Random-projection ensemble dimension reduction

Wenxing Zhou and Timothy I. Cannings

School of Mathematics, University of Edinburgh

Can many random projections recover the right low-dimensional view of the data?

Goal: find a low-dimensional representation that preserves E(Y | X)

- High-dimensional regression suffers from the curse of dimensionality.

Goal: find a low-dimensional representation that preserves E(Y | X)

- High-dimensional regression suffers from the curse of dimensionality.
- We do not want arbitrary compression.

$$\mathbb{E}(Y \mid X = x) = g(A^\top x), \qquad d < p$$

Goal: find a low-dimensional representation that preserves E(Y | X)

- High-dimensional regression suffers from the curse of dimensionality.
- We do not want arbitrary compression.

$$\mathbb{E}(Y \mid X = x) = g(A^\top x), \qquad d < p$$

$A^T X$ preserves all the information in X

about the conditional mean of Y.

We want the smallest such dimension d_0 and the corresponding directions A_0 .

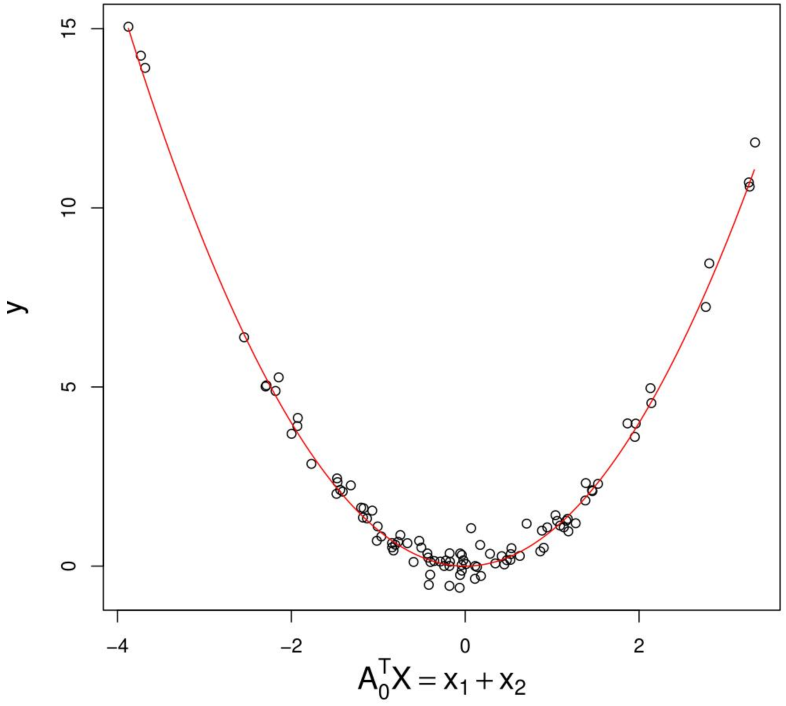
A single random projection can fail

many random projections

...

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

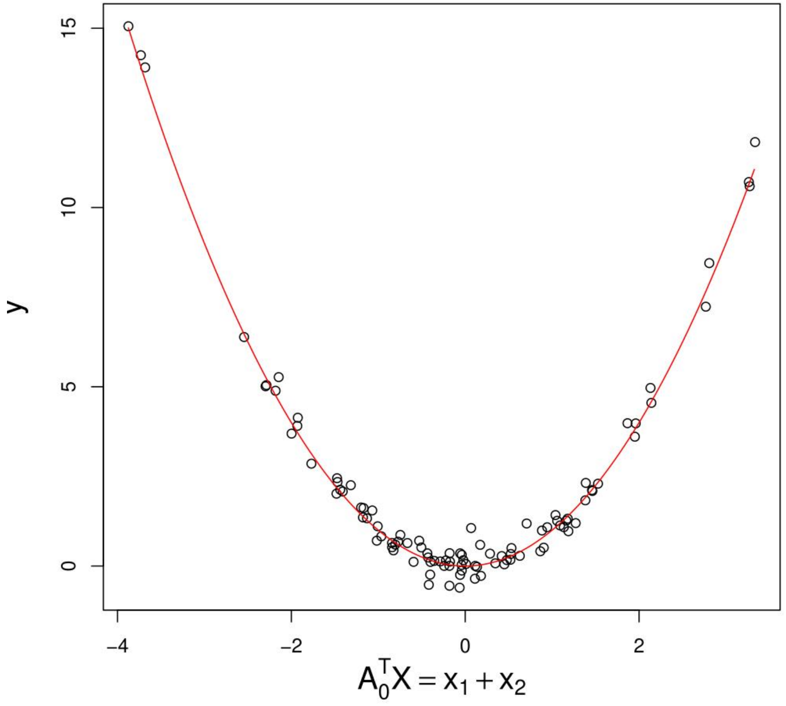
$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$



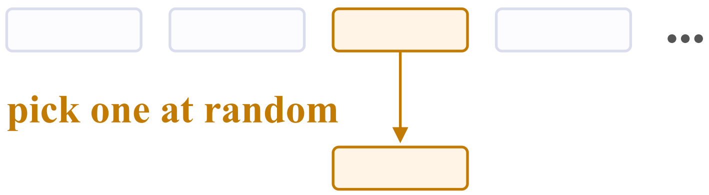
A single random projection can fail

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \dots, 0)^\top / \sqrt{2}$



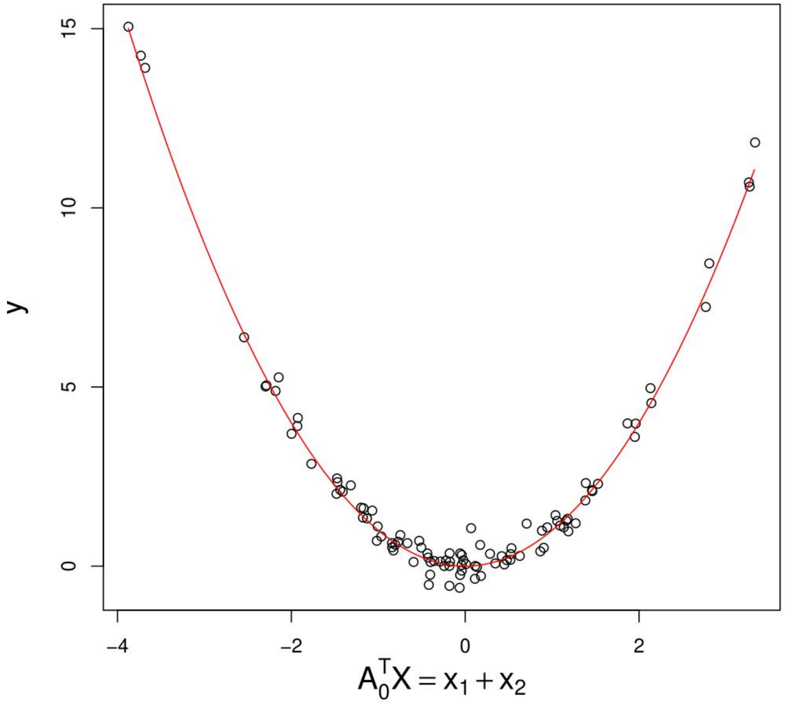
many random projections

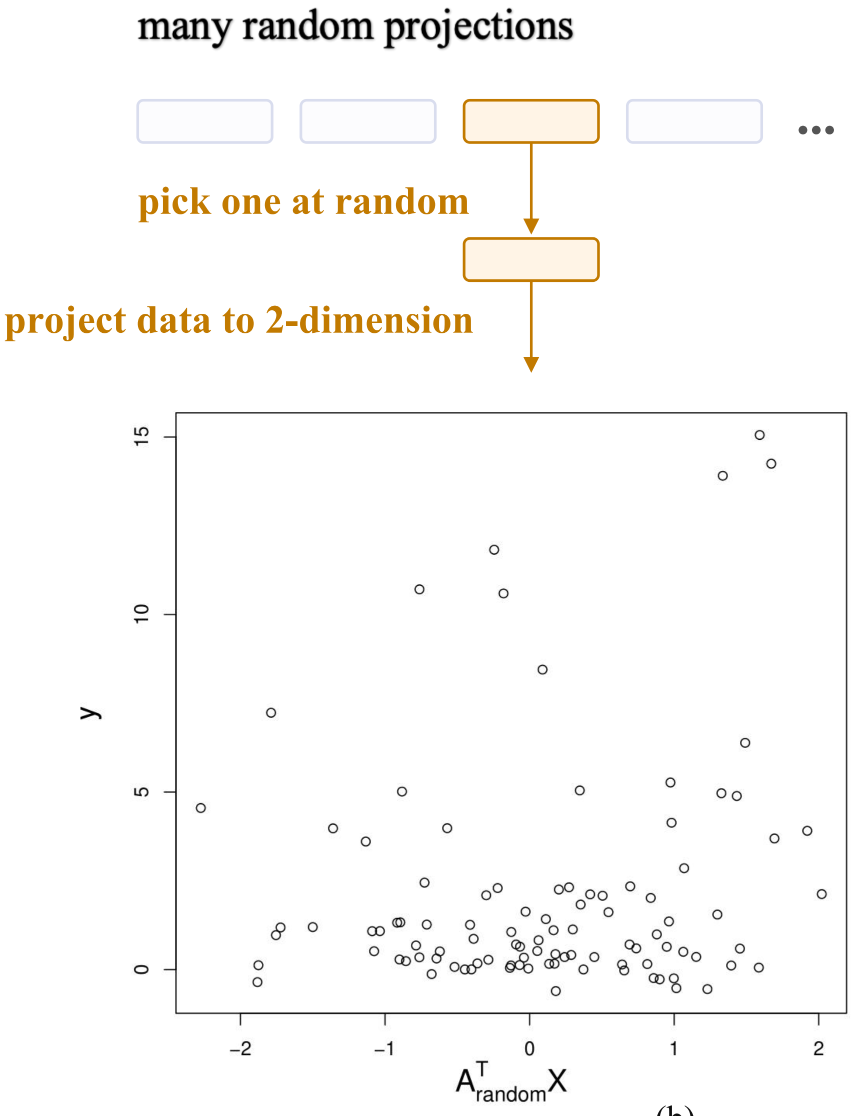


A single random projection can fail

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \dots, 0)^\top / \sqrt{2}$$



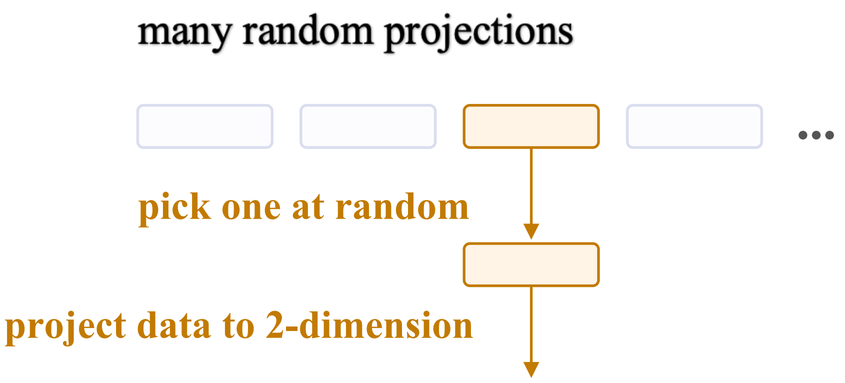


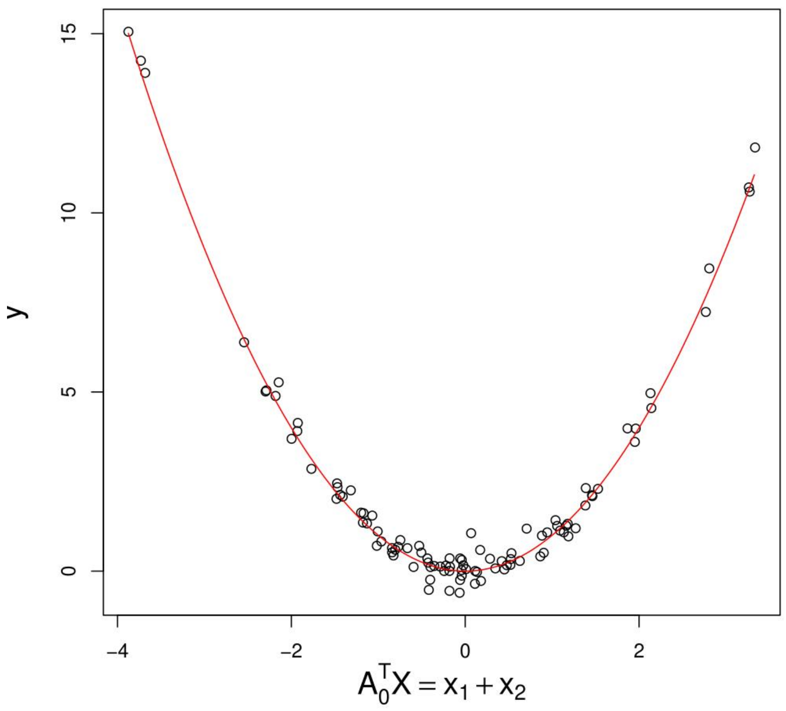
(b)

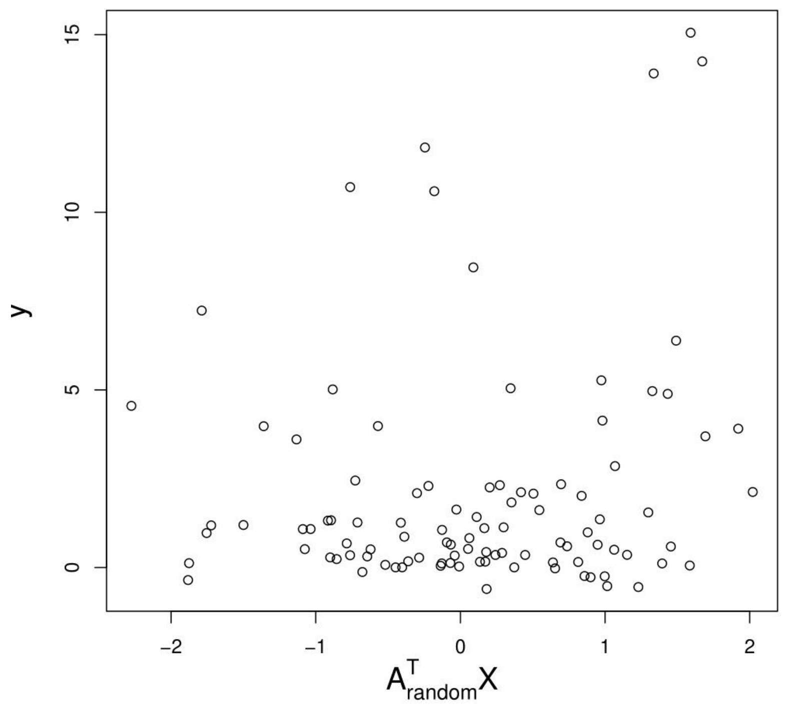
A single random projection can fail

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$





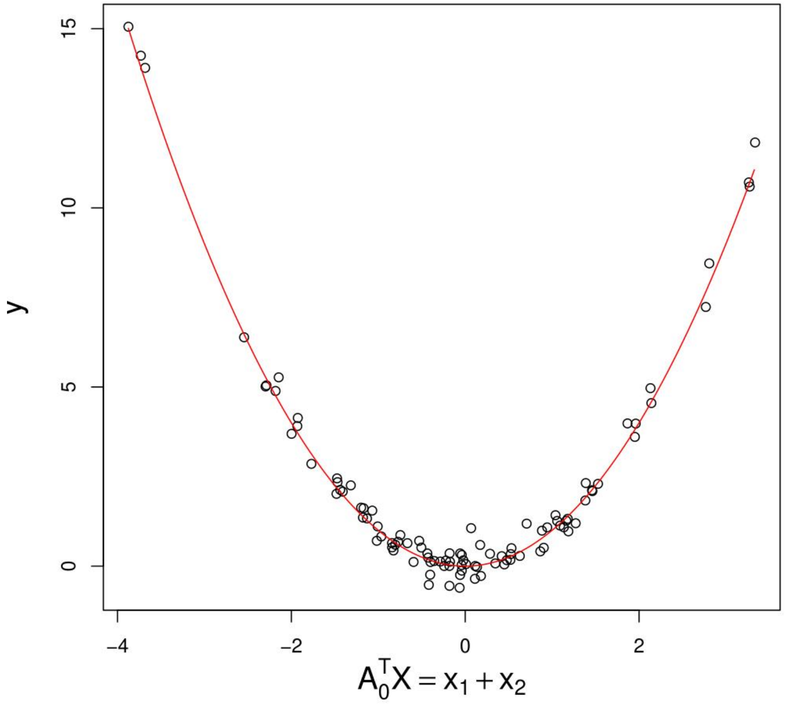


A random projection is easy to get, but usually not useful!

Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$



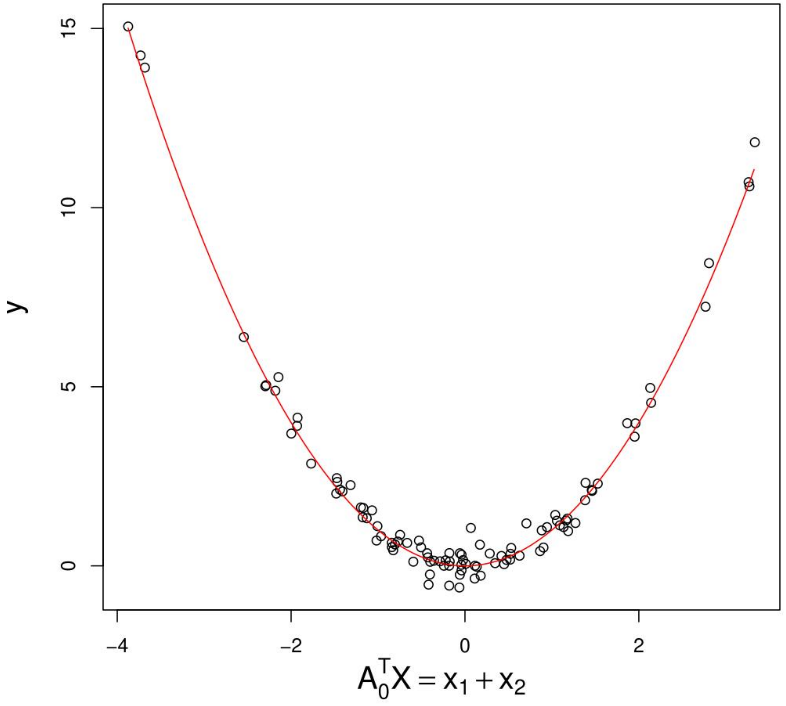
one group of candidate projections

- candidate 1
- candidate 2
- candidate 3
- candidate 4

Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$



one group of candidate projections

candidate 1 → project + fit → error = 0.42

candidate 2

candidate 3

candidate 4

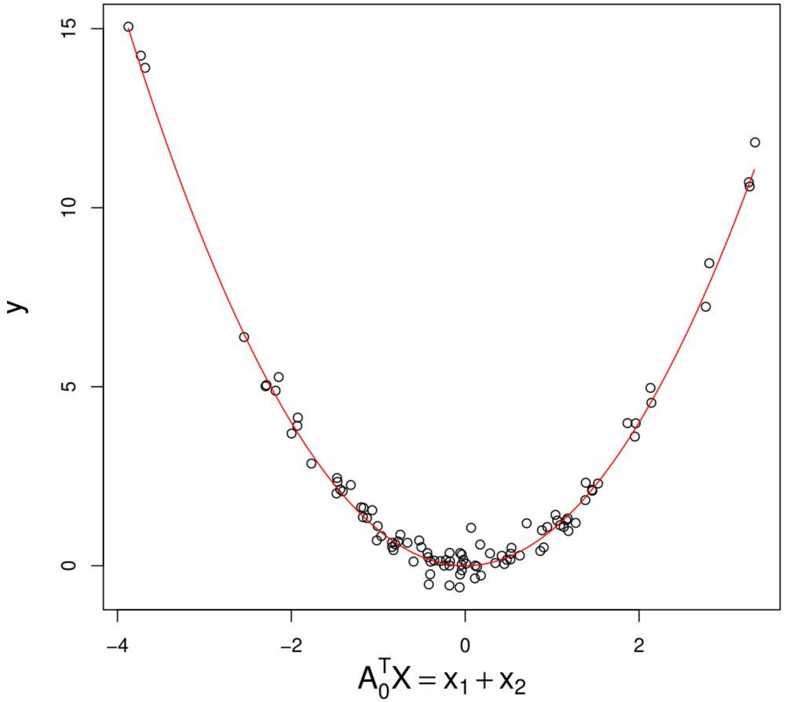
empirical error

0.42

Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$



one group of candidate projections

candidate 1

candidate 2 → **project + fit** → **error = 0.17**

candidate 3

candidate 4

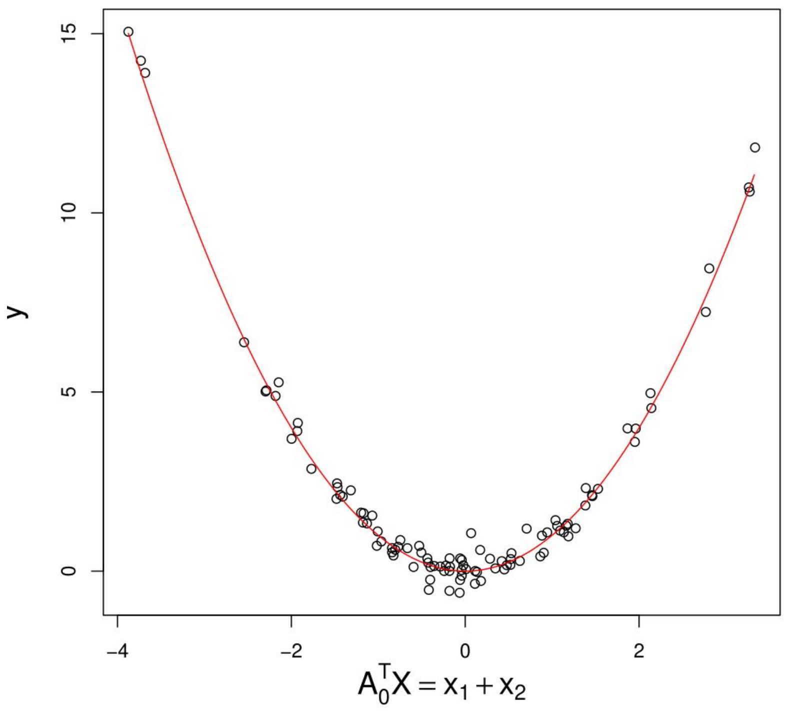
empirical error

0.42
0.17

Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$



one group of candidate projections

candidate 1

candidate 2

candidate 3 → **project + fit** → **error = 0.31**

candidate 4

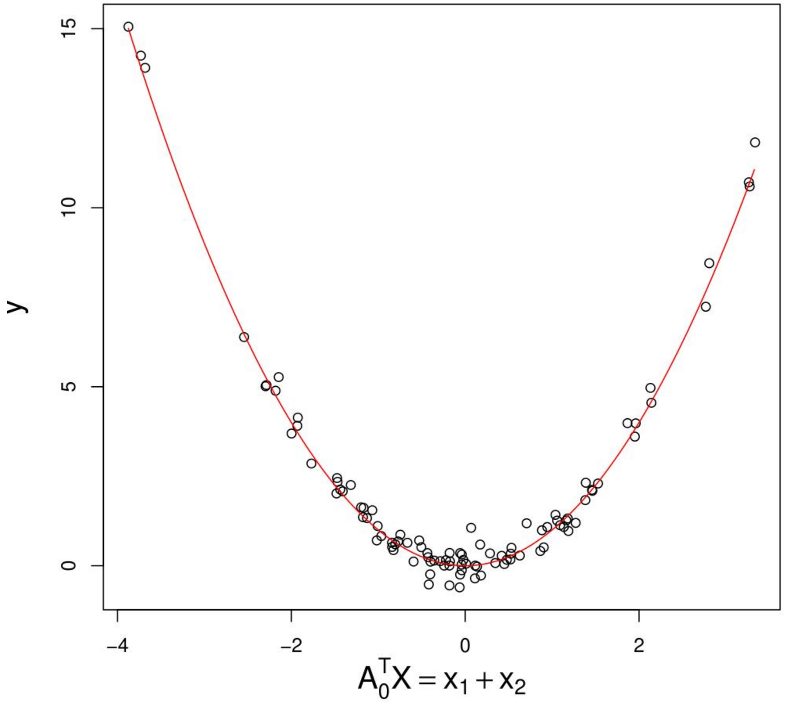
empirical error
0.42
0.17
0.31

Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$

oracle view

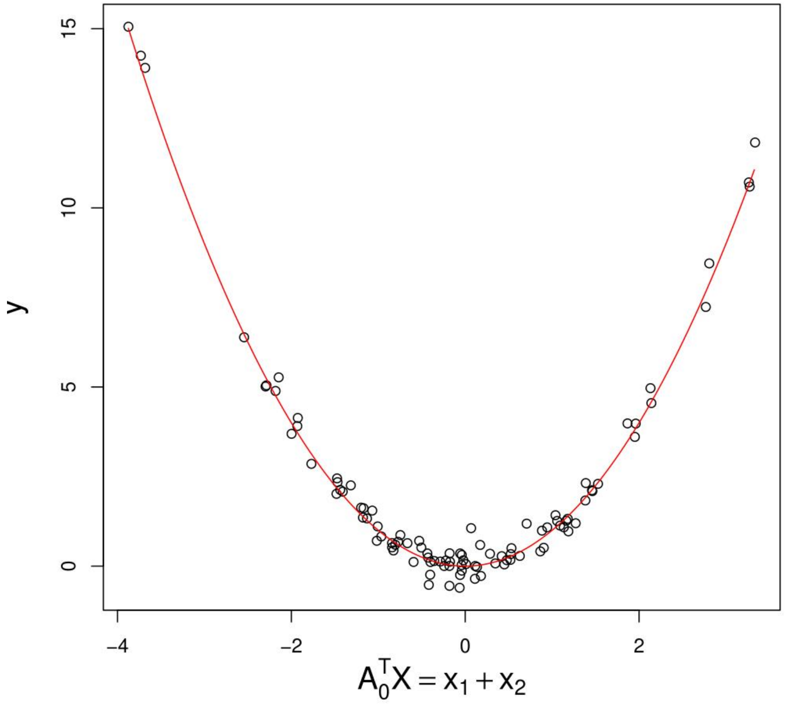


one group of candidate projections			empirical error
candidate 1			0.42
candidate 2			0.17
candidate 3			0.31
candidate 4 →	project + fit	error = 0.09	**0.09**

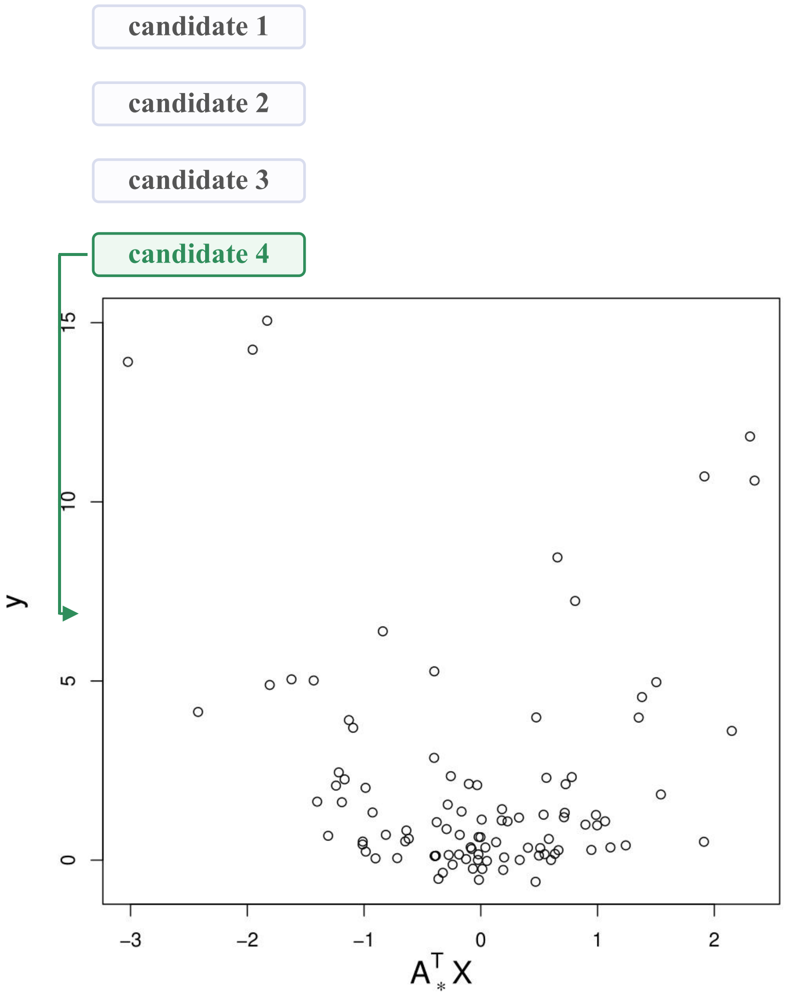
Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$



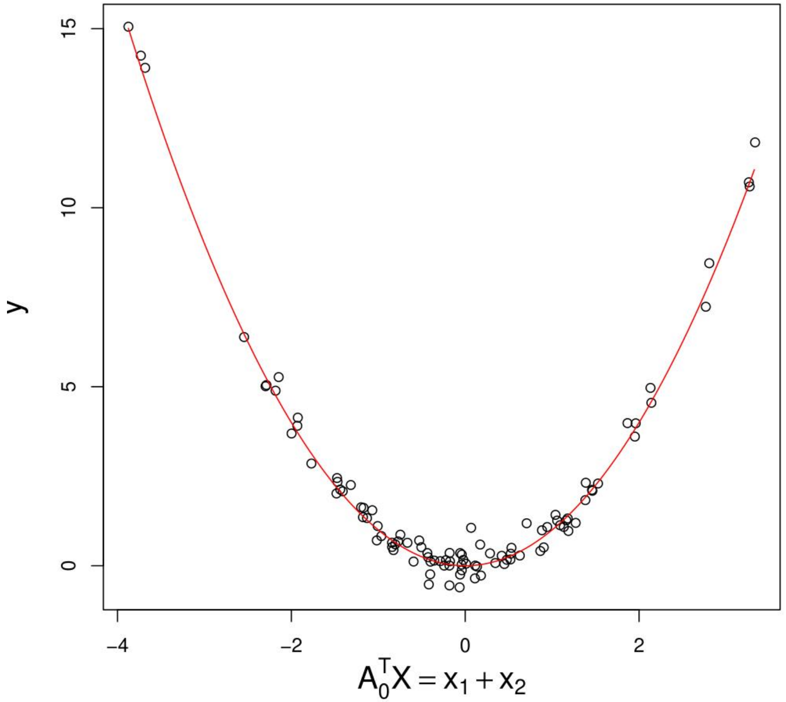
keep the smallest empirical error



Selecting the best projection in one group

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$



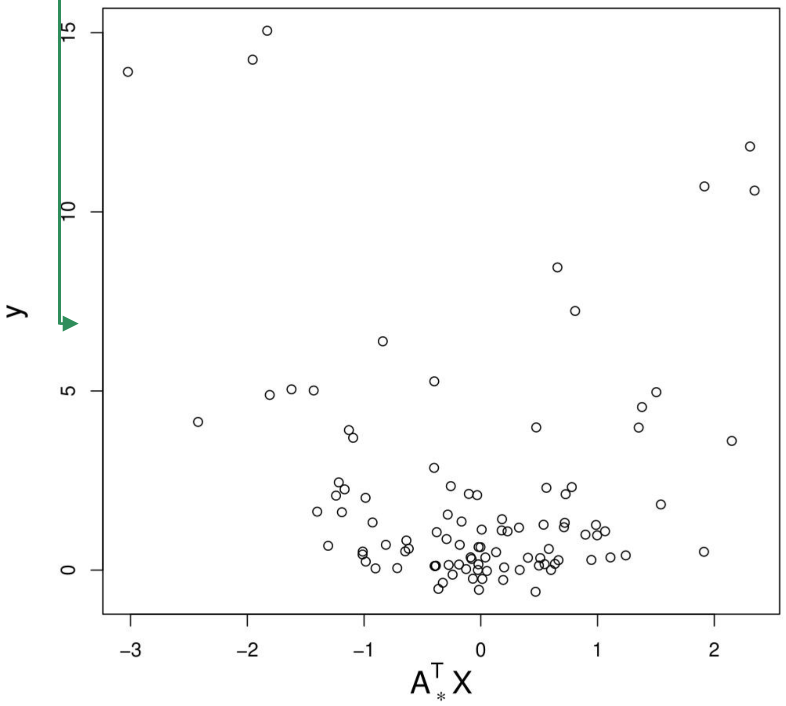
keep the smallest empirical error

candidate 1

candidate 2

candidate 3

candidate 4



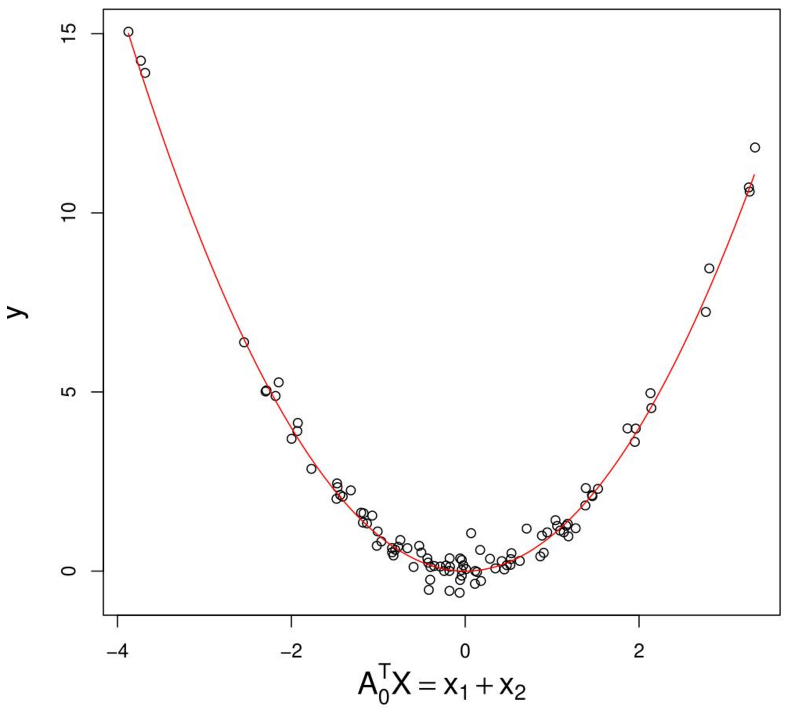
Choose the best projection in a group yields better representation.

Repeat the same selection across many groups

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$

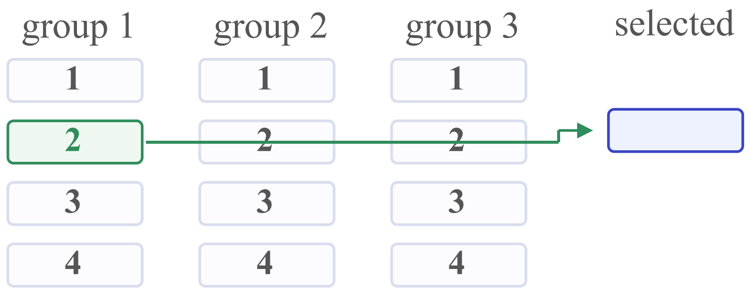
group 1	group 2	group 3
1	1	1
2	2	2
3	3	3
4	4	4

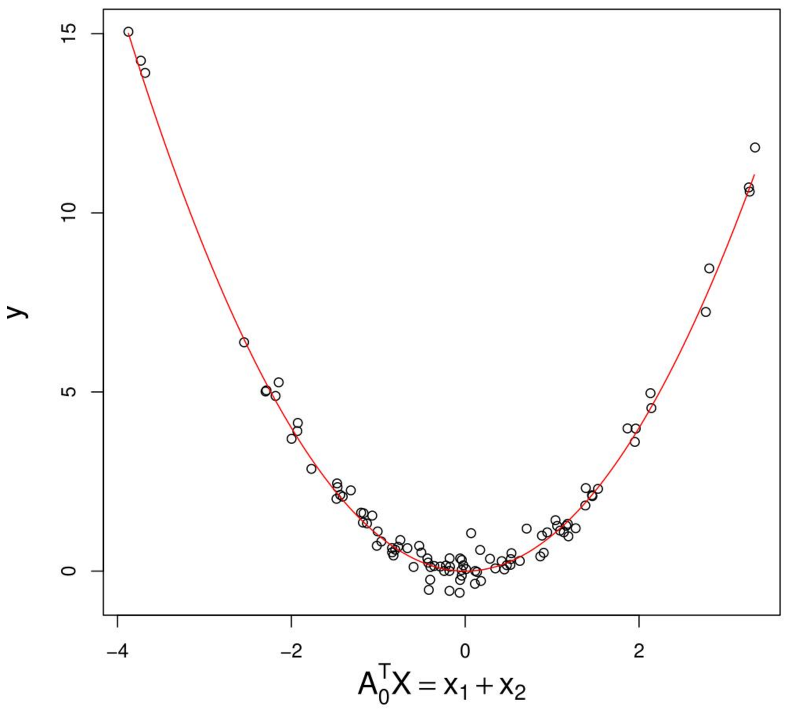


Repeat the same selection across many groups

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$

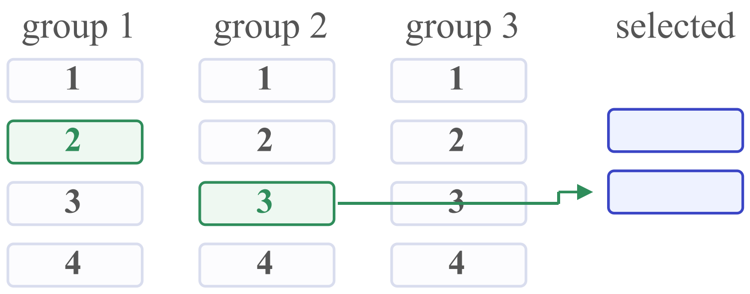


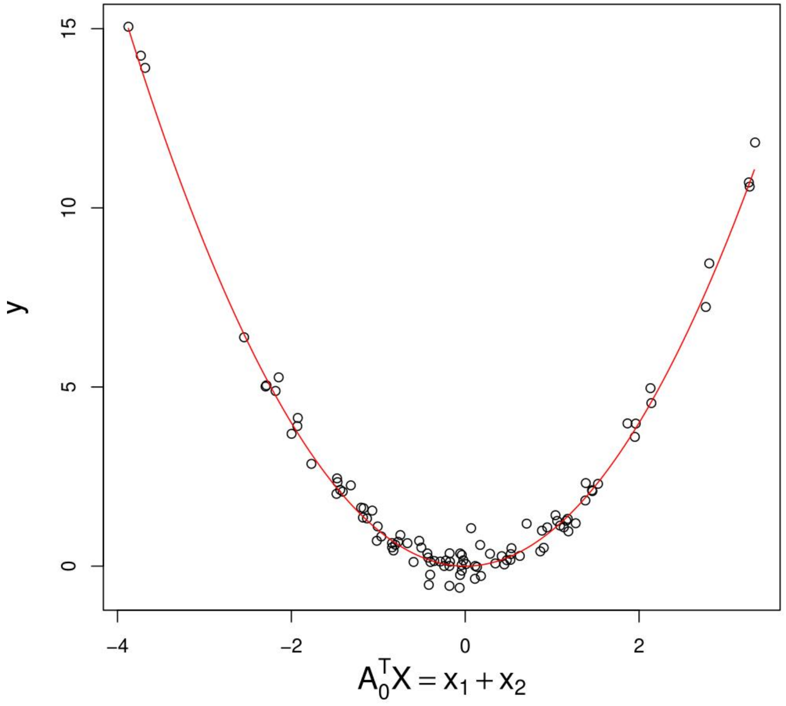


Repeat the same selection across many groups

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top/\sqrt{2}$

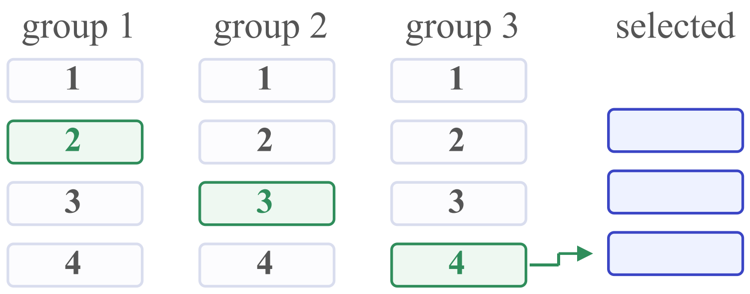


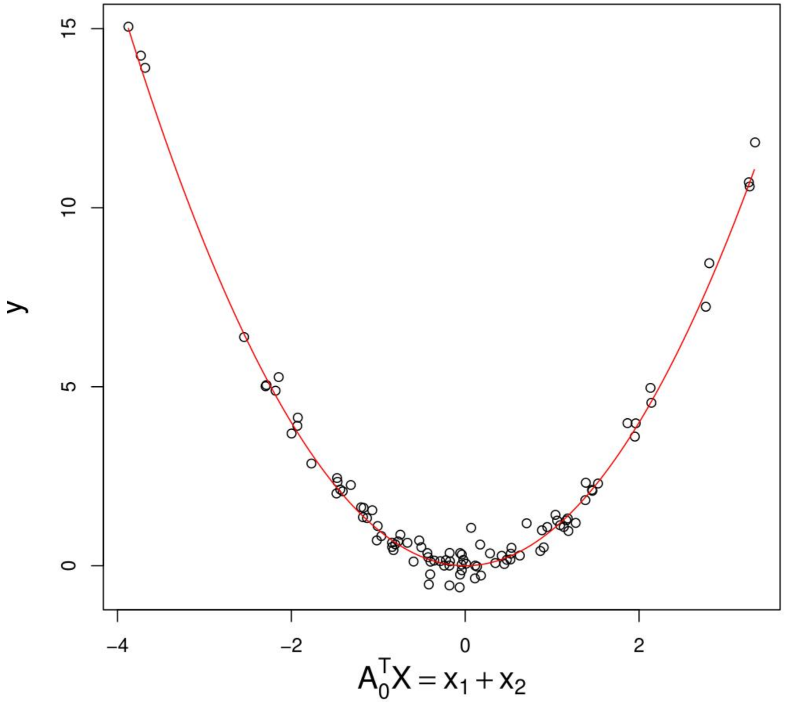


Repeat the same selection across many groups

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$

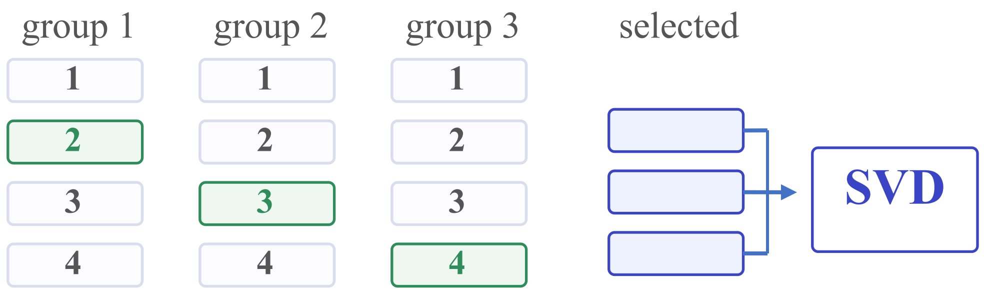


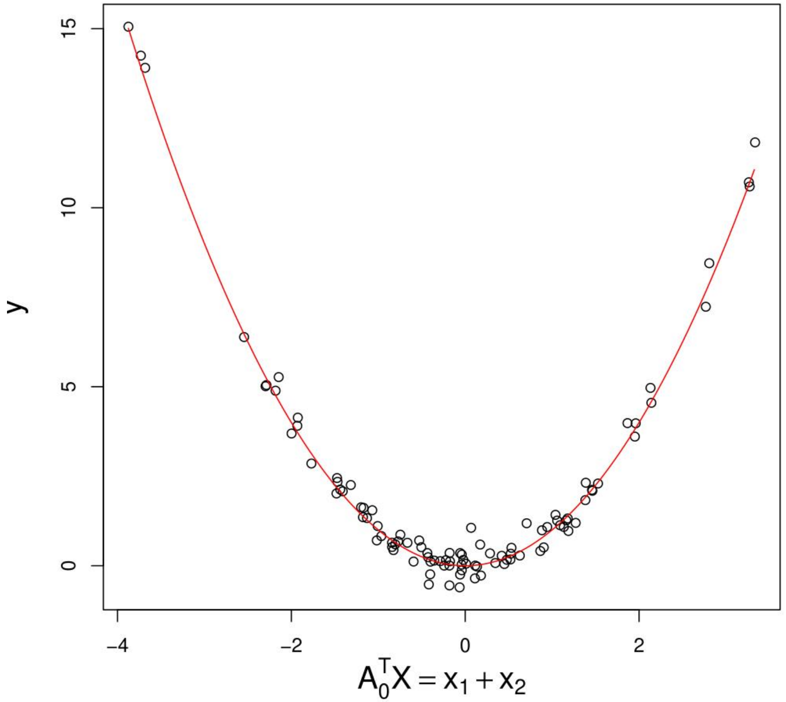


Aggregate the selected projections by SVD

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \ldots, 0)^\top / \sqrt{2}$

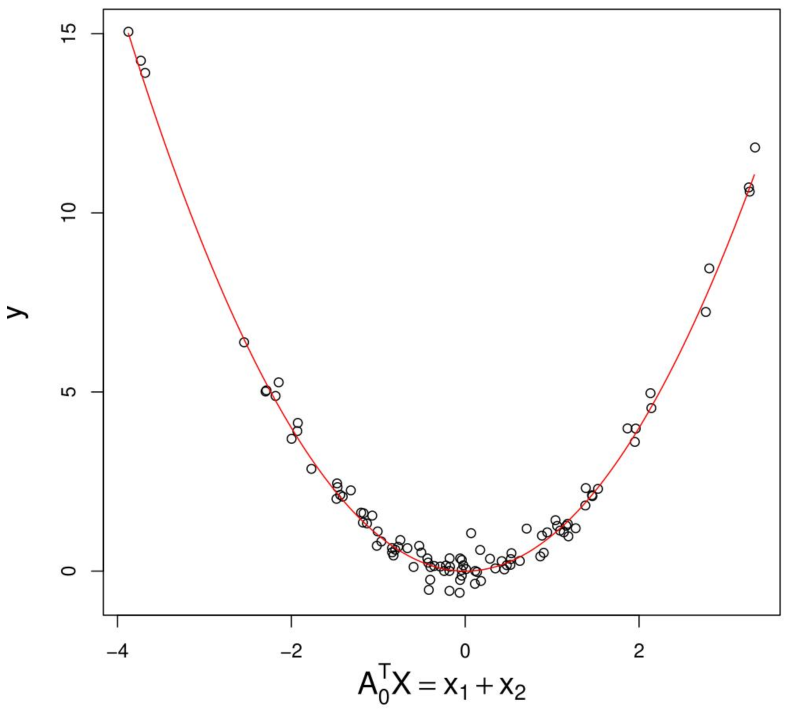


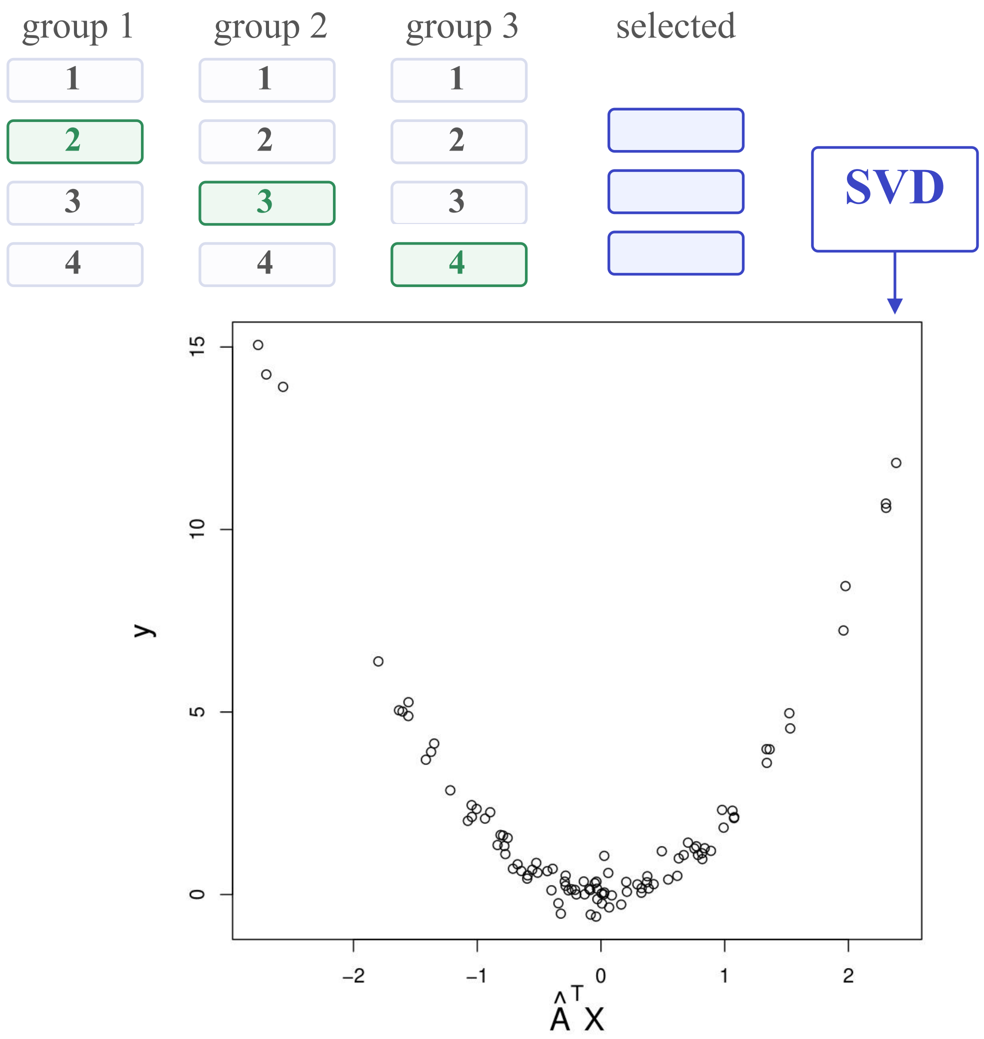


SVD recovers the signal directions

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \dots, 0)^\top/\sqrt{2}$

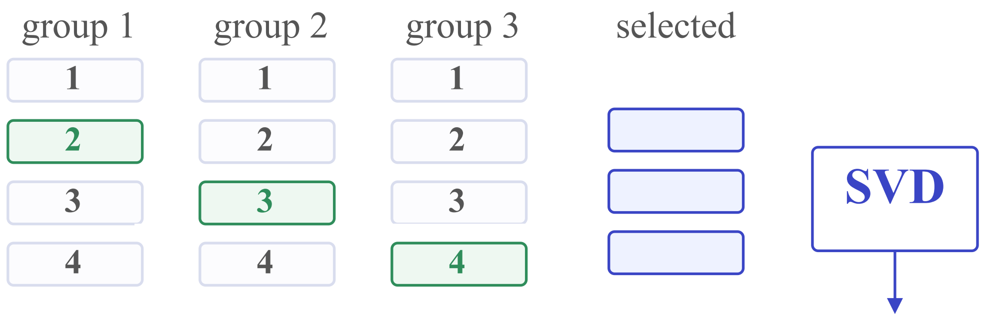


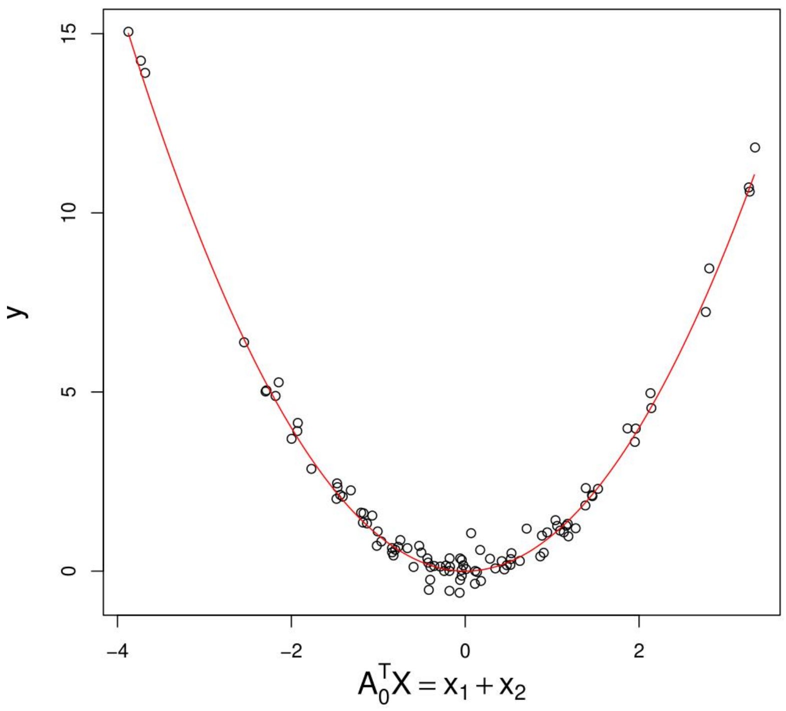


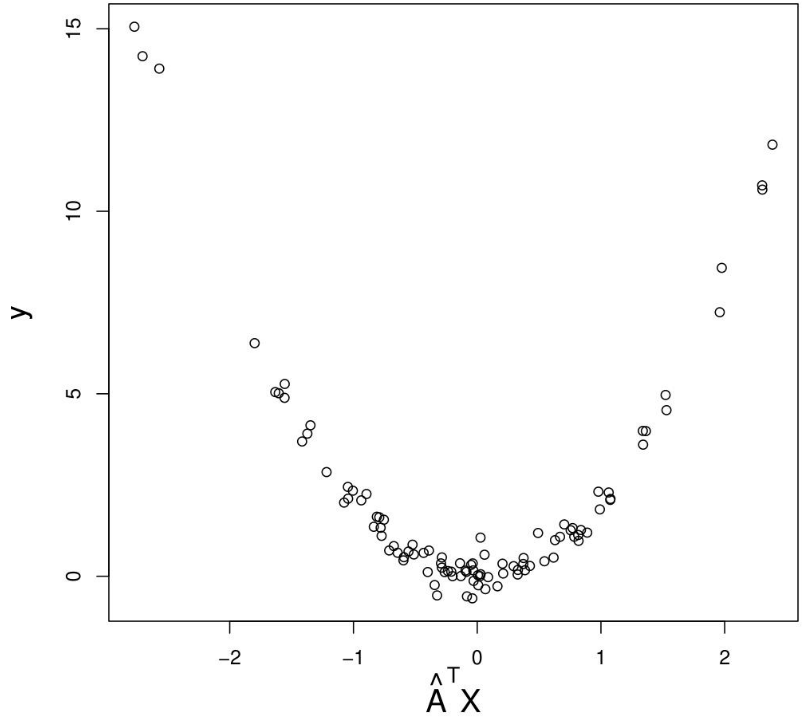
SVD recovers the signal directions

Toy example: $Y = (A_0^\top X)^2 + 0.3\varepsilon$

$n = 100, \quad p = 20, \quad X \sim N_{20}(0, I), \quad A_0 = (1, 1, 0, \dots, 0)^\top / \sqrt{2}$



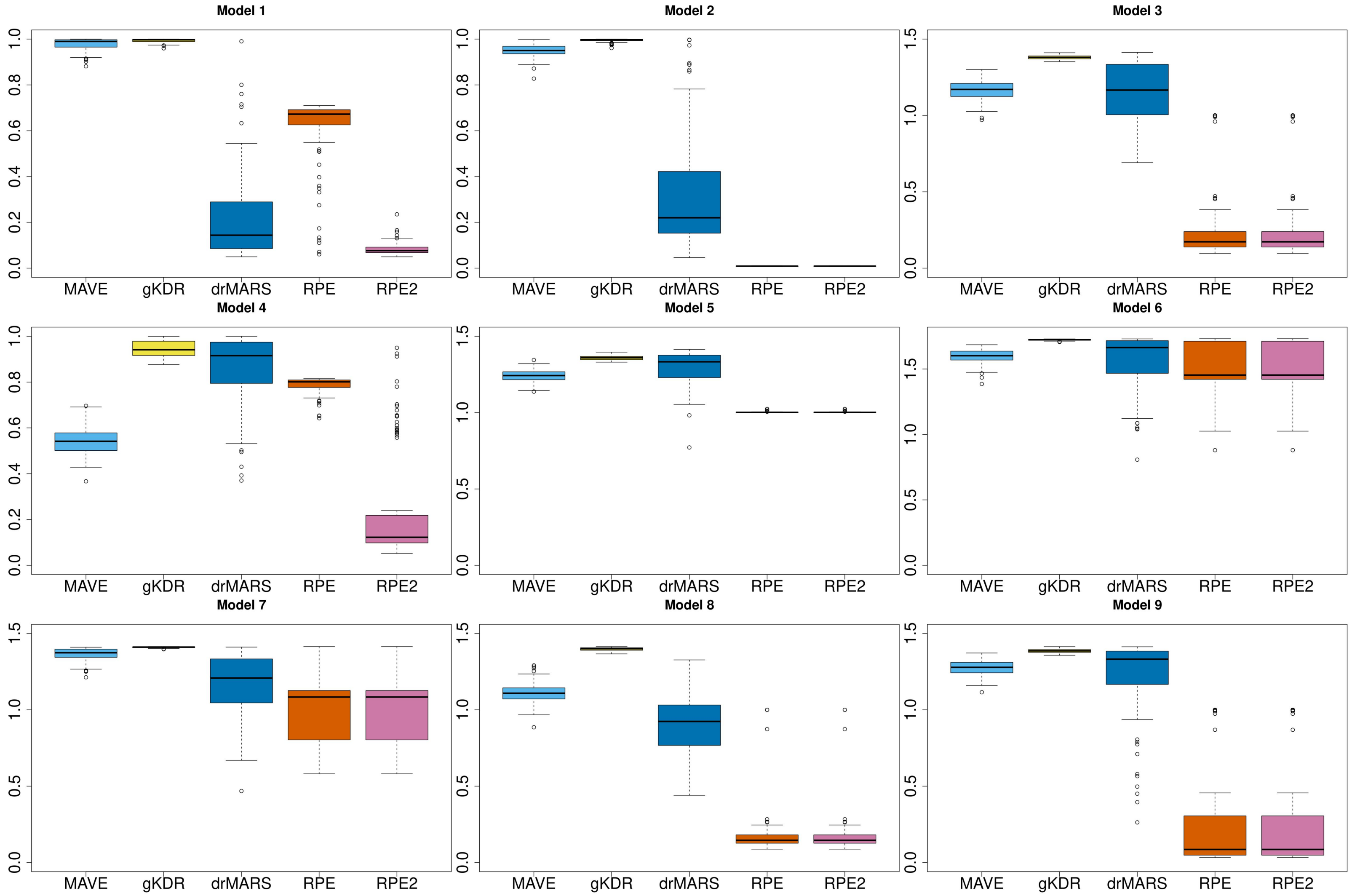




Using our method retains almost all the structure observed after oracle projection!

High-dimensional results: Figure 7

At n = 200 and p = 500, RPE-based methods are best overall; RPE2 gives an extra gain in Models 1a and 4.



Takeaway

- Single random projections are unreliable.

Takeaway

- Single random projections are unreliable.
- Selection makes randomness useful.

Takeaway

- Single random projections are unreliable.
- Selection makes randomness useful.
- Ensembling the selected projections recovers the signal directions.

Takeaway

- Single random projections are unreliable.
- Selection makes randomness useful.
- Ensembling the selected projections recovers the signal directions.
- More groups L help in theory; more results in the paper.

Main idea: randomness alone is weak;
selection + ensembling makes it informative.

Scan for the latest version