Application of Linear System in LArTPC reconstruction

Chao Zhang

BROOKHAVEN NATIONAL LABORATORY

The Most Important Math Equation in Human History

The Most Important Math Equation in Human History

$$y = Ax$$

The Most Important Math Equation in Human History

- ~2nd century: First appearance in an antient Chinese book: "The Nine Chapters on the Mathematical Art (九章算术)"
- 16 -18th century: Systematically studied by many famous European mathematicians: *Descartes, Leibniz, Cramer, Gauss, Grassmann, …*
- Now: modern applications everywhere

$$y = Ax$$

- Many problems can be reduced to System of Linear Equations
  - or its variant forms
- 80% of the work is to figure out how to write out the equations:
  - y: the measurements
  - x: the unknowns
  - A: the connection between y and x.

A Classical Application: Digital Signal Processing

- Fundamental theory of linear time-invariant (LTI) system:
  - Any LTI system can be characterized entirely by a single function called the system's **impulse response**

$$y = Ax$$

- y: measured discrete-time signal
- x: the (unknown) true signal
- A:??

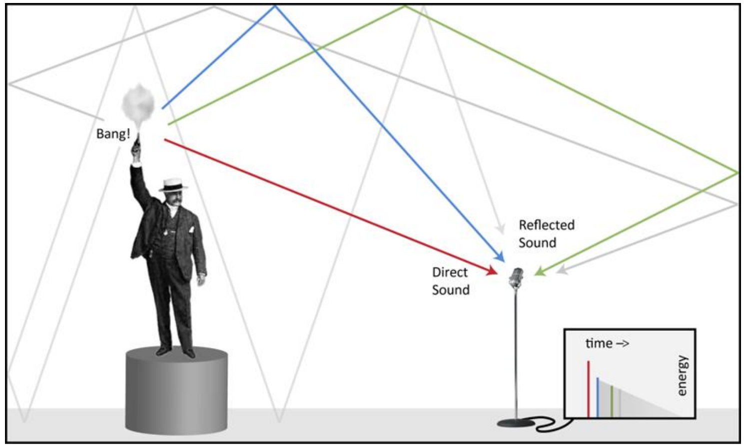
$$x(t) \longrightarrow \boxed{h(t)} \longrightarrow y(t) = h(t) * x(t)$$

$$= \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau$$

digitize

Impulse response

Room acoustic response



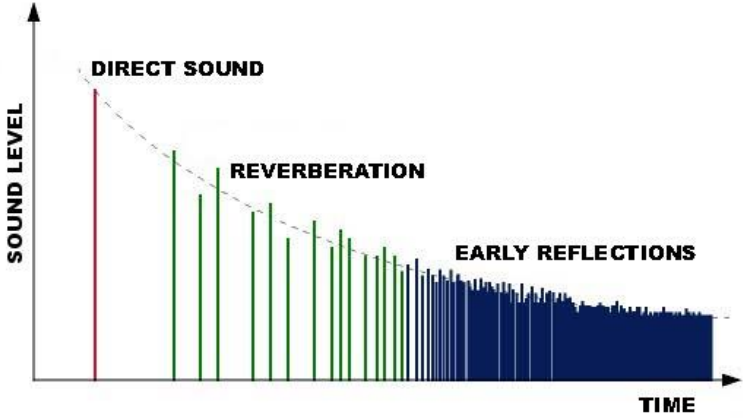
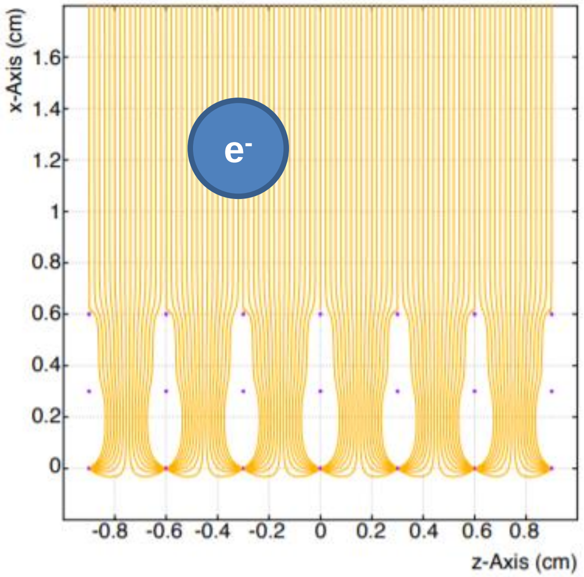
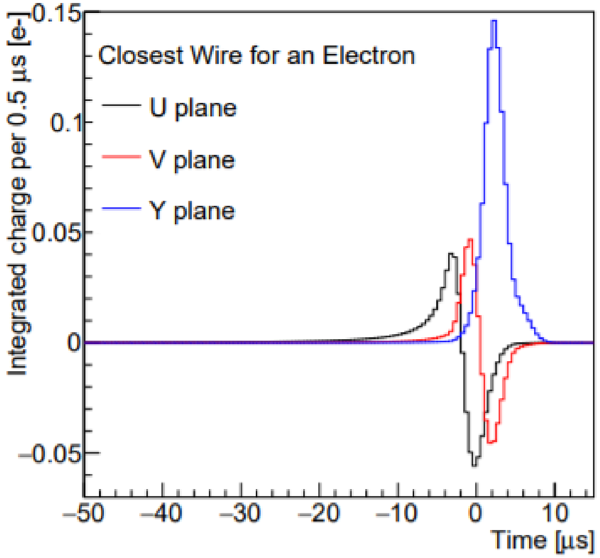


Image credit: www.prosoundweb.com

$$\boxed{h(t)}$$

LArTPC
field + electronic
response

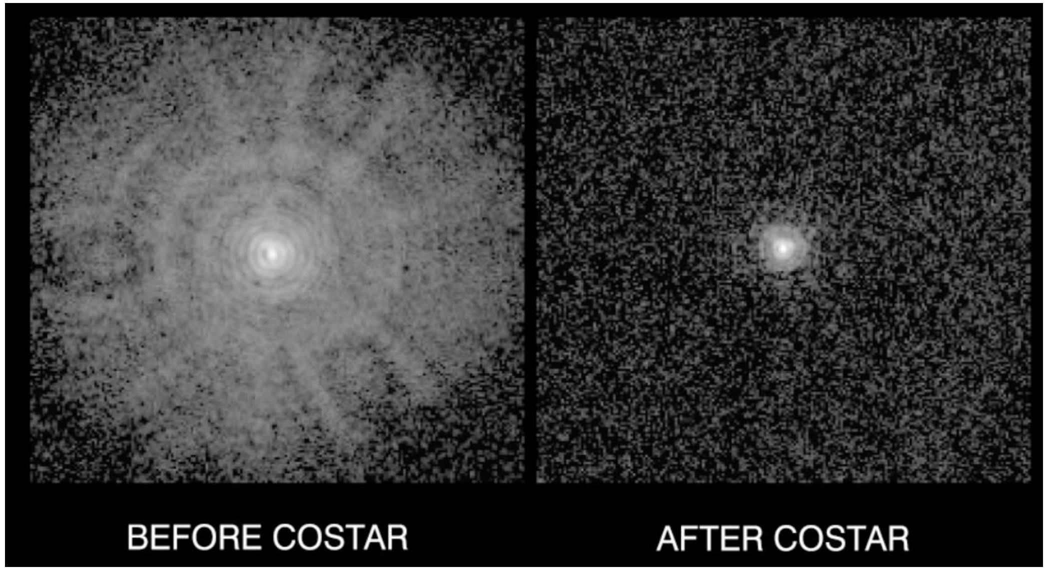




JINST 13 P07007 (2018)

2D Impulse response

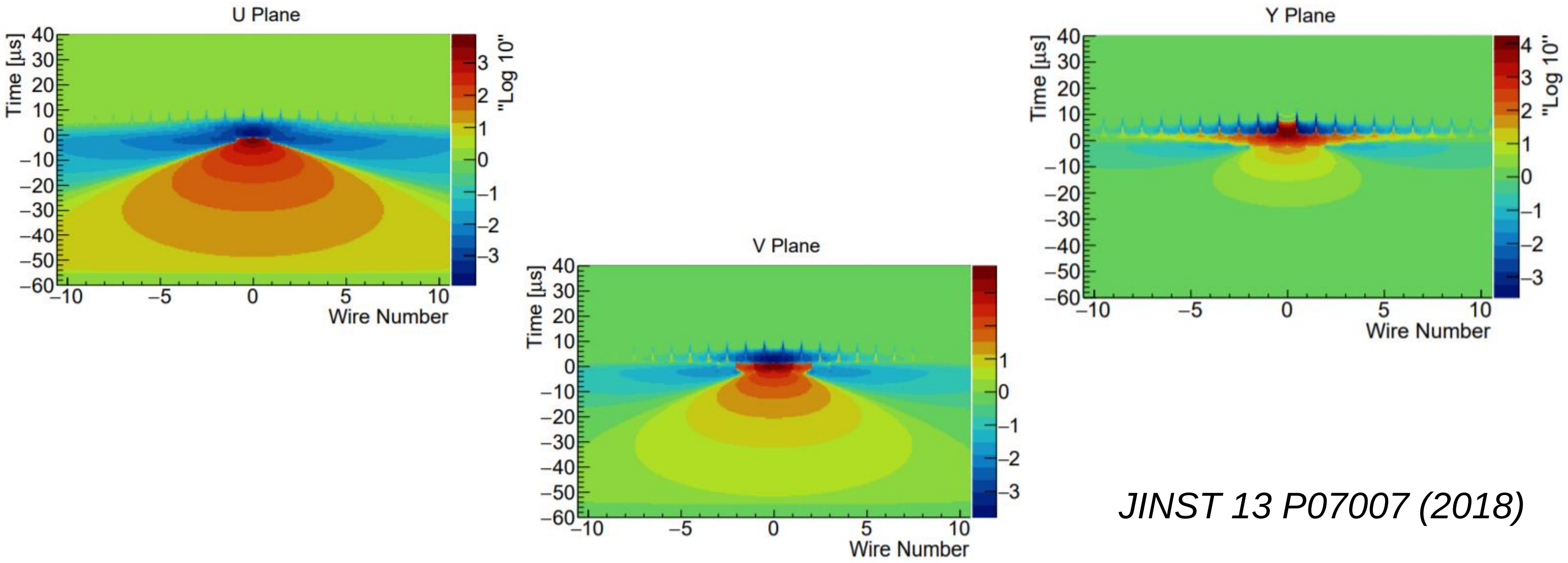
Camera optical response



Hubble telescope impulse response (Image credit: http://web.mit.edu)

$$\boxed{h(x, y)}$$

LArTPC 2D response:
y: t * drift velocity
x: wire number * pitch



JINST 13 P07007 (2018)

Change of Basis

$$\begin{pmatrix} & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \end{pmatrix}$$

Time Domain

$$\mathrm{x}(t,w) \rightarrow \boxed{h(t,w)} \rightarrow \mathrm{y}(t,w) = h(t,w) * x(t,w)$$

IDFT. DFT DFT

Frequency Domain

$$X(f,s) \rightarrow \boxed{H(f,s)} \rightarrow \mathrm{Y}(f,s) = H(f,s) \cdot X(f,s)$$

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

- Discrete Fourier Transform;
  - Convolution -> multiplication: matrix diagonalization
  - FFT algorithm: $O(N^2) \rightarrow O(N \log N)$: fast computation
  - Frequency-domain filters: reduce noise, regularize fluctuations

Three Types of Linear Problems

$$y = Ax \begin{cases} \dim(y) = \dim(x) \\ \dim(y) > \dim(x) \\ \dim(y) < \dim(x) \end{cases}$$

One-to-one transformation
Convolution, DFT

exact solutions

Over-determined
N(measurement) >> N(unknown)

optimized solutions

Under-determined
N(measurement) << N(unknowns)

- Physicists desire an overdetermined system: reduce systematics
- Reality (technology / economics constraints): experiments with underdetermined systems are still very common

Overdetermined linear system with uncertainties

$$y = Ax \xrightarrow[\text{(maximum likelihood)}]{\text{weighted least square}} \chi^2 = \left(\frac{y - Ax}{\sigma}\right)^2$$

$\dim(y) > \dim(x)$

More general: $$\chi^2 = (y - Ax)^T \cdot V^{-1} \cdot (y - Ax)$$

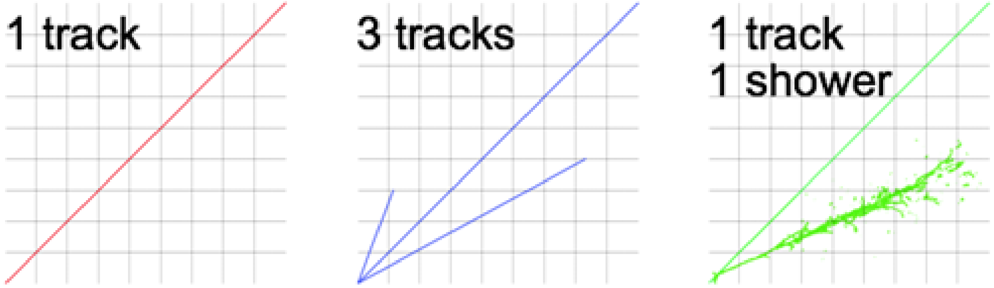
minimize

Solution: $$x = (A^T V^{-1} A)^{-1} A^T V^{-1} \cdot y$$

- ❑ Usage: parameter optimization (aka. fitting). *See Xin's next talk for examples*
- ❑ Main challenge: fast computation when dimension is large

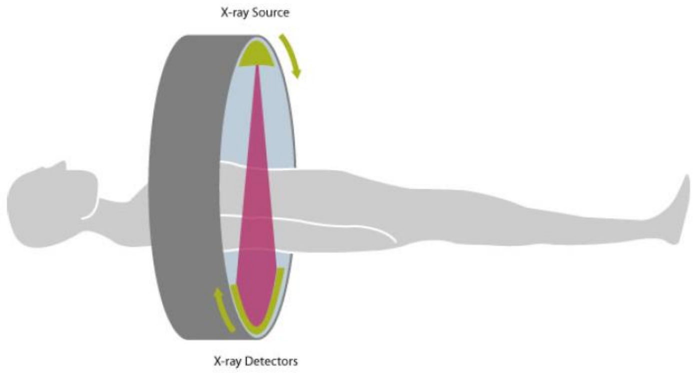
Underdetermined system: TPC 3D image reconstruction

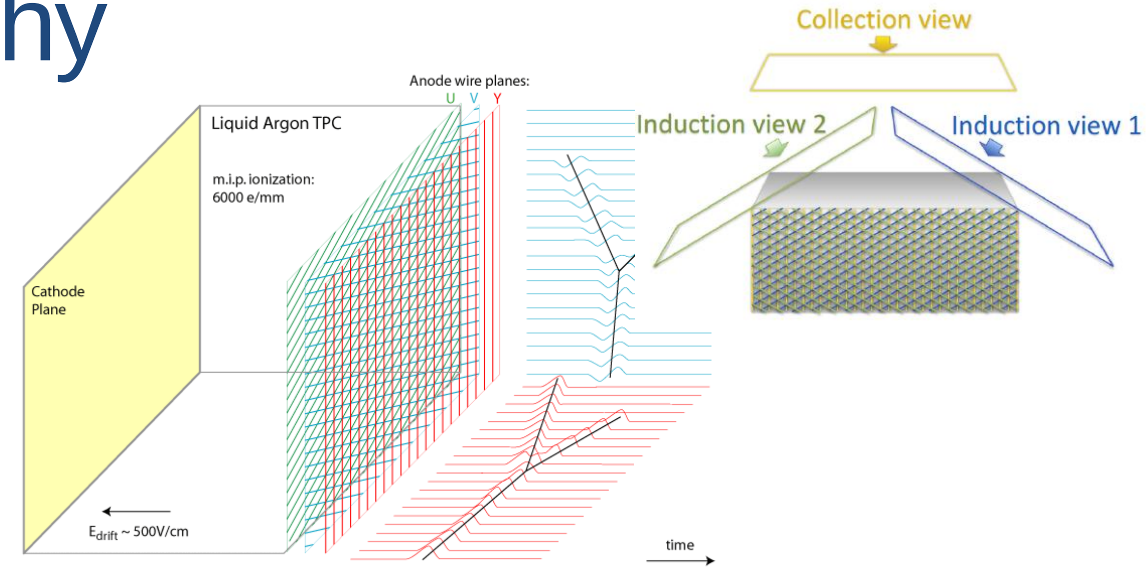
- Pad-readout TPC: <u>one-to-one / over-determined system</u>
  - A 3-D image (2D pixels + time) can be directly obtained without ambiguity.
- Wire-readout TPC: <u>under-determined system</u>
  - Large LArTPCs typically use wire readout due to the cost and power consumption constraints.
  - Limited number of wire planes (typically 3): information loss (n^2 - > 3n)
  - Difficulty is topology dependent
    - <u>worst case:</u> isochronous event where electrons arrive at the wire planes at the same time.



same wires are hit, which topology is the true event?

LArTPC vs X-ray Tomography





❑ CAT Scan

- Detector (x-ray generator/receiver) moves across the object (body)
- Axial projections (~180) by detector rotation
- Cross section can be reconstructed at each position along detector movement

❑ LArTPC

- Objects (ionizing electrons) move across detectors (wire planes)
- Axial projections (~3) by wire orientation
- Cross section can be reconstructed at each time slice along electron drift

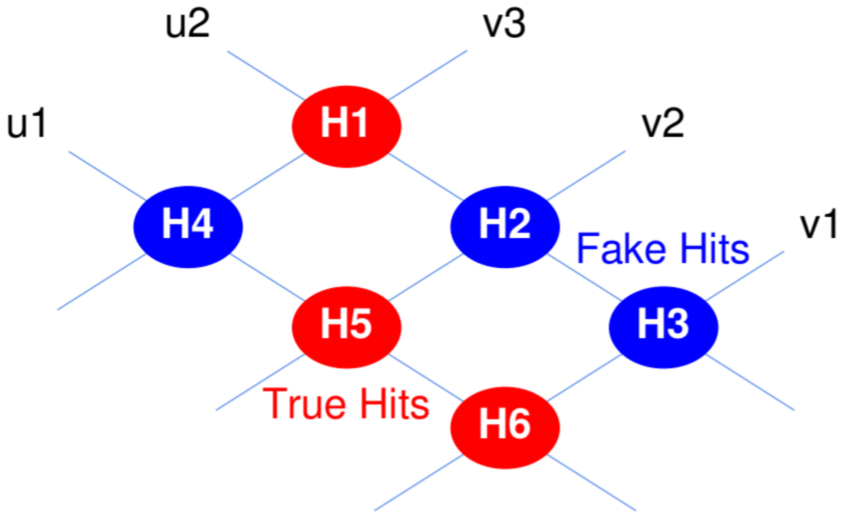
Construct Linear Equations

Wire-Cell

$$y = Ax$$

$$\begin{pmatrix} u1 \\ u2 \\ v1 \\ v2 \\ v3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} H1 \\ H2 \\ H3 \\ H4 \\ H5 \\ H6 \end{pmatrix}$$

Use two planes as an illustration



At one time slice

y: measured charge signal on each wire
x: the (unknown) true charge deposition in each possible cell
A: bi-adjacency matrix connecting wires and cells (determined solely by wire geometry)

Desired Solution $x = \begin{pmatrix} H1 \\ 0 \\ 0 \\ 0 \\ H5 \\ H6 \end{pmatrix}$

Solve underdetermined linear problem: regularization

- Previous example has 6 unknowns, 5 equations: under-determined system

$$x = \boxed{(A^T V^{-1} A)}^{-1} A^T V^{-1} \cdot y$$

↓

non-invertible, 2 zero-eigenvalues out of 6.

- Adding constraints: find the sparsest solution (applies to most physics events): L0-regularization

minimize $||x||_0$, subject to: $y = Ax$

(L0-norm: number non-zero elements)

NP-hard!

Procedure

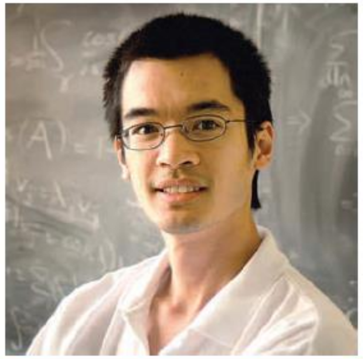
- Remove unknowns until equations can be solved, then find the best solution with the minimum χ^2
- a combinatorial problem
  - 2 out of 6: 15 combinations
  - 10 out of 40: 0.8 billion combinations

Compressed Sensing (L1-regularization)

- Breakthrough: mathematical proof that L0 problem can be well approximated by the L1 problem (Compressed Sensing, *Candes, Romberg, and Tao, 2005*.)

Emmanuel Candes. *(Photo courtesy of Emmanuel Candes.)*

Justin Romberg. *(Photo courtesy of Justin Romberg.)*



Terence Tao. *(Photo courtesy of Reed Hutchinson/UCLA.)*

https://arxiv.org/abs/math/0503066

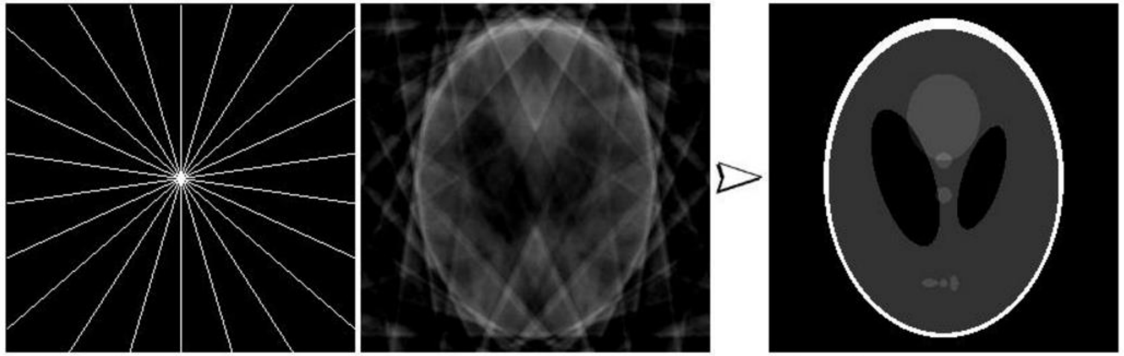
minimize $||x||_1$, subject to: $y = Ax$

(L1-norm: sum of absolute values of the elements)

Or, equivalently, minimize

$$\chi^2 = (y - Ax)^T \cdot V^{-1} \cdot (y - Ax) + \lambda||x||_1$$

Sparse projections: 11 radial lines



available portion of the spectrum (11 radial lines) | Back-projection estimate | Estimate after convergence (exact reconstruction)

Tomography: reconstruct image with far less projections

L1 Regularization

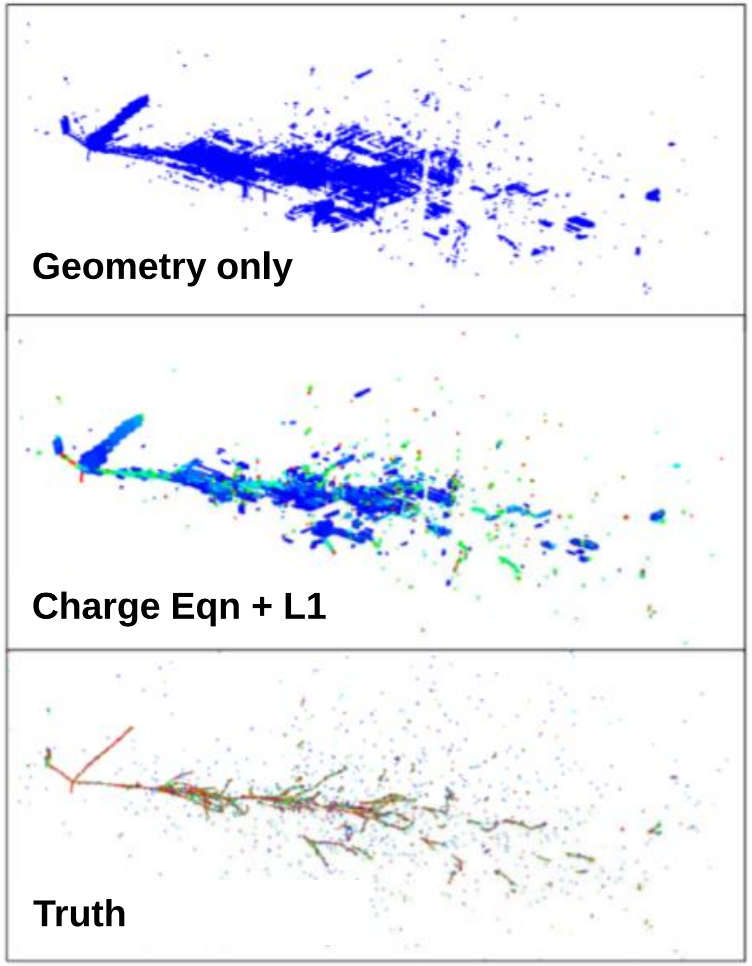
Brief History of L1 regularization

- Wavelet soft thresholding *(Donoho and Johnstone 1994)*
- Lasso regression *(Tibshirani 1995)*
- Same idea in Basis Pursuit *(Chen, Donoho and Saunders)*
- Extended to many linear-model settings, e.g. Survival models *(Tibshirani, 1997)*
- Gives to a new field Compressed Sensing *(Candes, Romberg, Tao, 2005)*: near exact recovery of sparse signals in very high dimensions

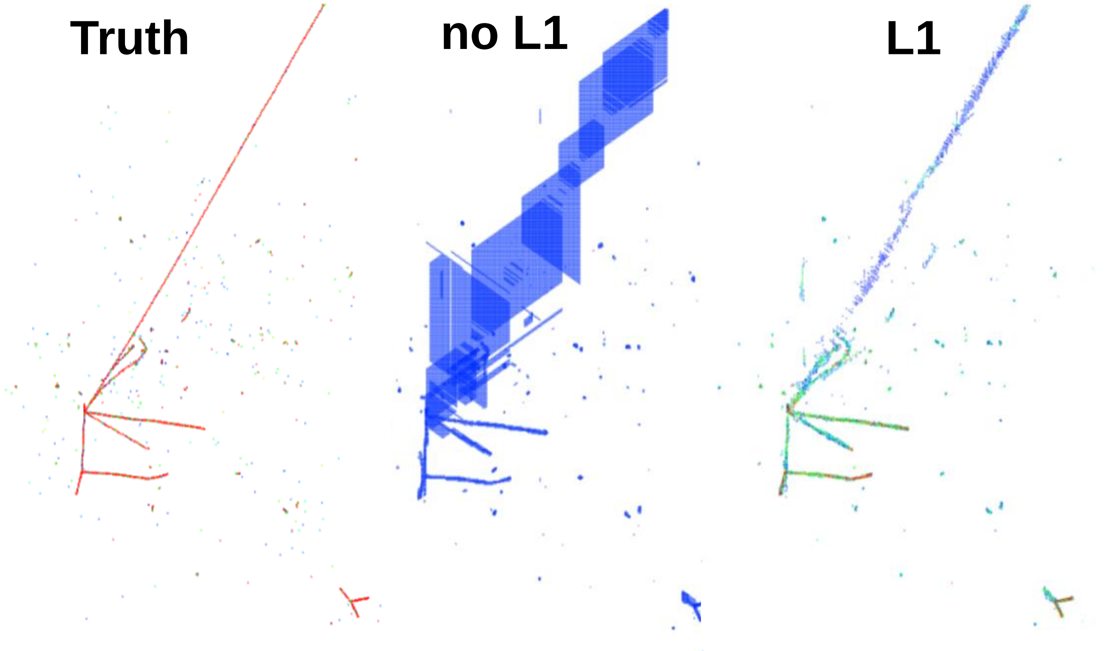
Features of L1 regularization

- Convex problem
  - local minimum == global minimum
- Fast minimization
  - efficient algorithms to quickly find the global minimum e.g. coordinate descent
- Shrink the irrelevant variables to exactly zero leading to the desired sparse solution, due to the specific shape of the χ^2 function
  - compared with L2 regularization (RIDGE), which also shrinks the variables, but only to small non-zero values

Performance in Wire-Cell



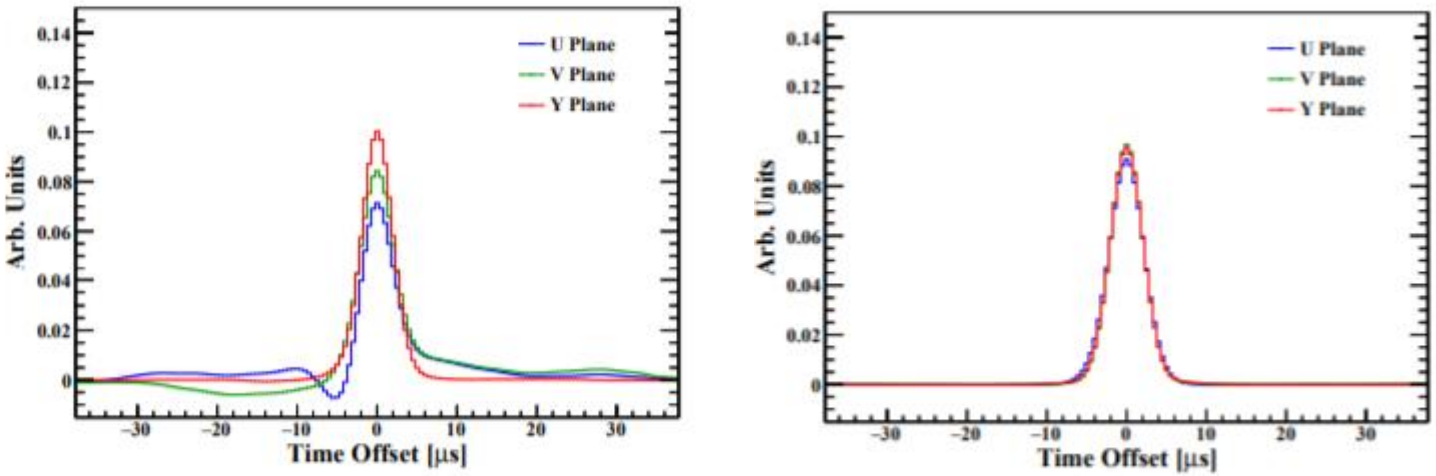
JINST 13, P05032 (2018)



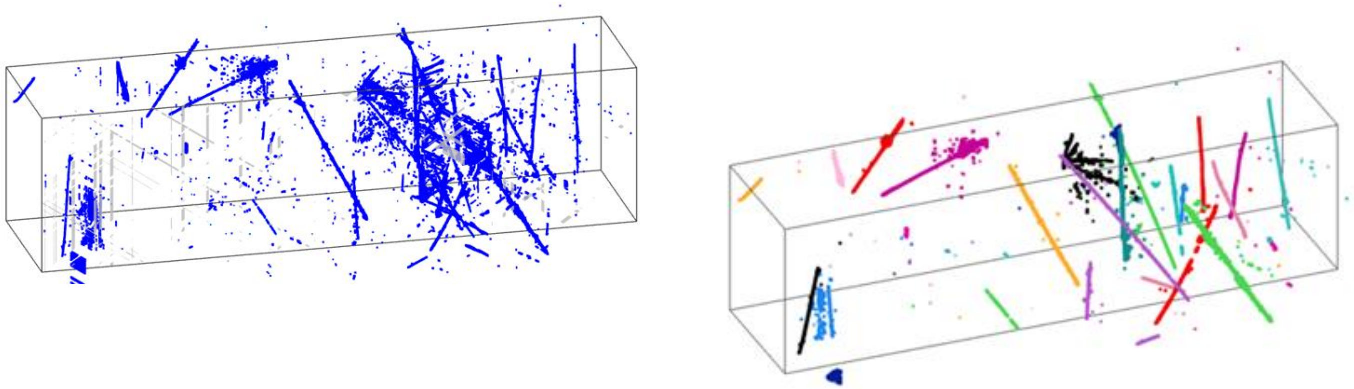
- Typically ~tens of seconds to reconstruct the whole 3D image (originally a few hours)

Practical Lessons from MicroBooNE

- Expectation (from LArTPC principle): same charge measured three times by each wire plane
  - Reality: takes two years to improve signal processing to achieve charge matching among different planes
- Expectation: 3 wire measurements everywhere
  - Reality: non-functional channels cause in some regions only 2 or even 1 wire-plane measurement available → lots of noises and ghost tracks



(e) 1D deconvolution, $30° < \theta_{xz} < 50°$.
(f) 2D deconvolution, $30° < \theta_{xz} < 50°$.



The devil is in the detail

More on Compressed Sensing

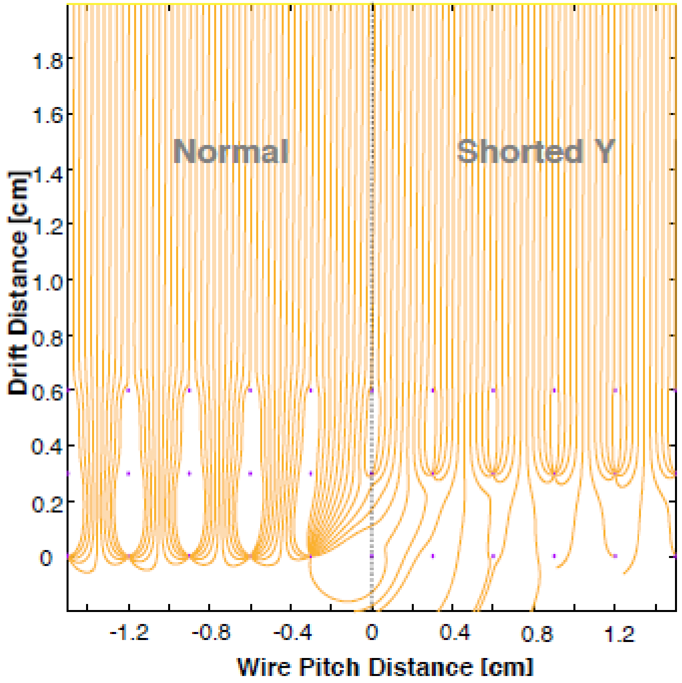
Advanced L1 regularization

- Many ways to extend the simplest L1 regularization, by adding additional known constraints in the $\chi^2 = (y - Ax)^2 + \lambda ||x||_1$
  - Grouped Lasso: add penalty based on group
  - Fused Lasso: add penalty based on connectivity
  - Other customized penalty terms in the χ2 definition
  - Or even generalize to non-linear case:
    $$\chi^2 = -2\log L(y, x) + \lambda ||x||_1$$
- The added constraints improves the results by avoiding random fluctuations.
  - Typically need new minimization algorithms to run fast

General application in HEP

- Compressed sensing is a general mathematical technique to solve for sparse signal in underdetermined linear system
  - Prove applicable in LArTPC 3D reconstruction, easily extendable to other wire-based tracking detectors.
- Provide a solution to some of the previously intractable problems
- Cost reduction: use less measurement to obtain high performance

One Example: deconvolution with multiple impulse responses

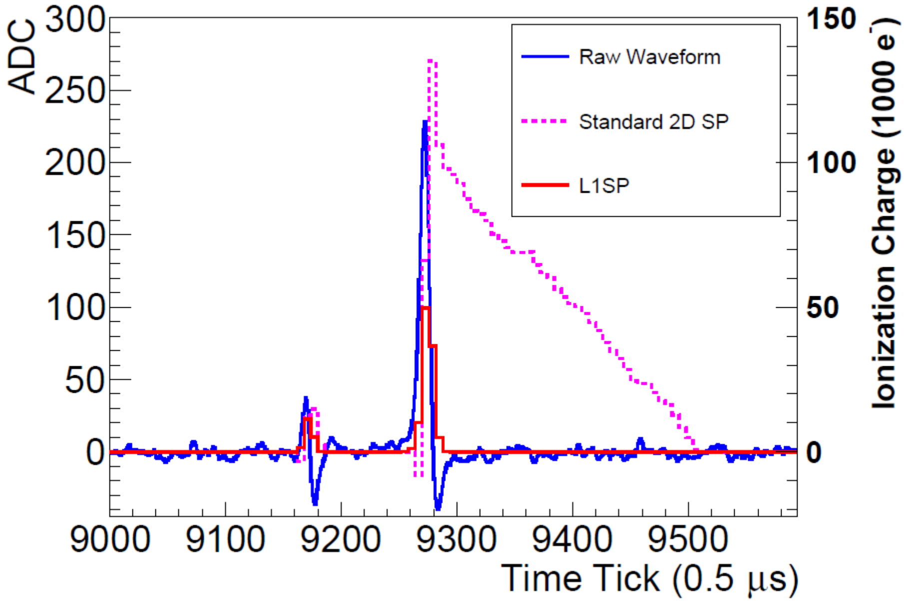


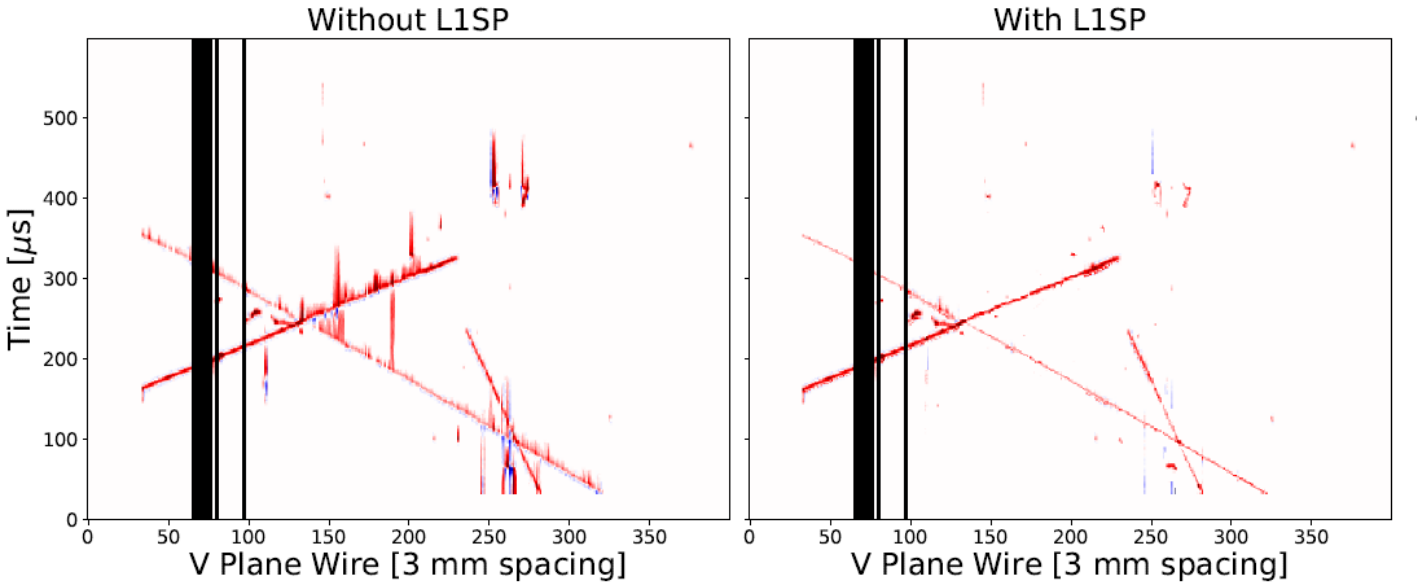
Two possible (position dependent) impulse responses on a single wire due to shorted regions

$$y = Ax \qquad y = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\dim(y) = N; \dim(x) = 2N$$

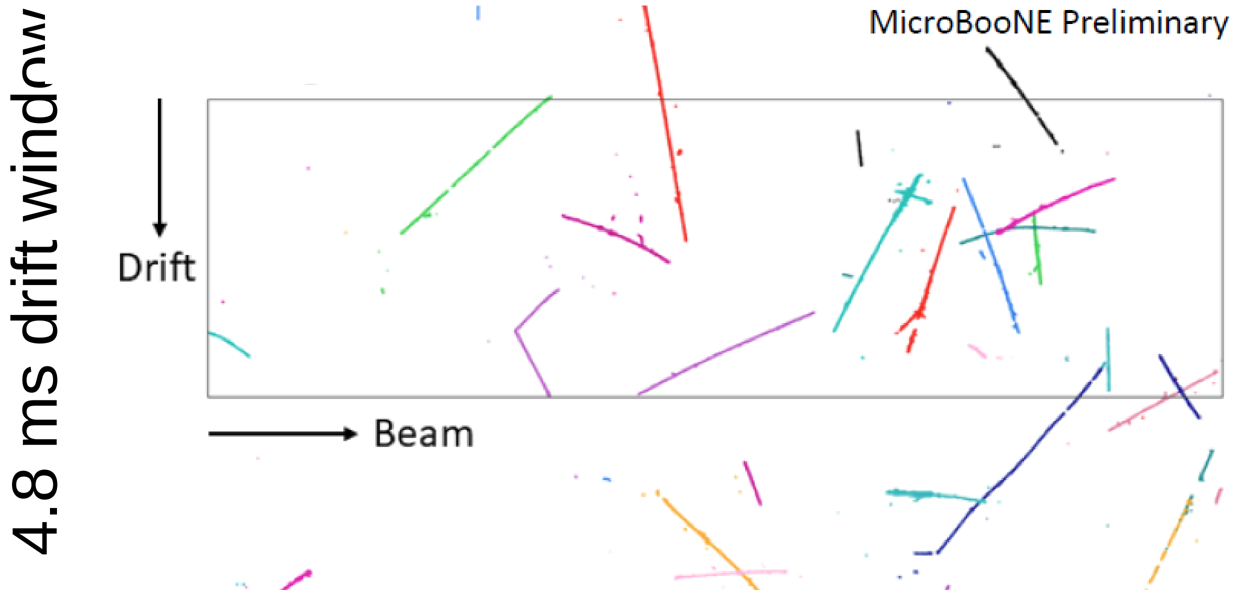
$\longrightarrow$ L1 SP!



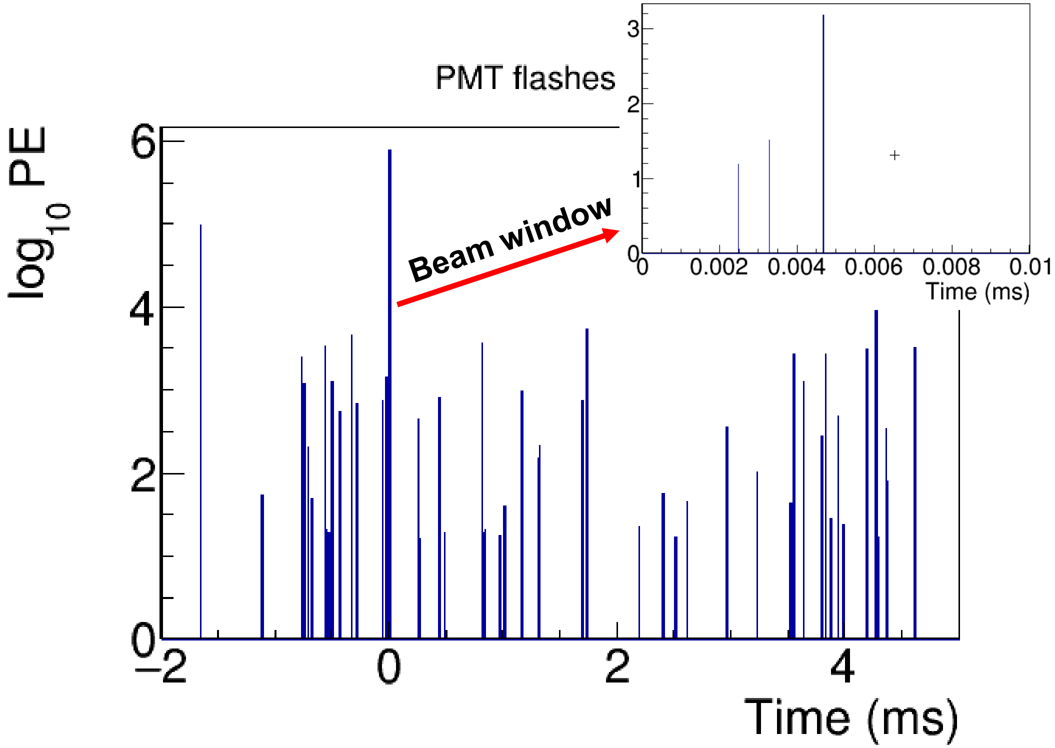


JINST 13 P07007 (2018)

Another example: Light Matching in LArTPC



20-30 TPC activities

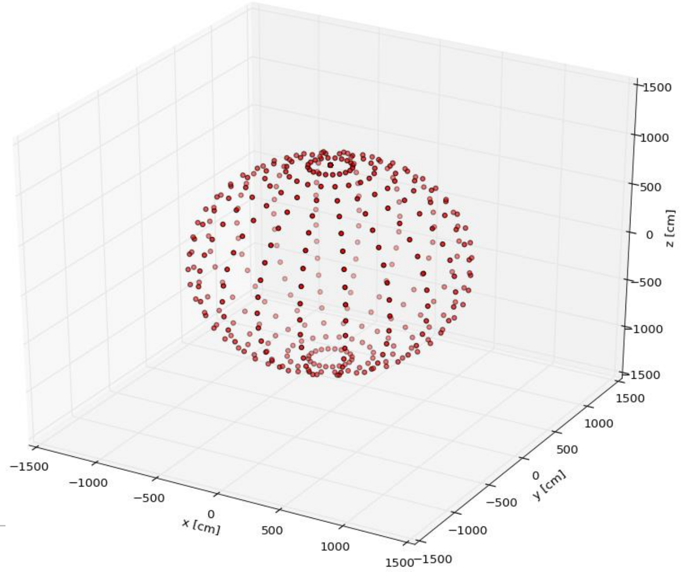


40-50 PMT activities

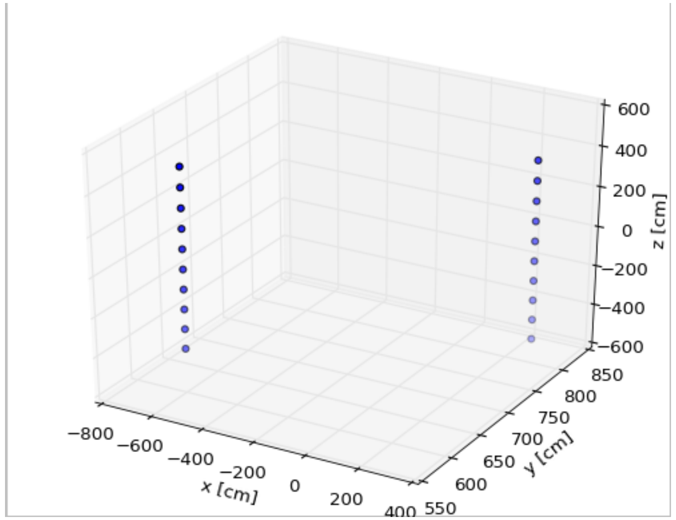
- associate the correct T_0 (light flash) to the corresponding TPC cluster
  - combinatorial problem -> compressed sensing! (Q: what's the y=Ax equation here?)

Third Example: Reconstruct multiple tracks in LS

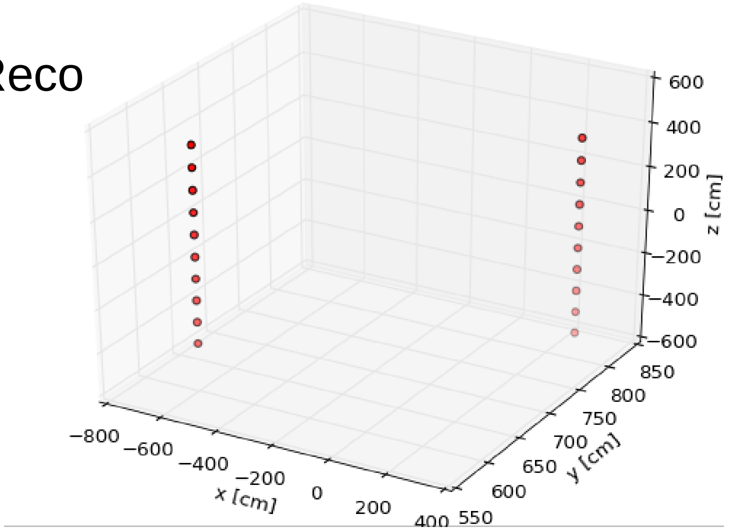
- Test of principle: JUNO-like LS detector with radius 10 m.
- 400 PMTs uniformly distributed
- Reconstruction in 1m x 1m x 1m voxels (4139 in total)
- Response matrix *A* included the photon $1/r^2$ spread and the exponential attenuation from absorption.
- Put in two "vertical muon tracks", reconstructed successfully



Truth



Reco



(Q: what's the y=Ax equation here?)

Summary

- Linear system is an important concept that finds many applications in LArTPC and other physics topics
- The three types linear system each has its own applications
  - One-to-one transform: DSP
  - Overdetermined system: parameter optimization
  - Underdetermined system: LArTPC imaging
    - L1-regularization (compressed sensing)
- Apply the linear system concept to your own problems

Backups

Coordinate Descent

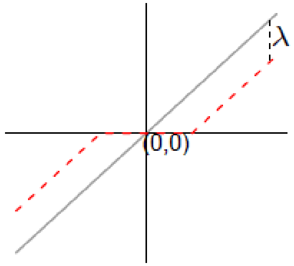
$$\min_\beta \frac{1}{2N}\sum_{i=1}^N (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Suppose the p predictors and response are standardized to have mean zero and variance 1. Initialize all the $\beta_j = 0$.

Cycle over $j = 1, 2, \ldots, p, 1, 2, \ldots$ till convergence:

- Compute the partial residuals $r_{ij} = y_i - \sum_{k \neq j} x_{ik}\beta_k$.
- Compute the simple least squares coefficient of these residuals on jth predictor: $\beta_j^* = \frac{1}{N}\sum_{i=1}^N x_{ij} r_{ij}$
- Update β_j by *soft-thresholding*:

$$\begin{aligned} \beta_j &\leftarrow S(\beta_j^*, \lambda) \\ &= \text{sign}(\beta_j^*)(|\beta_j^*| - \lambda)_+ \end{aligned}$$



- Advantage of coordinate descent
  - Fast convergence due to the function shape and soft-thresholding
  - Easy to enforce "non-negative" constraint (only descent toward positive side)
  - Easy to parallelize (can solve huge matrix)
- Existing software packages:
  - Python: scikit-learn (LASSO)
  - R: lars, glmnet, l1logreg
  - Matlab: admm,
  - or code the algorithms yourself:
    - https://github.com/BNLIF/wire-cell-ress (simple implementation)

Tuning of regularization parameters

$$minimize\ \chi^2 = (y - Ax)^2 + \lambda||x||_1\ .$$

- The “regularization” parameter λ is a free parameter that needs to be tuned to the data to produce good results
  - Too large λ: too much weight in L1 will lead to over-sparse solutions (too many zero x’s)
  - Too small λ: too little weight in L1 will not lead to good sparse solutions
  - On the other hand, not too sensitive to λ (a set of λ’s could yield the same results)
  - Optimal λ is typically obtained from scanning through a path.

Flash matching using L1

- Flash matching is to create associations between flashes (light) and TPC objects (tracks, showers)
- Example: 2 flashes (F1, F2), 3 tracks (T1, T2, T3)
  - 6 possible combinations (x1 … x6)

"->" denotes "Track contributes to flash"

F: PE's on each PMT
a_{ij}: predicted PE's on each PMT give one hypothesis
Chi-square is constructed by summing over all PMTs

$$\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{24} & a_{25} & a_{26} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} \begin{array}{l} \text{T1 -> F1} \\ \text{T2 -> F1} \\ \text{T3 -> F1} \\ \text{T1 -> F2} \\ \text{T2 -> F2} \\ \text{T3 -> F2} \end{array}$$

- This way naturally solves:
  - Multiple tracks contribute to one flash, like this example
  - Non-matching tracks (will become 0, since other tracks can explain the flashes)
  - On track contribute to multiple flashes due to late light (x will become less than 1)

3D Imaging in LS, WC, and WbLS detectors

- Liquid Scintillator detectors usually only considered as calorimeters. Event reconstruction typically only restricted to simple point-like low energy event or single-track (muon) event.
- With Compressed Sensing, one can do 3D imaging in LS: $y = Ax$
  - y would be the charge/time from each PMT
  - x would be the true charge (#photons) in discrete voxels
  - A would be the response matrix to populate each voxel to the PMT, which one can pre-calculate based on geometry and physics
- Similar for WC and WbLS. Since Cerenkov radiation is directional, one needs to add two more dimensions in angular space.
  - Computation requirements grows quickly with dimensions of the matrix.