Topics in Numerical Analysis II
Computational Inverse Problems

Bangti Jin (b.jin@cuhk.edu.hk)

Chinese University of Hong Kong

November 9, 2022

Outline

problem setup

finite-dimensional formulation

$$b = Ax^* + \eta,$$

- $x^* \in \mathbb{R}^p$: the unknown signal
- $\eta \in \mathbb{R}^n$: additive Gaussian noise; $\epsilon = \|\eta\|$: noise level
- $A \in \mathbb{R}^{n \times p}$, $p \gg n$: (normalized column), i.e., $\|A_i\| = 1$

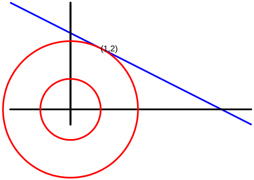
The problem has infinitely many solutions (if it has one), which one shall we take ?

insights from “exact data”

toy example: find a “reasonable” solution to the problem

$$x_1 + 2x_2 = 5$$

There are infinitely many solutions.

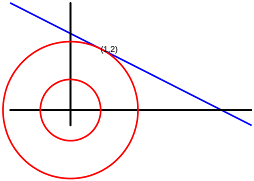


Which one shall we take ?
convention: least-squares Gauss 1809

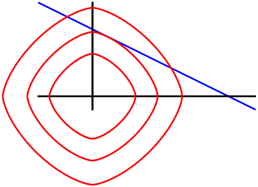
$$\min |x_1|^2 + |x_2|^2$$
$$\text{s.t.} \quad x_1 + 2x_2 = 5$$

generalized “minimum-energy” solution

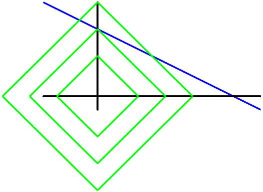
$$\min |x_1|^p + |x_2|^p, 0 \leq p \leq 2$$
$$\text{s.t.} \quad x_1 + 2x_2 = 5$$



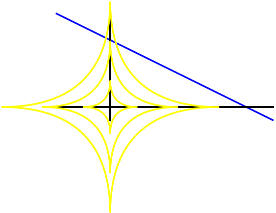
$p = 2$



$p = 3/2$



$p = 1$



$p = 1/2$

in case of noisy data: Tikhonov regularization

$$\frac{1}{2}\|Ax - b\|^2 + \alpha\psi(x)$$

two possible choices of $\psi(x)$ (convexity ...)

- classical Tikhonov regularization

$$\psi(x) = \frac{1}{2}\|x\|_2^2 =: \frac{1}{2}\sum_i |x_i|^2$$

- sparsity regularization

$$\psi(x) = \|x\|_p^p =: \frac{1}{p}\sum_i |x_i|^p, \ p \in [0, 1]$$

- general analogues ...

in case of noisy data: Tikhonov regularization

$$\frac{1}{2}\|Ax - b\|^2 + \alpha\psi(x)$$

assumption: i.i.d. additive Gaussian noise on the data

$$b_i = b_i^\dagger + \xi_i, \quad \xi_i \sim N(0, \sigma^2)$$

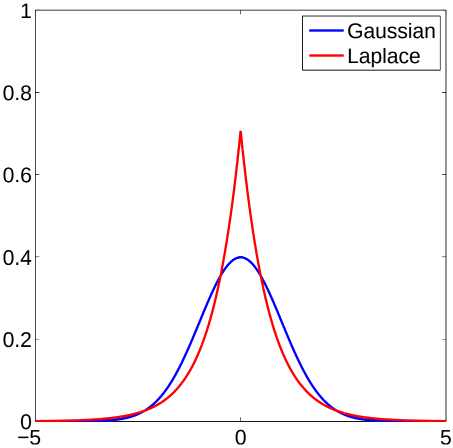
$\Rightarrow$likelihood

$$p(b|x) \propto e^{-\frac{1}{2\sigma^2}\|Ax-b\|_2^2}$$

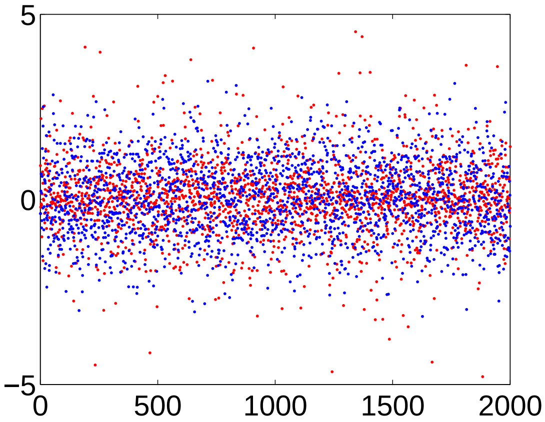
assumption: prior knowledge

$$p(x) \propto e^{-\lambda\psi(x)}$$

- classical Tikhonov regularization $\Leftrightarrow$ Gaussian prior distribution
- sparsity regularization $\Leftrightarrow$ Laplace distribution



Gaussian v.s. Laplace distribution



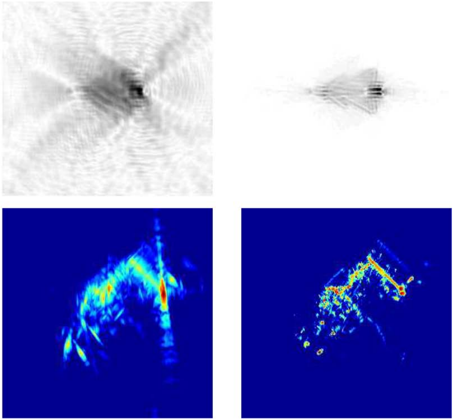
Gaussian v.s. Laplace

The energy can be more general:

$$\psi(x) = \Psi(Wx),$$

under certain transformation, e.g., wavelet, framelet, curvelet, shearlet ...
The discussions below extend to these more complex cases

an example from radar imaging



conventional v.s. sparsity

©Potter et al 2010 passive imaging, backhoe data natural idea for sparse solution is to penalize the number of unknowns

$$\frac{1}{2}\|Ax - b\|^2 + \alpha\|x\|_0$$

where

$$\|x\|_0 = \#(\text{nonzeros in } x)$$

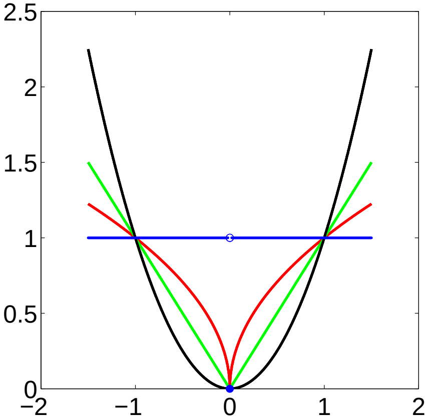
conceptually intuitive, but computationally very challenging:
approximations:

- bridge penalty

$$\|x\|_q^q = \sum |x_i|^q, \quad q \in (0, 1)$$

- l1 penalty

$$\|x\|_1 = \sum |x_i|$$


2.5
2
1.5
1
0.5
0
−2
−1
0
1
2

- The l0 penalty is the genuine choice, but VERY challenging there are different ways to approximate it ...
- lq is an approximation, and there are many others
- especially l1 is very popular, since l1 is convex
- further, there is a solid theory

Regression shrinkage and selection via the lasso

R Tibshirani - Journal of the Royal Statistical Society: Series B …, 1996 - Wiley Online Libr

We propose a new method for estimation in linear models. The 'lasso' minimizes the residua sum of squares subject to the sum of the absolute value of the coefficients being less than .

☆ Save 99 Cite Cited by 48004 Related articles All 49 versions Web of Science:

Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information EJ Candès, J Romberg, T Tao IEEE Transactions on information theory 52 (2), 489-509	18665	2006
Stable signal recovery from incomplete and inaccurate measurements EJ Candes, JK Romberg, T Tao	11708 *	2006

notation:

- $S = \{1, \ldots, p\}$
- $I \subset S$, x_I: subvector consisting of entries of x indexed by $i \in I$
- $I \subset S$, A_I: submatrix consisting of columns of A indexed by $i \in I$

restricted isometry property (RIP)

- RIP of order s, if $\exists$ a $\delta_s \in (0,1)$ s.t.

$$(1-\delta_s)\|c\|^2 \le \|A_I c\|^2 \le (1+\delta_s)\|c\|^2 \quad \forall I \subset S, |I| \le s.$$

  with δ_s being the smallest constant for which RIP holds

  $\delta_s := \inf\{\delta : (1-\delta)\|c\|^2 \le \|A_I c\|^2 \le (1+\delta)\|c\|^2\, \forall |I| \le s, \forall c \in \mathbb{R}^{|I|}\}$

  denoted by RIP (s, δ_s)
- $\delta_s \le \delta_{s'}$ if $s < s'$
- RIP $(s, \delta_s) \Rightarrow$

$$1-\delta_s \le \lambda_{min}(A_I^* A_I) \le \lambda_{max}(A_I^* A_I) \le 1+\delta_s$$

  the submatrix A_I is fairly well-conditioned
- RIP is difficult to compute

Matrices satisfying RIP

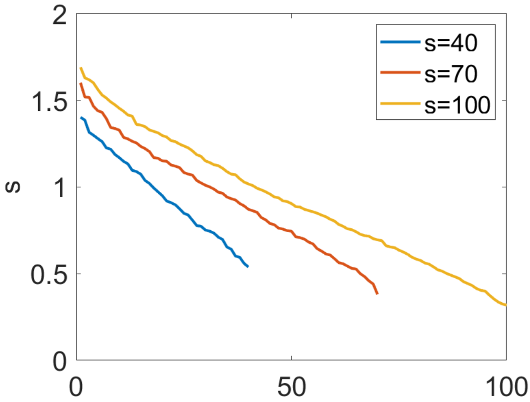
- random matrices with i.i.d. Gaussian/Bernoulli with zero mean and variance $1/p$
  RIP holds with overwhelming probability if

$$s \leq cn/\log(p/n)$$

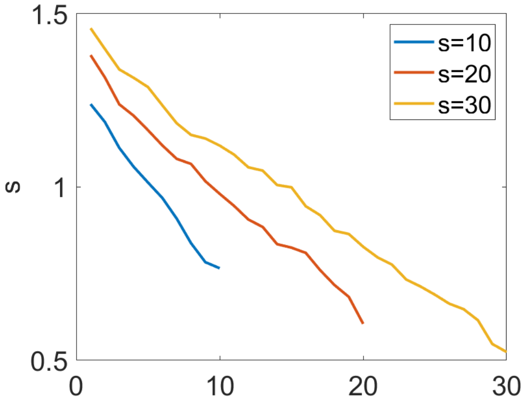
- random matrices from Fourier ensemble: randomly select n rows from $p \times p$ discrete Fourier transform matrix *uniformly at random*, and then re-normalized
  RIP holds with overwhelming probability if

$$s \leq cn/(\log(p))^6$$

DFT matrix is used in Magnetic Resonance Imaging random Gaussian matrix, $n = 200, p = 1000$



randomly subsampled Fourier matrix, $n = 100, p = 1000$



- mutual incoherence (MC) ν (easier to compute)

$$\nu = \max_{i \neq j} |\psi_i^t \psi_j| \ll 1$$

the cosine of the angle between two distinct columns

- The smaller is ν, the closer is it to be orthonormal ...
- the result in terms ν is often not sharp, but convenient to check

Fix $\delta \in (0, 1)$. For an $n \times p$ Bernoulli matrix, the coherence statistics $\nu \leq 2n^{-1/2}(\ln \frac{d}{\delta})^{1/2}$ with prob. exceeding $1 - \delta^2$.

D. Donoho, X. Huo, IEEE Trans. Inform. Theory 47(7), 2845–2862, 2001

J. Tropp, A. Gilbert. IEEE Trans. Inform. Theory 53(12), 4655, 2007

RIP and MIC do not hold for many practical inverse problems!

under certain conditions on the matrix Ψ and the true solution $x^\dagger$:

$$\|x^\dagger - x_\alpha^\delta\| \leq c\delta$$

conditions

- α by the discrepancy principle

$$\min \|x\|_{\ell^1} \qquad \text{s.t. } \|\Psi x - b^\delta\| \leq \delta$$

- with $s = \|x^\dagger\|_0$, the result holds on $\delta_{3s} + 3\delta_{4s} < 2$
- n is nearly of order s, i.e., $n \sim s$ up to log factors for random Ψ
- the constant c depends on RIP constant
- the reconstruction error is of the same order as data error δ
  much better than the classical inverse problems $\sim$ sublinear
  $\Leftarrow$ much stronger conditions

there are other methods achieving similar errors

E.J Candes, J. Romberg, T. Tao. Commun. Pure Appl. Math. 59(8), 1207–1223, 2006

benchmark performance:
suppose an oracle tells us the support $\mathcal{A}^\dagger$ of $x^\dagger$ $\Rightarrow$ oracle solution x^o

$$x^o = \begin{cases} (\Psi^t_{\mathcal{A}^\dagger}\Psi_{\mathcal{A}^\dagger})^{-1}\Psi^t_{\mathcal{A}^\dagger} b^\delta & \text{on } \mathcal{A}^\dagger, \\ 0 & \text{otherwise} \end{cases}$$

Then $x^o - x^\dagger = 0$ on $\mathcal{I}^\dagger = [n] \setminus \mathcal{A}^\dagger$, while on $\mathcal{A}^\dagger$

$$x^\dagger - x^o = (\Psi^t_{\mathcal{A}^\dagger}\Psi_{\mathcal{A}^\dagger})^{-1}\Psi^t_{\mathcal{A}^\dagger}\eta$$

By the RIP property

$$\|x^\dagger - x^o\| \approx \|\Psi^t_{\mathcal{A}^\dagger}\eta\| \approx \|\eta\|$$

Message: No recovery method can perform fundamentally better.

proof sketch

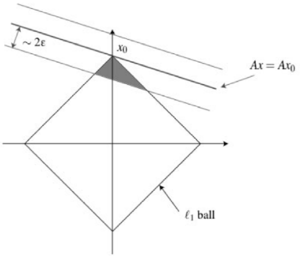
crucial observations:

- tube constraint

$$\|\Psi x^\dagger - \Psi x_\alpha^\delta\| \leq \|\Psi x^\dagger - b^\delta\| + \|\Psi x_\alpha^\delta - b^\delta\| \leq 2\delta$$

- cone constraint: with $x_\alpha^\delta = x^\dagger + h$, there holds $\|h_{\mathcal{A}^\dagger}\|_1 \geq \|h_{\mathcal{I}^\dagger}\|_1$

$$\begin{aligned}\|x^\dagger\|_1 \geq \|x_\alpha^\delta\|_1 &= \|x^\dagger + h\|_1 \\ &= \|(x^\dagger + h)_{\mathcal{A}^\dagger}\|_1 + \|h_{\mathcal{I}^\dagger}\|_1 \\ &\geq \|x^\dagger\|_1 - \|h_{\mathcal{A}^\dagger}\|_1 + \|h_{\mathcal{I}^\dagger}\|_1\end{aligned}$$



- $\|h\| \sim \|\Psi h\| \sim \delta$?!

  generally this does not hold, but under constraint + RIP, it is true proof construction

- decompose h into $\mathcal{A}^\dagger$ and $\mathcal{I}^\dagger$ (further)
  - $T_0 = \mathrm{supp}(x^\dagger)$
  - divide the set T_0^c into subsets of size M
    $T_j = \{n_\ell, (j-1)M + 1 \leq \ell \leq jM\}$
  - T_1 contains the indices of M largest coefficients of $h_{T_0^c}$, etc.

- the ℓ^2-norm of h concentrates on $T_{01} = T_0 \cup T_1$

$$\|h\|^2 \le \left(1 + \frac{|T_0|}{M}\right)\|h_{T_{01}}\|^2$$

- the kth largest value of $h_{T_0^c}$ obeys

$$|h_{T_0^c}|(k) \le \frac{\|h_{T_0^c}\|_{\ell^1}}{k}$$

$\Rightarrow$

$$\|h_{T_{01}^c}\|^2 \le \|h_{T_0^c}\|_{\ell^1}^2 / M$$

- ℓ^1 cone constraint

$$\|h_{T_{01}^c}\|^2 \le \frac{\|h_{T_0}\|_{\ell^1}^2}{M} \le \frac{\|h_{T_0}\|^2 |T_0|}{M}$$

- triangle inequality + RIP

$$\|\Psi h\| \geq \|\Psi_{T_{01}} h_{T_{01}}\| - \sum_{j\geq 2} \|\Psi_{T_j} h_{T_j}\|$$

$$\geq \sqrt{1-\delta_{M+|T_0|}}\|h_{T_{01}}\| - \sqrt{1+\delta_M}\sum_{j\geq 2}\|h_{T_j}\|$$

- $|h_{T_{j+1}}(t)| \leq \|h_{T_j}\|_{\ell^1}/M \quad \Rightarrow \quad \|h_{T_{j+1}}\|^2 \leq \|h_{T_j}\|_{\ell^1}^2/M \quad \Rightarrow$

$$\sum_{j\geq 2}\|h_{T_j}\| \leq \sum_{j\geq 1}\frac{\|h_{T_j}\|_{\ell^1}}{\sqrt{M}} \leq \frac{\|h_{T_0}\|_{\ell^1}}{\sqrt{M}}$$

$$\leq \sqrt{\rho}\|h_{T_0}\| \leq \sqrt{\rho}\|h_{T_{01}}\|,$$

with $\rho = |T_0|/M$

- crucial inequality

$$\|\Psi h\| \geq ((1-\delta_{M+|T_0|})^{1/2} - \sqrt{\rho}(1+\delta_M)^{1/2})\|h_{T_{01}}\|$$

- tube condition

$$\|h\| \leq \sqrt{1+\rho}\|h_{T_{01}}\| \leq \frac{\sqrt{1+\rho}}{C_{T_0,M}}\|\Psi h\| \leq \frac{2\sqrt{1+\rho}}{C_{T_0,M}}\delta$$

fill in the details of the proof.

convex function

convex functions: $f(x)$ is convex over its domain $\mathrm{dom}(f)$ if

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad \forall \lambda \in [0,1], x_1, x_2 \in \mathrm{dom}(f)$$

- f is concave if $-f$ is convex
- f is strictly convex if

  $$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2) \quad \forall \lambda \in (0,1), x_1 \neq x_2 \in \mathrm{dom}(f),$$

- if f differentiable

  $$f(x_2) \geq f(x_1) + (\nabla f(x_1), x_2 - x_1)$$

  first-order Taylor exp. is a global under-estimator

how to verify:

- by definition
- if f is twice differential: convex $\equiv f'' \geq 0$

if f is convex, then for $x_1, x_2 \in \mathrm{dom}(f)$ and $\lambda \in (0,1)$,
$x_1 + \lambda(x_2 - x_1) \in \mathrm{dom}(f)$
by the convexity of f

$$f(x_1 + \lambda(x_2 - x_1)) \leq \lambda f(x_2) + (1-\lambda) f(x_1),$$

i.e.,

$$f(x_2) \geq f(x_1) + \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda}$$

sending $\lambda \to 0^+$

$$f(x_2) \geq f(x_1) + (\nabla f(x_1), x_2 - x_1)$$

examples

- $f(x) = x^2$, $f(x) = |x|$, $f(x) = \|Ax - b\|^2$
- exponential function e^{ax} with $a \in \mathbb{R}$
- entropy function $x \log x$ on $\mathbb{R}^+$ is convex
- nonconvex: $f(x) = |x|^{1/2}$
- $\log x$, x^α, $\alpha \in [0, 1]$ is concave on $\mathbb{R}^+$

properties: convexity preserves under

- nonnegative weighted sum
- pointwise maximization/superemum
- composition

composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$

$$f(x) = h(g(x))$$

f is convex

- g convex, h convex, h nondecreasing
- g concave, h convex, h nonincreasing

$$f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x)$$

example: $e^{g(x)}$ if g is convex, $1/g(x)$ is g is concave and positive

ℓ^1 term is not differentiable, but a generalized derivative exists

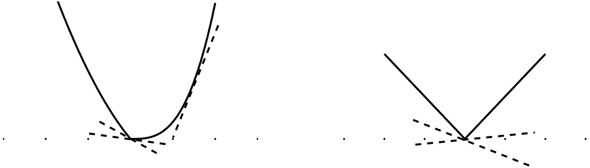
- a vector $g \in \mathbb{R}^n$ is a subgradient of a convex function $f(x) : \mathbb{R}^n \to \mathbb{R}$ at x^0 if

$$f(x) - f(x^0) \geq \langle x - x^0, g \rangle \quad \forall x \in \mathrm{dom}(f)$$

  i.e.,

$$f(x) \geq f(x_0) + \langle x - x^0, g \rangle \quad \forall x \in \mathrm{dom}(f)$$

- the set of subgradient at x^0 is denoted by $\partial f(x^0)$
- if f is differentiable at x^0, then it is identical with $f'(x^0)$



the subdifferential of $f(t) = |t|$

- at $t \neq 0$, f is differentiable, $\partial f(t) = \{f'(t)\}$, i.e.,

$$\partial f(t) = \mathrm{sign}(t), \quad t \neq 0$$

- at $t = 0$, $f(t)$ is not differentiable: any constant c s.t.

$$|t| = f(t) \geq f(0) + c(t - 0) = ct \quad \forall t \in \mathbb{R}$$

$\Rightarrow -1 \leq c \leq 1$, i.e. $(\partial|t|)(0) = [-1, 1]$

Hence, $\partial|t|$

$$\partial(|t|) = \begin{cases} 1, \ t > 0, \\ -1, \ t < 0, \\ [-1, 1], \ t = 0. \end{cases}$$

property

- x^* is a minimizer to f if and only if $0 \in \partial f(x^*)$
- sum rules (under certain mild conditions)