Discussion on the Causal Discovery

IA pour la Science

January 8th, 2026

$do(X_{t-p-2} = x_0)$

$$(X_{t-p+1}, X_{t-p-2}, \dots, X_{t-1}, X_t) \longrightarrow Y = X_{t+1}$$

$X_{t-p-2} = x_0$

$S^1 : (\qquad) \qquad (\quad)$

$S^2 : (\qquad) \qquad (\quad)$

$\vdots$

$S^n : (\qquad) \qquad (\quad)$

sample (training data)

model: $Y = \sigma(WX)$

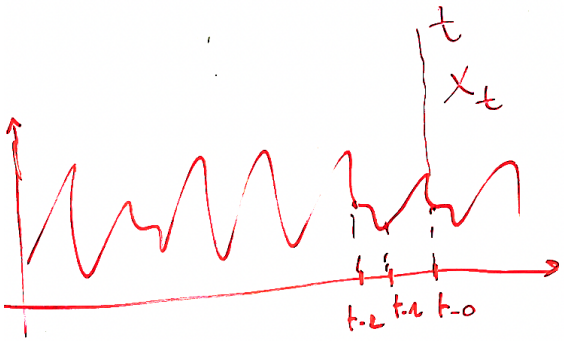
$$W = \begin{bmatrix} W_1^T \\ \\ W_k^T \end{bmatrix}, \quad W \in \mathbb{R}^{k \times n_x}, \quad W_i \in \mathbb{R}^n \ (1 \le i \le k)$$

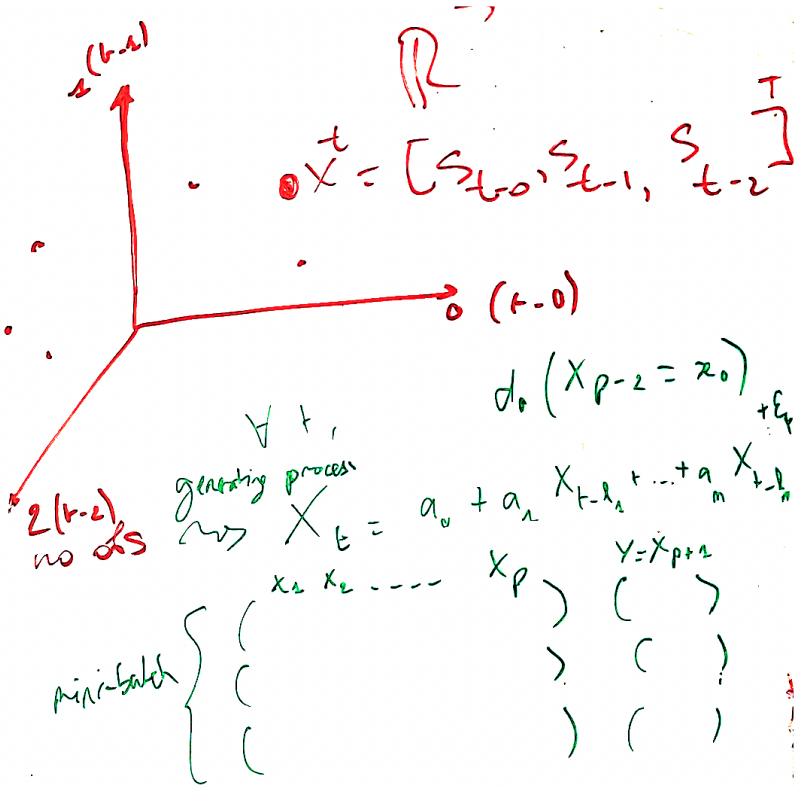
$$X \in \mathbb{R}^{n_x \times 1}$$

? $P(W \mid X)$

? $P(W \mid do(X = x_0))$

$$KL\left(P(W \mid X) \,\|\, P(W \mid do(X = x_0))\right)$$





$x^t = [s_{t-0}, s_{t-1}, s_{t-2}]^T$

$d_0(X_{p-2} = x_0) + \varepsilon_t$

$\forall t,$ generating process $\rightsquigarrow X_t = a_0 + a_1 X_{t-l_1} + \dots + a_n X_{t-l_n}$

$d_0(X_{t-2} = x_0)$ $\quad y = f(w^* x)$ $\quad \hat{y} = f(w, x)$ $\quad$ model

$$MB^{(1)} \left\{ \begin{array}{cccc} x & & y & \hat{y} \\ (x_1^{(1)}, \cancel{x_2^{(1)}}^{x_0}, \dots, x_p^{(1)}) & (x_{p+1}^{(1)}) & (\quad) \\ (x_1^{(2)}, \cancel{x_2^{(2)}}^{x_0}, \dots, x_p^{(2)}) & (x_{p+1}^{(2)}) & (\quad) \\ \vdots & & \\ (x_1^{(d)}, \cancel{x_2^{(d)}}^{x_0}, \dots, x_p^{(d)}) & (x_{p+1}^{(d)}) & (\quad) \end{array} \right.$$

$|y - \hat{y}|$

$MB^{(1)}$

$$\text{mini-batch} \left\{ \begin{array}{cc} x_1 \; x_2 \; \dots \; x_p & y = x_{p+1} \\ (\qquad\qquad) & (\quad) \\ (\qquad\qquad) & (\quad) \\ (\qquad\qquad) & (\quad) \end{array} \right.$$

$$w^* = \underset{w}{\operatorname{argmax}} \; P(w \mid y_{1:p}, X(d_0(X_{t-2} = x_0)))$$

1. Sample $x \to s_1, \dots, s_n$

2. Optimize $\mathcal{L}(w) \mapsto p_{MAP}(w \mid x$

3. $\hat{w} = \arg\max \; (\dots$

4. $[\hat{w}_1, \dots, \hat{w}_n]^T = \hat{W}$

mini batch

$p(y \mid x$

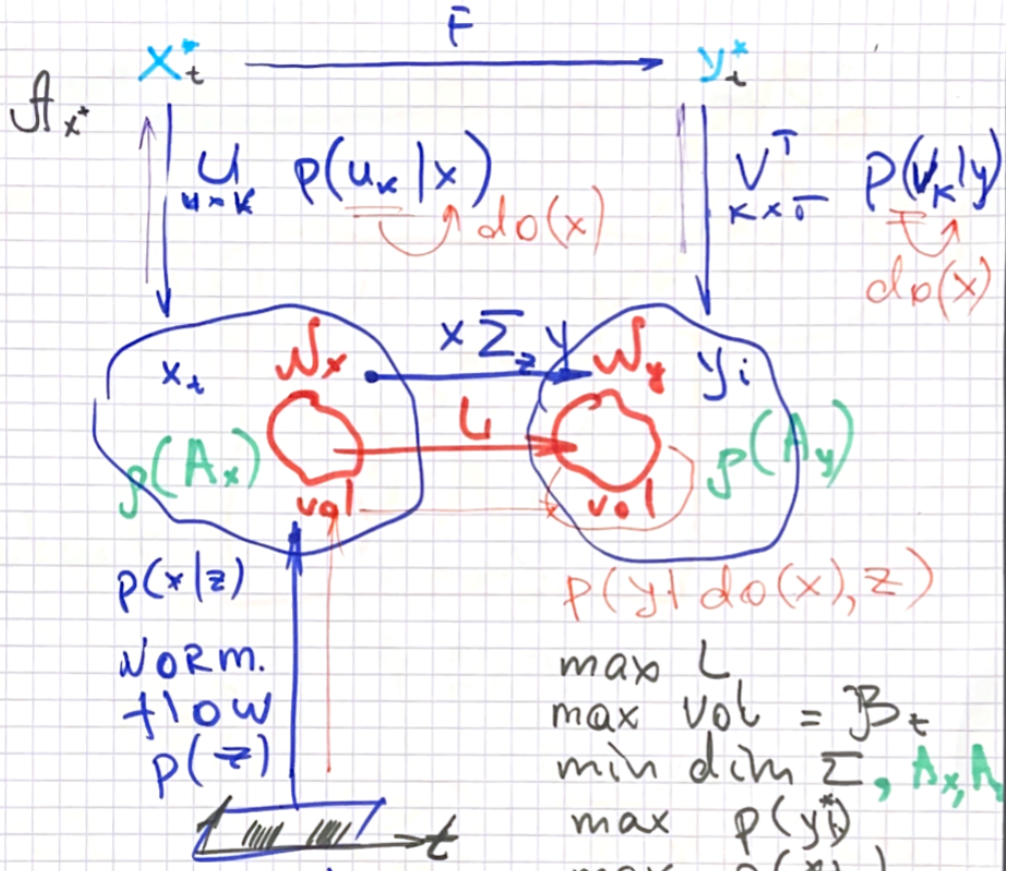
5. $E(w) = \frac{1}{n} \sum \hat{w}_i$

$\hat{A}(w) = \frac{1}{n-1} \hat{W}^T \hat{W}$

PSD

$\to p(w \mid x, y) =$

obs, given
to infer
time, space

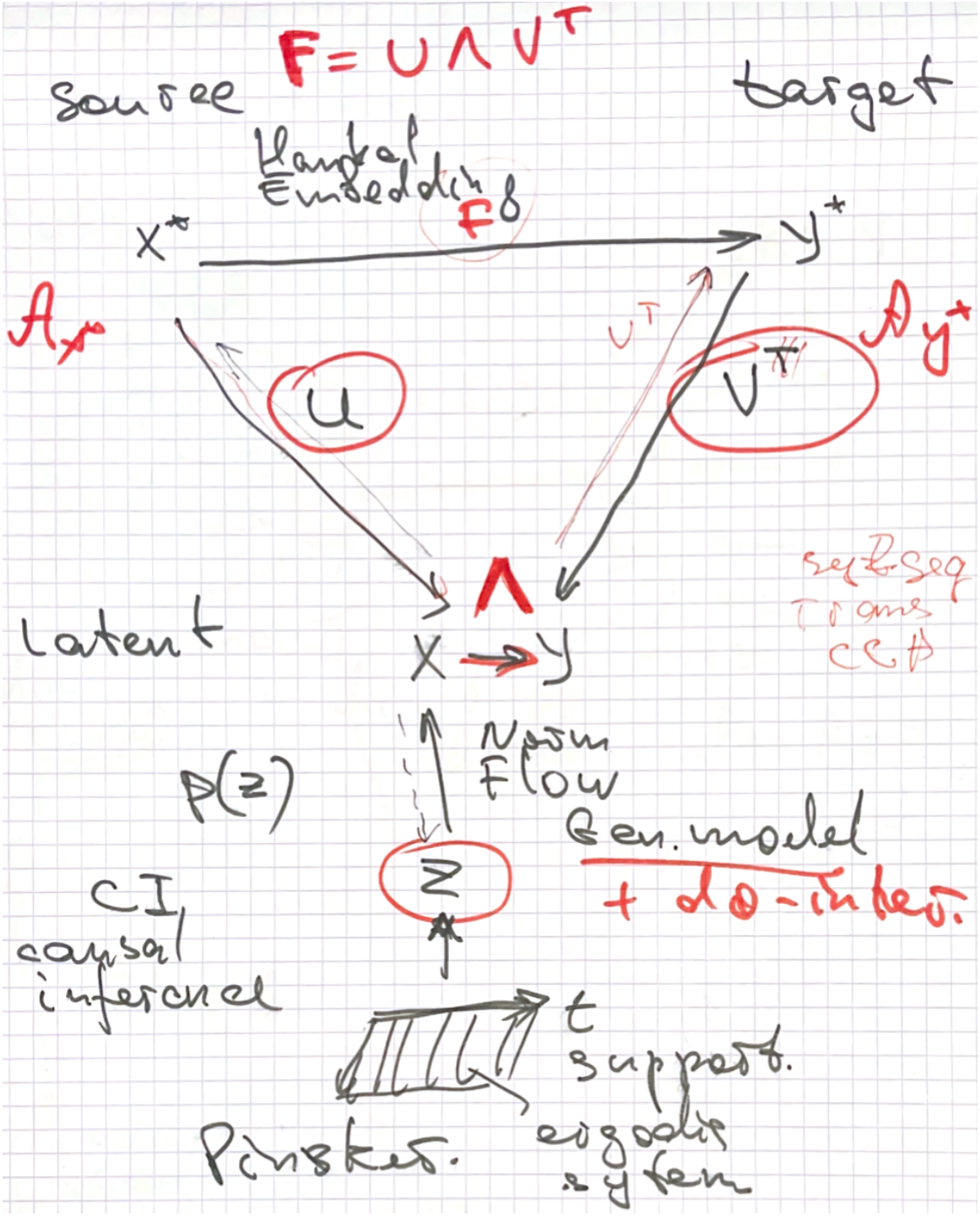


max L
max vol $= \mathcal{B}_t$
min dim Σ, A_x, A_y
max $p(y_i^*)$
~~max $p(x_i..)$~~

support
ergodic, low dim
space time

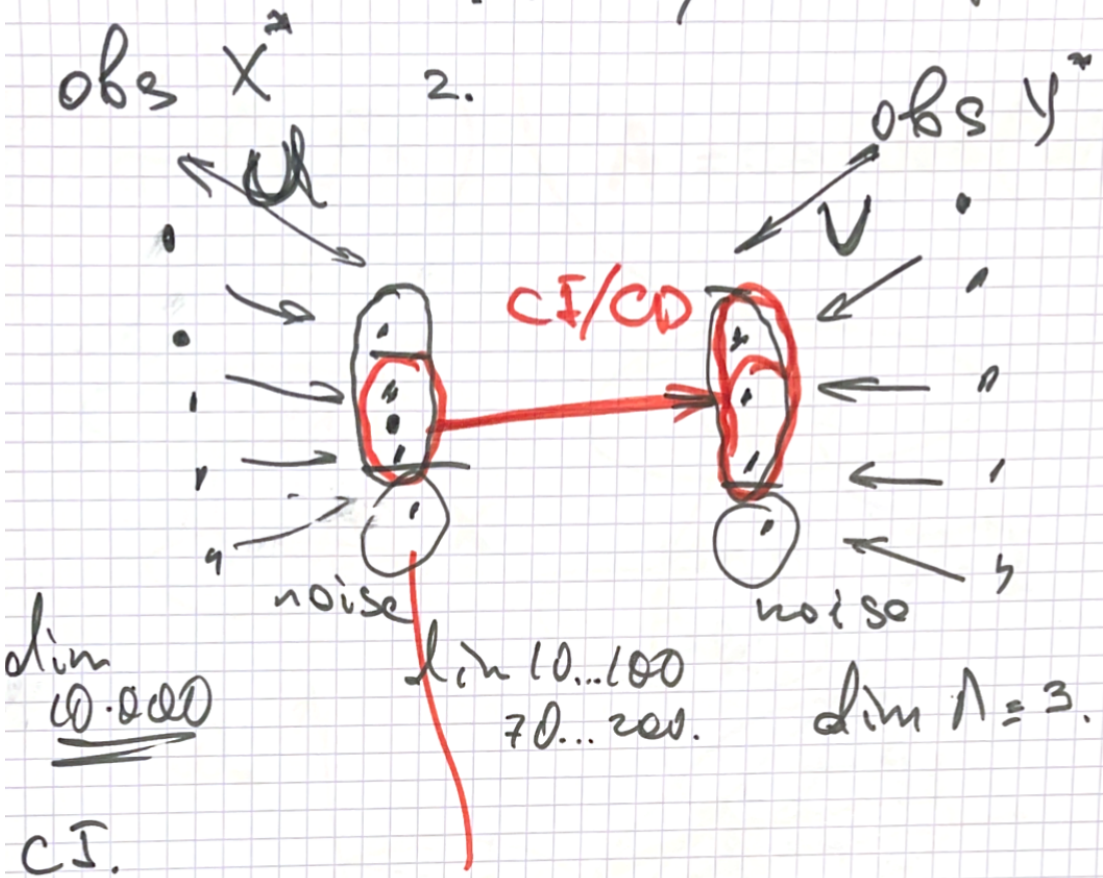
$\mathcal{A}_{x^*} = \bigcup \{$ Reconstructed, caused to y, noise, other. $\}$
indexes

1. Reconstruct x^*, y^*
2. Set initial π
3. Max Vol by $do(x) \leftarrow z$
4. Connect spaces.


F = U Λ V^T
source
target
Kernel Embedding
F
X*
Y*
A_X*
A_Y*
U
V^T
U^T
Λ
Latent
X → Y
seq-to-seq
Trans
CCA
Norm Flow
p(z)
Gen. model
+ do-interv.
Z
CI,
causal
inference
t
support.
ergodic
system
Pinsker.

1. Trans/CCA param.

2.



dim 10.000

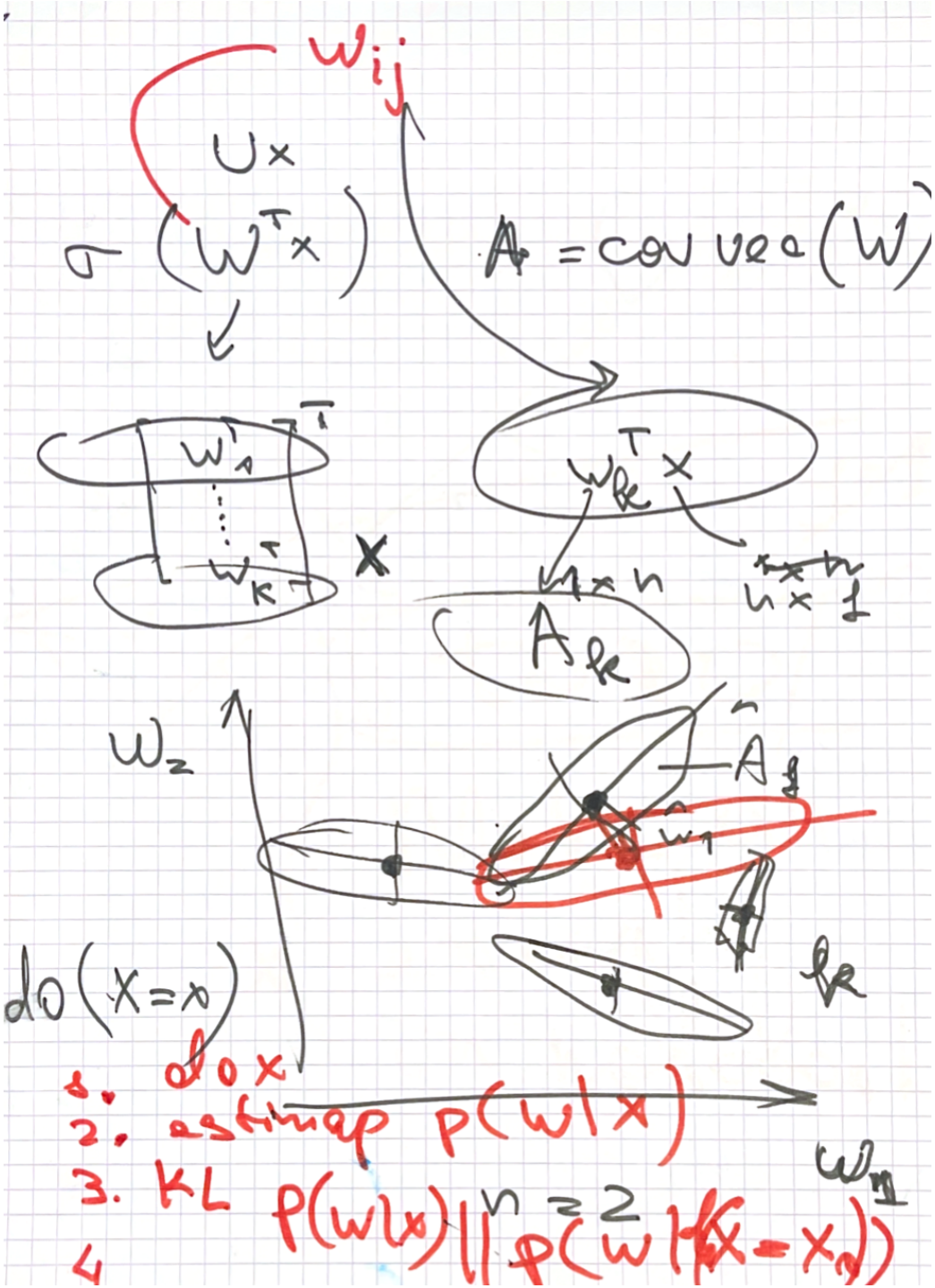
dim 10...100
70...200.

dim Λ = 3.

CI.

Step 1.

do calculus
~~LinFAM~~,
causal discovery CD


w_{ij}
Ux
$\sigma(W^T x)$
$A = \text{convex}(W)$
w_1^T
w_k^T
X
$w_k^T x$
$n \times n$
$n \times 1$
A_k
w_2
A_1
w_1
k
$do(X=x)$
1. do x
2. estimate $p(w|x)$
3. KL $p(w|x) \| p(w|X=x_0)$
4
w_n
$n=2$