Balancing Covariates in Randomized Experiments using the Gram-Schmidt Walk

Peng Zhang

joint with Chris Harshaw, Fredrik Sävje, Dan Spielman

Yale University

STOC 2020 Workshop: Recent advances in Discrepancy and Applications

Balancing Covariates in Randomized Experiments using the Gram-Schmidt Walk ★

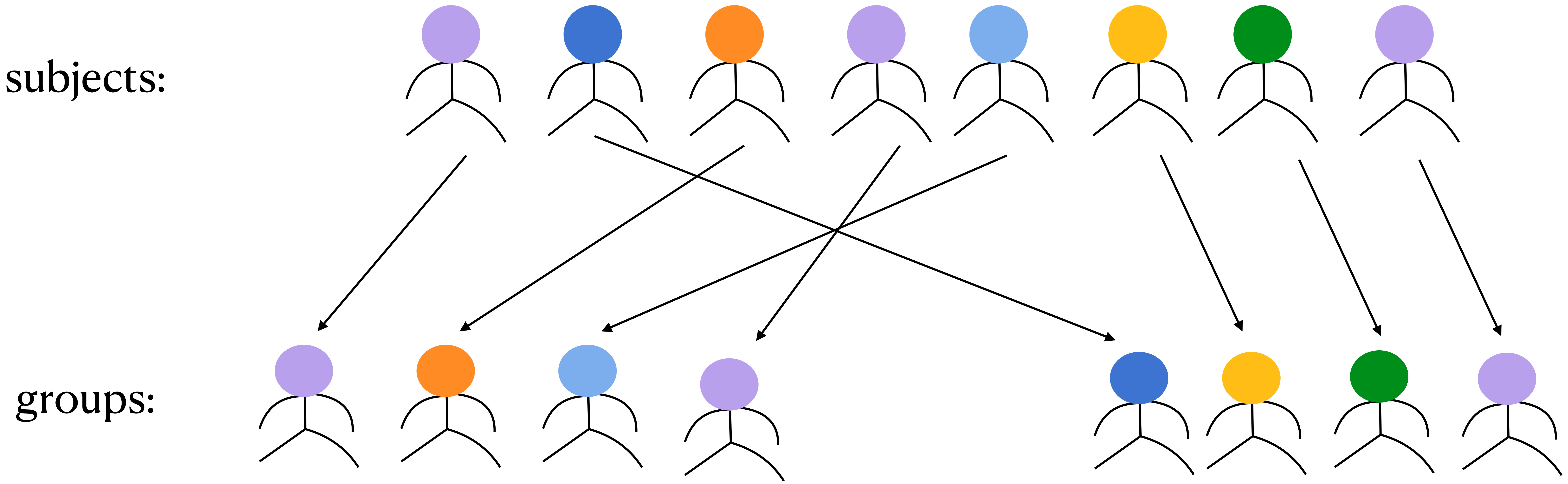
Peng Zhang

joint with Chris Harshaw, Fredrik Sävje, Dan Spielman

Yale University

★ Bansal, Dadush, Garg, and Lovett, 2018

Discrepancy theory & Randomized experiments



Discrepancy theory usually outputs deterministic assignments..

Want **random** assignments! ⟵ the Gram-Schmidt walk [BDGL'18] does this

The most valuable resource is subjects..

Want small **constants**!

Neyman-Rubin Causal Model

n subjects -> treatment or control

For each subject i, there are two **potential outcomes**:

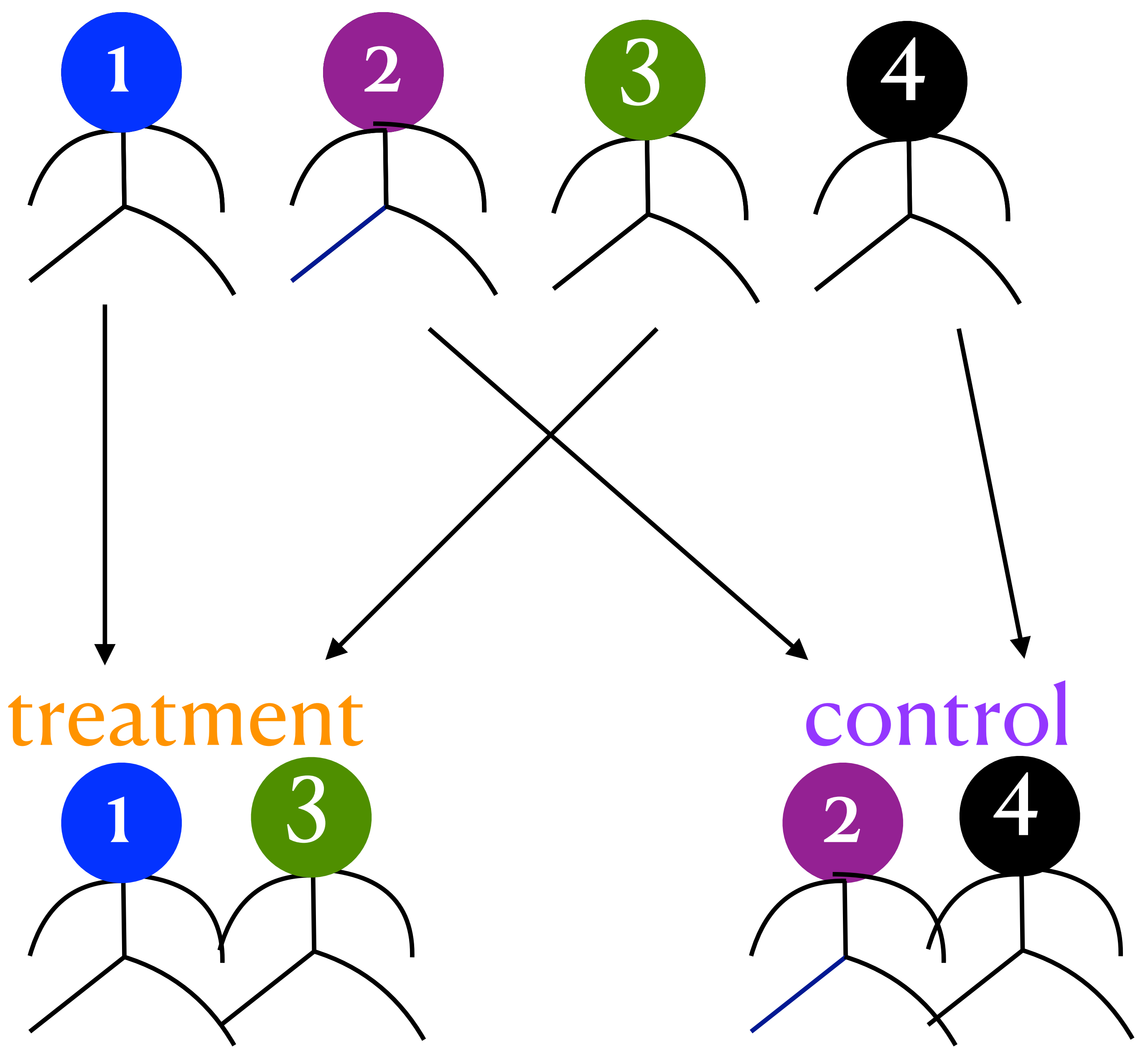
a_i, if subject i in treatment

b_i, if subject i in control

a_i, b_i are fixed and not affected by assignments

causal effect: $a_i - b_i$

Can only observe one of a_i, b_i



Never know subject-level causal effect

Estimate the average

unit	treatment	control	Casual effect
1	a1	??	??
2	??	b2	??
3	a3	??	??
4	??	b4	??

Neyman-Rubin Causal Model

n subjects

Want to output random assignments: $\mathbf{z} = (z_1, \ldots, z_n), z_i \in \{-1, 1\}$ The design is the distribution of $\mathbf{z}$

This talk assumes $\Pr(z_i = 1) = \Pr(z_i = -1) = 1/2$

observe a_i if $z_i = 1$; observe b_i if $z_i = -1$

Potential outcomes: $\mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n)$

a_i, b_i are fixed

The only randomness is $\mathbf{z}$

The average treatment effect:

$$\tau = \frac{1}{n} \sum_{i=1}^{n} (a_i - b_i)$$

The Horvitz-Thompson estimator (HT'52):

$$\hat{\tau} = \frac{2}{n}\Big(\sum_{i:z_i=1} a_i - \sum_{i:z_i=-1} b_i \Big)$$

Accuracy of the HT estimator

Error $\hat{\tau} - \tau = \frac{1}{n}\mu^\top \mathbf{z}$,

where $\mu = \mathbf{a} + \mathbf{b}$: the potential outcomes, $\mathbf{z}$: random assignments

Don't know μ

Unbiased estimator: $\mathbb{E}[\hat{\tau}] = \tau$

Variance: $\mathrm{Var}(\hat{\tau}) = \mathbb{E}[(\hat{\tau} - \tau)^2] = \frac{1}{n^2}\mu^\top \mathbb{E}[\mathbf{z}\mathbf{z}^\top]\mu$

Tails: want σ-subgaussian:

$\Pr(|\hat{\tau} - \tau| > \lambda n) < 2\exp(-\frac{\lambda^2}{2\sigma^2}), \forall \lambda$

Want small constant σ —> narrow confidence interval

Covariates

subject i -> covariates $\mathbf{x}_i \in \mathbb{R}^d$ (Euclidean norm 1). Write $\mathbf{X} = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n)$

Covariates can help

if $\mu = \mathbf{a} + \mathbf{b}$ is close to row space of $\mathbf{X}$

Ideally, $\mu = \mathbf{X}^\top \beta$

$$n^2 \operatorname{Var}(\hat{\tau}) = \mu^\top \mathbb{E}[\mathbf{z}\mathbf{z}^\top]\mu \leq \|\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\| \, \|\mu\|^2$$

$$= \beta^\top \mathbb{E}[(\mathbf{X}\mathbf{z})(\mathbf{X}\mathbf{z})^\top] \, \beta \leq \|\mathbb{E}[(\mathbf{X}\mathbf{z})(\mathbf{X}\mathbf{z})^\top]\| \, \|\beta\|^2$$

Helps if $\|\beta\| \ll \|\mu\| = \|\mathbf{X}^\top \beta\|$:

e.g., $d \ll n$, $\mathbf{X}^\top$ tall matrix with orthogonal columns and unit rows

The Gram-Schmidt Walk [BDGL'18]

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ with norm ≤ 1.

Output: random $\mathbf{z} \in \{-1,1\}^n$ s.t. $\mathbf{Xz}$ is σ-subgaussian, where $\sigma = \sqrt{40} \approx 6.3$

1. $\|\mathbb{E}[(\mathbf{Xz})(\mathbf{Xz})^\top]\| \leq \sigma^2$
2. $\Pr(|\theta^\top \mathbf{Xz}| > \lambda) \leq 2\exp(-\frac{\lambda^2}{2\sigma^2}), \forall \theta$ with norm $1, \lambda$

GSW provides a poly-time alg + [Dadush, Garg, Lovett, and Nikolov'16] —> Banaszczyk'97

O(1)-subgaussian is sufficient

Improving the analysis of GSW

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ with norm ≤ 1.

Output: random $\mathbf{z} \in \{-1,1\}^n$ s.t. $\mathbf{X}\mathbf{z}$ is a 1-subgaussian

$\sigma = 1$ is tight!

Less # of subjects!

1. $\|\mathbb{E}[(\mathbf{X}\mathbf{z})(\mathbf{X}\mathbf{z})^\top]\| \leq 1$
2. $\Pr(|\theta^\top \mathbf{X}\mathbf{z}| > \lambda) \leq 2\exp(-\frac{\lambda^2}{2}), \forall \theta$ with norm $1, \lambda$

Covariates in general case

subject i -> covariates $\mathbf{x}_i \in \mathbb{R}^d$ (Euclidean norm 1). Write $\mathbf{X} = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n)$

In general, $\mu = \mathbf{X}^\top \beta + \epsilon$

w. $\beta = \arg\min_{\beta} \|\mu - \mathbf{X}^\top \beta\|$, ϵ orth. to the row space of $\mathbf{X}$

$$n^2 \operatorname{Var}(\hat{\tau}) = \beta^\top \mathbf{X}\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\mathbf{X}^\top \beta + \epsilon^\top \mathbb{E}[\mathbf{z}\mathbf{z}^\top]\epsilon + 2\beta^\top \mathbf{X}\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\epsilon$$

$$\leq \|\mathbb{E}[(\mathbf{X}\mathbf{z})(\mathbf{X}\mathbf{z})^\top]\| \|\beta\|^2 + \|\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\| \|\epsilon\|^2 + \cdots$$

Want small $\|\mathbb{E}[(\mathbf{X}\mathbf{z})(\mathbf{X}\mathbf{z})^\top]\|$ && $\|\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\|$

The Gram-Schmidt Walk Design

Choose parameter $\phi \in (0,1)$

Run the GSW with input $\begin{pmatrix} \sqrt{\phi}\mathbf{X} \\ \sqrt{1-\phi}\mathbf{I} \end{pmatrix}$, output random $\mathbf{z} \in \{-1,1\}^n$

We show:

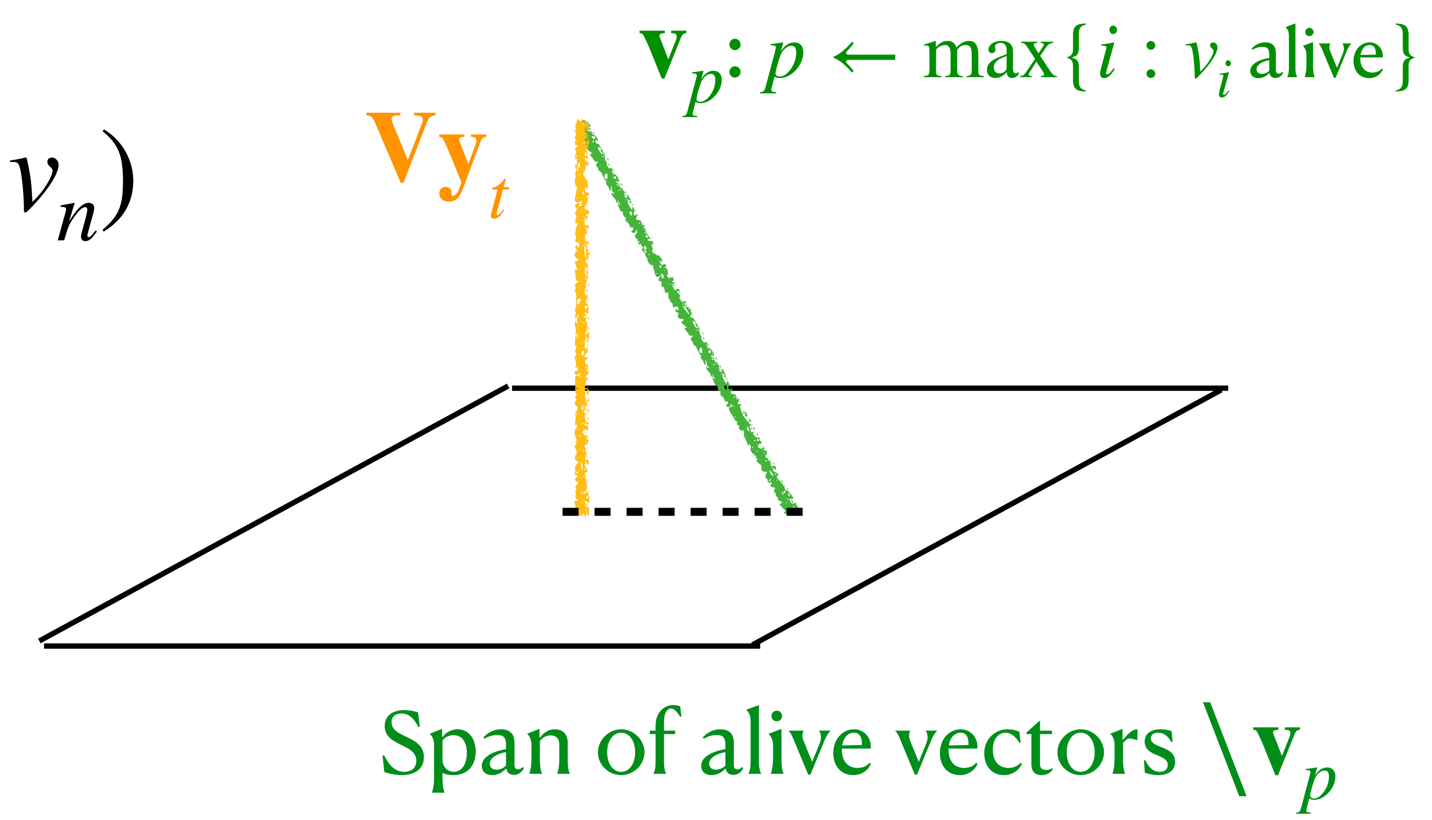
1. $\|\mathbb{E}(\mathbf{Xz})(\mathbf{Xz})^\top\| \leq \frac{1}{\phi}$, $\|\mathbb{E}\mathbf{zz}^\top\| \leq \frac{1}{1-\phi}$
2. $\mathrm{Var}(\hat{\tau}) \leq \frac{L_\phi}{n^2}$, where $L_\phi = \min_\beta\{\frac{1}{1-\phi}\|\mu - X^\top\beta\|^2 + \frac{1}{\phi}\|\beta\|^2\}$
3. $\Pr(|\hat{\tau} - \tau| > \lambda) \leq 2\exp(-\frac{\lambda^2 n^2}{2L_\phi}), \forall\lambda$

The Gram-Schmidt Walk [BDGL'18]

Input: $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ with norm ≤ 1

Initially $\mathbf{z}_0 \in [-1,1]^n$

$V = (v_1, \ldots, v_n)$



$\mathbf{z} \in [-1,1]^n$

Once $z_i = \pm 1$, never change

$\mathbf{v}_i$ is ***alive*** at iter t if $|\mathbf{z}_t(i)| < 1$,

frozen otherwise

For $t = 1,2,\ldots$

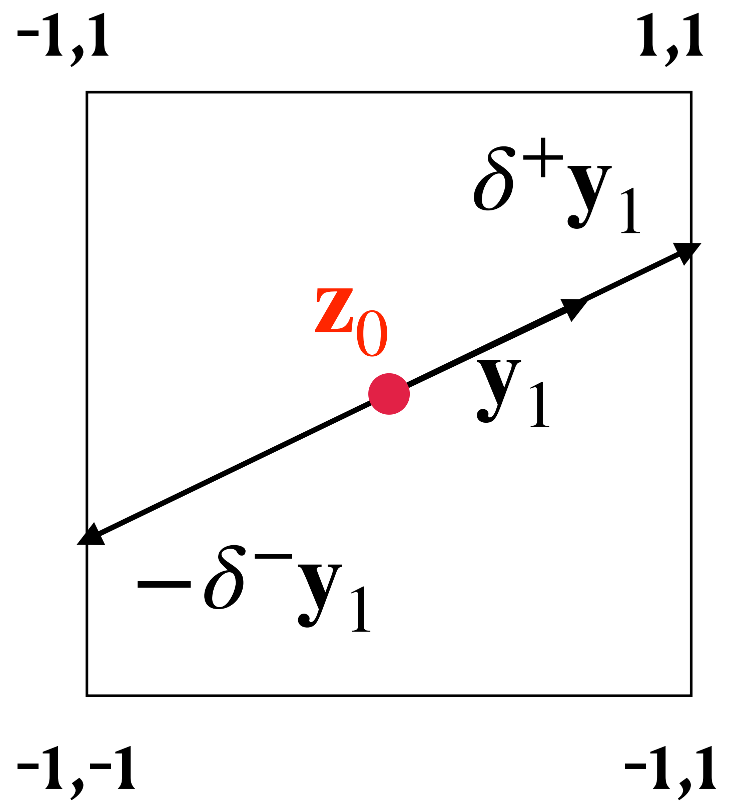
pivot $p_t \leftarrow \max\{i : v_i \text{ alive}\}$

$\mathbf{y}_t \leftarrow \arg\min\{\|\mathbf{V}\mathbf{y}\| : y_p = 1, y_i = 0 \text{ for all frozen } i\}$

$\delta^+ \leftarrow |\max \Delta|, \delta^- \leftarrow |\min \Delta|$, where $\Delta = \{\delta : \mathbf{z}_{t-1} + \delta \mathbf{y}_t \in [-1,1]^n\}$

$\delta_t \leftarrow \delta^+$ w.p. $\dfrac{\delta^-}{\delta^+ + \delta^-}$, $\delta_t \leftarrow -\delta^-$ w.p. $\dfrac{\delta^+}{\delta^+ + \delta^-}$

$\mathbf{z}_t \leftarrow \mathbf{z}_{t-1} + \delta_t \mathbf{y}_t$



$\mathbb{E}[\delta_t] = 0$

1-subgaussian Intuition

Goal:

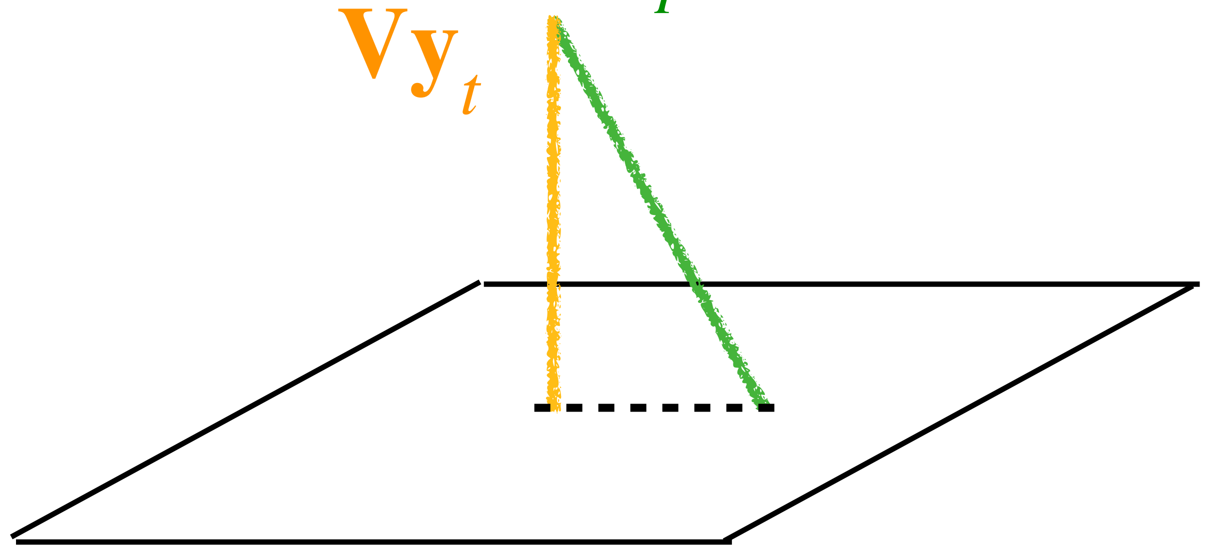
$$\mathbb{E}[\exp(\theta^\top Vz)] \leq \exp(\frac{1}{2}\|\theta\|^2), \forall\theta$$

$\mathbf{z} = \delta_1\mathbf{y}_1 + \delta_2\mathbf{y}_2 + \delta_3\mathbf{y}_3 + \cdots$ form a **martingale**

Break down iterations to **phases** [BDGL'18]: each phase is a maximal consecutive sequence of steps with the same pivot.

$\mathbf{V}\mathbf{y}_t \perp \mathbf{V}\mathbf{y}_{t'}$ if t, t' in different phases

The difficulty: $\mathbf{V}\mathbf{y}_t$'s in a single phase are NOT orthogonal



More about 1-subgaussian

Goal:

$$\mathbb{E}[\exp(\sum_{t=1}^{n} \delta_t \theta^\top \mathbf{V}\mathbf{y}_t)] \leq \exp(\frac{1}{2}\|\theta\|^2), \forall \theta$$

A special case: the alg has 1 phase

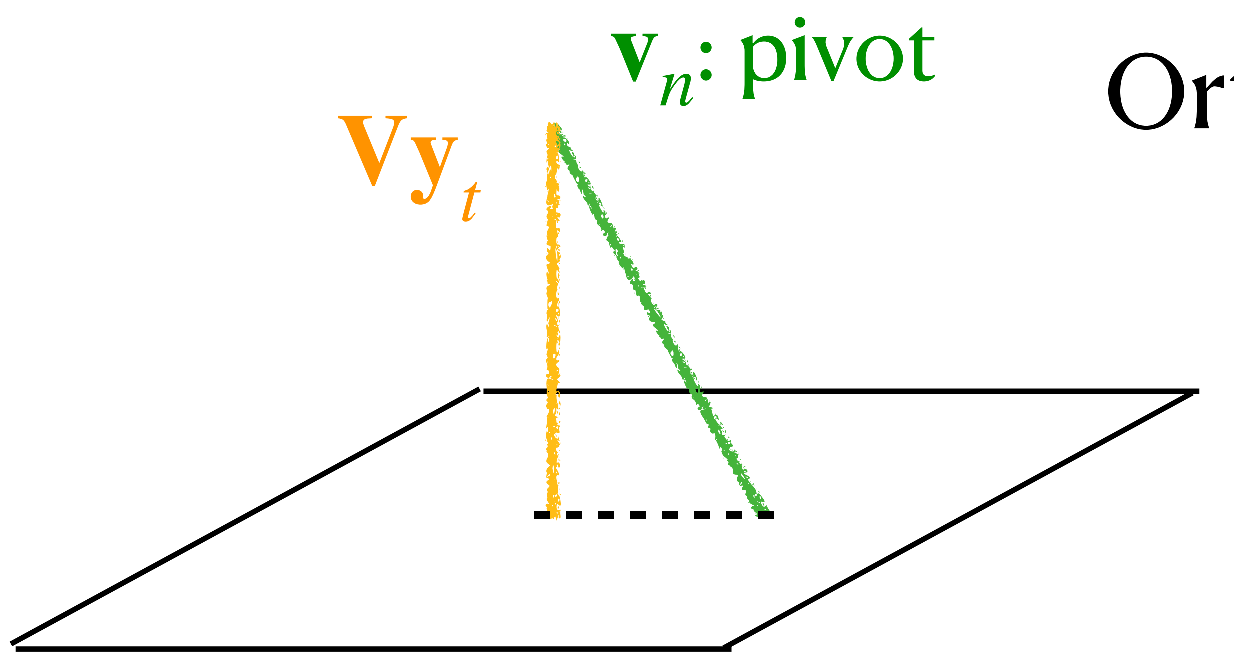
freezing order ←

input vectors: $\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_{n-1} \quad \mathbf{v}_n$

the Gram-Schmidt orth.: $\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_{n-1} \quad \mathbf{b}_n$

→

Orthonormal basis



Span of $\mathbf{v}_1, \dots, \mathbf{v}_{n-t}$

$$\mathbf{V}\mathbf{y}_1 = (\mathbf{v}_n^\top \mathbf{b}_n)\mathbf{b}_n$$

$$\mathbf{V}\mathbf{y}_2 = (\mathbf{v}_n^\top \mathbf{b}_n)\mathbf{b}_n + (\mathbf{v}_n^\top \mathbf{b}_{n-1})\mathbf{b}_{n-1}$$

$$\mathbf{V}\mathbf{y}_3 = (\mathbf{v}_n^\top \mathbf{b}_n)\mathbf{b}_n + (\mathbf{v}_n^\top \mathbf{b}_{n-1})\mathbf{b}_{n-1} + (\mathbf{v}_n^\top \mathbf{b}_{n-2})\mathbf{b}_{n-2}$$

......

$$\|\theta\|^2 = (\theta^\top \mathbf{b}_n)^2 + (\theta^\top \mathbf{b}_{n-1})^2 + \dots + (\theta^\top \mathbf{b}_1)^2$$

More about 1-subgaussian

Goal:

$$\mathbb{E}[\exp(\sum_{t=1}^{n} \delta_t \theta^\top \mathbf{V}\mathbf{y}_t)] \le \exp(\frac{1}{2}\|\theta\|^2), \forall \theta$$

$$\theta^\top V y_t = \alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_t\beta_t$$

$$\|\theta\|^2 = \beta_1^2 + \beta_2^2 + \cdots + \beta_n^2$$

$$\ge (\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2)(\beta_1^2 + \beta_2^2 + \cdots + \beta_n^2)$$

$$\alpha_i \leftarrow \mathbf{v}_n^\top \mathbf{b}_{n-i+1}$$

$$\beta_i \leftarrow \theta^\top \mathbf{b}_{n-i+1}$$

More about 1-subgaussian

Goal:

$$\mathbb{E}[\exp(\sum_{t=1}^{n}\delta_t\sum_{i=1}^{t}\alpha_i\beta_i-\frac{1}{2}(\sum_{i=1}^{n}\alpha_i^2)(\sum_{i=1}^{n}\beta_i^2))]\leq 1$$

$\sum_{i=1}^{n}\alpha_i^2\leq 1,\sum_{i=1}^{n}\beta_i^2\leq 1$

δ_t determines α_{t+1},β_{t+1}

$\mathbb{E}\delta_t=0$

$\sum_{t=1}^{n}\delta_t=\pm 1$

Proof. **Backward induction.**

Conditioned on $\delta_1,\ldots,\delta_{n-1}$,

$$\text{LHS}=\exp(\sum_{t=1}^{n-1}\delta_t\sum_{i=1}^{t}\alpha_i\beta_i)\exp(-\frac{1}{2}(\sum_{i=1}^{n}\alpha_i^2)(\sum_{i=1}^{n}\beta_i^2))\mathbb{E}_{n-1}[\exp(\delta_n(\sum_{i=1}^{n}\alpha_i\beta_i))]$$

maximized

when $\alpha_n=\beta_n=0$

$$\sum_{t=1}^{n}\delta_t=\pm 1,\mathbb{E}[\delta_n]=0$$

Future work

- Better tails $\Pr(|\hat{\tau} - \tau| > \lambda n) \leq \frac{1}{\lambda} \exp(-\frac{\lambda^2}{2}), \forall \lambda$?? —> better confidence interval
- Online design
- More treatment groups

Paper: https://arxiv.org/abs/1911.03071

(version 1: CS; version 2: statisticians, experimenters)

Code: https://github.com/crharshaw/GSWDesign.jl