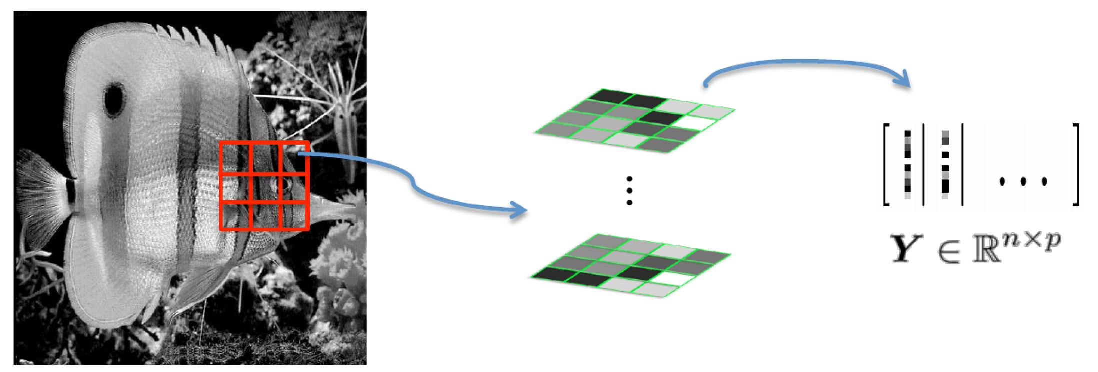
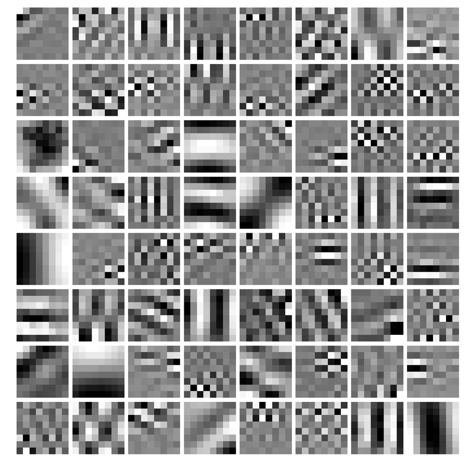
Complete Dictionary Recovery over the Sphere

Ju Sun, Qing Qu, and John Wright, Department of Electrical Engineering, Columbia University

Motivation: Dictionary Learning

Given $\boldsymbol{Y}$, find $(\boldsymbol{A}, \boldsymbol{X})$ such that $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{X}$, with $\boldsymbol{X}$ as sparse as possible.





Very successful in classical image processing, visual recognition, compressive signal requisition, and recently into deep architectures for signal classification ... limited theoretical understanding.

Problem: Dictionary Recovery

Given $n \times p$ data matrix $\boldsymbol{Y} = \boldsymbol{A}_0\boldsymbol{X}_0$ with $\boldsymbol{X}_0$ sparse, recover $\boldsymbol{A}_0$ and $\boldsymbol{X}_0$.

- Even $\boldsymbol{A}_0$ is known, seeking $\boldsymbol{X}_0$ is generally hard (sparse recovery problem!)
- Recovery only up to sign, permutation, and scale (as $\boldsymbol{A}_0\boldsymbol{X}_0 = \boldsymbol{A}_0\boldsymbol{\Pi}\Sigma * \Sigma^{-1}\boldsymbol{\Pi}^*\boldsymbol{X}_0$ for any permutation $\boldsymbol{\Pi}$ and full rank diagonal Σ)
- Due to the symmetry, hard to convexify the problem!

We focus on the case $\boldsymbol{A}_0$ is complete (square and invertible), and the coefficients $\boldsymbol{X}_0$ obeys a Bernoulli-Gaussian model: $[X_0]_{ij} = V_{ij}B_{ij}$, with $V_{ij} \sim \mathcal{N}(0,1)$ and $B_{ij} \sim \mathrm{Ber}(\theta)$.

Main Result (informal)

For any $\theta \in (0, 1/3)$, given $\boldsymbol{Y} = \boldsymbol{A}_0\boldsymbol{X}_0$ with $\boldsymbol{A}_0$ a complete dictionary and $\boldsymbol{X}_0 \sim_{i.i.d.} \mathrm{BG}(\theta)$, there is a polynomial-time algorithm that recovers $\boldsymbol{A}_0$ and $\boldsymbol{X}_0$ with high probability (at least $1 - O(p^{-6})$) whenever

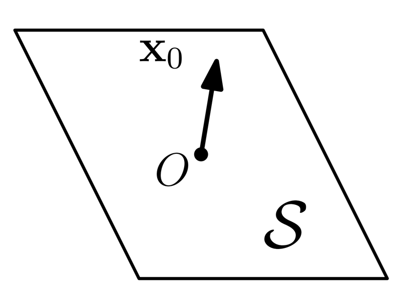
$$p \geq p_\star\left(n, 1/\theta, \kappa\left(\boldsymbol{A}_0\right), 1/\mu\right)$$

for a fixed polynomial $p_\star(\cdot)$, where $\kappa(\boldsymbol{A}_0)$ is the condition number of $\boldsymbol{A}_0$ and μ is a smoothing parameter which can be set as $\mu = cn^{-5/4}$.

Main Ingredients I - A Nonconvex Formulation

- When $\boldsymbol{A}_0$ is complete, $\mathrm{row}(\boldsymbol{Y}) = \mathrm{row}(\boldsymbol{X}_0)$.
- Rows of $\boldsymbol{X}_0$ are sparse vectors in $\mathrm{row}(\boldsymbol{Y})$. When $p \geq \Omega(n \log n)$, they are also the sparest ones! [Spielman et al '12]

Find sparsest vectors in a given linear subspace ...



Natural formulation:

$$\text{minimize} \ \ \|\boldsymbol{q}^*\boldsymbol{Y}\|_0 \quad \text{subject to} \quad \boldsymbol{q} \neq \boldsymbol{0}.$$

Convex relaxation: [Spielman et al'12]

$$\text{minimize} \ \ \|\boldsymbol{q}^*\boldsymbol{Y}\|_1 \quad \text{subject to} \quad \|\boldsymbol{q}^*\boldsymbol{Y}\|_\infty = 1.$$

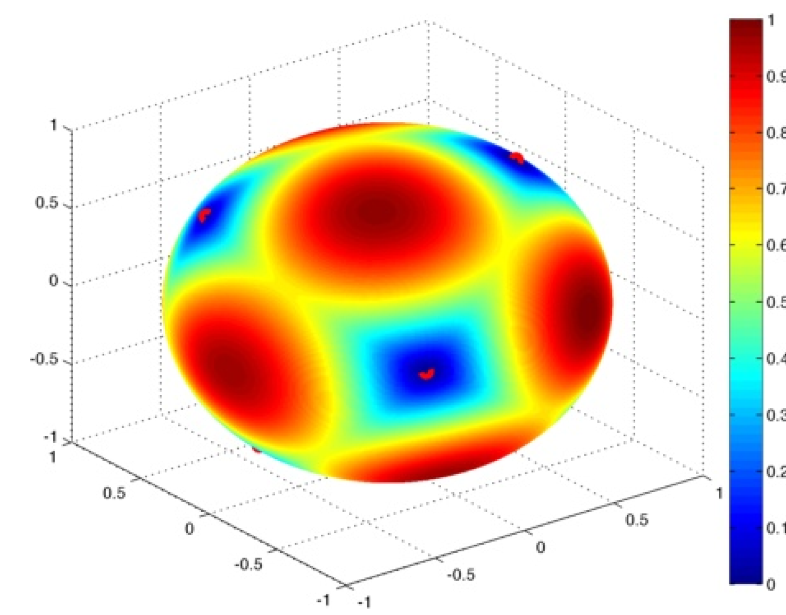
Convex relaxation is known to break down when each column of $\boldsymbol{X}_0$ contains more than $O(\sqrt{n})$ nonzeros. We look at a nonconvex "relaxation":

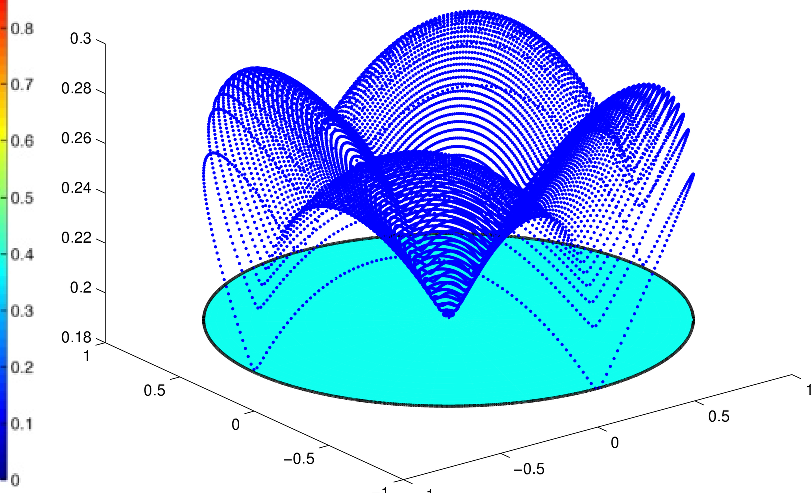
$$\boxed{\text{minimize} \ \ f(\boldsymbol{q};\boldsymbol{Y}) \doteq \frac{1}{p}\sum_{k=1}^{p} h_\mu(\boldsymbol{q}^*\boldsymbol{y}_k) \quad \text{subject to} \quad \|\boldsymbol{q}\|_2 = 1.}$$

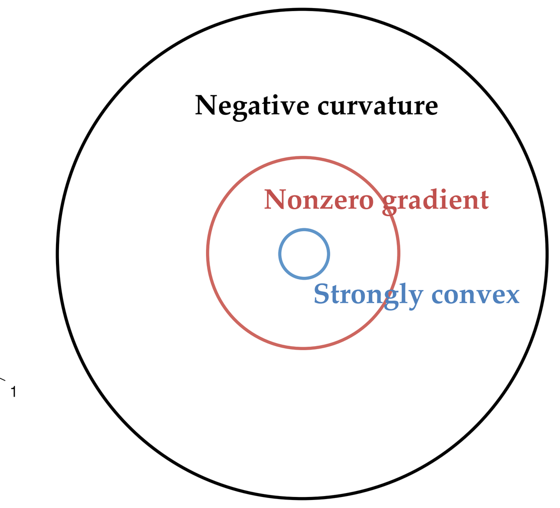
with $h_\mu(\cdot)$ a smooth approximation to the $|\cdot|$ function:

$$h_\mu(z) = \mu \log\left(\frac{\exp(z/\mu) + \exp(-z/\mu)}{2}\right) = \mu \log\cosh(z/\mu)$$

Main Ingredients II - A Glimpse into High-dimensional Geometry







Theorem: (informal) Suppose $\boldsymbol{A}_0 = \boldsymbol{I}$ and hence $\boldsymbol{Y} = \boldsymbol{A}_0\boldsymbol{X}_0 = \boldsymbol{X}_0$. For $\theta \in (0, 1/2)$ and μ sufficiently small $O(n^{-c})$, whenever $p \geq \frac{C}{\mu^2\theta^2}n^3\log\frac{n}{\mu\theta}$, the following hold simultaneously w.h.p.:

$$\nabla^2 g(\boldsymbol{w};\boldsymbol{X}_0) \succeq \frac{c_\star\theta}{\mu}\boldsymbol{I} \qquad \forall\, \boldsymbol{w} \quad \text{s.t.} \quad \|\boldsymbol{w}\| \leq \frac{\mu}{4\sqrt{2}},$$

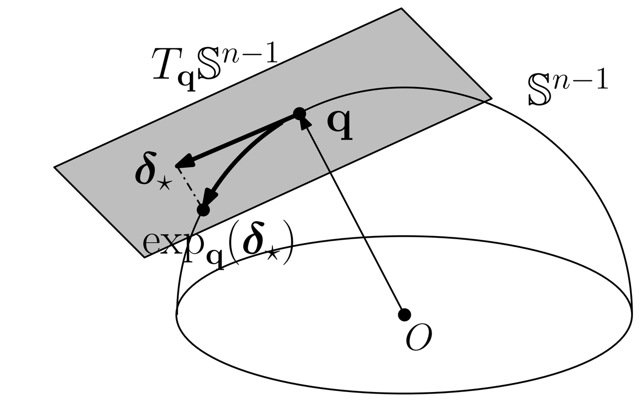
$$\frac{\boldsymbol{w}^*\nabla g(\boldsymbol{w};\boldsymbol{X}_0)}{\|\boldsymbol{w}\|} \geq c_\star\theta \qquad \forall\, \boldsymbol{w} \quad \text{s.t.} \quad \frac{\mu}{4\sqrt{2}} \leq \|\boldsymbol{w}\| \leq \frac{1}{20\sqrt{5}}$$

$$\frac{\boldsymbol{w}^*\nabla^2 g(\boldsymbol{w};\boldsymbol{X}_0)\boldsymbol{w}}{\|\boldsymbol{w}\|^2} \leq -c_\star\theta \qquad \forall\, \boldsymbol{w} \quad \text{s.t.} \quad \frac{1}{20\sqrt{5}} \leq \|\boldsymbol{w}\| \leq \sqrt{\frac{4n-1}{4n}},$$

and the function $g(\boldsymbol{w};\boldsymbol{X}_0)$ has exactly one local minimizer $\boldsymbol{w}_\star$ over the open set $\Gamma \doteq \left\{\boldsymbol{w} : \|\boldsymbol{w}\| < \sqrt{\frac{4n-1}{4n}}\right\}$, which satisfies

$$\|\boldsymbol{w}_\star - \boldsymbol{0}\| \ \leq \ \min\left\{\frac{c_c\mu}{\theta}\sqrt{\frac{n\log p}{p}}, \frac{\mu}{16}\right\}.$$

Main Ingredients III - A Riemannian Trust-region Algorithm on Sphere



Consider $\boldsymbol{q} \in \mathbb{S}^{n-1}$; for $\boldsymbol{\delta} \perp \boldsymbol{q}$, calculus gives

$$f(\exp_{\boldsymbol{q}}(\boldsymbol{\delta})) = f(\boldsymbol{q}) + \langle\boldsymbol{\delta}, \nabla f(\boldsymbol{q})\rangle + \frac{1}{2}\boldsymbol{\delta}^*\left(\nabla^2(\boldsymbol{q}) - \langle\boldsymbol{q}, \nabla f(\boldsymbol{q})\rangle\right)\boldsymbol{\delta} + O(\|\boldsymbol{\delta}\|^3)$$
$$= \widehat{f}(\boldsymbol{\delta};\boldsymbol{q}) + O(\|\boldsymbol{\delta}\|^3)$$

where $\exp_{\boldsymbol{q}}(\boldsymbol{\delta}) \doteq \boldsymbol{q}\cos\|\boldsymbol{\delta}\| + \frac{\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|}\sin\|\boldsymbol{\delta}\|$.

Basic Riemannian trust-region method:

$$\boxed{\begin{aligned}\boldsymbol{\delta}_\star &\in \underset{\boldsymbol{\delta}\in T_{\boldsymbol{q}_k}\mathbb{S}^{n-1}, \|\boldsymbol{\delta}\|\leq\Delta}{\arg\min} \ \widehat{f}(\boldsymbol{\delta};\boldsymbol{q}_k)\\ \boldsymbol{q}_{k+1} &= \exp_{\boldsymbol{q}_k}(\boldsymbol{\delta}_\star).\end{aligned}}$$

The trust-region subproblem involves a (possibly nonconvex) quadratic objective and one norm constraint. Solvable in polynomial time by root finding [More+Sorensen'83] or SDP relaxation.

References

Ju Sun, Qing Qu, John Wright. Complete dictionary recovery over the sphere. Available online: `http://arxiv.org/abs/1504.06785`

Email: <js4038, qq2105, jw2966>@columbia.edu