Abstract—We study the problem of compressed sensing off-the-grid in 
the multivariate and general operator setting. Under the Fisher metric, 
we establish stable recovery of a sparse measure when the number of 
random measurement scales linearly with sparsity up to log factors.

I. INTRODUCTION

Let \( \mathcal{X} \subseteq \mathbb{R}^d \), \( \mathcal{M}(\mathcal{X}) \) denote the space of complex-valued Radon 
measures on \( \mathcal{X} \), \( |\mu|(\mathcal{X}) \) denote the total variation norm for \( \mathcal{M}(\mathcal{X}) \). 

We consider the problem of compressed sensing off-the-grid, where 
we aim to recover a sparse measure \( \mu_0 = \sum_{j=1}^{K} a_j \delta_{x_j} \) given 
measurements \( y = \Phi \mu_0 + w \), where \( w \in \mathbb{C}^m \) is the additive 
noise with \( \|w\| \leq \delta \) and \( \Phi : \mathcal{M}(\mathcal{X}) \to \mathbb{C}^m \) is defined by 
\( \Phi_{ij}(\mu) = \int \varphi_{\omega}(x) d\mu(x) \), where \( \varphi_{\omega} \in \mathcal{C}(\mathcal{X}) \) are smooth functions 
associated to a random variable \( \omega \) in probability space \( \Omega \) with density 
\( \Lambda \) and \( \omega_1, \ldots, \omega_m \overset{iid}{\sim} \Lambda \). We assume that \( \mathbb{E}_{\omega}[|\varphi_{\omega}(x)|^2] = 1 \) for all 
\( x \). We consider the solutions to the following minimisation problem:

\[
\min_{\mu \in \mathcal{M}(\mathcal{X})} |\mu|(\mathcal{X}) + \frac{1}{2\lambda} \|\Phi \mu - y\|^2_2. 
\] (1)

A. Previous works and contributions

Previous studies [2, 7], which focus on specific sampling operators 
have established that a minimum separation condition (often imposed 
via some ad-hoc metric) between the spike positions \( \{x_j\}_j \) is 
necessary for recovery guarantees of (1). As a first contribution, 
we show that this separation distance can be imposed naturally using 
the Fisher metric associated with \( \Phi \). This allows for the first time, 
quantitative results in the study of non-translation invariant operators 
(e.g. sampling the Laplace transform). For the specific problem of 
random subsampling, to our knowledge, the only previous work is 
for the case of sampling Fourier coefficients [6] in the noisless setting, 
where it is shown that a sparse measure can be recovered using 
\( \mathcal{O}(s \log(s) \log(j)) \) provided that the spike positions are sufficiently 
separated, and \( \text{sign}(a) \) is drawn uniformly randomly iid from the 
complex unit circle.

Under a sufficient separation using the Fisher metric, we extend this 
result to the general setting and establish support stability under small noise. 
Our main result then shows that the random sign condition can be removed while still retaining the 
optional \( s \) scaling, albeit under optimal partial transport stability.

II. THE FISHER METRIC

The kernel associated to \( \Phi \) is the bivariate function 
\( K(x, x') = \frac{1}{m} \sum_{k=1}^{m} \varphi_{\omega_k}(x) \varphi_{\omega_k}(x') \) and its limit is defined as 
\( K(x, x') = \mathbb{E}[K(x, x')] \). In previous studies [2, 7], the kernel is translation 
invariant so \( K(x, x') = k(x - x') \) for some \( k : \mathcal{X} \to \mathbb{R} \), 
and the minimum separation distance is naturally measured in the Euclidean metric 
and is related to the “width” of the kernel. In the case where \( K \) is no longer translation invariant, 
a natural separation distance can be derived by considering the metric 
\( H_x = \nabla_1 \nabla_2 K(x, x) \). Assuming that this is positive definite, it induces a geodesic distance 
\( d_H \) between points on \( \mathcal{X} \). Note that \( H_x \) can be shown to be 4 times 
the Fisher information metric of \( f(\omega, x) = |\varphi_{\omega}(x)|^2 \) (interpreted as 
a probability density function for \( \omega \) conditional on parameter \( x \)) in 
the case where \( \varphi_{\omega} \) is real-valued.

In the following, \( D_r [f] \) denotes the \( r^{th} \) derivative of \( f \) normalised 
by the metric \( H_x \), so that \( D_1 [f](x) = H_x^{-\frac{1}{2}} \nabla f(x) \), \( D_2 [f](x) = H_x^{-\frac{1}{2}} \nabla^2 f(x) H_x^{-\frac{1}{2}} \) and so on, and \( K^{(s)} \) denotes \( K \) after taking the 
\( i^{th} \) and \( j^{th} \) metric normalised derivatives in the first and second variables respectively.

We say that \( K \) is admissible with respect to \( s_{\text{max}} \in \mathbb{N} \) (for 
maximum number of spikes), \( \Delta > 0 \) (for width of the kernel), 
\( \varepsilon_0, \varepsilon_2, r > 0 \) (for local curvature), \( B_{ij} > 0 \) and \( C_{ij} \geq 1 \) if

1) Uniform bounds: For \( i + j \leq 3 \), \( \sup_{x, x' \in \mathcal{X}} \|K^{(i,j)}(x, x')\| \leq B_{ij} \); 
2) Local curvature: For all \( x, x' \in \mathcal{X} \) with \( d_H(x, x') \leq r \), 
\( K^{(2)}(x, x') \leq \varepsilon_2 I \) and \( \|H_x^{-\frac{1}{2}} \nabla^2 f(x) H_x^{-\frac{1}{2}} - I\| \leq C_{ij} d_H(x, x') \);
and for \( d_H(x, x') \geq r \), \( |K(x, x')| \leq 1 - \varepsilon_0 \).

Our main result gives recovery guarantees for \( K \) and \( s_{\text{max}} \leq s \).

We let \( L \overset{\text{def.}}{=} \max_{\omega \in \Omega, x, x' \in \mathcal{X}} \max_{j=1}^{3} \|D_j [\varphi_{\omega}](x)\| \).

III. RESULTS

A first result, previously published in [5], formulates support stability at small noise.

**Theorem 1.** Assume that \( \text{sign}(a) \) is distributed uniformly iid on 
the complex unit circle, and the number of measurements \( m \) satisfies:

\[ m \gtrsim s \cdot C \cdot \log(s) \log \left( \frac{N^d}{\rho} \right), \]

where \( C \overset{\text{def.}}{=} \mathcal{E}^2 \max \left( \frac{\rho^2}{\rho \log(\rho)}, 1 \right) \) and \( N \overset{\text{def.}}{=} \frac{dN^2 \log(\rho)}{\rho} \), and that 
\( w \sim \mathcal{N}(0, \Delta) \), \( \lambda \sim \varepsilon \min_{i} |a_i|^2 \) (up to log factors). Then with probability

\[ \left( \sum_{i=1}^{a} |a_i - \hat{a}_i|^2 + d_H(x_i, x_i') \right)^{1/2} \lesssim \sqrt{\lambda + \|w\|_2} \min_{i} |a_i| \]

Our main result gives recovery guarantees for arbitrary \( \text{sign}(a) \).

**Theorem 2.** Assume the number of measurements \( m \) satisfies:

\[ m \gtrsim \sigma \cdot \left( \log(s) \log \left( \frac{N^d}{\rho} \right) \right), \]

then with probability at least \( 1 - \rho \), any solution to \( \mu \) to (1) with 
\( \|w\| \sim \delta \), \( \lambda \sim \delta / \sqrt{\delta} \) (up to log factors) is approximately sparse in 
the sense

\[ \mathcal{T}_{d_H}^2 \left( \|\hat{\mu} - \mu\|_1 \sum_{j=1}^{\rho} \|\mu| (B_0(x_i)) \|_1 d_{x_j} \right) \lesssim \delta, \]

where \( \mathcal{T}_{d_H}^2 \) is the \( W_{d_H} \) optimal partial transport distance [4].
<table>
<thead>
<tr>
<th>$\mathcal{X}$</th>
<th>Torus $\mathbb{T}^d$</th>
<th>$\mathcal{X} \subset \mathbb{R}^d$</th>
<th>$\mathcal{X} \subset \mathbb{R}^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi_\omega$</td>
<td>$e^{i2\pi(\omega, x)}$</td>
<td>$e^{i(\omega, x)}$</td>
<td>$\prod_{i=1}^d \sqrt{2(\alpha_i + \alpha_i^*)} e^{i(\omega, x)}$</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>weighted sampling of $\mathbb{Z}^d \cap [-f_c, f_c]^d$</td>
<td>$\mathcal{N}(0, \Sigma^{-1})$</td>
<td>$\prod_{i=1}^d (2\alpha_i) e^{-(\alpha, x)}$</td>
</tr>
<tr>
<td>$d_H(x, x')$</td>
<td>$|x - x'|_2$, $k \sim f_c$</td>
<td>$\sqrt{\Sigma^{-1}(x - x'), x - x'}$</td>
<td>$\sum_i \log \left( \frac{\alpha_i + \alpha_i^*}{\pi^2} \right)$</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>$\sqrt{\tau_{\max}} \sqrt{d}$</td>
<td>$\sqrt{\log(\tau_{\max})}$</td>
<td>$d + \log(d\tau_{\max})$</td>
</tr>
<tr>
<td>$C$ (Thm. 1, 2)</td>
<td>$\mathcal{O}(d^2)$</td>
<td>$\mathcal{O}(d^2)$</td>
<td>$\mathcal{O}(d^6)$</td>
</tr>
</tbody>
</table>

**TABLE I**

Three examples of admissible kernels: a multivariate extension of the Discrete Fourier model on the torus $\mathbb{T}$ (where the weighting in $\Lambda$ is chosen as in the paper [2]), a Fourier model with Gaussian frequencies on a compact $\mathcal{X} \subset \mathbb{R}^d$ (which bears connection with mixture model learning), and a Laplace transform model used in fluorescence microscopy [3]. Note that, strictly speaking, the Gaussian and Laplace model do not have bounded features $L < \infty$. However it is possible to refine the analysis to replace $L$ by a stochastic quantity bounded with high probability, such that all the theory stays valid with the quantities indicated above (which we do not describe here for simplicity). See [5] for more details.

**ACKNOWLEDGMENT**

This work was partly funded by the CFM-ENS chair “Modèles et Sciences des données” and the European Research Council, NORIA project.

**REFERENCES**


