An Optimizer Formulation for Tolerant Junta Testing

Working note

July 3, 2026

Abstract

This note formulates an optimizer-based route toward tolerant junta testing. The starting point is the distance-estimation problem studied by Chen, Patel, and Servedio: given query access to a Boolean function

$$f : \{\pm 1\}^n \to \{\pm 1\},$$

estimate its distance to the class of k-juntas. Their algorithm gives an adaptive query bound of $2^{\widetilde{O}(k^{1/3})}$. Here we isolate a continuous relaxation over the junta polytope and ask whether the paper's coordinate-refinement machinery can be interpreted as a specialized stochastic optimizer for that relaxation.

1 The Testing Problem

Let $\mathcal{J}_k$ denote the class of all Boolean-valued functions on $\{\pm 1\}^n$ that depend on at most k coordinates. The tolerant junta testing problem is to estimate

$$\mathrm{dist}(f, \mathcal{J}_k) = \min_{g \in \mathcal{J}_k} \mathbb{P}_{x \sim \{\pm 1\}^n} \big[f(x) \neq g(x) \big].$$

Equivalently, define the junta correlation over a fixed coordinate set $A \subseteq [n]$ by

$$\mathrm{corr}(f, \mathcal{J}_A) = \max_{g : \{\pm 1\}^A \to \{\pm 1\}} \mathbb{E}_x [f(x) g(x)].$$

Then

$$\mathrm{corr}(f, \mathcal{J}_k) = \max_{\substack{A \subseteq [n] \\ |A| \leq k}} \mathrm{corr}(f, \mathcal{J}_A),$$

and for Boolean-valued f,

$$\mathrm{dist}(f, \mathcal{J}_k) = \frac{1}{2} \left(1 - \mathrm{corr}(f, \mathcal{J}_k) \right).$$

Thus the core hidden optimization problem is

$$\boxed{\text{estimate } \max_{\substack{A \subseteq [n] \\ |A| \leq k}} \mathrm{corr}(f, \mathcal{J}_A)}$$

from black-box queries to f.

2 Fourier View

Write the Fourier expansion of f as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x), \qquad \chi_S(x) = \prod_{i \in S} x_i.$$

For a fixed set A, the conditional expectation of f onto the coordinates A is

$$f_A(x) = \mathbb{E}[f(y) \mid y_A = x_A] = \sum_{S \subseteq A} \widehat{f}(S)\chi_S(x).$$

The quantity

$$\|f_A\|_2^2 = \sum_{S \subseteq A} \widehat{f}(S)^2$$

measures how much Fourier mass of f is captured by coordinates in A. This is not identical to junta correlation, which involves an L_1 norm, but it is a natural smooth surrogate and is tightly aligned with the spectral language used in tolerant junta testing.

3 A Continuous Relaxation

Define the relaxed junta polytope

$$\Delta_k = \left\{ w \in [0,1]^n : \sum_{i=1}^n w_i \leq k \right\}.$$

Here w_i is interpreted as the soft weight assigned to coordinate i. For $w \in \Delta_k$, define the Fourier-energy surrogate

$$\Phi_f(w) = \sum_{S \subseteq [n]} \widehat{f}(S)^2 \prod_{i \in S} w_i.$$

If $w = \mathbf{1}_A$ is the indicator vector of a coordinate set A, then

$$\Phi_f(\mathbf{1}_A) = \sum_{S \subseteq A} \widehat{f}(S)^2 = \|f_A\|_2^2.$$

Therefore the discrete captured-Fourier-mass problem is exactly the restriction of Φ_f to integral points of Δ_k.

Problem 1 (Optimizer formulation)**.** *Given query access to $f : \{\pm 1\}^n \to \{\pm 1\}$, design a stochastic optimization algorithm over*

$$\Delta_k = \{w \in [0,1]^n : \sum_i w_i \leq k\}$$

that approximately maximizes $\Phi_f(w)$, or a related smooth surrogate for $\mathrm{corr}(f, \mathcal{J}_k)$*, using query complexity comparable to*

$$2^{\widetilde{O}(k^{1/3})}.$$

4 Gradient Structure

The formal partial derivative of Φ_f is

$$\frac{\partial \Phi_f}{\partial w_i}(w) = \sum_{\substack{S \subseteq [n] \\ i \in S}} \widehat{f}(S)^2 \prod_{j \in S \setminus \{i\}} w_j.$$

At $w = \mathbf{1}_A$, this becomes

$$\frac{\partial \Phi_f}{\partial w_i}(\mathbf{1}_A) = \sum_{T \subseteq A} \widehat{f}(T \cup \{i\})^2, \qquad i \notin A.$$

This is an influence-like signal: it measures how much additional Fourier mass would become visible if coordinate i were added to the current candidate set A.

This gives the heuristic optimizer picture:

$$\begin{aligned} \text{candidate weights} &\longrightarrow \text{estimate influence-like gradient signal} \\ &\longrightarrow \text{project back to } \Delta_k \\ &\longrightarrow \text{round to a coordinate set.} \end{aligned}$$

5 Relation to Chen–Patel–Servedio

Chen, Patel, and Servedio do not run generic gradient descent. Their method is a specialized adaptive Fourier search. Still, the pieces look optimizer-like:

1. The hidden objective is to identify a coordinate set A whose junta correlation with f is nearly optimal.

2. Normalized influences provide a query-accessible signal about which coordinates participate in important Fourier coefficients.

3. Their coordinate-refinement procedure repeatedly improves candidate sets, functioning like a combinatorial analogue of a noisy ascent method.

4. Their sharp noise operator suppresses Fourier mass outside the current candidate set, making local estimation feasible.

5. Local estimators approximate the value of candidate junta correlations without reconstructing the best junta.

Thus one possible interpretation is:

$$\text{their tolerant tester} \approx \text{a provable, Fourier-adapted stochastic optimizer over sparse supports.}$$

6 A Theorem Template

The strongest useful statement would connect the relaxed optimizer to the original tolerant testing quantity.

Conjecture 1 (Optimizer-to-tester route)**.** *There is a query-efficient stochastic optimization procedure which, given black-box access to* $f : \{\pm 1\}^n \to \{\pm 1\}$*, returns* $w \in \Delta_k$ *such that after rounding* w *to a set* $A \subseteq [n]$ *with* $|A| \leq k$*,*

$$\operatorname{corr}(f, \mathcal{J}_A) \geq \operatorname{corr}(f, \mathcal{J}_k) - \varepsilon.$$

Consequently,

$$\operatorname{dist}(f, \mathcal{J}_A) \leq \operatorname{dist}(f, \mathcal{J}_k) + O(\varepsilon).$$

A weaker but perhaps more tractable version would use the Fourier-energy surrogate first.

Conjecture 2 (Surrogate version)**.** *Suppose* f *is* δ*-close to a* k*-junta. Then a stochastic optimizer for* Φ_f*, using query-estimated influence gradients, returns a set* A *with* $|A| \leq k$ *such that*

$$\sum_{S \subseteq A} \widehat{f}(S)^2 \geq 1 - O(\delta) - \varepsilon.$$

If true, this would certify that almost all Fourier energy of f is captured by A. One would then need an additional argument converting this energy certificate into a junta-distance or junta-correlation certificate.

7 Immediate Obstacles

There are several technical obstacles to a direct gradient-descent theorem.

1. $\Phi_f(w)$ is high-degree, multilinear, and generally nonconcave.
2. The gradient is not directly observable; it must be estimated from queries to f.
3. Junta correlation is fundamentally an L_1-type quantity, while Φ_f is an L_2-Fourier-energy surrogate.
4. Worst-case query guarantees are much stronger than heuristic optimizer performance.
5. Rounding a fractional w to a set A may lose information unless the relaxation has additional structure.

8 Proposed Research Direction

The problem can be sharpened into three subproblems.

Problem 2 (Gradient estimation)**.** *Give a query algorithm that estimates*

$$\nabla \Phi_f(w)$$

or a normalized-influence substitute at the precision needed for projected ascent over Δ_k*.*

9 Solution to the Gradient Estimation Problem

The gradient of Φ_f is query-estimable. The key observation is that each coordinate derivative is an anisotropic noisy influence.

For $i \in [n]$, define the discrete derivative

$$D_i f(x_{-i}) = \frac{f(x_{-i}, +1) - f(x_{-i}, -1)}{2}.$$

Its Fourier expansion is

$$D_i f = \sum_{T \subseteq [n] \setminus \{i\}} \widehat{f}(T \cup \{i\}) \chi_T.$$

For $w \in [0,1]^n$, set

$$\rho_j = \sqrt{w_j}.$$

Let $T_\rho^{(-i)}$ be the anisotropic noise operator on all coordinates except i, with Fourier multiplier

$$\chi_T \mapsto \left(\prod_{j \in T} \rho_j \right) \chi_T.$$

Then

$$\frac{\partial \Phi_f}{\partial w_i}(w) = \sum_{T \subseteq [n] \setminus \{i\}} \widehat{f}(T \cup \{i\})^2 \prod_{j \in T} w_j = \left\| T_\rho^{(-i)} D_i f \right\|_2^2.$$

Thus the gradient coordinate is the squared L_2 norm of a noisy derivative, and this norm has a direct query estimator.

Theorem 1 (Coordinate-gradient estimator)**.** *Fix $w \in [0,1]^n$, $i \in [n]$, and accuracy/confidence parameters $\eta, \delta \in (0,1)$. There is a randomized query algorithm using*

$$O\left(\eta^{-2} \log(1/\delta)\right)$$

queries to f that outputs $\widehat{g}_i$ satisfying

$$\left| \widehat{g}_i - \frac{\partial \Phi_f}{\partial w_i}(w) \right| \le \eta$$

with probability at least $1 - \delta$.

Proof. Repeat the following experiment M times. First sample $x_{-i} \in \{\pm 1\}^{n-1}$ uniformly. Then independently sample y_{-i} and z_{-i} from the anisotropic noisy distribution around x_{-i}: for every $j \neq i$, set the coordinate equal to x_j with probability ρ_j, and otherwise resample it uniformly from $\{\pm 1\}$. Equivalently,

$$\mathbb{E}[\chi_T(y_{-i}) \mid x_{-i}] = \left(\prod_{j \in T} \rho_j \right) \chi_T(x_{-i}),$$

and the same holds independently for z_{-i}.

Query f four times to compute

$$Y = D_i f(y_{-i}) = \frac{f(y_{-i}, +1) - f(y_{-i}, -1)}{2},$$

and

$$Z = D_i f(z_{-i}) = \frac{f(z_{-i}, +1) - f(z_{-i}, -1)}{2}.$$

Output the empirical average of YZ.

Since y_{-i} and z_{-i} are conditionally independent given x_{-i},

$$\mathbb{E}[YZ \mid x_{-i}] = \left(T_\rho^{(-i)} D_i f(x_{-i})\right)^2.$$

Averaging over x_{-i},

$$\mathbb{E}[YZ] = \left\|T_\rho^{(-i)} D_i f\right\|_2^2 = \frac{\partial \Phi_f}{\partial w_i}(w).$$

Also $Y, Z \in [-1, 1]$, so $YZ \in [-1, 1]$. Hoeffding's inequality gives

$$\mathbb{P}\left[\left|\frac{1}{M}\sum_{r=1}^{M} Y_r Z_r - \frac{\partial \Phi_f}{\partial w_i}(w)\right| > \eta\right] \leq 2\exp\left(-\frac{M\eta^2}{2}\right).$$

Taking

$$M \geq 2\eta^{-2}\log(2/\delta)$$

proves the claim. □

Theorem 2 (Full-gradient estimator on a reduced coordinate universe)**.** *Suppose the active coordinate universe has size m. There is a randomized query algorithm using*

$$O\left(m\eta^{-2}\log(m/\delta)\right)$$

queries to f that outputs $\widehat{g} \in \mathbb{R}^m$ satisfying

$$\|\widehat{g} - \nabla\Phi_f(w)\|_\infty \leq \eta$$

with probability at least $1 - \delta$.

Proof. Run the coordinate estimator for every $i \in [m]$ with failure probability δ/m, and take a union bound. □

This is the precision guarantee needed by projected or conditional-gradient ascent. Indeed, for any two feasible points $u, v \in \Delta_k$,

$$|\langle \widehat{g} - \nabla\Phi_f(w), u - v\rangle| \leq \|\widehat{g} - \nabla\Phi_f(w)\|_\infty \|u - v\|_1 \leq 2k\eta.$$

Thus if we choose

$$\eta \leq \frac{\varepsilon}{8k},$$

then every linearized ascent decision over Δ_k is distorted by at most $\varepsilon/4$. In particular, a linear maximization oracle using the estimated gradient,

$$\widehat{s} \in \arg\max_{s \in \Delta_k} \langle \widehat{g}, s\rangle,$$

satisfies

$$\langle \nabla\Phi_f(w), \widehat{s}\rangle \geq \max_{s \in \Delta_k} \langle \nabla\Phi_f(w), s\rangle - 2k\eta.$$

Since maximizing a linear function over Δ_k simply chooses the coordinates with largest positive gradient values, the above estimator gives a query-efficient noisy ascent direction.

Equivalently, define the anisotropic influence substitute

$$\mathrm{Inf}_i^{(w)}(f) := \left\| T_{\sqrt{w}}^{(-i)} D_i f \right\|_2^2 .$$

Then

$$\mathrm{Inf}_i^{(w)}(f) = \frac{\partial \Phi_f}{\partial w_i}(w).$$

So the normalized-influence substitute required for ascent is not merely analogous to the gradient; for this surrogate objective it is exactly the gradient coordinate.

Problem 3 (Landscape guarantee)**.** *Identify structural conditions under which Φ_f has no bad stationary points on Δ_k, or under which noisy projected ascent reaches a point whose rounded support captures nearly optimal junta correlation.*

Problem 4 (Connection to tolerant testing)**.** *Prove that the optimizer certificate implies*

$$\left| \mathrm{dist}(f, \mathcal{J}_k) - \frac{1}{2}\left(1 - \widehat{\mathrm{corr}}\right) \right| \leq \varepsilon$$

for an efficiently computable estimate $\widehat{\mathrm{corr}}$.

The clean conceptual target is therefore:

$$\boxed{\text{reinterpret adaptive tolerant junta testing as stochastic optimization over sparse Fourier support.}}$$

If this can be made rigorous, it would explain the Chen–Patel–Servedio algorithm as more than a clever querying procedure: it would become an instance of a broader optimizer principle for noisy Boolean structure testing.