Fast & Efficient Data Science Techniques for COVID-19 Group Testing

Varlam Kutateladze, Ekaterina Seregina {varlam.kutateladze, ekaterina.seregina}@email.ucr.edu

Department of Economics, University of California, Riverside

Group Testing

Test ~~individual~~ pooled samples. ✔FDA authorized

"Because samples are pooled together, ultimately fewer tests are run overall, meaning fewer testing supplies are used, and results can be returned to patients more quickly in most cases."

FDA

- Increased testing throughput
- Limited use of chemical reagents
- Higher overall testing capacity

Problem

Pool N individuals according to a pooling matrix A:

	🕴	🕴	🕴	🕴	🕴	…	🕴		PCR
Test 1	●	○	●	○	○	…	○		26.1
Test 2	○	●	○	●	○	…	○		39.9
Test 3	○	○	●	○	●	…	○	$\Longrightarrow$	39.9
Test 4	●	○	○	○	●	…	●		21.5
						⋱			39.9
Test m	○	○	●	○	○	…	●		39.9

Can we detect k positives using only $m \ll N$ tests?

That is, observe $y = Ax + \epsilon$, want to infer x.

Known methods

Other non-adaptive algorithms:

- COMP (Combinatorial Orthogonal Matching Pursuit)
- DD (Definite Defectives)
- CBP (Combinatorial Basis Pursuit)
- SCOMP (Sequential COMP)

Proposed method

(1) Pooling matrix A with constant column weight.

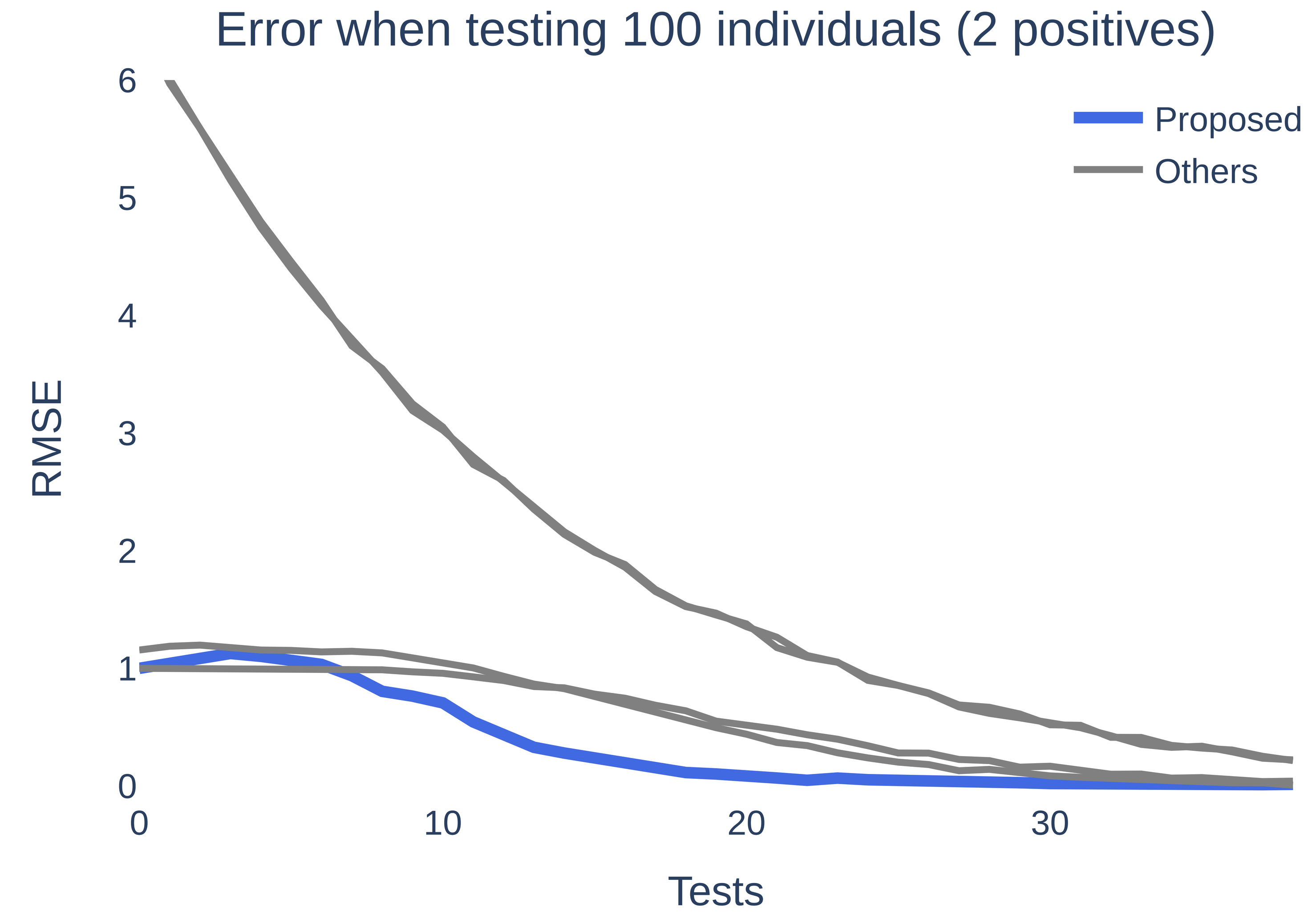
✓ Theoretical justification
✓ Avoid too much dilution
✓ Better performance

(2) Infer x with ℓ_1-norm sparse recovery:

$$\boxed{\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \epsilon, \quad x \geq 0}$$

✓ Guaranteed to recover the sparsest solution

Results



Summary

✔ With 20 test kits can accurately test 100 individuals*

✔ 5x improvement factor over individual testing with 95% sensitivity & specificity*

✔ Fast and efficient, $m = O(klog(N))$

*At 2% disease prevalence

Advantages

- One-round testing
- # of tests scales as $O(k \log(N))$
- Inputs real-numbered readouts
- Reconstructs viral loads
- Adapts to estimation noise

Theoretical contribution

Proposed pooling matrix satisfies the k-Restricted Isometry Property whp:

$$(1 - \delta_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k) \|x\|_2^2,$$

for all k-sparse $x \in \mathbb{R}^n$

Proof strategy: (1) concentration of $\|Ax\|_2^2$ around its mean, $\|x\|_2^2$, (2) covering numbers for finite-dimensional balls in Euclidean space.

Learn more

Code

Paper