CMSC 510 – L17
Regularization Methods for Machine Learning

Instructor:
Dr. Tom Arodz

Detour: Compressed Sensing

- We have so far talked about making predictions that match the class
  - $y_i \sim w^T x_i$
    - For the whole training set,
      we're looking to find a vector of weights w,
      based on a matrix X of training samples,
      and class vector y

$$y \sim Xw$$

  - Nothing changes is we use different letters:

$$y \sim Ax$$

    - vector y and matrix A are known, x is unknown

Compressed Sensing

- Our problem:

$$y \sim Ax$$

  - vector y and matrix A are known, x is unknown
- Let's have a different view of what these mean:
  - x – some unknown signal (e.g. image, linearized from 2D to 1D)
  - A – some known matrix that transforms the signal
  - y – what we measure and have available for inspection
- Assume x has more dimensions than y
  - E.g. we have a point (x) in 3D, we project it on 2D (y)
    - 2 x 3 matrix A describes the angles of the projection
- If we know A and y, can we reconstruct x?

Compressed Sensing

- Let's first ask for exact equality:

  $$y = Ax$$

  - we first have some unknown physical signal x^*,
  - then some physical process creates $y=Ax^*$
  - we don't know what is x^*, but we can measure y
  - From there, can we get x such that $y=Ax$, and $x=x^*$
- Assume x has more dimensions (n) than y (m)
  - A is n by m: $y=Ax$ is underdetermined system of linear eqs.
  - In underdetermined case, there are many x that have $y=Ax$
  - Can we guess which one was the x^* that created y?
  - Impossible!

Compressed Sensing

- Let's first ask for exact equality:

$$Ax^* = y = Ax$$

  - Find x such that $y=Ax$ and $x=x^*$
- It would be possible is matrix A was square, and invertible (e.g. orthogonal columns)
  - $x^*=A^{-1}y$
- But it's impossible if A has more columns (n) than rows (m)
- Compressed sensing / sparse recovery:
  - We can do it, if x is sparse (has very few non-zeros)
  - And if matrix A is of some special type

Compressed Sensing

- Let's first ask for exact equality:

$$Ax^* = y = Ax$$

  - Find x such that $y=Ax$ and $x=x^*$
- Compressed sensing / sparse recovery:
  - A is m x n, ideally with $m<<n$
    - "Compressed": measure n-dim vector using $m<<n$ measured values, resulting in an m-dim vector y
  - x^* is guaranteed to be k-sparse
    - x^* has at most k nonzeros, e.g. $x^*=[0.1\ 0\ 0\ -2\ 0\ 0]$
  - Then, there are compressed sensing matrices for which recovering $x=x^*$ becomes possible

Compressed Sensing

$$Ax^* = y = Ax$$

- Knowing y but not x^*, find x such that $y=Ax$ and $x=x^*$

- Compressed sensing / sparse recovery:
  - There are matrices A that allow for recovering any k-sparse x^* from measurement $y=Ax^*$
- Two interrelated problems:
  - How to construct such matrices
    - construct A given m, n, k (if possible)
    - or given k, m, construct A with largest n
    - or given k, n, construct A with smallest m
  - How to use A (and y) to find x^*

Compressed Sensing

$$Ax^* = y = Ax$$

- Knowing y but not x^*, find x such that $y=Ax$ and $x=x^*$

- Compressed sensing / sparse recovery:
  - How to use A (and y) to find x^*
  - Return x with smallest L_1 norm
    - $x^* = \text{argmin } ||x||_1$
      subject to $y=Ax$
    - Or if measurement y is a bit noisy (up to ε error)
      - $x^* = \text{argmin } ||x||_1$
        subject to $||y - Ax|| <= \varepsilon$

Compressed Sensing

$$Ax^* = y = Ax$$

- Knowing y but not x^*, find x such that $y=Ax$ and $x=x^*$

- Compressed sensing / sparse recovery:
  - How to use A (and y) to find x^*
  - Return x with smallest L_1 norm
    - $x^* = \text{argmin} \, ||x||_1$ subject to $y=Ax$
  - Why? We can construct matrices A where: if x^* is k-sparse and $y=Ax^*$ then all other x' with $y=Ax'$ have higher L_1 norm

Compressed Sensing

- We can construct matrices A where
  if x* is k-sparse and y=Ax*
  then all other x' with y=Ax' have higher L_1 norm
- Matrix A has Nullspace Property
  if
  - For each v such that Av=0 (except v=0):
    - Let v_k be a vector resulting from v by keeping k coefficients, and setting every other coefficient to zero
    - Let $v_{\sim k} = v - v_k$ (i.e., the rest)
      - E.g. v=[1 2 3] for k=2: v_k=[1 2 0], or [1 0 3], or [0 2 3]
        $v_{\sim k}$=[0 0 3], or [0 2 0], or [1 0 0]
    - We require that always $||v_k||_1 < ||v_{\sim k}||_1$

Compressed Sensing

- Nullspace Property of matrix A
  - For each v such that Av=0 (except v=0)
    - Let v_k be a vector resulting from v by keeping up to k coefficients, and setting every other coefficient to zero
    - We require that $||v_{\sim k}||_1 > ||v_k||_1$ where $v_{\sim k} = v - v_k$
- Why is this sufficient?
  - y=Ax* where x* is k-sparse
  - There are other x' with y=Ax'
  - Define v=x*-x', then Av = Ax* - Ax' = y-y = 0
  - Pick v_k that has zeros where x* has zeros, we have $||v_{\sim k}||_1 > ||v_k||_1$
  - Then $v_{\sim k} = x^*_{\sim k} - x'_{\sim k} = -x'_{\sim k}$ and $v_k = x^*_k - x'_k$
  - $||x'||_1 = ||x'_k||_1 + ||x'_{\sim k}||_1 = ||x'_k||_1 + ||v_{\sim k}||_1 > ||x'_k||_1 + ||v_k||_1$
  - But $||x'_k||_1 + || x^*_k - x'_k ||_1 >= || x^*_k ||_1$
    - from triangle inequality: $|| x^*_k || + ||-x'_k|| <= || x^*_k - x'_k ||$
- So, alternative solutions x' will have higher L_1 norm than x*

x*=[1 1 0 0]
x' =[0 -1 3 -5]
v=[1 2 -3 5],
v_k=[1 2 0 0]
$v_{\sim k}$=[0 0 -3 5]

Compressed Sensing

- Nullspace Property of matrix A
  - For each v such that Av=0 (except v=0)
    - Let v_k be a vector resulting from v by keeping up to k coefficients, and setting every other coefficient to zero
    - We require that $||v_{\sim k}||_1 > ||v_k||_1$ where $v_{\sim k} = v - v_k$
- NSP is one of several properties that guarantee sparse recovery for a given k
- Another one is Restricted Isometry Property (RIP)
  - For every 2k-sparse vector v, we have:

$$(1-\delta)||v||_2^2 \leq ||Av||_2^2 \leq (1+\delta)||v||_2^2$$

  with some small δ

- There is also RIP-1

$$(1-\delta)||v||_1 \leq ||Av||_1 \leq ||v||_1.$$

Compressed Sensing

$$Ax^* = y = Ax$$

- Knowing y but not x^*, find x such that $y=Ax$ and $x=x^*$
- Compressed sensing / sparse recovery:
  - There are matrices A that allow for recovering any k-sparse x^* from measurement $y=Ax^*$
- Two interrelated problems:
  - How to construct such matrices
    - construct A given m, n, k (if possible)
    - or given k, m, construct A with largest n
    - or given k, n, construct A with smallest m
  - How to use A (and y) to find x^*

Compressed Sensing

$$Ax^* = y = Ax$$

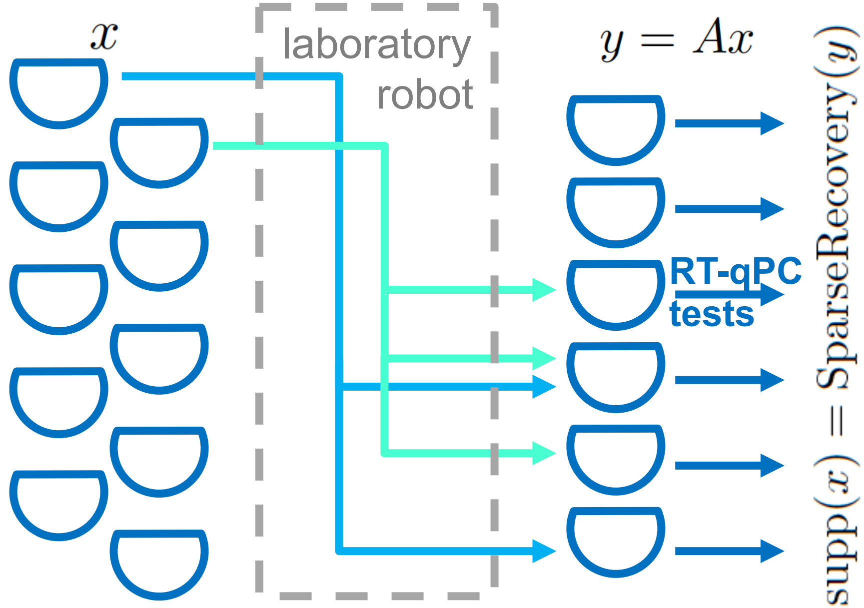
- Knowing y but not x^*, find x such that $y=Ax$ and $x=x^*$
- Compressed sensing / sparse recovery:
  - There are matrices A that allow for recovering any k-sparse x^* from measurement $y=Ax^*$
- Two interrelated problems:
  - How to construct such matrices
    - construct A given m, n, k (if possible)
    - or given k, m, construct A with largest n
    - or given k, n, construct A with smallest m
  - Just find a matrix that has NSP (or RIP, or RIP-1) property

Sparse recovery in viral testing

- Recently, compressed sensing has been proposed as a way to increase throughput in viral testing
  - x_i = viral load in a sample collected from person i
  - y_j = measurement from test j

$$A = \mathrm{RecoveryMatrix}(m = 6, k = 2, d = 3)$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$



Nonnegative sparse recovery

Compressed sensing has been explored in depth for arbitrary unknown vectors x, with both positive and negative elements x_i

Compressed sensing in which unknown vectors are nonnegative, $x \geq 0$, has not received as much attention

In viral testing, we deal with nonnegative values in all elements of the $y = Ax$ equation

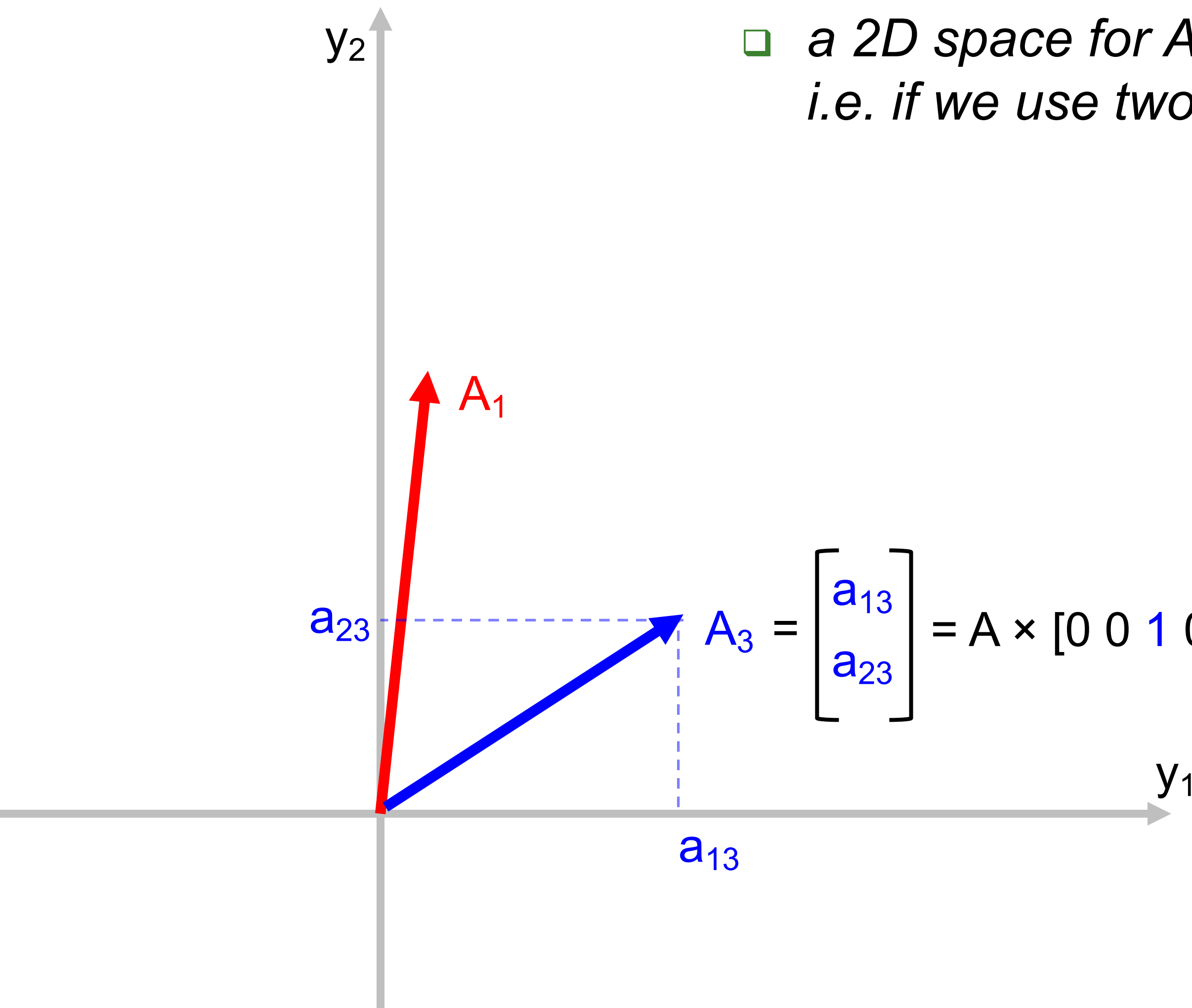
- unknown viral loads x_i are nonnegative
- known measurement matrix A is nonnegative
- known quantitative test results y_j are nonnegative

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \tfrac{1}{2} & 1 & 0 \\ \tfrac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \tfrac{1}{2} & 1 & 0 \\ \tfrac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ -1 \end{bmatrix}$$

Measurement space *y*

- Columns of matrix A are vectors in the *measurement space*, the space of all possible measurement outcomes *y*
    - *a 2D space for A with two rows, i.e. if we use two qPCR tests*

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{ qPCR}$$

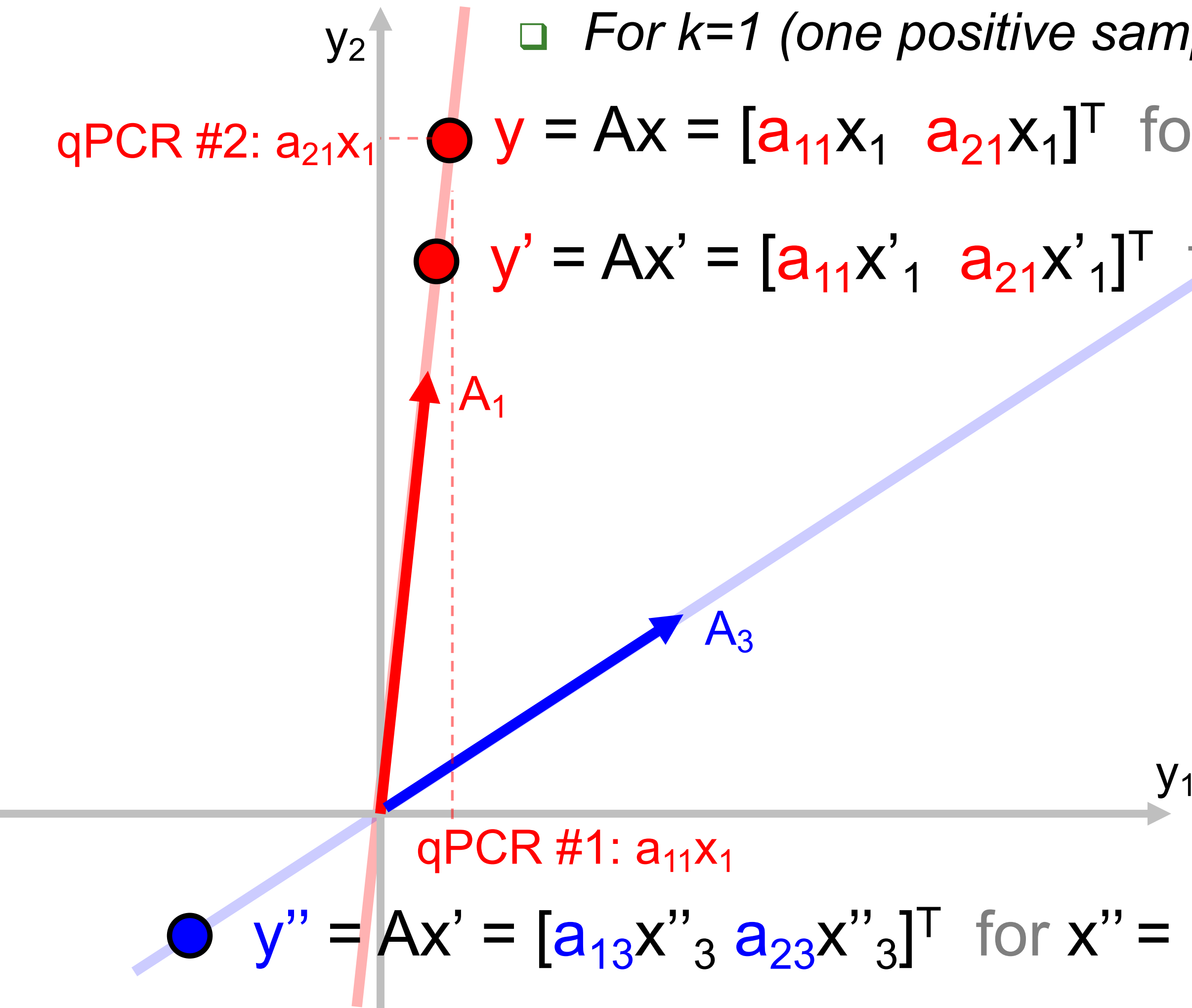


$$A_3 = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix} = A \times [0\ 0\ 1\ 0\ ]^T$$

$$A = \begin{array}{c} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \\ \begin{array}{cccc} A_1 & \quad & A_3 & \quad \end{array} \end{array}$$

k-sparse signals in *y* space

- all *k*-sparse signals *x* with the same *support* (0s/non-0s) form *k*-dimensional hyperplane in measurement space *y*
  - *For k=1 (one positive sample), we have lines*



$y = Ax = [a_{11}x_1 \;\; a_{21}x_1]^T$ for $x = [x_1 \; 0 \; 0 \; 0\,]^T$

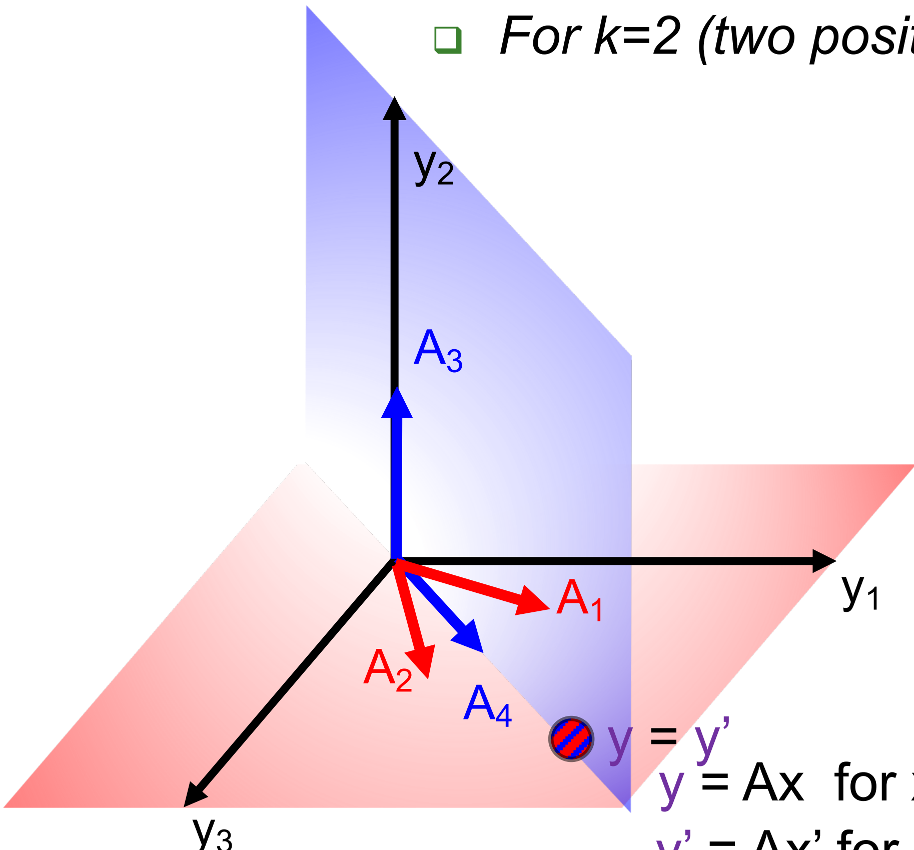
$y' = Ax' = [a_{11}x'_1 \;\; a_{21}x'_1]^T$ for $x' = [x'_1 \; 0 \; 0 \; 0\,]^T$

$$A = \begin{array}{c} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \\ \begin{array}{cccc} A_1 & \quad & A_3 & \quad \end{array} \end{array}$$

$y'' = Ax' = [a_{13}x''_3 \; a_{23}x''_3]^T$ for $x'' = [0 \; 0 \; x''_3 \; 0\,]^T$

Ambiguity of measurements

- all k-sparse signals x with the same support form k-dimensional hyperplane in measurement space y
  - *For k=2 (two positives), we have 2D planes*



$$
\begin{array}{cc}
A= & \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \\
 & \begin{matrix} A_1 & \;\; A_2 & \;\; A_3 & \;\; A_4 \end{matrix}
\end{array}
$$

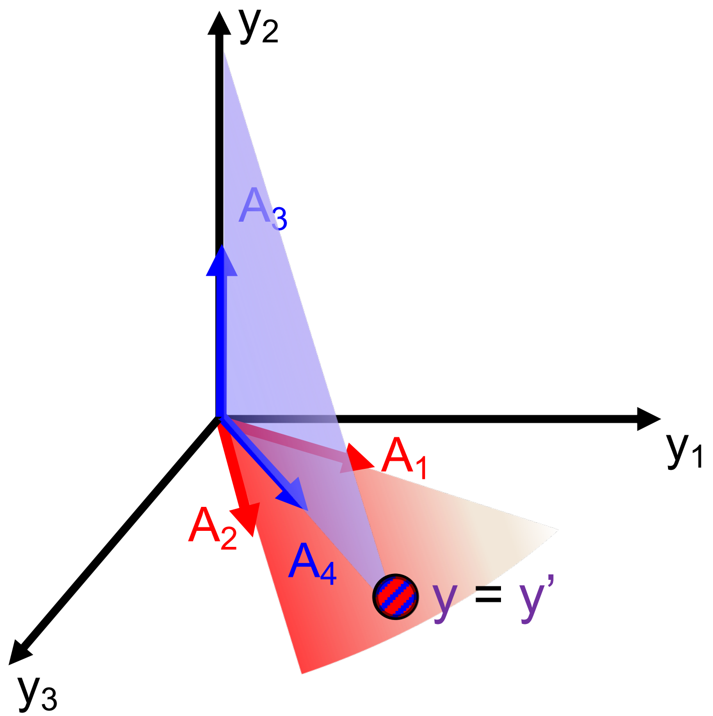
  - *Two vectors x, x'*
    with different support
    (i.e., two different pairs of positives)
    lead to the same
    measurements y=y'

$y = Ax$ for $x = [x_1\ x_2\ 0\ 0\ ]$

$y' = Ax'$ for $x' = [0\ 0\ x_3\ x_4\ ]$

Nonnegative signals

- k-sparse signals $x \geq 0$ with the same support form k-dimensional cones in the measurement space y
  - *For k=2, we have 2D cones*

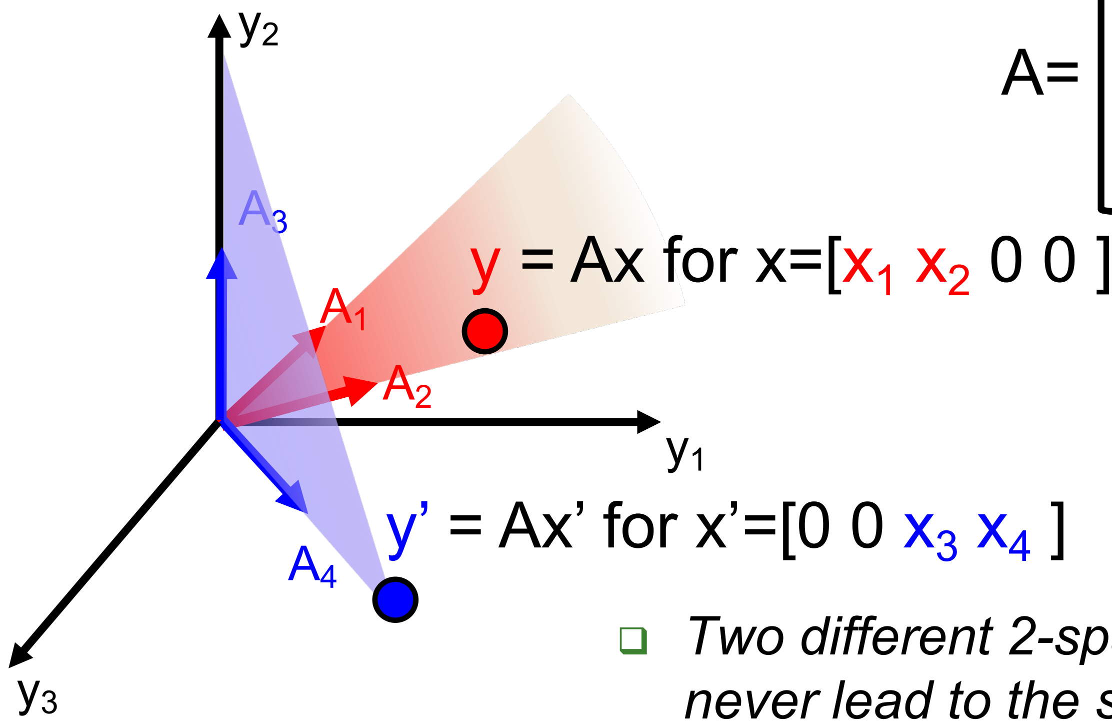


- *Cones intersect – same problem:*
- *two different vectors x, x' same measurements y=y'*

Why nonnegativity matters?

- Cones are easier to arrange than hyperplanes, it's easier to prevent intersections
  - **Goal: find matrices A with columns that define non-intersecting cones**

$$A = \begin{bmatrix} a'_{11} & a'_{12} & a_{13} & a_{14} \\ a'_{21} & a'_{22} & a_{23} & a_{24} \\ a'_{31} & a'_{32} & a_{33} & a_{34} \end{bmatrix}$$

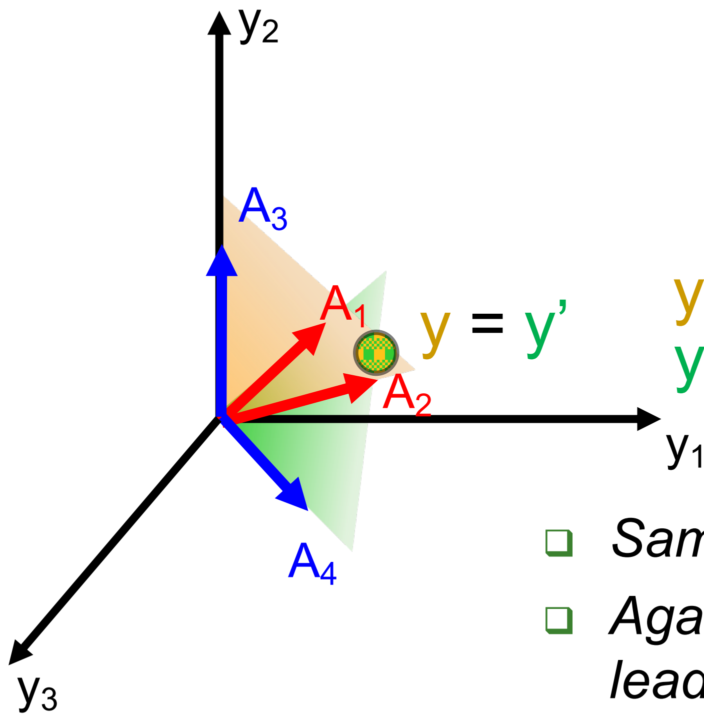


  - *Two different 2-sparse vectors x, x' never lead to the same measurements*

Is this a good set of columns?

- We have to consider all possible combinations of pairs (for *k*=2) of columns
  - The four columns A_1-A_4 from previous slide do not actually work well

$$A = \begin{bmatrix} a'_{11} & a'_{12} & a_{13} & a_{14} \\ a'_{21} & a'_{22} & a_{23} & a_{24} \\ a'_{31} & a'_{32} & a_{33} & a_{34} \end{bmatrix}$$



$y = Ax$ for $x=[0\ x_2\ x_3\ 0\ ]$
$y' = Ax'$ for $x'=[x_1\ 0\ 0\ x_4\ ]$

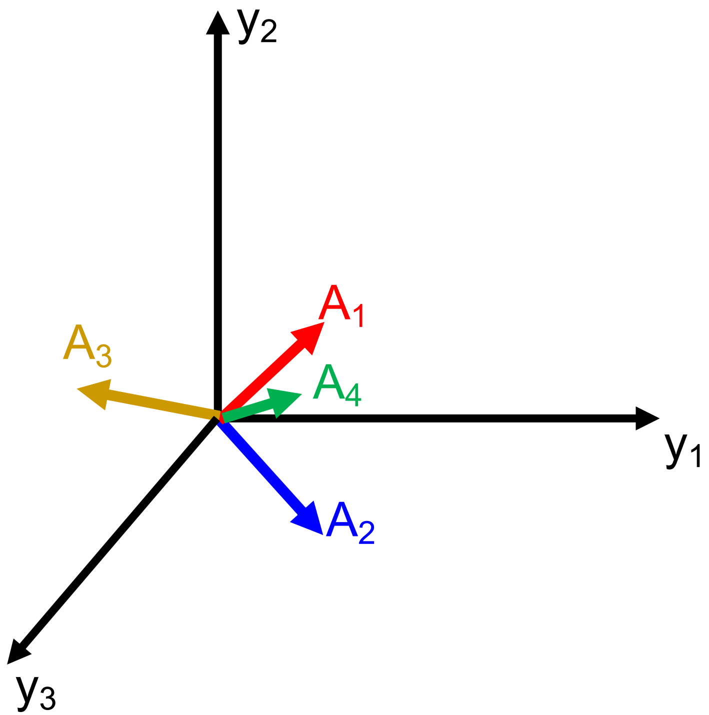
  - *Same matrix, different pairing of columns*
  - *Again, two different vectors x, x' lead to the same measurements y=y'*

Finding good matrices A

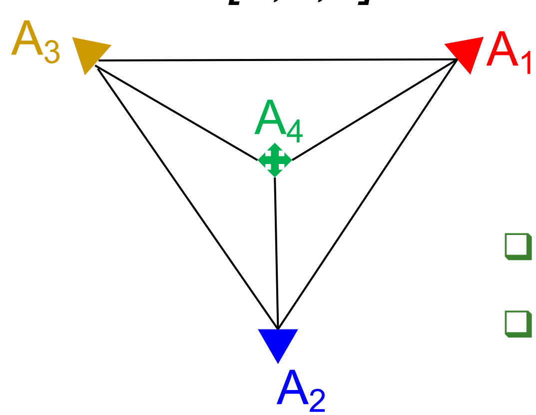
- Can we construct a 3 x 4 matrix (4 samples, 3 tests) with no intersection of any 2D-cones?

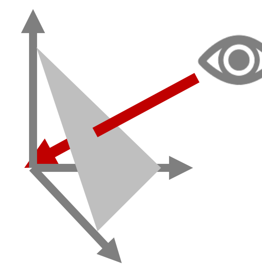
$$A = \begin{bmatrix} \tfrac{1}{2} & \tfrac{1}{2} & 0 & \tfrac{1}{3} \\ \tfrac{1}{2} & 0 & \tfrac{1}{2} & \tfrac{1}{3} \\ 0 & \tfrac{1}{2} & \tfrac{1}{2} & \tfrac{1}{3} \end{bmatrix}$$

  - Yes!



  - *2D view from [1,1,1] towards [0,0,0]*





  - *Edge ~ cone*
  - *No two edges intersect!*

How good is this matrix A?

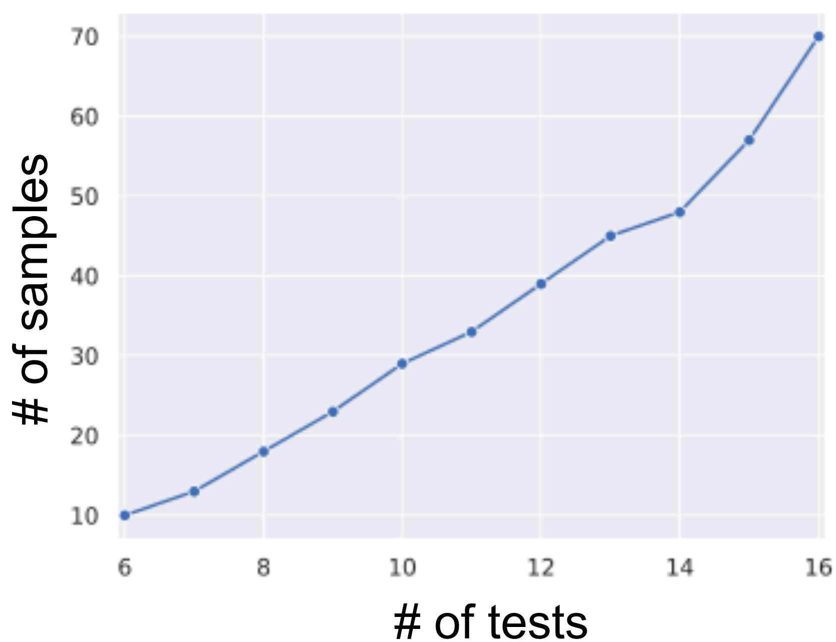
- For k=2, any k-sparse nonnegative vector cannot be mistaken for any other k-sparse vector
  - y=[2 1 1] $\Longrightarrow$ x=[2 2 0 0]
  - y=[2 5 5] $\Longrightarrow$ x=[0 0 2 3]

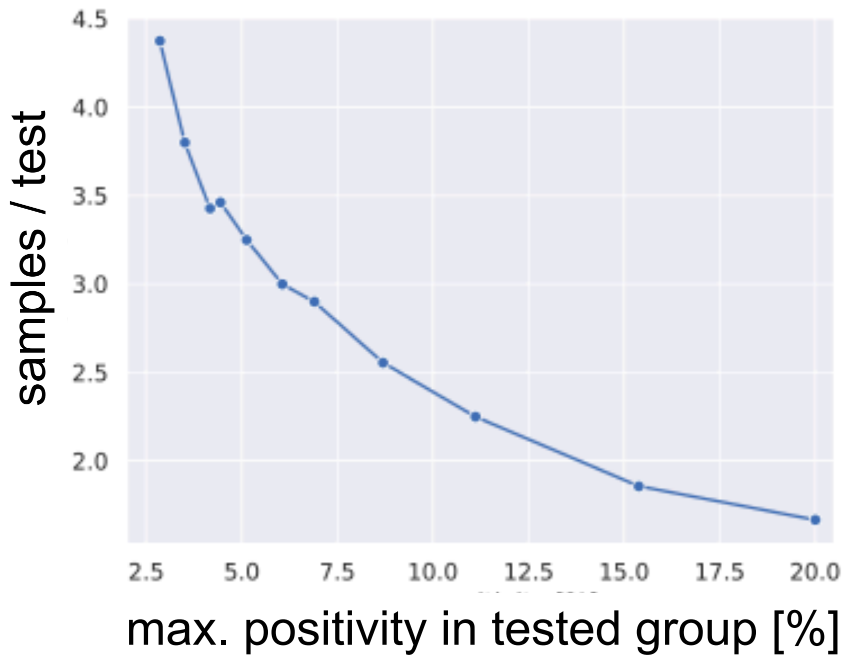
$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

- But, a k-sparse vector can still be mistaken for a non-k-sparse decoy vector
  - y=[2 2 2] $\Longrightarrow$ x=[0 0 0 6] (2-sparse: at most 2 non-zeros)
  - y=[2 2 2] $\Longrightarrow$ x=[2 2 2 0] (not 2-sparse: more than 2 non-zeros)
- We can do better:
  - RIP-1 (Berinde et al. 2008): No decoys with *sparsity*(original vector + decoy) > 2k
    - Matrix A is not RIP-1 for k=2
  - NNSP (Saeedi et al., under review): No decoys at all

Uniqueness: no decoys possible

- We show a way to construct small matrices where
  - If the unknown vector x is *k*-sparse
    there are no other nonnegative solutions to y=Ax (no decoys)
- We constructed a series of “no-decoys” matrices





(Saeedi, Serrano, Buck, Brooks & Arodz, under review)

Caveats

- The quality of sparse recovery deteriorates
  if we make an error in setting sparsity parameter *k*
  - E.g. matrix assumes k=2, signal has >2 non-zeros: decoys reappear
    - this will happen if *k* is set assuming <5% positivity rate
      but the tested group has 6%

- The quality of sparse recovery deteriorates
  with experimental measurement noise
  - Measurement has noise: decoys reappear
  - We are working towards better
    theoretical understanding of noise robustness
    for nonnegative sparse recovery
    (Saeedi, Yang & Arodz, in preparation)