The Compressed Sensing Paradigm

MA4291 Presentation Report 1

Ang Yan Sheng
A0144836Y

Motivation The digital camera is the product of an arms race: between increasing the number of megapixels during image capture, and throwing out most of the raw data during compression.

Since camera sensors are made of silicon, camera manufacturers can piggyback onto Moore's Law to cram exponentially many sensors onto a single chip. However, there are many other domains where measurement is expensive—such as infrared cameras, astronomy, and medical imaging—where this wasteful model is not viable. For these applications, we would like to minimise the amount of redundant information captured at the sensor end.

In general, given a data vector $x \in \mathbb{R}^n$, we can model a set of m measurements by a measurement matrix $A \in \mathbb{R}^{m\times n}$, and we want to recover x from Ax. Of course, this is impossible in general if $m < n$, since A cannot be injective. However, for most domains of interest, there is an appropriate basis in which the data is (nearly) k-*sparse* for some $k \ll n$, meaning that at most k components are nonzero.

Write $\Sigma_k \coloneqq \{z \in \mathbb{R}^n : \|z\|_0 \leqslant k\}$, the set of k-sparse vectors in $\mathbb{R}^n$. We can now formulate the central problem of compressed sensing:

Problem. *Given $k \ll n$, find a measurement matrix $A \in \mathbb{R}^{m\times n}$ with $m \ll n$, such that any $x \in \Sigma_k$ can be efficiently recovered from Ax.*

For any hope of a solution, we clearly need A to be injective on Σ_k, in which case x is the unique solution of the optimisation problem

$$\min_z \|z\|_0 \qquad \text{with} \quad Az = Ax. \tag{P_0}$$

The issue is that the above ℓ_0 minimisation problem is NP-hard. Conventionally, (P_0) is relaxed to the least-squares problem

$$\min_z \|z\|_2 \qquad \text{with} \quad Az = Ax, \tag{P_2}$$

which can be solved very quickly. However, in general the solution is not sparse, and the reconstructions are rather poor in practice.

The ℓ_1 variant of the above problems,

$$\min_z \|z\|_1 \qquad \text{with} \quad Az = Ax, \tag{P_1}$$

is a convex optimisation problem, and hence can also be solved efficiently. In 2004, it was noticed that (P_1) gave *exact* reconstructions of certain MRI test images with only 2% of the Fourier coefficients needed conventionally (Candès-Romberg-Tao 2006). The theoretical analysis of this phenomenon gave rise to the field of compressed sensing. We shall present some of its key concepts below.

The restricted isometry property Motivated by the above discussion, we want to find A such that (P_1) has unique solution $z = x$ for all $x \in \Sigma_k$. We shall do this by reducing to a condition which holds for random Gaussian matrices A with high probability.

Definition. A matrix $A \in \mathbb{R}^{m\times n}$ is said to have the *null space property of order* k (NSP_k) if for all $v \in \ker A\backslash\{0\}$, and $T \subseteq [n]$ of size k, we have $\|v_T\|_1 < \|v_{T^c}\|_1$.

In other words, the largest k coordinates of v (by absolute value) contains less than half the mass of v.

An easy consequence of the definition is the following:

Proposition. *(P_1) has unique solution $z = x$ for all $x \in \Sigma_k \iff A$ has NSP_k.*

However, verifying NSP_k is hard, so we introduce the following condition:

Definition. Given $A \in \mathbb{R}^{m\times n}$, the *restricted isometry constant of order* k is the minimal constant $\delta_k(A) \geqslant 0$ such that

$$\left|\|Ax\|_2^2 - \|x\|_2^2\right| \leqslant \delta_k\|x\|_2^2 \qquad \text{for all } x \in \Sigma_k.$$

If the restricted isometry constant is small, then A almost preserves the inner product. Combining this fact with some elementary estimates, we obtain:

Theorem. *If $\delta_{2k}(A) < 1/3$, then A has NSP_k.*

The main result of the presentation is the following:

Theorem. *Let $n \geqslant m \geqslant k \geqslant 1$, $0 < \varepsilon < 1$, $0 < \delta < 1$. Then there exists an absolute $C > 0$ such that if*

$$m \geqslant C\delta^{-2}(k\ln(en/k) + \ln(2/\varepsilon)),$$

then with $A = \frac{1}{\sqrt{m}}(\omega_{ij})_{i,j=1}^{m,n}$, where $\omega_{ij} \sim N(0,1)$,

$$\mathbb{P}(\delta_k(A) \leqslant \delta) \geqslant 1 - \varepsilon.$$

Proof (sketch). By usual concentration bounds on $\sum_i \omega_i^2$ with $\omega_i \sim N(0,1)$, we have

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geqslant \delta\right) \leqslant 2\exp(-C_0\delta^2 m)$$

for fixed $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$.

For each of the $\binom{n}{k}$ k-dimensional subspaces in Σ_k, cover its unit sphere $\mathbb{S}^{k-1}$ with 9^k balls of radius $1/4$, by a greedy choice. Then we can use the set of centres S as anchor points by the estimate

$$\sup_{z\in\mathbb{S}^{k-1}} \left| \|Az\|_2^2 - 1 \right| \leqslant 2 \sup_{x_i\in S} \left| \|Ax_i\|_2^2 - 1 \right|.$$

Finally, we succeed if none of the anchor points fails, so by the union bound,

$$\begin{aligned}\mathbb{P}(\delta_k(A) > \delta) &\leqslant \binom{n}{k}(9^k)(2\exp(-C_0(\delta/2)^2 m)) \\ &\leqslant \exp\left(k\ln(\frac{en}{k}) + k\ln 9 + \ln 2 - \frac{C_0\delta^2}{4}m\right),\end{aligned}$$

which can be rearranged to the desired result. □

When $k = O(n^c)$ for $0 < c < 1$, this theorem says that we can reconstruct the data from $O(k \ln n)$ random measurements with high probability. Observe that even if the support of the data is known, we still need k measurements to reconstruct the data, so this result is only a logarithmic factor from optimal.*

Noise The above results do not directly apply to real-life scenarios, since it is not clear that they are either

- *stable* (works when x is almost, rather than exactly, k-sparse), or
- *robust* (works when measurements $y = Ax + e$ contain noise $\|e\|_2 \leqslant \eta$).

For these problems, the appropriate optimisation problem to consider is

$$\min_z \|z\|_1 \qquad \text{with} \quad \|Az - y\|_2 \leqslant \eta, \tag{$P_{1,\eta}$}$$

and we have the following noise-aware analogue of the main result:

Theorem. *If $\delta_{2k} < 1/3$, then ($P_{1,\eta}$) has solution z with*

$$\|z - x\|_2 \leqslant \frac{C}{\sqrt{k}} \inf_{\hat{x}\in\Sigma_k} \|\hat{x} - x\|_1 + D\eta,$$

with $C, D > 0$ depending only on δ_{2k}.

Hence the error of recovery is linear in the distance of x to Σ_k and the noise level of the measurements, which is as good as one can hope for.

Matrix completion There is also a matrix analogue to compressed sensing which has arisen in applications such as recommender systems and global positioning. In these contexts, the data is a matrix $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$ ($n_1 \leqslant n_2$), and we

*In fact, when the recovery is required to be stable (see next paragraph), the above bound is known to be optimal up to a constant factor, by a combinatorial argument.

observe its values at a subset of indices $\Omega \subseteq [n_1] \times [n_2]$. We can represent this by a linear operator on $\mathbb{R}^{n_1 \times n_2}$:

$$\mathcal{R}_\Omega(\mathbf{M}) = \sum_{(i,j)\in\Omega} m_{ij}\mathbf{e}_i\mathbf{e}_j^\mathsf{T}.$$

We replace the sparsity assumption for compressed sensing by a low-rank assumption on $\mathbf{M}$, to arrive at the *matrix completion problem*:

Problem. *Given* $r \ll n_1 \leqslant n_2$, *find the smallest* m *such that for random* $\Omega \subseteq [n_1] \times [n_2]$ *of size* m, *any matrix* $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ *of rank at most* r *can be recovered efficiently from* $\mathcal{R}_\Omega(\mathbf{M})$ *with high probability.*

We first note that this is not a well-posed problem for all $\mathbf{M}$: for instance, if $\mathbf{M}$ is zero in every row except the first, then we cannot reconstruct the first row unless we see all its entries, so $m = (1 - o(1))n_1 n_2$! The problem here is that we gain no information about the first row by looking at other rows. As such, we need some technical restrictions on $\mathbf{M}$.

Definition. For a subspace $U \subseteq \mathbb{R}^n$ with $\dim(U) = r$, the *coherence* of U is

$$\mu(U) = \frac{n}{r} \max_{1 \leqslant i \leqslant n} \|\mathbf{P}_U \mathbf{e}_i\|_2^2.$$

We have the following *incoherence conditions*:

A0: $\mu(\mathrm{RS}(\mathbf{M})), \mu(\mathrm{CS}(\mathbf{M})) \leqslant \mu_0$.

A1: If $\mathbf{M}$ has singular value decomposition $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\mathsf{T}$, with $\Sigma \in \mathbb{R}^{r \times r}$, then the entries of $\mathbf{U}\mathbf{V}^\mathsf{T}$ have absolute value $\leqslant \mu_1 \sqrt{r/n_1 n_2}$.

As before, $\mathbf{M}$ is the solution to the NP-hard optimisation problem

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \qquad \text{with} \quad \mathcal{R}_\Omega \mathbf{X} = \mathcal{R}_\Omega \mathbf{M},$$

which we can relax to the convex optimisation problem

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \qquad \text{with} \quad \mathcal{R}_\Omega \mathbf{X} = \mathcal{R}_\Omega \mathbf{M}, \tag{P_*}$$

where $\|\cdot\|_*$ is the nuclear norm (sum of singular values).

Theorem (Recht 2011). *If* $\beta > 1$, $\mathbf{M}$ *satisfies* ***A0*** *and* ***A1***, *and*

$$m \geqslant 32 \max(\mu_1^2, \mu_0) r(n_1 + n_2)\beta \ln^2(2n_2),$$

then (P_*) *has unique solution* $\mathbf{X} = \mathbf{M}$ *with probability at least*

$$1 - 6(n_1 + n_2)^{2-2\beta} \ln(n_2) - n_2^{2-2\sqrt{\beta}}.$$

By the coupon collector problem, we need $O(n_2 \ln n_2)$ observations to see at least one entry in every column; hence the above bound can be improved by a factor of at most $O(\ln n_2)$.

Proof (idea). Similar to previous approaches, Recht wants to construct an approximate solution to the dual optimisation problem. To avoid dealing with messy decoupling inequalities such as in Candès-Recht (2009), we will choose Ω uniformly with replacement from $[n_1] \times [n_2]$.

Recht reduces the problem to checking that the following hold w.h.p.:

$$\frac{n_1 n_2}{m} \left\| \mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T - \frac{m}{n_1 n_2} \mathcal{P}_T \right\| \leqslant \frac{1}{2}, \tag{1a}$$

$$\|\mathcal{R}_\Omega\| \leqslant \frac{8}{3} \sqrt{\beta} \ln(n_2), \tag{1b}$$

and there exists $\mathbf{Y} \in \mathrm{Ran}(\mathcal{R}_\Omega)$ with

$$\|\mathcal{P}_T(\mathbf{Y}) - \mathbf{U}\mathbf{V}^*\|_F \leqslant \sqrt{\frac{r}{2n_2}}, \qquad \|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < \frac{1}{2}, \tag{2}$$

for certain projection operators $\mathcal{P}_T, \mathcal{P}_{T^\perp}$ defined in terms of $\mathbf{U}$ and $\mathbf{V}^T$.

(1a) states that $\mathcal{R}_\Omega$ is nearly (a multiple of) an isometry on T. This is proven with the matrix Bernstein inequality.

(1b) is a bound on the maximum number of repetitions of any entry in Ω, which follows from a standard Chernoff bound.

For (2), $\mathbf{Y}$ is defined recursively by splitting Ω into blocks Ω_k, using the fact that each $\mathcal{R}_{\Omega_k}$ is nearly an isometry. □

References

H. Boche, et al., "A Survey of Compressed Sensing." *Compressed Sensing and its Applications*, edited by H. Boche, et al., Birkhäuser Basel, 2015, 1–37.

E. Candès, B. Recht, "Exact Matrix Completion via Convex Optimization." *Found. of Comput. Math.*, **9**(6), 2009, 717–772.

E. Candès, J. Romberg, T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information." *IEEE Trans. Inform. Theory*, **52**, 2006, 489–509.

B. Recht, "A Simpler Approach to Matrix Completion." *J. Mach. Learn. Res.*, **12**, 2011, 3413–3430.