High-Rate Sparse Superposition Codes with Iteratively Optimal Estimates

Andrew Barron, Sanghee Cho
Department of Statistics
Yale University

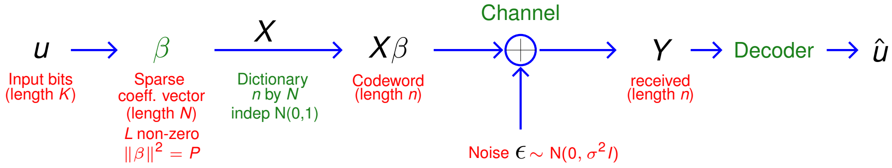
2012 IEEE International Symposium on Information Theory
July 2, 2012 MIT

Sparse Superposition Code for the Gaussian Channel

$$u \longrightarrow \beta \xrightarrow{\quad X \quad} X\beta \longrightarrow \overset{\text{Channel}}{\oplus} \longrightarrow Y \longrightarrow \text{Decoder} \longrightarrow \hat{u}$$

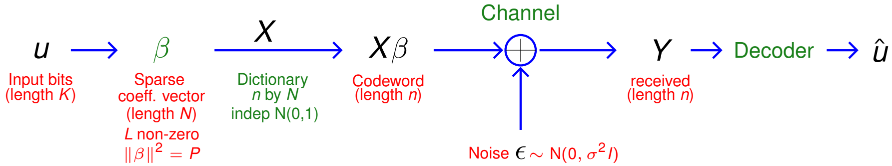
- u: Input bits (length K)
- β: Sparse coeff. vector (length N), L non-zero, $\|\beta\|^2 = P$
- X: Dictionary n by N indep N(0,1)
- $X\beta$: Codeword (length n)
- Channel: Noise $\epsilon \sim$ N(0, $\sigma^2 I$)
- Y: received (length n)

Sparse Superposition Code for the Gaussian Channel



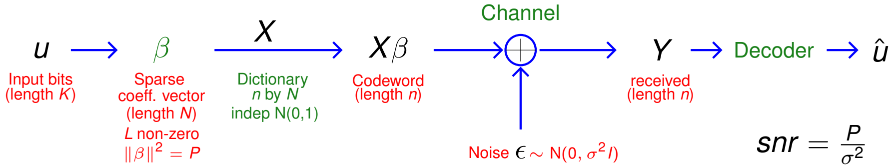
Linear Model $\quad Y = X\beta + \epsilon$

Sparse Superposition Code for the Gaussian Channel



- Partitioned Coef.: $\beta = (00{*}0000,\ 000{*}000, \ldots, 0{*}00000)$
- L sections of size $M = N/L$, one non-zero in each

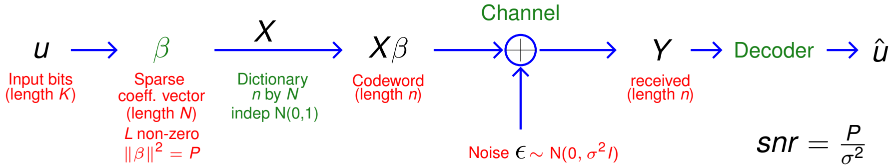
Sparse Superposition Code for the Gaussian Channel



$$snr = \frac{P}{\sigma^2}$$

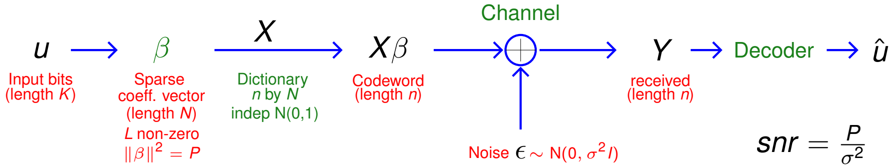
- Partitioned Coef.: $\beta = (00{*}0000,\ 000{*}000, \ldots, 0{*}00000)$
- L sections of size $M = N/L$, one non-zero in each
- Rate $R = \frac{K}{n} = \frac{L \log M}{n}$, Capacity $C = \frac{1}{2}\log(1 + snr)$

Sparse Superposition Code for the Gaussian Channel



- Partitioned Coef.: $\beta = (00{*}0000,\ 000{*}000, \ldots, 0{*}00000)$
- L sections of size $M = N/L$, one non-zero in each
- Rate $R = \frac{K}{n} = \frac{L \log M}{n}$, Capacity $C = \frac{1}{2}\log(1 + snr)$
- Ultra-sparse case: Impractical $M = 2^{nR/L}$ with L constant
  (successive decoder reliable for $R < C$: Cover 1972 IT)
- Moderately-sparse: $M = L^a$ with $n = (L \log M)/R$

Sparse Superposition Code for the Gaussian Channel



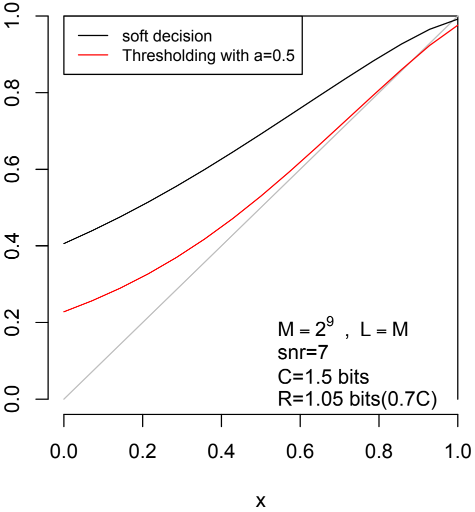
- Partitioned Coef.: $\beta = (00{*}0000, 000{*}000, \ldots, 0{*}00000)$
- L sections of size $M = N/L$, one non-zero in each
- Rate $R = \frac{K}{n} = \frac{L \log M}{n}$, Capacity $C = \frac{1}{2}\log(1 + snr)$
- Ultra-sparse case: Impractical $M = 2^{nR/L}$ with L constant
  (successive decoder reliable for $R < C$: Cover 1972 IT)
- Moderately-sparse: $M = L^a$ with $n = (L \log M)/R$
  (reliable for $R < C$)

  Maximum likelihood decoder (Joseph & Barron 2010a ISIT, 2012a IT)

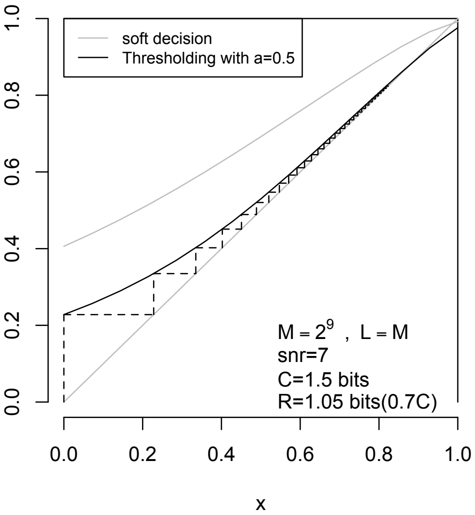
  Adaptive successive decoder with threshold (J&B 2010b ISIT, 2012b)

  Adaptive successive decoder with soft decision (B&C, this talk)

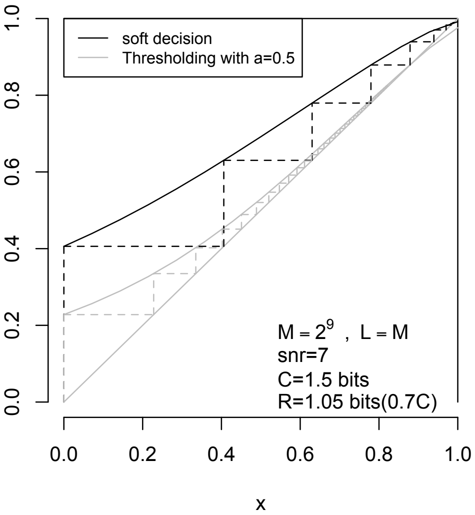
Progression of success rate



Progression of success rate



Progression of success rate

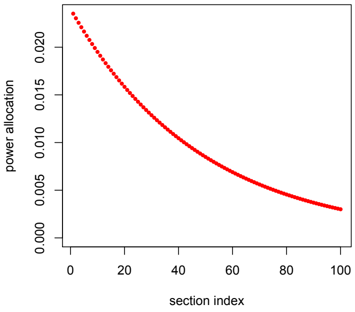


Power Allocation

- Power control: $\sum_{\ell=1}^{L} P_\ell = P \qquad \|\beta\|^2 = P$
- Special choice: P_ℓ proportional to $e^{-2C\ell/L}$ for $\ell = 1, \ldots, L$

Power Allocation

- Power control: $\sum_{\ell=1}^{L} P_\ell = P \qquad \|\beta\|^2 = P$
- Special choice: P_ℓ proportional to $e^{-2C\ell/L}$ for $\ell = 1, \ldots, L$



Coefficient vectors β

- Power control: $\sum_{\ell=1}^{L} P_\ell = P \qquad\qquad \|\beta\|^2 = P$
- Special choice: P_ℓ proportional to $e^{-2C\ell/L}$ for $\ell = 1, \ldots, L$
- Coeff. sent: $\beta \; = (00\sqrt{P_1}0000, 000\sqrt{P_2}000, \ldots, 0\sqrt{P_L}00000)$
- Terms sent: $(j_1, j_2, \ldots, j_L)$
- $\beta_j = \sqrt{P_\ell}\,1_{\{j=j_\ell\}}$ for j in section ℓ , for $\ell = 1, \ldots, L$
- $\mathcal{B} =$ set of such allowed vectors β for codewords $X\beta$

Coefficient Estimates $\hat{\beta}$

- Power control: $\sum_{\ell=1}^{L} P_\ell = P \qquad \|\beta\|^2 = P$
- Special choice: P_ℓ proportional to $e^{-2C\ell/L}$ for $\ell = 1, \ldots, L$
- Coeff. sent: $\beta \; = (00\sqrt{P_1}0000, 000\sqrt{P_2}000, \ldots, 0\sqrt{P_L}00000)$
- Terms sent: $(j_1, j_2, \ldots, j_L)$
- $\beta_j = \sqrt{P_\ell}\, 1_{\{j=j_\ell\}}$ for j in section ℓ , for $\ell = 1, \ldots, L$
- $\mathcal{B} =$ set of such allowed vectors β for codewords $X\beta$
- $\hat{\beta}_j$ restricted to $\mathcal{B}$ or the convex hull of $\mathcal{B}$
- $\hat{\beta}_j = \sqrt{P_\ell}\, \hat{w}_j$ for j in sec_ℓ , with $\hat{w}_j \geq 0$, $\sum_{j \in sec_\ell} \hat{w}_j = 1$

Iterative Estimation

For $k \geq 1$

- Coefficient fits: $\hat{\beta}_{k,j}$ (initially 0)
- Codeword fits: $F_k = X\hat{\beta}_k$
- Vector of statistics: $stat_k =$ function of $(X, Y, F_1, \ldots, F_k)$
- e.g. $stat_{k,j}$ proportional to $X_j^T(Y - F_k)$
- Update $\hat{\beta}_{k+1}$ as a function of $stat_k$

Iterative Estimation

For $k \geq 1$

- Coefficient fits: $\hat{\beta}_{k,j}$ (initially 0)
- Codeword fits: $F_k = X\hat{\beta}_k$
- Vector of statistics: $stat_k =$ function of $(X, Y, F_1, \ldots, F_k)$
- e.g. $stat_{k,j}$ proportional to $X_j^T(Y - F_k)$
- Update $\hat{\beta}_{k+1}$ as a function of $stat_k$
  - Thresholding: Adaptive Successive Decoder
    $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}$ if $stat_{k,j}$ is above threshold in sections ℓ not previously decoded

Iterative Estimation

For $k \geq 1$

- Coefficient fits: $\hat{\beta}_{k,j}$ (initially 0)
- Codeword fits: $F_k = X\hat{\beta}_k$ also $F_{k,-j} = X\hat{\beta}_{k,-j}$
- Vector of statistics: $stat_k =$ function of $(X, Y, F_1, \ldots, F_k)$
- e.g. $stat_{k,j}$ proportional to $X_j^T(Y - F_{k,-j})$
- Update $\hat{\beta}_{k+1}$ as a function of $stat_k$
  - Thresholding: Adaptive Successive Decoder

    $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}$ if $stat_{k,j}$ is above threshold in sections ℓ not previously decoded
  - Soft decision:

    $\hat{\beta}_{k+1,j} = \mathbb{E}[\beta_j | stat_k]$

    with thresholding on the last step

Statistics

- $stat_k$ = function of $(X, Y, F_1, \ldots, F_k)$ $\qquad F_k = X\hat{\beta}_k$
- Orthogonalization : Let $G_0 = Y$ and for $k \geq 1$

$$G_k = \text{part of } F_k \text{ orthogonal to } G_0, G_1, \ldots, G_{k-1}$$

- Components of statistics

$$\mathcal{Z}_{k,j} = \frac{X_j^T G_k}{\|G_k\|}$$

- Class of statistics $stat_k$ formed by combining $\mathcal{Z}_0, \ldots, \mathcal{Z}_k$

Statistics

- $stat_k$ = function of $(X, Y, F_1, \ldots, F_k)$ $\qquad F_k = X\hat{\beta}_k$
- Orthogonalization : Let $G_0 = Y$ and for $k \geq 1$

$$G_k = \text{part of } F_k \text{ orthogonal to } G_0, G_1, \ldots, G_{k-1}$$

- Components of statistics

$$\mathcal{Z}_{k,j} = \frac{X_j^T G_k}{\|G_k\|}$$

- Class of statistics $stat_k$ formed by combining $\mathcal{Z}_0, \ldots, \mathcal{Z}_k$

$$stat_{k,j} = \mathcal{Z}_{k,j}^{comb} + \frac{\sqrt{n}}{\sqrt{c_k}}\,\hat{\beta}_{k,j}$$

where $\mathcal{Z}_k^{comb} = \sqrt{\lambda_{k,0}}\,\mathcal{Z}_0 - \sqrt{\lambda_{k,1}}\,\mathcal{Z}_1 - \ldots - \sqrt{\lambda_{k,k}}\,\mathcal{Z}_k$

with $\quad \lambda_{k,0} + \lambda_{k,1} + \ldots + \lambda_{k,k} = 1$

Statistics based on residuals

Let $stat_{k,j}$ be proportional to $X_j^T(Y - X\hat{\beta}_{k,-j})$

$$stat_{k,j} = \frac{X_j^T(Y - X\hat{\beta}_k)}{\sqrt{nc_k}} + \frac{\|X_j\|^2}{\sqrt{nc_k}}\,\hat{\beta}_{k,j}$$

Arises with $\underline{\lambda}_k$ proportional to

$$\Big(\,(\|Y\| - \mathcal{Z}_0^T\hat{\beta}_k)^2,\ (\mathcal{Z}_1^T\hat{\beta}_k)^2, \ldots, (\mathcal{Z}_k^T\hat{\beta}_k)^2\,\Big)$$

and $nc_k = \|Y - X\hat{\beta}_k\|^2$.

Here, c_k is typically between σ^2 and $\sigma^2 + P$

Idealized Statistics

$\underline{\lambda}_k$ exists yielding $stat_k^{ideal}$ with distributional representation

$$\frac{\sqrt{n}}{\sqrt{\sigma^2 + \|\beta - \hat{\beta}_k\|^2}}\beta \; + \; Z_k^{comb}$$

with $\qquad Z_k^{comb} \sim N(0, I)$.

This is a normal shift that improves with decreasing $\|\beta - \hat{\beta}_k\|^2$.

Idealized Statistics

$\underline{\lambda}_k$ exists yielding $stat_k^{ideal}$ with distributional representation

$$\frac{\sqrt{n}}{\sqrt{\sigma^2 + \|\beta - \hat{\beta}_k\|^2}}\,\beta \;+\; Z_k^{comb}$$

with $\qquad Z_k^{comb} \sim N(0, I)$.

This is a normal shift that improves with decreasing $\|\beta - \hat{\beta}_k\|^2$.

For terms sent the shift $\alpha_{\ell,k}$ has an effective *snr* interpretation

$$\alpha_{\ell,k} \;=\; \sqrt{n\,\frac{P_\ell}{\sigma^2 + P_{remaining,k}}}$$

where $\quad P_{remaining,k} = \|\beta - \hat{\beta}_k\|^2$.

Distributional Analysis

Lemma 1: shifted normal conditional distribution

Given $\mathcal{F}_{k-1} = (\|G_0\|, \ldots, \|G_{k-1}\|, \mathcal{Z}_0, \mathcal{Z}_1, \ldots, \mathcal{Z}_{k-1})$, the $\mathcal{Z}_k$ has the distributional representation

$$\mathcal{Z}_k = \frac{\|G_k\|}{\sigma_k} b_k + Z_k$$

- $\|G_k\|^2/\sigma_k^2 \sim \text{Chi-square}(n-k)$
- $Z_k \sim N(0, \Sigma_k)$ indep of $\|G_k\|$

Distributional Analysis

Lemma 1: shifted normal conditional distribution

Given $\mathcal{F}_{k-1} = (\|G_0\|, \ldots, \|G_{k-1}\|, \mathcal{Z}_0, \mathcal{Z}_1, \ldots, \mathcal{Z}_{k-1})$, the $\mathcal{Z}_k$ has the distributional representation

$$\mathcal{Z}_k = \frac{\|G_k\|}{\sigma_k} b_k + Z_k$$

- $\|G_k\|^2/\sigma_k^2 \sim \text{Chi-square}(n-k)$
- $Z_k \sim N(0, \Sigma_k)$ indep of $\|G_k\|$
- $b_0, b_1, \ldots, b_k$ the successive orthonormal components of

$$\begin{bmatrix} \beta \\ \sigma \end{bmatrix}, \begin{bmatrix} \hat{\beta}_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \hat{\beta}_k \\ 0 \end{bmatrix} \qquad (*)$$

- $\Sigma_k = I - b_0 b_0^T - b_1 b_1^T - \ldots - b_k b_k^T$
  $= \text{projection onto space orthogonal to } (*)$
- $\sigma_k^2 = \hat{\beta}_k^T \Sigma_{k-1} \hat{\beta}_k$

Distribution of $\mathcal{Z}_k^T = \left(\frac{X_1^T G_k}{\|G_k\|}, \ldots, \frac{X_N^T G_k}{\|G_k\|}, \frac{\epsilon^T G_k}{\sigma\|G_k\|}\right)$

Lemma 1: shifted normal conditional distribution

Given $\mathcal{F}_{k-1} = (\|G_0\|, \ldots, \|G_{k-1}\|, \mathcal{Z}_0, \mathcal{Z}_1, \ldots, \mathcal{Z}_{k-1})$, the $\mathcal{Z}_k$ has the distributional representation

$$\mathcal{Z}_k = \frac{\|G_k\|}{\sigma_k} b_k + Z_k$$

- $\|G_k\|^2/\sigma_k^2 \sim \text{Chi-square}(n-k)$
- $Z_k \sim N(0, \Sigma_k)$ indep of $\|G_k\|$
- $b_0, b_1, \ldots, b_k$ the successive orthonormal components of

$$\begin{bmatrix} \beta \\ \sigma \end{bmatrix}, \begin{bmatrix} \hat{\beta}_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \hat{\beta}_k \\ 0 \end{bmatrix} \qquad (*)$$

- $\Sigma_k = I - b_0 b_0^T - b_1 b_1^T - \ldots - b_k b_k^T$
  $= \text{projection onto space orthogonal to } (*)$
- $\sigma_k^2 = \hat{\beta}_k^T \Sigma_{k-1} \hat{\beta}_k$

Idealized Statistics

Weights of combination based on $\underline{\lambda}_k$ proportional to

$$\Big((\sigma_Y - b_0^T \hat{\beta}_k)^2,\ (b_1^T \hat{\beta}_k)^2, \ldots, (b_k^T \hat{\beta}_k)^2 \Big)$$

produces the desired distributional representation

$$stat_k^{ideal} = \frac{\sqrt{n}}{\sqrt{\sigma^2 + \|\beta - \hat{\beta}_k\|^2}}\,\beta + Z_k^{comb}$$

with $Z_k^{comb} \sim N(0, I)$ and $\sigma_Y^2 = \sigma^2 + P$.

Idealized Statistics

Weights of combination based on $\underline{\lambda}_k$ proportional to

$$\Big((\sigma_Y - b_0^T \hat{\beta}_k)^2,\ (b_1^T \hat{\beta}_k)^2, \ldots, (b_k^T \hat{\beta}_k)^2 \Big)$$

produces the desired distributional representation

$$stat_k^{ideal} = \frac{\sqrt{n}}{\sqrt{\sigma^2 + \|\beta - \hat{\beta}_k\|^2}}\,\beta + Z_k^{comb}$$

with $Z_k^{comb} \sim N(0, I)$ and $\sigma_Y^2 = \sigma^2 + P$.

- $\|\beta - \hat{\beta}_k\|^2$ is close to its known expectation
- This provides approximation of the distribution of the $stat_{k,j}$ as independent shifted normals.

Relationship between statistics

The stats based on residuals estimate the idealized statistics.

Why? For $stat_k^{ideal}$ the $\underline{\lambda}_k$ are proportional to

$$n\left((\sigma_Y - b_0^T\hat{\beta}_k)^2,\ (b_1^T\hat{\beta}_k)^2, \ldots, (b_k^T\hat{\beta}_k)^2\right)$$

whereas, for the residual-based $stat_k$ they are proportional to

$$\left((\|Y\| - \mathcal{Z}_0^T\hat{\beta}_k)^2,\ (\mathcal{Z}_1^T\hat{\beta}_k)^2, \ldots, (\mathcal{Z}_k^T\hat{\beta}_k)^2\right)$$

Here $\mathcal{Z}_{k'}^T\hat{\beta}_k/\sqrt{n}$ is approximately $b_{k'}^T\hat{\beta}_k$ for $k' \leq k$.

Indeed, with the chi-square factor replaced by its expectation,

$$\mathcal{Z}_{k'}^T\hat{\beta}_k/\sqrt{n} = b_{k'}^T\hat{\beta}_k + Z_{k'}^T\hat{\beta}_k/\sqrt{n}.$$

The $Z_{k'}^T\hat{\beta}_k$ has mean 0 and is stochastically dominated by $Z_{k'}^T\beta$.

Iteratively Bayes optimal coefficient estimates

With prior $j_\ell \sim \mathit{Unif}$ on sec_ℓ, the Bayes estimate based on stat_k

$$\hat{\beta}_{k+1} = \mathbb{E}[\beta | \mathit{stat}_k]$$

has representation $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\, \hat{w}_{k,j}$ with

$$\hat{w}_{k,j} = \mathit{Prob}\{j_\ell = j | \mathit{stat}_k\}.$$

Iteratively Bayes optimal coefficient estimates

With prior $j_\ell \sim Unif$ on sec_ℓ, the Bayes estimate based on $stat_k$

$$\hat{\beta}_{k+1} = \mathbb{E}[\beta | stat_k]$$

has representation $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\, \hat{w}_{k,j}$ with

$$\hat{w}_{k,j} = Prob\{j_\ell = j | stat_k\}.$$

Here, when the $stat_{k,j}$ are independent $N(\alpha_{\ell,k} 1_{\{j=j_\ell\}}, 1)$, we have the logit representation

$$\hat{w}_{k,j} = \frac{e^{\alpha_{\ell,k}\, stat_{k,j}}}{\sum_{j \in sec_\ell} e^{\alpha_{\ell,k}\, stat_{k,j}}}.$$

In our setting, $\alpha_{\ell,k}$ is the shift given by

$$\alpha_{\ell,k} = \sqrt{\frac{n\, P_\ell}{\sigma^2 + \mathbb{E}\|\beta - \hat{\beta}_k\|^2}}$$

Relating error rate and squared distance

- Error of posterior weight is $(1 - \hat{w}_{k,j_\ell})$ if j_ℓ is sent.
- The power-weighted error

$$\sum_{\ell=1}^{L} P_\ell \, (1 - \hat{w}_{k,j_\ell})$$

- Squared distance from $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\,\hat{w}_{k,j}$ to $\beta_j = \sqrt{P_\ell}\,1_{\{j=j_\ell\}}$

$$\|\hat{\beta}_{k+1} - \beta\|^2$$

Relating error rate and squared distance

- Error of posterior weight is $(1 - \hat{w}_{k,j_\ell})$ if j_ℓ is sent.
- The power-weighted error

$$\sum_{\ell=1}^{L} P_\ell \, (1 - \hat{w}_{k,j_\ell})$$

- Squared distance from $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\hat{w}_{k,j}$ to $\beta_j = \sqrt{P_\ell}1_{\{j=j_\ell\}}$

$$\|\hat{\beta}_{k+1} - \beta\|^2$$

Lemma 2

- The power-weighted error and the squared distance have the same expectation.
- Equivalently, the success rate $\sum_{\ell=1}^{L}(P_\ell/P)\hat{w}_{k,j_\ell}$ which is $\beta^T\hat{\beta}_{k+1}/P$ and $\|\hat{\beta}_{k+1}\|^2/P$ have the same expectation.

Relating error rate and squared distance

- Error of posterior weight is $(1 - \hat{w}_{k,j_\ell})$ if j_ℓ is sent.
- The power-weighted error

$$\sum_{\ell=1}^{L} P_\ell \left(1 - \hat{w}_{k,j_\ell}\right)$$

- Squared distance from $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\,\hat{w}_{k,j}$ to $\beta_j = \sqrt{P_\ell}\,1_{\{j=j_\ell\}}$

$$\|\hat{\beta}_{k+1} - \beta\|^2$$

Lemma 2

- The power-weighted error and the squared distance have the same expectation.
- Equivalently, the success rate $\sum_{\ell=1}^{L}(P_\ell/P)\hat{w}_{k,j_\ell}$ which is $\beta^T\hat{\beta}_{k+1}/P$ and $\|\hat{\beta}_{k+1}\|^2/P$ have the same expectation.

Proof: Use $\hat{\beta}_{k+1} = \mathbb{E}\left[\beta | stat_k\right]$.

Relating error rate and squared distance

- Error of posterior weight is $(1 - \hat{w}_{k,j_\ell})$ if j_ℓ is sent.
- The power-weighted error

$$\sum_{\ell=1}^{L} P_\ell (1 - \hat{w}_{k,j_\ell})$$

- Squared distance from $\hat{\beta}_{k+1,j} = \sqrt{P_\ell}\,\hat{w}_{k,j}$ to $\beta_j = \sqrt{P_\ell}1_{\{j=j_\ell\}}$

$$\|\hat{\beta}_{k+1} - \beta\|^2$$

Lemma 2

- The power-weighted error and the squared distance have the same expectation.
- Equivalently, the success rate $\sum_{\ell=1}^{L}(P_\ell/P)\hat{w}_{k,j_\ell}$ which is $\beta^T\hat{\beta}_{k+1}/P$ and $\|\hat{\beta}_{k+1}\|^2/P$ have the same expectation.

Proof: Use $\hat{\beta}_{k+1} = \mathbb{E}[\beta|stat_k]$.

Expected success rate: $x_{k+1} = \sum_{\ell=1}^{L}(P_\ell/P)\,\mathbb{E}\left[\hat{w}_{k,j_\ell}\right]$

Consequence for expected success rate

If the expected success rate was x_k, then using the $stat_{k,j}$ representation $\alpha_{\ell,k}\, 1_{\{j=j_\ell\}} + Z_{k,j}$ with

$$\alpha_{\ell,k} = \sqrt{nP_\ell/(\sigma^2 + P(1-x_k))},$$

then at the next step we have

$$x_{k+1} = g(x_k)$$

where $g(x)$ is the success update function

$$g(x) = \sum_{\ell=1}^{L} \frac{P_\ell}{P}\, success(\alpha_\ell(x))$$

where

$$success(\alpha) = \mathbb{E}\left[\frac{e^{\alpha^2+\alpha Z_1}}{e^{\alpha^2+\alpha Z_1} + \sum_{j=2}^{M} e^{\alpha Z_j}}\right]$$

evaluated at $\alpha_\ell(x) = \sqrt{nP_\ell/(\sigma^2 + P(1-x))}$

assuming w.l.o.g. that first term is sent in each section.

Decoding progression

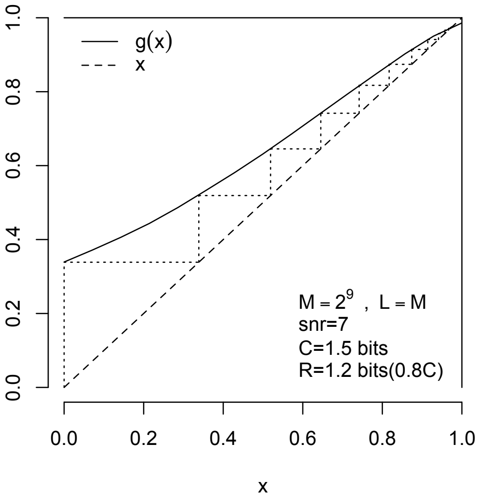


Figure: Plot of $g(x)$ and the sequence x_k.

Integral representation of $g(x)$

Change of variables from $t = \ell/L$ to

$$u = \frac{1 - e^{-2Ct}}{1 - e^{-2C}} \qquad \sim \quad \text{Uniform on } [0, 1],$$

$\alpha_\ell(x)$ becomes

$$\alpha(u, x) \;=\; \tau \sqrt{\frac{C}{R} \; \frac{1 + snr(1-u)}{1 + snr(1-x)}}$$

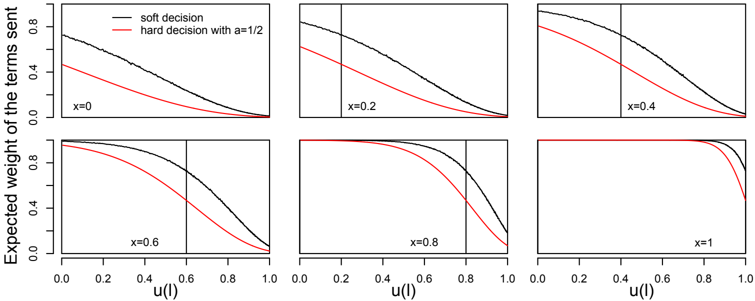
which can be compared to $\tau = \sqrt{2 \log M}$.

We have the integral representation of $g(x)$

$$g(x) = \mathbb{E}_U \left[g(U, x) \right] = \int_0^1 g(u, x) du$$

where $g(u, x) = success(\alpha(u, x))$

Transition plots



Setting: $M = 2^9$, $L = M$, $C = 1.5$ bits and $R = 0.8C$.
Plot of $g(u, x)$ for $x = 0, 0.2, 0.4, 0.6, 0.8, 1$.
Horizontal axis: depicts $u(\ell) = (1 - e^{-2C\ell/L})/(1 - e^{-2C})$.
Black curves: our soft decision decoder
Red curves: thresholding decoder with threshold $\sqrt{2 \log M} + a$

The area under the curve is $g(x)$.

Lowerbound for update function

Using Jensen's inequality, we have

$$
\begin{aligned}
\mathit{success}(\alpha) &= \mathbb{E}\left[\frac{e^{\alpha^2+\alpha Z_1}}{e^{\alpha^2+\alpha Z_1}+\sum_{j=2}^{M} e^{\alpha Z_j}}\right] \\
&\geq \mathbb{E}\left[\frac{e^{\alpha^2+\alpha Z_1}}{e^{\alpha^2+\alpha Z_1}+(M-1)e^{\alpha^2/2}}\right]
\end{aligned}
$$

so that

$$
g(x) \geq \mathbb{P}\{\xi \leq \alpha_U^2/2 - \tau^2/2 + \alpha_U Z\}
$$

where $\quad \xi \sim \mathit{logistic}(0,1) \quad$ and $\alpha_u = \alpha(u,x)$

The Logit representation

- By McFadden(1974),

  Let $s_1, \ldots, s_m$ be a fixed sequence and ϵ_j be independent Gumbel distributed random variable. Then,

$$\mathbb{P}\{s_1 + \epsilon_1 \geq \max_{2 \leq j \leq m} (s_j + \epsilon_j)\} = \frac{e^{s_1}}{\sum_{j=1}^{m} e^{s_j}}$$

  Thus, we can write g(x) as

$$g(x) = \mathbb{P}\left\{\alpha_U^2 + \alpha_U Z_1 + \epsilon_1 \geq \max_{2 \leq j \leq m} (\alpha_U Z_j + \epsilon_j)\right\},$$

Extreme value representation of the update function

- Using the logit representation: Approximation of the update function

$$g(x) = \mathbb{P}\{V_1 \le \alpha_U\},$$

where

$$V_1 = \max_{2 \le j \le m} \left\{ -\frac{Z_1 - Z_j}{2} + \sqrt{\left[\epsilon_j - \epsilon_1 + \frac{(Z_1 - Z_j)^2}{4}\right]_+} \right\}.$$

- For the lowerbound

$$g(x) \ge \mathbb{P}\{V_2 \le \alpha_U\}$$

where

$$V_2 = -Z_1 + \sqrt{(\tau^2 + 2\xi + Z_1^2)_+}.$$

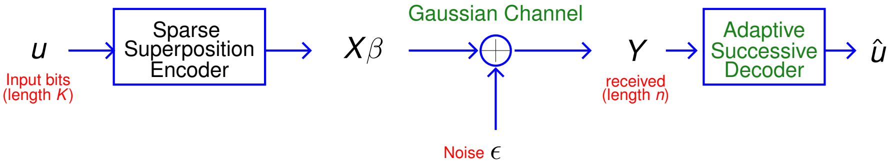
Analysis of Update function

- x^* solves $g(x) = x$, yields mistake rate $1 - x^*$
- Communication rate $R = C/(1 + r/\tau^2)$
- with $r = \mathbb{E}[(V_+^2 - \tau^2)1_B]$, mistake rate

$$1 - x^* = \frac{1}{snr} \, \frac{r}{\tau^2}$$

- Here r grows no faster than order of τ
- $B = \{\alpha(1, x^*) \leq V \leq \alpha(0, x^*)\}$

Summary



Reliable for rates $R < C$

For the adaptive success decoder

- with thresholding (J&B 2010b ISIT, 2012b)
- with iteratively optimal soft decision (shown here)

Update fuctions

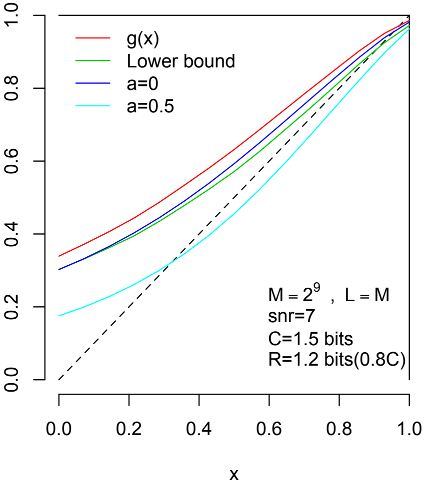


Figure: Comparison of update functions. Blue and light blue lines indicates $\{0, 1\}$ decision using the threshold $\sqrt{2 \log M} + a$ with respect to the value a as indicated.

Update fuctions

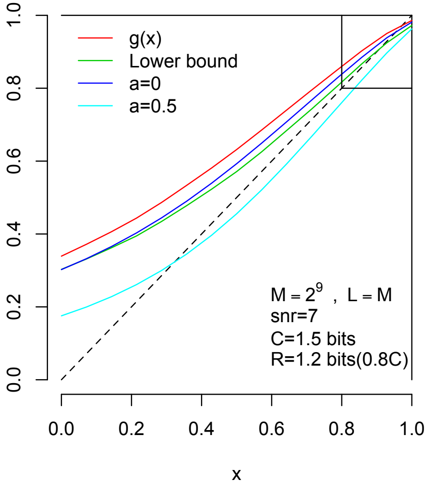


Figure: Comparison of update functions. Blue and light blue lines indicates $\{0, 1\}$ decision using the threshold $\sqrt{2 \log M} + a$ with respect to the value a as indicated.

Update fuctions

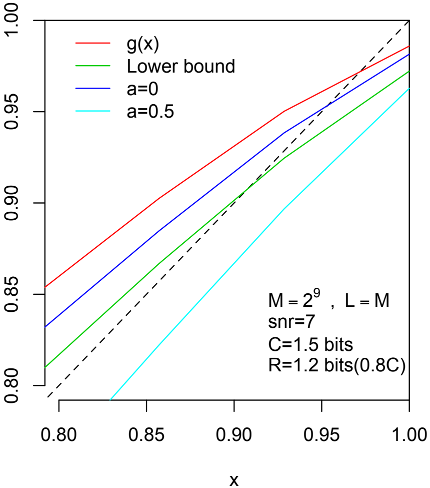


Figure: Comparison of update functions. Blue and light blue lines indicates $\{0, 1\}$ decision using the threshold $\sqrt{2\log M} + a$ with respect to the value a as indicated.

Iteratively Bayes optimal coefficient estimates

With prior $j_\ell \sim Unif$ on sec_ℓ, the Bayes estimate based on $stat_k$

$$\begin{aligned}\hat{\beta}_{k+1} &= \mathbb{E}[\beta | stat_k] \\ &\cong \mathbb{E}[\beta | stat_k, stat_{k-1}, \ldots, stat_1] \\ &\cong \mathbb{E}[\beta | \mathcal{F}_k]\end{aligned}$$

where $\mathcal{F}_k = \{$ Standardized inner products of X columns with Y and with components of the fits $F_1, \ldots, F_k\}$

Here, $F_k = X\hat{\beta}_k$

Fraction of Mistakes

Translating power-weighted value of $(1 - \hat{w}_{k,j_\ell})$

into fraction of occurrences of $\{1 - \hat{w}_{k,j_\ell} \geq 1/2\}$

$$\delta_{mis} \leq \frac{snr}{C}(1 - x^*) = \frac{r}{C\tau^2}$$

at rate $\quad R = \frac{C}{1+r/\tau^2}$