# Stability of low-rank matrix recovery and its connections to Banach space geometry


J. Alejandro Chávez-Domínguez

Department of Mathematics
The University of Texas at Austin

Joint work with Denka Kutzarova (UIUC)


First Brazilian Workshop in Geometry of Banach spaces
Maresias, Brazil August 29th, 2014

http://arxiv.org/abs/1406.6712

# Practical problems

- Inferring quantities of interest from measured information; Compressing information.

$$x \in \mathbb{R}^n \text{ unknown,} \quad \text{measurement } y = Ax$$

where

$$A : \mathbb{R}^n \to \mathbb{R}^m \text{ is a linear map,} \qquad n > m$$

We have $y$, we want to recover $x$.

# Practical problems

- Inferring quantities of interest from measured information; Compressing information.

$$x \in \mathbb{R}^n \text{ unknown}, \quad \text{measurement } y = Ax$$

  where

$$A : \mathbb{R}^n \to \mathbb{R}^m \text{ is a linear map}, \qquad n > m$$

  We have $y$, we want to recover $x$.

- Impossible! Too many degrees of freedom.

# Practical problems

- Inferring quantities of interest from measured information; Compressing information.

  $$x \in \mathbb{R}^n \text{ unknown}, \qquad \text{measurement } y = Ax$$

  where

  $$A : \mathbb{R}^n \to \mathbb{R}^m \text{ is a linear map}, \qquad n > m$$

  We have $y$, we want to recover $x$.

- Impossible! Too many degrees of freedom.

**The magic of compressed sensing**

It becomes possible if $x$ is *sparse*, i.e. has few nonzero coordinates.

# Sparse recovery problem

- Given the measurement $y$, what's a reasonable guess for $x$?

# Sparse recovery problem

- Given the measurement $y$, what's a reasonable guess for $x$?
- IDEA: Find the sparsest vector that is consistent with the measurement.

$$\text{minimize } \|z\|_0 \quad \text{subject to } Az = y,$$

where

$$\|z\|_0 = \text{ number of nonzero coordinates of } z.$$

# Sparse recovery problem

- Given the measurement $y$, what's a reasonable guess for $x$?
- IDEA: Find the sparsest vector that is consistent with the measurement.

$$\text{minimize } \|z\|_0 \quad \text{subject to } Az = y,$$

where

$$\|z\|_0 = \text{ number of nonzero coordinates of } z.$$

- This is NP-hard! [Natarajan 1995].

# Norm minimization

- Instead of

$$\text{minimize } \|z\|_0 \quad \text{subject to } Az = y,$$

we would like to consider

$$\text{minimize } \|z\|_{\ell_1} \quad \text{subject to } Az = y.$$

- This is a *convex problem* that can be solved efficiently.

## Moral

If the original vector $x$ is sparse enough, both problems have $x$ as solution (as long as we choose the linear map $A$ wisely).

# Norm minimization

- Instead of
$$\text{minimize } \|z\|_0 \quad \text{subject to } Az = y,$$
we would like to consider
$$\text{minimize } \|z\|_{\ell_1} \quad \text{subject to } Az = y.$$
- This is a *convex problem* that can be solved efficiently.

**Moral**

If the original vector $x$ is sparse enough, both problems have $x$ as solution (as long as we choose the linear map $A$ wisely).

**(Rough) Example of a Result**

For a fixed $k$-sparse vector $x$ and a random Gaussian matrix $A$, $\ell_1$-minimization exactly recovers $x$ with high probability if $m > 2k\ln(n/k)$.

# Stability

- True sparsity: only in idealized situations.
- More realistic: the unknown vector is close to sparse vectors.

# Stability

- True sparsity: only in idealized situations.
- More realistic: the unknown vector is close to sparse vectors.
- Examples: Image or sound compression.

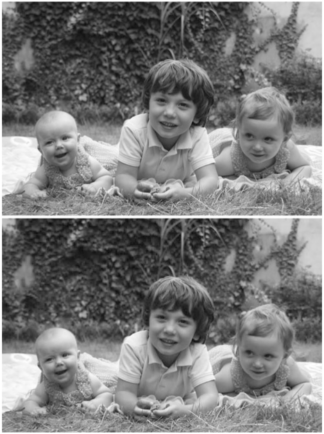
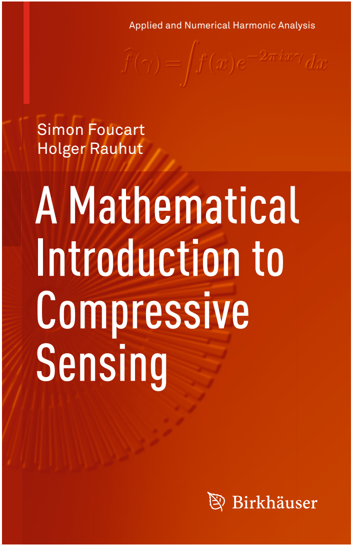


**Fig. 1.1** Antonella, Niels, and Paulina. *Top*: Original Image. *Bottom*: Reconstruction using 1% of the largest absolute wavelet coefficients, i.e., 99 % of the coefficients are set to zero

# Unpaid advertisement



# Stability

We would like to recover a vector $x$ with an error controlled by its distance to $k$-sparse vectors.

$$\rho_k(x)_{\ell_p} := \inf \left\{ \left\| x - x' \right\|_{\ell_p} \ : \ \left\| x' \right\|_0 \leq k \right\}.$$

Denote by $\Delta_p(y)$ a solution to

$$\text{minimize } \|z\|_{\ell_p} \quad \text{subject to } Az = y,$$

## (Rough) Example of a Result

There are linear maps $A : \mathbb{R}^n \to \mathbb{R}^m$ such that for every vector $x$

$$\|x - \Delta_1(Ax)\|_{\ell_1} \leq 6\rho_k(x)_{\ell_1}.$$

# Stability and geometry

**Theorem (Kashin-Temlyakov 2007)**

*Linear map* $A : \mathbb{R}^n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. *TFAE (up to a change in the constant):*

1. *For all* $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(Ax)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. *For all* $x \in \ker(A)$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2}\|x\|_{\ell_1}$$

# Stability and geometry

**Theorem (Kashin-Temlyakov 2007)**

*Linear map* $A : \mathbb{R}^n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. *TFAE (up to a change in the constant):*

1. *For all* $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(Ax)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. *For all* $x \in \ker(A)$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2}\|x\|_{\ell_1}$$

**Remark**

The latter is a statement about the norm of the identity map $\ell_1^n \to \ell_2^n$ restricted to the subspace $\ker(A)$.

# Gelfand numbers

- Gelfand $m$-number of a linear map $T : X \to Y$

$$c_m(T) := \inf \left\{ \left\| T\big|_L \right\| \ : \ L \text{ subspace of } X \text{ with } \operatorname{codim}(L) < m \right\}$$

- **Punchline**: How small can the norm of the operator be if we restrict ourselves to subspaces of a given codimension?

- The sequence $\left(c_m(T)\right)_{m=1}^{\infty}$ is a measure of the compactness of $T$.

- **Kashin-Temlyakov:**
  Results of [Kashin 1977; Garnaev-Gluskin 1984] on Gelfand numbers immediately imply the existence of good measurement maps for sparse recovery through $\ell_1$-minimization.

# Gelfand: numbers vs. widths

- In approximation theory/compressed sensing they normally use **Gelfand widths** instead of **Gelfand numbers**.
- See Pietsch's book *History of Banach spaces and linear operators* for an argument of why the numbers "won" in Banach space theory.
- In the cases we are considering, both concepts coincide.
- For geometric conditions guaranteeing the coincidence in more general situations, see [Edmunds-Lang 2013].

# Extra structure: matrices

- Sometimes the space of unknown vectors has an extra matricial structure.

## Matrix completion (Netflix problem)

- TASK: fill in missing entries of a matrix.
- Online store sells products indexed by the rows, consumers indexed by the columns rate some of these products.

  For purposes of individualized advertisement, the store is interested in predicting the whole matrix of consumer ratings.

# Low-rank recovery

- To stand a chance of success, the unknown matrix should have low rank (corresponding to sparsity).

**Low-rank recovery problem**

Solve

$$\text{minimize}\,\text{rank}(Z) \quad \text{subject to } \mathcal{A}Z = y,$$

where $\mathcal{A} : M^n \to \mathbb{R}^m$ is a linear map, $n^2 > m$.

- This is also NP-hard.

# Norm minimization

- Can some sort of norm minimization help?
- Natural candidate: Schatten $p$-norm.

$$\|X\|_{S_p} := \ell_p\text{-norm of the vector of singular values of } X$$

$S_p^n = (M_n, \|\cdot\|_{S_p})$ is considered a noncommutative version of $\ell_p^n$.

- Instead of

$$\text{minimize}\,\text{rank}(Z) \quad \text{subject to } \mathcal{A}Z = y,$$

we would like to consider

$$\text{minimize}\,\|Z\|_{S_p} \quad \text{subject to } \mathcal{A}Z = y.$$

# Plenty of recent work

- Fazel 2002.
- Candès-Recht 2009.
- Candès-Tao 2010.
- Recht-Fazel-Parrilo 2010.
- Dvijotham-Fazel 2010
- Candès-Plan 2011.
- Fornasier-Rauhut-Ward 2011.
- Recht-Hu-Hassibi 2011.
- Oymak-Mohan-Fazel-Hassibi 2011.
- Kong-Xiu 2013.
- Cai-Zhang 2014.
- And more...

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Kashin-Temlyakov Theorem

Linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $x \in \mathbb{R}^n$,
$$\|x - \Delta_1(Ax)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$
2. For all $x \in \ker(A)$,
$$\|x\|_{\ell_2} \leq Ck^{-1/2} \|x\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(Ax)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. For all $x \in \ker(A)$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2} \|x\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(Ax)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. For all $x \in \ker(A)$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2}\|x\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(\mathcal{A}x)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. For all $x \in \ker(\mathcal{A})$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2} \|x\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

**Noncommutative Kashin-Temlyakov Theorem**

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $x \in \mathbb{R}^n$,

$$\|x - \Delta_1(\mathcal{A}x)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(x)_{\ell_1}$$

2. For all $x \in \ker(\mathcal{A})$,

$$\|x\|_{\ell_2} \leq Ck^{-1/2} \|x\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(X)_{\ell_1}$$

2. For all $X \in \ker(\mathcal{A})$,

$$\|X\|_{\ell_2} \leq Ck^{-1/2} \|X\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{\ell_2} \leq Ck^{-1/2}\rho_k(X)_{\ell_1}$$

2. For all $X \in \ker(\mathcal{A})$,

$$\|X\|_{\ell_2} \leq Ck^{-1/2} \|X\|_{\ell_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq Ck^{-1/2}\rho_k(X)_{S_1}$$

2. For all $X \in \ker(\mathcal{A})$,

$$\|X\|_{S_2} \leq Ck^{-1/2}\|X\|_{S_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,
$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq Ck^{-1/2}\rho_k(X)_{S_1}$$
2. For all $X \in \ker(\mathcal{A})$,
$$\|X\|_{S_2} \leq Ck^{-1/2}\|X\|_{S_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/log(m/n)$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq Ck^{-1/2}\rho_k(X)_{S_1}$$

2. For all $X \in \ker(\mathcal{A})$,

$$\|X\|_{S_2} \leq Ck^{-1/2}\|X\|_{S_1}$$

# Goal: Noncommutative version of the Kashin-Temlyakov theorem

## Noncommutative Kashin-Temlyakov Theorem

Linear map $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp ?????$. TFAE (up to a change in the constant):

1. For all $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq Ck^{-1/2} \rho_k(X)_{S_1}$$

2. For all $X \in \ker(\mathcal{A})$,

$$\|X\|_{S_2} \leq Ck^{-1/2} \|X\|_{S_1}$$

# Known results

**Theorem (Oymak-Mohan-Fazel-Hassibi 2011)**

*Let $C$ be a constant. Then the following are equivalent:*

(i) *For any $X$ and $X'$ with $\|X'\|_{S_1} \leq \|X\|_{S_1}$ and $\mathcal{A}X = \mathcal{A}X'$,*

$$\left\|X - X'\right\|_{S_2} \leq Ck^{-1/2}\rho_k(X)_{S_1}$$

(ii) *For any $Y \in \ker(\mathcal{A})$ we have*

$$\left\|Y - Y_{[k]}\right\|_{S_1} - \left\|Y_{[k]}\right\|_{S_1} \geq \frac{2\sqrt{k}}{C}\,\|Y\|_{S_1}$$

Where $Y_{[k]}$ is the $k$-spectral truncation of $Y$ (i.e. keeping just the $k$ largest singular values).

# Known results II

**Theorem (Oymak-Mohan-Fazel-Hassibi 2011)**

*Suppose that for all $X \in \ker(\mathcal{A})$ we have*

$$\|X\|_{S_2} \leq \frac{1}{\sqrt{D}} \|X\|_{S_1}$$

*Then for any $X$ and $X'$ with $\|X'\|_{S_1} \leq \|X\|_{S_1}$ and $\mathcal{A}X = \mathcal{A}X'$, it holds that*

$$\left\|X - X'\right\|_{S_1} \leq \frac{2}{1 - 2\sqrt{k/D}} \rho_k(X)_{S_1}.$$

# Noncommutative Kashin-Temlyakov

## Theorem (CD-K)

*Linear map* $\mathcal{A} : M_n \to \mathbb{R}^m$, $k \asymp n/m$. *TFAE (up to a change in the constant):*

1. *For all* $X \in M_n$,

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq Ck^{-1/2}\rho_k(X)_{S_1}$$

2. *For all* $X \in \ker(\mathcal{A})$,

$$\|X\|_{S_2} \leq Ck^{-1/2} \|X\|_{S_1} .$$

# Turning the tables

Use compressed sensing ideas to prove results about Gelfand numbers.

**Theorem (Foucart-Pajor-Rauhut-Ullrich 2010)**

*For* $0 < p \leq 1$ *and* $p < q \leq 2$*, if* $m < n$*, then*

$$c_m(Id : \ell_p^n \to \ell_q^n) \asymp \min\left\{1, \frac{\ln(n/m)+1}{m}\right\}^{1/p-1/q}.$$

# Goal, Part II

- Noncommutative version of the Foucart-Pajor-Rauhut-Ullrich theorem: calculate the Gelfand numbers of identity maps

$$Id : S_p^n \to S_q^n$$

for $0 < p \leq 1$ and $p < q \leq 2$.

# Goal, Part II

- Noncommutative version of the Foucart-Pajor-Rauhut-Ullrich theorem: calculate the Gelfand numbers of identity maps

$$Id : S_p^n \to S_q^n$$

  for $0 < p \leq 1$ and $p < q \leq 2$.

- Known result: [Carl-Defant 1997]

$$c_m(Id : S_1^n \to S_2^n) \asymp \min\left\{1, \frac{n}{m}\right\}^{1/2}.$$

  Their approach is based on tensor product arguments plus a deep inequality of Pajor and Tomczak-Jaegermann.

# Main result

We are able to calculate Gelfand numbers in a noncommutative version of the Foucart-Pajor-Rauhut-Ullrich result.

## Theorem (CD-K)

*For* $0 < p \leq 1$ *and* $p < q \leq 2$, *if* $1 \leq m < n^2$, *then*

$$c_m(Id : S_p^n \to S_q^n) \asymp \min\left\{1, \frac{n}{m}\right\}^{1/p-1/q}.$$

# Technical tool: Restricted Isometry Property

A condition on the measuring map that guarantees stable recovery.

# Technical tool: Restricted Isometry Property

A condition on the measuring map that guarantees stable recovery.

## RIP [Candès-Tao 2005]

$A : \mathbb{R}^n \to \mathbb{R}^m$ has the RIP of order $k$ with constant $\delta_k > 0$ if for all vectors $z \in \mathbb{R}^n$ of sparsity at most $k$,

$$(1-\delta_k)^{1/2} \|z\|_{\ell_2} \leq \|Az\|_{\ell_2} \leq (1+\delta_k)^{1/2} \|z\|_{\ell_2}$$

# Technical tool: Restricted Isometry Property

A condition on the measuring map that guarantees stable recovery.

## RIP [Candès-Tao 2005]

$A : \mathbb{R}^n \to \mathbb{R}^m$ has the RIP of order $k$ with constant $\delta_k > 0$ if for all vectors $z \in \mathbb{R}^n$ of sparsity at most $k$,

$$(1-\delta_k)^{1/2} \|z\|_{\ell_2} \leq \|Az\|_{\ell_2} \leq (1+\delta_k)^{1/2} \|z\|_{\ell_2}$$

**Note:** There are strong connections between the RIP and the Johnson-Lindenstrauss lemma [Baraniuk-Davenport-DeVore-Wakin 2008, Krahmer-Ward 2011].

# Noncommutative Restricted Isometry Property

## Matrix RIP [Recht-Fazel-Parrilo 2010]

$\mathcal{A} : M_n \to \mathbb{R}^m$ has the RIP of order $k$ with constant $\delta_k > 0$ if for all matrices $Z \in M_n$ of rank at most $k$,

$$(1 - \delta_k)^{1/2} \|Z\|_{S_2} \leq \|\mathcal{A}Z\|_{\ell_2} \leq (1 + \delta_k)^{1/2} \|Z\|_{S_2}$$

# Noncommutative Restricted Isometry Property

**Matrix RIP [Recht-Fazel-Parrilo 2010]**

$\mathcal{A} : M_n \to \mathbb{R}^m$ has the RIP of order $k$ with constant $\delta_k > 0$ if for all matrices $Z \in M_n$ of rank at most $k$,

$$(1 - \delta_k)^{1/2} \|Z\|_{S_2} \leq \|\mathcal{A}Z\|_{\ell_2} \leq (1 + \delta_k)^{1/2} \|Z\|_{S_2}$$

- We use a modified version inspired by [Foucart-Lai 2009].

# Stability and Robustness: technical result

## Theorem (CD-K)

*Given* $0 < p \leq 1$, *if for integers* $t \geq k$ *we have*

$$\frac{2\delta_{2t}}{1-\delta_{2t}} < 4(\sqrt{2}-1)\left(\frac{t}{k}\right)^{1/p-1/2},$$

*then a solution* $X^*$ *of*

$$\text{minimize } \|Z\|_{S_p} \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \beta_{2k} \cdot \theta.$$

*approximates the original matrix* $X$ *with errors*

$$\|X - X^*\|_{S_p} \leq C_1 \rho_k(X)_{S_p} + D_1 \cdot k^{1/p-1/2} \cdot \theta,$$

$$\|X - X^*\|_{S_2} \leq C_2 \frac{\rho_k(X)_{S_p}}{t^{1/p-1/2}} + D_2 \cdot \theta.$$

# Idea of the proof

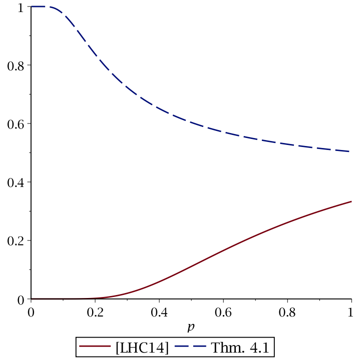
- It follows the general strategy of [Foucart-Lai 2009].
- BIG ISSUE: if $P : \mathbb{R}^n \to \mathbb{R}^n$ is a projection over a subset of coordinates, then for any vector $x \in \mathbb{R}^n$,

$$\|x\|_{\ell_p}^p = \|Px\|_{\ell_p}^p + \|x - Px\|_{\ell_p}^p .$$

  This is not the case for matrices.
- We overcome this difficulty by modifying some matrix decompositions from [Kong-Xiu 2013].

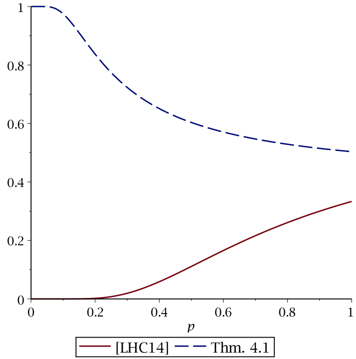
# Comparison to related results

- A recent result similar to ours has appeared in [Liu-Huang-Chen 2014]. Their conclusions have the same form, but our hypothesis are weaker.



# Comparison to related results

- A recent result similar to ours has appeared in [Liu-Huang-Chen 2014]. Their conclusions have the same form, but our hypothesis are weaker.
- [MalekMohammadi-BabaieZadeh-Skoglund] have independently obtained the same stability result as us.
- Neither one of those papers calculate Gelfand widths.



THANKS!