3 Paving Small Matrices and The Kadison-Singer Extension Problem

Gary Weiss* and Vrej Zarikian

May 22, 2006

States λ:
positive linear functionals on a VNA with $\lambda(I) = 1$.

By Krein-Milman:
Convex + Compact in w*-topology $\Rightarrow$
closure of the convex hull of the extreme pts
(these are the **pure states**)

FACT:
$B(H)$-extreme pts = all vector functionals ω_x
$(A \to (Ax, x)$ with $||x|| = 1)$

States on $\mathbb{D}$ (diagonals of $B(\ell^2)$) extend trivially to states on $B(\ell^2)$: $\lambda(A) := \lambda(diag\, A)$

$$\begin{aligned} |\lambda(A)| = |\lambda(diag\, A)| &\leq ||\lambda||_{\mathbb{D}^*} ||diag\, A||_{\mathbb{D}} \\ &\leq ||\lambda||_{\mathbb{D}^*} ||A||_{B(\ell^2)} \end{aligned}$$

In particular, pure states on $\mathbb{D}$ extend to $B(\ell^2)$.
But more–to a pure state on $B(\ell^2)$.

Kadison-Singer Problem [KS] (1959): Unique?

Equivalent to many important problems
(cf. Casazza's invited address at GPOTS 2005-
The KS Problem in Mathematics and Engineering:
op thy, frame thy, Hilb sp thy, Ban sp thy, harm anal, time-requency anal and engineering).

E.g., relative Dixmier property for $\mathbb{D}$:
$\mathbb{D}' \cap \overline{co}\{UAU^* \mid U \text{ unitary in } \mathbb{D}\} \neq \phi \quad \forall A \in B(H)$
(Halpern, Kaftal, W 86-
Rel Dixmier property in discrete crossed products)

Important reformulation
Anderson's Paving Problem: Fix $\epsilon > 0$, $\exists?$ $k \in \mathbb{N}$
$\Rightarrow$ $\forall\ n \in \mathbb{N}$ and each $A \in M_n(\mathbb{C})$ with zero diagonal,
$\exists$ diag $\perp$ proj (d.d.) $P_1 + P_2 + ... + P_k = I$ for which

$$\|P_jAP_j\| \leq \epsilon\|A\|,\ 1 \leq j \leq k.$$

Observed in (HKW87-Matrix pavings in B(H))
and probably long before:
Paving fails for $k = 2$,
e.g., $P_1 + P_2 = I \Rightarrow \underset{i=1,2}{max}\{||P_iAP_i||\} = 1 = ||A||$:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Question (HKW87-ibid): Does $k = 3$ suffice?
We know of no concrete evidence against this.

Bourgain-Tzafriri (91):
Using advanced probabilistic techniques
solved an analogous single compression problem
and the paving problem for large classes of matrices

Reformulation of Corollary 1.2 and
remark following it states:
$\exists$ a universal $C > 0$ for which
every zero-diagonal matrix of size $n \geq \frac{1}{C}$ has a
large diagonal projection with small compression.

Quantitatively: $||PAP|| \leq \epsilon ||A||$, with $\operatorname{rank} P \geq C\epsilon^2 n$.

This cannot be improved because of
(Berman-HKW88-Matrix norm inequalities
and the relative Dixmier property)
but BHKW did implicitly obtain
for non-negative entried matrices $\operatorname{rank} P \geq \frac{\epsilon n}{2}$.

BHKW88-Theorem 3.4: For every k,
(i) For zero-diagonal nonnegative entried A,
$\underset{1 \leq i \leq k}{max} \{||P_i A P_i||\} \leq \frac{2}{k}||A||$ (open question: sharp?)
for some $P_1 + P_2 + ... + P_k = I$
(ii) For s.a. zero-diagonal nonnegative entried A,
$\underset{1 \leq i \leq k}{max} \{||P_i A P_i||\} \leq \frac{1}{k}||A|| \ldots$ and $\frac{1}{k}$ is sharp.

Bourgain-Tzafriri in 91 paragraph before Cor 1.4:
"...reasonably good evidence [that KS is true]"),
KS in 1959 p. 397
"We incline to the view that
such extension is non-unique"
and Casazza presently believes false

BT91-Theorem 2.3: paveable if matrix is sufficiently large (depending on ϵ) and its entries are $O(\frac{1}{(log\, n)^{1+\varepsilon}})$ (ε depending also on ϵ) (lacks uniform bdd k)

HKW86-Matrix pavings and Laurent operators
Corollary 4.2 and Theorem 4.5:
Laurent operators with Riemann integrable symbol are uniformly paveable (infinite matrix setting)
(projections are infinite arithmetic progressions)
but not all Laurents are uniformly paveable

BT91 on Laurent operators-Theorem 4.1:
Laurent $\cap$ Besov ops (weighted ℓ^2 fourier series)
(multiplication by L^∞ funcs with $\sum_{\mathbb{Z}} |\widehat{\phi}(n)|^2 |n|^\tau < \infty$ for some $\tau > 0$) are paveable.

Pavings in C_p-norms

Motivation: $s_1^p \leq \sum_1^n s_i^p \leq n\, s_1^p$ is
$||A|| \leq ||A||_p \leq n^{1/p}||A||$

Hence the variational approach:
set $p = log\, n$ so $n^{1/log\, n} = e^{\frac{log\, n}{log\, n}} = e$
and hence $||A|| \leq ||A||_p \leq e||A||$ and likewise for smaller rank A's,
so C_p-paveability for p sufficiently large in a sense is equivalent to norm paveability.

BHKW90-Some C_4 and C_6 norm inequalities relating to the paving problem

$||\sum_k^{\oplus} P_i A P_i||_2 \leq k^{-1/2}||A||_2$ (sharp for Hadamards)

$||\sum_k^{\oplus} P_i A P_i||_4 \leq k^{-1/4}||A||_4$
(combinatorics improve coefficient to $2^{1/4}k^{-3/8}$)

$||\sum_k^{\oplus} P_i A P_i||_6 \leq 5.3^{1/6}\, k^{-1/3}||A||_6$

BT90 exploits deeper phenomena of this type

e.g., his random paving principle:

an L_p-mean ($p = log\, n$) of single compression norms is small

implying k-paveability where k depends on n

Below we use C_p-norms to yield $B(H)$-norm bounds for 3-pavings.

Negative evidence

BT91-Example 2.2:
zero-diagonal matrices exist for which the BT probabilistic techniques fail (i.e., averages are too large)

Quantitatively:
$\forall\ 0 < \delta < 1,\ \exists\ N(\delta)$ where for each $n \geq N(\delta)$ can construct A with entries $|a_{ij}| \leq 2log\frac{1}{\delta}/log\, n$ with large paving average (expectation)

Bad average example: finite "unitary" shift
Stirling set number-counting number of n-partitions into k-nonempty subsets
$S(n,k) = \frac{1}{k!}\sum_0^{k-1}(-1)^i \binom{k}{i}(k-i)^n$
$\leq$ # of partitions of m not nec empty subsets
$S(n,3) = \frac{1}{6}(3^n - 6\cdot 2^n + 3) = O(3^n)$
partitions not containing any $\{i, i+1\}$ nor $\{1, n\}$:
$\frac{1}{6}3\cdot 2\cdot 1\cdot 2\cdots 2 = 2^{n-3}$
$\therefore$ Ave paving approaches 1 as $n \to \infty$

BT goal for Example 2.2 was to get uniformly small entries-so not this example

Preliminary Report on Bad 3-Pavings

Setup for Bad Paver (i.e., ground up) approach:

For a zero-diagonal matrix
(finite or infinite and bounded),

$$\alpha_3(A) := \underset{d.d.}{inf} \; ||\sum_3^{\oplus} P_i A P_i|| = \underset{d.d.}{inf} \; \underset{3}{max} \, ||P_i A P_i||$$

normalized: $\widetilde{\alpha}_3(A) := \alpha_3(A)/||A||$

Easy facts about α:

$\widetilde{\alpha}_k(A) \leq 1$ and paving problem is
whether or not for some k, $\underset{A}{sup} \; \widetilde{\alpha}_k(A) < 1$

$\alpha_k(\sum^{\oplus} A_i) = \underset{i}{sup} \; \alpha_k(A_i)$ and

$\alpha_k(UAU^*) = \alpha_k(A)$ for all permutations U.

Consequence: KS $\Leftrightarrow$ there is a universal k for which
$\widetilde{\alpha}_k(A) < 1$ for every zero-diagonal $A \in B(\ell^2)$

Bad pavers are extremals A for which
$\alpha_3(A) = max\,||P_i A P_i||$ ($\exists$ via fin dim compactness)

Bad Paver approach:
find bad pavers and study their properties

E.g., preliminary data suggests the best d.d. for bad pavers is uniform, e.g., $3-3-4$ for a 10×10 (open: prove it)

HKW87-Example $(2n+2) \times (2n+2)$ unitary (then remove diagonal) $\widetilde{\alpha}_3(A) \sim 2/3$

$$\frac{1}{n+1}\begin{pmatrix} 1 & -n & 1 & 1 & \cdots \\ 1 & 1 & -n & 1 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 1 & 1 & 1 & \cdots & -n \\ -n & 1 & 1 & \cdots & 1 \end{pmatrix} (Toeplitz)$$

Largest P has PAP norm $\geq \frac{2n+2}{3(n+1)} = 2/3.$

So beating $2/3$ is a starting goal and a problem posed in HKW87-Matrix pavings in B(H).

Bad 4×4 pavers:

idea-first renormalize to insist $\alpha_3(A) = 1$.
$\therefore$ each entry or its diag reflection are ≥ 1.

Associate to A the undirected graph on 4 pts with edge (i, j) when i,j-entry or its reflection is ≥ 1.

All possible such graphs after
left and right multiplication by a diagonal unitary
possess a submatrix which entrywise dominates
$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ which has norm $=$
$\sqrt{\frac{3+\sqrt{5}}{2}} = \frac{1+\sqrt{5}}{2} \approx 1.618$- Fibonnacci's golden ratio.

By searching with our proprietary software:
norm $\frac{1+\sqrt{5}}{2}$ with constraint $\alpha_3(A) = 1$ is attainable.

$$\begin{pmatrix} 0 & 1 & 1 & -\frac{2}{1+\sqrt{5}} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \therefore\ \tilde{\alpha}_3(A) = \frac{2}{1+\sqrt{5}} \approx .618$$

$5 \times 5 = 4 \times 4 \oplus 0_1$

6×6: $\tilde{\alpha}_3(A) = \frac{\sqrt{2}}{2} \approx$ **.7071!**

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ -\frac{1}{2} & 1 & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & -\frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} (\sqrt{2}\cdot\text{unitary})$$

Unique up to basis permutation! (We suspect)
I.e., rare.

Our proof required similar analysis as in the 4 case but the number of isomorphic 6 graphs = 156 (7=1044)

Found graph-axioms and constraints that a bad (optimal) paver must have.
E.g., insure $\alpha_3(A) = 1$ and involving lower bound knowledge of $\tilde{\alpha}_3$.

Reduced # of isomorphic 6 graph optimal candidates to 1 (again rare)
for which other matrix techniques applied.

Software and heuristics:
found **initial conditions crucial**-required theoretical insights

Steepest Ascent Method

Steepest ascent is a method for finding the local maximum of a function. Simply follow the gradient.

Conjugate Gradient Method

Conjugate gradient is a method for finding the local maximum of a function. It is a "weighted" steepest ascent method.

Simulated Annealing

Simulated annealing is a method for finding the global maximum of a function. It (roughly) proceeds as follows: A random change is made to the state of the system. If the change produces an increase in the value of the function, then the change is accepted. If, on the other hand, the change produces a decrease in the value of the function, then the change is accepted with some probability. Usually, the probability decreases exponentially with the decrease in the function value.

• Decide on n and search.
We find many stalls at false peaks so initial conditions are crucial. Theoretical work impacts our choice of initial condition.

• Interpret software findings-translate into useful mathematical objects.

• Continue the loop until example and theory yield optimality.

• Advantaged classes for possible counterexamples (offers special features or merely reduces complexity)

Laurents (Toeplitz) (by BT-non Besov is a place to aim),

Ramanujan matrices (suggested by Davidson/Szarek that incidence matrices of Ramanujan graphs should be bad pavers. This was not our experience.),

conference matrices (zero-diag Hadamard),

circulant matrices
(fin lin combos of U^n, $U =$ the finite unilateral shift)

3-pavings summary

n	General	SA complex	Real symmetric	$a_{ij} \geq 0$
4	0.6180	0.5774	0.4472	0.5550
5	0.6180	0.5774	0.4472	?
6	0.7071	0.5774	?	?
7	$[\mathbf{.8026}, 1]$	?	$[0.6667, 0.7559]$	?
8	?	?	?	?
9	?	?	?	?
10	?	?	$[0.7454, 1]$	?

Single listings = Sharp examples (i.e., theoretically provable):

4×4, 5×5 general:

$$A = \begin{pmatrix} 0 & 1 & 1 & -\frac{2}{1+\sqrt{5}} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

4×4, 5×5, 6×6 SA:

$$A = \begin{pmatrix} 0 & i & 1 & 1 \\ -i & 0 & 1 & -1 \\ 1 & 1 & 0 & i \\ 1 & -1 & -i & 0 \end{pmatrix}$$

4×4, 5×5 symmetric:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \end{pmatrix}$$

4×4 non-negative:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

6×6 general $\tilde{\alpha}_3(A) = \frac{\sqrt{2}}{2} \approx .7071$:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ -\frac{1}{2} & 1 & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & -\frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

7×7 general (lower bound): $\widetilde{\alpha}_3(A) \approx \mathbf{.8026}$

$A \approx$

$$\begin{pmatrix} 0 & 79.7 & 67.2 & 35.9 & -67.5 & 56.1 & -27.8 \\ -27.8 & 0 & 79.7 & 67.2 & 35.9 & -67.5 & 56.1 \\ 56.1 & -27.8 & 0 & 79.7 & 67.2 & 35.9 & -67.5 \\ -67.5 & 56.1 & -27.8 & 0 & 79.7 & 67.2 & 35.9 \\ 35.9 & -67.5 & 56.1 & -27.8 & 0 & 79.7 & 67.2 \\ 67.2 & 35.9 & -67.5 & 56.1 & -27.8 & 0 & 79.7 \\ 79.7 & 67.2 & 35.9 & -67.5 & 56.1 & -27.8 & 0 \end{pmatrix}$$

Obtained by starting with a combination of the "fat" operator and the shift, and searching.

7×7 symmetric (lower bound $2/3 = \widetilde{\alpha}_3(A)$):

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 0 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 0 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 0 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

This example is trying (unsuccessfully, of course) to be a conference matrix (zero-diagonal Hadamard).

Very recent:
7-by-7 matrix (circulant + unitary)
which 3-paves to 0.8231
(previous best 0.8029).

Possible (??) simple argument that shows that NO 7-by-7, circulant, unitary paves to 1.000.

10×10 real symmetric (lower bound .7454 $\approx \tilde{\alpha}_3(A)$):

$$C10 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 0 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 0 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & 1 & 0 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 0 \end{pmatrix}$$

Most attention paid: 7×7 symmetric

Why?: first case where 3×3 compressions are unavoidable
KS $\Leftrightarrow$ paving problem holds for self-adjoints

$2/3 \leq \widetilde{\alpha}_3(A) \leq \frac{2}{\sqrt{7}} \approx .7559$

The 7×7 symmetric upper bound comes from the following fact:

If A is 7×7 s.a. zero-diagonal,
and **every 3-compression of A has norm** ≥ 1
(analog to 2-compression constraint of smaller cases),

then $\|A\| \geq \frac{\sqrt{7}}{2}$. (see last slide)

Now some of the theoretical tools used for these examples

Analytic norm 1 criteria

Recall for 2×2 matrix (and hence rank 2 matrices)

$$||A_{2\times 2}|| = \frac{||A||_2^2}{2} + \sqrt{\frac{||A||_2^4}{4} - |det\ A|^2}$$

and so

$$||A_{2\times 2}|| = 1 \Leftrightarrow |det\ A|^2 + 1 = ||A||_2^2 \leq 2$$

NASC for norm one 3×3 zero-trace **S.A.** matrix (so rank 3 matrix & not nec zero-diagonal)

$$||A|| = 1 \Leftrightarrow \frac{||A||_2^2}{2} + |Det\, A| = 1.$$

For $>, <$ 1, the resp. conditions also equivalent.
Nec cond for norm 1: $3/2 \leq ||A||_2^2 \leq 2$.

NASC for norm one 3×3 zero-trace matrix
(so rank 3 matrix & not nec zero-diagonal)

$$1 \leq ||A||_2^2 \leq 2 + |Det\ A|^2$$

and

$$||A||_2^2 - \frac{1}{2}(||A||_2^4 - ||S||_4^4) + |Det\ A|^2 = 1$$

or if you prefer

$$||A||_4^4 + 2|Det\ A|^2 = (||A||_2^2 - 1)^2 + 1$$

Necessary conditions

$$|Det\ A| \leq 1 \quad \text{and} \quad |Det\ A| \leq \frac{||A||_2^2 - 1}{2}$$

Observe: in all these cases criteria involves only the determinant and the C_2, C_4-norms. Advantageous for averaging.

Problem: NASC for larger ranks
For generalizing-symmetric function representations of coefs for the general char poly appear relevant. (Cayley Hamilton also appears relevant.)

4×4 trace-zero S.A. norm 1-exists but hard

S.A. with large 3-compressions

$n \times n$ S.A. zero-trace

$$\|A\|^2 \leq \frac{n-1}{n}\|A\|_2^2.$$

Easy from
$ns_1^2 = n\|A\|^2 \leq (n-1)\|A\|_2^2 \underset{trace-0\&Holder}{\Leftrightarrow}$
$s_1 \leq \sum_2^n |s_j| \leq (n-1)^{1/2}(\sum_2^n |s_j|^2)^{1/2}$

$$\|A\|_2^2 \leq \begin{cases} n\|A\|^2 & n \text{ even} \\ (n-1)\|A\|^2 & n \text{ odd} \end{cases}.$$

Even case is simply Holder and holds for all n but **odd case is harder and may be new**

$\Downarrow$

Zero-diag w/ all 3-compression having norm ≥ 1

$$\|A\| \geq \begin{cases} \frac{\sqrt{n-1}}{2} & n \text{ even} \\ \frac{\sqrt{n}}{2} & n \text{ odd} \end{cases}.$$

by a Hilbert-Schmidt averaging argument

$n = 7 \Rightarrow \tilde{\alpha}_3(A) = \frac{\alpha_3(A)}{||A||} \geq \frac{1}{\frac{\sqrt{7}}{2}} = \frac{2}{\sqrt{7}} \approx .7559$
bound we saw above.