Linear and Nonlinear Iterative Multiuser Detection

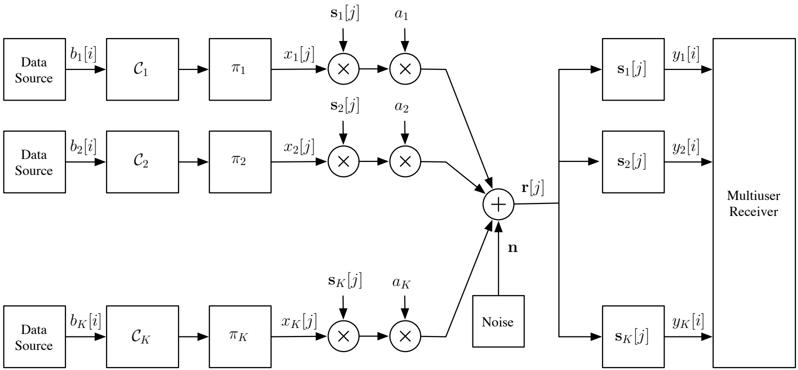
Alex Grant and Lars Rasmussen

University of South Australia

October 2011

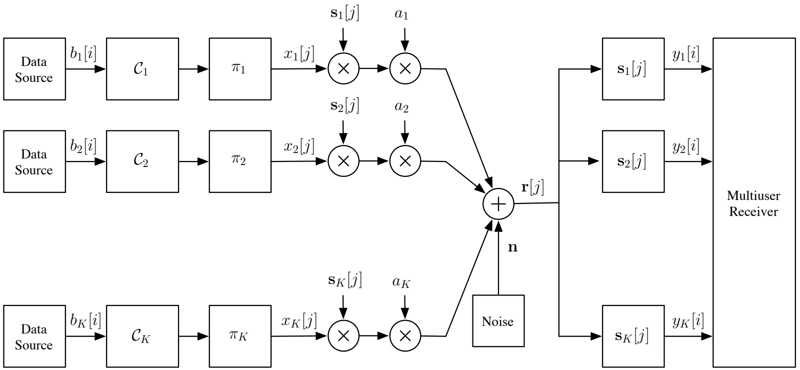
Outline

Outline



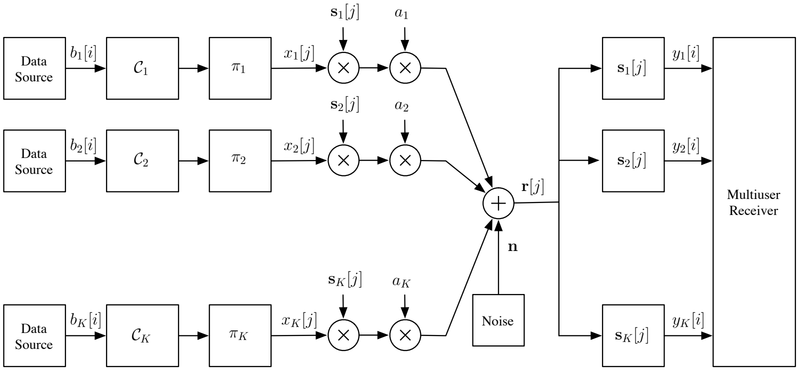
- BPSK data $x_k[j] \in \{-1, 1\}$
- Spreading $\mathbf{s}_k[i] \in \left\{-1/\sqrt{N}, +1/\sqrt{N}\right\}^N$
- AWGN E $\left\{\mathbf{n}[i]\mathbf{n}^t[i]\right\} = \sigma^2\mathbf{I}$

System Model



$$\mathbf{r}[i] = \sum_{k=1}^{K} \mathbf{s}_k[i] a_k x_k[i] + \mathbf{n}[i]$$
$$= \mathbf{S}[i]\mathbf{A}\mathbf{x}[i] + \mathbf{n}[i]$$

System Model



$$
\begin{aligned}
\mathbf{y}[i] &= \mathbf{S}^t[i]\mathbf{r}[i] \\
&= \mathbf{S}^t[i]\mathbf{S}[i]\mathbf{A}\mathbf{x}[i] + \mathbf{S}^t[i]\mathbf{n}[i] \\
&= \mathbf{R}[i]\mathbf{A}\mathbf{x}[i] + \mathbf{z}[i]
\end{aligned}
$$

Matched Filter Ouput

$$\begin{aligned}\mathbf{y}[i] &= \mathbf{S}^t[i]\mathbf{r}[i] \\ &= \mathbf{S}^t[i]\mathbf{S}[i]\mathbf{A}\mathbf{x}[i] + \mathbf{S}^t[i]\mathbf{n}[i] \\ &= \mathbf{R}[i]\mathbf{A}\mathbf{x}[i] + \mathbf{z}[i]\end{aligned}$$

- Matched filter output is a sufficient statistic
- $\mathbf{R}[i] = \mathbf{S}^t[i]\mathbf{S}[i]$ is the correlation matrix
- $\mathbf{z}[i]$ is colored Gaussian noise, $\mathsf{E}\left\{\mathbf{z}[i]\mathbf{z}^t[i]\right\} = \sigma^2\mathbf{R}[i]$
- Assume unit norm modulation, $R_{ii} = 1$ and $|R_{ij}| \leq 1$, $i \neq j$.

Outline

Optimal Joint Detection

- Minimum error probability optimal multiuser detector

$$\hat{x}_{\text{MAP}} = \arg \max_{\mathbf{x} \in \{-1,1\}^K} \Pr(\mathbf{x}|\mathbf{r})$$
$$= \arg \max_{\mathbf{x} \in \{-1,1\}^K} \mathrm{p}(\mathbf{r}|\mathbf{x}) \Pr(\mathbf{x}).$$

- Equiprobable data - maximum-likelihood detector,

$$\hat{x}_{\text{ML}} = \arg \max_{\mathbf{x} \in \{-1,1\}^K} \mathrm{p}(\mathbf{r}|\mathbf{x})$$
$$= \arg \max_{\mathbf{x} \in \{-1,1\}^K} \exp\left(-\frac{1}{2}||\mathbf{r} - \mathbf{SAx}||^2\right)$$
$$= \arg \max_{\mathbf{x} \in \{-1,1\}^K} \exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathbf{RAx}\right)\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{RAx}\right)^t\right).$$

- Complexity $\mathcal{O}\left(2^K\right)$

Optimal Joint Detection

W. van Etten.
Maximum likelihood receiver for multiple channel transmission systems.
IEEE Trans. Commun., 24(2):276–283, February 1976.

K. S. Schneider.
Optimum detection of code division multiplexed signals.
IEEE Trans. Aerosp. Electron. Systems, 15:181–185, January 1979.

R. Kohno, M. Hatori, and H. Imai.
Cancellation techniques of co-channel interference in asynchronous spread spectrum multiple access systems.
Electronics and Commun., 66-A:20–29, 1983.

S. Verdú.
Minimum probability of error for asynchronous Gaussian multiple–access channels.
IEEE Trans. Inform. Theory, 32(1):85–96, January 1986.

Decorrelator

- Suppose correlation matrix $\mathbf{R} = \mathbf{S}^t\mathbf{S}$ is invertible

$$
\begin{aligned}
\hat{\mathbf{x}} &= \left(\mathbf{S}^t\mathbf{S}\right)^{-1}\mathbf{S}^t\mathbf{r} \\
&= \mathbf{R}^{-1}\mathbf{y} \\
&= \mathbf{A}\mathbf{x} + \mathbf{R}^{-1}\mathbf{z}, \quad \mathsf{E}\left\{\mathbf{R}^{-1}\mathbf{z}\mathbf{z}^t\mathbf{R}^{-1}\right\} = \sigma^2\mathbf{R}^{-1}
\end{aligned}
$$

- $\mathbf{y}$ consists of correlated data $\mathbf{RAx}$ and correlated noise $\mathbf{R}$,
- $\hat{\mathbf{x}}$ consists of independent data $\mathbf{Ax}$ and correlated noise $\mathbf{R}^{-1}$.
- Multiple-access interference elimination vs. noise enhancement
- ML when relaxing $\mathbf{x} \in \{-1, 1\}^K$ to $\mathbf{x} \in \mathbb{R}^K$.
- ML for unknown $\mathbf{A}$ (estimates $\mathbf{Ax}$ rather than $\mathbf{x}$).

R. Lupas and S. Verdú.
Linear multiuser detectors for synchronous code–division multiple–access channels.
IEEE Trans. Inform. Theory, 35(1):123–136, January 1989.

Minimum Mean Squared Error Estimation

- Let $\mathbf{x}$ and $\mathbf{y}$ be random vectors with

$$\begin{aligned}\bar{\mathbf{x}} &= \mathsf{E}\left\{\mathbf{x}\right\} \\ \bar{\mathbf{y}} &= \mathsf{E}\left\{\mathbf{y}\right\} \\ \mathbf{G}_{xy} &= \left(\mathsf{Cov}\left\{\mathbf{y},\mathbf{y}\right\}\right)^{-1}\mathsf{Cov}\left\{\mathbf{y},\mathbf{x}\right\}\end{aligned}$$

- LMMSE estimate of $\mathbf{x}$ given $\mathbf{y}$ is

$$\bar{\mathbf{x}} + \mathbf{G}_{xy}^{t}\left(\mathbf{y} - \bar{\mathbf{y}}\right).$$

- For jointly Gaussian $\mathbf{x}, \mathbf{y}$ linear estimate in fact minimizes mean squared error.

H. V. Poor.
An Introduction to Signal Detection and Estimation.
Springer-Verlag, 1994.

Linear Minimum Mean Square Error Detectors

- Chip-level (considering $\mathbf{x}, \mathbf{r}$) or symbol-level (considering $\mathbf{x}, \mathbf{y}$)

$$\hat{\mathbf{x}}^r = \mathsf{E}\left\{\mathbf{x}\right\} + \mathbf{G}_{xr}^t \left(\mathbf{r} - \bar{\mathbf{r}}\right)$$
$$\hat{\mathbf{x}}^y = \mathsf{E}\left\{\mathbf{x}\right\} + \mathbf{G}_{xy}^t \left(\mathbf{y} - \bar{\mathbf{y}}\right).$$

- Independent data with $\mathsf{E}\left\{\mathbf{x}\right\} = \bar{\mathbf{x}}$ and $\mathsf{Cov}\left\{\mathbf{x}, \mathbf{x}\right\} = \mathbf{I} - \mathrm{diag}\left(\bar{\mathbf{x}}\bar{\mathbf{x}}^t\right) = \mathbf{V}$ results in

$$\begin{aligned}
\hat{\mathbf{x}}^r &= \bar{\mathbf{x}} + \mathbf{VAS}^t \left(\mathbf{SAVAS}^t + \sigma^2\mathbf{I}\right)^{-1} \left(\mathbf{r} - \mathbf{SA}\bar{\mathbf{x}}\right) \\
\hat{\mathbf{x}}^y &= \bar{\mathbf{x}} + \mathbf{VAR}\left(\mathbf{RAVAR} + \sigma^2\mathbf{R}\right)^{-1} \left(\mathbf{y} - \mathbf{RA}\bar{\mathbf{x}}\right) \\
&= \bar{\mathbf{x}} + \mathbf{A}^{-1}\left(\mathbf{R} + \sigma^2\left(\mathbf{AVA}\right)^{-1}\right)^{-1} \left(\mathbf{y} - \mathbf{RA}\bar{\mathbf{x}}\right).
\end{aligned}$$

- Zero mean data results in

$$\hat{\mathbf{x}}^r = \mathbf{AS}^t\left(\mathbf{SA}^2\mathbf{S}^t + \sigma^2\mathbf{I}\right)^{-1}\mathbf{r}$$
$$\hat{\mathbf{x}}^y = \mathbf{A}^{-1}\left(\mathbf{R} + \sigma^2\mathbf{A}^{-2}\right)^{-1}\mathbf{y}.$$

Linear Minimum Mean Square Error Detectors

Z. Xie, R. T. Short, and C. K. Rushforth.
A family of suboptimum detectors for coherent multiuser communications.
IEEE J. Select. Areas Commun., 8(4):683–690, May 1990.

P. B. Rapajic and B. S. Vucetic.
Adaptive receiver structures for asynchronous CDMA systems.
IEEE J. Select. Areas Commun., 12(4):685–697, May 1994.

U. Madhow and M. L. Honig.
MMSE interference suppression for direct–sequence spread–spectrum CDMA.
IEEE Trans. Commun., 42(12):3178–3188, December 1994.

Outline

Interference Cancellation

- Matched filter output for user k

$$
\begin{aligned}
y_k &= \mathbf{s}_k^t \mathbf{r} \\
&= a_k x_k + \mathbf{s}_k^t \left(\sum_{j \neq k} \mathbf{s}_j a_j x_j + \mathbf{n} \right) \\
&= a_k x_k + \underbrace{\sum_{j \neq k} R_{kj} a_j x_j}_{\text{multiple-access interference}} + z_k
\end{aligned}
$$

- User k could subtract an estimate of the MAI
- Estimate will not be perfect leaving residual MAI.
- This motivates an *iterative* cancellation approach

Iterative Interference Cancellation

- Let estimate of user k at iteration n be $\hat{x}_k^{(n)}$.

$$\hat{x}_k^{(n+1)} = a_k^{-1}\mathbf{s}_k^t \left(\mathbf{r} - \sum_{j\neq k}\mathbf{s}_j a_j \hat{x}_j^{(n)}\right) \quad \text{chip-rate}$$

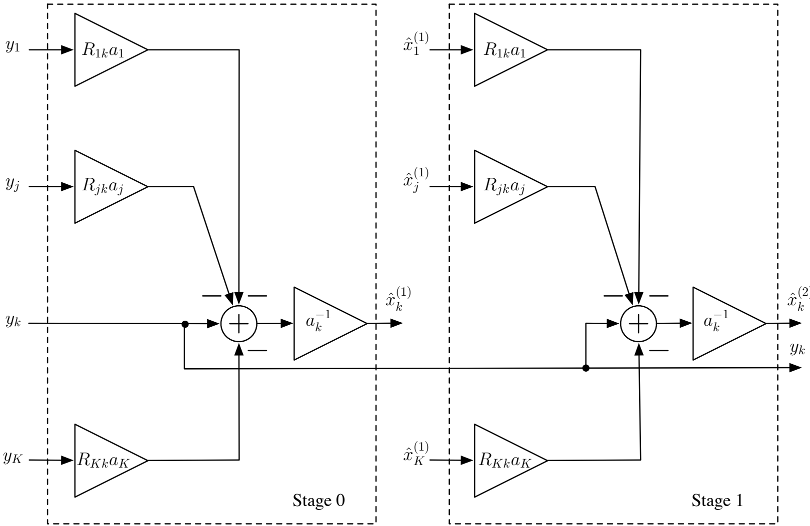
$$= a_k^{-1} \left(y_k - \sum_{j\neq k} R_{jk} a_j \hat{x}_j^{(n)}\right) \quad \text{symbol-rate}$$

- Choosing initial estimate $\hat{x}_k^{(0)} = 0$ yields $\hat{x}_k^{(1)} = a_k^{-1} y_k$.
- If $\hat{\mathbf{x}}^{(n)} = \mathbf{x}$ cancellation is perfect,

$$\hat{x}_k^{(n+1)} = x_k + \frac{z_k}{a_k}$$

Parallel Cancellation

$$\hat{\mathbf{x}}^{(n+1)} = \mathbf{A}^{-1}\left(\mathbf{y} - (\mathbf{R} - \mathbf{I})\,\mathbf{A}\hat{\mathbf{x}}^{(n)}\right).$$



Serial Cancellation

- Cancel estimates as soon as they are available

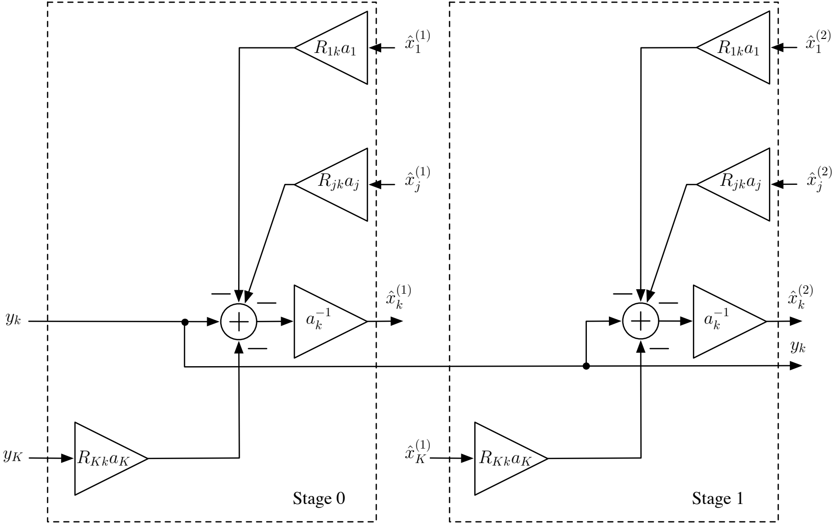
$$\hat{x}_k^{(n+1)} = a_k^{-1}\mathbf{s}_k^t \left(\mathbf{r} - \sum_{j=1}^{k-1} \mathbf{s}_j a_j \hat{x}_j^{(n+1)} - \sum_{j=k+1}^{K} \mathbf{s}_j a_j \hat{x}_j^{(n)} \right)$$
$$= a_k^{-1} \left(y_k - \sum_{j=1}^{k-1} R_{jk} a_j \hat{x}_j^{(n+1)} - \sum_{j=k+1}^{K} R_{jk} a_j \hat{x}_j^{(n)} \right).$$

- Let $\mathbf{R} = \mathbf{L} + \mathbf{L}^t + \mathbf{I}$ where L is strictly triangular

$$\hat{\mathbf{x}}^{(n+1)} = \mathbf{A}^{-1} \left(\mathbf{y} - \mathbf{L}\mathbf{A}\hat{\mathbf{x}}^{(n+1)} - \mathbf{L}^t\mathbf{A}\hat{\mathbf{x}}^{(n)} \right).$$

Serial Cancellation

$$\hat{\mathbf{x}}^{(n+1)} = \mathbf{A}^{-1}\left(\mathbf{y} - \mathbf{L}\mathbf{A}\hat{\mathbf{x}}^{(n+1)} - \mathbf{L}^t\mathbf{A}\hat{\mathbf{x}}^{(n)}\right).$$



Implementation via Residual Error Update

- Chip-level parallel interference canceller

$$\hat{x}_k^{(n+1)} = a_k^{-1}\mathbf{s}_k^t \left(\mathbf{r} - \sum_{j=1}^{K} \mathbf{s}_j a_j \hat{x}_j^{(n)} + \mathbf{s}_k a_k \hat{x}_k^{(n)}\right)$$

$$= \hat{x}_k^{(n)} + a_k^{-1}\mathbf{s}_k^t \left(\mathbf{r} - \sum_{j=1}^{K} \mathbf{s}_j a_j \hat{x}_j^{(n)}\right).$$

$$= \hat{x}_k^{(n)} + a_k^{-1}\mathbf{s}_k^t \boldsymbol{\eta}^{(n)}$$

- *Residual error*, or noise hypothesis

$$\boldsymbol{\eta}^{(n)} = \mathbf{r} - \sum_{j=1}^{K} \mathbf{s}_j a_j \hat{x}_j^{(n)}$$

- Perfect cancellation, residual is thermal noise, $\boldsymbol{\eta}^{(n)} = \mathbf{n}$.

Implementation via Residual Error Update

- Can do same thing for serial cancellation

$$\hat{x}_k^{(n+1)} = \hat{x}_k^{(n)} + a_k^{-1}\mathbf{s}_k^t\boldsymbol{\eta}_k^{(n)},$$

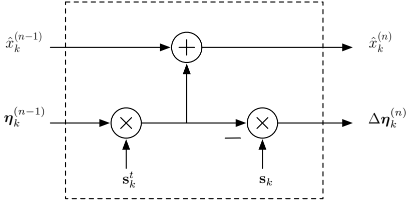
- Residual error seen by user k after iteration n.

$$\boldsymbol{\eta}_k^{(n)} = \mathbf{r} - \sum_{j=1}^{k-1}\mathbf{s}_j a_j \hat{x}_j^{(n+1)} - \sum_{j=k}^{K}\mathbf{s}_j a_j \hat{x}_j^{(n)}$$

Implementation via Residual Error Update

- Can write iteration in terms of an update to residual error

$$\Delta \boldsymbol{\eta}_k^{(n)} = \mathbf{s}_k a_k \left(\hat{x}_k^{(n-1)} - \hat{x}_k^{(n)} \right)$$
$$= -\mathbf{s}_k \mathbf{s}_k^t \boldsymbol{\eta}_k^{(n-1)}.$$

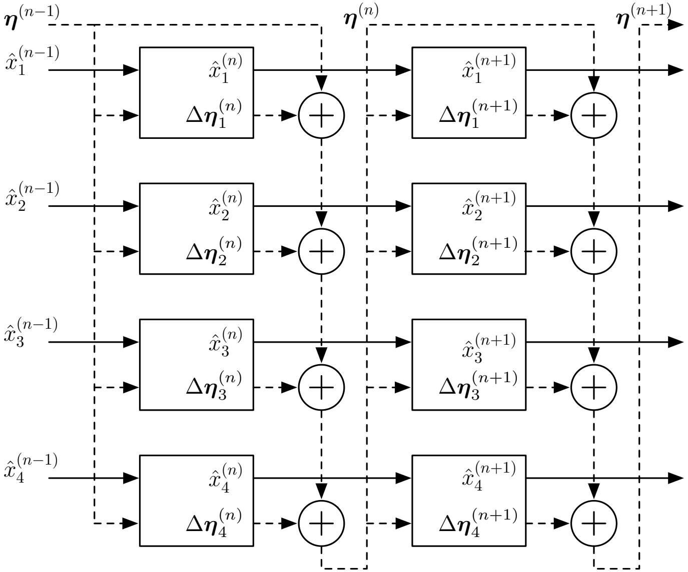


Interference cancellation module.

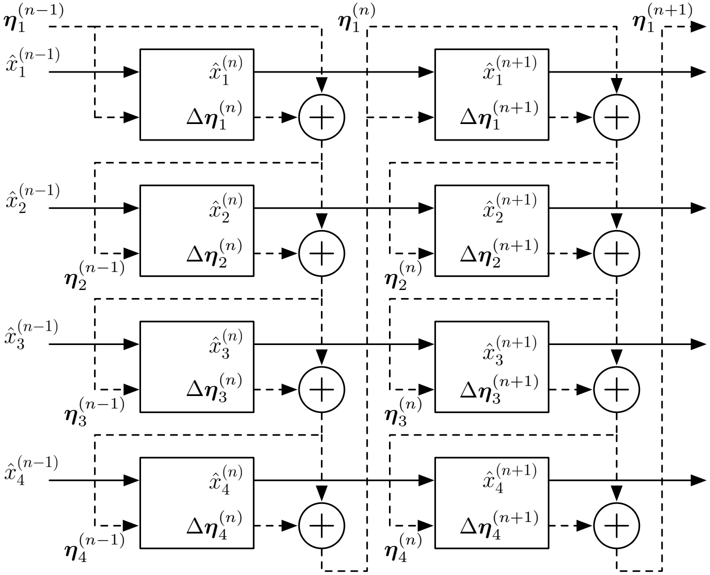
$$\boldsymbol{\eta}^{(n)} = \boldsymbol{\eta}^{(n-1)} + \sum_{j=1}^{K} \Delta \boldsymbol{\eta}_j^{(n)}, \quad \text{(Parallel)}$$

$$\boldsymbol{\eta}_k^{(n)} = \begin{cases} \boldsymbol{\eta}_K^{(n-1)} + \Delta \boldsymbol{\eta}_K^{(n)} & k = 1 \\ \boldsymbol{\eta}_{k-1}^{(n)} + \Delta \boldsymbol{\eta}_{k-1}^{(n+1)} & k > 1. \end{cases} \quad \text{(Serial)}$$

Modular Parallel Cancellation

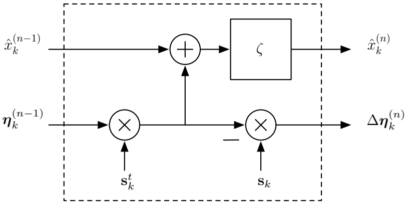


Modular Serial Cancellation

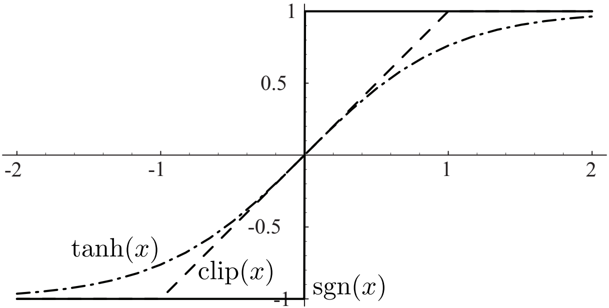


Tentative Decision Functions

- Transmitted symbols discrete $\{-1, +1\}$
- Estimates $\hat{x}_k^{(n)}$ could be any real number.
- What if $\left|\hat{x}_k^{(n)}\right| \gg 1$?
- Non-linear *tentative decision function* $\zeta : \mathbb{R} \mapsto [-1, +1]$



Tentative Decision Functions



Outline

Linear Methods

- Both decorrelator and LMMSE require $K \times K$ matrix inversion
- Complexity scales $\mathcal{O}\left(K^3\right)$
- There are many lower complexity approaches for solving linear systems
- Series expansions
- Iterative matrix inversion
- Gradient descent
- These can all be implemented as interference cancellation

Direct Solution Methods

- Output of decorrelator or LMMSE can be written

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathbb{R}^K} \|\mathbf{M}\mathbf{x} - \mathbf{y}\|_2^2$$

- $\mathbf{M} = \mathbf{R}$ for the decorrelator,
- $\mathbf{M} = \mathbf{R}\mathbf{A}$ for the normalized decorrelator,
- $\mathbf{M} = \mathbf{R} + \sigma^2 \mathbf{A}^{-2}$ for LMMSE.
- Solution to the *unconstrained* optimization problem is

$$\mathbf{M}\hat{\mathbf{x}} = \mathbf{y}.$$

Direct Solution

- Gaussian elimination followed by back-substitution.
- Symmetric $\mathbf{M}$ - equivalent to Cholesky factorization

$$\mathbf{M} = \mathbf{F}\mathbf{F}^t$$

followed by forward and backward substitution,

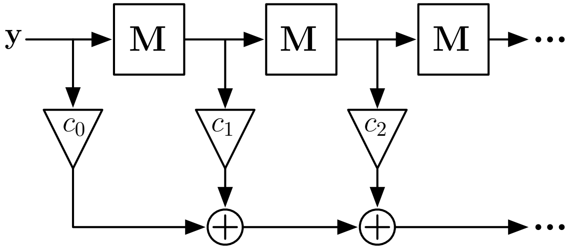
$$\mathbf{F}\mathbf{z} = \mathbf{y} \qquad \text{forward substitution}$$
$$\mathbf{F}^t\hat{\mathbf{x}} = \mathbf{z} \qquad \text{backward substitution}$$

- Cholesky factorization is $\mathcal{O}\left(K^3/3\right)$,
- Each substitution step is $\mathcal{O}\left(K^2/2\right)$.
- If $M_{ij} = 0$ for $|i - j| > b$, Cholesky decomposition is $\mathcal{O}\left(K(b^2 + 3b)\right)$ and substitution steps are $\mathcal{O}\left(Kb\right)$

Series Expansions

- Find coefficients c_n such that

$$\mathbf{M}^{-1} = \sum_n c_n \mathbf{M}^n.$$



- With K terms this is $\mathcal{O}\left(K^3\right)$.
- Truncate series to $n \ll K$.

Series Expansions

- S. Moshavi, E. G. Kanterakis, and D. L. Schilling.
  Multistage linear receivers for DS-CDMA systems.
  Int. J. Wireless Inform. Networks, 3:1–17, January 1996.
- D. Guo, L. K. Rasmussen, S. Sun, , and T. J. Lim.
  A matrix-algebraic approach to linear parallel interference cancellation in CDMA.
  IEEE Trans. Commun., 41(1):152–161, Jan. 2000.
- R. R. Müller and S. Verdú.
  Design and analysis of low-complexity interference mitigation on vector channels.
  IEEE J. Select. Areas Commun., 19:1429–1441, August 2001.
- D. Guo, S. Verdú, and L. K. Rasmussen.
  Asymptotic normality of linear multiuser receiver outputs.
  IEEE Trans. Inform. Theory, 48(12):3080–3095, December 2002.

Cayley Hamilton - Finite Series

Theorem (Cayley Hamilton)

$$p_{\mathbf{M}}(\mathbf{M}) = \sum_{n=0}^{K} (-1)^{K-n} c_{K-n} \mathbf{M}^n = 0$$

- Can write any power of $\mathbf{M}$ as a linear combination of $\mathbf{M}^n, n = 0, 1, \ldots, K$.

$$\mathbf{M}^{-1} = \frac{1}{(-1)^K \det(\mathbf{M})} \sum_{n=1}^{K} (-1)^{K-n} c_{K-n} \mathbf{M}^{n-1}$$

- K-stage multistage implementation.
- Computation of c_n is as complex as matrix inversion
- There exists coefficients such that a finite power series implements matrix inversion exactly.

Taylor Series

- Taylor series

$$(\mathbf{I}+\mathbf{X})^{-1}=\sum_{n=0}^{\infty}(-\mathbf{X})^n,$$

- Convergent if spectral radius satisfies $\rho\left(\mathbf{X}\right)<1$
- Setting $\mathbf{X}=\mathbf{M}-\mathbf{I}$,

$$\mathbf{M}^{-1}=\sum_{n=0}^{\infty}(-1)^n(\mathbf{M}-\mathbf{I})^n,$$

- Convergent if $\rho\left(\mathbf{M}\right)<2$.

Taylor Series

- First order truncation results in the *approximate decorrelator*

$$\hat{\mathbf{x}}^{(1)} = (2\mathbf{I} - \mathbf{R})\mathbf{y} = \mathbf{y} - \underbrace{(\mathbf{R} - \mathbf{I})\mathbf{y}}_{\text{Interference Estimate}}$$

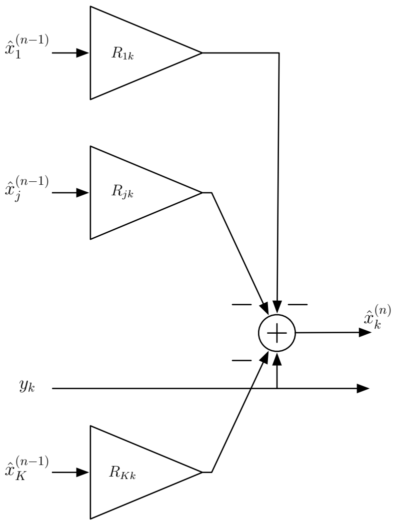
- Higher-order truncation

$$\hat{\mathbf{x}}^{(n)} = \mathbf{y} + (\mathbf{I} - \mathbf{R})\mathbf{y} + (\mathbf{I} - \mathbf{R})^2\mathbf{y} + \cdots + (\mathbf{I} - \mathbf{R})^n\mathbf{y} = \mathbf{y} - (\mathbf{R} - \mathbf{I})\hat{\mathbf{x}}^{(n-1)}$$

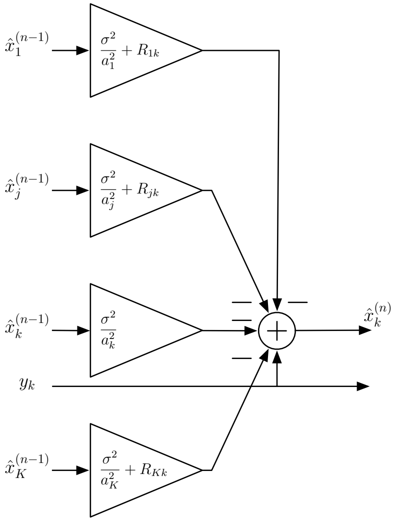
- This is parallel interference cancellation:

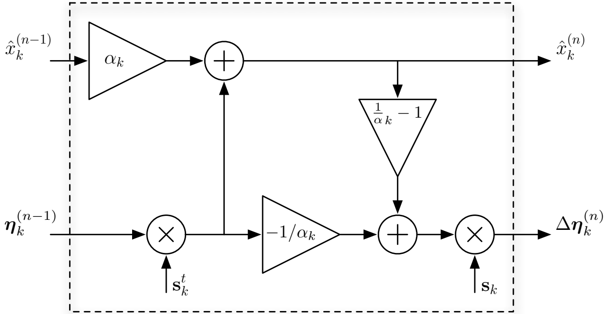
$$\hat{x}_k^{(n)} = \underbrace{y_k}_{\text{User } k \text{ matched filter output}} - \underbrace{\sum_{k' \neq k} R_{kk'} \hat{x}_{k'}^{(n-1)}}_{\text{Interference estimate from previous stage}} .$$

Taylor Series: Decorrelator



Taylor Series: LMMSE





Decorrelator $\alpha_k = 1$ **and LMMSE,** $\alpha_k = 1 - \sigma^2/a_k^2$

Taylor Series

- Infinte series rather than finite series.
- Convergent only if $\rho\left(\mathbf{M}\right) < 2$.
- Easy computation of series cofficients.
- LMMSE only marginally more complex than decorrelator.

- S. Verdú.
  Multiuser Detection.
  Cambridge University Press, Cambridge, 1998.

- N. B. Mandayam and S. Verdú.
  Analysis of an approximate decorrelating detector.
  Wireless Personal Commun., 6:97–111, June 1998.

- S. Moshavi, E. G. Kanterakis, and D. L. Schilling.
  Multistage linear receivers for DS-CDMA systems.
  Int. J. Wireless Inform. Networks, 3:1–17, January 1996.

Iterative Solution Methods

- Let $\mathbf{M} = \mathbf{M}_1 - \mathbf{M}_2$. Then $(\mathbf{M}_1 - \mathbf{M}_2)\hat{\mathbf{x}} = \mathbf{y}$,
- Fixed point equation,

$$\mathbf{M}_1\hat{\mathbf{x}} = \mathbf{y} + \mathbf{M}_2\hat{\mathbf{x}}.$$

- Motivates iteration of the form

$$\mathbf{M}_1\mathbf{x}^{(n+1)} = \mathbf{y} + \mathbf{M}_2\mathbf{x}^{(n)}.$$

- Require:
  1. Easy to solve $\mathbf{M}_1\mathbf{x} = \mathbf{z}$. E.g. $\mathbf{M}_1$ triangular or diagonal.
  2. Choose $\mathbf{M}_1$ and $\mathbf{M}_2$ so iteration converges quickly.

- Let $\mathbf{e}^{(n)} = \hat{\mathbf{x}} - \mathbf{x}^{(n)}$

$$\|\mathbf{e}^{(n)}\| = \|(\mathbf{M}_1^{-1}\mathbf{M}_2)^n\mathbf{e}^{(0)}\| \leq \|(\mathbf{M}_1^{-1}\mathbf{M}_2)^n\|\|\mathbf{e}^{(0)}\|.$$

Theorem

A necessary and sufficient condition for convergence of in any norm is

$$\rho\left(\mathbf{M}_1^{-1}\mathbf{M}_2\right) < 1.$$

Jacobi Iteration

- Choose

$$\mathbf{M}_1 = \omega \mathbf{D}$$
$$\mathbf{M}_2 = \omega \mathbf{D} - \mathbf{M}.$$

With $\mathbf{x}^{(0)} = \mathbf{y}$, iteration becomes

$$\mathbf{x}^{(n+1)} = \frac{\mathbf{D}^{-1}}{\omega} \left(\mathbf{y} - (\mathbf{M} - \omega \mathbf{D}) \mathbf{x}^{(n)} \right).$$

- Taylor series expansion of $(\omega \mathbf{D} + \mathbf{M})^{-1}$.
- Multistage parallel interference cancellation.

Theorem

The Jacobi implementation of the decorrelator with $\mathbf{M}_1 = \omega \mathbf{I}$ is convergent for any $\omega > 0$ such that $\rho(\mathbf{R}) < 2\omega$.

Jacobi Iteration: LMMSE

- For $\omega = 1$, the LMMSE Jacobi iteration is

$$\mathbf{x}^{(n+1)} = \left(\mathbf{I} + \mathbf{A}^{-2}\sigma^2\right)^{-1}\left(\mathbf{y} - (\mathbf{R} - \mathbf{I})\mathbf{x}^{(n)}\right).$$

- Per-user signal-to-noise ratio scaling each iteration.

Theorem

The Jacobi iterative implementation of the LMMSE filter is convergent if and only if

$$\rho_J = \rho\left(\left(\mathbf{I} + \mathbf{A}^{-2}\sigma^2\right)^{-1}(\mathbf{I} - \mathbf{R})\right) < 1$$

Jacobi Iteration: LMMSE

Theorem

The Jacobi iterative implementation of the LMMSE filter is convergent if

$$\rho\left(\mathbf{R}-\mathbf{I}\right) < 1+\gamma_{\max}^{-1}$$

where $\gamma_{\max}=\max_k A_k^2/\sigma^2$*. The iteration is not convergent if*

$$\rho\left(\mathbf{R}-\mathbf{I}\right) > 1+\gamma_{\min}^{-1}$$

where $\gamma_{\min}=\min_k A_k^2/\sigma^2$*.*

Gauss-Seidel Iteration

- Choose

$$\mathbf{M}_1 = \frac{1}{\omega}\mathbf{D} + \mathbf{L}$$
$$\mathbf{M}_2 = \frac{1-\omega}{\omega}\mathbf{D} - \mathbf{L}^t$$

- Results in the following iteration (with $\mathbf{x}^{(0)} = \mathbf{y}$)

$$\left(\frac{1}{\omega}\mathbf{D} + \mathbf{L}\right)\mathbf{x}^{(n+1)} = \mathbf{y} + \left(\frac{1-\omega}{\omega}\mathbf{D} - \mathbf{L}^t\right)\mathbf{x}^{(n)}.$$

- Successive cancellation

Theorem

Gauss-Seidel iteration is convergent for symmetric positive definite $\mathbf{M}$ and $\omega \in (0, 2)$.

Descent Algorithms

- For symmetric positive definite $\mathbf{M}$, define

$$\|\mathbf{x}\|_{\mathbf{M}^{-\frac{1}{2}}} = \|\mathbf{M}^{-\frac{1}{2}}\mathbf{x}\|_2 = \mathbf{x}^t\mathbf{M}^{-1}\mathbf{x},$$

- Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{M}\mathbf{x} - \mathbf{y}\|_{\mathbf{M}^{-\frac{1}{2}}}$. Least-squares minimization is solution to

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^K} f(\mathbf{x}).$$

- Note

$$\nabla(\mathbf{x}) \triangleq \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_K}\right)^t = \mathbf{M}\mathbf{x} - \mathbf{y}.$$

- Gradient is equal to the error vector $\mathbf{e} = \mathbf{M}\mathbf{x} - \mathbf{y}$, and unique stationary point is $\mathbf{M}\mathbf{x} = \mathbf{y}$.

Descent Algorithms

- Descent algorithms take the form

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + t_n \mathbf{d}^{(n)}$$

- Direction $\mathbf{d}^{(n)}$ chosen to reduce objective function
- Step size t_n chosen each iteration to minimize objective function in the direction of $\mathbf{d}^{(n)}$.

Steepest Descent

- Search direction is negative gradient of objective function, $\mathbf{d}^{(n)} = -\nabla(\mathbf{x}^{(n)})$, resulting in

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n+1)} - t_n \nabla(\hat{\mathbf{x}}^{(n)}) \tag{1}$$

$$= t_n \mathbf{y} - (t_n \mathbf{M} - \mathbf{I})\hat{\mathbf{x}}^{(n)}, \tag{2}$$

where the optimal step size is

$$t_n = \frac{\|\mathbf{e}^{(n)}\|_2}{\|\mathbf{M}^{\frac{1}{2}}\mathbf{e}^{(n)}\|_2}. \tag{3}$$

- Jacobi and Gauss-Seidel are steepest descent algorithms with suboptimal t_n.

Steepest Descent

Theorem

The error norm of the steepest descent algorithm with optimal step size decreases geometrically with rate at least

$$\left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)$$

where $\lambda_{\min}$ *and* $\lambda_{\max}$ *are the smallest and largest eigenvalues of* $\mathbf{M}$.

Conjugate Gradient

- Better approach: new search direction *orthogonal* to all previous directions $(\mathbf{d}^{(n+1)}, \mathbf{M}\mathbf{d}^{(j)}) = 0, \quad j = 0, 1, \ldots, n$.
- Choose a linear combination of current error vector (steepest descent) and previous direction, where the combining coefficient β_n ensures orthogonality.

$$\begin{aligned}
\mathbf{d}^{(0)} &= -\mathbf{e}^{(0)} \\
\mathbf{d}^{(n+1)} &= -\mathbf{e}^{(n+1)} + \beta_n \mathbf{d}^{(n)} \quad \text{where} \\
\beta_n &= \frac{(\mathbf{e}^{(n+1)}, \mathbf{M}\mathbf{d}^{(n)})}{(\mathbf{d}^{(n)}, \mathbf{M}\mathbf{d}^{(n)})}.
\end{aligned}$$

Conjugate Gradient

Theorem

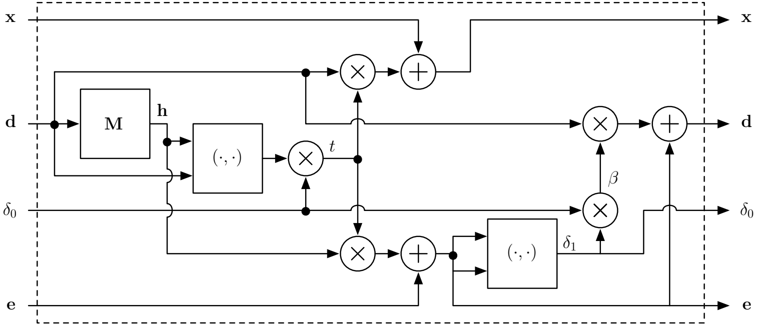
The error norm for the conjugate gradient algorithm decreases geometrically with rate at least

$$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)$$

where $\kappa = \lambda_{\max}/\lambda_{\min}$ *is the condition number of* $\mathbf{M}$*.*

Conjugate Gradient

- Efficient implementation: Only two extra inner products and one extra vector addition compared to parallel cancellation.



Outline

Constrained Optimization

- Linear detectors were solutions to unconstrained optimization

$$\hat{\mathbf{x}} = \arg\min \frac{1}{2} \|\mathbf{M}\mathbf{x} - \mathbf{y}\|_{\mathbf{M}^{-\frac{1}{2}}}.$$

- Ignores discrete alphabet of $\mathbf{x}$.
- Idea: Introduce manageable constraints to approximate discrete alphabet.

$$\begin{aligned} \min f(\mathbf{x}) \quad & \text{subject to} \\ g_i(\mathbf{x}) \leq 0 \quad & i = 1, 2, \ldots, m \end{aligned}$$

Constrained Optimization

- Use barrier functions to replace constrained probem with equivalent unconstrained problem

$$\min f(\mathbf{x}) + \underbrace{\sum_{i=1}^{m} I\left(g_i(\mathbf{x})\right)}_{\text{penalty function}}$$

where the ideal barrier function I is defined as

$$I(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0. \end{cases}$$

Constrained Optimization

- Approximate $I(\cdot)$ by differentiable approximation $b(u) \approx I(u)$
- Define *penalty function*

$$\varphi(\mathbf{x}) = \sum_{i=1}^{m} b(g_i(\mathbf{x})) \tag{4}$$

- New optimization problem

$$\min f(\mathbf{x}) + \varphi(\mathbf{x}),$$

Gradient Descent

- Gradient descent

$$\hat{\mathbf{x}}^{(n+1)} = t_n\mathbf{y} - (t_n\mathbf{M} - \mathbf{I})\hat{\mathbf{x}}^{(n)} - t_n\varphi'\left(\hat{\mathbf{x}}^{(n)}\right).$$

- Setting $t_n = 1$

$$\hat{\mathbf{x}}^{(n+1)} + \varphi'\left(\hat{\mathbf{x}}^{(n)}\right) = \mathbf{y} - (\mathbf{M} - \mathbf{I})\hat{\mathbf{x}}^{(n)}.$$

- Supposing $\xi(\mathbf{x}) = \mathbf{x} + \varphi'(\mathbf{x})$ has inverse function $\zeta = \xi^{-1}$,

$$\hat{\mathbf{x}}^{(n+1)} = \zeta\left(\mathbf{y} - (\mathbf{M} - \mathbf{I})\hat{\mathbf{x}}^{(n)}\right),$$

- Nonlinear parallel interference cancellation!

Nonlinear Iteration

Theorem

The non-linear iteration

$$\hat{\mathbf{x}}^{(n+1)} = \zeta\left(\mathbf{y} - (\mathbf{M} - \mathbf{I})\hat{\mathbf{x}}^{(n)}\right)$$

is a gradient method for numerical solution of

$$\min \frac{1}{2}\|\mathbf{M}\mathbf{x} - \mathbf{y}\|_{\mathbf{M}^{-\frac{1}{2}}} + \varphi(\mathbf{x})$$

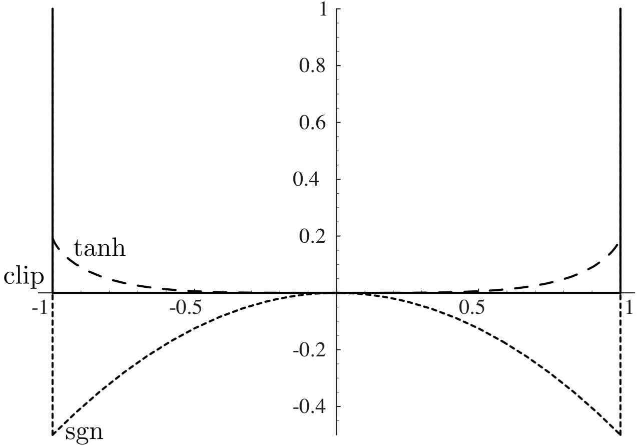
where φ *satisfies*

$$\zeta^{-1}(\mathbf{x}) = \varphi'(\mathbf{x}) + \mathbf{x}.$$

If φ *is convex, then the iteration is convergent to the unique point* $\hat{\mathbf{x}}$ *satisfying*

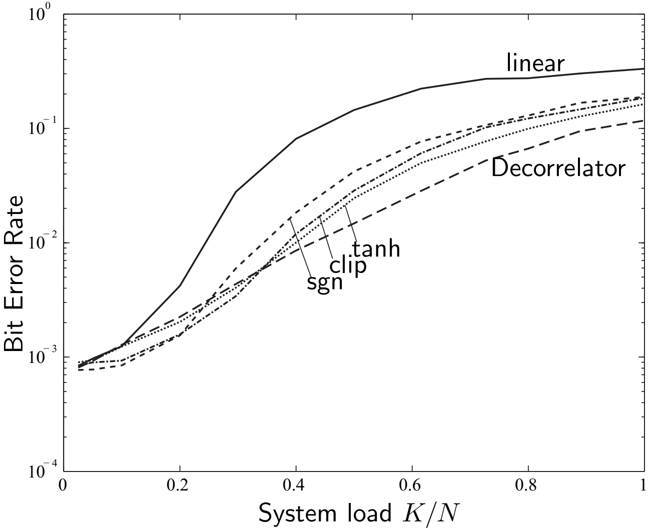
$$\mathbf{M}\hat{\mathbf{x}} + \varphi'(\hat{\mathbf{x}}) = \mathbf{y}.$$

Tentative Decision Functions

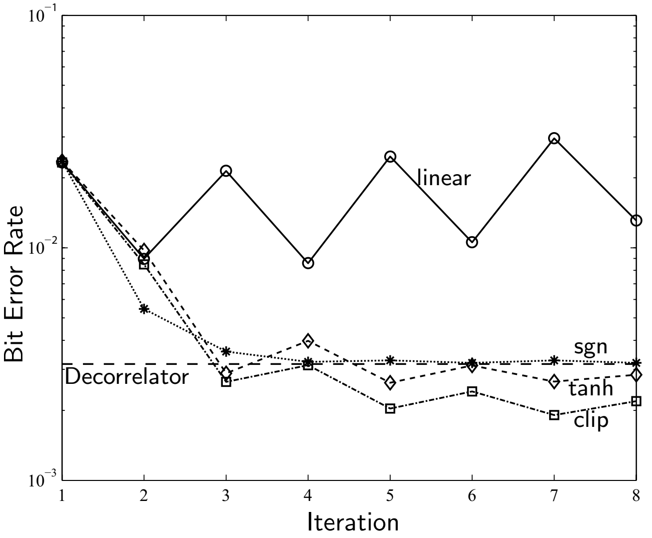


Outline

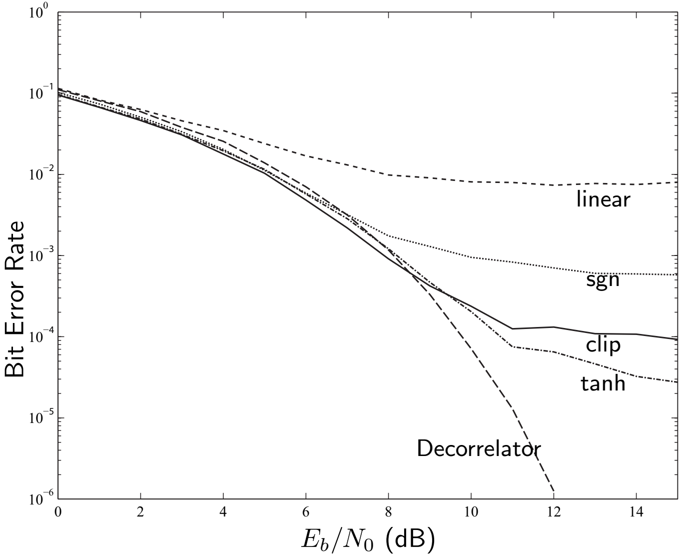
Parallel cancellation decorrelator, $K = 8$, $E_b/N_0 = 7$ dB.



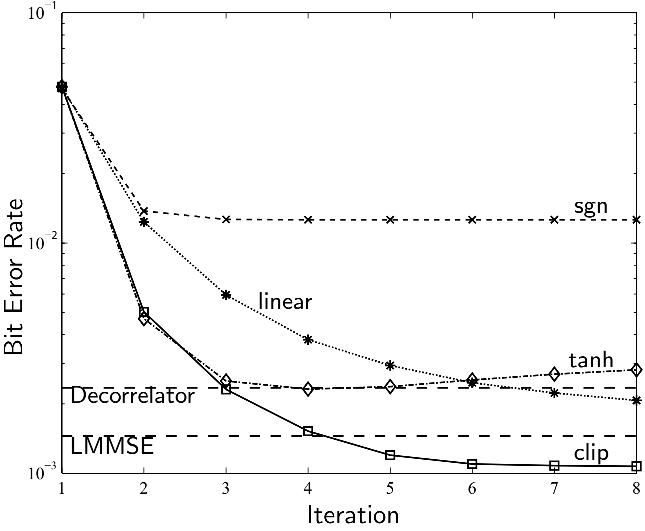
Parallel cancellation, $K = 8$, $N = 32$, $E_b/N_0 = 7$ dB.



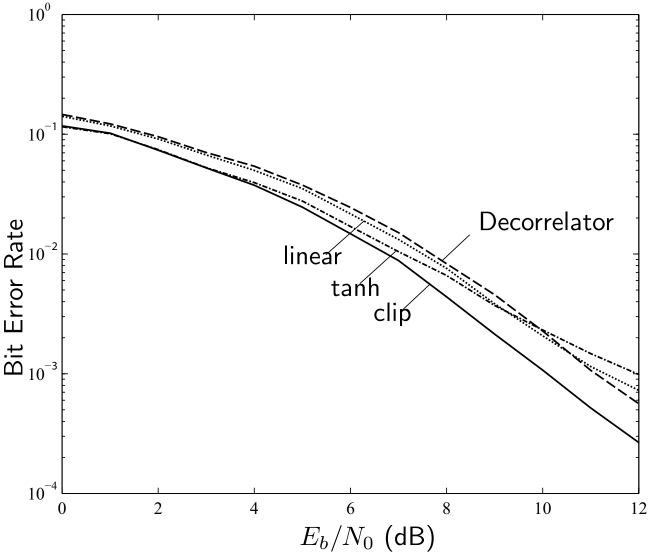
Parallel, $K = 8$, $N = 32$, 8 iterations.



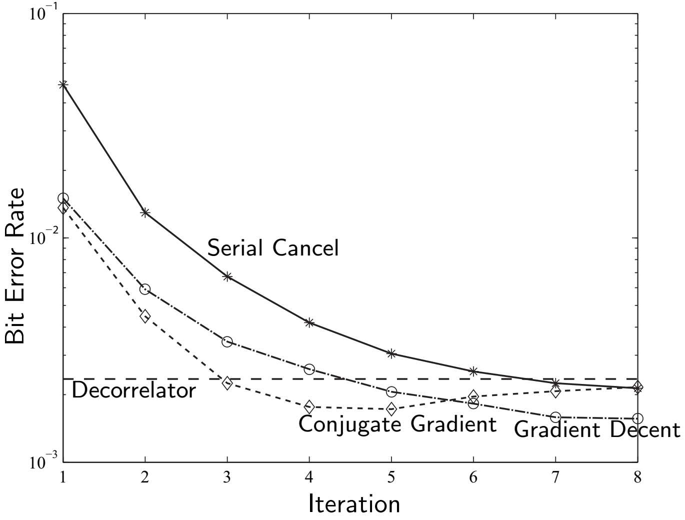
Serial, $K = 8$, $N = 16$, $E_b/N_0 = 10$ dB.



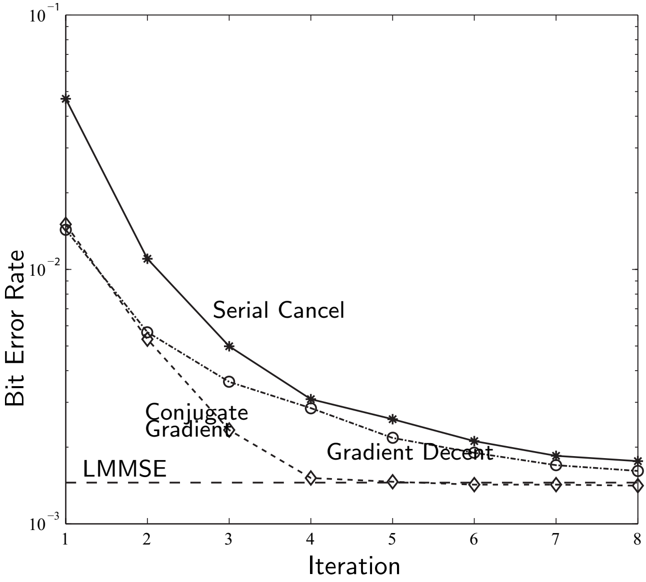
Serial, $K = 8$, $N = 16$, 8 iterations.



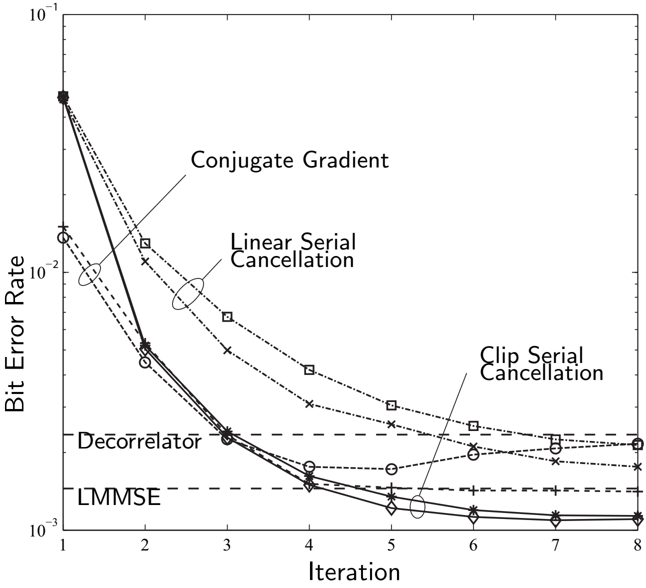
Descent algorithms, $K = 8$, $N = 16$, $E_b/N_0 = 10$ dB



Descent, LMMSE, $K = 8$, $N = 16$, $E_b/N_0 = 10$ dB.



Decorrelator & LMMSE, $K = 8$, $N = 16$, $E_b/N_0 = 10$ dB.



For more information...

A. Grant and L. Rasmussen,
"Iterative Techniques,"
Chapter 3 in *Advances in Multiuser Detection*, Wiley, 2009.