

Tolerant testers for self-correctable properties

Arnab Bhattacharyya
abhattach@mit.edu

May 17, 2010

The purpose of this short note is to record a simple but useful observation made by Madhu Sudan in a conversation in 2009 (which may already have been folklore earlier) that self-correctable properties are testable tolerantly. Let us start by making the necessary definitions. A property \mathcal{P} consists of functions f defined over some finite domain \mathcal{D} of size n . The distance between two functions f and g over the domain \mathcal{D} is:

$$\text{dist}(f, g) \stackrel{\text{def}}{=} \frac{|\{x \in \mathcal{D} : f(x) \neq g(x)\}|}{n}$$

The distance of f to a given property \mathcal{P} is defined to be $\text{dist}(f, \mathcal{P}) \stackrel{\text{def}}{=} \min_{g \in \mathcal{P}} \text{dist}(f, g)$. The *unique decoding radius* of a property \mathcal{P} is the maximum $\delta \in (0, 1)$ such that for any f satisfying $\text{dist}(f, \mathcal{P}) \leq \delta$, there is a unique $g \in \mathcal{P}$ satisfying $\text{dist}(f, g) = \text{dist}(f, \mathcal{P})$, and we say f is *uniquely decodable to g* in this case.

Definition 1 (Self-correctability) *Let \mathcal{P} be a property consisting of functions over \mathcal{D} . A self-corrector for \mathcal{P} with query complexity q is a randomized algorithm A that, when given oracle access to a function f that is uniquely decodable to $g \in \mathcal{P}$, has the following behavior. On any input $x \in \mathcal{D}$, A^f evaluates f on at most q elements of \mathcal{D} , perhaps chosen adaptively, and returns y with the guarantee that $\Pr[y = g(x)] \geq \frac{2}{3}$, where the probability is over the randomness of A .*

Next, we move on to the notion of tolerant testing.

Definition 2 (Tolerant testability) *Let \mathcal{P} be a property consisting of functions over \mathcal{D} and let $0 \leq \epsilon_1 < \epsilon_2 \leq 1$. An (ϵ_1, ϵ_2) -tolerant tester for \mathcal{P} with query complexity q is a randomized algorithm T that, when given oracle access to a function f defined over \mathcal{D} , has the following behavior. T^f evaluates f on at most q elements of \mathcal{D} , perhaps chosen adaptively, and then accepts f with probability $\geq 2/3$ if $\text{dist}(f, \mathcal{P}) \leq \epsilon_1$ and rejects f with probability $\geq 2/3$ if $\text{dist}(f, \mathcal{P}) \geq \epsilon_2$, where the probabilities are over the randomness of the tester.*

Our main claim is the following.

Theorem 3 *Let \mathcal{P} be a property consisting of functions over \mathcal{D} with unique decoding radius δ . Suppose there is a self-corrector for \mathcal{P} with query complexity $q(n)$. Then, for every ϵ_1, ϵ_2 such that $0 \leq \epsilon_1 < \epsilon_2 < \delta$, there is an (ϵ_1, ϵ_2) -tolerant tester for \mathcal{P} with query complexity $O(q(n) \cdot \frac{1}{(\epsilon_2 - \epsilon_1)^2} \log \frac{1}{\epsilon_2 - \epsilon_1})$.*

Proof: The (ϵ_1, ϵ_2) -tolerant tester T we construct, when given oracle access to a function f over \mathcal{D} , estimates $\text{dist}(f, \mathcal{P})$ that is correct to within additive error $(\epsilon_2 - \epsilon_1)/2$ with probability at least $2/3$ and then accepts or rejects based on whether the estimated distance is closer to ϵ_1 or ϵ_2 accordingly. The correctness of T is clear.

Now we describe how to estimate the distance to \mathcal{P} . Let $d = \text{dist}(f, \mathcal{P})$, let $g \in \mathcal{P}$ be the function that f uniquely decodes to, and let A be the self-corrector for \mathcal{P} with query complexity $q(n)$. Observe that for any $x \in \mathcal{D}$ and for any $\alpha \in (0, 1)$, we can obtain a value $\tilde{g}(x)$ such that $\Pr[\tilde{g}(x) \neq g(x)] < \alpha$ by invoking the self-corrector $O(\log 1/\alpha)$ many times with independent random coins and letting $\tilde{g}(x)$ be the value output by A majority of the times; this is seen by an application of Chernoff's bound. We will fix a value for α later. Now, T operates by picking s elements $x_1, \dots, x_s \in \mathcal{D}$ uniformly and independently at random (with the number s again to be specified later) and setting $\tilde{d} = \frac{|\{i \in [s]: \tilde{g}(x_i) \neq f(x_i)\}|}{s}$. We argue that \tilde{d} is our desired estimate of d .

To see this, let $U(x)$ be the indicator variable of the event that $\tilde{g}(x) \neq f(x)$. Then, $d - \alpha \leq \mathbb{E}_{\$,x}[U(x)] \leq d + \alpha$, where the expectation is over uniformly chosen $x \in \mathcal{D}$ and the randomness $\$$ for the invocations of T . Then, because x_1, \dots, x_s are uniformly chosen at random and because the invocations of A are always made with independent random coins, we can use the Chernoff bound to claim that $\Pr[|\sum_{i \in [s]} U(x_i) - ds| > 2\alpha s] < e^{-2\alpha^2 s}$. So, if we take $s = O(1/\alpha^2)$ and $\alpha = \frac{\epsilon_2 - \epsilon_1}{4}$, we are guaranteed that with probability at least $2/3$, \tilde{d} is within $2\alpha = \frac{\epsilon_2 - \epsilon_1}{2}$ of d . ■

Theorem 3 thus yields tolerant testers for the many properties already known to be self-correctable, such as the property of being a low degree polynomial over a finite field.

Another remark is that in certain cases, one can generalize Theorem 3 to obtain tolerant testers that work correctly when the input is farther away from the property than the unique decoding radius. This brings us to the notion of a *local list corrector*¹, which we formally define below.

Definition 4 (Local list corrector) *Let \mathcal{P} be a property consisting of functions over \mathcal{D} . A local list corrector for \mathcal{P} with list size L , query complexity q and radius δ is a randomized algorithm A that when given oracle access to a function f over \mathcal{D} such that $\text{dist}(f, \mathcal{P}) \leq \delta$, A^f outputs a list of randomized algorithms $\{A_1, \dots, A_L\}$ such that:*

- *For each $i \in [L]$ and every input $x \in \mathcal{D}$, the algorithm A_i^f evaluates f on at most q elements of \mathcal{D}*
- *With probability at least $2/3$, for every $g \in \mathcal{P}$ such that $\text{dist}(f, g) \leq \delta$, there exists an index $i \in [L]$ such that for every input $x \in \mathcal{D}$, $\Pr[A_i^f(x) = g(x)] \geq 2/3$.*

If \mathcal{P} has a local list corrector with list size L , query complexity q and radius δ , then for every ϵ_1, ϵ_2 such that $0 \leq \epsilon_1 < \epsilon_2 < \delta$, there exists an (ϵ_1, ϵ_2) -tolerant tester for \mathcal{P} with query complexity $O(qL \cdot \text{poly}(1/(\epsilon_2 - \epsilon_1)))$. The proof is very similar to that of Theorem 3, the main difference being that the tester searches for the function in \mathcal{P} closest to the input function f by invoking the local list corrector several times and going through all the possible elements in the list. δ can potentially be much larger than the unique decoding radius. Note though that here we want both q and L to be sublinear in n , whereas in the context of coding, L is only required to be $\text{poly}(n)$.

¹What we call “local list correctors” are more conventionally called “local list decoders.” We choose this different terminology because in the context of coding, “local list decoding” also (perhaps more often?) refers to list decoding to the message, instead of to the codeword. And so, in analogy to how “locally correctable codes” and “locally decodable codes” are related, we use the term “local list correctors” here to avoid confusion.

Acknowledgements

Other than the obvious thanks to Madhu, I also want to thank Kevin Matulef for some related discussions.