Stability and instance optimality for Gaussian measurements in compressed sensing*

P. Wojtaszczyk†
Institut of Applied Mathematics
University of Warsaw
ul. Banacha 2; 02-097 Warszawa
Poland

February 14, 2008

Abstract

In compressed sensing we seek to gain information about vector $x \in \mathbb{R}^N$ from $d << N$ nonadaptive linear measurements. Candes, Donoho, Tao et. al. (see e.g. [2, 4, 8]) proposed to seek good approximation to x via ℓ_1 minimisation. In this paper we show that in the case of Gaussian measurements it recovers the signal well from inacurate measurements, thus improving result from [4]. We also show that with big probability it gives information comparable with best k term approximation in euclidean norm, $k \sim d/\ln N$. This provides the first numerically friendly algorithm to do so, see [7].

1 Introduction

Compressed sensing is a new scheme which shows that some signals can be reconstructed from fewer measurements that previously were considered necessary. The mathematical formulation is the following. Our signal is a vector $x \in \mathbb{R}^N$. We have a $N \times d$ matrix Φ called *measurement matrix* and our measurements are represented by $y = \Phi(x) \in \mathbb{R}^d$. We also need a decoder Δ (which maybe non-linear) which produces $\Delta(y) \in \mathbb{R}^N$ which should be an approximation to x. The main point in compressed sensing as expressed in recent papers is that it is actually possible to recover the essential information about x from relatively few non-adaptive measurements $d << N$. Substantial progress have been made in

*Mathematics Subject Classification 68P30, 68W20, 41A46

†Author would like to express his gratitude to Piotr Mankiewicz for answering questions about convex geometry and to Albert Cohen, Wolfgang Dahmen and Ron DeVore for teaching him compressed sensing. Thanks are also due to Rachel Ward for interesting discussion about stability results. This research was made possible by EC Marie Curie ToK program SPADE-2 at IMPAN

recent years in understanding the performance of various measurement matrices Φ and decoders Δ. Generally we have also an integer $k \leq d$ which measures the amount of information we wish to recover. The standard initial requirement is that for every k–sparse vector (i.e. $x \in \Sigma_k$) we have $\Delta(\Phi(x)) = x$. This clearly forces $\Phi|\Sigma_k$ to be one to one. But foe Δ to be numerically friendly we must have the corresponding systems of equations well conditioned. This leads to the restricted isometry property RIP (also called in the literature uniform uncertainity property – UUP). This was introduced in [6, 5].

By Σ_μ we will mean the set of all vectors from $\mathbb{R}^s$ (where s should be clear from the context) which have at most μ non-zero coefficients. We say that matrix Φ satisfies RIP(k, δ) where $0 < \delta < 1$ and $k \in \mathbb{N}$ if

$$(1-\delta)\|c\|_2 \leq \|\Phi(c)\|_2 \leq (1+\delta)\|c\|_2 \tag{1}$$

for all vectors $c \in \Sigma_k$. Results of [6, 8] (see also Theorem 2.2) imply in particular that when Φ satisfies RIP$(2k, 0.4)$ then *every* $x \in \Sigma_k$ can be exactly recovered with

$$\Delta_1(x) = \text{Argmin}\{\|z\|_1 \; : \; \Phi(x) = \Phi(z)\}. \tag{2}$$

This is a numerically tractable decoder.

Current arguments for existence of RIP matrices with optimal bounds are probabilistic. This is a line of arguments present in Banach space theory and in approximation theory from the 70's. For a streamlined presentation of the proof and historical comments see [1]. More general results are in [12, 13]. If Φ_ω is either independent gaussian matrix or its columns are drawn independently from the uniform distribution on the unit sphere in $\mathbb{R}^d$ then there exists constants $c_1, c_2 > 0$ depending only on δ such that matrix Φ_ω satisfies (1) for any $k \leq c_1 d/\ln(N/k)$ with probability $\geq 1 - \exp -c_2 d$. Let us note that even when we have a concrete relatively big matrix it is practically impossible to check if it satisfies RIP.

Clearly there are various ways to evaluate the efficiency of given measurement–decoder pair (Φ, Δ). In [7] A. Cohen, W. Dahmen and R. DeVore proposed to look for *instance optimality* i.e. they wanted a constant C such that

$$\|x - \Delta(\Phi x)\|_2 \leq C_0 \sigma_k^2(x) \tag{3}$$

for all $x \in \mathbb{R}^N$ where

$$\sigma_\mu^2(x) = \inf\{\|x - v\|_2 \; : \; v \in \Sigma_\mu\} \tag{4}$$

Actually they proved that instance optimality with $k = 1$ forces d to be proportional with N which is not what we look for, so they suggested to consider instance optimality in probability. They consider a random matrix Φ_ω and ask for the following: for any $x \in \mathbb{R}^N$ the inequality

$$\|x - \Delta(\Phi_\omega x)\|_2 \leq C_0 \sigma_k^2(x) \tag{5}$$

holds for this particular x with high probability. It is important that the set of ω's where (5) holds depend on x.

One of the aims of this paper is to point out the first numerically feasible measurement matrix and decoder for which instance optimality in probability holds. To do this we introduce a new property LQ(α) of a measurement matrix. We note (it is basically a know fact in Banach space theory) that LQ($\mu/\sqrt{k}$) is satisfied with great probability by random gaussian matrices as explained in section 4. So as a corollary we prove that if Φ_ω is either independent gaussian matrix or its columns are drawn independently from the uniform distribution on the unit sphere in $\mathbb{R}^d$ then the decoder Δ_1 (see (2)) gives instance optimality in probability. For measurement matrices considered here it answers the question formulated in [7]. We also discuss the case when our measuremant is corrupted by noise i.e. we apply the decoder to vector $\Phi_\omega(x)+e$. We prove that LQ($\mu/\sqrt{k}$) gives very good stability result in this setting. An interesting feature of our Theorem 3.3 is that it proves stability for the decoder Δ_1. Similar results were proved in [4] but there the decoder had to be modified according to our estimate of the noise magnitude.

2 Preliminaries

In the space $\mathbb{R}^\mu$ we will consider two norms: the usual euclidean norm $\|.\|_2$ and ℓ_1 norm $\|.\|_1$. Unit balls in $\mathbb{R}^\mu$ in those norms will be denoted by B_2^μ and B_1^μ respectively. By $\sigma_\mu^1(x)$ we will denote the error of best μ–term approximation in ℓ_1 norm defined as in (4) but with 2 replaced by 1.

Definition 2.1. *We say that a matrix $\Phi : \mathbb{R}^N \to \mathbb{R}^d$ satisfies ℓ_1 quotient property with constant $\alpha > 0$ (LQ(α) for short) if $\Phi(B_1^N) \supset \alpha B_2^d$.*

A fundemental role in our considerations will be played by the following Theorem of E.Candes, J.Romberg and T.Tao [4]:

Theorem 2.2. *Suppose the matrix Φ satisfies RIP($2k, \delta$) with $\delta < \sqrt{2} - 1$. Then there exists a constant C such that*

$$\|x - \Delta_1(\Phi x)\|_2 \leq \frac{C\sigma_k^1(x)}{\sqrt{k}} \tag{6}$$

for all $x \in \mathbb{R}^N$.

This Theorem was formulated without proof in [3]. It appeared for the first time in [4] but with the assumption that Φ satisfies RIP($3k, \delta_1$) and RIP($4k, \delta_2$) and $\delta_1 + 3\delta_2 < 2$.

Motivated by this Theorem and following [7] we will say that the pair (Φ, Δ) is $(2, 1)$ instance optimal if there exists a constant K such that

$$\|x - \Delta(\Phi x)\|_2 \leq \frac{K\sigma_k^1(x)}{\sqrt{k}} \tag{7}$$

for all $x \in \mathbb{R}^N$.

Our notation is rather standard. Note however that if we have a vector in $\mathbb{R}^\mu$ and a subset S of the index set which we assume to be $\{1, 2, \ldots, \mu\}$ then $x|S$ denotes the vector which coincides with x on S and has zero coordinates outside S.

3 Stability results

We begin this section with some general deterministic results for measurement matrices which satisfie *both* RIP and LQ. Any decoder aims at recovering signal $x \in \mathbb{R}^N$ using only vector $y = \Phi(x)$ so our decoding does not distinguish vectors $x_1, x_2 \in \mathbb{R}^N$ such that $\Phi(x_1) = \Phi(x_2)$. However in the proofs we may exploit some apriori properties of x. Thus it is natural to try for the sake of the proof to modify our unknown x so it will have some extra properties. In the following Lemma we use LQ property to give such a modification.

Lemma 3.1. *Suppose that matrix Φ satisfies RIP(k, δ) and LQ$(\mu/\sqrt{k})$. For every $x \in \mathbb{R}^N$ there exists $\tilde{x} \in \mathbb{R}^N$ such that*

$$\Phi(x) = \Phi(\tilde{x}) \tag{8}$$

$$\|\tilde{x}\|_1 \leq \frac{\sqrt{k}}{\mu}\|\Phi(x)\|_2 \tag{9}$$

$$\|\tilde{x}\|_2 \leq C(\delta, \mu)\|\Phi(x)\|_2. \tag{10}$$

Proof. Vector $\tilde{x}$ satisfying (8) and (9) we get directly from LQ condition. To estimate $\|\tilde{x}\|_2$ we split the set $\{1, 2, \ldots, N\}$ into disjoint k–element sets $S_0, S_1, \ldots$ such that $|\tilde{x}_j| \geq |\tilde{x}_l|$ whenever $j \in S_\nu$ and $l \in S_{\nu+1}$. Clearly we have

$$\|\tilde{x}|S_{\nu+1}\|_2 \leq \frac{1}{\sqrt{k}}\|\tilde{x}|S_\nu\|_1. \tag{11}$$

From (9) and (11) we get

$$\|\tilde{x}|S_0^c\|_2 \leq \sum_{\nu=1} \|\tilde{x}|S_\nu\|_2 \leq \frac{1}{\sqrt{k}}\|\tilde{x}\|_1 \leq \frac{\|\Phi(x)\|_2}{\mu}. \tag{12}$$

Also using (12) and RIP condition we get

$$\|\Phi(\tilde{x}|S_0^c)\|_2 \leq \sum_{\nu=1} \|\Phi(\tilde{x}|S_\nu)\|_2 \leq \frac{1}{1-\delta}\sum_{\nu=1} \|\tilde{x}|S_\nu\|_2 \leq \frac{\|\Phi(x)\|_2}{(1-\delta)\mu}. \tag{13}$$

Now we use RIP condition and (13) to get

$$\begin{aligned} \|\tilde{x}|S_0\|_2 &\leq \frac{1}{1-\delta}\|\Phi(\tilde{x}|S_0)\|_2 = \frac{1}{1-\delta}\|\Phi(\tilde{x}) - \Phi(\tilde{x}|S_0^c)\|_2 \\ &\leq \frac{1}{1-\delta}\left(\|\Phi(x)\|_2 + \|\Phi(\tilde{x}|S_0^c)\|_2\right) \\ &\leq \frac{1}{1-\delta}\left(\|\Phi(x)\|_2 + \frac{\|\Phi(x)\|_2}{(1-\delta)\mu}\right). \end{aligned}$$

which together with (12) gives (10)

$\square$

Remark 3.1 Spliting of the support of $\tilde{x}$ into k-elements blocks of decreasing coefficients was used in [4] and also in [7]. The above argument shows that we can take

$$C(\delta,\mu)=\frac{1}{\mu}+\frac{1+\mu(1-\delta)}{\mu(1-\delta)^2}.$$

Now let us state the following abstract result

Theorem 3.2. *Suppose that the matrix Φ satisfies RIP(k,δ) and LQ$(\mu/\sqrt{k})$. If the pair (Φ,Δ) is $(2,1)$ instance optimal with constant K then*

1. *For any $x\in\mathbb{R}^N$ and any $r\in\mathbb{R}^d$*

$$\|\Delta(\Phi(x)+r)-x\|_2\leq C\Big(\|r\|_2+\frac{\sigma_k^1(x)}{\sqrt{k}}\Big). \tag{14}$$

2. *For any $x\in\mathbb{R}^N$*

$$\|\Delta\Phi(x)-x\|_2\leq C\Big(\sigma_k^2(x)+\|\Phi(x|S^c)\|_2\Big) \tag{15}$$

where S is a k–elements set such that $\|x|S^c\|_2=\sigma_k^2(x)$.

3. *For any $x\in\mathbb{R}^N$ and any $r\in\mathbb{R}^d$*

$$\|\Delta(\Phi(x)+r)-x\|_2\leq C\Big(\|r\|_2+\sigma_k^2(x)+\|\Phi(x|S^c)\|_2\Big) \tag{16}$$

where S is a k–elements set such that $\|x|S^c\|_2=\sigma_k^2(x)$.

Proof. Let us start with the proof of (14). From LQ we infer that there exists $z\in\mathbb{R}^N$ such that $\Phi(z)=r$. From Lemma 3.1 we infer that we can choose z such that $\|z\|_1\leq\frac{\sqrt{k}}{\mu}\|r\|_2$ and $\|z\|_2\leq C_1\|r\|_2$. Since $\Phi(x+z)=\Phi(x)+r$ from $(2,1)$–instance optimality we get

$$\|\Delta(\Phi(x)+r)-(x+z)\|_2\leq K\frac{\sigma_k^1(x+z)}{\sqrt{k}}$$

so

$$\begin{aligned}\|\Delta(\Phi(x)+r)-x\|_2 &\leq \|z\|_2+K\frac{\sigma_k^1(x+z)}{\sqrt{k}}\\ &\leq C_1\|r\|_2+K\frac{\sigma_k^1(x)+\|z\|_1}{\sqrt{k}}\\ &\leq \Big(C_1+\frac{K}{\mu}\Big)\|r\|_2+K\frac{\sigma_k^1(x)}{\sqrt{k}}.\end{aligned}$$

Now we prove (16) which for $r = 0$ becomes (15). Let us fix a k–element set S such that $\sigma_k^2(x) = \|x|S^c\|_2$. From Lemma 3.1 like above we get $v, z \in \mathbb{R}^N$ such that

$$\begin{array}{ccc} \Phi(v) = r & \|v\|_1 \leq \frac{\sqrt{k}}{\mu}\|r\|_2 & \|v\|_2 \leq C\|r\|_2 \\ \Phi(z) = \Phi(x|S^c) & \|z\|_1 \leq \frac{\sqrt{k}}{\mu}\|\Phi(x|S^c)\|_2 & \|z\|_2 \leq C\|\Phi(x|S^c)\|_2 \end{array}$$

Clearly $\Phi(x|S + z + v) = \Phi(x) + r$ so $(2, 1)$ instance optimality gives

$$\|\Delta(\Phi(x) + r) - (x|S + z + v)\|_2 \leq K\frac{\sigma_k^1(x|S + z + v)}{\sqrt{k}}$$

so we get

$$\begin{aligned} \|\Delta(\Phi(x) + r) - x\|_2 &\leq \|x|S^c - z - v\|_2 + K\frac{\sigma_k^1(x|S + z + v)}{\sqrt{k}} \\ &\leq \sigma_k^2(x) + \|z\|_2 + \|v\|_2 + K\frac{\|z\|_1 + \|v\|_1}{\sqrt{k}} \\ &\leq \sigma_k^2(x) + C'\|r\|_2 + C''\|\Phi(x|S^c)\|_2 \end{aligned}$$

□

Remark 3.2 Note that we may derive (15) directly applying (14) for $x|S$ and $r = \Phi(x|S^c)$ like we do to prove Corollary 3.5. This shows formally that if a pair (Φ, Δ) satisfies (14) it also satisfies (15).

Let us observe that assumptions in Lemma 3.1 and Theorem 3.2 are somewhat contradictory: if we have $\mathrm{RIP}(k, \delta)$ then we have $\mathrm{RIP}(l, \delta)$ for all $l \leq k$ while if we have our assumption about LQ for k we have it also for all $l \geq k$. Thus at best we are dealing with k in certain range. Observe however that if we have (14) for some k it also holds for $k' \leq k$.

For independent Gaussian or uniform on the sphere measurement ensemble Corollary 4.3 shows that the assumptions of the above Theorem 3.2 are satisfied with $k \sim d/\ln N$. Thus we have

Theorem 3.3. *Suppose that Φ_ω is either independent Gaussian or uniform on the sphere $N \times d$ measurement ensemble. There exists a $k_0 \sim d/\ln N$ such that for $k \leq k_0$ we have*

a) There exists a set Ω_1 with $\mathbb{P}(\Omega_1) \geq 1 - e^{-cd}$ such that for any $x \in \mathbb{R}^N$ and any $r \in \mathbb{R}^d$ and $k \leq k_0$ we have

$$\|\Delta_1(\Phi_\omega(x) + r) - x\|_2 \leq C\Big(\|r\|_2 + \frac{\sigma_k^1(x)}{\sqrt{k}}\Big) \tag{17}$$

for $\omega \in \Omega_1$.

b) *For any $x \in \mathbb{R}^N$ there exists a set $\Omega_1(x)$ with $\mathbb{P}(\Omega_1(x)) \geq 1 - e^{-cd}$ such that for any $r \in \mathbb{R}^d$ and any $k \leq k_0$ we have*

$$\|\Delta_1(\Phi_\omega(x) + r) - x\|_2 \leq C(\|r\|_2 + \sigma_k^2(x)) \tag{18}$$

for $\omega \in \Omega_1(x)$.

Proof. For k_0 statement *a*) follows directly from Theorems 2.2, Theorem 3.2 and Corollary 4.3. For $k \leq k_0$ it follows because $\frac{\sigma_k^1(x)}{\sqrt{k}}$ is a decreasing sequence in k. Analogously the second statement for k_0 follows from Theorems 2.2 and 3.2, Corollary 4.3 and Lemma 4.1[1.] aplied for $x|S^c$ where S is a k_0 element set such that $\sigma_{k_0}^2 = \|x|S^c\|_2$. For $k \leq k_0$ we use monotonicity of $\sigma_k^2(x)$. □

As one of our aims is to provide examples of instance optimality in probability let us discuss this aspect of recent results of D. Needell and R. Vershynin [14, 15]. They constructed a very interesting decoder which for consistency I will denote as Δ_{ROMP} based on a greedy type algorithm which they call ROMP. About this decoder they proved the following (see [15, Corollary 1.3] and remarks after it)

Theorem 3.4. *Suppose Φ satisfies RIP$(8k, \epsilon)$ with $\epsilon = 0.01/\sqrt{\log k}$. Then for $x \in \mathbb{R}^N$ and $r \in \mathbb{R}^d$ we have*

$$\|x - \Delta_{ROMP}(\Phi(x) + r)\|_2 \leq 160\sqrt{\log 2k}\Big(\|r\|_2 + \frac{\sigma_k^1(x)}{\sqrt{k}}\Big). \tag{19}$$

Now suppose we have an arbitrary vector $x \in \mathbb{R}^N$ and we fix a k-element set such that $\|x|S^c\|_2 = \sigma_k^2(x)$. Applying Theorem 3.4 for $x|S$ and treating $\Phi(x|S^c) + r$ as a measurement error we get

$$\|x|S - \Delta_{ROMP}(\Phi(x) + r)\|_2 \leq 160\sqrt{\log 2k}(\|r\|_2 + \|\Phi(x|S^c)\|_2)$$

which yields

$$\|x - \Delta_{ROMP}(\Phi(x) + r)\|_2 \leq 161\sqrt{\log 2k}(\|r\|_2 + \|\Phi(x|S^c)\|_2 + \sigma_k^2(x)).$$

Thus we get

Corollary 3.5. *Suppose Φ_ω is a random measurement ensemble, $\Phi_\omega : \mathbb{R}^N \to \mathbb{R}^d$. Assume also that*

1. *for each $x \in \mathbb{R}^N$ with big probability we have $\|\Phi_\omega(x)\|_2 \leq 2\|x\|_2$*
2. *k is such that with big probability Φ_ω satisfies RIP$(8k, \epsilon)$ with $\epsilon = \frac{0.01}{\sqrt{\log k}}$*

Then for each $x \in \mathbb{R}^N$ with big probability we have for any $r \in \mathbb{R}^d$

$$\|x - \Delta_{ROMP}(\Phi(x) + r)\|_2 \leq 161\sqrt{\log 2k}(\|r\|_2 + 3\sigma_k^2(x)). \tag{20}$$

This is almost instance optimality, the problem is the logarithmic factor in the estimate (20). Another disadvantage is the fact that in random setting to ensure RIP$(8k, \epsilon)$ we must choose d greater than $ck\sqrt{\log k}\log N$ which is slightly bigger then $ck \log N$ which is optimal for a fixed δ. The great advantage of this fact and generally of the Δ_{ROMP} decoder is that it applies to many random ensembles (c.f. [1]).

4 ℓ_1 quotient property

In this section we want to discuss property LQ for random matrices considered in compressed sensing literature. Let us start with the Gaussian ensemble. Let g be the standard, normalised Gaussian variable. The following estimate is elementary and classical

$$\left(1 - \frac{1}{\lambda^2}\right)\sqrt{\frac{2}{\pi}}\frac{1}{\lambda}\exp -\frac{\lambda^2}{2} \leq \mathbb{P}(|g| \geq \lambda) \leq \sqrt{\frac{2}{\pi}}\frac{1}{\lambda}\exp -\frac{\lambda^2}{2} \tag{21}$$

for each $\lambda > 0$.

By $\Phi(\omega)$ we denote an $N \times d$ matrix whose entries are $\left(\frac{1}{\sqrt{d}}g_{i,j}\right)_{i=1}^{d}{}_{j=1}^{N}$ where $g_{i,j}$ are independent standard normalised gaussian variables. Columns of this matrix denoted $(\Phi_j(\omega))_{j=1}^N$ are independent Gaussian vectors in $\mathbb{R}^d$. Another case which we will consider is the matrix $\tilde{\Phi}$ whose columns are normalised vectors $\tilde{\Phi}_j = \|\Phi_j(\omega)\|^{-1}\Phi_j(\omega)$. One easily sees that those columns are independent norm one vectors drawn from the uniform distribution on the unit sphere in $\mathbb{R}^d$. By Φ_ω we denote the operator from $\mathbb{R}^N$ into $\mathbb{R}^d$ given by $\Phi_\omega(x) = \sum_{j=1}^N x_j\Phi_j(\omega)$. Analogously $\tilde{\Phi}(\omega)(x) = \sum_{j=1}^N x_j\|\Phi_j(\omega)\|^{-1}\Phi_j(\omega)$.

For reference let me state the following well known facts

Lemma 4.1. *For for both those measurement ensemblies we have:*

1. *There exists a constant $c > 0$ such that for each $x \in \mathbb{R}^N$ there exists set Ω_2 with $\mathbb{P}(\Omega_2) \geq 1 - e^{-cd}$ and on this set we have $\|\Phi_\omega(x)\|_2 \leq 1.5\|x\|_2$. The same holds for $\tilde{\Phi}_\omega$.*
2. *There exists constants $c_1, c_2 > 0$ depending only on δ such that matrix Φ_ω satisfies RIP(k, δ) for any $k \leq c_1 d/\ln(N/k)$ with probability $\geq 1 - \exp -c_2 d$.*

The following result is basically known as a folklore example in Banach space theory cf. [9]. More general (but also requiring more sophisticated tools) proof is given in [9]. An argument for $N = d^2$ is given in [11]. In this situation for the sake of the reader we decided to present the selfcontained standard argument, c.f. [11].

Proposition 4.2. *Let $0 < \mu < \frac{1}{\sqrt{2}}$ and let $e^{Cd} \geq N \geq C_1 d(\ln d)^\xi$ for some $\xi > (1 - 2\mu^2)^{-1}$ and some constants $C, C_1 > 0$. There exists a constant $c > 0$*

such that the set Ω_μ *of those* ω*'s that* Φ_ω *satisfies* $LQ(\frac{\mu}{\sqrt{d}}\sqrt{\ln\frac{N}{d}})$ *i.e.*

$$\Phi_\omega(B_1^N) \supset \frac{\mu}{\sqrt{d}}\sqrt{\ln\frac{N}{d}}B_2^d. \tag{22}$$

has the probability $\geq 1 - \exp -cd$*. The same is true for* $\tilde{\Phi}_\omega$*.*

Proof. Let us start from the Gaussian case. From RIP we know that there exists a set Ω_1 with $\mathbb{P}(\Omega_1) \geq 1 - e^{-cd}$ such that $\|\Phi_j(\omega)\| \leq 2$ for $\omega \in \Omega_1$ and $j = 1, 2, \ldots, N$. To simplyfy the notation we put $\alpha = \frac{\mu}{\sqrt{d}}\sqrt{\ln\frac{N}{d}}$. Let Ω_0 denotes the set that (22) does not hold. If $\omega \in \Omega_0$ then there exists a vector $x \in \mathbb{R}^d$ with $\|x\| = 1$ such that

$$\sup_{j=1,\ldots,N} |\langle x, \Phi_j(\omega)\rangle| < \alpha \tag{23}$$

From properties of Gaussian variables we know that $\langle x, \Phi_j(\omega)\rangle$ has the same distribution as $\frac{g}{\sqrt{d}}$. Using independence and (21) we get

$$\mathbb{P}(\sup_{j=1,\ldots,N} |\langle x, \Phi_j(\omega)\rangle| < \alpha) = \left[\mathbb{P}(\frac{|g|}{\sqrt{d}} < \alpha)\right]^N \tag{24}$$

$$\leq \left[1 - (1 - \alpha^{-2}d^{-1})\frac{1}{\alpha\sqrt{d}}\sqrt{\frac{2}{\pi}}\exp -\frac{d\alpha^2}{2}\right]^N \tag{25}$$

Now let $\mathcal{N}$ be an $\frac{\alpha}{2}$–net in the unit sphere of $\mathbb{R}^d$. It is well known (see e.g. [10, Ch. 15 Prop. 1.3]) that we can find such a net with cardinality not bigger then $\left(\frac{6}{\alpha}\right)^d$. If $x_0 \in \mathcal{N}$ is such that $\|x - x_0\| \leq \frac{\alpha}{2}$ then for $\omega \in \Omega_1$ we have

$$\begin{aligned} |\langle x_0, \Phi_j(\omega)\rangle| &= |\langle x, \Phi_j(\omega)\rangle - \langle x - x_0, \Phi_j(\omega)\rangle| \\ &\leq |\langle x, \Phi_j(\omega)\rangle| + \|x - x_0\|\|\Phi_j(\omega)\| \leq 2\alpha \end{aligned}$$

so if there exists x as in (23) then there exists $x_0 \in \mathcal{N}$ such that

$$\sup_{j=1,\ldots,N} |\langle x_0, \Phi_j(\omega)\rangle| < 2\alpha. \tag{26}$$

This shows that

$$\mathbb{P}(\Omega_0 \cap \Omega_1) \leq \#\mathcal{N}\left[1 - \left(1 - \frac{1}{4\alpha^2 d}\right)\frac{1}{2\alpha\sqrt{d}}\sqrt{\frac{2}{\pi}}\exp -2d\alpha^2\right]^N. \tag{27}$$

Since $(1 - \frac{1}{4\alpha^2 d})\sqrt{\frac{2}{\pi}} \geq \frac{1}{2}$ if $N \geq C_\mu d$ where C_μ is a constant dependent on μ we can replace this part of (27) by $\frac{1}{2}$. Using this and elementary inequality $1 - \xi \leq e^{-\xi}$ we get

$$\mathbb{P}(\Omega_0 \cap \Omega_1) \leq \#\mathcal{N}\exp -N\left(\frac{1}{4\alpha\sqrt{d}}\exp -2d\alpha^2\right) \tag{28}$$

Substituting the value for α and $\#\mathcal{N}$ we see that we must estimate from above

$$\left(\frac{6}{\mu}\sqrt{\frac{d}{\ln(N/d)}}\right)^d \exp -\frac{N^{1-2\mu^2}d^{2\mu^2}}{4\mu\sqrt{\ln(N/d)}} \tag{29}$$

Converting to exponentials we see that we must show that

$$d\left(\ln\left(\frac{6}{\mu}\sqrt{\frac{d}{\ln(N/d)}}\right) - \left(\frac{N}{d}\right)^{1-2\mu^2}\frac{1}{4\mu\sqrt{\ln(N/d)}}\right) \leq -cd \tag{30}$$

so we need

$$\ln\left(\frac{6}{\mu}\sqrt{\frac{d}{\ln(N/d)}}\right) + c \leq \left(\frac{N}{d}\right)^{1-2\mu^2}\frac{1}{4\mu\sqrt{\ln(N/d)}} \tag{31}$$

Now we see that when d and N grow the left hand side grows no faster that $a + \frac{1}{2}\ln d$ while the right hand side grows no slower than $(N/d)^{1-2\mu^2-\epsilon} \geq (\ln d)^{\xi(1-2\mu^2-\epsilon)}$ so we infer that there exists $c > 0$ such that

$$\mathbb{P}(\Omega_0 \cap \Omega_1) \leq e^{-cd}. \tag{32}$$

This clearly finishes the Gaussian case. To see the uniform case it suffices to make the trivial observation that for $\omega \in \Omega_1$ we have $\Phi_\omega(B_1^N) \subset 2\tilde{\Phi}_\omega(B_1^N)$. $\square$

Before we proceed let us comment how this Proposition relates to our problem. We know that to get RIP(k, δ) with big probability we must have $d \geq ck\log(N/k)$. Since N and d are fixed by our measurement we choose $k \sim d/\log(N/k)$. In this case we have

$$\frac{\mu}{\sqrt{d}}\sqrt{\ln\frac{N}{d}} \sim \frac{\mu'}{\sqrt{k}}.$$

Also we assume (otherwise the results a non-interesting) that k is a natural number. This implies that restrictions on dimensions in Proposition 4.2 are satisfied and using Lemma 4.1 we get

Corollary 4.3. *Let us consider either Gaussian or uniform random $N \times d$ measurement matrix Φ_ω. There exists positive constants δ, μ, α, c such that for $k = \alpha d/\ln N$ with probability greater then $1 - \exp -cd$ matrices Φ_ω satisfy both RIP(k, δ) and LQ$(\mu/\sqrt{k})$.*

This means that we are in the situation covered by Theorem 3.2.

Other widely used random sets of measurement matrices Φ_ω are Bernoulli or Fourier (see e.g. [2, 7, 8]). All they have the property that each entry is $\leq C/\sqrt{d}$. This implies that each vector in $\Phi_\omega(B_1^N)$ will have each coordinate bounded by $c/\sqrt{d}$, so the best one can hope for is that such a matrix satisfies LQ$(c/\sqrt{d})$.

References

[1] R.Baraniuk, M.Davenport, R.DeVore and M.Wakin, *A Simple proof of the Restricted Isometry Property for Random Matrices*, http://www.dsp.ece.rice.edu/cs/

[2] E.Candes, *Compressive sampling*, Proc. Intl. Congress Math. Madrid 2006

[3] E.Candes, *"People Hearing Without Listening':' An introduction to compressive sampling*, IEEE Signal Processing Magazine (to appear)

[4] E.Candes, J.Romberg and T.Tao, *Stable signal recovery from incomplete and inacurate measurements*, Comm. Pure Appl. Math., 59 1207-1223.

[5] E. Candes, T.Tao, *Near optimal recovery from random projections: universal encoding strategies* IEEE Trans Inform. Theory 52 (2004) 5406-5425.

[6] E.Candes, T.Tao, *Decoding by linear programing*, IEEE Trans. Inform. Theory 51 (2005) 4203-4215

[7] A.Cohen, W.Dahmen and R.DeVore, *Compressed sensing and best* $k-$ *term approximation*, http://www.dsp.ece.rice.edu/cs/

[8] D. Donoho, *Compressed sensing* IEEE Trans. Inform. Theory 52 (2006) 1289–1306

[9] A.Litvak, P.Mankiewicz and N. Tomczak–Jaegermann, *Randomized Isomorphic Dvoretzky Theorem*, C. R. Math. Acad. Sci. Paris 335 (2002), no. 4, 345–350.

[10] G.G.Lorentz, M. v.Golitschek, Y. Makovoz, **Constructive Approximation; Advanced Problems** Springer Verlag 1996

[11] P.Mankiewicz, *Compact Groups of operators on Subproportional Quotiens of* ℓ_1^m, Canad. J. Math. 52 (2000), no. 5, 999–1017.

[12] S.Mendelson, A.Pajor, N. Tomczak-Jaegermann, *Reconstruction and subgaussian processes*, C.R. Acad. Sci. Paris, Ser. I Math. 340 (2005), 885-888

[13] S.Mendelson, A.Pajor, N. Tomczak-Jaegermann, *Uniform uncertainity principle for Bernoulli and subgaussian ensembles*, Constructive Approximation (to appear)

[14] D.Needell, R.Vershynin, *Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit*, http://www.dsp.ece.rice.edu/cs/

[15] D.Needell, R.Vershynin, *Signal recovery from incomplete and inacurate measurements via regularised orthogonal matching pursuit*, http://www.dsp.ece.rice.edu/cs/