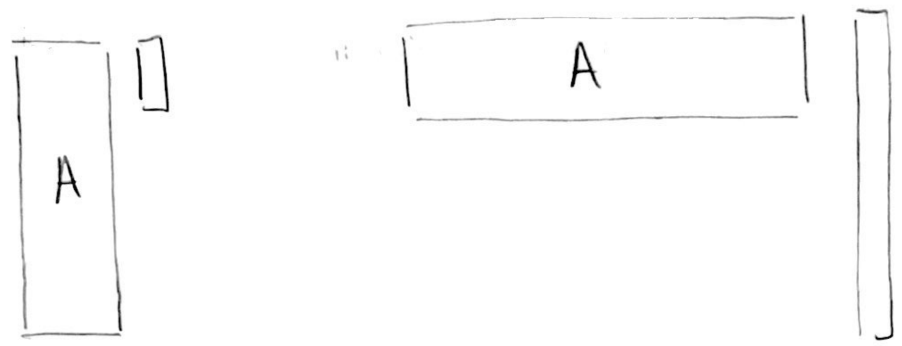
REGRESSION AND MODEL SELECTION

WE WANT TO SOLVE THE PROBLEM:

$$A x = b$$

BUT USUALLY, IN DATASCIENCE WE HAVE 'TALL-SKINNY' OR 'SHORT FAT MATRIX' $\underline{A}$ AND NOT SQUARE



OVERDETERMINED SYSTEM

↓

NO SOLUTION

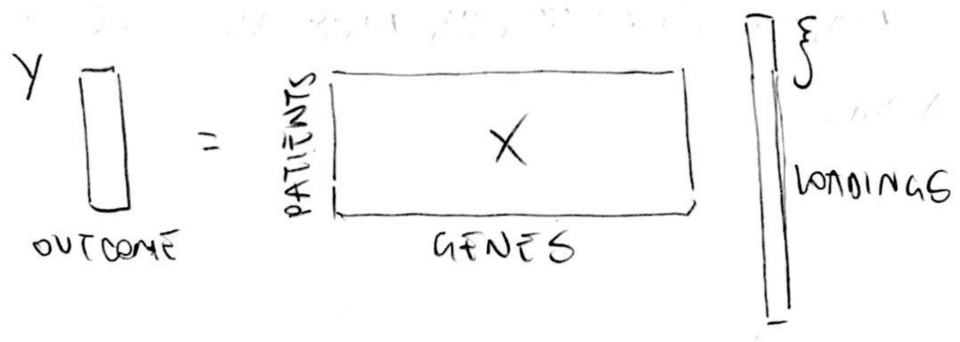
UNDERDETERMINED SOLUTION

↓

∞ - SOLUTIONS

BUT IF WE DO FOR EXAMPLE IN MATLAB $x = A \backslash b$ WE OBTAIN A SOLUTION IN BOTH CASES.

LET'S CONSIDER A BIOLOGICAL EXAMPLE.



WE WANT TO UNDERSTAND WHICH OF THESE GENES IS RESPONSIBLE FOR AN OUTCOME

$$Y = f(X, \xi)$$

WE CAN OBTAIN A SOLUTION IN WHICH ALL THE ξ_i ARE NON ZERO AND

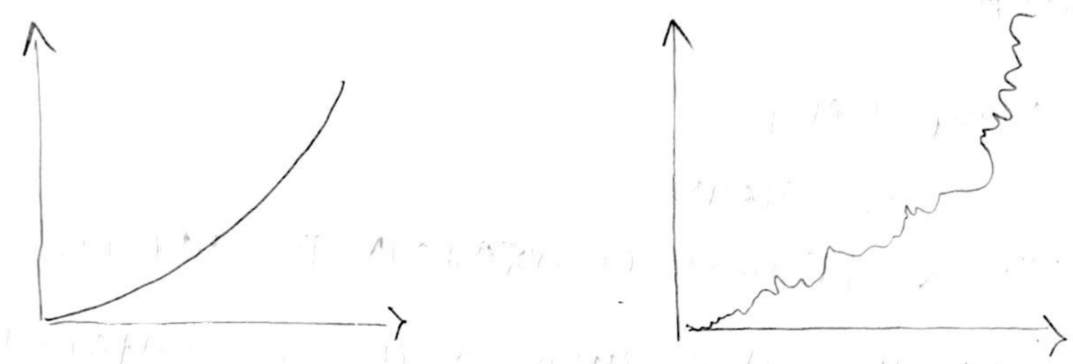
2.

SMALL AS POSSIBLE, FOR EXAMPLE WITH LEAST SQUARES, OR WE CAN TRY TO FIND A SPARSE SOLUTION IN WHICH JUST FEW OF THE ξ_i ARE NON ZERO

A SPARSE SOLUTION IS OBTAINED BY PENALIZING $\|\xi\|_1$, WHICH IS A PROXY FOR SPARSITY.

WE HAVE A FUNCTION $f(x) = x^2$ BUT WE MEASURE IT WITH SOME NOISE

$$y = x^2 + \mathcal{N}(0, \sigma)$$



WE CAN FOR EXAMPLE FIT A POLYNOMIAL TO THIS DATA:

$$f(x) = a_0 + a_1 x + \dots + a_{19} x^{19}$$

IN THIS CASE WE HAVE 20 UNKNOWNS: a_j $j = 0, \dots, 19$ THE LOADINGS AND WE BUILD THE MATRIX A AS:

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} | & | & & | \\ 1 & x & \dots & x^{19} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \\ \xi \\ \\ \end{bmatrix}$$

SPARSE SOLUTIONS, OBTAINED WITH LASSO REGRESSION ALLOW TO AUTONOMOUSLY DISCOVER WHICH ARE THE IMPORTANT FEATURES TO CONNECT INPUT AND OUTPUTS.

BRIEF RECAP:

$$\dot{x} = f(x) \qquad \text{DMD: } x = \sum b_k \psi_k e^{w_k t} \qquad \dot{x} = Ax$$

$$X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_m \\ | & & | \end{bmatrix} \qquad \text{KOOPMAN} \quad y = g(x) \qquad y = \sum b_k \psi_k e^{w_k t} \qquad \dot{y} = Ky$$

WE WILL NOW CONSIDER SINDY: "Sparse identification of nonlinear dynamics" WITH THE GOAL OF DISCOVER THE GOVERNING EQUATIONS FROM DATA.

WE WANT A MODEL THAT IS:

1. PARSIMONIOUS ← FEW TERMS
2. INTERPRETABLE ← WE WANT TO UNDERSTAND THE PHYSICS.

THE PROBLEM WITH DMD AND KOOPMAN IS THE GENERALIZATION, ACTUALLY THE MODEL IS NOT $\dot{x} = f(x)$ BUT $\dot{x} = f(x, \beta)$ AND IF THE PARAMATER CHANGE THE MODEL STOP WORKING

BY DISCOVERING THE ACTUAL PHYSICS WE CAN EXTRAPOLATE AND NOT JUST INTERPOLATE.

WE HAVE DATA AND WE WANT TO DISCOVER THE GOVERNING EQUATIONS $\dot{x} = ?$ WE HAVE A LOT OF EQUATIONS IN OUR BOOKS AND SO, WE WANT TO BUILD A LIBRARY OF MODELS Θ:

$$\Theta = \begin{bmatrix} | & | & | & & | & & | \\ 1 & x & x^2 & \dots & \cos x & \dots & e^x \\ | & | & | & & | & & | \end{bmatrix}$$

↑ POTENCIAL RIGHT HAND SIDE OF OUR PROBLEM

$$\dot{x} = \Theta \xi$$

PHYSICS IS PARSIMONIOUS AND WE WANT TO DISCOVERE IT BY PROMOTING SPARSITY.

WE CAN HAVE DIFFERENT ALGORITHMS:

- LEAST SQUARE REGRESSION: $\min \| Ax - b \|_2$
- LASSO: $\min \| Ax - b \| + \lambda_1 \| x \|_1$
- RIDGE: $\min \| Ax - b \| + \lambda_2 \| x \|_2$
- ELASTIC-NET: $\min \| Ax - b \| + \lambda_1 \| x \|_1 + \lambda_2 \| x \|_2$

▷ DISCREPANCY MODELING

THE REAL SYSTEM IS DESCRIBED BY:

$$\dot{x} = \underbrace{f(x)}_{\text{OBSERVED PART}} + \underbrace{g(x)}_{\text{UNOBSERVED/ UNMODELED PART}} = f(x)$$

AN EXAMPLE CAN BE THE EXPLANATION OF THE GRAVITY LAW: WE HAVE $F = mg$ WHERE WE SUPPOSE g TO BE COSTANT BUT IF WE LEFT TWO DIFFERENT BODY FALL DOWN THEY WILL HIT THE GROUND IN DIFFERENT TIMES. THIS HAPPEN BECAUSE WE HAVE ALSO ANOTHER FORCE, THE DRAG, WHICH IS REPRESENTED BY $g(x)$ IN THE EQUATION.

LET'S CONSIDER:

$$\dot{u} = u_{xx} + u^3 - u$$

WE BUILD A LIBRARY:

$$\Theta = \begin{bmatrix} \bar{u}_1 & & & \\ \bar{u}_2 & u_2^2 & u_x & \dots \\ \vdots & & & \\ \bar{u}_m & & & \end{bmatrix}$$

WE STACK IN A COLUMN THE TIME HISTORY OF THE VARIABLES