A Convex Program for Binary Tomography

Ajinkya Kadu
Utrecht University, Netherlands

joint work with **Tristan van Leeuwen**

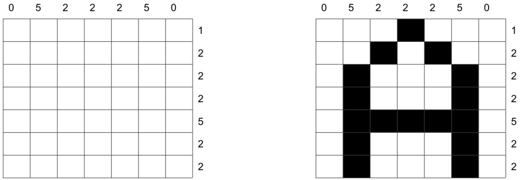
Applied Math Seminar, Utrecht University

November 21, 2019

Motivation

Consider a simple form of **nonogram**

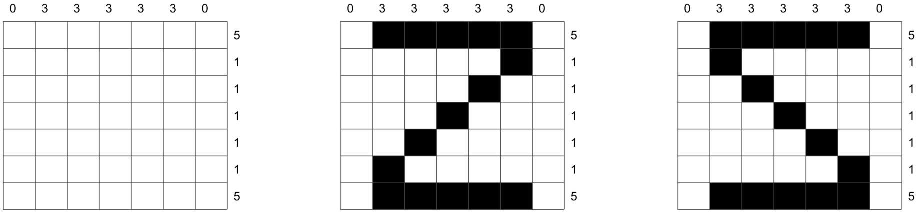
Example 1



Motivation

Consider a simple form of **nonogram**

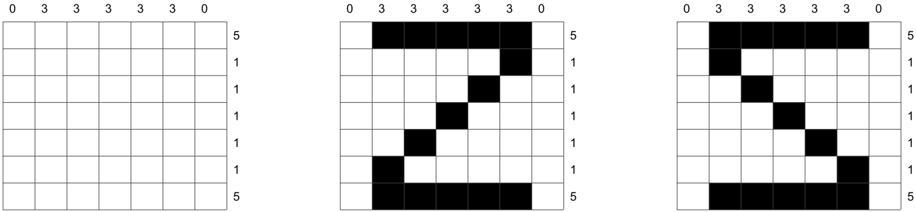
Example 2



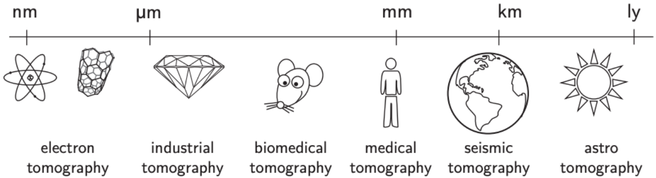
Motivation

Consider a simple form of **nonogram**

Example 2



Goal: find a *binary* solution from the projection data (*e.g.*, row and column sums).



Outline

Binary tomography

Convex formulation

Small-scale experiments

X-Ray tomography experiments

Conclusions

Rubicon Proposal

The Problem

- recover *binary* images from a few of their projections

$$\text{find} \quad \mathbf{x} \in \mathcal{U}^n \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}$$

  - discrete set, $\mathcal{U} = \{u_0, u_1\}$
  - tomographic projection matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \ll n$
  - tomographic data, $\mathbf{y} \in \mathbb{R}^m$, (*e.g.*, row and column sums)

- constrained least-squares formulation

$$\mathbf{x}^\star = \underset{\mathbf{x} \in \mathcal{U}^n}{\operatorname{argmin}} \quad \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$$

Challenges

$$\mathbf{x}^\star = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{U}^n} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$$

- constraints are non-convex ($\mathcal{U}$ is a discrete set)
- Large null-space of $\mathbf{A}$ ($m \ll n$)
- solving constrained-LS exactly (*i.e.*, getting $\mathbf{x}^\star$) is not trivial
- Noise in $\mathbf{y}$ affects $\mathbf{x}^\star$
- considered as NP-hard (for $m \geq 3, n > 2$)

Current state-of-the-art

- **Algebraic methods**
  (iterative algorithms similar to *Kaczmarz* method)

- **Stochastic sampling methods**
  (based on sampling the *pdf* on the space of discrete images)

- **Relaxation methods**
  - Convex relaxations (*e.g.*, TV regularization)
  - Non-convex relaxations (*e.g.*, level-set method, non-convex regularizers)

- **Heuristic algorithms**
  (practical methods, *e.g.*, DART)

Outline

Invertible $\mathbf{A}$

$$\min_{\mathbf{x}\,,\,\boldsymbol{\phi}} \quad \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 \quad \text{subject to} \quad \mathbf{x} = \mathbf{sign}\,\boldsymbol{\phi}$$

- **Lagrangian** $$\mathcal{L}(\mathbf{x},\boldsymbol{\phi},\boldsymbol{\nu}) = \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \boldsymbol{\nu}^T(\mathbf{x}-\mathbf{sign}\,\boldsymbol{\phi})$$

- **Dual function** $$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{\mathbf{x}}\left\{\tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \boldsymbol{\nu}^T\mathbf{x}\right\} + \inf_{\boldsymbol{\phi}}\left\{-\boldsymbol{\nu}^T\,\mathbf{sign}\,\boldsymbol{\phi}\right\} \\ &= \left(-\tfrac{1}{2}\|\mathbf{A}^T\mathbf{y}-\boldsymbol{\nu}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \tfrac{1}{2}\mathbf{y}^T\mathbf{y}\right) - \|\boldsymbol{\nu}\|_1 \end{aligned}$$

- **Dual problem** $$\min_{\boldsymbol{\nu}} \quad \tfrac{1}{2}\|\boldsymbol{\nu}-\mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \|\boldsymbol{\nu}\|_1$$

Dual Problem

$$\textbf{(Primal)} \qquad \min_{\mathbf{x},\boldsymbol{\phi}} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 \quad \text{subject to} \quad \mathbf{x} = \mathbf{sign}\,\boldsymbol{\phi}$$

$$\textbf{(Dual)} \qquad \min_{\boldsymbol{\nu}} \quad \frac{1}{2}\|\boldsymbol{\nu}-\mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \|\boldsymbol{\nu}\|_1$$

- a LASSO (ℓ_1-regularized least-squares) problem
- Convex program. But can have multiple solutions.
- Many existing techniques available to solve the problem numerically.

a *primal* **solution** is retrieved using $\quad \mathbf{x}^\star = \mathbf{sign}\,(\boldsymbol{\phi}^\star) = \mathbf{sign}\,(\boldsymbol{\nu}^\star)$

General $\mathbf{A}$

Dual Problem for rank-deficient A

The dual objective for general $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given by

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} - \|\boldsymbol{\nu}\|_1 + \frac{1}{2}\mathbf{y}^T\mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \text{otherwise} \end{cases}$$

This leads to an optimization problem: $\min_{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}} \quad \frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + \|\boldsymbol{\nu}\|_1.$

Limited Tomography ($m \ll n$)

If $\mathbf{rank}(\mathbf{A}) = m$ then the dual problem takes the following form:

$$\min_{\boldsymbol{\mu} \in \mathbb{R}^m} \quad \frac{1}{2}\|\boldsymbol{\mu} - \mathbf{y}\|^2 + \|\mathbf{A}^T\boldsymbol{\mu}\|_1.$$

The primal solution is retrieved using $\mathbf{x}^\star = \mathbf{sign}\left(\mathbf{A}^T\boldsymbol{\mu}^\star\right)$

Natural Extensions

Binary Tomography with grey levels

The dual problem for a binary tomography problem with grey levels $u_0 < u_1$ is given by:

$$\min_{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}} \quad \frac{1}{2} \|\boldsymbol{\nu} - \mathbf{A}^T \mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + p(\boldsymbol{\nu}),$$

where $p(\boldsymbol{\nu}) = \sum_i |u_0| \max(-\nu_i, 0) + |u_1| \max(\nu_i, 0)$ is an asymmetric one-norm.

Poisson Noise

In case of Poisson noise, BT problem takes constrained weighted least-squares form:

$$\min_{\mathbf{x}\,,\,\boldsymbol{\phi}} \quad \frac{1}{2} \|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|^2_{\boldsymbol{\Lambda}} \quad \text{subject to} \quad \mathbf{x} = \mathbf{sign}(\boldsymbol{\phi}),$$

The (dual) optimization problem is $\displaystyle \min_{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}} \quad \frac{1}{2} \|\boldsymbol{\nu} - \mathbf{A}^T \boldsymbol{\Lambda} \mathbf{y}\|^2_{(\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A})^\dagger} + \|\boldsymbol{\nu}\|_1.$

Solving the dual problem

- **exact** solution (generally a bit slow)

  Convert LASSO to a SDP, solve it using interior point method.
  Implemented in CVX toolbox.

- **approximate** solution (iterative methods)

  – Approximate the ℓ_1 norm
  (solve resulting problem using gradient-based method)

  – Proximal gradient method (*e.g.*, ISTA, FISTA)

  – Splitting methods (*e.g.*, ADMM, Primal-Dual)

ADMM

$$\min_{\boldsymbol{\nu}} \quad \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|_{\mathbf{K}}^2 + \|\boldsymbol{\nu}\|_1$$

Augmented Lagrangian $\quad \mathcal{L}_\rho(\boldsymbol{\nu}, \boldsymbol{\lambda}, \mathbf{p}) = \quad \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|_{\mathbf{K}}^2 + \|\boldsymbol{\lambda}\|_1 + \tfrac{\rho}{2}\|\boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{p}\|^2$

Iterations $(t = 0, \ldots, T)$:

$$\boldsymbol{\nu}^{t+1} \triangleq \operatorname{argmin}\left\{\tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|_{\mathbf{K}}^2 + \frac{\rho}{2}\|\boldsymbol{\nu} - \boldsymbol{\lambda}^t + \mathbf{p}^t\|^2\right\}$$

$$\boldsymbol{\lambda}^{t+1} \triangleq \operatorname{argmin}\left\{\|\boldsymbol{\lambda}\|_1 + \frac{\rho}{2}\|\boldsymbol{\lambda} - \boldsymbol{\nu}^{t+1} - \mathbf{p}^t\|^2\right\} \qquad = S_{1/\rho}\left(\boldsymbol{\nu}^{t+1} - \mathbf{p}^t\right)$$

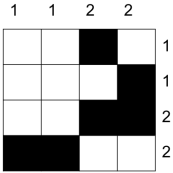
$$\mathbf{p}^{t+1} \triangleq \mathbf{p}^t + \boldsymbol{\nu}^{t+1} - \boldsymbol{\lambda}^{t+1}$$

Outline

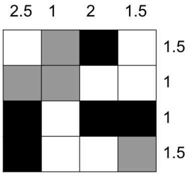
Unique solution

- **Example I**

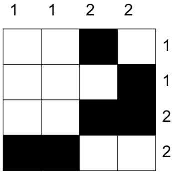
True solution



(Pseudo-Inverse)$_\tau$

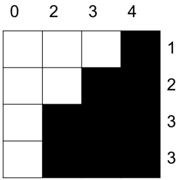


Dual

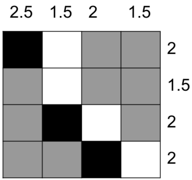


- **Example II**

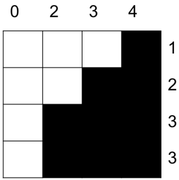
True solution



(Pseudo-Inverse)$_\tau$

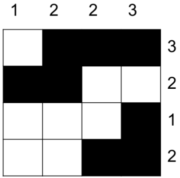


Dual

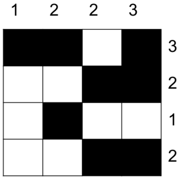


Multiple solutions

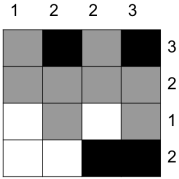
Solution 1



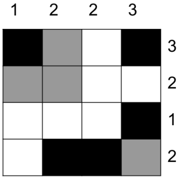
Solution 2



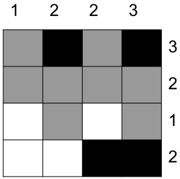
Intersection



(Pseudo-Inverse)$_\tau$



Dual



Experiments

	n	total	unique	multiple
$m = 2$	2×2	16	14/14	2/2
	3×3	512	230/230	282/282
	4×4	65536	6902/6902	58541/58634*
$m = 3$	2×2	16	16/16	0/0
	3×3	512	496/496	16/16
	4×4	65536	54272/54272	10813/11264*
$m = 4$	2×2	16	16/16	0/0
	3×3	512	512/512	0/0
	4×4	65536	65024/65024	512/512

n denotes the length of image. $m = 2$ - row and column sums, $m = 3$ - row, column and diagonal sums, $m = 4$ - row, column, diagonal and off-diagonal sums. Results shows the success times of the dual approach vs total possibilities. * denotes issues with CVX.

Main Result

Conjecture: Recovery of Solution

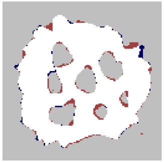
- If the problem has a **unique** solution then the dual approach **retrieves** it.
- If the problem has **multiple** solutions then the dual approach retrieves the **intersection** of all solutions.

There is a *subclass* of the described binary tomography problem that is not **NP-hard**.

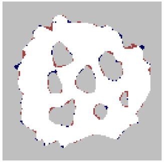
Outline

Sparse Projection Test

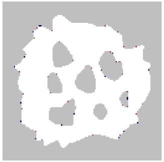
(10 projection angles from 0° to 180°)



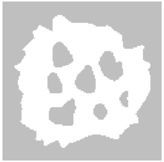
$(\mathsf{LSQR})_\tau$



$(\mathsf{TV})_\tau$



DART

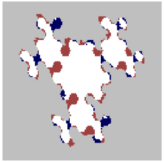


Dual

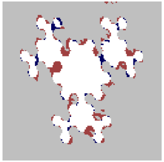
Images are 128×128 pixels. We use parallel beam geometry with 128 detectors. LSQR and TV reconstruction are thresholded using Otsu's method. blue pixel denote underestimated pixel and red denote overestimated pixel. Lesser the blue and red pixels, better the reconstruction.

Limited Angle Test

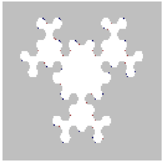
(10 projection angles from 0° to 90°)



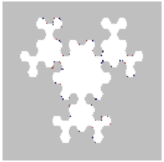
$(\text{LSQR})_\tau$



$(\text{TV})_\tau$



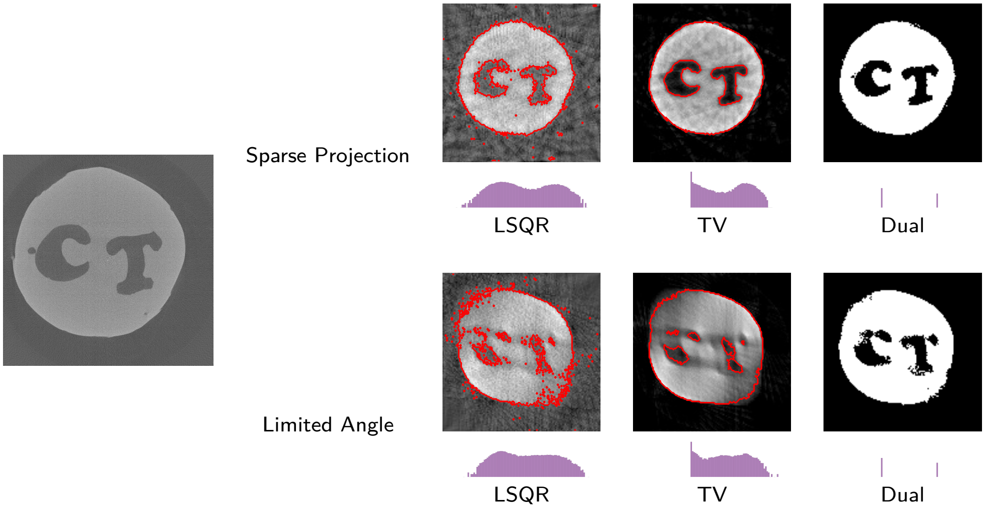
DART



Dual

Images are 128×128 pixels. We use parallel beam geometry with 128 detectors. LSQR and TV reconstruction are thresholded using Otsu's method. blue pixel denote underestimated pixel and red denote overestimated pixel. Lesser the blue and red pixels, better the reconstruction.

X-Ray Cheese Dataset [FIPS]



FBP reconstruction (left) is of size 2048×2048 pixels. In Sparse projection, 10 projection angles from $0°$ to $180°$ are used. Images are of size 128×128 pixels. In limited angle, the angle is restricted from $0°$ to $90°$.

Outline

To summarize

- We propose a **convex program** to retrieve *binary* solution from its tomographic projections.

- From small-scale experiments, we conjecture that
  1. If a **unique** solution then the dual approach **retrieves** it.
  2. If **multiple** solutions then the dual approach retrieves the **intersection** of all solutions.

- The proposed method compares *favorably* to existing approaches (*e.g.*, TV and DART).

Open Questions:

- Does the strong *duality* hold?
- Proving the conjecture
- Extension to discrete tomography (more than 2 grey levels)

Thanks!

Read:

- A. Kadu, T. van Leeuwen
  A convex formulation for Binary Tomography.
  IEEE Transactions on Computational Imaging 2019 [arXiv:1807.09196]

Email:

- `a.a.kadu@uu.nl`

Web:

- `https://ajinkyakadu125.github.io`
- `https://github.com/ajinkyakadu125/` (code will be available soon)

Special Thanks to:
Nick Luiken, K. Joost Batenburg

Outline

Binary tomography

Convex formulation

Small-scale experiments

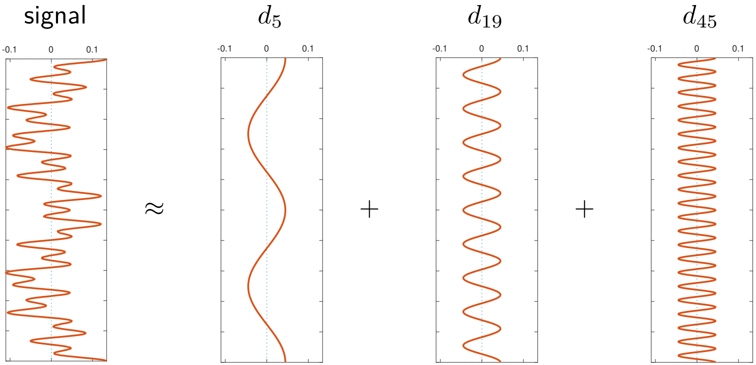
X-Ray tomography experiments

Conclusions

Rubicon Proposal

Signal Representation - Binary Coding

Aim: reduce the signal $x(t)$ to the sum of atoms $d_i(t)$

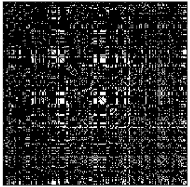


$$x(t) = \sum_{i=1}^{m} w_i d_i(t) \quad \text{with } w_i \in \{0, 1\}$$

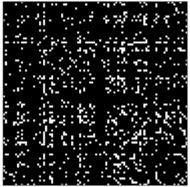
Binary Matrix

a matrix with entries from the Boolean domain $\mathcal{B} = \{0, 1\}$

Friendship



Jazz collaboration



Senate Joint Press

Examples:

- Gene Expressions
- Document Clustering
- web click-stream data
- Market basket data

http://moreno.ss.uci.edu/data.html

Matrix Factorization

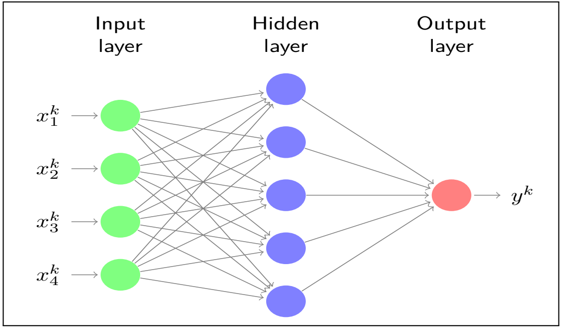
decompose binary matrix $\mathbf{X}$ into product of two binary matrices

$$\text{find} \quad \mathbf{U} \in \mathcal{B}^{m\times k}, \mathbf{V} \in \mathcal{B}^{n\times k} \quad \text{subject to} \quad \mathbf{X} = \mathbf{U} \circ \mathbf{V}^T$$

$$\begin{bmatrix} x_{11} & \dots & \dots & x_{1n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{m1} & \dots & \dots & x_{mn} \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_k \\ | & & | \end{bmatrix} \circ \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_k \\ | & & | \end{bmatrix}^T$$

- Low-rank factorization: $k \ll \min(m, n)$
- Operation $\circ$: Linear or Boolean or Modulo-2

Neural Networks



$$\text{define} \quad g_i(\mathbf{x}) = h(\mathbf{W}_i \mathbf{x} + \mathbf{b}_i) \qquad i = 1, \ldots, k$$

$$\text{find} \quad (\mathbf{W}_i, \mathbf{b}_i)_1^k \quad \text{such that} \quad y^{(j)} \approx g_k \circ g_{k-1} \circ \cdots \circ g_1 \left(\mathbf{x}^{(j)}\right) \qquad j = 1, \ldots, N$$

Binary Networks

networks with weights restricted to $\{0, 1\}$

Advantages

- **Storage**: need to store in a bit (0 or 1) instead of single precision - 32 bits
- **Computation**: bit-wise computations are faster (Logic gates)
- **Robust**: may avoid over-fitting
- **Inference**: quick inference due to small number of operations

Table 1: Comparison of different methods on ImageNet dataset. The compression rate (denoted as Comp. rate in the table) considers the overall parameters of the network. Both top-5 and top-1 accuracies are presented in the last column.

Methods	Model size	Comp. rate	Accuracy(%)
AlexNet	232 MB[1]	$1\times$	$80.2/56.6$
BNN	22.6 MB	$10.3\times$	$50.4/27.9$
XNOR-net	22.6 MB	$10.3\times$	$69.2/44.2$
DoReFa-net	22.6 MB	$10.3\times$	$-\ /49.8$[2]
Our Method	1.23 MB	$23.6\times$	$75.6/51.4$

Proposal Summary

- **Signal Representation**
  - find $\mathbf{w} \in \{0,1\}^m$ such that $x(t) \approx \sum_{i=1}^{m} w_i d_i(t)$.
  - **Blind representation**: given a set of point $\{\mathbf{x}_i\}$, find a dictionary $\mathbf{D}$ and corresponding binary weights matrix $\mathbf{W}$.
- **Matrix Decomposition**
  - given a binary matrix $\mathbf{X}$, decompose into binary matrices $\mathbf{U}$ and $\mathbf{V}$ such that $\mathbf{X} \approx \mathbf{U}\mathbf{V}^T$.
  - **Matrix completion**: fill up a binary matrix $\mathbf{X}$ with assumption of low-rank structure.
- **Deep Networks**
  - A network can be compressed using quantized weights.
  - find a suitable **training** procedure for quantized weights without sacrificing the accuracy.
- **Some Applications:**
  - Wireless Communications
  - Unsupervised learning
  - Light-weight Nets