Sparse Optimization and Discrete-Valued Signal Reconstruction

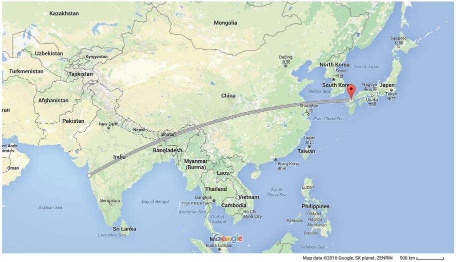
Masaaki Nagahara[1,2]

[1]The University of Kitakyushu
nagahara@kitakyu-u.ac.jp
[2]Visiting Faculty at IITB
(SysCon, 1st floor, visiting faculty room)

15 February 2017, IIT Bombay

Where is Kitakyushu?!

- North part of Kyushu island, Japan. ("Kita"="North")
- Far east from India.
- Very different culture from India:
  - Bushido (Samurai Spirit[1]), Buddhism (from India), Chop Sticks (from China), Animation ("Anime" in short), TV Games (Nintendo, Sega, Capcom, etc), . . .



[1]A samurai (soldier) who fights for family, friends, society, etc is much stronger than one who fights for himself.

The University of Kitakyushu

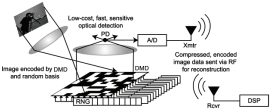
- I am working with The University of Kitakyushu
- Faculty of Environmental Engineering
- Control theory, signal/image processing, artificial intelligence, autonomous vehicles (including drones), and so on.
- *We welcome foreign students for master and PhD degrees.*
  - Short term visit is also OK (There are some funding schemes).
  - If you are interested, please visit my office (1st floor, faculty staff room), or send me email (nagahara@ieee.org).

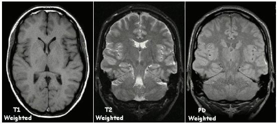
Table of Contents

Table of Contents

Sparsity in Engineering

- Image processing
  - single-pixel camera, compressed sensing MRI
- Statistics
  - big data analysis





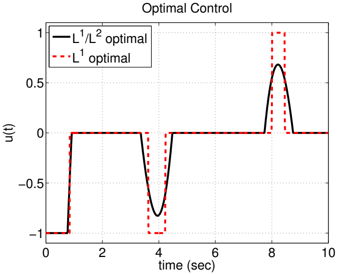
"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

Sparsity in Engineering

- Image processing
  - single-pixel camera, compressed sensing MRI
- Statistics
  - big data analysis
- *Discrete signal processing*
  - binary image reconstruction, digital communications

Sparsity in Engineering

- Image processing
  - single-pixel camera, compressed sensing MRI
- Statistics
  - big data analysis
- *Discrete signal processing*
  - binary image reconstruction, digital communications
- *Control* [2]
  - networked control, sparse control, discrete-valued control



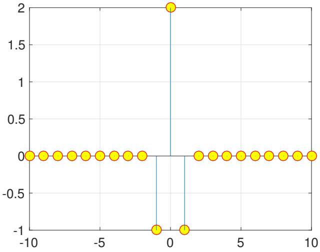
[2]This will be presented in my second seminar on 01/Mar/2017.

What is sparsity?

- A vector x in $\mathbb{R}^n$ is *sparse* if it contains many 0's, or has small ℓ^0 norm

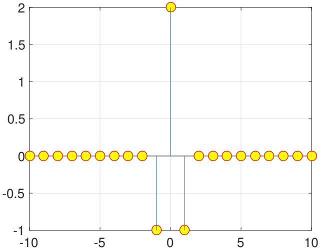
$$\|x\|_0 \triangleq \text{the number of the nonzero elements in } x.$$

- Examples of sparse vectors
  - Frequency domain data of natural signals and images; almost all of them are nearly 0 except for low-frequency data.
  - Pulse signals; they are sparse in the time domain.

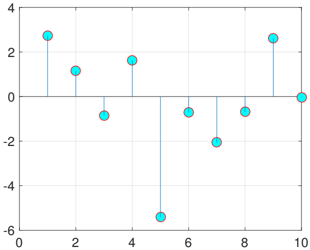


Sparse signal reconstruction

- Suppose that a sparse signal $x \in \mathbb{R}^n$ is measured by linear measurements
$$y = \Phi x \in \mathbb{R}^m,$$
where $\Phi \in \mathbb{R}^{m \times n}$ is a known matrix (we assume Φ has full row rank).
- Finding the original x is ill-posed if $m < n$.
- To determine *one* vector from y, we adopt *optimization*.



$\xrightarrow{\Phi}$



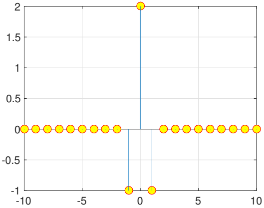
Sparse optimization

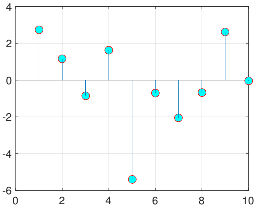
- The following optimization will do for sparse signal reconstruction:

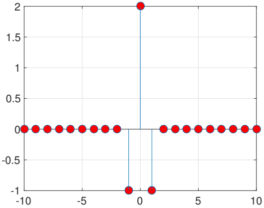
$$\min_{z \in \mathbb{R}^n} \|z\|_0 \text{ subject to } y = \Phi z.$$

- This gives the exact reconstruction (with assumptions on z and Φ).
- However, it is hard to solve if n is very large (e.g. 1 milion).
- In many cases, the following ℓ^1 optimization solves the problem:

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$







How to solve this?

ℓ^1 Optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

Answer

ℓ^1 Optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

Use MATLAB CVX[3]

```
cvx_begin
    variable z(n)
    minimize( norm(z, 1) )
    subject to
        y == Phi * z
cvx_end
```

[3] M. Grant & S. Boyd, http://cvxr.com/cvx, 2013.

Answer

ℓ^1 Optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

Use MATLAB CVX[3]

```
cvx_begin
    variable z(n)
    minimize( norm(z, 1) )
    subject to
        y == Phi * z
cvx_end
```

This is nice for small or middle scale problems.

[3] M. Grant & S. Boyd, http://cvxr.com/cvx, 2013.

Fast Algorithms for ℓ^1 optimization

ℓ^1 Optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

- General purpose toolbox (MATLAB CVX, Python CVXPY, etc) is very useful but relatively slow.
- For large-scale problems that need real-time computation, you may need a *custom-made* algorithm.
- Fast algorithms for ℓ^1 optimization

Preliminaries

effective domain

The effective domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathrm{dom}(f) \triangleq \left\{ z \in \mathbb{R}^n : f(z) < \infty \right\}$$

Preliminaries

effective domain

The effective domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathrm{dom}(f) \triangleq \{z \in \mathbb{R}^n : f(z) < \infty\}$$

epigraph

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathrm{epi}(f) \triangleq \{(z, t) \in \mathbb{R}^n \times \mathbb{R} : f(z) \leq t\}$$

Preliminaries

effective domain

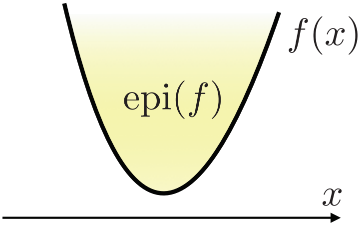
The effective domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathrm{dom}(f) \triangleq \{z \in \mathbb{R}^n : f(z) < \infty\}$$

epigraph

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

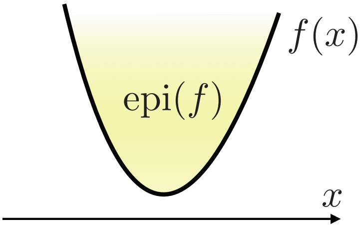
$$\mathrm{epi}(f) \triangleq \{(z, t) \in \mathbb{R}^n \times \mathbb{R} : f(z) \leq t\}$$



Proper, closed and convex function

Let us consider a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

1. f is convex iff $\mathrm{epi}(f)$ is convex.
2. f is closed iff $\mathrm{epi}(f)$ is closed.
3. f is proper iff $\mathrm{epi}(f)$ is non-empty



Preliminaries

Proximal operator

Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper, closed, and convex function. The proximal operator $\mathrm{prox}_{\gamma f}$ with parameter $\gamma > 0$ is defined by

$$\mathrm{prox}_{\gamma f}(v) \triangleq \underset{z\in\mathrm{dom}(f)}{\arg\min}\left\{ f(z) + \frac{1}{2\gamma}\|z-v\|_2^2 \right\}.$$

Preliminaries

Proximal operator

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, closed, and convex function. The proximal operator $\mathrm{prox}_{\gamma f}$ with parameter $\gamma > 0$ is defined by

$$\mathrm{prox}_{\gamma f}(v) \triangleq \underset{z \in \mathrm{dom}(f)}{\arg\min} \left\{ f(z) + \frac{1}{2\gamma} \|z - v\|_2^2 \right\}.$$

OK. But, what is it?!

Proximal operator

Proximal operator

$$\mathrm{prox}_{\gamma f}(v) \triangleq \underset{z \in \mathrm{dom}(f)}{\arg\min} \left\{ f(z) + \frac{1}{2\gamma} \|z - v\|_2^2 \right\}.$$

- $\gamma = \infty$: Minimizer of $f(z)$:

$$\mathrm{prox}_{\gamma f}(v) = \underset{z \in \mathrm{dom}(f)}{\arg\min}\ f(z)$$

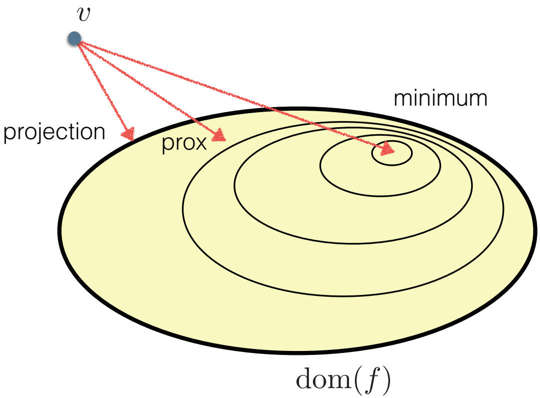
- $\gamma = 0$: Projection onto $\mathrm{dom}(f)$:

$$\mathrm{prox}_{\gamma f}(v) = \underset{z \in \mathrm{dom}(f)}{\arg\min}\ \frac{1}{2\gamma} \|z - v\|_2^2$$

- $\gamma \in (0, \infty)$: the mix of those.

Proximal operator

$$\mathrm{prox}_{\gamma f}(v) \triangleq \operatorname*{arg\,min}_{z \in \mathrm{dom}(f)} \left\{ f(z) + \frac{1}{2\gamma} \|z - v\|_2^2 \right\}.$$

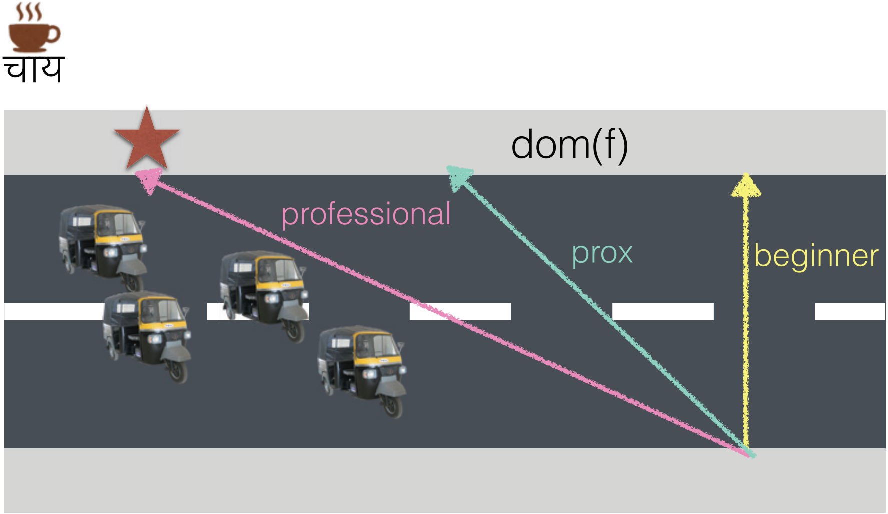


Proximal operator

The "crossing the street" problem.

Proximal operator

The "crossing the street" problem.



Indicator function and its prox

Indicator function

For a subset C in $\mathbb{R}^n$, the indicator function is defined by

$$I_C(z) = \begin{cases} 0, & z \in C \\ +\infty, & z \notin C \end{cases}$$

If C is a non-empty, closed, and convex set, then $I_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed, and convex function.

Let $C \subset \mathbb{R}^n$ be a non-empty, closed, and convex set. Then

$$\begin{aligned} \operatorname{prox}_{\gamma I_C}(v) &= \operatorname*{arg\,min}_{z \in \operatorname{dom}(I_C)} I_C(z) + \frac{1}{2\gamma}\|z - v\|_2^2 \\ &= \operatorname*{arg\,min}_{z \in C} \frac{1}{2\gamma}\|z - v\|_2^2 \\ &= P_C(v) \quad : \text{the projection operator onto } C \end{aligned}$$

ℓ^1 norm and its prox

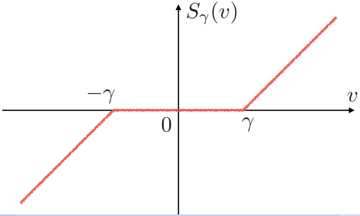
$$\mathrm{prox}_{\gamma\|\cdot\|_1}(v) = \arg\min_z \left\{ \|z\|_1 + \frac{1}{2\gamma}\|z - v\|_2^2 \right\}$$

This can be solved in a closed form:

$$\left[\mathrm{prox}_{\gamma\|\cdot\|_1}(v)\right]_i = S_\gamma(v_i),$$

where $S_\gamma : \mathbb{R} \to \mathbb{R}$ is the soft-thresholding operator defined by

$$S_\gamma(v) = \begin{cases} v - \gamma, & v \geq \gamma, \\ 0, & -\gamma < v < \gamma, \\ v + \gamma, & v \leq -\gamma. \end{cases}$$



Fixed-point of algorithm

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, closed, and convex function. Fix $\gamma > 0$ arbitrarily. Then

$$z^* = \arg\min_z f(z) \quad \text{iff} \quad z^* = \mathrm{prox}_{\gamma f}(z^*)$$

Fixed-point of algorithm

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, closed, and convex function. Fix $\gamma > 0$ arbitrarily. Then

$$z^* = \arg\min_z f(z) \quad \text{iff} \quad z^* = \mathrm{prox}_{\gamma f}(z^*)$$

This leads to the fixed-point algorithm:

$$z[k+1] = \mathrm{prox}_{\gamma_k f}(z[k]), \quad k = 0, 1, 2, \ldots$$

Fixed-point of algorithm

Theorem

Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper, closed, and convex function. Fix $\gamma > 0$ arbitrarily. Then

$$z^* = \arg\min_z f(z) \quad \text{iff} \quad z^* = \mathrm{prox}_{\gamma f}(z^*)$$

This leads to the fixed-point algorithm:

$$z[k+1] = \mathrm{prox}_{\gamma_k f}(z[k]), \quad k = 0, 1, 2, \ldots$$

Now, let us go back to our ℓ^1 optimization problem!

ℓ^1 optimization problem

ℓ^1 optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

- The indicator function for $C \triangleq \{z \in \mathbb{R}^n : y = \Phi z\}$:

$$I_C(z) = \begin{cases} 0, & \text{if } \ y = \Phi z, \\ +\infty, & \text{otherwise} \end{cases}$$

ℓ^1 optimization problem

ℓ^1 optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

- The indicator function for $C \triangleq \{z \in \mathbb{R}^n : y = \Phi z\}$:

$$I_C(z) = \begin{cases} 0, & \text{if } \; y = \Phi z, \\ +\infty, & \text{otherwise} \end{cases}$$

- Equivalent unconstrained problem:

$$\min_{z \in \mathbb{R}^n} f(z), \quad f(z) \triangleq \|z\|_1 + I_C(z).$$

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed, and convex function.

ℓ^1 optimization problem

ℓ^1 optimization

$$\min_{z\in\mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

- The indicator function for $C \triangleq \{z \in \mathbb{R}^n : y = \Phi z\}$:

$$I_C(z) = \begin{cases} 0, & \text{if } \ y = \Phi z, \\ +\infty, & \text{otherwise} \end{cases}$$

- Equivalent unconstrained problem:

$$\min_{z\in\mathbb{R}^n} f(z), \quad f(z) \triangleq \|z\|_1 + I_C(z).$$

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed, and convex function.

- Proximal operator $\mathrm{prox}_{\gamma f}$:

$$\mathrm{prox}_{\gamma f}(v) = \arg\min_z \Big\{ \|z\|_1 + I_C(z) + \frac{1}{2\gamma}\|z - v\|_2^2 \Big\}$$

ℓ^1 optimization problem

ℓ^1 optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

- The indicator function for $C \triangleq \{z \in \mathbb{R}^n : y = \Phi z\}$:

$$I_C(z) = \begin{cases} 0, & \text{if } \ y = \Phi z, \\ +\infty, & \text{otherwise} \end{cases}$$

- Equivalent unconstrained problem:

$$\min_{z \in \mathbb{R}^n} f(z), \quad f(z) \triangleq \|z\|_1 + I_C(z).$$

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed, and convex function.

- Proximal operator $\mathrm{prox}_{\gamma f}$:

$$\mathrm{prox}_{\gamma f}(v) = \arg\min_z \Big\{ \|z\|_1 + I_C(z) + \frac{1}{2\gamma}\|z - v\|_2^2 \Big\}$$

- *But, nothing is solved?!*

The secret of proximal algorithm

- Why the proximal method so useful??

The secret of proximal algorithm

- Why the proximal method so useful??
- The proximal method is useful and leads to fast algorithms *only when the prox can be obtained in a closed form*.

The secret of proximal algorithm

- Why the proximal method so useful??
- The proximal method is useful and leads to fast algorithms *only when the prox can be obtained in a closed form*.
- Unfortunately, there is no closed form for the prox of $\|z\|_1 + I_C(z)$.

The secret of proximal algorithm

- Why the proximal method so useful??
- The proximal method is useful and leads to fast algorithms *only when the prox can be obtained in a closed form*.
- Unfortunately, there is no closed form for the prox of $\|z\|_1 + I_C(z)$.
- But, don't mind. Try *splitting*.

Douglas-Rachford splitting algorithm

Douglas-Rachford splitting algorithm

For $\min_z \{f(z) + g(z)\}$, we have

$$z[k] = \mathrm{prox}_{\gamma_k f}(y[k])$$
$$y[k+1] = y[k] + \mathrm{prox}_{\gamma_k g}(2z[k] - y[k]) - z[k]$$

For appropriately chosen γ_k, $z[k]$ converges (one of) the optimal solution(s).

- Now, $f(z) = \|z\|_1$ and $g(z) = I_C(z)$.
- $\mathrm{prox}_{\gamma\|\cdot\|_1}(v)$ is given by the soft-thresholding operator $S_\gamma(v)$.
- $\mathrm{prox}_{\gamma I_C(z)}(v)$ is the projection operator onto $\{z : y = \Phi z\}$:

$$P_C(v) = \Phi^\top(\Phi\Phi^\top)^{-1}\Phi v + \Phi^\top(\Phi\Phi^\top)^{-1} y$$

An algorithm for ℓ^1 optimization

ℓ^1 optimization

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } y = \Phi z.$$

Algorithm

For $k = 0, 1, 2, \ldots$

$$\begin{aligned} z[k] &= S_{\gamma_k}(y[k]) \\ y[k+1] &= y[k] + M(2z[k] - y[k]) + w - z[k] \end{aligned}$$

where

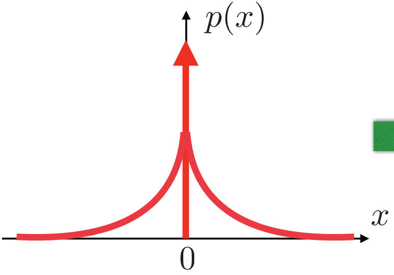
$$M \triangleq \Phi^\top (\Phi \Phi^\top)^{-1} \Phi, \quad w \triangleq \Phi^\top (\Phi \Phi^\top)^{-1} y$$

Note that this algorithm only requires the element-wise thresholding in S_{γ_k} and matrix-vector multiplications at each step.

Table of Contents

Sparse signals

- Probability distribution of sparse vectors
  - Dirac delta at $x = 0$ (discrete distribution)
  - continuous distribution for $x \neq 0$

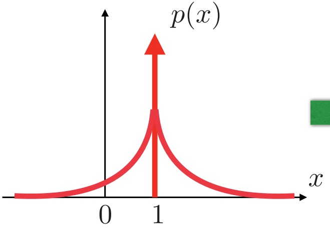


$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

$$\|z\|_0 : \text{small}$$

Signals that contain many 1's

- Probability distribution of many-1 vectors
  - Dirac delta at $x = 1$ (discrete distribution)
  - continuous distribution for $x \neq 1$



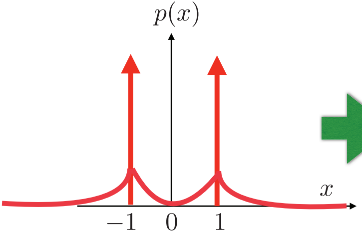
$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

$$\|z - 1\|_0 : \text{small}$$

$z - 1$: subtracts scalar 1 from each element z_i of vector z

Signals that contain many binary values ± 1

- Probability distribution
  - Dirac deltas at $x = \pm 1$ (discrete distribution)
  - continuous distribution for $x \neq \pm 1$
- If $\mathbb{P}[x = 1] = \mathbb{P}[x = -1]$, then



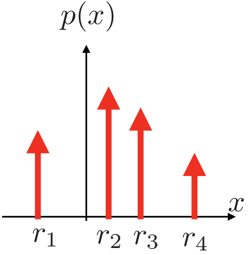
$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

$$\frac{1}{2}\|z - 1\|_0 + \frac{1}{2}\|z + 1\|_0 : \text{small}$$

Discrete signals

- Discrete signal z on a finite alphabet, $\{r_1, r_2, \ldots, r_L\}$
- Probability distribution is Dirac deltas at $x = r_1, r_2, \ldots, r_L$.

$$\mathbb{P}[x = r_j] = p_j, \quad p_j > 0, \quad p_1 + p_2 + \cdots + p_L = 1.$$



- The weighted sum of ℓ^0 norms

$$p_1\|z - r_1\|_0 + p_2\|z - r_2\|_0 + \cdots + p_L\|z - r_L\|_0$$

is small.

Discrete-valued signal reconstruction

- A binary signal $x \in \{1, -1\}^n$ whose entries are drawn from

$$\mathbb{P}[x = \pm 1] = 1/2.$$

- Incomplete linear measurement

$$y = \Phi x \in \mathbb{C}^m, \quad \text{with } m \ll n$$

- Reconstruct x from y (discrete signal reconstruction)

Sum-of-absolute-values optimization

- Observing that

$$\frac{1}{2}\|x-1\|_0 + \frac{1}{2}\|x+1\|_0$$

is small, we can say that *the sum of absolute values (SOAV)*

$$\frac{1}{2}\|x-1\|_1 + \frac{1}{2}\|x+1\|_1$$

is also small.

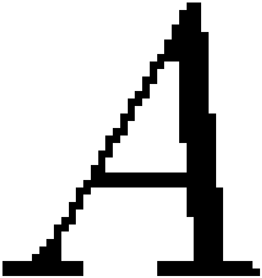
- Solve the SOAV optimization

$$\min_{z\in\mathbb{R}^n} \frac{1}{2}\|z-1\|_1 + \frac{1}{2}\|z+1\|_1 \text{ subject to } y = \Phi z$$

- In many cases, this will also do!
  - See [Nagahara, IEEE SPL, Oct. 2015]

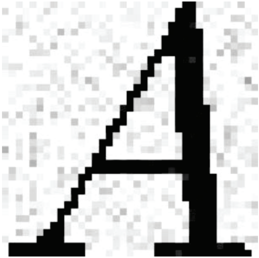
Binary image reconstruction

- Original image



Binary image reconstruction

- Original image disturbed by Gaussian noise



- Measurement: FFT and downsampling by 2
  - incomplete linear measurement

Binary image reconstruction

- Reconstruction by SOAV

Binary image reconstruction

- Reconstruction by Basis Pursuit (ℓ^1 optimization)

Discrete signal reconstruction

- Binary (or low-bit) image reconstruction [Nagahara IEEE SPL 2015]
- Digital communications [Sasahara, Hayashi, Nagahara, IEEE SPL 2016]
- Discrete-valued control [Ikeda, Nagahara, Ono, IEEE TAC 2017] (to appear)

Conclusion

- Sparsity plays an important role in signal/image processing.
- Sparse optimization can be *efficiently* solved via ℓ^1 optimization.
- Connection between sparsity and discreteness.
- Applications to *control* (the topic of the next seminar).

Conclusion

- Sparsity plays an important role in signal/image processing.
- Sparse optimization can be *efficiently* solved via ℓ^1 optimization.
- Connection between sparsity and discreteness.
- Applications to *control* (the topic of the next seminar).

ध्यान देने के लिए आपका धन्यवाद