

Pooled Coronavirus Testing with Side Information

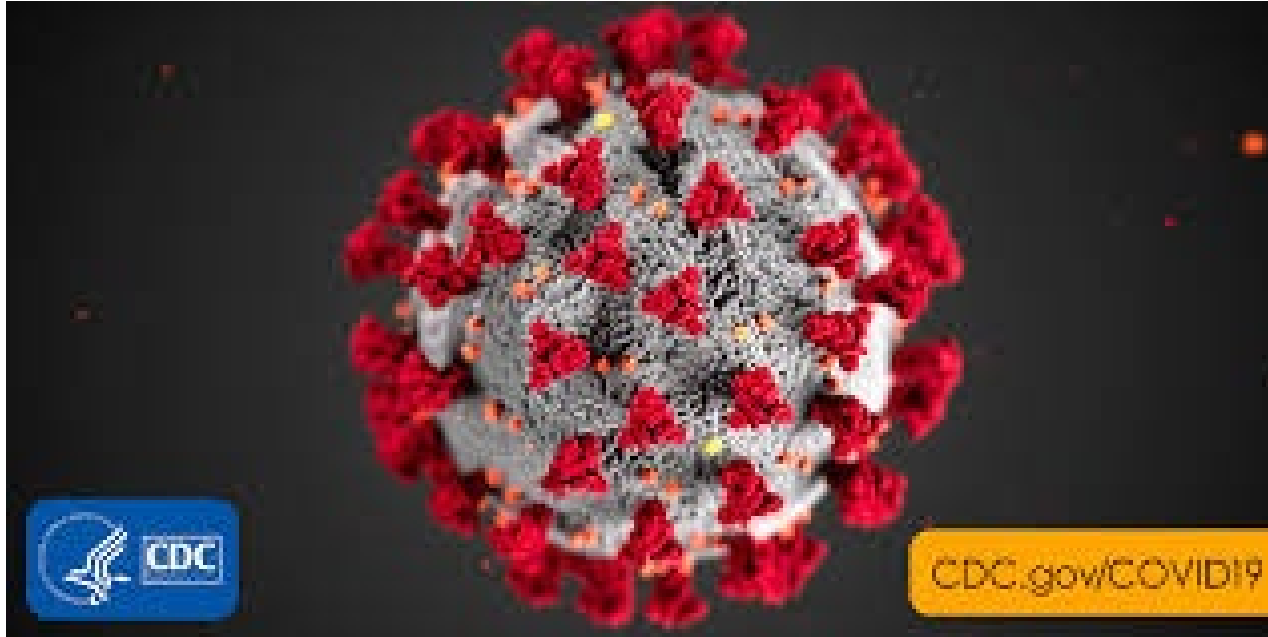


Junan Zhu,^h Kristina Rivera,ⁿ Cindy Rush,^c and Dror Baronⁿ
^hHarvest Fund, ⁿNC State University, and ^cColumbia University

Paris Machine Learning Meetup
May 2020



Motivation



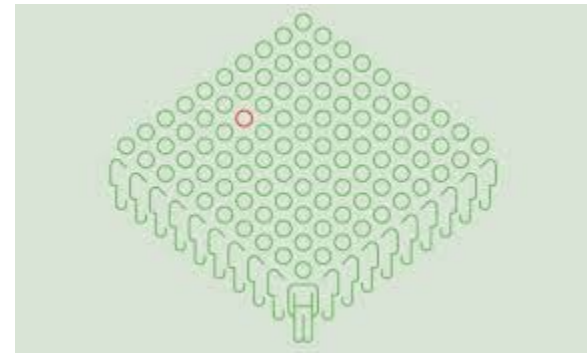
- Fast / efficient / affordable testing of large populations

Conventional testing



- Swab patient for mucus / saliva
- Amplify genetic material with reverse transcription polymerase chain reaction (RT-PCR)
- Test for viral material
- Challenges
 - False negatives / positives
 - Time / resource intense
- Want fast / efficient / affordable testing

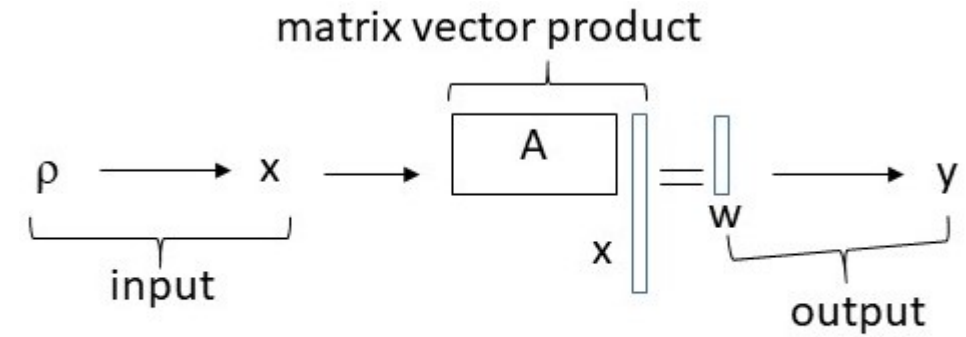
Pooled/group testing



- Suppose few people (0.1%? 1%?) are sick
- Pool group of (10?) people's samples together
- All healthy $\rightarrow$ negative pooled test $\rightarrow$ rules out group
- Any sick $\rightarrow$ positive $\rightarrow$ need more information
- Demonstrated for COVID-19 [Kishony et al. 2020]
- Can optimize pool size [Hanel & Thurner 2020]
- Non-adaptive approaches (Tapestry; IIT Bombay; [Ghosh et al. 2020])
- *Fragile to false negatives / positives*

Proposed approach

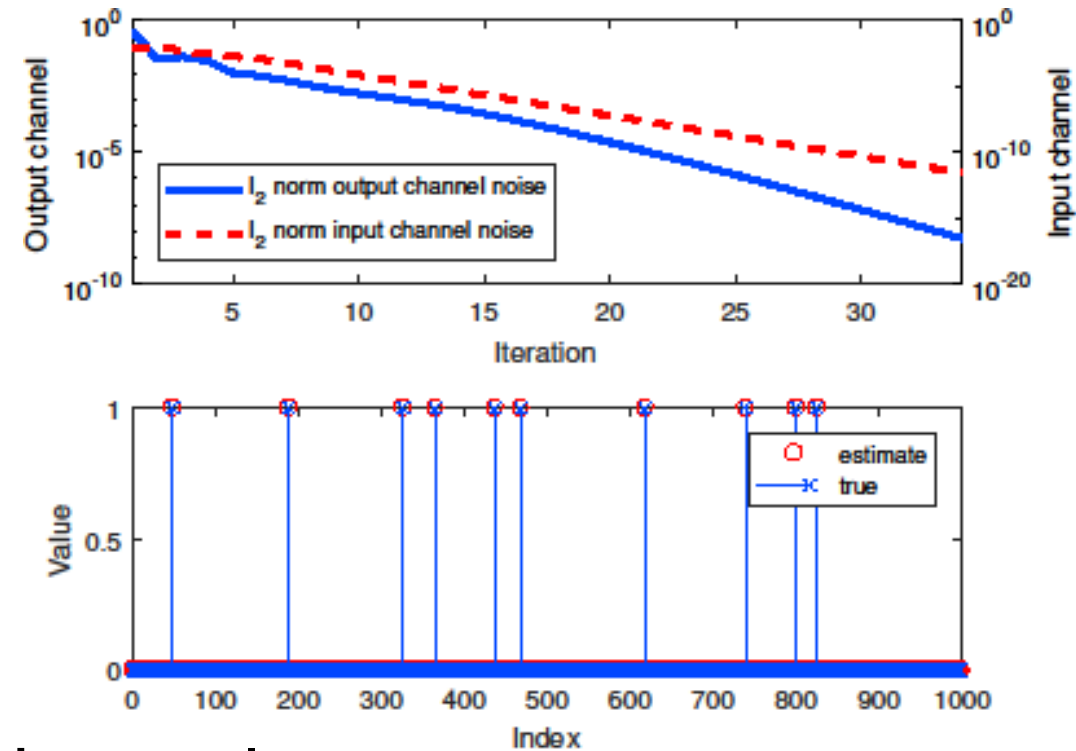
- Key idea: Convert to linear algebra



- ρ sickness prevalence
- x binary *input vector*
 - $x_n=1$ sick; $x_n=0$ healthy
- Multiply x by binary *measurement matrix* A ; rows/cols correspond to measurements / patients
- Matrix vector product w ; w_m #sick in measurement m
- *Noisy* y_m depends on w_m zero / nonzero
- Goal: Estimate x from y , A , statistical info (e.g., ρ)

Preliminary numerical result

- Contribution: GAMP-based algorithm estimates x with iterative *denoising* [Rangan 2011]
- $N=5000$ patients; $\rho=1\%$ prevalence; $M=1000$ measurements
- $R=M/N=20\%$ measurement rate



- Accurate reconstruction
- Fast (~ 1 sec on laptop)
- Iteratively denoises input/output channels

Can we do better?



GAMP with *side information*

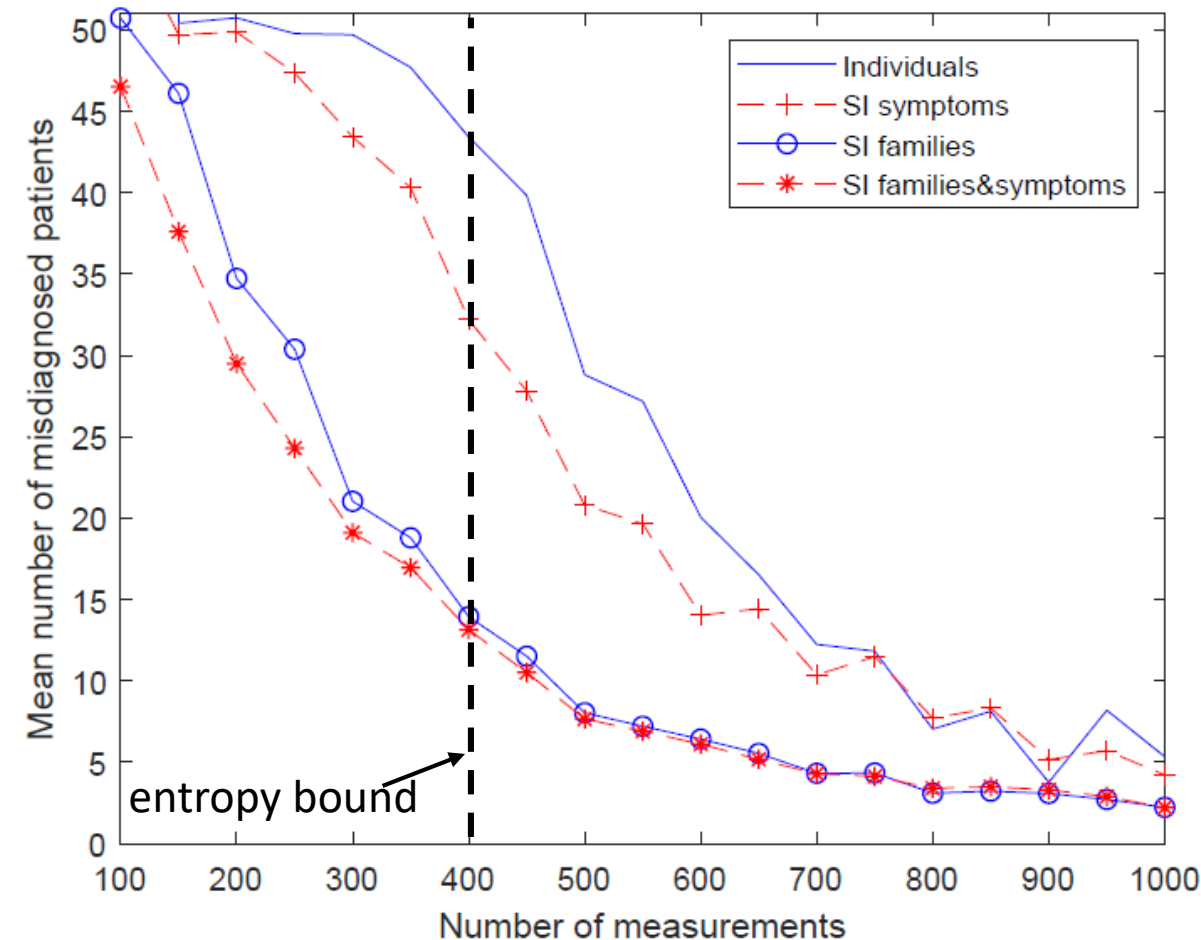
- Earlier goal: Estimate x from y , A , statistical info
- Side information (SI) often available
 - Symptoms affect probability of infection
 - Family members w/ correlated infection status
 - Coworkers sitting nearby
 - Contact tracing



- AMP can use SI in denoiser [Baron et al. 2017, Ma et al. 2019]
- Vector denoisers support dependencies between patients [Donoho et al. 2013, Ma et al. 2014]
- Contribution: Vector denoisers with SI in GAMP

Preliminary numerical result

- As before, $N=5000$ patients; $\rho=1\%$ prevalence
- Families SI (family size $F=4$)
 - $\rho_f=1.5\%$ of families infected; $\rho_i=2/3$ of individuals within families
- Symptoms SI
 - Families w/symptoms $\rho_1=5\%$
 - Families without $\rho_2=1\%$
- *R below entropy bound*
 - SI reduces entropy
- Both types of SI help
- Dependencies more useful



Discussion

- Contributions
 - Convert pooled tests to noisy linear inverse problem
 - GAMP solver
 - Use SI in GAMP; supports dependencies between patients
- Future directions
 - Matrix design
 - Support *very low* false negative rates
 - Part1: use “regular” matrix
 - Part2: retest healthy patients with modified matrix
- More data; more SI; collaborate with PCR people

Thanks!



More details in our paper



Supported in part by National Science Foundation Grant EECS 1611112