Uncertainty Principles, Extractors, and Explicit Embeddings of $L_2$ into $L_1$

Piotr Indyk
MIT
Uncertainty principles (UP)

• Consider a vector \( x \in \mathbb{R}^n \) and a Fourier matrix \( F \)

• UP: either \( x \) or \( Fx \) must have “many” non-zero entries (for \( x \neq 0 \) )

• History:
  – Physics: Heisenberg principle
  – Signal processing [Donoho-Stark’89]:
    • Consider any \( 2n \times n \) matrix \( A=[I \ B]^T \) such that
      – \( B \) is orthonormal
      – For any distinct rows \( A_i, A_j \) of \( A \) we have
        \[ |A_i^* * A_j| \leq M \]
    • Then for any \( x \in \mathbb{R}^n \) we have that
      \[ ||x||_0 + ||Bx||_0 > 1/M \]
  – E.g., if \( A=[I \ H]^T \), where \( H \) is a normalized \( n \times n \) Hadamard matrix (orthogonal, entries in \( \{-1/n^{1/2}, 1/n^{1/2}\} \)):
    • \( M=1/n^{1/2} \)
    • \( Ax \) must have \( >n^{1/2} \) non-zero entries
Extractors

- Expander-like graphs:
  - $G=(A, B, E)$, $|A|=a$, $|B|=b$
  - Left degree $d$
- Property:
  - Consider any distribution $P=(p_1, \ldots, p_a)$ on $A$ s.t. $p_i \leq 1/k$
  - $G(P)$: a distribution on $B$:
    - Pick $i$ from $P$
    - Pick $t$ uniformly from $[d]$
    - $j$ is the $t$-th neighbor of $i$
  - Then $||G(P)-\text{Uniform}(B)||_1 \leq \varepsilon$
- Equivalently, can require the above for $p_i = 1/k$
- Many explicit constructions
- Holy grail:
  - $k=b$
  - $d=O(\log a)$
- Observation: w.l.o.g. one can assume that the right degree is $O(ad/b)$
(Norm) embeddings

• Metric spaces $M=(X,D)$, $M'=(X',D')$
  (here, $X=\mathbb{R}^n$, $X'=\mathbb{R}^m$, $D=||.||_X$ and $D=||.||_{X'}$)
• A mapping $F: M \rightarrow M'$ is a $c$-embedding if for any $p \in X$, $q \in X$ we have
  $D(p,q) \leq D'(F(p),F(q)) \leq c \cdot D(p,q)$
  (or, $||p-q||_X \leq ||F(p-q)||_{X'} \leq c \cdot ||p-q||_X$)

• History:
  – Mathematics:
    • [Dvoretzky'59]: there exists $m(n,\varepsilon)$ s.t., for any $m>m(n,\varepsilon)$ and any space $M'=(\mathbb{R}^m,||.||_{X'})$
      there exists a $(1+\varepsilon)$-embedding of an $n$-dimensional Euclidean space $l_2^n$ into $M'$
    • In general, $m$ must be exponential in $n$
    • [Milman'71]: probabilistic proof
    • ........
    • [Figiel-Lindenstrauss-Milman'77, Gordon]: if $M'=l_1^m$, then $m \approx n/\varepsilon^2$ suffices
      That is, $l_2^n$ $O(1)$-embeds into $l_1^{O(n)}$
      A.k.a. Dvoretzky's theorem for $l_1$
  – Computer science:
    • [Linial-London-Rabinovich'94]: [Bourgain'85] for sparsest cut, many other tools
    • ........
    • [Dvoretzky, FLM] used for approximate nearest neighbor [IM'98, KOR'98], hardness of
      lattice problems [Regev-Rosen'06], etc.
Recap

• Uncertainty principles
• Extractors
• (Norm) embeddings
Dvoretzky theorem, ctd.

• Since [Milman’71], almost all proofs in geometric functional analysis use probabilistic method
  – In particular, the [FLM] embedding of $l_2^n$ into $l_1^{O(n)}$ uses random linear mapping $A: \mathbb{R}^n \rightarrow \mathbb{R}^{O(n)}$

• **Question**: can one construct such embeddings **explicitly**?
  [Milman’00, Johnson-Schechtman’01]
### Embedding \( l_2^n \) into \( l_1 \)

<table>
<thead>
<tr>
<th>Distortion</th>
<th>Dim. of ( l_1 )</th>
<th>Type</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1+\sigma )</td>
<td>( O(n/\sigma^2) )</td>
<td>Probabilistic</td>
<td>[FLM,Gordon]</td>
</tr>
<tr>
<td>( O(1) )</td>
<td>( O(n^2) )</td>
<td>Explicit</td>
<td>[Rudin’60,LLR]</td>
</tr>
<tr>
<td>( 1+1/n )</td>
<td>( n^{O(\log n)} )</td>
<td>Explicit</td>
<td>[Indyk’00]</td>
</tr>
<tr>
<td>( 1+1/\log n )</td>
<td>( n^{2^{O(\log \log n)^2}} )</td>
<td>Explicit</td>
<td>This talk</td>
</tr>
</tbody>
</table>
Other implications

• The embedding takes time $O(n^{1+o(1)})$, as opposed to $O(n^2)$ [FLM]

• Similar phenomenon discovered for Johnson-Lindenstrauss dimensionality reduction lemma [Ailon-Chazelle’06],
  – Applications to approximate nearest neighbor problem, Singular Value Decomposition, etc
  – More on this later
Embedding of $l_2^n$ into $l_1$: overview

• How does [FLM] work?
  – Choose a “random” matrix $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m=O(n)$
    • E.g., the entries $A_{ij}$ are i.i.d. random variables with normal distribution ($\text{mean}=0$, $\text{variance}=1/m$)
  – Then, for any unit vector $x$, $(Ax)_i$ is distributed according to $N(0,1/m)$
  – With prob. $1-\exp(-m)$, constant fraction of such variables is $\Theta(1/m^{1/2})$. I.e.,
    $$Ax = (\approx1/m^{1/2}, \ldots, \approx1/m^{1/2}, \ldots, \ldots, \approx1/m^{1/2})$$
  – This implies $\|Ax\|_1 = \Omega( m^{1/2} \|x\|_2 )$
  – Similar arguments give $\|Ax\|_1 = O( m^{1/2} \|x\|_2 )$
  – Can extend to the whole $\mathbb{R}^n$
Overview, ctd.

- We would like to obtain something like

\[ Ax = (\approx 1/m^{1/2}, \ldots, \approx 1/m^{1/2}, \ldots, \ldots, \approx 1/m^{1/2}) \]

\[
\begin{array}{cccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

\[ \approx 0^{1/4} \quad \approx 0^{1/4} \quad 0 \quad \approx 0^{1/4} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \]

UP: \( n^{1/2} \) non-zero entries

Extractor

\[ \approx 1/n^{1/4} \quad 0 \quad 0 \quad \approx 1/n^{1/4} \quad 0 \quad 0 \quad 0 \quad 0 \quad \approx 1/n^{1/4} \]

* constant

* \( d = \log^{O(1)} n \)

Repeat \( \log \log n \) times

Total blowup: \( (\log n)^{O(\log \log n)} \)
Part I:

- Lemma:
  - Let \( A = [H_1 \ H_2 \ldots \ H_L]^T \), such that:
    - Each \( H_i \) is an \( n \times n \) orthonormal matrix
    - For any two distinct rows \( A_i, A_j \) we have \( |A_i^*A_j| \leq M \)
    - \( M \) is called coherence
  - Then, for any \( x \in \mathbb{R}^n \), and set \( S \) of coordinates, \( |S| = s \):
    \[
    \| (Ax)_S \|_2^2 \leq 1 + Ms
    \]
    (note that \( \| (Ax) \|_2^2 = L \))

- Proof:
  - Take \( A_S \)
  - \( \max_{\|x\| = 1} \|A_S x\|_2^2 = \lambda(A_S \times A_S^T) \)
  - But \( A_S \times A_S = I + E \), \( |E| \leq M \)
  - Since \( E \) is an \( s \times s \) matrix, \( \lambda(E) \leq Ms \)

- Suppose that we have \( A \) s.t. \( M \leq 1/n^{1/2} \). Then:
  - For any \( x \in \mathbb{R}^n \), \( |S| \leq n^{1/2} \), we have \( \| (Ax)_S \|_2^2 \leq 2 \)
  - At the same time, \( \| (Ax) \|_2^2 = L \)
  - Therefore, \((1-2/L)\) fraction of the “mass” \( \|Ax\|_2^2 \) is contained in coordinates \( i \) s.t. \( (Ax)_i^2 \leq 1/n^{1/2} \)
Part II:

- Let \( y=(y_1, \ldots, y_n') \)
- Define probability distribution
  \[
  P = (\frac{y_1^2}{||y||_2^2}, \ldots, \frac{y_n^2}{||y||_2^2})
  \]
- Extractor properties imply that, for “most” buckets \( B_i \), we have
  \[
  ||G(y)|_{B_i}||_2^2 \approx \frac{||G(y)||_2^2}{\#buckets}
  \]
- After \( \log \log n \) steps, “most” entries will be around \( 1/n^{1/2} \)
Incoherent dictionaries

• Can we construct $A = [H_1 \ H_2 \ ... \ H_L]^T$ with coherence $1/n^{1/2}$?
  - For $L=2$, take $A = [I \ H]^T$
  - Turns out $A$ exists for $L$ up to $n/2+1$
    • $H_i = H \times D_i$, $D_i$ has ±1 on the diagonal and 0’s elsewhere
    • [Calderbank-Cameron-Cantor-Seidel] for more (Kerdock codes)
Digression

- **Johnson-Lindenstrauss’84:**
  - Take a “random” matrix $A: \mathbb{R}^n \rightarrow \mathbb{R}^{m/\varepsilon^2} (m << n)$
  - For any $\varepsilon > 0, x \in \mathbb{R}^n$ we have
    \[ ||Ax||_2 = (1 \pm \varepsilon)||x||_2 \]
    with probability $1-\exp(m)$
  - $Ax$ can be computed in $O(mn/\varepsilon^2)$ time

- **Ailon-Chazelle’06:**
  - Essentially: take $B = A \times P \times (H \times D_i)$, where
    - $H$: Hadamard matrix
    - $D_i$: random $\pm 1$ diagonal matrix
    - $P$: projection on $m^2$ coordinates
    - $A$ as above (but $n$ replaced by $m/\varepsilon^2$)
  - $Ax$ can be computed $O(n\log n + m^3/\varepsilon^2)$
Conclusions

• Extractors+UP → Embedding of $l_2^n$ into $l_1$
  – Dimension almost as good as for the probabilistic result
  – Near-linear in $n$ embedding time

• Questions:
  – Remove $2^{O(\log \log n)^2}$ ?
  – Making other embeddings explicit ?
  – Any particular reason why both [AC’06] and this paper use $H \times D_i$ matrices ?