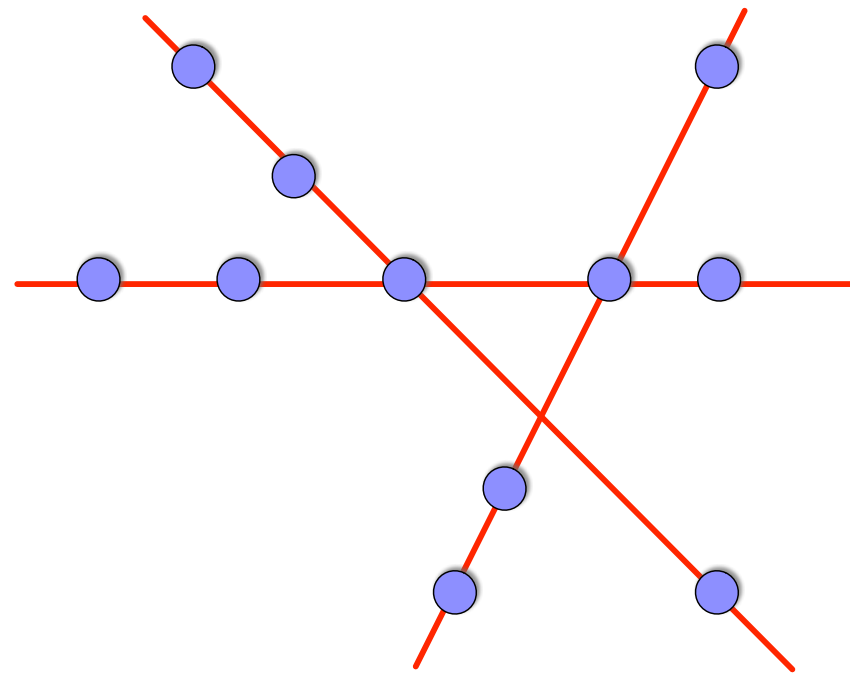
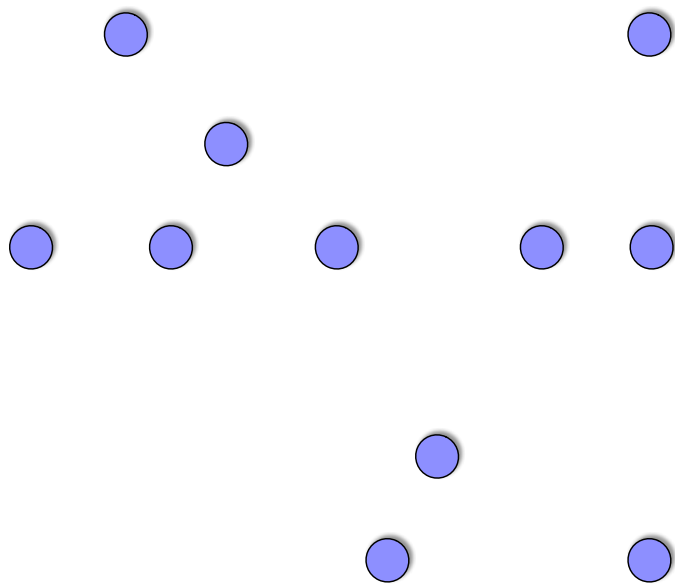


Line Cover and Related Problems

Matthias Bentert, Fedor Fomin, Petr Golovach, Souvik Saha, Sanjay Seetharaman, and Anannya Upasana

Point Line Cover

Task: Given a set P of n points in the plane and an integer k , find k lines that cover all the points.



Point Line Cover

Task: Given a set P of n points in the plane and an integer k , find k lines that cover all the points.

- is NP-complete [Megiddo, Tamir, 1982]
- admits a kernel with $O(k^2)$ points
- solvable in time $k^{O(k)} \cdot n^{O(1)}$
- Hyperplane Cover FPT($k+d$) [Langerman and Morin, 2005]

Line Clustering

LINE CLUSTERING

Input: A multiset $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of points in $\mathbb{R}^2$ and integers k and B .

Question: Are there k lines $L_1, L_2, \dots, L_k$ such that $\sum_{i=1}^n \min_{1 \leq j \leq k} \text{dist}(\mathbf{x}_i, L_j)^2 \leq B$?

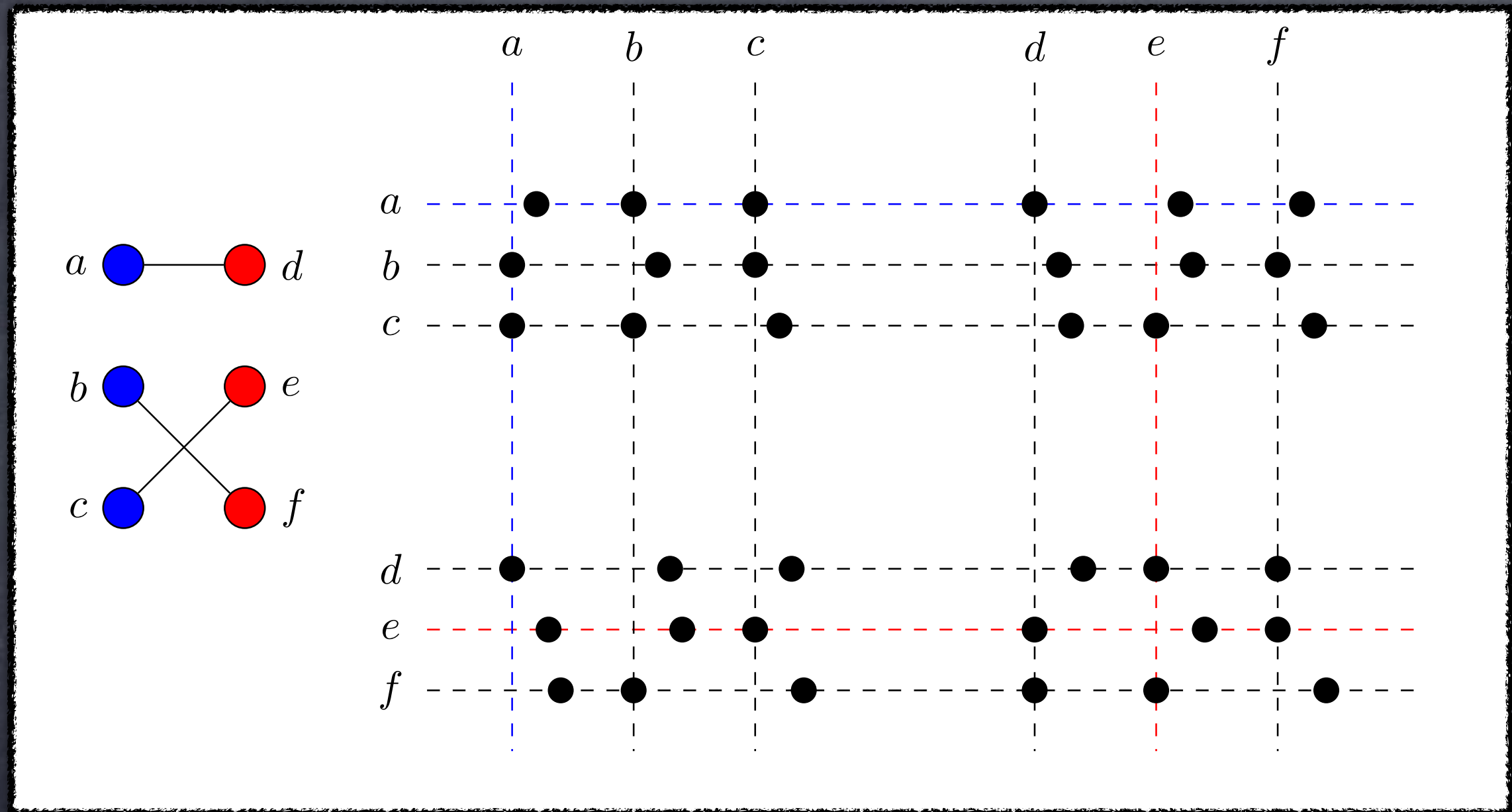
Our results

Line clustering is $W[1]$ -hard parameterized by k

Projective clustering is XP parameterized by k
and dimension d

► **Theorem 1.** *LINE CLUSTERING is $W[1]$ -hard parameterized by k . Assuming the ETH, it cannot be solved in $n^{o(k)}$ time.*

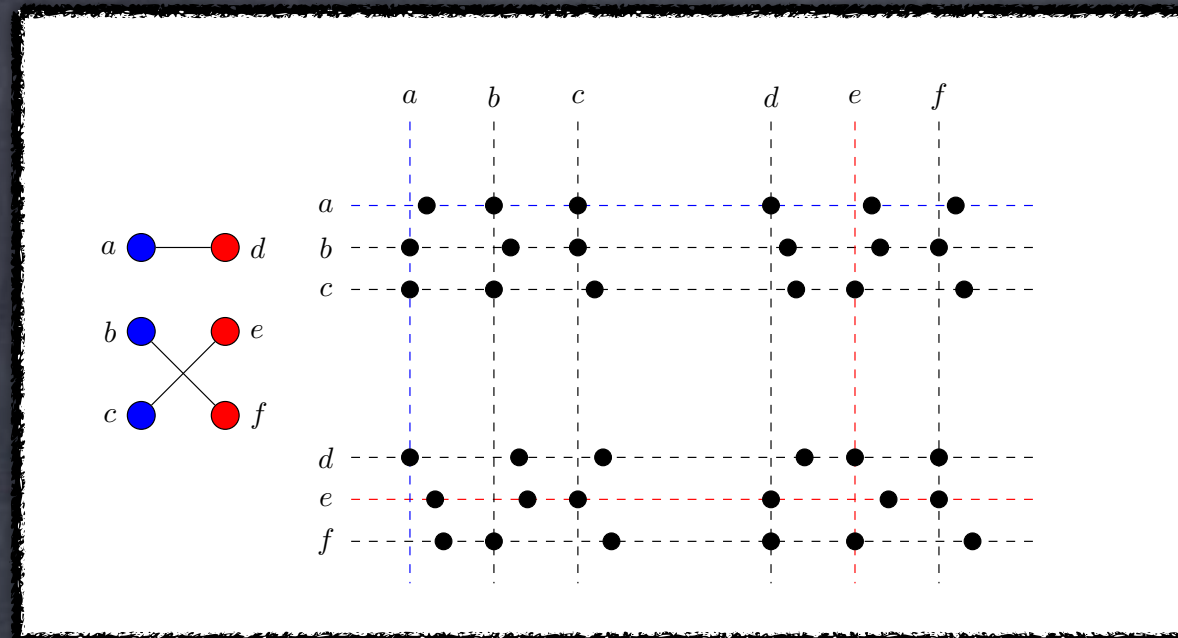
Reduction idea



Regular Multicolored Independent Set of size k

k vertical and k horizontal lines hitting max number of points

Reduction idea



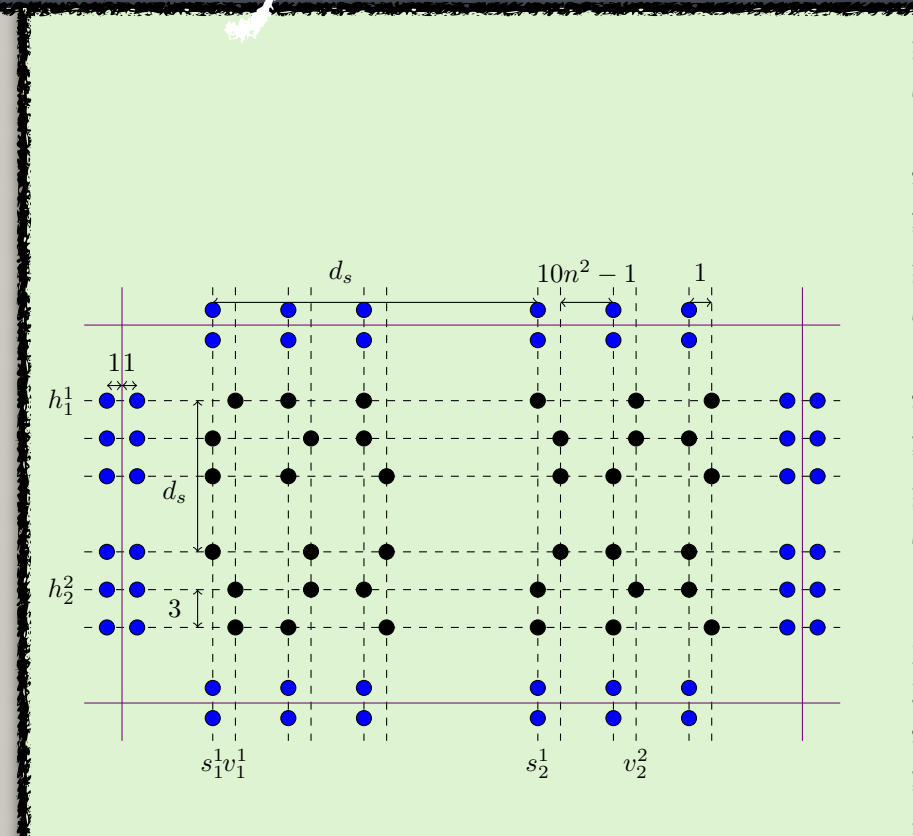
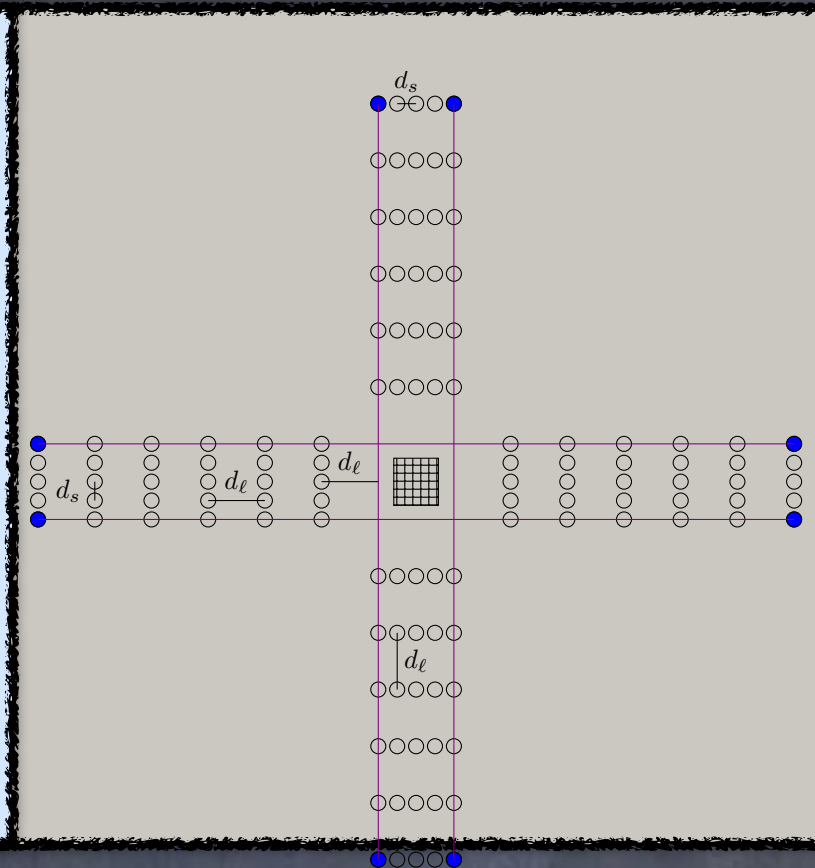
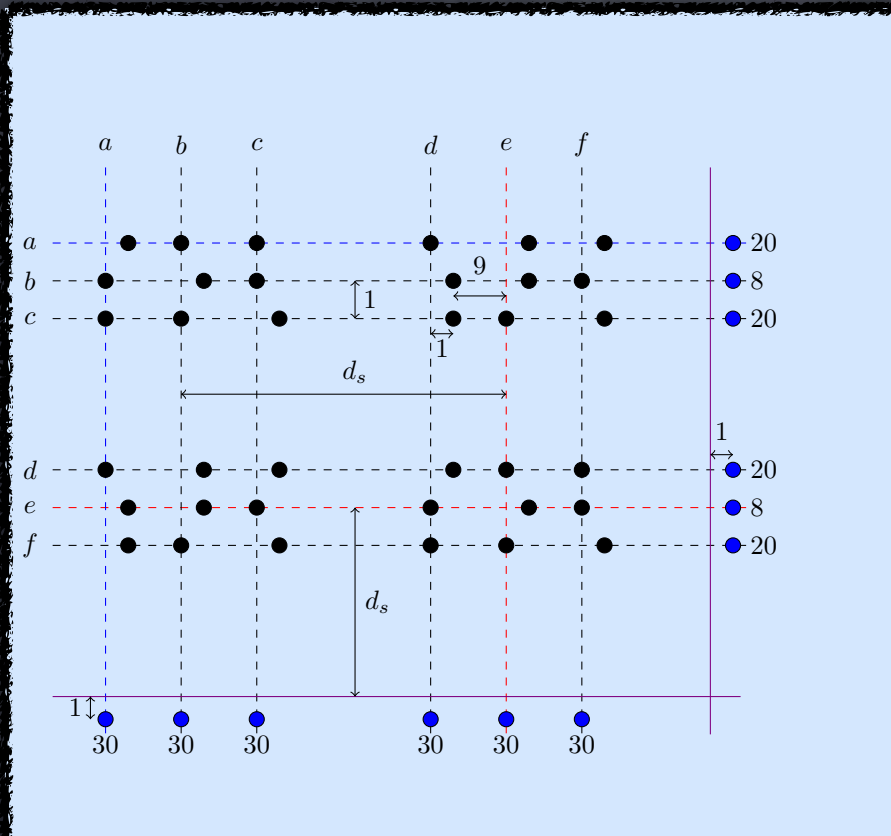
Vertex $\leftrightarrow$ horizontal and vertical line

Two vertices cannot be in Colored IS $\leftrightarrow$ intersection of the corresponding lines contains a point

Regular Multicolored Independent Set of size k

k vertical and k horizontal lines hitting max number of points

Reduction idea: not that easy...

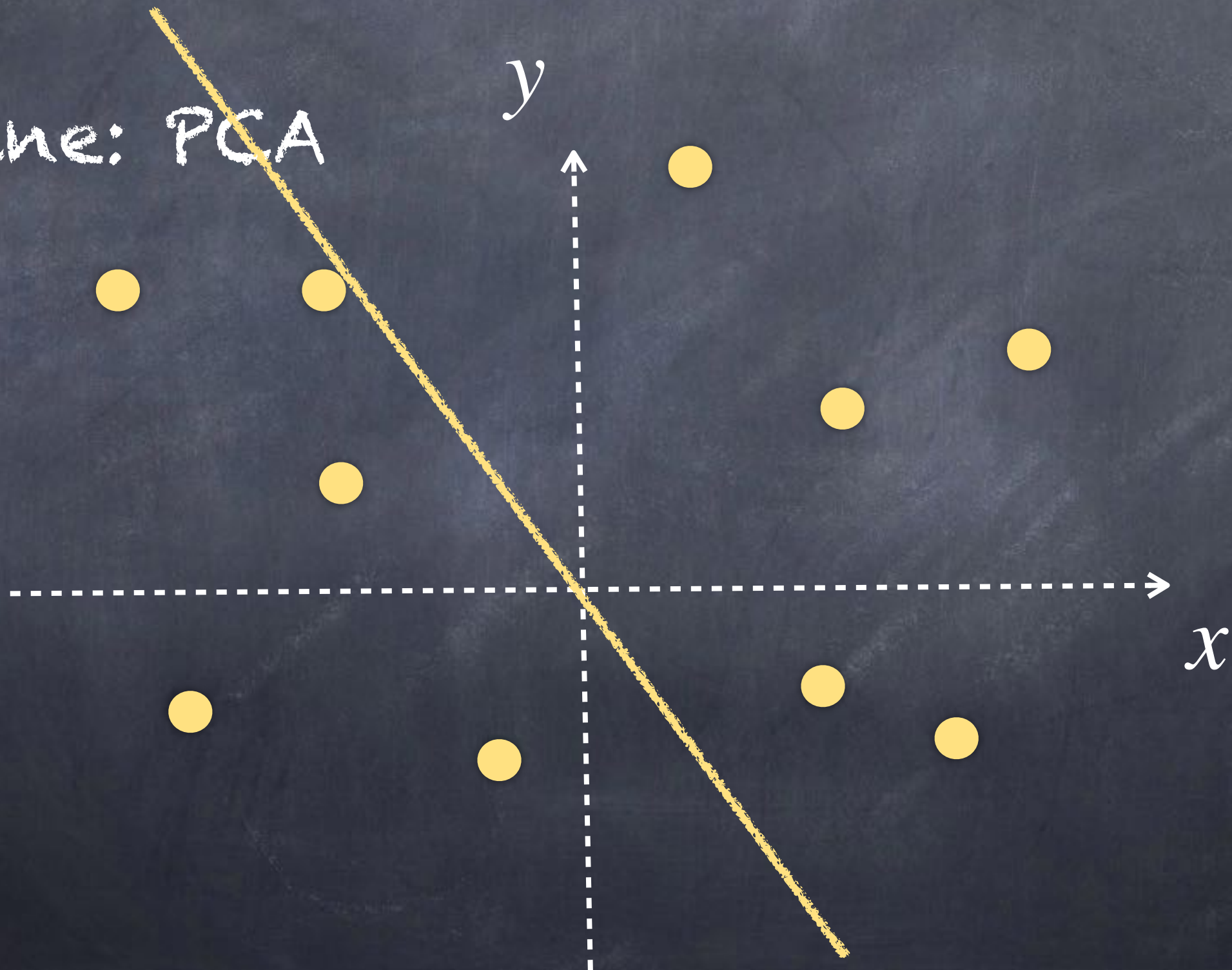


- How to ensure that lines are vertical and horizontal?
- We do not cover points but minimize sum of squares of distances
- at the end becomes pretty technical...

► **Theorem 1.** *LINE CLUSTERING is $W[1]$ -hard parameterized by k . Assuming the ETH, it cannot be solved in $n^{o(k)}$ time.*

What about XP?

One Line: PCA



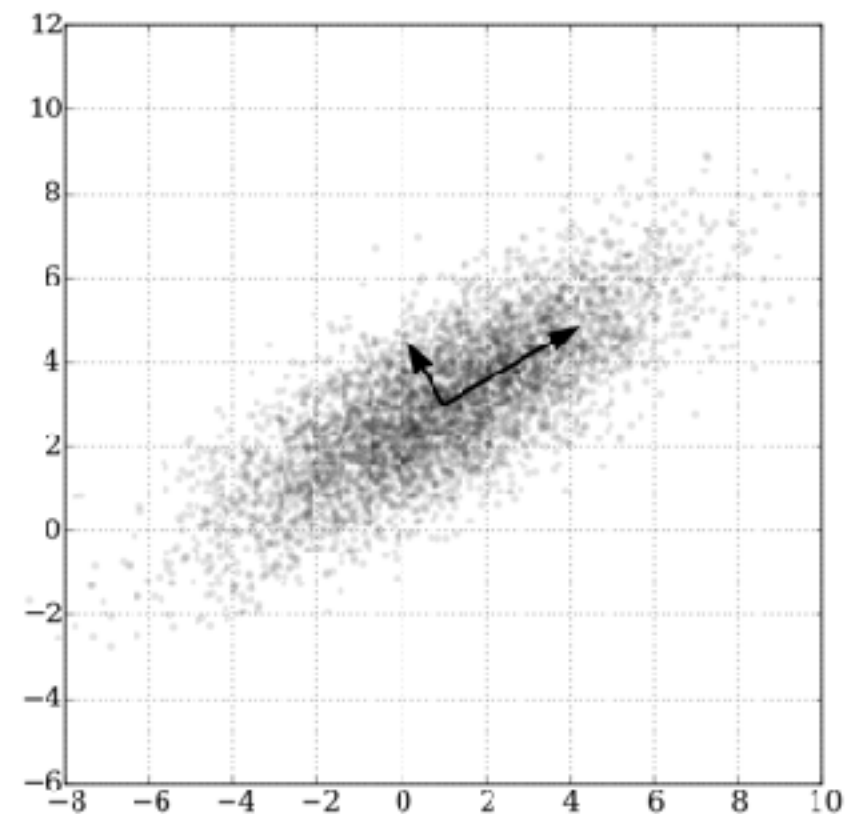
Principal component analysis (PCA)

minimize $\|M - L\|_F^2$
subject to $\text{rank}(L) \leq r$.

// $\|A\|_F^2 = \sum_{i,j} a_{ij}^2$
Frobenius norm

Sum of the squares of the Euclidean distances between rows (columns) of $M-L$

[Eckart & Young, 1936]

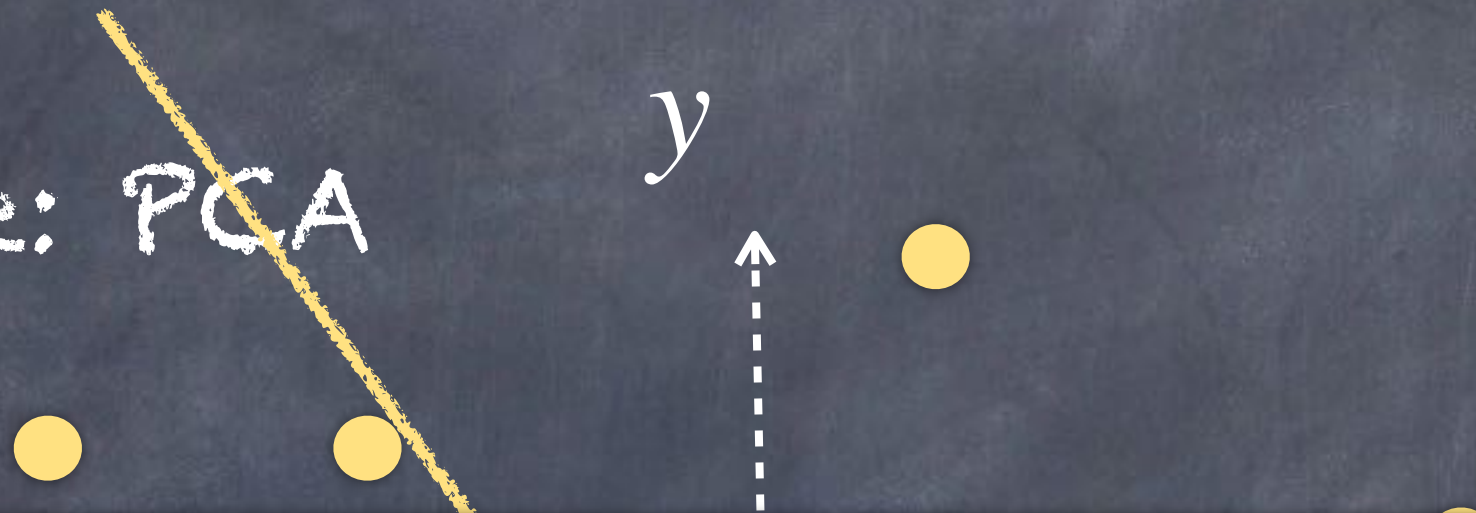


Source: Wikipedia

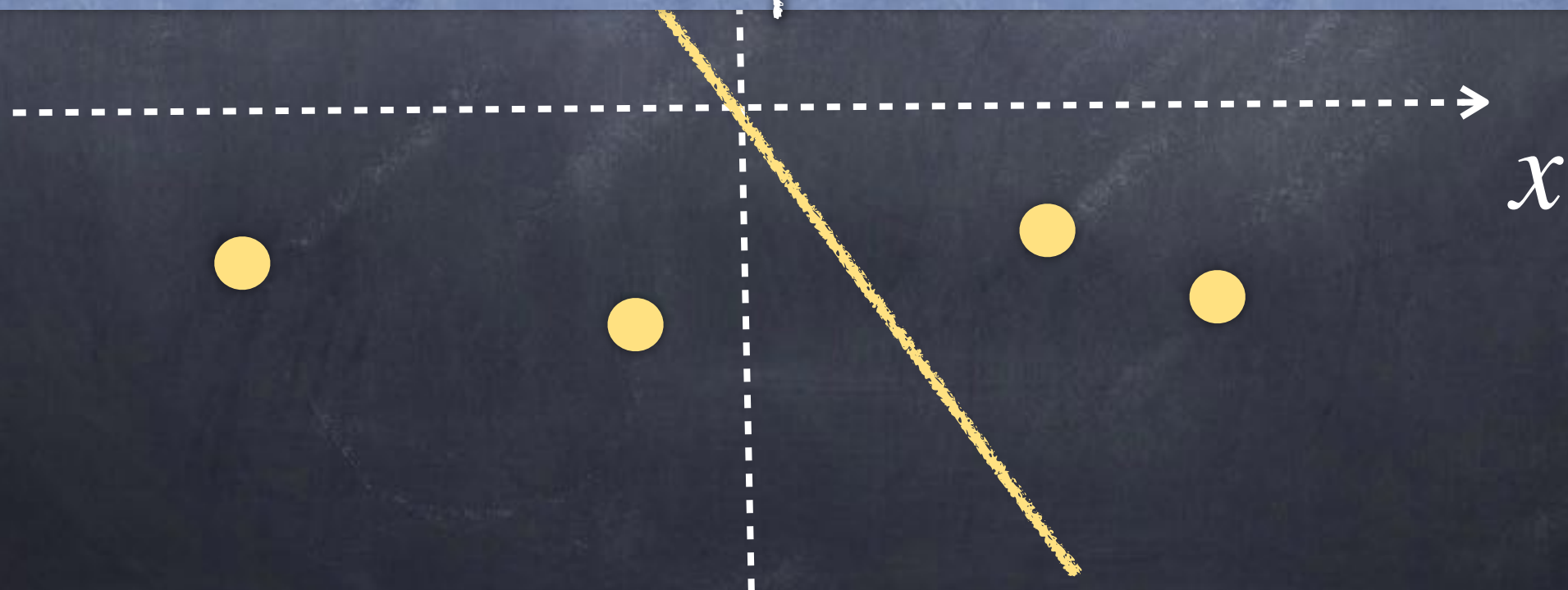
► **Theorem 1.** *LINE CLUSTERING is $W[1]$ -hard parameterized by k . Assuming the ETH, it cannot be solved in $n^{o(k)}$ time.*

What about XP?

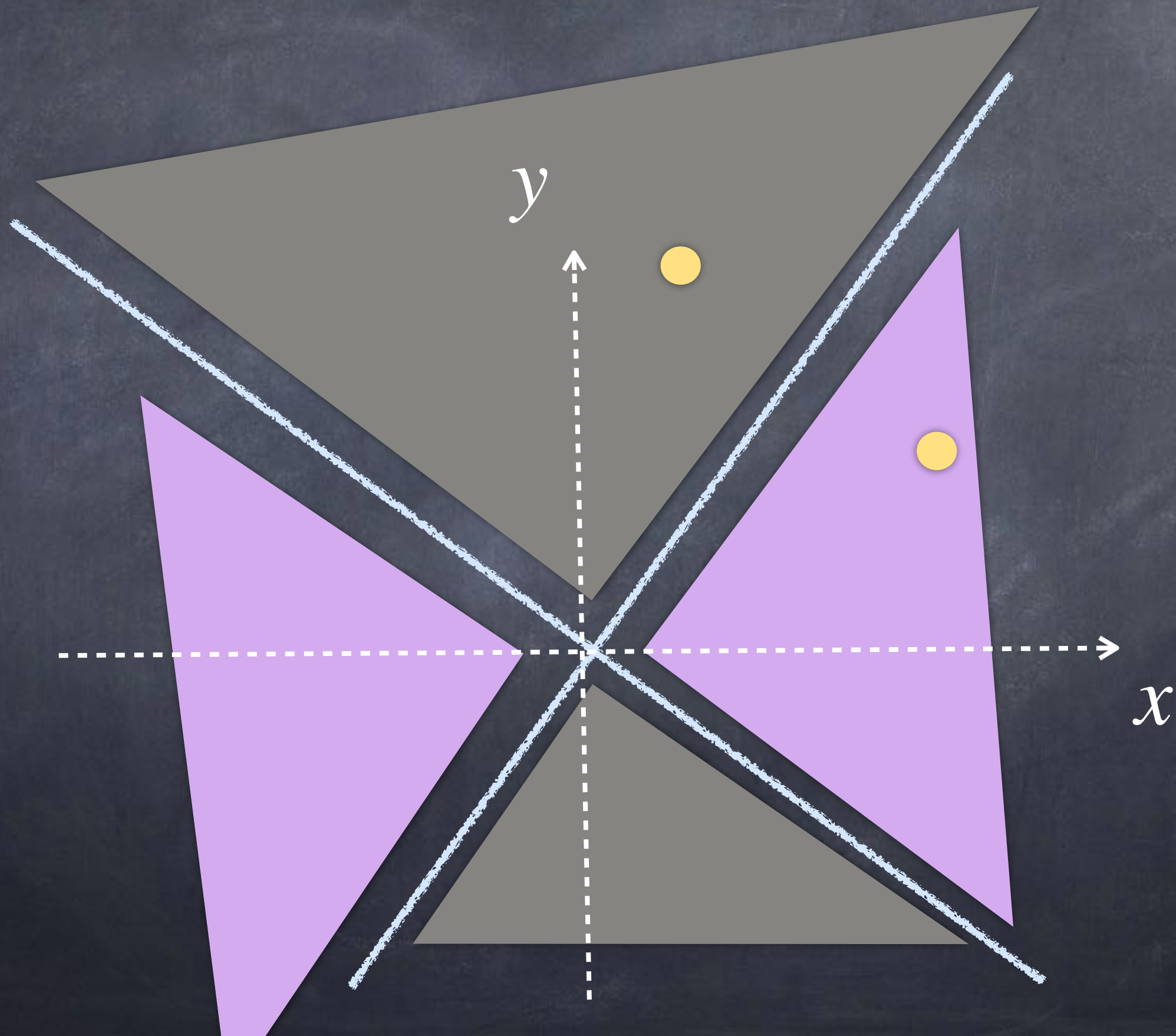
One Line: PCA



If we know how to partition points in clusters, we solve the problem



Intuition: Voronoi



Intuition

- Number of different clusters (for 2 lines) depend only on the distances between pairs of points
- Leaves us with $O(n^2)$ clusters
- $n^{O(k)}$ clusters for k lines
- $n^{O(k)}$ calls of PCA

PROJECTIVE CLUSTERING

Input: A multiset $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of points in $\mathbb{R}^d$ and integers r , k , and B .

Question: Are there k r -dimensional affine subspaces $A_1, A_2, \dots, A_k$ whose union H satisfies $\sum_{i=1}^n \text{dist}(\mathbf{x}_i, H)^2 \leq B$?

- PCA for $k = 1$
- Line Clustering: $d = 2$ and $r = 1$
- k -means clustering: $r = 0$
- Hyperplane Cover: $B = 0$

PROJECTIVE CLUSTERING

Input: A multiset $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of points in $\mathbb{R}^d$ and integers r , k , and B .

Question: Are there k r -dimensional affine subspaces $A_1, A_2, \dots, A_k$ whose union H satisfies $\sum_{i=1}^n \text{dist}(\mathbf{x}_i, H)^2 \leq B$?

Theorem

PROJECTIVE CLUSTERING can be solved in $n^{\mathcal{O}(\min\{dk(r+1), dk(d-r+1)\})}$ time.

Main idea

Encode the information about equivalent partitions of points by a family of nk^2 polynomials of degree 4 and $O(dkr)$ variables

Changes of signs (algebraic cells) of this system correspond to equivalent partitions of points

Use the powerful tools from algebraic geometry (Algorithm of Basu-Pollack-Roy) to identify algebraic cells in time $n^{O(dkr)}$

Run PCAs for each of the partitions

Open questions

FPT approximation of Projective Clustering:
approximate B with ratio $f(k,d)$ in time $g(k,d)$
 $\text{poly}(n)$?

Deshpande, Rademacher, Vempala, and Wang
(2006): PTAS when r and k are constants

Thank you!

