Compressed Sensing

1 Signal compressibility

Let $\mathcal{L}^* = A^* \mathbb{Z}^d$ be a lattice with the reciprocal $\mathcal{L} = A \mathbb{Z}^d$. Given a signal $f : \mathbb{R}^d \rightarrow \mathbb{R}$, its sampling on $\mathcal{L}^*$ can be viewed as projections onto shifted Diracs,

$$f[n] = f(A^* n) = (\tau_{A^* n}\delta)(f) = \langle \tau_{A^* n}\delta, f \rangle$$

for every $n \in \mathbb{Z}^d$. This set of measurements can be viewed as the orthogonal projection of $f$ onto the space of $\mathcal{L}$-band limited functions spanned by the shifts of sinc,

$$\mathbb{B}_\mathcal{L} = \text{sp}\{\text{sinc}(A^T x - n)\}_{n \in \mathbb{Z}^d}.$$

In our notations of dictionaries, the synthesis operator $\Phi : \ell^2(\mathbb{Z}^d) \rightarrow L^2(\mathbb{R}^d)$ is given by

$$\Phi c = \sum_{n \in \mathbb{Z}^d} c_n \phi_n$$

with $\phi_n(x) = \text{sinc}(A^T x - n)$. The analysis operator $\Psi : L^2(\mathbb{R}^d) \rightarrow \ell^2(\mathbb{Z}^d)$ is given by

$$\Psi f = \langle \psi_n, f \rangle_{L^2(\mathbb{R}^d)},$$

where $\psi_n = \tau_{A^* n}\delta$. From this perspective, classical sampling is nothing but representation of the signal in a shift-invariant dictionary of shifted sincs. For a general signal not belonging to $\mathbb{B}_\mathcal{L}$, the projection onto the dictionary introduces error (high frequency loss and aliasing); however, for a function belonging to the subspace $\mathbb{B}_\mathcal{L}$ of $L^2(\mathbb{R}^d)$, the projection is loss-less (and, hence, the reconstruction is exact). In our terms, classical sampling makes a prior assumption: band limitedness.

However, band limitedness is problematic from two perspectives. Firstly, many naturally occurring signals (images in primis) are far from being band limited. Secondly, the sampled signal still contains a significant amount of structure that can be exploited to further compress it. For example, speech and music signals exhibit strong harmonic structure, which is manifested by energy concentration at a sparse set of coefficients in the windowed Fourier transform domain. In other words, given a sampled signal $f[n]$, its inner products

$$c_{k,m} = \langle f, \phi_{k,m} \rangle_{\ell^2(\mathbb{Z}^d)}$$

with the windowed shifted harmonics

$$\phi_{k,m}[n] = e^{2\pi i m^T (n-k)} w[n - k]$$
is a sparse sequence. This means that the harmonic signal is compressible in the windowed Fourier domain – we can through a certain amount of the coefficients (those with the smallest energy) and still reconstruct it accurately. This in turn implies that our sampling was wasteful in first place – if we could incorporate the stronger prior of harmonic structure directly into the sampling scheme, we could represent the signal more efficiently (with less measurements).

2 Compressed sensing

The idea of incorporating the prior leading to signal compressibility directly into the sampling scheme is usually referred to as compressed sensing or sampling (CS). In its basic setting, CS can be viewed as replacing the analysis (a.k.a. sensing) operator

$$\Psi f = \langle \psi_n, f \rangle_{L^2(\mathbb{R}^d)}$$

with some other operator containing different sensing functions \(\{\psi_n\}\). The sensing functions need not be localized anymore. Essentially, \(\Psi\) can realize any linear measurements (i.e., such measurements that can be expressed as linear functionals of the data). We can think of the sensing operator as of a linear encoder whose task is to capture maximum information about the signal with minimum measurements.

Let us now introduce a prior on the class of measured signal: as before, we will assume that \(f\) admits a sparse approximation in a dictionary \(\Phi\) (it can be a local shift-invariant dictionary or some global dictionary such as the wavelet transform). We can therefore substitute \(f = \Phi c\), where \(c\) is the sequence of representation coefficients. Recovering the signal \(f\) from its measurements \(y = \Psi f\) can therefore be written as the following inverse problem

$$\hat{f} = \Phi \hat{c},$$

where

$$\hat{c} = \arg \min_c \|\Psi \Phi c - y\|_{L^2(\mathbb{R}^d)}^2 + \lambda \|c\|_1.$$  

Note that the data term is now expressed in the measurement space. The mapping from \(y\) to \(\hat{f}\) can be thought of as a decoder; unlike the encoder, the decoder is highly non-linear. A surprising fact is that a random projection acts as a universally good sensing operator. Compressed sensing theory claims that an \(n\)-dimensional \(k\)-sparse signal can be recovered from \(\mathcal{O}(k \log n)\) measurements.