Low–Complexity Model Uncertainties in Compressed Sensing with Application to Sporadic Communication

Peter Jung

Technische Universität Berlin
Fachgebiet Informationstheorie und theoretische Informationstechnik

* joint work with *Philipp Walk*, Technische Universität München
Lehrstuhl für Theoretische Informationstechnik

Matheon Workshop 2013, 9-13.Dec, *Compressed Sensing and its Applications*

[revised version Mar2014]

Motivation

Why Sporadic Communication

- H2H = main driver for 2G/3G, human generated traffic for 4G
- Machine type communication (MTC) will be a major 5G issue, part of the "*internet of things*" vision
- MTC traffic: low data rate but *sporadic* in nature
- *One-stage random transmission of short payload* is much more efficient than two-stage designs
- For payload of around 200bits twice the no. of devices can be supported by one–stage strategies, (Dhillon & Huang, 2013)
- Smartphone traffic is sporadic (fast dormancy)
- LTE has no physical channel for short and sporadic messages; proposed for 5G
- Theoretical challenges: understand **noncoherent compressive reception** of sporadic signals

Motivation

Compressed Sensing in Wireless Communication

- Compressibility is relevant issue in future wireless communication, but has not yet sufficiently exploited
  1. user/device distribution (near/far)
  2. data payload sensor application
  3. multipath channels are sparse in delay and Doppler
  4. multiple antenna, correlation (low rank) due to collocation
  5. user activity

- Non-adaptivity is a also desired feature in wireless communication
  1. feedback protocols
  2. simple relaying without overhead
  3. universal sampling and data acquisition

Outline

- A blind compressed sensing problem
- Main result and the approach
- Random demodulation and universality
- Application: Sporadic communication and compressed blind deconvolution

Motivation

Compressed Sensing in Wireless Communication

A typical problem: data y, compressible a known domain, is transmitted through an unknown doubly–dispersive mobile communication channel. Discrete formulation of the signal at the channel output looks like:

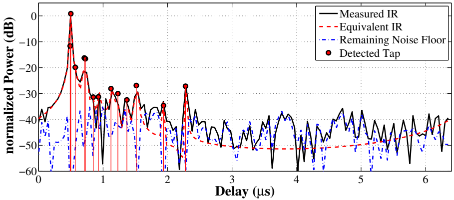
$$\sum_{(k,\bar{k})\in I} x_{(k,\bar{k})} y_{l-k} e^{i2\pi l\bar{k}/n}$$

whereby channel coefficients $\{x_{(k,\bar{k})}\}$ are supported in "small box I" but, in addition, are compressible in a known domain.

At moderate mobility the dominating effect is single–dispersive, i.e. the convolution:

$$(x * y)_l = \sum_k x_k y_{l-k}$$

source: S. Jaeckel, Heinrich-Hertz-Institute (Jaeckel *et al.* , 2009)



Goals: From a single or a few transmission frames...

1. (multi–user): *identify and separate multiple contributions of type $x * y$*
2. *estimate data y* and *channel x*

- Common model: $\Phi \in \mathbb{C}^{m \times n}$ and

$$b = \Phi \Psi y + e$$

  basis $\Psi \in \mathbb{C}^{n \times n}$ is assumed to be known *perfectly*.

- $\Psi + E$ (Herman & Strohmer, 2010)
- However, what happens if Ψ is *only known to have a compressible representation* $\Psi = \sum_i x_i \Psi_i$ such that, for example, x is sparse ?

$$b = \Phi(\sum_i x_i \Psi_i) y + e =: \Phi B(x, y) + e$$

Questions:

- Universal sensing for instantaneous decoding data frame without knowing the model x ?
- Universal sensing, storage and relaying for later decoding of multiple data frames at sink possibly knowing already the model ?

- In such a model the mapping B is bilinear:

$$b = \Phi B(x, y) + e = \Phi B(x \otimes y) + e \tag{1}$$

- $x \otimes y \in \Sigma_s \otimes \Sigma_f \Rightarrow \text{vec}(x \otimes y) \in \Sigma_{sf}$, depending on ΦB, $\mathcal{O}(sf \log n)$ measurements.
- But, only $s + f$ independent parameters, i.e. when $m = \mathcal{O}((s + f) \log n)$ is possible ?
- x or y is known before sampling $\rightarrow$ no. of measurements can be reduced
- x or y is known after sampling $\rightarrow$ conventional CS algorithms, but sampling has still to cope with all sparse bilinear–combinations.

Low–Dimensional Embedding

RIP–like Conditions on B

- B should be well–conditioned on all sparse rank-one matrices $u \in \Sigma_s \otimes \Sigma_f$,

$$\alpha\|u\| \leq \|B(u)\| \leq \beta\|u\| \tag{2}$$

since $\|x \otimes y\| = \|x\| \cdot \|y\|$ called *restricted norm multiplicativity* (RNMP) in (Walk & PJ, 2012).

RNMP is sufficient for distinguishing special rank–one cases, like for example:

1. $x \otimes y_1 - x \otimes y_2 = x \otimes (y_1 - y_2)$
2. $x \otimes x - y \otimes y = (x + y) \otimes (x - y)$ - relevant for phase retrieval (Walk & PJ, 2013)

- B should be well–conditioned for $u \in \Sigma_s \otimes \Sigma_f - \Sigma_s \otimes \Sigma_f$,

$$\alpha\|u\| \leq \|B(u)\| \leq \beta\|u\| \tag{3}$$

Bi–Lipschitz on $\Sigma_s \otimes \Sigma_f$:

Note that, in general, $\alpha = \alpha(s, f, \text{ambient dimensions})$ (same for β)

Theorem

Let $B : \mathbb{C}^{n\times n} \to \mathbb{C}^n$ *linear and there exist* (α, β) *independent of* n *such that:*

$$\alpha\|u\| \leq \|B(u)\| \leq \beta\|u\| \quad \text{for all} \quad u \in U$$

where (i) $U = \Sigma_s^n \otimes \Sigma_f^n$ *or (ii)* $U = \Sigma_s^n \otimes \Sigma_f^n - \Sigma_s^n \otimes \Sigma_f^n$. *Let* $\Phi \in \mathbb{C}^{m\times n}$ *with* $\Pr(\{|\|\Phi r\| - \|r\|| \leq \frac{\delta}{2}\|r\|\}) \geq 1 - e^{-c\delta^2 m}$ *for any* $r \in V = B(U)$.

There are constants $c', c'' > 0$ *(depending on* (α, β)*) such that if:*

$$m \geq c''\delta^{-2}(s+f)\log(n/\min(s,f))$$

there holds:

$$\Pr(\{\forall v \in V \;:\; |\|\Phi v\| - \|v\|| \leq \delta\|v\|\}) \geq 1 - e^{-c'\delta^2 m}$$

- case (i) (Walk&PJ, 2012) .
- (α, β) independent of n can be achieved for case (i) and $B(x, y) = x * y$ (Walk&PJ, 2013)

Low–Dimensional Embedding

Adapt arguments of (Baraniuk *et al.* , 2008) for $V \subseteq \mathbb{C}^n$ with low–dimensional structure

- Usually $v - r \notin V$ for generic $v, r \in V \Rightarrow$ intrinsic description required
- *compact manifolds* (Baraniuk & Wakin, 2009)
- **Here:** generic decompositions $\{v_i\} \subset V$ with $v - r = \sum_i v_i$. *Intrinsic distance* $d(v, r)$:

$$\|v - r\| \leq \sum_i \|v_i\| =: d(v, r)$$

remark: *projective/atomic norm*: $\inf\{\sum_i \|v_i\| \, : \, v - r = \sum_i v_i \text{ with } v_i \in V\}$

Lemma

Let V be a subset of a finite normed space and $d : V \times V \to \mathbb{R}$ be an intrinsic distance. Let $\Phi : V \to W$ be a linear map and $r \in V$ such that $|\|\Phi r\| - \|r\|| \leq \frac{\delta}{2}\|r\|$. Then for all $v \in V$ with $d(v, r) \leq \epsilon\|v\|$ it holds:

$$|\|\Phi v\| - \|v\|| \leq \delta\|v\|$$

if $\epsilon < \delta/7$ ($\epsilon < \delta/4$ if $\|v\| = \|r\|$).

We assume $B : U \to V$ with the following two conditions:

$(\triangle^\sigma)$ differences in U can be *intrinsically* decomposed as $u - \rho = \sum_{i=1}^{k} u_i$ with $u_i \in U$ and:

$$\sum_{i=1}^{k} \|u_i\| \leq \sigma \| \sum_{i=1}^{k} u_i \|$$

with σ independent of ambient dimensions. For U=union of sets - assume here u, ρ in the same subset

$(\frac{\beta}{\alpha})$ that B has the property that exist $0 < \alpha \leq \beta < \infty$ such that for all $u \in U$:

$$\alpha\|u\| \leq \|B(u)\| \leq \beta\|u\|$$

"B(ϵ–coverings in U) are $\frac{\beta\sigma\epsilon}{\alpha}$–coverings in V in $\|\cdot\|$ *and the intrinsic distance* d " . . .

Lemma

Let $B : U \to V$ *be a linear map with properties* $(\triangle^\sigma)$ *and* $(\frac{\beta}{\alpha})$. *For any* $u, \rho \in U$ *with* $\|u - \rho\| \leq \epsilon\|u\|$ *it follows*

$$\|v - r\| \leq d(v, r) \leq \frac{\beta\sigma}{\alpha}\epsilon\|v\|$$

where $v = B(u)$ *and* $r = B(\rho)$.

- ϵ–nets $\{\rho\} \subset U$ for the $\|\cdot\|$–unit ball in U with minimal cardinality $\Rightarrow$ covering number $N_\epsilon(U)$
- (metric) entropy as: $H_\epsilon(U) := \log N_\epsilon(U)$
- We have: $H_\epsilon(V) \leq H_{\frac{\alpha\epsilon}{\beta\sigma}}(U)$

Theorem (PJ and P. Walk)

Let $\Phi : V \to W$ be a random linear map which obeys $\Pr(\{|\|\Phi r\| - \|r\|| \leq \frac{\delta}{2}\|r\|\}) \geq 1 - e^{-\gamma}$ *uniformly in $r \in V$. Let $V = B(U)$ for a linear map B. If conditions $(\triangle^\sigma)$ and $(\frac{\beta}{\alpha})$ are fulfilled there holds:*

$$\Pr(\{\forall v \in V \ : \ |\|\Phi v\| - \|v\|| \leq \delta\|v\|\}) \geq 1 - e^{-(\gamma - H_\epsilon(U))} \tag{4}$$

for $\epsilon < \frac{\alpha}{7\beta\sigma}\delta$.

- *Anisotropic situation*: measurements are due to $\Phi B : U \to W$. With $\alpha = (1-\eta)\xi$ and $\beta = (1+\eta)\xi$:

$$|\|\Phi B(u)\| - \xi\|u\|| \leq \xi(\delta + \eta(\delta + 1))\|u\| =: \xi\delta'\|u\| \tag{5}$$

- According to (Cai & Zhang, 2013) $\delta' < \frac{1}{3}$ is sufficient (and in a certain sense necessary) for recovery via ℓ_2–constrained ℓ_1–minimization ($U = \Sigma_{2k}$) or $\|\cdot\|_*$–minimization ($U = M_{2\kappa}$)

Random Low–Dimensional Embedding

- Practically important are structured measurements Φ with fast implementation (like FFT).
- From (Krahmer & Ward, 2011) follows that classical (k, δ_k)–RIP matrices Φ with additional column–sign randomization $D_\xi = \text{diag}(\xi)$ are JL–embeddings for $k = \mathcal{O}(H_\epsilon(U))$,

Corollary

Let Φ be (k, δ_k)–RIP with $\delta_k \leq \delta/8$ and $k \geq 40(\rho + H_\epsilon(U) + 3\log(2))$. Then

$$\Pr(\{\forall v \in V \,:\, |\|\Phi D_\xi v\| - \|v\|| \leq \delta\|v\|\}) \geq 1 - e^{-\rho} \tag{6}$$

where $\epsilon < \frac{\alpha}{7\beta\sigma}\delta$.

- Important consequences for the *random demodulator* concept

Let $B : U \to V$

- k–sparse signals $U = \Sigma_k^n - \Sigma_k^n \quad \to \quad \boxed{H_\epsilon \leq d\log(3/\epsilon) + d\log(en/d)}$ with $d = 2k$.
  Union of subspaces $\to \sigma = 1$.
- rank–κ matrices $U = M_\kappa^n - M_\kappa^n \to \boxed{H_\epsilon \leq (2n+1)2\kappa\log(9/\epsilon)}$ (Candes & Plan, 2011).
  rank-2κ matrix $u - \rho = u_1 + u_2$ has rank–κ orthogonal decomposition $\Rightarrow \sigma = \sqrt{2}$.
- Simultaneous (s,f)–sparse and κ–rank matrices $U = M^n_{(\kappa,s,f)} - M^n_{(\kappa,s,f)}$

$$\boxed{\mathrm{H}_\epsilon \leq (d(s+f)+1)d\log 9/\epsilon + d(s+f)\log\frac{e\cdot n}{d\min(s,f)}}$$

  with $d = 2\kappa$.

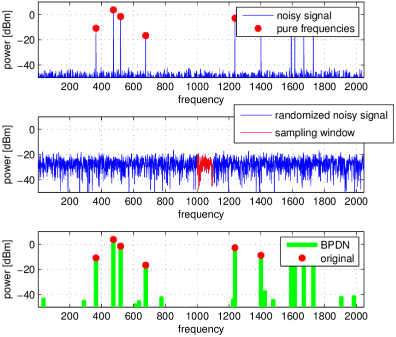
For $d \in \{1,2\}$ this is the bilinear case $B(x,y) = B(x \otimes y)$.

Random Demodulator

Idea

- Structural measurements are required to reduce signal processing requirements and provide efficient implementations

- random convolution $\Phi = P_{\Omega} \hat{D}_{\eta}$ (Romberg, 2009) for random Ω $\Rightarrow$ *universal*.
- *random demodulator* $\Phi = P_{\Omega} \hat{D}_{\eta} F$ (Tropp & Laska, 2010) for frequency–sparse signals $\Rightarrow$ *non–universal*
- *partial circulant matrix*, $m = \mathcal{O}((k \log n)^{3/2})$ (Rauhut *et al.* , 2012) for fixed Ω $\Rightarrow$ *non–universal*
- $m = \mathcal{O}(k(\log k \log n)^2)$, (Krahmer *et al.* , 2012) for fixed Ω $\Rightarrow$ *non–universal*



Achieving universality for fixed Ω ? . . .

Random Demodulator

Theorem

Consider random $m \times n$–matrices $\Phi = \frac{1}{|\Omega|} P_\Omega \hat{D}_\eta D_\xi$ with ξ and η being for example i.i.d. Rademacher sequences. If

$$m \geq 64c\delta^{-2}(\lambda + h_\epsilon) \max((\log(\lambda + h_\epsilon) \log(n))^2, \lambda + \log(2))$$

where $h_\epsilon = H_\epsilon(U) + 4\log(2)$. Then

$$\Pr(\{\forall v \in V \,:\, |\|\Phi v\| - \|v\|| \leq \delta \|v\|\}) \geq 1 - e^{-\lambda}$$

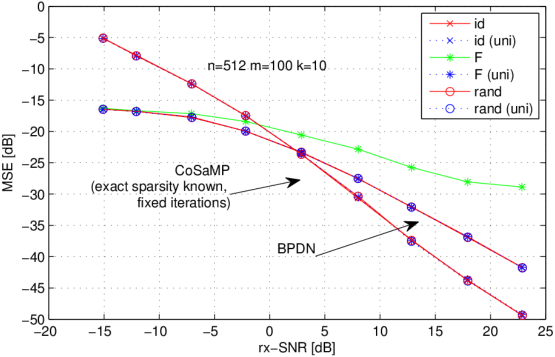
- Thus, $\Phi = \frac{1}{|\Omega|} P_\Omega \hat{D}_\eta D_\xi$ is **universal sampling scheme** (with entropy $2n$)

Universal Random Demodulator

Example

Here is an example for the conventional sparse case, $V = \Sigma_k^n - \Sigma_k^n$

- domain:
  $U \in \{\text{Id}, F, \text{random unitary}\}$
- conventional $P_\Omega \hat{D}_\eta$
- universal $P_\Omega \hat{D}_\eta D_\xi$



Open problem: optimal recovery asymptotics $(s+f)\log n$ can not be achieved with conventional convex methods (Oymak *et al.* , 2012), more precisely:

$$\hat{\rho} = \arg\min \|\rho\|_* + \lambda \cdot \text{sparsification} \quad \text{s.t.} \quad \|\Phi B(\rho) - b\|_2 \leq \epsilon$$

failure with $\geq 1 - e^{-c'm}$ for sparsification:

- $\|\rho\|_{1,1}$ and if $m \leq c \cdot \min(sf, n)$
- $\|\rho\|_{1,2} + \|\rho\|_{2,1}$ and if $m \leq c \cdot \min(s, f) \cdot n$

Sparse power factorization (Lee *et al.* , 2013)

Application to Sporadic Communication

Overview, Single–User

Goals: Estimate L data symbols $y = \{y_i\}_{i=1}^{L}$ and s–sparse channel coefficients $x = \{x_i\}_{i=1}^{T}$ instantaneously from m received samples b. The single user model is:

$$b = \Phi B(x, Cy) + e = \Phi B(1 \otimes C)(x \otimes y) + e$$

where a precoding via $C \in \mathbb{C}^{n \times L}$ is used.

- Blind deconvolution with random C as recovery of rank–one $\rho = x \otimes y$ proposed and numerically demonstrated in (Asif *et al.* , 2009).
- $A = B(1 \otimes C)$ for Gaussian C and $B(x, y) = x * y$ (Ahmed *et al.* , 2012)

  $$\text{If } n \geq c_a \cdot L \log(TL) \log(L) \;\Rightarrow\; A^*A|_{\text{rank–two}} \quad \text{is well conditioned}$$

  with probability $\geq 1 - 3(TL)^{-a}$ for any fixed $a \geq 1$ and $c_a = \mathcal{O}(a)$.
- Example: convex program for blind recovery of sparse channel x and non–sparse data y can be casted as:

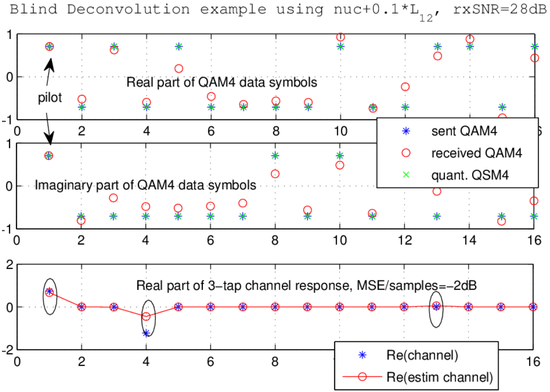
  $$\min \|\rho\|_* + \lambda \|\rho\|_{\ell_{1,2}} \quad \text{s.t.} \quad \|\Phi B(1 \otimes C)\rho - b\|_{\ell_2} \leq \epsilon$$

Application to Sporadic Communication

Blind Deconvolution, Single–User

Here is a small QAM4–example:

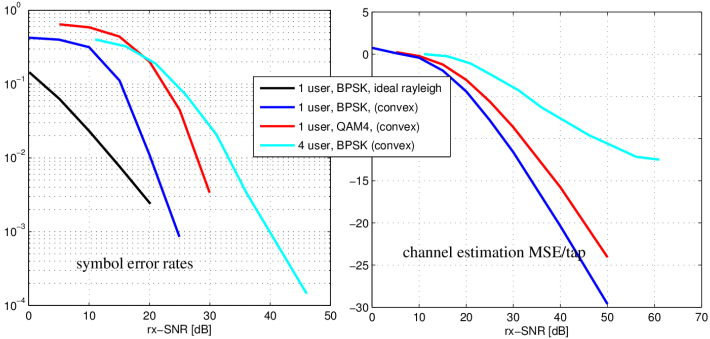
- data $y \in c \cdot \{\pm 1 \pm i\}^{16}$
- sparse channel $x \in \Sigma_3^{16}$
- random $\begin{pmatrix} C \\ 0 \end{pmatrix} \in \mathbb{C}^{2 \cdot 128 \times 16}$
- $\Phi = P_\Omega \hat{D}_\eta D_\xi \in \mathbb{C}^{96 \times 256}$
- receive SNR=28dB



Application to Sporadic Communication

Blind Deconvolution, Single–User

Here is a full BPSK&QAM4–example:

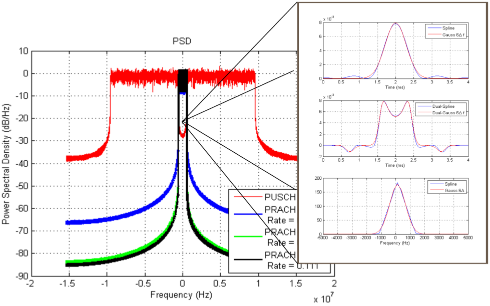


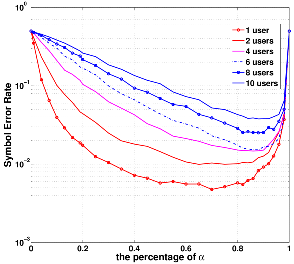
Application to Sporadic Communication

Compressive Random Access Ideas for 5G

* joint work with Gerhard Wunder, Fraunhofer Heinrich–Hertz Institute Berlin, 5GNOW EU-project

$$b = \Phi \sum_{k=1}^{K} B(x_k, y_k + p_k) + e \tag{7}$$





Conclusions

- Compressed sensing with sparse model uncertainty
- casted as anisotropic matrix recovery problem
- $m = \mathcal{O}((s+f)\log n)$ if B is well–conditioned on rank–1&2
- optimal recovery program still open
- compressive blind deconvolution via random demodulation
- application to sporadic communication

Thank You

Ahmed, Ali, Recht, Benjamin, & Romberg, Justin. 2012.
Blind deconvolution using convex programming.
arXiv preprint arXiv:1211.5608.

Asif, MS, Mantzel, W., & Romberg, J. 2009.
Random channel coding and blind deconvolution.
Allerton.

Baraniuk, R., & Wakin, M. 2009.
Random projections of smooth manifolds.
Foundations of Computational Mathematics, **9**(1), 51–77.

Baraniuk, R., Davenport, M., DeVore, R. A., & Wakin, M. 2008.
A Simple Proof of the Restricted Isometry Property for Random Matrices.
Constructive Approximation, **28**(3), 253–263.

Cai, T. Tony, & Zhang, Anru. 2013.
Sharp RIP bound for sparse signal and low-rank matrix recovery.
Applied and Computational Harmonic Analysis, **35**(1), 74–93.

Candes, EJ, & Plan, Y. 2011.
Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements.
Information Theory, IEEE Transactions on, 1–30.

Dhillon, HS, & Huang, HC. 2013.
Fundamentals of Throughput Maximization with Random Arrivals for M2M Communications.
arXiv preprint arXiv: . . . .

Herman, MA, & Strohmer, T. 2010.
General deviants: An analysis of perturbations in compressed sensing.
Selected Topics in Signal Processing, IEEE Journal, **4**(2).

Jaeckel, S, Jiang, L, Jungnickel, V, Thiele, L, Jandura, C, Sommerkorn, G, & Schneider, C. 2009.
Correlation properties of large and small-scale parameters from multicell channel measurements.
Proc. EUCAP '09, 3406–3410.

Krahmer, Felix, & Ward, Rachel. 2011.
New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property.
Tech. rept. 2.

Krahmer, Felix, Mendelson, Shahar, & Rauhut, Holger. 2012.
Suprema of chaos processes and the restricted isometry property.
arXiv preprint arXiv:1207.0235, 1–23.

Lee, Kiryung, Wu, Yihong, & Bresler, Yoram. 2013.
Near Optimal Compressed Sensing of Sparse Rank-One Matrices via Sparse Power Factorization.
1–41.

Oymak, Samet, Jalali, Amin, Fazel, Maryam, Eldar, YC, & Hassibi, B. 2012.
Simultaneously structured models with application to sparse and low-rank matrices.
arXiv preprint arXiv: . . . .

Rauhut, Holger, Romberg, Justin, & Tropp, Joel a. 2012.
Restricted isometries for partial random circulant matrices.
Applied and Computational Harmonic Analysis, **32**(2), 242–254.

Romberg, Justin. 2009.
Compressive Sensing by Random Convolution.
SIAM Journal on Imaging Sciences, **2**(4), 1098.

Tropp, JA, & Laska, JN. 2010.
Beyond Nyquist: Efficient sampling of sparse bandlimited signals.
Trans. on Information Theory, **56**(1), 520–544.

Walk, P., & Jung, P. 2012.
Compressed Sensing on the Image of Bilinear Maps.
In: Information Theory, 2012. ISIT 2012. Proceedings. International Symposium on.

Walk, P, & Jung, P. 2013a.
A Stability Result for Sparse Convolutions.
arXiv preprint arXiv:1312.2222.

Walk, P, & Jung, P. 2013b.
Stable Recovery from the Magnitude of Symmetrized Fourier Measurements.
arXiv preprint arXiv:1310.5895.

Walk, Philipp, & Jung, Peter. 2013c.
On a Reverse l2âInequality for Sparse Circular Convolution.
In: 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP).