AMATH 840: Advanced Numerical Methods for Computational and Data Sciences

Giang Tran

Department of Applied Mathematics, University of Waterloo

Jan ~~19~~ 24, 2021

Lecture 06:

ℓ_1-Minimization and Compressive Sensing (cont'd)

- Last time:
  - Different ℓ_1-Minimization Problems and Popular ℓ_1-Minimization Algorithms.
  - Examples of Sparse Optimization Problems: How to Construct Measurement Matrix A.
- Today: Exact Recovery of Sparse Vectors via Basis Pursuit.
  - Null Space Property
  - Stability
  - Robustness

Chapter 4
"A Math. Introduction to Compressive Sensing"
by S. Foucart & H. Rauhut.

ℓ_1-Minimization Problems

- Models:
  - Basis pursuit:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad s.t. \quad Az = y. \qquad (BP)$$

$\lambda \to 0$

  - Basis pursuit denoising:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad s.t. \quad \|Az - y\|_2 \leq \sigma, \qquad (BP_\sigma)$$

  or

$$\min_{z \in \mathbb{C}^n} \frac{1}{2}\|Az - y\|_2^2 + \lambda \|z\|_1. \qquad (QP_\lambda)$$

  - Lasso:

$$\min_{z \in \mathbb{C}^n} \frac{1}{2}\|Az - y\|_2^2 \quad s.t. \quad \|z\|_1 \leq \tau. \qquad (LS_\tau)$$

- With suitable σ, λ, τ, the solutions of $BP_\sigma, QP_\lambda, LS_\tau$ coincide.
- BP vs QP_λ: $\lim_{\lambda \to 0^+} z_{QP_\lambda} = z_{bp}$, provided that the (BP) has a unique solution z_{bp}.
- Algorithms: SPGL1, SpaRSA, Primal-Dual,FISTA, Nesterov's 2nd method, Augmented Lagrangian/Split-Bregman, coordinate descent,...

ℓ_1-Minimization and Regression Problem

- **Input:** $\{x_k, y_k\}_{k=1}^m \subset \mathbb{R}^d \times \mathbb{R}$. Learn $f : \mathbb{R}^d \to \mathbb{R}$ s.t. $f(x_k) \approx y_k,\ \forall k = 1, \ldots, m$.
- **Step 1:** Split data in training and testing data.
- **Step 2:** Choose a model:
  - Linear model: Assume $f(x) = x^T w + w_0$, where $x, w \in \mathbb{R}^d, w_0 \in \mathbb{R}$.
  - Nonlinear models: Assume f can be approximated by multivariate polynomials or orthogonal bases,

$\vec{x} = (x_1, x_2, \ldots, x_d)^T$

linearized problem

$$f(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \ldots + w_{2d} x_1 x_d + w_{2d+1} x_2^2 + \ldots + w_r x_d^p.$$

p: hyperparameter

or the random feature approach - choose n randomized vectors $\omega_k \in \mathbb{R}^d$ and fix those weights:

$n \gg 1$

$$f(x) = \sum_{k=1}^{n} w_k \sigma(\langle x, \omega_k \rangle).$$

$\in \mathbb{R}^d$

Random

fix ω_k

nonlinear act.

ℓ_1-Minimization and Regression Problem

- **Step 3:** Formulate the optimization problem with mean-squared error loss – Impose sparsity constraint.

m = # training data

$$\min_{w} \left\{ \underbrace{\frac{1}{m} \sum_{k=1}^{m} |f(x_{(k)}) - y_{(k)}|^2}_{\text{Matrix mult}} + \underbrace{R(w)}_{\lambda \|w\|_1} \right\}$$

w is sparse

$$Aw \approx y$$

linear

$$f(x_k) = x_{(k)}^T \vec{w} \approx y_k$$

$$\begin{pmatrix} - & x_{(1)}^T & - \\ - & x_{(2)}^T & - \\ & \vdots & \\ - & x_{(m)}^T & - \end{pmatrix} \vec{w} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

ℓ_0 ℓ_1

- **Step 4:** Solve the optimization problem using spgl1, omp, LASSO...

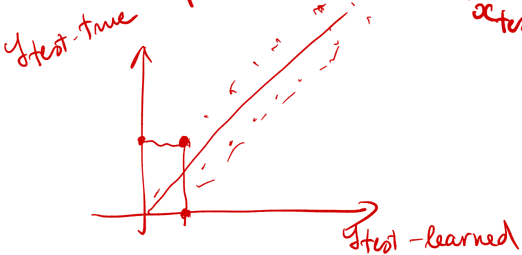
ℓ_1 # scikitlearn

- **Output:** The learned weights w and the learned function f.
- **Final Step:** Compute test errors and suitable plots.

$$MSE = \frac{1}{|test|} \| f^{\#}(x_{test}) - y_{test} \|_2^2$$

$$\text{Relative test error} = \frac{\| f^{\#}(x_{test}) - y_{test} \|_2^2}{\| y_{test} \|_2^2}$$

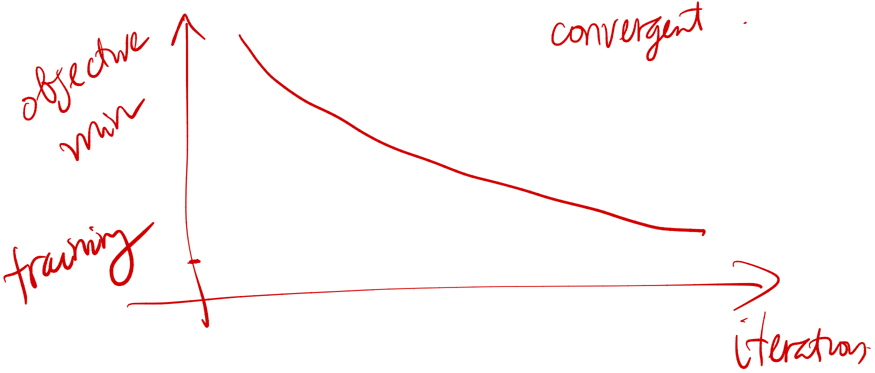
Plot



Simulated Data → Real Data

Know true soln W

plot true coef vs. learned coef.



Basis Pursuit: Reconstruction Guarantees

$Aw = Az$

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad \text{s.t.} \quad y = Az. \qquad (BP)$$

Question: Study conditions on A that ensure exact reconstruction of every sparse vector $w \in \mathbb{C}^n$ as a solution of (BP) with the vector $y \in \mathbb{C}^m$ obtained as $y = Aw$.

Ground-truth

$$w \longrightarrow y = Aw \xrightarrow{BP} \tilde{w}$$

Given

s-sparse

$\tilde{w} = w$?

$\stackrel{?}{=}$

Null Space Property

- Provide a necessary and sufficient condition for exact recovery of sparse vectors via basis pursuit.

Definition 6.1. A matrix $A \in \mathbb{K}^{m \times n}$ is said to satisfy

- The null space property (NSP) relative to a set $S \subset [n]$ if

$$\|v_S\|_1 < \|v_{S^c}\|_1 \quad \forall\ v \in \ker A \setminus \{0\}.$$

- The null space property of order s if

$$\|v_S\|_1 < \|v_{S^c}\|_1, \quad \forall\ v \in \ker A \setminus \{0\},\ \forall S \subset [n] \text{ with } |S| \leq s.$$

Here $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Null Space Property – Equivalent Conditions

"Given" $w \in \mathbb{C}^n$ sparse, solve:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad \text{s.t.} \quad Aw = Az.$$

Theorem 6.1. The following statements are equivalent:

- The matrix A satisfies the NSP of order s.
- $\|v_S\|_1 < \|v_{S^c}\|_1$ for all $v \in \ker A \setminus \{0\}$, $\forall S \subset [n]$ with $|S| \le s$.
- $2\|v_S\|_1 < \|v\|_1$ for all $v \in \ker A \setminus \{0\}$, $\forall S \subset [n]$ with $|S| \le s$.
- $2\|v_S\|_1 < \|v\|_1$ for all $v \in \ker A \setminus \{0\}$ and
  $S = \{\text{indices of } s \text{ largest ab. entries of } v\}$.
- $\|v\|_1 < 2\|v_{S^c}\|_1$ for all $v \in \ker A \setminus \{0\}$, $\forall S \subset [n]$ with $|S| \le s$.
- $\|v\|_1 < 2\sigma_s(v)_1,$ for all $v \in \ker A \setminus \{0\}$.

Recall: The ℓ_p error of best s-term approximation to a vector x is given by $\sigma_s(x)_p := \inf\{\|x - z\|_p : \quad z \in \mathbb{C}^n \text{ is } s\text{-sparse}\}$.

Null Space Property - Exact Recovery Theorem

Chapter 4

Theorem 6.2. Given $A \in \mathbb{K}^{m \times n}$, every s-sparse vector $w \in \mathbb{K}^n$ is the unique solution of

$$\min_{z \in \mathbb{K}^n} \|z\|_1 \quad \text{s.t.} \quad Aw = Az$$

if and only if A satisfies the null space property of order s.

See Theorem 4.4 in the Reference.

($\Leftarrow$) Suppose A satisfies the NSP. $\|v_S\|_1 < \|v_{S^c}\|_1 \ \forall v \in \ker A - \{0\}$

let $w \in \mathbb{C}^n$ be a s-sparse vector. $\operatorname{supp} w = S$, $|S| \leq s$

$z \in \mathbb{C}^n$: $Az = Aw$.

Claim : $\|w\|_1 < \|z\|_1$.

Notation $z_S \in \mathbb{C}^n$: 0 outside the index set S

$z_{S^c} \in \mathbb{C}^n$: 0 outside S^c

Let $v = z - w \neq 0 \in \mathbb{C}^n$. , $v \in \ker A - \{0\}$

$v_S = z_S - w_S = z_S - w$

$v_{S^c} = z_{S^c} - w_{S^c}$

$= z_{S^c}$

$$\|w\|_1 \le \|w - z_{\underline{S}}\|_1 + \|z_{\underline{S}}\|_1 = \|v_{\underline{S}}\|_1 + \|z_{\underline{S}}\|_1$$

$$\wedge$$

$$\|v_{\underline{S}^c}\|_1 + \|z_{\underline{S}}\|_1$$

$$\|$$

$$\|z_{S^c}\|_1 + \|z_{\underline{S}}\|_1 = \|z\|_1$$

($\Rightarrow$) Let $\underline{S} \subset [n]$, $|\underline{S}| \leq s$.

Take $v \in \ker A - \{0\}$. <u>Claim</u> $\|v_{\underline{S}}\|_1 < \|v_{\underline{S}^c}\|_1$.

Consider $\min_{z \in \mathbb{C}^n} \|z\|_1$ s.t $Az = Av_{\underline{S}}$ (*)

Then $v_{\underline{S}} \in \mathbb{C}^n$ is a s-sparse solution of (*)

Null Space Property - ℓ_0 and ℓ_1 Models

$\ell_1 \Rightarrow \ell_0$

Theorem 6.3. If $A \in \mathbb{K}^{m\times n}$ satisfies the NSP of order s then for every $y = Aw$ with s-sparse w, the basis pursuit problem solves the ℓ_0-minimization problem. That is, the solution of the basis pursuit problem is the solution of the ℓ_0-minimization problem.

Proof $w = \operatorname{argmin}_{Az=Aw} \|z\|_1$, w is s-sparse. $\in \mathbb{C}^n$

let $z_* \in \operatorname{argmin}_{Az=Aw} \|z\|_0$.

$\|z_*\|_0 \leq \|w\|_0 \leq s$

z_* is also a solution of BP

uniqueness of BP

$z_* = w$

Null Space Property

Theorem 6.4. If $A \in \mathbb{K}^{m\times n}$ satisfies the NSP of order s, the following matrices also satisfy the NSP of order s:

$$\hat{A} := GA, \quad \text{where } G \in \mathbb{K}^{m\times m} \text{ is some invertible matrix,}$$

$$\widetilde{A} := \begin{bmatrix} A \\ B \end{bmatrix}, \quad \text{where } \; B \in \mathbb{K}^{m'\times n}.$$

Remark:

- If $A \in \mathbb{K}^{m\times n}$ satisfies the NSP of order s, there exists matrix $H \in \mathbb{K}^{n\times n}$ such that AH does not satisfy the NSP.
- The above theorem indicates that the sparse recovery property of basis pursuit is preserved if some measurements are rescaled, reshuffled, or added.

$$\hat{A} = GA. \qquad \operatorname{Ker} \hat{A} = \operatorname{Ker} A$$

$$\tilde{A} = \left[\begin{array}{c} A \\ \hline B \end{array}\right] \qquad \operatorname{Ker} \tilde{A} \subset \operatorname{Ker} A.$$

Stable Null Space Property

Study $\min_{z \in \mathbb{C}^n} \|z\|_1$ s.t $Az = Aw$ ($y = Aw$)

where $w \in \mathbb{C}^n$ may not be sparse

Definition 6.2. A matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy

- The stable null space property with constant $0 < \rho < 1$ relative to a set $S \subset [n]$ if

NSP: $\|v_S\|_1 < \|v_{S^c}\|_1$

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 \quad \forall v \in \ker A.$$

- The stable null space property of order s with constant $0 < \rho < 1$ if

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 \quad \forall v \in \ker A, \; \forall S \subset [n] \text{ with } |S| \leq s.$$

Stable Null Space Property - Verification Theorem

Theorem 6.5. The matrix $A \in \mathbb{C}^{m \times n}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$ relative to a set $S \subset [n]$ if and only if

$$\|z - x\|_1 \leq \frac{1+\rho}{1-\rho} \left(\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1 \right),$$

for all vectors $x, z \in \mathbb{C}^n$ with $Az = Ax$.

See Theorem 4.14. in the Reference.

$$\|a+b\|_1 \leq \|a\|_1 + \|b\|_1$$

Stable Sparse Recovery

Theorem 6.6. Suppose that $A \in \mathbb{C}^{m\times n}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the basis pursuit,

$$\min_z \|z\|_1 \quad s.t. \quad Az = Aw,$$

note w can be dense

approximates the vector w with ℓ_1-error:

$$\|w - w^\#\|_1 \le \frac{2(1+\rho)}{1-\rho} \sigma_s(w)_1.$$

$= \inf \{ \|w - z\|_1$ z: s-sparse vector $\}$

$\to 0$ when w is well approx by best s-term approx

See Theorem 4.12 in the reference.

Remark: If $A \in \mathbb{C}^{m\times n}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$, the basis pursuit may have more than one solution.

Robust Null Space Property

Study $\min \|z\|_1$ s.t $\|y-Az\| \leq \eta$.

Definition 6.3. A matrix $A \in \mathbb{C}^{m\times n}$ is said to satisfy

- The robust null space property w.r.t. $\|\cdot\|$ with constants $0 < \rho < 1$ and $\tau > 0$ relative to a set $S \subset [n]$ if

$$\|v_S\|_1 \leq \rho\|v_{S^c}\|_1 + \tau\|Av\| \quad \forall v \in \mathbb{C}^n.$$

- The stable null space property of order s with constant $0 < \rho < 1$ if

$$\|v_S\|_1 \leq \rho\|v_{S^c}\|_1 + \tau\|Av\| \quad \forall v \in \mathbb{C}^n,\ \forall S \subset [n] \text{ with } |S| \leq s.$$

Robust Sparse Recovery

Theorem 6.7. Suppose a matrix $A \in \mathbb{C}^{m \times n}$ satisfies the robust null space property of order s with constant $0 < \rho < 1$ and $\tau > 0$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the BPDN:

$$\min_z \|z\|_1 \quad s.t. \quad \|Az - y\| \le \eta,$$

with $y = Aw + e$ and $\|e\| \le \eta$ approximates the vector w with ℓ_1-error:

$$\|w - w^\#\|_1 \le \frac{2(1+\rho)}{1-\rho}\sigma_s(w)_1 + \frac{4\tau}{1-\rho}\eta.$$

ℓ_2-Robust Null Space Property

RNSP : $\|v_S\|_1 \leq \rho\|v_{S^c}\|_1 + \tau\|Av\|$

Definition 6.4. A matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\| \cdot \|$ with constants $0 < \rho < 1$ and $\tau > 0$ if

2

$$\|v_S\|_2 \leq \frac{\rho}{s^{1/2}}\|v_{S^c}\|_1 + \tau\|Av\| \quad \forall v \in \mathbb{C}^n, \ \forall S \subset [n] \text{ with } |S| \leq s.$$

Theorem 6.8. Suppose the matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\| \cdot \|_2$ with constants $0 < \rho < 1$ and $\tau > 0$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the BPDN:

$$\min_z \|z\|_1 \quad s.t. \quad \|Az - y\|_2 \leq \eta,$$

with $y = Aw + e$ and $\|e\|_2 \leq \eta$ approximates the vector w with ℓ_p-error:

$$\|w - w^\#\|_p \leq \frac{C}{s^{1-1/p}}\sigma_s(w)_1 + Ds^{1/p-1/2}\eta, 1 \leq p \leq 2,$$

for some constants C, D depending only on ρ and τ.

ℓ_2-Robust Null Space Property

Definition 6.4. A matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\|\cdot\|$ with constants $0 < \rho < 1$ and $\tau > 0$ if

$$\|v_S\|_2 \leq \frac{\rho}{s^{1/2}} \|v_{S^c}\|_1 + \tau \|Av\| \quad \forall v \in \mathbb{C}^n, \ \forall S \subset [n] \text{ with } |S| \leq s.$$

Theorem 6.8. Suppose the matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\|\cdot\|_2$ with constants $0 < \rho < 1$ and $\tau > 0$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the BPDN:

$$\min_z \|z\|_1 \quad s.t. \quad \|Az - y\|_2 \leq \eta,$$

with $y = Aw + e$ and $\|e\|_2 \leq \eta$ approximates the vector w with ℓ_p-error:

$$\|w - w^\#\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(w)_1 + D s^{1/p-1/2} \eta, 1 \leq p \leq 2,$$

for some constants C, D depending only on ρ and τ.

$$\|w - w^{\#}\|_1 \leq C\, \rho_s(w)_1 + D\sqrt{s}\,\eta$$

$$\|w - w^{\#}\|_2 \leq \frac{C}{\sqrt{s}}\, \rho_s(w)_1 + D\eta$$

For supervised learning, we are interested in generalization error;

$$\|f^{\text{learned}} - f^{\text{true}}\|_2 \leq \, ?$$

Can use to estimate upper bound of $\|f^{\text{learned}} - f^{\text{true}}\|$