Improved Bounds on the Sign-Rank of AC^0

Mark Bun[1] Justin Thaler[2]

[1]Harvard University $\Longrightarrow$ Princeton University [2]Yahoo Labs

What is Sign-Rank?

- The sign-rank of a matrix $A = [A_{ij}]$ with entries in $\{+1, -1\}$ is the least rank of a real matrix $B = [B_{ij}]$ with $A_{ij} \cdot B_{ij} > 0$ for all i, j.

Motivation and Prior Work

Why Study Sign-Rank?

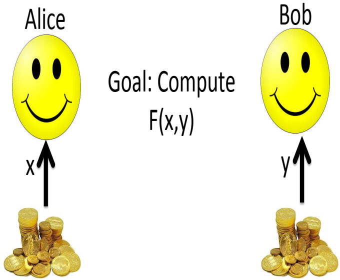
- **Learning Theory.** Sign-rank upper bounds underly the fastest known PAC learning algorithms.
  - E.g., Fastest known algorithm for PAC learning DNF formulae (Klivans and Servedio, 2003).
- **Circuit Complexity**. Sign-rank lower bounds on a matrix $A = [f(x,y)]_{x,y \in \{-1,1\}^n}$ imply lower bounds on Threshold-of-Majority circuits computing f.
- **Communication Complexity**. Sign-rank characterizes the communication model UPP^{cc} (Paturi and Simon, 1984).
  - UPP^{cc} is the most powerful communication model against which we know how to prove lower bounds.

Why Study Sign-Rank?

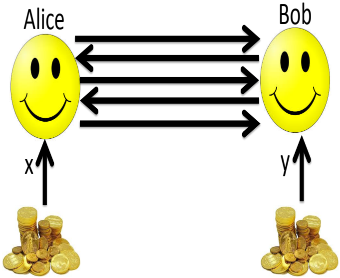
- **Learning Theory.** Sign-rank upper bounds underly the fastest known PAC learning algorithms.
  - E.g., Fastest known algorithm for PAC learning DNF formulae (Klivans and Servedio, 2003).
- **Circuit Complexity**. Sign-rank lower bounds on a matrix $A = [f(x, y)]_{x,y \in \{-1,1\}^n}$ imply lower bounds on Threshold-of-Majority circuits computing f.
- **Communication Complexity**. Sign-rank characterizes the communication model UPP^{cc} (Paturi and Simon, 1984).
  - UPP^{cc} is the most powerful communication model against which we know how to prove lower bounds.
  - It is a "communication complexity analogue" of the Turing Machine complexity class PP, which is the decisional variant of the "counting" class #P.

Definition of the UPP^{cc} Communication Model

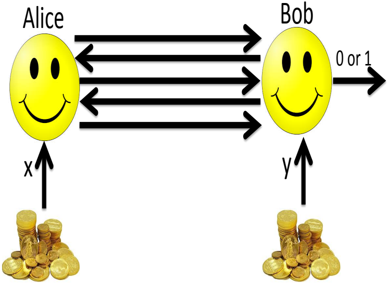
Definition of the UPP^{cc} Communication Model



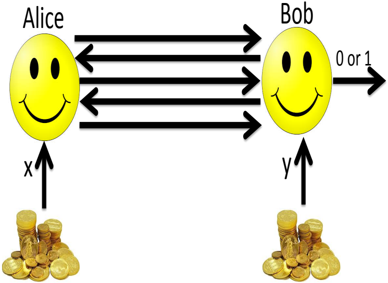
Definition of the UPP^{cc} Communication Model



Definition of the UPP^{cc} Communication Model

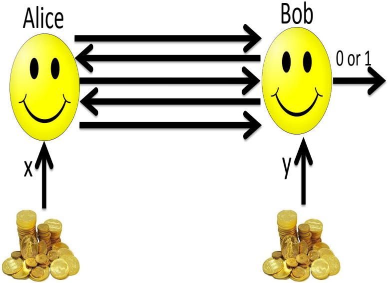


Definition of the UPP^{cc} Communication Model



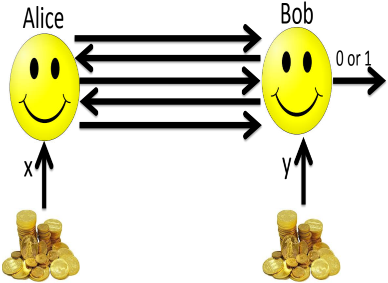
- Protocol is said to compute F if on every input (x, y), the output is correct with probability greater than $1/2$.
- The cost of a protocol is the worst-case number of bits exchanged on any input (x, y).

Definition of the UPP^{cc} Communication Model



- Protocol is said to compute F if on every input (x, y), the output is correct with probability greater than $1/2$.
- The cost of a protocol is the worst-case number of bits exchanged on any input (x, y).
- UPP$^{cc}(F)$ is the least cost of a protocol that computes F.
- UPP^{cc} is the class of all F computed by UPP^{cc} protocols of polylogarithmic cost.

Definition of the UPP^{cc} Communication Model



- Protocol is said to compute F if on every input (x, y), the output is correct with probability greater than $1/2$.
- The cost of a protocol is the worst-case number of bits exchanged on any input (x, y).
- UPP$^{cc}(F)$ is the least cost of a protocol that computes F.
- UPP^{cc} is the class of all F computed by UPP^{cc} protocols of polylogarithmic cost.
- Paturi and Simon showed that UPP$^{cc}(F) \approx \log(\text{sign-rank}(F))$.

A Brief History of Sign-Rank Lower Bounds

- Alon et al. (1985) proved optimal lower bounds on the sign rank of random matrices.
- Forster (2001) nearly-optimal lower bounds on the sign-rank of Hadamard matrices.
  - More generally, for any Boolean matrix with exponentially small spectral norm.
  - Implies optimal UPP^{cc} and circuit lower bounds on the "inner-product mod 2" function.
- Sherstov (2008) proved tight sign-rank lower bounds on symmetric predicates, i.e., matrices of the form $[D(\sum_i x_i \wedge y_i)]_{x,y \in \{-1,1\}^n}$.
- Razborov and Sherstov (2008) proved exponential sign-rank lower bound for a function in AC^0 (more context to follow).

A Motivating Question for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by #P and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathsf{PH} \subseteq \mathsf{P}^{\mathsf{PP}}$.
- But it is open how much of PH is contained in PP itself.

A Motivating Question for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by #P and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathsf{PH} \subseteq \mathsf{P}^{\mathsf{PP}}$.
- But it is open how much of PH is contained in PP itself.
- Babai, Frankl, and Simon (1986) introduced the communication analogues of Turing Machine complexity classes.
- Main question they left open was the relationship between PH^{cc} and UPP^{cc}.
  - Is $\mathsf{PH}^{cc} \subseteq \mathsf{UPP}^{cc}$?
  - Is $\mathsf{UPP}^{cc} \subseteq \mathsf{PH}^{cc}$?

Prior Work By Razborov and Sherstov (2008)

- Razborov and Sherstov (2008) resolved the first question left open by Babai, Frankl, and Simon!
- They gave a function F in PH^{cc} (actually, in Σ_2^{cc}) such that F has sign-rank $\exp(\Omega(n^{1/3}))$.
- Their proof is heavily tailored to this specific F.

Our Results

Summary of our Results

- We generalize Razborov and Sherstov's result, giving exponential sign-rank lower bounds for a broad class of functions in AC^0 and PH^{cc}.
  - Our class includes the function used by Razborov and Sherstov.
- As a corollary of our general result, we improve their lower bound on the sign-rank of Σ_2^{cc}, from $\exp(\Omega(n^{1/3}))$ to $\exp(\Omega(n^{2/5}))$.

Summary of our Results

- We generalize Razborov and Sherstov's result, giving exponential sign-rank lower bounds for a broad class of functions in AC^0 and PH^{cc}.
  - Our class includes the function used by Razborov and Sherstov.
- As a corollary of our general result, we improve their lower bound on the sign-rank of Σ_2^{cc}, from $\exp(\Omega(n^{1/3}))$ to $\exp(\Omega(n^{2/5}))$.
  - Upcoming work with Bun and Chen: Applies our methods to exhibit a problem in AM^{cc} that is not in UPP^{cc}.
  - This answers a question of Göös, Pitassi, and Watson (ICALP 2016).

Techniques

Outline for the Remainder of the Talk

- Background:
  - Threshold degree and its relation to sign rank.
  - The Pattern Matrix Method (PMM).
  - Combining PMM with "smooth dual witnesses" to prove sign-rank lower bounds.
- Our results: new construction of smooth dual witnesses, to give stronger and more general sign-rank lower bounds.

Threshold Degree

- A real polynomial p sign-represents $f\colon \{-1,1\}^n \to \{-1,1\}$ if

$$p(x) \cdot f(x) > 0 \quad \forall x \in \{-1,1\}^n$$

- $\deg_{\pm}(f) =$ minimum degree needed to sign-represent f

Communication Upper Bounds from Threshold Degree Upper Bounds

- Let $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$.
- Claim: Let $d = \deg_{\pm}(F)$. There is a UPP^{cc} protocol of cost $O(d \log n)$ computing $F(x,y)$.

Communication Upper Bounds from Threshold Degree Upper Bounds

- Let $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$.
- Claim: Let $d = \deg_{\pm}(F)$. There is a UPP^{cc} protocol of cost $O(d \log n)$ computing $F(x,y)$.
- Proof: Let $p(x,y) = \sum_{|T| \le d} c_T \cdot \chi_T(x,y)$ sign-represent F.
- Alice chooses a parity T with probability proportional to $|c_T|$, and sends to Bob T and $\mathrm{sgn}(c_T) \cdot \chi_{T \cap [n]}(y)$.
- From this, Bob can compute and output $\mathrm{sgn}(c_T) \cdot \chi_T(x,y)$.

Communication Upper Bounds from Threshold Degree Upper Bounds

- Let $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$.
- Claim: Let $d = \deg_{\pm}(F)$. There is a UPP^{cc} protocol of cost $O(d \log n)$ computing $F(x,y)$.
- Proof: Let $p(x,y) = \sum_{|T| \leq d} c_T \cdot \chi_T(x,y)$ sign-represent F.
- Alice chooses a parity T with probability proportional to $|c_T|$, and sends to Bob T and $\mathrm{sgn}(c_T) \cdot \chi_{T \cap [n]}(y)$.
- From this, Bob can compute and output $\mathrm{sgn}(c_T) \cdot \chi_T(x,y)$.
- Since p sign-represents F, the output is correct with probability strictly greater than $1/2$.
- Communication cost is clearly $O(d \log n)$.

Communication Lower Bounds from Threshold Degree Lower Bounds

- The previous slide showed that threshold degree upper bounds for $F(x, y)$ imply communication upper bounds for $F(x, y)$.
- Can we use threshold degree lower bounds for $F(x, y)$ to establish communication lower bounds for $F(x, y)$?

Communication Lower Bounds from Threshold Degree Lower Bounds

- The previous slide showed that threshold degree upper bounds for $F(x, y)$ imply communication upper bounds for $F(x, y)$.
- Can we use threshold degree lower bounds for $F(x, y)$ to establish communication lower bounds for $F(x, y)$?
- Answer: No. Bad Example: The parity function has linear threshold degree, but constant communication complexity.

Communication Lower Bounds from Threshold Degree Lower Bounds

- The previous slide showed that threshold degree upper bounds for $F(x, y)$ imply communication upper bounds for $F(x, y)$.
- Can we use threshold degree lower bounds for $F(x, y)$ to establish communication lower bounds for $F(x, y)$?
- Answer: No. Bad Example: The parity function has linear threshold degree, but constant communication complexity.
- Next Slide: Something almost as good.
  - A way to turn threshold degree lower bounds for f into communication lower bounds for a related function $F(x, y)$.

The Pattern Matrix Method (Sherstov, 2008)

- Goal: Take a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with threshold degree (at least) d, and turn it into a $2^{2n} \times 2^{2n}$ matrix F of sign-rank at least 2^d.

The Pattern Matrix Method (Sherstov, 2008)

- Goal: Take a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with threshold degree (at least) d, and turn it into a $2^{2n} \times 2^{2n}$ matrix F of sign-rank at least 2^d.
- (Sherstov, 2008) comes close to this, but falls a little short.
  - Sherstov turns f into a matrix F, called the "pattern matrix" of f, satisfying the following property:
    - Any randomized communication protocol that computes F correctly with probability $p = 1/2 + 2^{-d}$ has cost at least d.

The Pattern Matrix Method (Sherstov, 2008)

- Goal: Take a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with threshold degree (at least) d, and turn it into a $2^{2n} \times 2^{2n}$ matrix F of sign-rank at least 2^d.
- (Sherstov, 2008) comes close to this, but falls a little short.
  - Sherstov turns f into a matrix F, called the "pattern matrix" of f, satisfying the following property:
    - Any randomized communication protocol that computes F correctly with probability $p = 1/2 + 2^{-d}$ has cost at least d.
    - Note: to get a sign-rank/UPP^{cc} lower bound, we would need the above to hold for <u>any</u> $p > 1/2$.

The Pattern Matrix Method (Sherstov, 2008)

- Goal: Take a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with threshold degree (at least) d, and turn it into a $2^{2n} \times 2^{2n}$ matrix F of sign-rank at least 2^d.
- (Sherstov, 2008) comes close to this, but falls a little short.
  - Sherstov turns f into a matrix F, called the "pattern matrix" of f, satisfying the following property:
    - Any randomized communication protocol that computes F correctly with probability $p = 1/2 + 2^{-d}$ has cost at least d.
    - Note: to get a sign-rank/UPP^{cc} lower bound, we would need the above to hold for any $p > 1/2$.
  - Specifically, $F(x,y)$ is set to f evaluated at an input derived from (x,y) in a simple way.

The Pattern Matrix Method (Sherstov, 2008)

- Goal: Take a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with threshold degree (at least) d, and turn it into a $2^{2n} \times 2^{2n}$ matrix F of sign-rank at least 2^d.
- (Sherstov, 2008) comes close to this, but falls a little short.
  - Sherstov turns f into a matrix F, called the "pattern matrix" of f, satisfying the following property:
    - Any randomized communication protocol that computes F correctly with probability $p = 1/2 + 2^{-d}$ has cost at least d.
    - Note: to get a sign-rank/UPP^{cc} lower bound, we would need the above to hold for any $p > 1/2$.
  - Specifically, $F(x, y)$ is set to f evaluated at an input derived from (x, y) in a simple way.
  - y "selects" n bits of x, flips some of them, and feeds the result into f.

Proof Sketch for the Pattern Matrix Method: Dual Witnesses

- By linear programming duality: $f\colon \{-1,1\}^n \to \{-1,1\}$ has threshold degree at least $d \iff \exists$ a distribution μ on $\{-1,1\}^n$ under which f is uncorrelated with any polynomial of degree at most d.

Proof Sketch for the Pattern Matrix Method: Dual Witnesses

- By linear programming duality: $f\colon \{-1,1\}^n \to \{-1,1\}$ has threshold degree at least $d \iff \exists$ a distribution μ on $\{-1,1\}^n$ under which f is uncorrelated with any polynomial of degree at most d.
- Think of μ as a dual "witness" to the fact that the threshold degree of f is large.

Proof Sketch for the Pattern Matrix Method: Dual Witnesses

- By linear programming duality: $f\colon \{-1,1\}^n \to \{-1,1\}$ has threshold degree at least $d \iff \exists$ a distribution μ on $\{-1,1\}^n$ under which f is uncorrelated with any polynomial of degree at most d.
- Think of μ as a dual "witness" to the fact that the threshold degree of f is large.
- Sherstov shows that μ can be "lifted" into a distribution over $\{-1,1\}^{2n} \times \{-1,1\}^{2n}$ under which $F(x,y)$ cannot be computed with probability $1/2 + 2^{-d}$, unless the communication cost is at least d.

Smooth Dual Witnesses Imply Sign-Rank Lower Bounds

- Let $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfy $\deg_{\pm}(f) \geq d$.
- Razborov and Sherstov showed that if there is a dual witness μ for f that additionally satisfies a smoothness condition, then the pattern matrix F of f actually has sign-rank at least 2^d.

Smooth Dual Witnesses Imply Sign-Rank Lower Bounds

- Let $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfy $\deg_{\pm}(f) \geq d$.
- Razborov and Sherstov showed that if there is a dual witness μ for f that additionally satisfies a smoothness condition, then the pattern matrix F of f actually has sign-rank at least 2^d.
  - Specifically, μ is said to be smooth if $\mu(x) > 2^{-O(d)} \cdot 2^{-n}$ for all but a 2^{-d} fraction of inputs x.

Smooth Dual Witnesses Imply Sign-Rank Lower Bounds

- Let $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfy $\deg_{\pm}(f) \geq d$.
- Razborov and Sherstov showed that if there is a dual witness μ for f that additionally satisfies a smoothness condition, then the pattern matrix F of f actually has sign-rank at least 2^d.
  - Specifically, μ is said to be smooth if $\mu(x) > 2^{-O(d)} \cdot 2^{-n}$ for all but a 2^{-d} fraction of inputs x.
- The bulk of Razborov-Sherstov is showing that there is a DNF formula f with large threshold degree and smooth dual witness to this fact.
- Since f is computed by a DNF formula, its pattern matrix is easily seen to be in Σ_2^{cc}.

A DNF With a Smooth Dual Witness

- Minsky and Papert (1969) famously exhibited a DNF formula $f = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ with threshold degree $\Omega(n^{1/3})$.
- Razborov and Sherstov show that f has a smooth dual witness to this fact.
  - They did not explicitly construct the smooth dual witness; they just showed that one exists.

Our Results: New Construction of Smooth Duals

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.
- Examples:
  - AND_k is in $\mathcal{C}_d$ for $d = \sqrt{k}$.
  - By setting $k = n^{2/3}$, we recover Razborov and Sherstov's result for the function $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.
- Examples:
  - AND_k is in $\mathcal{C}_d$ for $d = \sqrt{k}$.
  - By setting $k = n^{2/3}$, we recover Razborov and Sherstov's result for the function $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$.
  - The function ED_k is in $\mathcal{C}_d$ for $d = k^{2/3}$.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.
- Examples:
  - AND_k is in $\mathcal{C}_d$ for $d = \sqrt{k}$.
  - By setting $k = n^{2/3}$, we recover Razborov and Sherstov's result for the function $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$.
  - The function ED_k is in $\mathcal{C}_d$ for $d = k^{2/3}$.
  - ED_k is computed by a CNF with logarithmic bottom fan-in $\Longrightarrow$ its pattern matrix is in Σ_2^{cc}.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.
- Examples:
  - AND_k is in $\mathcal{C}_d$ for $d = \sqrt{k}$.
  - By setting $k = n^{2/3}$, we recover Razborov and Sherstov's result for the function $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$.
  - The function ED_k is in $\mathcal{C}_d$ for $d = k^{2/3}$.
  - ED_k is computed by a CNF with logarithmic bottom fan-in $\Longrightarrow$ its pattern matrix is in Σ_2^{cc}.
  - Let $f = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$. Our result implies that there is smooth dual witness for the fact that $\deg_{\pm}(f) \geq n^{2/5}$.

Our Results and Methods

- Let OR_d denote the OR function on d bits.
- We identify a class of functions $\mathcal{C}_d$, so that if $h \in \mathcal{C}_d$, then there is a smooth dual witness for the fact that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$.
- Roughly, $\mathcal{C}_d$ corresponds to the set of functions h that cannot be uniformly approximated to error $1/3$ by degree d polynomials.
- Examples:
  - AND_k is in $\mathcal{C}_d$ for $d = \sqrt{k}$.
  - By setting $k = n^{2/3}$, we recover Razborov and Sherstov's result for the function $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$.
  - The function ED_k is in $\mathcal{C}_d$ for $d = k^{2/3}$.
  - ED_k is computed by a CNF with logarithmic bottom fan-in $\Longrightarrow$ its pattern matrix is in Σ_2^{cc}.
  - Let $f = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$. Our result implies that there is smooth dual witness for the fact that $\deg_{\pm}(f) \geq n^{2/5}$.
  - Hence, the pattern matrix of f has sign-rank $\exp(\Omega(n^{2/5}))$.

Smooth Witness Construction

- (Sherstov, STOC 2014) gave a dual witness showing that $\deg_{\pm}(\mathrm{OR}_d \circ h) \geq d$ for any $h \in \mathcal{C}_d$.
- But his dual witness isn't smooth.
- We substantially modify his construction to give a smooth dual witness.