

Invariance in Property Testing

Madhu Sudan
MIT

Joint work with Tali Kaufman (IAS).

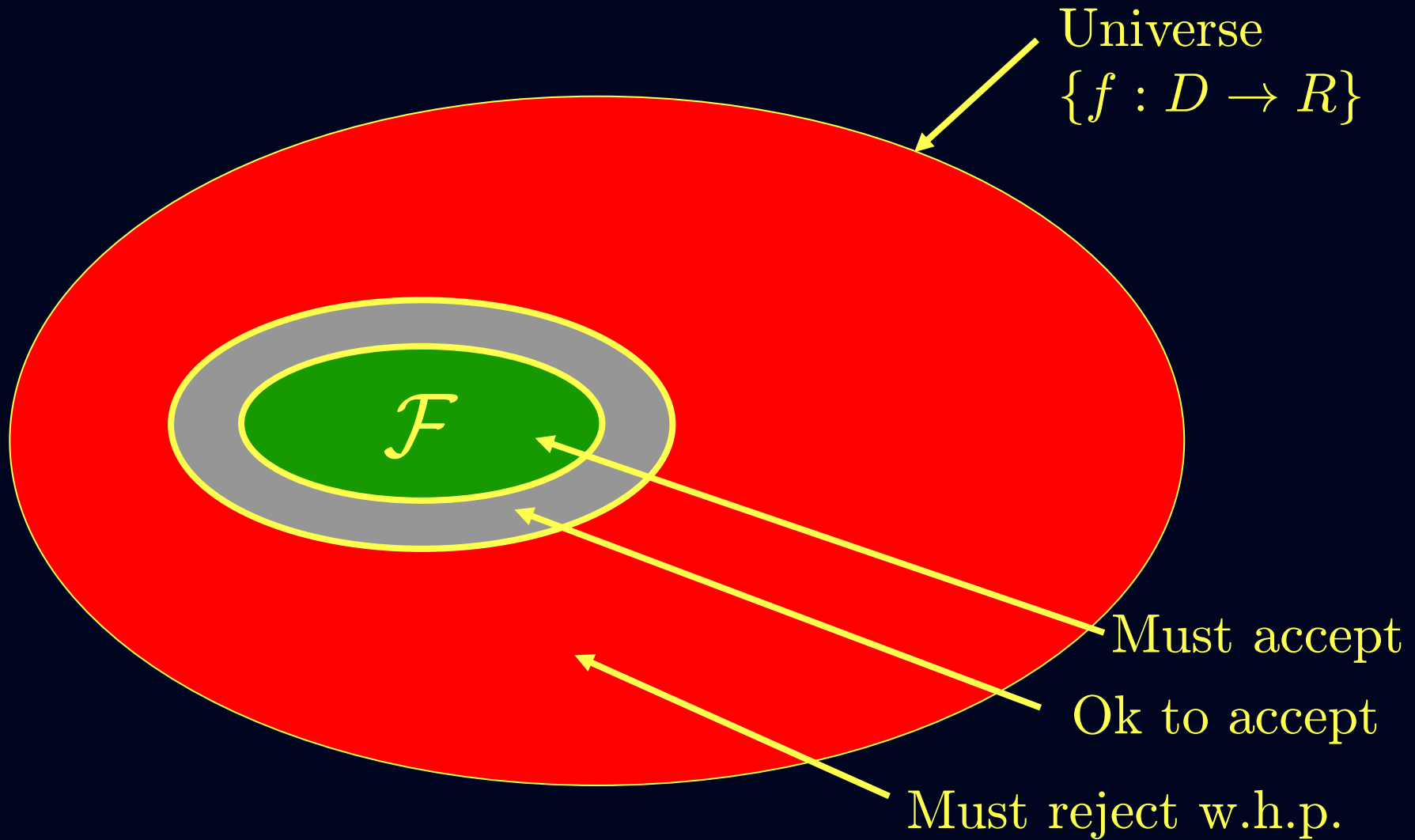
Property Testing

- Goal: “Efficiently” determine if some “data” “essentially” satisfies some given “property”.
- Formalism:
 - Data: $f : D \rightarrow R$ given as oracle
 D finite, but huge. R finite, possibly small
 - Property: Given by $\mathcal{F} \subseteq \{f : D \rightarrow R\}$
 - Efficiently: $o(D)$ queries into f . Even $O(1)$!
 - Essentially: **Must** accept if $f \in \mathcal{F}$
Ok to accept if $f \approx g \in \mathcal{F}$.

Property Testing

- **Distance:** $\delta(f, g) = \Pr_{x \in D}[f(x) \neq g(x)]$
 $\delta(f, \mathcal{F}) = \min_{g \in \mathcal{F}} \{\delta(f, g)\}$
 $f \approx_\epsilon g$ if $\delta(f, g) \leq \epsilon$.
- **Definition:**
 \mathcal{F} is (q, α) -locally testable if
 \exists a q -query tester that
accepts $f \in \mathcal{F}$ with probability one
rejects $f \notin \mathcal{F}$ with probability $\geq \alpha \cdot \delta(f, \mathcal{F})$.
- **Notes:** q -locally testable implies $\exists \alpha > 0$
locally testable implies $\exists q = O(1)$
Weaker testing: can reject $f \in \mathcal{F}$ with small prob.

Property Testing (Pictorially)



Example: Pre-election Polling

- Domain = Population
Range = $\{0, 1\}$
- Property: \mathcal{F} = functions with majority 1
- Essentially:
Must reject w.h.p. if $\Pr_{x \in D}[f(x) = 1] \leq 1/2 - \epsilon$
- Efficiency? Can test weakly with $\tilde{O}(1/\epsilon^2)$ queries.
Chernoff bounds.

Modern Day Example: Testing Linearity

- Domain = Vector space \mathbb{F}_2^n
Range = Field \mathbb{F}_2
- Property: \mathcal{F} = linear functions
i.e., $\{f(x) = \langle a, x \rangle \mid a \in \mathbb{F}_2^n\}$
- Theorem [Blum, Luby, Rubinfeld '89]:
Linearity is 3-locally testable.
- Test: Pick $x, y \in \mathbb{F}_2^n$ uniformly.
Accept iff $f(x) + f(y) = f(x + y)$

Property Testing: Abbreviated History

- Prehistoric: Statistical sampling
 - E.g., “Majority = 1?”
- Linearity Testing [BLR'90], Multilinearity Testing [Babai, Fortnow, Lund '91].
- Graph/Combinatorial Property Testing [Goldreich, Goldwasser, Ron '94].
 - E.g., Is a graph “close” to being 3-colorable.
- Algebraic Testing [GLRSW,RS,FS,AKKLR,KR,JPSZ]
 - Is multivariate function a polynomial (of bounded degree).
- Graph Testing [Alon-Shapira, AFNS, Borgs et al.]
 - Characterizes graph properties that are testable.

Quest for this talk

- What makes a property testable?
- In particular for algebraic properties:
 - Current understanding:
 - Low-degree multivariate functions are testable.
 - Different proofs for different cases.
 - Linear functions
 - Low-degree polynomials
 - Higher degree polynomials over \mathbb{F}_2
 - Higher degree polynomials over other fields

Necessary Conditions for Testability

- One-sided error and testability:
 - Suppose f is rejected by a k -query 1-sided tester.
Suppose queried points are $x_1, \dots, x_k \in D$.
Let $f(x_i) = \alpha_i$.
 - Then for every function $g \in \mathcal{F}$,
 $\langle g(x_1), \dots, g(x_k) \rangle \neq \langle \alpha_1, \dots, \alpha_k \rangle$.
- Constraint: $C = \langle x_1, \dots, x_k \rangle; S \subsetneq R^k$
 g satisfies C if $\langle g(x_1), \dots, g(x_k) \rangle \in S$
 \mathcal{F} satisfies C if every $g \in \mathcal{F}$ satisfies C .
- Conclusion: Testability implies Constraints.

Necessary Conditions for Testability

- One-sided error and testability:
 - Suppose f is rejected by a k -query 1-sided tester.
Suppose queried points are $x_1, \dots, x_k \in D$.
Let $f(x_i) = \alpha_i$.
 - Then for every function $g \in \mathcal{F}$,
 $\langle g(x_1), \dots, g(x_k) \rangle \neq \langle \alpha_1, \dots, \alpha_k \rangle$.
- Constraint: $C = \langle x_1, \dots, x_k \rangle; S \subsetneq R^k$
 g satisfies C if $\langle g(x_1), \dots, g(x_k) \rangle \in S$
 \mathcal{F} satisfies C if every $g \in \mathcal{F}$ satisfies C .
- Conclusion: Testability implies Constraints.

Constraints, Characterizations, Testing

- Strong testing:

Every $f \notin \mathcal{F}$ rejected by some k -local constraint.

Set of k -local constraints characterize \mathcal{F} .

$\exists C_1, \dots, C_m$ s.t. $f \in \mathcal{F} \Leftrightarrow f$ satisfies C_j for every j .

- Conclusion: Testability implies Local Characterizations.

- Example:

$f \subseteq \{\mathbb{F}_2^n \rightarrow \mathbb{F}_2\}$ is linear iff

for all $x, y \in \mathbb{F}_2^n$, f satisfies $C_{x,y}$ where

$C_{x,y} = \langle x, y, x + y \rangle; S = \{000, 011, 101, 110\}$.

Characterizations Sufficient?

- NO! [Ben-Sasson, Harsha, Raskhodnikova]
 - Random 3-locally characterized error-correcting codes (“Expander Codes”) are not $o(D)$ -locally testable.
 - Property:
 $D = [n]$; $R = \{0, 1\}$;
 $\mathcal{F} =$ set of functions that satisfy some random 3-ary \mathbb{F}_2 -linear constraints.
- Criticism: Random constraints too “asymmetric”.
- Perhaps should consider more “symmetric” properties.

Invariance & Property testing

- Invariances (Automorphism groups):

For permutation $\pi : D \rightarrow D$, \mathcal{F} is π -invariant if
 $f \in \mathcal{F}$ implies $f \circ \pi \in \mathcal{F}$.

$\text{Aut}(\mathcal{F}) = \{\pi \mid \mathcal{F} \text{ is } \pi\text{-invariant}\}$

Forms group under composition.

- Hope: If Automorphism group is “large” (“nice”),
then property is testable.

Examples

- Majority:
 - Aut group = S_D (full group).
 - Easy Fact: If $\text{Aut}(\mathcal{F}) = S_D$ then \mathcal{F} is $\text{poly}(R, 1/\epsilon)$ -locally testable.
- Graph Properties:
 - Aut. group given by renaming of vertices
 - [AFNS, Borgs et al.] implies *regular* properties with this Aut group are testable.
- Matrix Properties: Have lots of symmetries – do they suffice?
- Algebraic Properties: What symmetries do they have? Will focus on this today.

Algebraic Properties & Invariances

- Properties:

$D = \mathbb{F}^n$, $R = \mathbb{F}$ (Linearity, Low-degree, Reed-Muller)

Or $D = \mathbb{K} \supseteq \mathbb{F}$, $R = \mathbb{F}$ (Dual-BCH) (\mathbb{K}, \mathbb{F} finite fields)

- Automorphism groups?

Linear transformations of domain.

$\pi(x) = Ax$ where $A \in \mathbb{F}^{n \times n}$ (Linear-Invariant)

- Additional restriction: Linearity

$f, g \in \mathcal{F}$ and $\alpha, \beta \in \mathbb{F}$ implies $\alpha f + \beta g \in \mathcal{F}$

- Question: Are Linear, Linear-Invariant, Locally Characterized Properties Testable?

Linear-Invariance & Testability

- Question: Are Linear, Linear-Invariant, Locally Characterized Properties Testable?
 - Why?
 - Unifies previous results on Prop. Testing.
 - (Will show it also is non-trivial extension)
 - Nice family of 2-transitive group of symmetries.
 - Conjecture [Alon, Kaufman, Krivelevich, Litsyn, Ron] : Linear code with k -local constraint and 2-transitive group of symmetries must be testable.

Our Results

- Theorem 1: $\mathcal{F} \subseteq \{\mathbb{K}^n \rightarrow \mathbb{F}\}$ linear, linear-invariant, k -locally characterized implies \mathcal{F} is $f(\mathbb{K}, k)$ -locally testable.
- Theorem 2: $\mathcal{F} \subseteq \{\mathbb{K}^n \rightarrow \mathbb{F}\}$ linear, *affine*-invariant, has k -local *constraint* implies \mathcal{F} is $f(\mathbb{K}, k)$ -locally testable.
- Other stuff: Study of Linear-invariant Properties.

Linear Invariant Properties

Examples of Linear-Invariant Families

- Polynomials in $\mathbb{F}[x_1, \dots, x_n]$ of degree at most d
- Traces of Poly in $\mathbb{K}[x_1, \dots, x_n]$ of degree at most d
- (Traces of) Homogenous polynomials of degree d
- $\mathcal{F}_1 + \mathcal{F}_2$, where $\mathcal{F}_1, \mathcal{F}_2$ are linear-invariant.
Polynomials supported by degree 2, 3, 5, 7 monomials.

What Dictates Locality of Characterizations?

- Precise locality not yet understood:
 - Depends on p -ary representation of degrees.
 - Example: \mathcal{F} supported by monomials $x^{p^i + p^j}$ behaves like degree two polynomial
- For affine-invariant family dictated (coarsely) by highest degree monomial in family
- For some linear-invariant families, can be *much* less than the highest degree monomial.

Example: $\mathbb{K} = \mathbb{F} = \mathbb{F}_7$; $\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2$

$\mathcal{F}_1 =$ poly of degree at most 16

$\mathcal{F}_2 =$ poly supported on monomials of degree $3 \pmod 6$.

$\text{Degree}(\mathcal{F}) = \Omega(n)$; $\text{Locality}(\mathcal{F}) \leq 49$.

Analysis Ingredients

- Monomial Extraction:

E.g., $xy^2 + xyz + x^4 \in \mathcal{F}$ implies $xyz \in \mathcal{F}$

- Monomial Spread:

$x^5 \in \mathcal{F}$ implies x^4y, x^3y^2 also in \mathcal{F} (if $\text{char}(\mathbb{F})$ large)

Suffices for affine-invariant families.

For linear-invariant families, need to define the right parameter and bound locality weakly in terms of it.

Local Testing

Key Notion: Formal Characterization

- \mathcal{F} is formally characterized if
 \exists a *single* constraint $C = (\langle x_1, \dots, x_k \rangle, S)$ such that
 $\{C \circ \pi\}_{\pi \in \text{Aut}(\mathcal{F})}$ characterize \mathcal{F} .

Theorem: If \mathcal{F} is formally characterized by a k -local constraint (with some restrictions) then it is k -locally testable.

BLR (and our) analysis

BLR Analysis: Outline

- Have f s.t. $\Pr_{x,y}[f(x) + f(y) \neq f(x+y)] = \delta < 1/20$.
Want to show f close to some $g \in \mathcal{F}$.
- Define $g(x) = \text{most likely}_y \{f(x+y) - f(y)\}$.
- If f close to \mathcal{F} then g will be in \mathcal{F} and close to f .
- But if f not close? g may not even be uniquely defined!
- Steps:
 - Step 0: Prove f close to g
 - Step 1: Prove *most likely* is overwhelming majority.
 - Step 2: Prove that g is in \mathcal{F} .

BLR Analysis: Step 0

- Define $g(x) = \text{most likely}_y \{f(x+y) - f(y)\}$.

Claim: $\Pr_x[f(x) \neq g(x)] \leq 2\delta$

– Let $B = \{x \mid \Pr_y[f(x) \neq f(x+y) - f(y)] \geq \frac{1}{2}\}$

– $\Pr_{x,y}[\text{linearity test rejects} \mid x \in B] \geq \frac{1}{2}$

$$\Rightarrow \Pr_x[x \in B] \leq 2\delta$$

– If $x \notin B$ then $f(x) = g(x)$

$\text{Vote}_x(y)$

BLR Analysis: Step 1

- Define $g(x) = \text{most likely } y \{f(x + y) - f(y)\}$.
- Suppose for some x , \exists two equally likely values.
Presumably, only one leads to linear x , so which one?
- If we wish to show g linear,
then need to rule out this case.

Lemma: $\forall x, \Pr_{y,z}[\text{Vote}_x(y) \neq \text{Vote}_x(z)] \leq 4\delta$

$\text{Vote}_x(y)$

BLR Analysis: Step 1

- Define $g(x) = \text{most likely } y \{f(x + y) - f(y)\}$.
- Suppose for some x , \exists two equally likely values.
Presumably, only one leads to linear x , so which one?
- If we wish to show g linear,
then need to rule out this case.

Lemma: $\forall x, \Pr_{y,z}[\text{Vote}_x(y) \neq \text{Vote}_x(z)] \leq 4\delta$

BLR Analysis: Step 1

$\text{Vote}_x(y)$

- Define $g(x) = \text{most likely } y \{f(x + y) - f(y)\}$.

Lemma: $\forall x, \Pr_{y,z} [\text{Vote}_x(y) \neq \text{Vote}_x(z)] \leq 4\delta$

?	$f(y)$	$-f(x + y)$
$f(z)$	$f(y + z)$	$-f(y + 2z)$
$-f(x + z)$	$-f(2y + z)$	$f(x + 2y + 2z)$

Prob. Row/column sum non-zero $\leq \delta$.

BLR Analysis: Step 1

$\text{Vote}_x(y)$

- Define $g(x) = \text{most likely } y \{f(x + y) - f(y)\}$.

Lemma: $\forall x, \Pr_{y,z} [\text{Vote}_x(y) \neq \text{Vote}_x(z)] \leq 4\delta$

?	$f(y)$	$-f(x + y)$
$f(z)$	$f(y + z)$	$-f(y + 2z)$
$-f(x + z)$	$-f(2y + z)$	$f(x + 2y + 2z)$

Prob. Row/column sum non-zero $\leq \delta$.

BLR Analysis: Step 2 (Similar)

Lemma: If $\delta < \frac{1}{20}$, then $\forall x, y, g(x) + g(y) = g(x + y)$

$g(x)$	$g(y)$	$-g(x + y)$	Prob. Row/column sum non-zero $\leq 4\delta$.
$f(z)$	$f(y + z)$	$-f(y + 2z)$	
$-f(x + z)$	$-f(2y + z)$	$f(x + 2y + 2z)$	

Our Analysis: Outline

- f s.t. $\Pr_L[\langle f(L(x_1)), \dots, f(L(x_k)) \rangle \in V] = \delta \ll 1$.
- Define $g(x) = \alpha$ that maximizes
$$\Pr_{\{L|L(x_1)=x\}}[\langle \alpha, f(L(x_2)), \dots, f(L(x_k)) \rangle \in V]$$
- Steps:
 - Step 0: Prove f close to g
 - Step 1: Prove “most likely” is overwhelming majority.
 - Step 2: Prove that g is in \mathcal{F} .

Our Analysis: Outline

- f s.t. $\Pr_L[\langle f(L(x_1)), \dots, f(L(x_k)) \rangle \in V] = \delta \ll 1$.

- Define $g(x) = \alpha$ that maximizes

$$\Pr_{\{L|L(x_1)=x\}}[\langle \alpha, f(L(x_2)), \dots, f(L(x_k)) \rangle \in V]$$

- Steps:

- Step 0: Prove f close to g

- Step 1: Prove “most likely” is overwhelming majority.

- Step 2: Prove that g is in \mathcal{F} .

Same as before

$\text{Vote}_x(L)$

Matrix Magic?

- Define $g(x) = \alpha$ that maximizes

$$\Pr_{\{L \mid L(x_1) = x\}} [\langle \alpha, f(L(x_2)), \dots, f(L(x_k)) \rangle \in V]$$

Lemma: $\forall x, \Pr_{L,K} [\text{Vote}_x(L) \neq \text{Vote}_x(K)] \leq 2(k-1)\delta$

x	$L(x_2)$	\dots	$L(x_k)$
$K(x_2)$			
\vdots			
$K(x_k)$			

Matrix Magic?

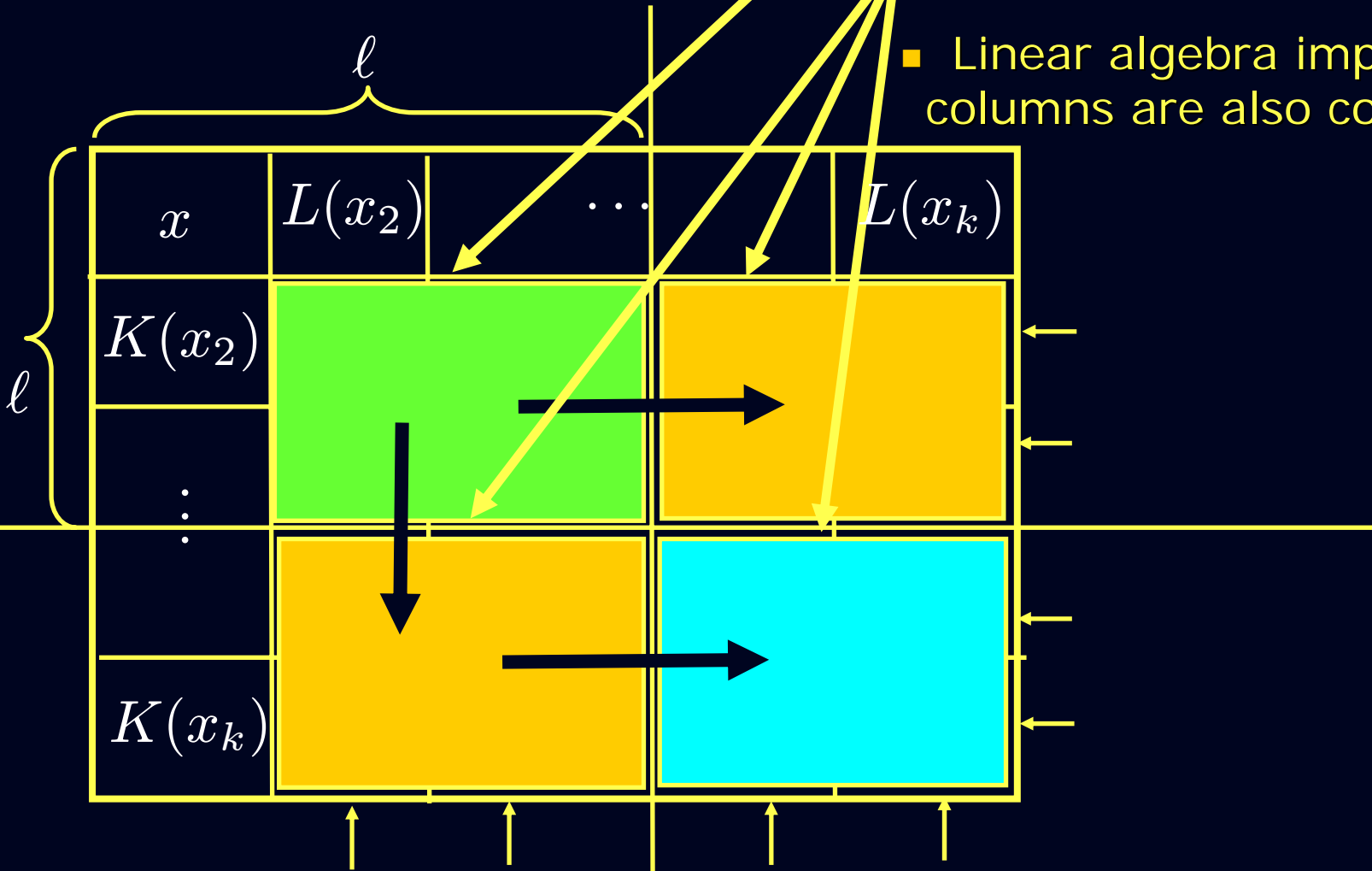
x	$L(x_2)$	\dots	$L(x_k)$	
$K(x_2)$				←
\vdots				←
$K(x_k)$				←

↑ ↑ ↑ ↑

- Want marked rows to be random constraints.
- Suppose x_1, \dots, x_ℓ linearly independent;
and rest dependent on them.

Matrix Magic?

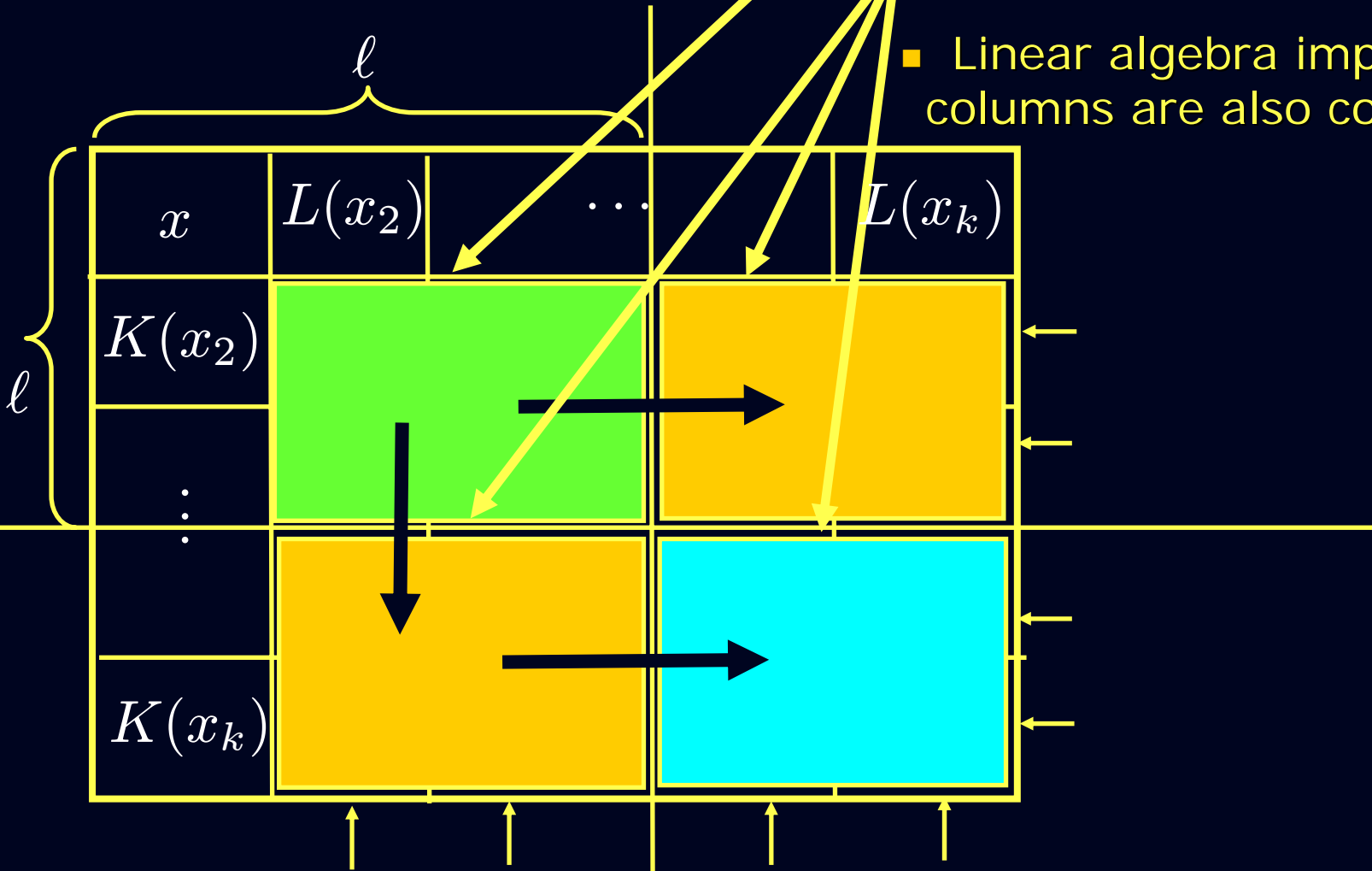
- Fill with random entries
- Fill so as to form constraints
- Linear algebra implies final columns are also constraints.



- Suppose x_1, \dots, x_ℓ linearly independent; and rest dependent on them.

Matrix Magic?

- Fill with random entries
- Fill so as to form constraints
- Linear algebra implies final columns are also constraints.



- Suppose x_1, \dots, x_ℓ linearly independent; and rest dependent on them.

Conclusions

- Invariance is important in property testing.
- Linear-invariance suffices to explain many algebraic tests (and shows some new ones).
- Future work: What are other invariances that lead to testability (from characterizations)?