}\mathbb{E}[\|g(X)\|].$$

Da $g(X)$ immer positiv semidefinit ist, gilt immer $\|g(X)\| \leq \operatorname{Tr}(g(X))$ (größter Eigenwert gegenüber Summe aller Eigenwerte). Wir erhalten

$$\frac{1}{\gamma}\mathbb{E}[\|g(X)\|] \leq \frac{1}{\gamma}\mathbb{E}[\operatorname{Tr}(g(X))] = \frac{1}{\gamma} \operatorname{Tr}(\mathbb{E}[g(X)]).$$

□

3.3 Haupttheorem

Die bekannte Chernoff-Ungleichung liefert eine obere Schranke für die Wahrscheinlichkeit, dass die Summe unabhängiger Zufallsvariablen von einem vorgegebenen Wert abweicht. Die Verallgemeinerung der Chernoff-Abschätzung

auf matrixwertige Zufallsvariablen stammt von Ahlswede und Winter [1]. Der Beweis stammt von Wigderson und Xiao [16]. Wir werden hier auf ihr Theorem eingehen. Bei den Sätzen 3.9 und 3.11 handelt es sich um Ahlswedes und Winters Abschätzungen. Sie werden aus dem Haupttheorem 3.8 abgeleitet. Dieses wird später auch genutzt, um pessimistische Schätzer für die Derandomisierung zu erhalten.

Satz 3.8. *Sei $f\colon [n] \to [-I_d, I_d]$ und seien $X_1, \ldots, X_k$ beliebige unabhängige Zufallsvariablen, die über $[n]$ verteilt sind. Dann gilt für alle $\gamma \in \mathbb{R}$*

$$\mathrm{P}\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] \leq d\,\mathrm{e}^{-t\gamma k} \prod_{j=1}^{k} \|\mathbb{E}[\exp(tf(X_j))]\|.$$

Beweis. Wir beginnen, indem wir das Ereignis auf der linken Seite umformen. Beide Seiten der Ungleichung in der eckigen Klammer werden mit kt multipliziert, wobei t eine beliebige aber feste positive Zahl ist. Dann wenden wir auf beiden Seiten der Ungleichung die Exponentialfunktion für Matrizen an und erhalten

$$\mathrm{P}\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] = \mathrm{P}\left[\exp(t\sum_{j=1}^{k} f(X_j)) \not\leq \mathrm{e}^{t\gamma k}\, I\right].$$

Hier ist wichtig zu bemerken, dass es sich bei den obigen Ungleichungen nicht um die „gewöhnliche“ Ordnung auf den reellen Zahlen handelt. Wir sollten also sicherstellen, dass die Aussage auch für unsere Matrizen-Ordung gilt. Sie tut es, denn für eine Matrix $A \in \mathrm{Sym}(d)$ und $\alpha \in \mathbb{R}$ bedeutet $A \not\leq \alpha I$, dass A einen Eigenwert größer als α besitzt. Dies ist äquivalent dazu, dass $\exp(A)$ einen Eigenwert größer e^{α} hat, wegen der Eigenwert-Transformation von exp.

Mit 3.7 erhalten wir

$$\mathrm{P}\left[\exp(t\sum_{j=1}^{k} f(x_j)) \not\leq \mathrm{e}^{t\gamma k}\, I\right] \leq \mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\mathbb{E}\left[\exp\left(t\sum_{j=1}^{k} f(X_j)\right)\right]\right).$$

Wir halten das folgende Zwischenergebnis fest:

$$\mathrm{P}\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] \leq \mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\mathbb{E}\left[\exp\left(t\sum_{j=1}^{k} f(X_j)\right)\right]\right). \tag{6}$$

Wir wenden nun Lemma 3.5 wiederholt an, um die obige Summe durch ein Produkt abzuschätzen. Dazu spalten wir zunächst den Erwartungswert in unabhängige Komponenten auf,

$$= \mathrm{e}^{-t\gamma k}\, \mathbb{E}_{X_1,\dots,X_{k-1}} \left[\mathrm{Tr}\left(\mathbb{E}_{X_k} \left[\exp\left(t \sum_{j=1}^{k-1} f(X_j) + t f(X_k) \right) \right] \right) \right],$$

um nach Anwendung von Lemma 3.5

$$\leq \mathrm{e}^{-t\gamma k}\, \mathbb{E}_{X_1,\dots,X_{k-1}} \left[\|\mathbb{E}_{X_k}[\exp(t f(X_k))]\| \cdot \mathrm{Tr}\left(\exp\left(t \sum_{j=1}^{k-1} f(X_j) \right) \right) \right]$$

zu erhalten. Der Inhalt der Norm hängt nur von t und X_k ab, also gilt nach Vertauschen von Spur und Erwartungswert

$$= \mathrm{e}^{-t\gamma k} \|\mathbb{E}[\exp(t f(X_k))]\| \cdot \mathrm{Tr} \left(\mathbb{E}_{X_1,\dots,X_{k-1}} \left[\exp\left(t \sum_{j=1}^{k-1} f(X_j) \right) \right] \right).$$

Diese Schritte wiederholen wir insgesamt k Mal, um die gesamte Summe zu faktorisieren, und erhalten

$$\leq \mathrm{e}^{-t\gamma k} \prod_{j=1}^{k} \|\mathbb{E}[\exp(t f(X_j))]\| \cdot \mathrm{Tr}(I).$$

Die $\mathrm{Tr}(I)$ am Ende ergibt sich wegen $\exp(0_d) = I_d$. Da $\mathrm{Tr}(I_d) = d$ ist, gilt

$$= d\, \mathrm{e}^{-t\gamma k} \prod_{j=1}^{k} \|\mathbb{E}[\exp(t f(X_j))]\|.$$

□

3.4 Wichtige Folgerungen

Die Folgerungen aus der Ahlswede-Winter-Ungleichung in diesem Abschnitt stellen verschiedene obere Schranken dar, mit denen sich Abweichungen vom Erwartungswert eingrenzen lassen. Sie lassen sich besser anwenden als das Haupttheorem.

Korollar 3.9. *Sei $f\colon [n] \to [-I_d, I_d]$. Sei X über $[n]$ verteilt, mit Erwar-*

tungswert $\mathbb{E}_X[f(X)] = 0$, *und seien* $X_1, \ldots, X_k$ *unabhängig und verteilt wie* X. *Dann gilt für alle* $1 > \gamma > 0$ *:*

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq \gamma I\right] \leq d\,\mathrm{e}^{-\gamma^2 k/4}.$$

Anmerkung. *Die Abschätzung*

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\geq -\gamma I\right] \leq d\,\mathrm{e}^{-\gamma^2 k/4}$$

erhalten wir sofort, indem wir statt f *die Funktion* $-f$ *betrachten.*

Beweis. Wir wenden zunächst Theorem 3.8 an. Da die X_i unabhängig und gleich verteilt sind, erhalten wir

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq \gamma I\right] \leq d\,\mathrm{e}^{-t\gamma k}\|\mathbb{E}[\exp(tf(X))]\|^k. \tag{7}$$

Nun schätzen wir die rechte Seite mit folgender Behauptung ab:

Behauptung 3.10. *Es gilt*

$$\|\mathbb{E}[\exp(tf(X))]\| \leq 1 + t^2 \text{ für } t \leq 1/2.$$

Beweis. Wir betrachten hier die Taylorentwicklung von exp:

$$\|\mathbb{E}[\exp(tf(X))]\| = \|\mathbb{E}[I + tf(X) + (tf(X))^2/2 + \cdots]\|$$

$$= \|I + t\mathbb{E}[f(X)] + \mathbb{E}[(tf(X))^2/2 + \cdots]\|. \tag{8}$$

Nun nutzen wir $\mathbb{E}[f(X)] = 0$, die Dreiecksungleichung und $\|f(X)\| \leq 1$ und erhalten

$$\|\mathbb{E}[\exp(tf(X))]\| \leq 1 + \sum_{j=2}^{\infty} \frac{t^j}{j!}.$$

Wegen $t \leq 1/2$ gilt

$$\leq 1 + t^2,$$

was die Behauptung beweist. □

Wir werden im Folgenden $t = \gamma/2$ wählen. Wenden wir Behauptung 3.10

auf Gleichung (7) an, erhalten wir

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq \gamma I\right] \leq d\,\mathrm{e}^{-t\gamma k}(1+t^2)^k.$$

Da $1 + x \leq \mathrm{e}^x$ für alle $x \in \mathbb{R}$, gilt $(1 + t^2) \leq \mathrm{e}^{t^2}$, also

$$\leq d\,\mathrm{e}^{-t\gamma k + t^2 k},$$

und wegen $t = \gamma/2$ schließlich

$$\leq d\,\mathrm{e}^{-\gamma^2 k/4}\,.$$

Damit haben wir Korollar 3.9 gezeigt. □

Korollar 3.11. *Sei $f\colon [n] \to [0, I_d]$. Sei X über $[n]$ verteilt, mit $M = \mathbb{E}_X[f(X)] \leq \mu I$ für ein $\mu \in (0, 1)$. Seien $X_1, \ldots, X_k$ unabhängig und verteilt wie X. Dann gilt für alle $0 \leq \gamma \leq \frac{1}{2}$*

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq (1-\gamma)\mu I\right] \leq d\,\mathrm{e}^{-\gamma^2\mu k/(2\ln 2)}\,.$$

Beweis. Wir nehmen ohne Beschränkung der Allgemeinheit an, dass $M = \mu I$ gilt. Umstellen der linken Seite ergibt

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq (1-\gamma)\mu I\right] = \mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} (I - f(X_i)) \not\leq (1-(1-\gamma)\mu) I\right].$$

Hierauf können wir Satz 3.8 anwenden und erhalten

$$\leq d\,\mathrm{e}^{-t(1-(1-\gamma)\mu)k} \|\mathbb{E}[\exp(t(I - F(X)))]\|^k$$

$$= d \cdot \|\mathbb{E}[\exp(-tf(X))]\,\mathrm{e}^{t(1-\gamma)\mu}\|^k.$$

Diese Größe wird weiter abgeschätzt durch die folgende Aussage aus dem Beweis von Satz 19 bei [1]. Sie wird hier ohne Beweis übernommen.

Behauptung 3.12. *Sei $t = \log\left(\frac{1-(1-\gamma)\mu}{1-\mu}\frac{1}{(1-\gamma)}\right)$, dann gilt*

$$\|\mathbb{E}[\exp(-tf(X))]\,\mathrm{e}^{t(1-\gamma)\mu}\| \leq \mathrm{e}^{-\gamma^2\mu/(2\ln n)}\,.$$

Anwenden der Behauptung schließt den Beweis ab. □

4 Probabilistische Beweise und Derandomisierung

4.1 Übersicht

Nehmen wir an, wir möchten die Existenz eines mathematischen Objekts (Zahl, Graph, etc.) beweisen, das eine bestimmte Eigenschaft hat. Dann genügt es zu zeigen, dass ein zufällig ausgewähltes Objekt diese Eigenschaft mit positiver Wahrscheinlichkeit besitzt. Dies ist die Idee hinter randomisierten Beweisen oder probabilistischen Algorithmen. Beispiele solcher Beweisführung findet man etwa in [4]. Es kann sehr schwierig sein, randomisierte Beweise zu entrandomisieren, das heißt das Objekt mit der gewünschten Eigenschaft explizit darzustellen. Dazu gibt es verschiedene Vorgehensweisen, eine davon ist die Methode der pessimistischen Schätzer (engl. method of pessimistic estimators). Sie wurde von Raghavan entwickelt [12].

4.2 Methode der pessimistischen Schätzer

Sei $\vec{X}$ eine Zufallsvariable mit $\operatorname{supp}(\vec{X}) \subseteq [n]^k$. Wir schreiben $\vec{X}$ als $\vec{X} = (X_1, \ldots, X_k)$, mit $X_i \in [n]$ für $i = 1, \ldots, k$. Sei $\sigma(\vec{X})$ ein Ereignis mit $\sigma\colon \operatorname{supp}(\vec{X}) \to \{0,1\}$ und $\mathrm{P}[\sigma(\vec{X}) = 1] > 0$. Das heißt, σ tritt für einen oder mehrere Ausgänge von $\vec{X}$ mit positiver Wahrscheinlichkeit ein. Die Aufgabe lautet nun, einen Vektor $\vec{x} \in \operatorname{supp}(\vec{X})$ zu finden, so dass $\sigma(\vec{x}) = 1$ gilt. Dazu suchen wir geeignete Werte $x_1 \in X_1$, $x_2 \in X_2, \ldots$ und erhalten das gewünschte $\vec{x} \in \vec{X}$.

Skizze. *Es gilt*

$$\mathrm{P}[\sigma(\vec{X}) = 0] = \mathbb{E}_{X_1}\Big[\mathrm{P}\big[\sigma(\vec{X}) = 0 \;\big|\; X_1\big]\Big].$$

Das bedeutet, dass ein Wert x_1 für X_1 existiert, so dass

$$\mathrm{P}\big[\sigma(\vec{X}) = 0 \;\big|\; X_1 = x_1\big] \leq \mathbb{E}_{X_1}\Big[\mathrm{P}\big[\sigma(\vec{X}) = 0 \;\big|\; X_1\big]\Big].$$

Ein solches x_1 existiert immer, da einer der Werte kleiner oder gleich dem Durchschnitt aller Werte sein muss (sonst wäre der Durchschnitt größer). Setzen wir also $X_1 = x_1$. Die Werte für $X_2, \ldots, X_k$ erhalten wir auf die

selbe Weise und vervollständigen so unser $\vec{x}$. Nun gilt folgendes:

$$\mathrm{P}\big[\sigma(\vec{x}=0)\big] \leq \mathrm{P}\big[\sigma(\vec{X}=0)\big] < 1.$$

Bei $\vec{x}$ handelt es sich um einen festen Vektor, es folgt also $\sigma(\vec{x}) = 1$.

Die Schwierigkeit dieses Algorithmus liegt in der Berechnung der bedingten Wahrscheinlichkeiten

$$\mathrm{P}_{X_{i+1},\ldots,X_k}\big[\sigma(\vec{X}) = 0 \;\big|\; X_1 = x_1, \ldots, X_i = x_i\big]$$

für alle $1 \leq i \leq k$ und $x_1, \ldots, x_i \in [n]$. Hier kommen die pessimistischen Schätzer ins Spiel. Sie geben uns obere Grenzen für diese Wahrscheinlichkeiten an.

Definition 4.1. Sei $\sigma\colon [n]^k \to \{0,1\}$ ein Ereignis auf einer Zufallsvariable $\vec{X}$ mit Werten in $[n]^k$, mit $\mathrm{P}\big[\sigma(\vec{X}) = 1\big] > 0$. Als *pessimistische Schätzer für σ* bezeichnen wir Funktionen $\phi_0, \ldots, \phi_k$ mit $\phi_i\colon [n]^i \to [0,1]$ (ϕ_0 ist eine Zahl aus $[0,1]$), wenn sie folgende Eigenschaften haben:

$$\mathrm{P}_{X_{i+1},\ldots,X_k}\big[\sigma(x_1, \ldots, x_i, X_{i+1}, \ldots, X_k) = 0\big] \leq \phi_i(x_1, \ldots, x_i) \tag{9}$$

$$\mathbb{E}_{X_{i+1}}\big[\phi_{i+1}(x_1, \ldots, x_i, X_{i+1})\big] \leq \phi_i(x_1, \ldots, x_i) \tag{10}$$

Außerdem sollen unsere Schätzer *effizient* und *nützlich* sein. Das heißt, sie lassen sich effizient berechnen und es gilt $\phi_0 < 1$.

Der folgende Satz zeigt, wie diese pessimistischen Schätzer angewandt werden.

Satz 4.2. *Wenn wir über effiziente und nützliche pessimistische Schätzer $(\phi_0, \ldots, \phi_k)$ für ein Ereignis σ verfügen, können wir effizient ein festes $\vec{x} \in [n]^k$ mit $\sigma(\vec{x}) = 1$ berechnen.*

Beweis. Die Eigenschaft $\phi_0 < 1$ stellt sicher, dass das Ereignis σ überhaupt eintreten kann. Wir werden Schritt für Schritt $x_1, \ldots, x_k$ berechnen. Dabei sind in Schritt i jeweils $x_1, \ldots, x_i$ festgelegt. Wir wählen ein x_{i+1} aus, für das

$$\phi_{i+1}(x_1, \ldots, x_{i+1}) \leq \phi_i(x_1, \ldots, x_i)$$

gilt. Dies folgt aus der zweiten Eigenschaft der Schätzer (10),

$$\mathbb{E}_{X_{i+1}}\left[\phi_{i+1}(x_1,\ldots,x_i,X_{i+1})\right] \leq \phi_i(x_1,\ldots,x_i).$$

Nach k Schritten erhalten wir also $\phi_k(\vec{x}) \leq \cdots \leq \phi_0 < 1$. Mit der ersten Eigenschaft der Schätzer (9) folgt

$$\mathrm{P}[\sigma(\vec{x}) = 0] \leq \phi_k(\vec{x}) < 1.$$

Da sich in jedem der k Schritte die Schätzer effizient berechnen lassen, handelt es sich um einen effizienten Algorithmus. □

Lemma 4.3 (Kombinieren von Schätzern)**.** *Für eine Zufallsvariable $\vec{X}$ mit Werten in $[n]^k$ seien zwei Ereignisse $\sigma, \tau\colon [n]^k \to \{0,1\}$ gegeben. Seien $(\phi_0,\ldots,\phi_k)$ und $(\psi_0,\ldots,\psi_k)$ pessimistische Schätzer für σ und τ. Dann sind die Summen der Schätzer $(\phi_0+\psi_0,\ldots,\phi_k+\psi_k)$ pessimistische Schätzer für das Ereignis $\sigma \cap \tau$.*

Beweis. Hierfür zeigen wir, dass die Eigenschaften von Definition 4.1 gelten.

1. Ergibt sich aus einer Abschätzung für das gemeinsame Ereignis,

$$\begin{aligned}&\mathrm{P}[(\sigma\cap\tau)(x_1,\ldots,x_i,X_{i+1},\ldots,X_k)=0]\\ &\leq \mathrm{P}[\sigma(x_1,\ldots,x_i,Xi+1,\ldots,X_k)=0]\\ &+\mathrm{P}[\tau(x_1,\ldots,x_i,Xi+1,\ldots,X_k)=0]\\ &\leq (\phi_i+\psi_i)(x_1,\ldots,x_i).\end{aligned}$$

2. Folgt direkt aus der Linearität des Erwartungswertes.

□

4.3 Derandomisierung der Ahlswede-Winter-Abschätzungen

Wir wenden nun das Prinzip der pessimistischen Schätzer auf die matrixwertigen Abschätzungen 3.9 und 3.11 an. Dadurch können wir gezielt ein Element angeben, welches die Ungleichung erfüllt.

Satz 4.4. *Sei $f\colon [n] \to [-I_d, I_d]$. Sei X über $[n]$ verteilt mit dem Erwartungswert $\mathbb{E}_X[f(X)] = 0$, und seien die Zufallsvariablen $X_1,\ldots,X_k$*

u.i.v.(unabhängig und identisch verteilt) wie X . Sei γ fest mit $1 > \gamma > 0$ und sei $t = \gamma/2$. Wir gehen davon aus, dass $\mathbb{E}[\exp(tf(X))]$ effizient zu berechnen ist. Wir schreiben $\vec{X} = (X_1, \ldots, X_k)$ mit $X_i \in [n]$ und wie oben sei $\sigma\colon [n]^k \to \{0,1\}$ das Ereignis mit $\sigma(\vec{x}) = 1$, falls $\frac{1}{k}\sum_{i=1}^{k} f(x_i) \leq \gamma I$ und $\sigma(\vec{x}) = 0$ sonst. Dann sind die folgenden $(\phi_0, \ldots, \phi_k)$, $\phi_i\colon [n]^i \to [0,1]$ effiziente pessimistische Schätzer für σ :

$$\phi_0 = d\,\mathrm{e}^{-t\gamma k}\|\mathbb{E}[\exp(tf(X))]\|^k$$

(Dieser Wert beträgt höchstens $d \cdot \mathrm{e}^{-\gamma^2 k/4}$),

$$\phi_i(x_1, \ldots, x_i) = d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\exp\left(t\sum_{j=1}^{i} f(x_j)\right)\right) \cdot \|\mathbb{E}[\exp(tf(X))]\|^{k-i}.$$

Beweis. Wir zeigen, dass die Eigenschaften von Definition 4.1 gelten.

1. Von der Ungleichung (6) erhalten wir

$$\mathrm{P}\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] \leq d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\mathbb{E}\left[\exp\left(t\sum_{j=1}^{k} f(X_j)\right)\right]\right)$$

$$\leq d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\mathbb{E}\left[\exp\left(t\sum_{j=1}^{k} f(X_j)\right)\right]\right)\prod_{j=i+1}^{k}\|\mathbb{E}[\exp(tf(X_j))]\|.$$

   Anmerkung: Dass $\|\mathbb{E}[\exp(tf(X))]\| \geq 1$ gilt, sieht man direkt an der Taylor-Entwicklung (8). Wir setzen $X_j = x_j$ für alle $j \leq i$, und erhalten

$$\mathrm{P}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq \gamma I \;\middle|\; X_1 = x_1, \ldots, X_i = x_i\right]$$

$$\leq d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\exp\left(t\sum_{j=1}^{i} f(x_j)\right)\right) \cdot \|\mathbb{E}(\exp(tf(X)))\|^{k-i}$$

$$= \phi_i(x_1, \ldots, x_i).$$

2. Mit Lemma 3.5 erhalten wir die folgende Ungleichung.

$$\mathbb{E}_{X_{i+1}}[\phi_{i+1}(x_1, \ldots, x_i, X_{i+1})]$$

$$= d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\mathbb{E}_{X_{i+1}}\left[\exp\left(t\sum_{j=1}^{i} f(x_i) + tf(X_{i+1})\right)\right]\right) \cdot \|\mathbb{E}(\exp(tf(X)))\|^{k-i-1}$$

$$\leq d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\exp(t\sum_{j=1}^{i} f(x_i))\right) \cdot \|\mathbb{E}(\exp(tf(X))\|^{k-i}$$

$$= \phi_i(x_1, \ldots, x_i).$$

Die ϕ_i sind effizient zu berechnen. Dies machen wir uns auf folgende Weise verständlich:

Ein Algorithmus zur Berechnung der Schätzer ϕ_i erhält als Eingabe die Funktion f als Liste von $d \times d$-Matrizen $f(1), \ldots, f(n)$ sowie 1^k. Die Laufzeit sollte also in $\mathrm{poly}(n, d, k)$ liegen. Um die Schätzer anhand von 4.4 zu berechnen, benötigen wir Matrixmultiplikation, -addition, Spur sowie Matrixexponential- und Normberechnungen. Multiplikation, Addition und Spur lassen sich offensichtlich effizient berechnen. Exponential und Norm ebenfalls, denn die Eigenwerte einer $d \times d$-Matrix lassen sich in $\mathcal{O}(d^3)$ ermitteln. Damit kann man die Matrix diagonalisieren und Norm und Exponential trivialerweise berechnen [7]. Noch zu beachten ist, dass die ϕ_i mit genügend Nachkommastellen berechnet werden, damit sich die auf Maschinengenauigkeit gerundeten Werte wie die reellwertigen Schätzer verhalten.

□

Korollar 4.4 liefert pessimistische Schätzer $(\phi_0, \ldots, \phi_k)$ für σ. Indem wir Satz 3.9 auf $-f$ anwenden, erhalten wir mit dem gleichen Beweis auch effiziente pessimistische Schätzer $(\psi_0, \ldots, \psi_k)$ für das Ereignis τ, bei dem $\tau(\vec{x}) = 1$ genau dann gilt, wenn $\frac{1}{k}\sum_{i=1}^{k} f(x_i) \geq -\gamma I$. Wir kombinieren diese Schätzer mit den ϕ_i und erhalten das folgende Korollar.

Korollar 4.5. *Sei $f\colon [n] \to [-I_d, -I_d]$. Sei X über $[n]$ verteilt mit Erwartungswert $\mathbb{E}_X[f(X)] = 0$, und seien $X_1, \ldots, X_k$ u.i.v. wie X. Sei $1 < \gamma < 0$ fest und $t = \gamma/2$. Wir gehen davon aus, dass $\mathbb{E}[\exp(tf(X))]$ und $\mathbb{E}[\exp(-tf(X))]$ effizient zu berechnen sind.*

Sei $\eta\colon [n]^k \to \{0, 1\}$ das Ereignis mit $\eta(\vec{x}) = 1$, falls $\|\frac{1}{k}\sum_{i=1}^{k} f(x_i)\| \leq \gamma$ und $\eta(\vec{x}) = 0$ sonst. Dann sind $(\phi_0 + \psi_0, \ldots, \phi_k + \psi_k)$ (siehe oben) effiziente pessimistische Schätzer für η.

Beweis. Es gilt $\eta = \sigma \cap \tau$. Effizienz ist offensichtlich, da die ϕ und ψ bereits effizient sind. Indem wir Lemma 4.3 anwenden, zeigen wir, dass

$(\phi_0 + \psi_0, \ldots, \phi_k + \psi_k)$ pessimistischer Schätzer für das Ereignis $\eta = \sigma + \tau$ sind. □

Hiermit können wir Theorem 3.9 effizient derandomisieren. Die einzige Bedingung, die wir hierfür an X stellen, ist, dass wir die Erwartungswerte $\mathbb{E}[\exp(tf(X))]$ und $\mathbb{E}[\exp(-tf(X))]$ berechnen können. Dies ist bereits gegeben, wenn X gleichverteilt ist oder wenn wir die Wahrscheinlichkeit $\mathrm{P}[X = x]$ für jedes $x \in [n]$ effizient berechnen können. Die genaue Verteilung von X ist nicht wichtig, da wir für jedes X_i den Raum aller möglichen Werte durchlaufen.

Satz 4.6. *Sei $f\colon [n] \to [-I_d, I_d]$ so, dass eine Verteilung X über $[n]$ existiert mit $\mathbb{E}[f(X)] = 0$. Dann können wir für $k = \mathcal{O}(\frac{1}{\gamma^2} \log d)$ effizient und deterministisch ein $\vec{x} \in [n]^k$ finden, für das $\|\frac{1}{k} \sum_{i=1}^{k} f(x_i)\| \leq \gamma$ gilt.*

Anmerkung. *Mit der Aussage $k = \mathcal{O}(\frac{1}{\gamma^2} \log d)$ ist hier gemeint, dass für große n die Größenordnung von k nur logarithmisch steigt. Für ein einzeln betrachtetes n ist die Aussage natürlich trivial.*

Beweis. Wir nutzen die effizienten pessimistischen Schätzer aus Korollar 4.5. Wir wählen $k = \mathcal{O}(\frac{1}{\gamma^2} \log d)$ so, dass $\phi_0 + \psi_0 < 1$ gilt und die Schätzer nützlich sind. Dann wenden wir Satz 4.2 an, um das gewünschte Ergebnis zu erhalten. □

Für Satz 3.11 lassen sich pessimistische Schätzer auf die selbe Weise konstruieren.

Satz 4.7. *Sei $f\colon [n] \to [0, I_d]$. Sei X über $[n]$ verteilt, mit $M = \mathbb{E}_X[f(X)] \geq \mu I$ für ein $\mu \in (0, 1)$. Seien $X_1, \ldots, X_k$ u.i.v. wie X. Wir fixieren*

$$t = \log\left(\frac{1 - (1-\gamma)\mu}{1 - \mu} \frac{1}{(1-\gamma)}\right).$$

Seien außerdem $\vec{X} = (X_1, \ldots, X_k)$ mit $X_i \in [n]$ und sei $\sigma\colon [n]^k \to \{0, 1\}$ das Ereignis $\sigma(\vec{x}) = 1$ wenn $\frac{1}{k} \sum_{i=1}^{k} f(x_i) \geq (1-\gamma)\mu I$ und $\sigma(\vec{x}) = 0$ sonst. Dann sind die folgenden $(\phi_0, \ldots, \phi_k), \phi_i\colon [n]^i \to [0, 1]$ effiziente pessimistische Schätzer für σ :

$$\phi_0 = d\, \mathrm{e}^{tk(1-\gamma)\mu} \|\mathbb{E}[\exp(-tf(X))]\|^k$$

(dieser Wert beträgt höchstens $d\,\mathrm{e}^{-\gamma^2\mu k/(2\ln 2)}$ *),*

$$\phi_i(x_1,\ldots,x_i) = d\,\mathrm{e}^{tk(1-\gamma)\mu}\,\mathrm{Tr}\left(\exp\left(-t\sum_{j=1}^{i} f(x_j)\right)\right)\cdot\|\mathbb{E}[\exp(-tf(X))]\|^{k-i}.$$

Beweis. Der Beweis erfolgt analog zum Beweis von Satz 4.4. □

Satz 4.8. *Sei* $f\colon [n]\to[0,I_d]$ *so, dass eine Verteilung* X *über* $[n]$ *und eine Zahl* $\mu\in(0,1)$ *existieren mit* $\mathbb{E}[f(X)]\geq\mu I$*. Dann gilt für* $k=\mathcal{O}(\frac{1}{\gamma^2\mu}\log d)$*, dass wir effizient und deterministisch ein* $\vec{x}\in[n]^k$ *mit*

$$\frac{1}{k}\sum_{i=1}^{k} f(x_i)\geq(1-\gamma)\mu I$$

finden können.

Beweis. Hierfür nutzen wir die Schätzer von Satz 4.7. Für unser k gilt $\phi_0<1$, so dass die Schätzer auch nützlich sind. Der Satz folgt dann aus Satz 4.2. □

5 Expander-Cayleygraphen für endliche Gruppen

Mit den bisherigen Ergebnissen lässt sich der probabilistische Beweis des Alon-Roichman-Theorems [3] entrandomisieren. Es besagt, dass der Cayleygraph einer endlichen Gruppe bezüglich einer Erzeugendenmenge logarithmischer Größenordnung mit positiver Wahrscheinlichkeit ein Expander mit vorgegebener Expansionskonstante ist. Die Entrandomisierung gestattet es, eine Erzeugendenmenge mit der gewünschten Eigenschaft explizit zu konstruieren. Der hier vorgestellte Beweis des Theorems stammt von Landau und Russell [9] (siehe auch [10]).

5.1 Definitionen

Definition 5.1. Eine *Multimenge* M besteht aus einer Menge M' und einer Abbildung $p\colon M'\to\mathbb{N}$. Für ein Element $m\in M'$ gibt $p(m)$ dabei an, wie oft das Element in der Multimenge enthalten sein soll.

Definition 5.2. Der *Cayley-Graph* $\mathrm{Cay}(H;S)$ einer Gruppe H bezüglich der generierenden Multimenge $S\subset H$ ist folgendermaßen definiert. Die Elemente von H entsprechen den Knoten des Graphen. Zwei Knoten h und h'

sind genau dann durch eine Kante verbunden, wenn es ein Element s in S gibt mit $h' = h \cdot s$. Da S eine Multimenge ist, können mehrfache Kanten vorkommen.

Um ungerichtete Cayleygraphen zu erhalten, sollte S symmetrisch sein, das heißt, für jedes $s \in S$ soll S einmal s^{-1} enthalten. Man kann dies erzwingen, indem man als generierende Multimenge den symmetrischen Abschluss $S \sqcup S^{-1}$ nutzt.

Sei $G = (V, E)$ ein ungerichteter d-regulärer Graph auf n Knoten. Wir definieren seine normalisierte Adjazenzmatrix A' mit $A'_{ij} = e_{ij}/d$, dabei sei e_{ij} die Anzahl der Kanten zwischen den Knoten i und j. Wir lassen Schleifen und Mehrfachkanten in G zu. A' ist reellwertig und symmetrisch. Wir gehen hier von zusammenhängendem G aus. Es ist bekannt, dass dann die Eigenwerte von A' die Form $1 = \lambda_1(A') > \lambda_2(A') \geq \cdots \geq \lambda_n(A')$ haben. Als Eigenwerte von G bezeichnen wir die Eigenwerte von A'. Die 1 ist ein einfacher Eigenwert, mit dem zugehörigen Eigenvektor $u = [1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$. Die Orthogonalprojektion auf den von u aufgespannten Unterraum wird durch die Matrix J/n gegeben, wobei J die Matrix voller Einsen sei. Der Cayleygraph $\mathrm{Cay}(H; S)$ einer Gruppe H bezüglich der generierenden Multimenge $S \subset H$ ist der Graph mit Knotenmenge H, bei dem zwei Knoten h und h' mit einer Kante verbunden sind, falls es ein $s \in S$ mit $h' = hs$ gibt. Mehrfach in S auftretende Elemente erzeugen Mehrfachkanten. Hier soll S symmetrisch sein, also für jedes $s \in S$ soll auch $s^{-1} \in S$ sein. Dies ist notwendig, damit unser Cayleygraph ungerichtet ist. Mit $\lambda(\mathrm{Cay}(H; S))$ bezeichnen wir den betragsmäßig zweitgrößten Eigenwert der normalisierten Adjazenzmatrix des Cayleygraphen.

Der hier diskutierte derandomisierte Algorithmus wird als Eingabe ein festes $\lambda < 1$ und die Multiplikationstabelle einer Gruppe H der Ordnung n erhalten und daraus effizient eine kleine Erzeugendenmenge S konstruieren, mit der Eigenschaft $\lambda(\mathrm{Cay}(H; S)) < \lambda$. Bevor wir dies beweisen, geben wir den randomisierten Algorithmus an.

5.2 Ein randomisierter Algorithmus

Satz 5.3. *Sei $0 < \lambda < 1$ fest, und sei H eine Gruppe der Ordnung n. Wir identifizieren H mit $[n]$. Seien $X_1, \ldots, X_k$ zufällig aus H ausgewählt, mit*

$k = \mathcal{O}(\frac{1}{\lambda^2} \log n)$. *Die Multimenge* S *sei* $(X_1, \ldots, X_k)$. *Dann gilt*

$$\mathrm{P}_{S \subseteq H}[\lambda(\mathrm{Cay}(H; S \sqcup S^{-1})) > \lambda] < 1,$$

wobei $S \sqcup S^{-1}$ *den symmetrischen Abschluss von* S *bezeichnet, das heißt, die Anzahl der* s *und* s^{-1} *in* $S \sqcup S^{-1}$ *ist gleich der Anzahl von* s *in* S.

Damit die Notationen in diesem und dem letzten Abschnitt zusammenpassen, soll ab jetzt S unserem $\vec{X}$ entsprechen, $|S| = k$, und es wird sich herausstellen, dass hier $n = d = |H|$ gilt.

Beweis. Für jedes $h \in H$ betrachten wir jeweils die $n \times n$-Matrix P_h. Diese sei die $n \times n$-Permutationsmatrix der Gruppenoperation auf H mittels Rechtsmultiplikation von h. Wir betrachten nun die Matrix $\frac{1}{2}(P_h + P_{h^{-1}})$, und stellen die normalisierte Adjazenzmatrix A' von $\mathrm{Cay}(H; S \sqcup S^{-1})$ dar als

$$A' = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2}(P_{X_i} + P_{X_i^{-1}}).$$

Unser Ziel ist es, $\lambda(A')$ nach oben abzuschätzen. Wir wissen bereits, dass der größte Eigenwert 1 beträgt und der Matrix J/n entspricht, wobei J die Matrix voller Einsen ist. Da wir etwas über den zweitgrößten Eigenwert erfahren wollen, betrachten wir stattdessen

$$(I - J/n)A' = \frac{1}{k} \sum_{i=1}^{k} (I - J/n)\frac{1}{2}(P_{X_i} + P_{X_i^{-1}}).$$

Der größte Eigenwert dieser Matrix sollte λ sein. Wir können nun unsere matrixwertige Funktion

$$f(h) = (I - J/n)\frac{1}{2}(P_h + P_{h^{-1}})$$

definieren, und erhalten

$$\lambda(A') = \|(I - J/n)A'\| = \|\frac{1}{k} \sum_{i=1}^{k} f(X_i)\|.$$

Wir überzeugen uns, dass f die erforderlichen Eigenschaften $f(h) \in \mathrm{Sym}(n)$, $\|f(h)\| \leq 1$ und $\mathbb{E}_{h \in H}[f(h)] = 0_d$ aufweist. Es ist $f(h) \in \mathrm{Sym}(n)$, da die Matrizen I, J und $(P_h + P_{h^{-1}})$ symmetrisch sind und $\mathrm{Sym}(n)$ unter Multi-

plikation abgeschlossen ist. $\|f(h)\| \leq 1$ ergibt sich direkt aus Betrachtung der Eigenwerte, und für den Erwartungswert $\mathbb{E}_{h\in H}[f(h)]$ machen wir uns klar, dass

$$\sum_{h\in H} P_h + P_{h^{-1}} = 2|H|J$$

ist, was uns zu

$$\mathbb{E}_{h\in H}[f(h)] = \frac{1}{n}\sum_{h\in H}(I - J/n)\frac{1}{2}(P_h + P_{h^{-1}}) = (I - J/n)\cdot J = 0$$

führt.

Dies reicht bereits aus, um Satz 3.9 anzuwenden, und wir erhalten

$$\mathrm{P}[\lambda(A') > \gamma] = \mathrm{P}\left[\|\frac{1}{k}\sum_{i=1}^{k} f(X_i)\| > \gamma\right] \leq 2n\,\mathrm{e}^{-\gamma^2|S|/4}\,.$$

Wählen wir also genug Erzeuger, in der Größenordnung $k = \mathcal{O}(\frac{1}{\gamma^2}\log n)$, beträgt diese Wahrscheinlichkeit weniger als 1, was den probabilistischen Beweis abschließt. □

5.3 Anwendung der Schätzer

Nun können wir zeigen, dass die Konstruktion der kleinen Erzeugendenmenge S auch explizit und effizient möglich ist.

Satz 5.4. *Sei $\lambda < 1$ fest. Dann existiert ein in Polynomialzeit abhängig von n arbeitender Algorithmus, der zu einer Gruppe H der Ordnung n eine symmetrische Teilmenge $S \subseteq H$ der Größe $|S| = \mathcal{O}(\frac{1}{\lambda^2}\log n)$ konstruiert, so dass $\lambda(\mathrm{Cay}(H;S)) \leq \lambda$.*

Beweis von Satz 5.4. Hierfür werden wir Satz 5.3 entrandomisieren. Korollar 4.5 liefert uns effiziente pessimistische Schätzer

$$(\phi_0 + \psi_0) = 2d\,\mathrm{e}^{-t\gamma k}\|\mathbb{E}[\exp(tf(X))]\|^k$$

$$(\phi_i + \psi_i)(x_1,\ldots,x_i) = 2d\,\mathrm{e}^{-t\gamma k}\,\mathrm{Tr}\left(\exp\left(t\sum_{j=1}^{i} f(x_j)\right)\right)\cdot\|\mathbb{E}[\exp(tf(X))]\|^{k-i}$$

für das Ereignis σ mit $\sigma(S) = 1$ genau dann, wenn

$$\|\frac{1}{k}\sum_{i=1}^{k} f(X_i)\| \leq \gamma$$

gilt. Wir halten ein ausreichend großes $k = \mathcal{O}(\frac{1}{\gamma^2}\log n)$ fest, so dass die Wahrscheinlichkeit des Ereignisses größer als Null ist. Damit sind unsere Schätzer auch nützlich. Nun wenden wir Satz 4.2 an, um sukzessiv Elemente aus H in S aufzunehmen, und erhalten den gewünschten Cayleygraphen mit Expandereigenschaft. □

6 Zusammenfassung

In der vorliegenden Arbeit wurde aufgezeigt, wie bestehende Erkenntnisse verschiedener Gebiete wie der Informationstheorie, Wahrscheinlichkeitstheorie oder linearen Algebra verbunden wurden und zu einem neuen Resultat mit Anwendungen u. a. in der Computeralgebra bzw. Graphentheorie führten. Die verallgemeinerte Chernoffabschätzung von Ahlswede und Winter wurde neu formuliert und mit Hilfe Raghavans pessimistischer Schätzer derandomisiert. Stellvertretend für die zahlreichen Anwendungen (siehe [16]) wurde die Derandomisierung der Konstruktion eines Cayleygraphen mit Expandereigenschaft vorgestellt.

Literaturverzeichnis

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[2] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

[3] N. Alon and Y. Roichman. Random Cayley Graphs and Expanders. *Random Structures & Algorithms*, 5, 1994.

[4] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley, 2000.

[5] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 2008.

[6] G. Davidoff, P. Sarnak, and A. Valette. *Elementary Number Theory, Group Theory, and Ramanujan Graphs*. Cambridge University Press, 2003.

[7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1989.

[8] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 248–253. IEEE Computer Society, 1989.

[9] Z. Landau and A. Russell. Random Cayley graphs are expanders: a simplified proof of the Alon-Roichman theorem. *The Electronic Journal of Combinatorics*, 11(1), 2004.

[10] P.-S. Loh and L. J. Schulman. Improved Expansion of random Cayley graphs. *Discrete Mathematics and Theoretical Computer Science*, 6(2):523–528, 2004.

[11] N. Nisan and A. Ta-Shma. Extracting randomness: A survey and new constructions. *J. Computer and System Sciences*, 58(1):148–173, 1999.

[12] P. Raghavan. Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *J. Computer and System Sciences*, 37(2):130–143, 1988.

[13] R. Shaltiel. Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 2002.

[14] D. Spielman. *Computationally Efficient Error-Correcting Codes and Holographic Proofs*. PhD thesis, M.I.T., 1995.

[15] T. Tao. The Golden-Thompson inequality. `http://terrytao.wordpress.com/2010/07/15/the-golden-thompson-inequality/`, 2010.

[16] A. Wigderson and D. Xiao. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory of Computing*, 4:53–76, 2008.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.

Marburg, den 7. August 2014

Christian Fetsch