Soft MIMO Detection at Fixed Complexity

Erik G. Larsson
Linköping University (LiU), Sweden
Dept. of Electrical Engineering (ISY)
Division of Communication Systems
www.commsys.isy.liu.se

(work with J. Jaldén, TU Vienna)
Model and problem formulation

\[ \mathbf{y} = \mathbf{H} \cdot \mathbf{s} + \mathbf{e}, \quad s_k = S(b_1, ..., b_m) \in S, \quad |S| = 2^m \]

\( \Rightarrow \) Bits \( b_i \) a priori indep. with \( L(b_i) = \log \left( \frac{P(b_i=1)}{P(b_i=0)} \right) \)

\( \Rightarrow \) Objective: Determine \( L(b_i|\mathbf{y}) = \log \left( \frac{P(b_i=1|\mathbf{y})}{P(b_i=0|\mathbf{y})} \right) \)

\( \Rightarrow \) Applications: multiuser detection, ISI, MIMO, crosstalk in cables, ...

The optimum receiver

\[ L(b_i | y) = \log \left( \frac{P(b_i = 1 | y)}{P(b_i = 0 | y)} \right) = \log \left( \frac{\sum_{s: b_i(s) = 1} P(s | y)}{\sum_{s: b_i(s) = 0} P(s | y)} \right) \]

\( \geq \log \left( \frac{\sum_{s: b_i(s) = 1} P(y | s) P(s)}{\sum_{s: b_i(s) = 0} P(y | s) P(s)} \right) \)

\( = \log \left( \frac{\sum_{s: b_i(s) = 1} p(y | s) \left( \prod_{i' = 1}^{nt} P(b_{i'} = b_{i'}(s)) \right)}{\sum_{s: b_i(s) = 0} p(y | s) \left( \prod_{i' = 1}^{nt} P(b_{i'} = b_{i'}(s)) \right)} \right) \cdot P(b_i = 1) \)

\( = \log \left( \frac{\sum_{s: b_i(s) = 1} p(y | s) \left( \prod_{i' = 1, i' \neq i}^{nt} P(b_{i'} = b_{i'}(s)) \right) \cdot P(b_i = 1)}{\sum_{s: b_i(s) = 0} p(y | s) \left( \prod_{i' = 1, i' \neq i}^{nt} P(b_{i'} = b_{i'}(s)) \right) \cdot P(b_i = 0)} \right) + L(b_i) \)

In Gaussian noise \( p(y | s) = \frac{1}{(2\pi \sigma^2)^{nr}} \exp \left( -\frac{1}{\sigma^2} \| y - Hs \|^2 \right) \) so

\[ L(b_i | y) = \log \left( \frac{\sum_{s: b_i(s) = 1} \exp \left( -\frac{1}{N_0} \| y - Hs \|^2 \right) \left( \prod_{i' = 1, i' \neq i}^{nt} P(b_{i'} = b_{i'}(s)) \right)}{\sum_{s: b_i(s) = 0} \exp \left( -\frac{1}{N_0} \| y - Hs \|^2 \right) \left( \prod_{i' = 1, i' \neq i}^{nt} P(b_{i'} = b_{i'}(s)) \right)} \right) + L(b_i) \]
With a priori equiprobable bits

\[
L(b_i | y) = \log \left( \frac{\sum_{s : b_i(s) = 1} \exp \left( -\frac{1}{\sigma^2} \| y - Hs \|^2 \right)}{\sum_{s : b_i(s) = 0} \exp \left( -\frac{1}{\sigma^2} \| y - Hs \|^2 \right)} \right)
\]

Problem: direct computation has complexity \(O(2^{mmt})\)!

Needs be solved \textit{in real time} (once per received \(y\))

Preferably, approximation solution should have parallel structure

For \(H \propto \) unitary (OSTBC), the problem is trivial.

- Focus on unstructured \(H\)
- Generally, slow fading (no time diversity) is the hard case
Approximations

**ZF approximation.** Use a ZF linear filter to obtain

\[ \tilde{y} \triangleq (H^H H)^{-1} H y = s + (H^H H)^{-1} H e = s + \tilde{e} \]

If we neglect the noise correlation this leads to a set of *scalar* channels

\[ \tilde{y}_k = s_k + \tilde{e}_k, \quad e_k \sim N(0, \sigma^2 \cdot [(H^H H)^{-1}]_{k,k}) \]

Then treat \( s \) componentwise,

\[
L(b_i|y) \approx \log \left( \frac{\sum_{s_k: b_i(s_k)=1} \exp \left( -\frac{|\tilde{y}_k - s_k|^2}{\sigma^2 \cdot [(H^H H)^{-1}]_{k,k}} \right)}{\sum_{s_k: b_i(s_k)=0} \exp \left( -\frac{|\tilde{y}_k - s_k|^2}{\sigma^2 \cdot [(H^H H)^{-1}]_{k,k}} \right)} \right)
\]

This can be acceptable in fast fading, but very poor in slow fading.
Log-max approximation

\[ L(b_i|y) \approx \log \left( \frac{\max_{s:b_i(s)=1} \exp \left( -\frac{1}{\sigma^2} \| y - Hs \|^2 \right)}{\max_{s:b_i(s)=0} \exp \left( -\frac{1}{\sigma^2} \| y - Hs \|^2 \right)} \right) \]

Must solve ICLS problem: \( \min_{s \in \mathbb{S}_{nt}} \| y - Hs \| \) \( \Rightarrow \) NP-hard

- Many methods that find optimum with (more or less) high probability:
  - ZF/MMSE, ZF/MMSE-DFE
  - sphere decoding, lattice reduction, SD relaxation

- Generally, these methods are either very complex or not very good. Many are sequential in nature, cannot be parallelized and suffer from random runtime.
  - Acceptable if we find the correct solution quickly, and with high probability

List decoders (e.g., list sphere decoding)

- Retain terms in \( \sum \) encountered in search for \( s \)


Our approach

Consider again $L(b_k|y) = \log \left( \frac{\sum_{s \in \mathcal{S}} b_k(s) = 1 \exp \left( -\frac{1}{\sigma^2} \| y - H s \|^2 \right)}{\sum_{s \in \mathcal{S}} b_k(s) = 0 \exp \left( -\frac{1}{\sigma^2} \| y - H s \|^2 \right)} \right)$

$= \log \left( \frac{\sum_{b_1} \cdots \sum_{b_{k-1}} \sum_{b_{k+1}} \cdots \sum_{b_{ntm}} \mu(b_1, \ldots, b_{k-1}, 1, b_{k+1}, \ldots, b_{ntm})}{\sum_{b_1} \cdots \sum_{b_{k-1}} \sum_{b_{k+1}} \cdots \sum_{b_{ntm}} \mu(b_1, \ldots, b_{k-1}, 0, b_{k+1}, \ldots, b_{ntm})} \right)$

where $\mu(b_1, \ldots, b_{ntm}) \triangleq \exp \left( -\frac{1}{\sigma^2} \| y - H s(b_1, \ldots, b_{ntm}) \|^2 \right)$

Exact marginalization over $r$ selected bits: $\sum \cdots \sum$

Approximate log-max over rest $(mn_t - r)$: $\sum \cdots \sum \approx \max(\cdot)$

Carefully select bits for exact marginalization
Implementation

✧ Choose an index permutation (bit ordering) \( \mathcal{I} \)

✧ Let \( l \) be the unique integer such that \( \mathcal{I}_l = k \). For \( k \notin \{\mathcal{I}_1, \ldots, \mathcal{I}_r\} \) we use

\[
\log \left( \sum_{b_{\mathcal{I}_1}=0}^{1} \cdots \sum_{b_{\mathcal{I}_r}=0}^{1} \left( \prod_{i=1}^{r} \max_{b_{\mathcal{I}_{i+1}}, \ldots, b_{\mathcal{I}_{l-1}}, b_{\mathcal{I}_{l+1}}, \ldots, b_{\mathcal{I}_{n_{tm}}}} \mu(b_1, \ldots, b_{k-1}, 1, b_{k+1}, \ldots, b_{n_{tm}}) \right) \right)
\]

✧ For \( k \in \{\mathcal{I}_1, \ldots, \mathcal{I}_r\} \)

\[
\log \left( \sum_{b_{\mathcal{I}_1}=0}^{1} \cdots \sum_{b_{\mathcal{I}_{l-1}}=0}^{1} \sum_{b_{\mathcal{I}_{l+1}}=0}^{1} \cdots \sum_{b_{\mathcal{I}_r}=0}^{1} \left( \prod_{i=1}^{r} \max_{b_{\mathcal{I}_{i+1}}, \ldots, b_{\mathcal{I}_{n_{tm}}}} \mu(b_1, \ldots, b_{k-1}, 1, b_{k+1}, \ldots, b_{n_{tm}}) \right) \right)
\]
Log-max problems solved approximately: When solving
\[
\max_{b_{I_{r+1}}, \ldots, b_{I_{ntm}}} \mu(\ldots, b_{l-1}, 1, b_{l+1}, \ldots)
\]
\(r\) of the bits are fixed.

Partition \(H\) and \(y\). For BPSK per real dimension:
\[
\begin{align*}
H_a & \triangleq [h_{I_1}, \ldots, h_{I_r}], & H_b & \triangleq [h_{I_{r+1}}, \ldots, h_{I_{nt}}] \\
S_a & \triangleq [s_{I_1}, \ldots, s_{I_r}]^T, & S_b & \triangleq [s_{I_{r+1}}, \ldots, s_{I_{nt}}]^T \\
y & = H_a S_a + H_b S_b + e
\end{align*}
\]

We solve via ZF-DFE (or a suboptimal method of choice)
\[
\min_{\hat{S}_b \in S_{ntm-r}} \|y - H_a S_a - H_b \hat{S}_b\|^2
\]
Implementation, cont.

Choosing $\mathcal{I}$:

- We use the order that increases $\kappa(H)$ fastest when columns are removed

- Other possible (suboptimal) orderings:
  - sorting of $[(H^H H)^{-1}]_{k,k}$,
  - random/natural order

Choosing $r$:

- Larger $r$ $\implies$ smaller $\kappa(H_b)$ $\implies$ max$(\cdot)$ problems solved more accurately

- $r = 0$: only approximate marginalization

- $r = n_t m$: equivalent to brute-force

- Complexity is $O(2^r)$, so ideally should get good performance for small $r$
Some analysis

Write $Q_b R_b = H_b$ and $\tilde{y} \triangleq y - H_a s_a$. Then for any $\hat{s}_b$

$$\|\tilde{y} - H_b \hat{s}_b\|^2 = \|Q_b^T \tilde{y} - R_b \hat{s}_b\|^2 + \|\Pi_{H_b} \tilde{y}\|^2$$

Thus

$$\frac{\|\tilde{y} - H_b \hat{s}_b, ZF\|^2}{\|\tilde{y} - H_b \hat{s}_b, ML\|^2} = \frac{\|Q_b^T \tilde{y} - R_b \hat{s}_b, ZF\|^2}{\|\Pi_{H_b} \tilde{y}\|^2} + \frac{1}{\|Q_b^T \tilde{y} - R_b \hat{s}_b, ML\|^2 + 1} \geq 1$$

We can show that

$$\|Q_b^T \tilde{y} - R_b \hat{s}_b, ML\|^2 \leq \|Q_b^T \tilde{y} - R_b \hat{s}_b, ZF\|^2 \leq \kappa(R_b) \|Q_b^T \tilde{y} - R_b \hat{s}_b, ML\|^2$$

and $\kappa(R_b)$ decreases when $r$ is increased

Additionally, no. of cols in $H_b$ decr. with $r$ \Rightarrow $\|\Pi_{H_b} \tilde{y}\|^2$ incr. with $\nu$
Soft MIMO detection at fixed complexity

$4 \times 4$ slow Rayleigh fading MIMO, QPSK, rate 1/3 CC

![Graph showing frame-error-rate (FER) versus normalized signal-to-noise-ratio (SNR) for different values of $r$. The graph compares ZF, Exact, LogMax, and Proposed methods.](image)
6 × 6 slow Rayleigh fading MIMO, QPSK
Extension 1: Optimal Metric for Imperfect CSI

(assert)

\(H\) not known perfectly \(\Rightarrow\) replacing \(H\) with \(\hat{H}\) in \(p(y|s, H)\) not optimal!

(assert)

Instead, need to work with \(p(y|s, \hat{H})\). Write

\[
y = Hs + e \iff y = (s^T \otimes I)h + e, \quad h = \text{vec}(H), \quad e = \text{vec}(E)
\]

Suppose \(\|s\|^2 = n_t\) and

\[
\begin{align*}
\{ & h \sim N(0, \rho^2 I), \quad e \sim N(0, \sigma^2 I) \\
& \hat{h} = h + \delta, \quad \delta \sim N(0, \epsilon^2 I)
\}
\]

Then

\[
\begin{bmatrix} y \\ \hat{h} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ \rho^2(s^* \otimes I) \end{bmatrix}, \begin{bmatrix} (n_t\rho^2 + \sigma^2)I & \rho^2(s^* \otimes I) \\ \rho^2(s^T \otimes I) & (\rho^2 + \epsilon^2)I \end{bmatrix} \right)
\]

so

\[
p(y|\hat{h}, s) = \frac{1}{\pi^n n_t \pi_t^2} \frac{\frac{1}{n_t \epsilon^2} + \rho^2}{\frac{1}{1+\epsilon^2/\rho^2} + \sigma^2} \exp\left( -\frac{1}{\frac{n_t \epsilon^2}{1+\epsilon^2/\rho^2} + \sigma^2} \left\| y - \left( \frac{\rho^2}{\rho^2 + \epsilon^2} \right) \hat{H}s \right\|^2 \right)
\]
4 × 4 slow Rayleigh fading MIMO, QPSK, est. \( H \)
$4 \times 4$ slow Rayleigh fading MIMO, QPSK, outdated $H$
Consider $s_k \in \{\pm 1\}$, and let

$$s_k = 2b_k - 1$$

$$\gamma_k \triangleq \frac{1}{2} \log (P(s_k = -1) P(s_k = 1)) = \frac{1}{2} \log (P(b_k = 0) P(b_k = 1))$$

$$\lambda_k \triangleq \log \left( \frac{P(s_k = 1)}{P(s_k = -1)} \right) = \log \left( \frac{P(b_k = 1)}{P(b_k = 0)} \right) = L(b_k)$$

Then the prior is linear in $s_k$:

$$\log (P(s_k = s)) = \frac{1}{2} \left\{ (1 + s) \log (P(s_k = 1)) + (1 - s) \log (P(s_k = -1)) \right\}$$

$$= \frac{1}{2} \gamma_k + \frac{1}{2} \lambda_k s_k$$
Write

\[
L(s_k|y) = \log \left( \frac{\sum_{s:s_k=1} \exp \left( -\frac{1}{\sigma^2} \| y - H s \|^2 + \frac{1}{2} \sum_{i=1,i\neq k}^{n_t} (\gamma_i + \lambda_i s_i) \right)}{\sum_{s:s_k=0} \exp \left( -\frac{1}{\sigma^2} \| y - H s \|^2 + \frac{1}{2} \sum_{i=1,i\neq k}^{n_t} (\gamma_i + \lambda_i s_i) \right)} \right) + \lambda_k
\]

Define \( \tilde{y} \triangleq [y^T 1 \cdots 1]^T \) and \( \tilde{H} \triangleq \begin{bmatrix} H \\ \Lambda_k \end{bmatrix} \) where

\[
\Lambda_k \triangleq \text{diag} \left\{ \frac{\sigma^2}{4} \lambda_1, \cdots, \frac{\sigma^2}{4} \lambda_{k-1}, \frac{\sigma^2}{4} \lambda_{k+1}, \cdots, \frac{\sigma^2}{4} \lambda_n \right\}
\]

Then

\[
L(s_k|y) = \log \left( \frac{\sum_{s:s_k=1} \exp \left( -\frac{1}{\sigma^2} \| \tilde{y} - \tilde{H}s \|^2 + \sum_{i=1,i\neq k}^{n_t} \left( \frac{\sigma^2 \lambda_i^2}{16} + \frac{\gamma_i}{2} \right) \right)}{\sum_{s:s_k=0} \exp \left( -\frac{1}{\sigma^2} \| \tilde{y} - \tilde{H}s \|^2 + \sum_{i=1,i\neq k}^{n_t} \left( \frac{\sigma^2 \lambda_i^2}{16} + \frac{\gamma_i}{2} \right) \right)} \right) + \lambda_k
\]

A priori information on \( s_k \) ➾ “virtual antennas”
Iter. decod.  $4 \times 4$ MIMO, $r = 1/2$-LDPC, 1000 bits

Frame-error-rate (FER) vs. Normalized signal-to-noise-ratio (SNR) [dB]

- Proposed, $r = 3$, no iteration
- Proposed, $r = 3$, 1 iteration
- Proposed, $r = 3$, 2 iterations
- Brute-force, no iteration
- Brute-force, 1 iteration
- Brute-force, 2 iterations
**Conclusions**

- Method for soft detection in MIMO systems
- Approximating $L(b_k|y)$ rather than hard detection of $s$
- Merits:
  - has fixed (non-random) complexity,
  - provides very good performance at low complexity,
  - is suitable for massively parallel hardware architectures
- Performance/complexity tradeoff by choosing $r$.
- Especially suitable for slow fading
  - precomputation of various inverses amortized over many $y$
  - slow fading is most critical (no time diversity)
Thank You
Backup slides
Some background on ICLS

Let \( H = QL \) where

- \( Q \in \mathbb{R}^{n_r \times n_t} \) is orthonormal \((Q^TQ = I)\), and
- \( L \in \mathbb{R}^{n_t \times n_t} \) lower triangular:

\[
\begin{bmatrix}
L_{11} & 0 & 0 \\
L_{21} & L_{22} & 0 \\
L_{31} & L_{32} & L_{33}
\end{bmatrix}
\]

\[
\min_{s \in S_{nt}} \|y - Hs\|^2 \iff \min_{s \in S_{nt}} \|\tilde{y} - Ls\| \quad \text{where} \quad \tilde{y} \triangleq Q^Ty
\]

Note that \( \|\tilde{y} - Ls\|^2 = f_1(s_1) + f_2(s_1, s_2) + \cdots + f_n(s_1, \ldots, s_n) \)

where \( f_k(s_1, \ldots, s_k) \triangleq \|(\tilde{y} - Ls)_k\|^2 = \left(\tilde{y}_k - \sum_{l=1}^{k} L_{k,l} s_l\right)^2 \)
Soft MIMO Detection at Fixed Complexity

Decision tree view

root node

\[ s_1 = -1 \]
\[ f_1(-1) = 1 \]
\[ f_1(1) = 5 \]

\[ s_2 = -1 \]
\[ f_2(-1, -1) = 2 \]
\[ f_2(-1, 1) = 1 \]

\[ s_3 = -1 \]
\[ f_3(\cdots) = 4 \]

leaves

\[ \{1, -1, -1\} \]
\[ \{1, -1, 1\} \]
\[ \{1, 1, -1\} \]
\[ \{1, 1, 1\} \]
Zero-Forcing

Let
\[ \tilde{s} \triangleq \arg \min_{s \in \mathbb{R}^n} \| y - Hs \| = \arg \min_{s \in \mathbb{R}^n} \| \tilde{y} - Ls \| = L^{-1} \tilde{y} \]

E.g., Gaussian elimination: \( \tilde{s}_1 = \tilde{y}_1 / L_{1,1} \)
\[
\tilde{s}_2 = (\tilde{y}_2 - \tilde{s}_1 L_{2,1}) / L_{2,2}
\]

\[ \vdots \]

Then project onto \( S \): \( \hat{s}_k = [\tilde{s}_k] \triangleq \arg \min_{s_k \in S} |s_k - \tilde{s}_k| \)

This works very poorly. Why?
Neglects correlation between the noises in \( \tilde{s}_k \) (\( \text{cov}(\tilde{s}) = \Sigma \triangleq \sigma \cdot (L^T L)^{-1} \))
Zero-Forcing with Decision Feedback

Consider the following improvement

i) Detect $s_1$ via: $\hat{s}_1 = \begin{bmatrix} \tilde{y}_1 \\ \frac{L_{1,1}}{L_{1,1}} \end{bmatrix} = \arg \min_{s_1 \in S} f_1(s_1)$

ii) Consider $s_1$ known and set $\hat{s}_2 = \begin{bmatrix} \tilde{y}_2 - \hat{s}_1 L_{2,1} \\ L_{2,2} \end{bmatrix} = \arg \min_{s_2 \in S} f_2(\hat{s}_1, s_2)$

iii) Continue for $k = 3, ..., n$:

$$\hat{s}_k = \begin{bmatrix} \tilde{y}_k - \sum_{l=1}^{k-1} L_{k,l} \hat{s}_l \\ L_{k,k} \end{bmatrix} = \arg \min_{s_k \in S} f_k(\hat{s}_1, ..., \hat{s}_{k-1}, s_k)$$

This works poorly too. Why? Error propagation. Incorrect decision on $s_i \Rightarrow$ most of the follows $s_k$ wrong as well.

Optimized detection order (start with the best) does not help (much).
ZF-DF
**Sphere decoding**

- Select a sphere radius, $R$. Then traverse the tree, but once encountering a node with cumulative metric $> R$, do not follow it down.

- Enumerates all leaf nodes which lie inside the sphere $\|\tilde{y} - Ls\| \leq R$.

- Improvements:
  - Pruning: At each leaf, update $R$ according to $R := \min(R, M)$.
  - Improvements: optimal ordering of $s_k$.
  - Branch enumeration (e.g., $s_k = \{-5, -3, -1, -1, 3, 5\}$ vs. $s_k = \{-1, 1, -3, 3, -5, 5\}$).

- Known facts:
  - The algorithm solves the problem, if allowed to finish.
  - Runtime is random and algorithm cannot be parallelized.
  - Under relevant circumstances, average runtime is $O(2^{\alpha n})$ for $\alpha > 0$. 


SD, without pruning, $R = 6$
SD, with pruning, $R = \infty$
“Fixed complexity” sphere decoding

- Select a user parameter $r$, $0 \leq r \leq n$

- For each node on layer $r$, consider $\{s_1, \ldots, s_r\}$ fixed and solve

\[
(*) \quad \min_{\{s_{r+1}, \ldots, s_n\}} \{ f_{r+1}(s_1, \ldots, s_{r+1}) + \cdots + f_n(s_1, \ldots, s_n) \}
\]

- Subproblem (*) solved using $|S|^r$ times

- Low-complexity approximation (e.g. ZF-DF) can be used. Why? (*) is overdetermined (equivalent $H$ is tall)

- Order can be optimized: start with the “worst”

- Fixed runtime, fully parallel structure
$\text{FCSD, } r = 1$
Semidefinite relaxation (for $s_k \in \{\pm 1\}$)

Let $\bar{s} \triangleq \begin{bmatrix} s \\ 1 \end{bmatrix}$, $S \triangleq \bar{s}\bar{s}^T = \begin{bmatrix} s \\ 1 \end{bmatrix} \begin{bmatrix} s^T & 1 \end{bmatrix}$, $\Psi \triangleq \begin{bmatrix} L^T L & -L^T \tilde{y} \\ -\tilde{y}^T L & 0 \end{bmatrix}$

Then

$$\|\tilde{y} - Ls\|^2 = \bar{s}^T \Psi \bar{s} + \|\tilde{y}\|^2 = \text{Trace}\{\Psi S\} + \|\tilde{y}\|^2$$

so the problem is to

$$\min_{\text{diag}\{s\} = \{1, \ldots, 1\}} \text{Trace}\{\Psi S\}$$

rank\{s\} = 1

$\bar{s}_{n+1} = 1$

SDR proceeds by relaxing $\text{rank}\{S\} = 1$ to $S$ positive semidefinite

Interior point methods used to find $S$

$s$ recovered, e.g., by taking dominant eigenvector and project onto $S_{nt}^{32}$
**Lattice reduction**

- Extend $S_{nt}$ to lattice. For example, if $S = \{-3, -1, 1, 3\}$, then $\tilde{S}_{nt} = \{\ldots, -3, -1, 1, 3, \ldots\} \times \cdots \times \{\ldots, -3, -1, 1, 3, \ldots\}$.

- Decide on orthogonal integer matrix $T \in \mathbb{R}^{n \times n}$ that maps $\tilde{S}_{nt}$ onto itself:
  \[ T_{k,l} \in \mathbb{Z}, \quad T^T T \propto I, \quad \text{and} \quad Ts \in \tilde{S}_{nt} \quad \forall s \in \tilde{S}_{nt} \]

- Find one such $T$ for which $LT \propto I$

- Then solve $\hat{s}' \triangleq \arg \min_{s' \in \tilde{S}^n} \|\tilde{y} - (LT)s'\|^2$, and set $\hat{s} = Ts'$

- Critical steps:
  - Find suitable $T$ (computationally costly, but amortize over many $y$)
  - $\hat{s} \in \tilde{S}_{nt}$, but $\hat{s} \notin S_{nt}$ in general, so clipping is necessary