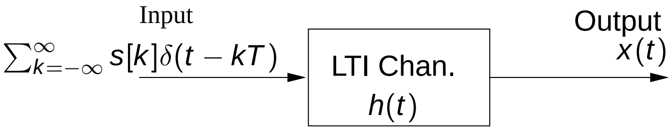
Blind Equalization

Christoph Krall
christoph.krall@tugraz.at

10. May 2006 / ASP Seminar

Outline

Definition of the Problem



- Received data in communications is distorted linearly by the channel.
- No Channel State Information (CSI) available at receiver.
- Additionally, we have also some noise $w[k]$.
- But, symbols $\{s[k]\}$ belong to the alphabet $\mathcal{A}$

Objective for the equalizer, $G(z, \theta)$

- Make something to get a distortionless channel, i.e.,

$$h[k] = a\delta[k - \Delta]$$

  such that the output data symbols are just scaled and delayed.
- Generally two approaches known (shown here in freq.-domain)
  - *Zero forcing* equalization

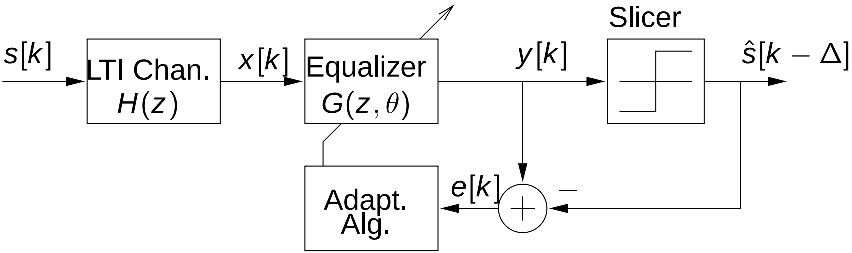
$$G_{zf}(z, \theta) = \frac{z^{-\Delta}}{H(z)}$$

  - *Minimum MSE* equalization

$$G_{mmse}(z, \theta) = \frac{H^*(z)z^{-\Delta}}{H(z)H^*(z^{-1}) + S_w(f)}$$

Objective for the *blind* equalizer

- find a parameter vector θ such that the overall response comes close to a distortion-less channel.
- Big problem: no training available (possible).
- One possible solution: decision directed learning



- Adaptation with symbol estimates $\hat{s}[k-\Delta] = Q(y[k])$, thus quality depends stronlgy on the estimator $Q()$.

Basic facts about blind equalization

- Second order statistics of $x[k]$ alone only provide the magnitude information of the channel and are insufficient for blind equalization of a mixed phase channel $H(z)$ containing zeros inside and outside the unit circle.
- A mixed phase linear channel $H(z)$ cannot be identified from its outputs when the input signal is i.i.d. Gaussian, since only second order statistics is available
- Although the exact inverse of a non-minimum phase channel is unstable, a truncated anti-causal expansion can be delayed by Δ to allow causal FIR approximation of a ZF equalizer.
- ZF equalizers cannot be implemented for channels $H(z)$ with zeros on the unit circle. Any FIR approximation will have unbounded approximation error.

Basic facts about blind equalization, cntd.

- Due to the absence of training, blind equalization has to exploit other available information contained in $x[k]$, i.e.,
  - Power Spectral Density (PSD) of $x[k]$ is given by

$$S_x(\omega) = S_s(\omega)|H(e^{j\omega})|^2 + S_w(\omega)$$

  - Higher Order Statistics (HOS), e.g. trispectrum

$$\begin{aligned} T_x(\omega_1,\omega_2,\omega_3) =& T_s(\omega_1,\omega_2,\omega_3) H(e^{j\omega_1}) H(e^{j\omega_2}) \\ & \times H(e^{j\omega_3}) H(e^{-j(\omega_1+\omega_2+\omega_3)}) + T_w(\omega_1,\omega_2,\omega_3) \end{aligned}$$

    for i.i.d. input and Gauss. noise the trispectra are const.,

$$T_s(\omega_1,\omega_2,\omega_3) = \gamma_s, \quad T_w(\omega_1,\omega_2,\omega_3) = 0$$

  - and phase information is obtained by

$$\begin{aligned} \angle T_x(\omega_1,\omega_2,\omega_3) =& \angle H(e^{j\omega_1}) + \angle H(e^{j\omega_2}) + \angle H(e^{j\omega_3}) \\ & - \angle H(e^{-j(\omega_1+\omega_2+\omega_3)}) + \angle \gamma_s \end{aligned}$$

FIR Linear Equalizers

- Generally two different approaches,
  - Stochastic Gradient Descent (SGD)
    Iteratively minimizes a cost function over all possible choices of equalizer coefficients
  - Stationary Statistics
    Sufficient statistics are collected over a block of data, and explotis HOS properties
- Some nomenclatures:
  - The equalizer is a FIR filter $G(z, \theta)$
  - parameter vector: $\theta = [\theta_0 \; \theta_1 \; \ldots \; \theta_m]^T$
  - block of received symbols:
    $\mathbf{x}[k] = [x[k] \; x[k-1] \; \ldots \; x[k-m]]$
  - Equalizer in frequency domain: $G(z, \theta) = \sum_{i=0}^{m} \theta_i z^{-i}$

General ideas

- We want to adjust the parameters $\theta(k)$ such that the output of the equalizer achieves a scaled, delayed version of the data symbols.
- We assume a finite alphabet of M levels

$$\{\pm(M-1)d,\ \pm(m-3)d,\ \ldots,\ \pm 3d,\ \pm d\}, \quad M \text{ even}$$

- Hint: We consider real signals, extension to complex signals is straightforward.
- We define a cost function which shall be a measure of the ISI in the equalizer output, i.e., the *mean cost function*

$$J(\theta) = E\{\Psi(y[k])\}$$

- $\Psi(\cdot)$ is a scalar function of the equalizer output, even to distinguish between $\pm$ levels.

General ideas, cntd.

- Not necessarily a min. MSE equalizer is found by optimization (minimum ISI equalizer).
- Stochastic gradient descent algorithm is then given as:

$$\begin{aligned}\theta(k+1) &= \theta(k) - \mu \frac{\partial}{\partial \theta(k)} \Psi(y[k]) \\ &= \theta(k) - \mu \Psi' \left(\theta(k)^T \mathbf{x}[k] \right) \mathbf{x}[k]\end{aligned}$$

- So we can define the derivative of Ψ (error function) as

$$\psi(x) = \Psi'(x) = \frac{\partial}{\partial x} \Psi(x)$$

- Thus a general *blind* algorithm is defined as

$$\theta(k+1) = \theta(k) - \mu \psi \left(\theta(k)^T \mathbf{x}[k] \right) \mathbf{x}[k]$$

Sato Algorithm (1975)

- Developed for multilevel PAM (same as DD when $s[k] = \pm 1$)
- Error function is given as

$$\psi(y[k]) = y[k] - R_1 \mathrm{sgn}(y[k])$$

where

$$R_1 = \frac{E\{|s[k]|^2\}}{E\{|s[k]|\}}$$

- Thus the parameter vector is updated as

$$\begin{aligned}\theta(k+1) &= \theta(k) - \psi(y[k])\mathbf{x}[k] \\ &= \theta(k) - [y[k] - R_1 \mathrm{sgn}(y[k])]\mathbf{x}[k]\end{aligned}$$

BGR Algorithms (app. 1980)

- Extension of Sato's Algorithm by *Benveniste, Goursat* and *Ruget*
- They are defined by a class of error functions

$$\psi(y[k]) = \tilde{\psi}(y[k]) - R_b \operatorname{sgn}(y[k])$$

where

$$R_b = \frac{E\{\tilde{\psi}(s[k])s[k]\}}{E\{|s[k]|\}}$$

- $\tilde{\psi}(y[k])$ is an odd and twice differentiable function

$$\tilde{\psi}''(x) \geq 0, \forall x \geq 0$$

- Sato would have $\tilde{\psi}(x) = x$

"Stop-and-Go" Algorithms (app. 1987)

- To avoid convergence to local minima in the cost function (poor performance) *Picchi* and *Prati* invented "stop-and-go" methodology
- Idea: continue adapting the filter when error function is more likely to have the correct sign for the gradient descent direction.
- Example:
  Consider two algorithms with error functions $\psi_1(y)$ and $\psi_2(y)$, respectively. "Stop-and-Go" would mean

$$\theta(k+1) = \begin{cases} \theta(k) - \mu\psi_1(y[k])\mathbf{x}[k], & \text{if } \operatorname{sgn}[\psi_1(y[k])] = \operatorname{sgn}[\psi_2(y[k])] \\ \theta(k), & \text{if } \operatorname{sgn}[\psi_1(y[k])] \neq \operatorname{sgn}[\psi_2(y[k])] \end{cases}$$

Constant Modulus (Godard) Algorithms (app. 1980)

- Integrating Sato's error function $\psi_1(x)$ yields

$$\Psi_1(y[k]) = 1/2(|y[k]| - R_1)^2$$

- Generalized by Godard to

$$\Psi_q(y[k]) = 1/(2q)(|y[k]|^q - R_q)^2, \quad q = 1, 2, \ldots$$

- Resulting in an update equation

$$\theta(k+1) = \theta(k) - \mu(|y[k]|^q - R_q)|y[k]|^{q-2} y[k]\mathbf{x}[k]$$

- for the special case of $q = 2$ the algorithm is called "Constant Modulus Algorithm", i.e. the channel signal has a constant modulus $|s[k]|^2 = R_2$

Constant Modulus (Godard) Algorithms (app. 1980), cntd.

- CMA penalizes all signals which are not constant modulus.
- Advantage: equalization and carrier recovery are independent.
- A carrier frequency offset Δ_f causes phase rotation of the output

$$y[k] = |y[k]| \exp[j(2\pi\Delta_f k + \varphi[k])]$$

- CMA cost function is insensitive to the phase!
- Possible application of CMA in analog modulation signals such as FM or PM.

Shalvi and Weinstein Algorithms (1990)

- Based on Higher Order Statistics (HOS)
- *Kurtosis* is defined as

$$K_y = E\{|y[k]|^4\} - 2(E\{|y[k]^2|\})^2 - |E\{y[k]^2\}|^2$$

- Minimization of $|K_y|$ w.r.t. a constant power constraint

$$E\{|y[k]^2|\} = E\{|s[k]^2|\}$$

- It can be shown that for $s[k]$ i.i.d.

$$E\{|y[k]^2|\} = E\{|s[k]^2|\} \sum_{i=-\infty}^{\infty} |c[i]|^2$$

where $c[i] = \theta * h$.

Shalvi and Weinstein Algorithms (1990), cntd.

- and for the Kurtosis

$$K_y = K_s \sum_i |c[i]|^4$$

- If $E\{|y[k]^2|\} = E\{|s[k]^2|\}$ then $|K_y| = |K_s|$ only when

$$\mathbf{c} = [0\,0\,0\,1\,0\,0\,]$$

- Thus the Shalvi-Weinstein equalizer operates according to the criterion

$$\text{maximize } \sum_i |c[i]|^4 \text{ w.r.t. } \sum_i |c[i]|^2 = 1$$

Summary

- Generally we can say that blind equalization is a very tricky task.
- However, we have seen that there are two main approaches:
  - Define a cost function and minimize it
  - HOS methods for blind deconvolution
- Popular Algorithms
  - Sato, BGR, Stop-and-Go, CMA, Shalvi-Weinstein Algorithm
- Hard to analyze convergence behavior of the algorithms
- Still ongoing research with all different kinds of error functions

Thanks

Thanks for your attention!