Supplementary materials for the paper: Sparse Recovery with Brownian Sensing

Proof of Proposition 1

First, we prove a very short Lemma describing some properties of the matrix A.

Lemma 1 *Let us consider M independent Brownian motions $(B^1, ..., B^M)$ on $\mathcal{X}$, and define the $M \times K$ matrix A with elements*

$$A_{m,k} = \frac{1}{\sqrt{M}} \Big(\int_{\mathcal{C}} \varphi_k(x) dB^m(x) \Big).$$

Then A is a centered Gaussian matrix where each row $A_{m,\cdot}$ is i.i.d. from $\mathcal{N}(0, \frac{1}{M} V_{\mathcal{C}})$, where $V_{\mathcal{C}}$ is the $K \times K$ covariance matrix of the basis, defined by its elements $V_{k,k'} = \int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx$.

Proof: Indeed, from the definition of stochastic integrals, each $A_{m,k} \sim \mathcal{N}(0, \frac{1}{M}\int_{\mathcal{C}} \varphi_k^2(x)dx)$, and $\text{Cov}(A_{m,k}, A_{m,k'}) = \frac{1}{M}\int_{\mathcal{C}} \varphi_k(x)\varphi_{k'}(x)dx$. Thus each row $A_{m,\cdot} \sim \mathcal{N}(0, \frac{1}{M} V_{\mathcal{C}})$ and are independent by independence of the Brownian motions. Additionally, we have

$$\mathbb{E}[(A^T A)_{k,k'}] = \mathbb{E}\Big[\frac{1}{M}\sum_{m=1}^{M} A_{m,k}A_{m,k'}\Big] = V_{k,k',\mathcal{C}}.$$

$\square$

Now let us define $B = AV_{\mathcal{C}}^{-1/2}$. Since each row of A is an independent draw of $\mathcal{N}(0, V_{\mathcal{C}})$, then each row of B is an independent draw of $\mathcal{N}(0, I)$. Thus B is a matrix with elements i.i.d. from $\mathcal{N}(0, 1)$. We thus can use the following result (as stated in [9], see also [14, 1]):

Theorem 5 *For $p' > 0$ and any integer $t > 0$, when $M > C'\delta^{-2}(t\log(K/t) + \log 1/p'))$, with C' being a universal constant, see [14, 1], then with probability at least $1 - p'$, there exists $\delta_t \le \delta$ (δ_t is the RIP constant of B for t-sparse vectors) such that for all $t-$sparse vectors $x \in \mathbb{R}^K$,*

$$(1 - \delta_t)\|x\|_2 \le \|Bx\|_2 \le (1 + \delta_t)\|x\|_2.$$

Since $V_{\mathcal{C}}$ is symmetric, it is possible to write $V_{\mathcal{C}} = UDU^T$ with U an orthogonal matrix and D a diagonal matrix with the eigenvalues of V as diagonal elements (SVD decomposition). Thus, $V^{1/2} = UD^{1/2}U^T$ where $D^{1/2}$ is the diagonal matrix with the square roots of the diagonal elements of D (i.e., the eigenvalues of $V_{\mathcal{C}}^{1/2}$).

Note that if U is an orthogonal matrix, BU is also RIP with the same constant as B (see [7] for the preservation of the RIP property to a change of orthonormal basis). Applying this and Theorem 5 with $\delta = 1/2$ for $2t$-sparse vectors, we have that whenever $M > 4C'(2t\log(K/2t) + \log 1/p')$, the RIP constant $\delta_{2t} \le 1/2$, i.e. for all $2t-$sparse vectors x,

$$\frac{1}{2}\|x\|_2 \le \|BUx\|_2 \le \frac{3}{2}\|x\|_2.$$

Now if we consider a $2t-$sparse vector x, then $D^{1/2}x$ is also $2t-$sparse with same support as x, and we also have that $\nu_{min,\mathcal{C}}\|x\|_2 \le \|D^{1/2}x\|_2 \le \nu_{max,\mathcal{C}}\|x\|_2$. Thus the matrix $BUD^{1/2}$ satisfies

$$\frac{\nu_{min,\mathcal{C}}}{2}\|x\|_2 \le \|BUD^{1/2}x\|_2 \le \frac{3\nu_{max,\mathcal{C}}}{2}\|x\|_2.$$

As mentioned before, the preservation of the RIP property to a change of orthonormal base (see [7]) can be applied with U and thus as $A = BV^{1/2} = BUD^{1/2}U^T$ to obtain:

$$\frac{1}{2}\nu_{\min,\mathcal{C}}\|x\|_2 \le \|Ax\|_2 \le \frac{3}{2}\nu_{\max,\mathcal{C}}\|x\|_2.$$

Proof of Proposition 2

We prove here without loss of generality (because of we can always parametrize the curve) the result for $\mathcal{X} = [0, l]$. Let us recall that f is (L, β)-Hölder and that we write $\sigma = \|\eta\|_2$. The estimation error $\varepsilon_m = b_m - \widehat{b_m}$, given the samples $(x_n, y_n)_n$, follows a centered Gaussian distribution (w.r.t. the choice of the Brownian B^m) with variance

$$
\begin{aligned}
\mathbb{V}(\varepsilon_m) &= \mathbb{V}\left(\frac{1}{\sqrt{M}}\Big(\int_0^l f(x)dB^m(x) - \sum_{n=0}^{N-1} y_n(B^m_{x_{n+1}} - B^m_{x_n})\Big)\right) \\
&= \frac{1}{M}\mathbb{V}\left(\int_0^l \big(f(x) - \sum_n (f(l\frac{(n+1)}{N}) + \eta_n)\mathbb{I}_{x\in[l\frac{n}{N};l\frac{(n+1)}{N}]}\big)dB^m(x)\right) \\
&= \frac{1}{M}\int_0^l \big(f(x) - \sum_n (f(l\frac{n}{N}) + \eta_n)\mathbb{I}_{x\in[l\frac{n}{N};l\frac{(n+1)}{N}]}\big)^2 dx \\
&= \frac{1}{M}\sum_n \int_{l\frac{n}{N}}^{l\frac{(n+1)}{N}} (f(x) - f(l\frac{n}{N}) - \eta_n)^2 dx \\
&\leq \frac{1}{MN}\sum_n (\frac{Ll^\beta}{N^\beta} + |\eta_n|)^2 dx \\
&= \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sum_n |\eta_n|^2\Big) \\
&\leq \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sigma^2\Big).
\end{aligned}
$$

We now wish to apply Bernstein's inequality in order to bound $\|\varepsilon\|_2$ in high probability. We recall the following result (see e.g. [2]):

Theorem 6 (Bernstein's inequality) *Let $(X_1, ...X_M)$ be independent real valued random variables and assume that there exist two positive numbers v and d such that: $\sum_{m=1}^M \mathbb{E}(X_m^2) \leq v$ and for all integers $r \geq 3$,*

$$\sum_{m=1}^M \mathbb{E}[(X_m)_+^r] \leq \frac{r!}{2} v d^{r-2}.$$

Let $S = \sum_{m=1}^M (X_m - \mathbb{E}(X_m))$, then for any $x \geq 0$, we have $\mathbb{P}(S \geq \sqrt{2vx} + dx) \leq \exp(-x)$.

Let us check that the assumptions for applying Bernstein's inequality hold with the choice $v = 8M(\mathbb{V}(\varepsilon_m))^2$ and $d = 2\mathbb{V}(\varepsilon_m)$. Indeed, since the ε_m are i.i.d. centered Gaussian, by writing $X_m = \varepsilon_m^2$, we have $X_m \geq 0$ and for any integer $r \geq 2$, $\mathbb{E}(X_m^r) = (\mathbb{V}(\varepsilon_m))^r \frac{(2r)!}{2^r r!}$. This gives $\sum_{m=1}^M \mathbb{E}[X_m^2] = 3M(\mathbb{V}(\varepsilon_m))^2 \leq v$, and for $r \geq 3$,

$$\sum_{m=1}^M \mathbb{E}[X_m^r] = M(\mathbb{V}(\varepsilon_m))^r \frac{(2r)!}{2^r r!} \leq M(\mathbb{V}(\varepsilon_m))^r \times 2^r r! \leq \frac{r!}{2} v d^{r-2}.$$

We thus apply Bernstein's inequality (and recall that $\mathbb{V}(\varepsilon_m) \leq \frac{2}{MN}\Big(\frac{L^2 l^{2\beta}}{N^{2\beta-1}} + \sigma^2\Big)$) to obtain that with probability at least $1 - p$,

$$\|\varepsilon\|_2^2 \leq 2\Big(\frac{L^2 l^{2\beta}}{N^{2\beta}} + \frac{\sigma^2}{N}\Big)\Big(1 + 4\sqrt{\frac{\log(1/p)}{M}} + 2\frac{\log(1/p)}{M}\Big).$$

Proof of Theorem 4

Following [10], we define $\alpha_t > 0$ (respectively $\beta_t > 0$) as the maximal (resp. minimal) values such that for all $x \in \mathbb{R}^K$ which are $t-$sparse,

$$\alpha_t \|x\|_2 \leq \|Ax\|_2 \leq \beta_t \|x\|_2. \tag{4}$$

We now define $\gamma_t = \frac{\beta_t}{\alpha_t}$ and use Theorem 3.1 of [10] applied to sparse vectors, in the case of ℓ_1 minimization, reminded below:

Theorem 7 (Foucart, Lai) *For any integer $S > 0$, for $t \geq S$, whenever $\gamma_{2t} - 1 \leq 4(\sqrt{2}-1)\sqrt{\frac{t}{S}}$, the solution $\widehat{\alpha}$ to the ℓ_1-minimization problem*

$$\min \|a\|_1, \text{ under the constraint } \|Aa - b\|_2^2 \leq \|\varepsilon\|_2^2,$$

satisfies $\|\alpha - \widehat{\alpha}\|_2 \leq \frac{D_2 \|\varepsilon\|_2}{\beta_{2S}}$, where D_2 is a constant which depends on γ_{2t}, S and t defined in [10].

In order to apply this results, we now provide conditions such that (4) holds, as well as an upper bound on the noise $\|\varepsilon^2\|$, and a lower bound on β_{2S}.

Step 1. Recovery Condition: We recall the results of Proposition 1 and have that (4) holds with $\alpha_{2t} \geq \frac{1}{2}\nu_{\min,\mathcal{C}}$ and $\beta_{2t} \leq \frac{3}{2}\nu_{\max,\mathcal{C}}$ with probability $1 - p'$ as long as $M > \frac{C'}{4}(t \log(K/t) + \log 1/p'))$. Thus $\gamma_{2t} \leq 3\frac{\nu_{\max,\mathcal{C}}}{\nu_{\min,\mathcal{C}}} = 3\kappa_{\mathcal{C}}$.

A sufficient condition for (7) is that $3\kappa_{\mathcal{C}} - 1 \leq 4(\sqrt{2}-1)\sqrt{\frac{t}{S}}$.

By defining $r = \left[(3\kappa_{\mathcal{C}} - 1)(\frac{1}{4\sqrt{2}-1})\right]^2$ (note that r only depends on $V_{\mathcal{C}}$), condition (7) holds whenever $t > Sr$, thus with probability $1 - p'$, whenever

$$M > 4C'\big(2\lceil Sr \rceil \log \frac{K}{2Sr} + \log 1/p'\big). \tag{5}$$

Note that this condition holds when the number of Brownian motions M (which can be chosen arbitrarily) is large enough (and does not depend on the number of observations N).

Step 2. Upper bound on $\|\varepsilon^2\|$: This is the result of Proposition 2.

Step 3. Lower bound on β_{2S} In order to apply Theorem 7, we now provide a lower bound on β_{2S}.

Lemma 2 *If*

$$M > C' \log 1/u, \tag{6}$$

then with probability $1 - u$ we have: $\beta_{2S} \geq \frac{1}{2}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2}$.

Proof: Let us define $i = \operatorname{argmax}_k \int_{\mathcal{C}} \varphi_k^2(x)dx$. Let us now consider the $1-$sparse vector a such that $a_i = 1$ and $a_k = 0$ otherwise. We have: $(Aa)_m = \int_{\mathcal{C}} \varphi_i(x) dB^m(x)$. So each $(Aa)_m$ is a sample drawn independently from $\mathcal{N}(0, \int_{\mathcal{C}} \varphi_i^2(x)dx)$.

By applying Theorem 5, with $S = K = 1$ and $\delta = 1/2$, when $M > C' \log 1/u$, then with probability $1 - u$,

$$\frac{1}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx}\|a\|_2 \leq \|Aa\|_2 \leq \frac{3}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx}\|a\|_2.$$

And since β_{2S} is the minimal constant such that for every $2S-$sparse vector x (in particular for a) we have $\|Ax\|_2 \leq \beta_{2S}\|x\|_2$, we deduce that

$$\beta_{2S} \geq \frac{1}{2}\sqrt{\int_{\mathcal{C}} \varphi_i^2(x)dx} = \frac{1}{2}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2(x)dx}.$$

$\square$

We now apply Theorem 7 and deduce that when M satisfies (5) (which implies that (6) also holds) using Lemma 2, with probability $1 - p' - u$,

$$\|\widehat{\alpha} - \alpha\|_2 \leq \frac{2D_2\tilde{\sigma}(N, M, p)}{\sqrt{N}\sqrt{\max_k \int_{\mathcal{C}} \varphi_k^2}} \tag{7}$$

Thus from Proposition 2, with probability $1 - p - p' - u$,

$$\|\widehat{\alpha} - \alpha\|_2^2 \leq \frac{8D_2^2\left(\frac{L^2}{N^{2\beta-1}}l^{2\beta} + \sigma^2\right)(1 + c(p, M))}{N(\max_k \int_{\mathcal{C}} \varphi_k^2)},$$

and from [10], we deduce that if we are able to recover $4S-$sparse vectors, i.e., if $M > 4C'\left(4Sr\log\frac{K}{4Sr} + \log 1/p'\right)$ then $D_2 \leq C\kappa_{\mathcal{C}}^2$ where C can be loosely bounded by 90, see [10] (note that this numerical constant can be greatly improved). The result follows with the choice $p = p' = u$.