F2E5216/TS1002 Adaptive Filtering and Change Detection

Fredrik Gustafsson (LiTH) and Bo Wahlberg (KTH)

Lecture 4

Adaptive filtering applications

- Summary
- Communication problems: blind and non-blind equalization
- Noise cancellation
- Vehicular problems: friction estimation

Signal Models

$$\begin{aligned} y_t &= G(q;\theta)u_t + H(q;\theta)e_t \\ A(q;\theta)y_t &= \frac{B(q;\theta)}{D(q;\theta)}u_t + \frac{C(q;\theta)}{F(q;\theta)}e_t \\ y_t &= \varphi_t^T\theta + e_t \text{ or } \varphi_t^T(\theta)\theta + e_t \end{aligned}$$

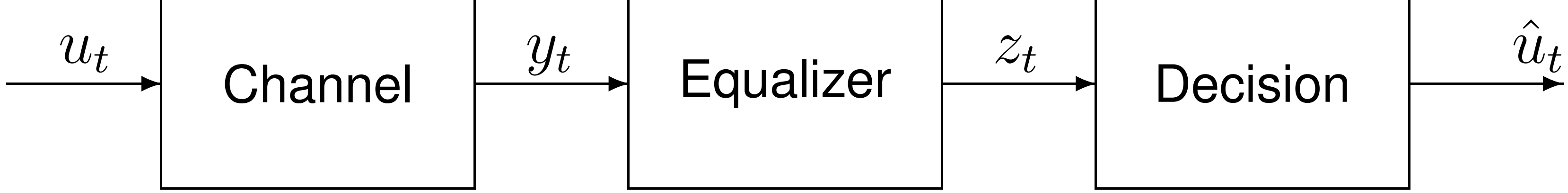
Optimization algorithms: Gauss-Newton, steepest descent

$$V(\theta) = \hat{\mathrm{E}}\varepsilon_t^2(\theta) = \hat{\mathrm{E}}(y_t - \varphi_t^T\theta)^2$$

Algorithms: LMS, NLMS, RLS, WLS, KF

$$\hat{\theta}_{t+1} = \hat{\theta}_t + K_t\varepsilon_t,$$

Adaptive Blind Equalization



The channel and equalizer are modeled as FIR filters

$$\begin{aligned} B(q) &= b_1(t)q^{-1} + b_2(t)q^{-2} ... + b_{n_b}(t)q^{-n_b} \\ C(q) &= c_1(t)q^{-1} + c_2(t)q^{-2} ... + c_{n_c}(t)q^{-n_c} \end{aligned}$$

Overall response

$$h(t) = b_{channel} * c_{equalizer}(t)$$

Good performance is defined by

$$h(t) \approx m\delta(t-k), \;\; |m| = 1$$

Steepest Descent Algorithms

Loss functions

$$V = \mathrm{E}[(1 - z^2)^2] \qquad \text{modulus restoral (Godard)}$$

$$V = \mathrm{E}[(\mathrm{sign}(z) - z)^2] \qquad \text{decision feedback (Sato)}$$

Algorithms

$$\begin{aligned}
\varphi_t &= (y_{t-1}, y_{t-2}, .., y_{t-n})^T \\
z_t &= \varphi_t^T \hat{\theta}_{t-1} \\
\varepsilon_t^{Sato} &= \mathrm{sign}(z_t) - z_t \\
\varepsilon_t^{Godard} &= z_t(1 - z_t^2) \\
\hat{\theta}_t &= \hat{\theta}_{t-1} + \mu \varphi_t \varepsilon_t \\
\hat{u}(t) &= |z(t)|
\end{aligned}$$

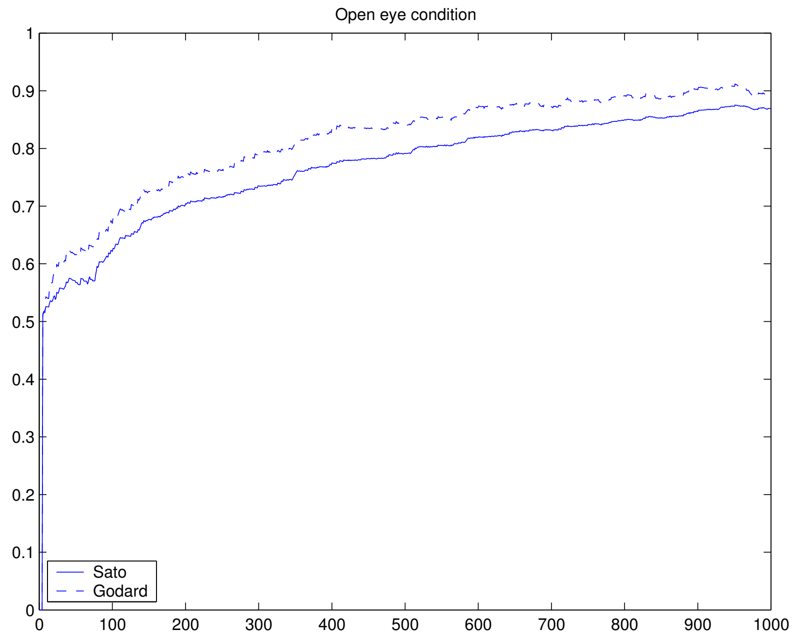
Performance Measure

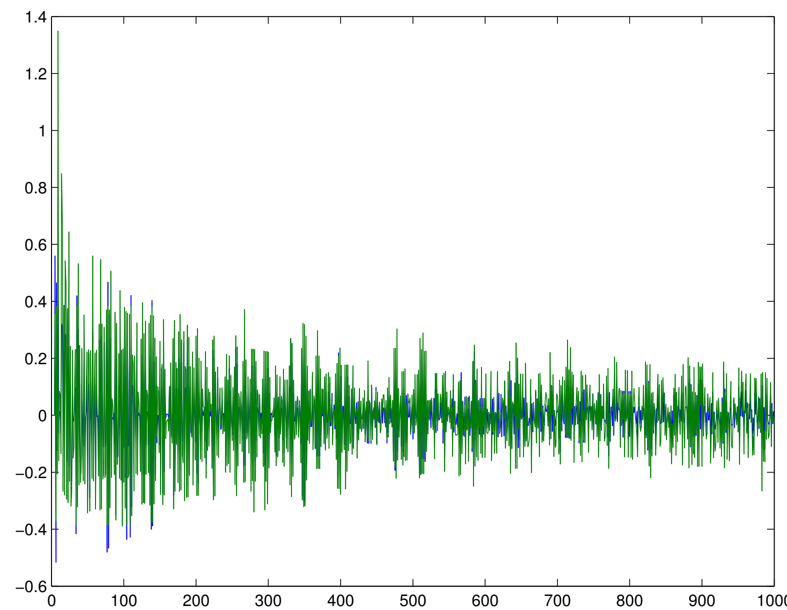
Consider the case of input alphabet $u(t) = \pm 1$. For successful demodulation when no noise is guaranteed if the largest component of $h(t)$ is larger than the sum of the other components. That is, $m(t) > 0$, where

$$m(t) = 2 - \frac{\sum(|h(t)|)}{\max_t |h(t)|}.$$

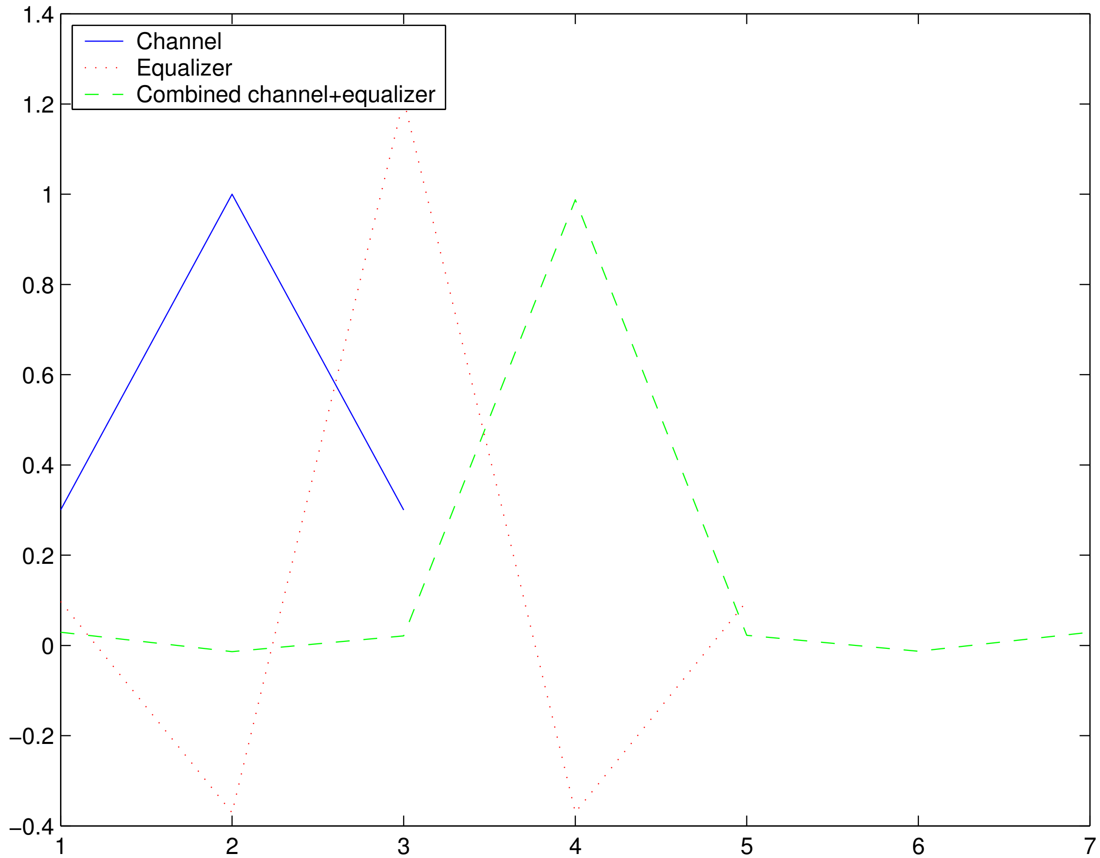
Perfect equalizer means $m(t) = 1$.
Open-eye condition corresponds to $m(t) > 0$.

Example





Open-eye measure and residuals for Sato's and Godard's algorithms.

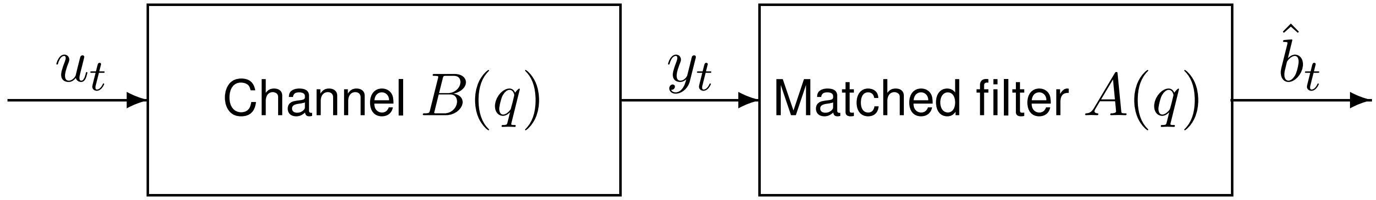


Impulse response of channel, equalizer and combined channel–equalizer (almost an impulse function).

Equalization

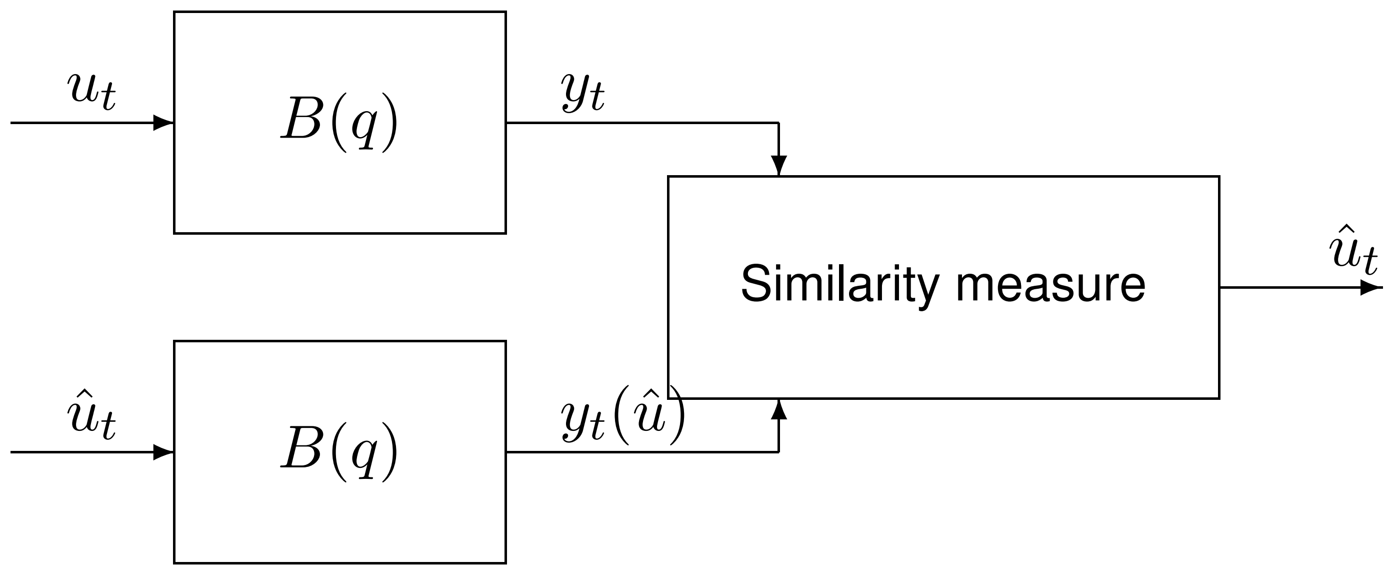
Two phases:

1. Channel estimation from training sequence.



Equalization, cont.

2. Input estimation using the Viterbi algorithm.



Channel Estimation

The standard model for channels is the FIR model

$$y_t = \sum_{k=1}^{n} b_k u_{t-k} = \varphi_t^T \mathbf{b}$$

where

$$\begin{aligned} \mathbf{b} &= (b_1, b_2, \ldots, b_n)^T \\ \varphi_t &= (u_{t-1}, u_{t-2}, \ldots, u_{t-n})^T \end{aligned}$$

Parameter estimation using LS:

$$\hat{\mathbf{b}} = \left(\sum_{t=1}^{N} \varphi_t \varphi_t^T \right)^{-1} \sum_{t=1}^{N} \varphi_t y_t = \sum_{t=1}^{N} \bar{\varphi}_t y_t$$

Matched Filter

Matched filter avoids the matrix inversion and matrix times vector multiplication, when the training sequence is known a prior.

Def. of matched filter: $(a * u)_t \approx \delta_0$.
Here

$$\hat{\mathbf{b}}(t) = (a * y)_t = (a * b * u)_t = (b * (a * u))_t \approx b_t$$

Identifying the equations gives:

$$(a * y)_t = \sum_{k=1}^{N} a_{t-k} y_k \;\Rightarrow\; a_{t-k} = \bar{\varphi}_t^{(k)}.$$

The Viterbi Algorithm

1. Enumerate all possible input sequences within the sliding window of size n (the channel model length). Append these candidates to the estimated sequence $\hat{u}^{t-n}$.

2. Filter all candidate sequences with $B(q)$.

3. Estimate the input sequence by the best possible candidate sequence. That is,

$$\hat{u} = \arg \min_{u_1, u_2, \ldots, u_N} \sum_{t=1}^{N} (B(q)u_t - y_t)^2$$

Theorem: This is the optimal ML estimator if noise is Gaussian and for a known FIR(n) channel.

In practise use an estimate of B.

Proof of the Viterbi Algorithm

From Bayes' rule

$$
\begin{aligned}
p(y^t|u^t) &= p(y_t|u^t, y^{t-1})p(y^{t-1}|u^t) \\
&= p(y_t|u^t_{t-n_b+1}, y^{t-1})p(y^{t-1}|u^{t-1}_{t-n_b+1}, u^{t-n_b})
\end{aligned}
$$

Suppose that we have computed the most likely sequence at time $t-1$ for each sequence $u^{t-1}_{t-n_b+1}$, so we have $p(y^{t-1}|u^t_{t-n_b+1}, u^{t-n_b})$ as a function $u^t_{t-n_b+1}$. Here we have also used the already calculated optimal u^{t-n_b}. The most likely sequence at time t can now be found by just maximizing over $u^t_{t-n_b+1}$ by using the already optimal u^{t-n_b}.

Noise Free Example

Channel: $y_t = u_t + 0.5u_{t-1}$

Possible $u_{t-1}^t = (1,1),\ (1,-1),\ (-1,1),\ (-1,1)$

True input : $u^t = (1,-1,1,1) \Rightarrow y^t = (-0.5, 0.5, 1.5)$

At time 1: Calculate $u_1 + 0.5u_0$ for each possible sequence u_0^1, $\Rightarrow$ $1.5, -0.5, 0.5, -1.5$ and compare with the measured $y_1 = -0.5$. This gives the optimal estimates $\hat{u}_0^1 = (1,-1)$. Save also the cost function for $\hat{u}_0^1 = (1,1)$.

At time 2: Calculate $u_2 + 0.5u_1$ for each possible sequence u_1^2. Calculate the cost function $\sum_{t=1}^{2}(B(q)u_t - y_t)^2$ as a function of u_1^2 $\Rightarrow \hat{u}_1^2 = (-1,1)$

At time 3: ...

BER

For evaluation, the bit error is commonly used, and it is defined as

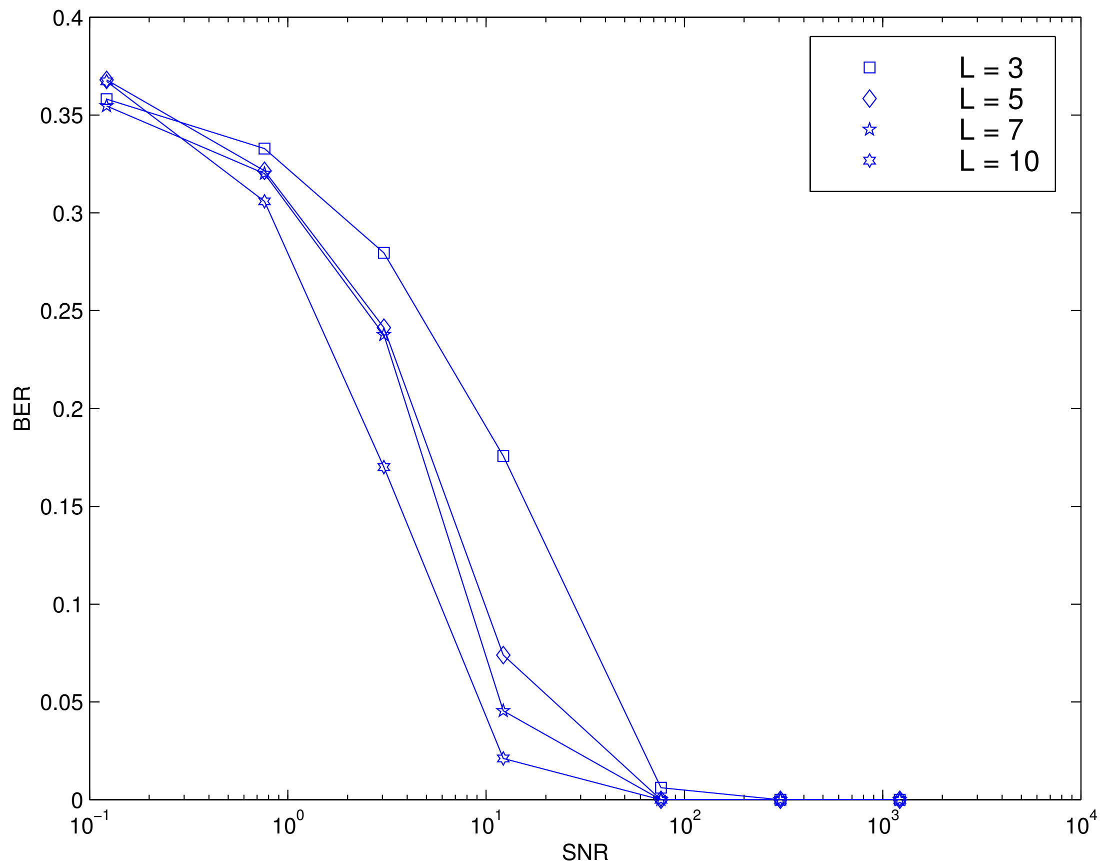
$$\text{BER} = \frac{\text{number of non-zero } (u_t - \hat{u}_t)}{N} \tag{1}$$

where trivial phase shifts (sign change, or modulus m) of the estimate should be discarded.

Example: Equalization

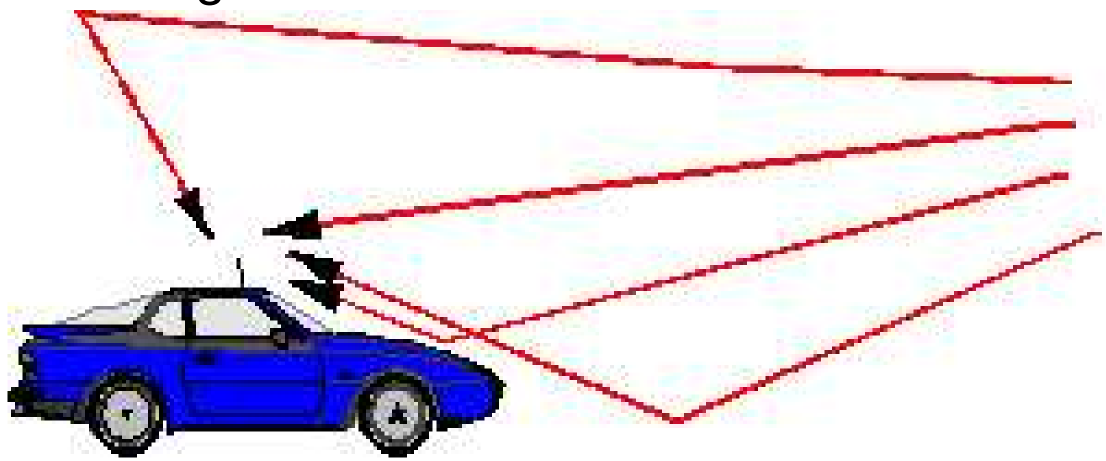
Channel $B(q) = q^{-1} + q^{-2} + q^{-3}$.

Investigate how signal to noise ratio influences BER for different lengths of the training sequence.


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
BER
10^-1
10^0
10^1
10^2
10^3
10^4
SNR
L = 3
L = 5
L = 7
L = 10

Equalization in GSM

Equalization of fading

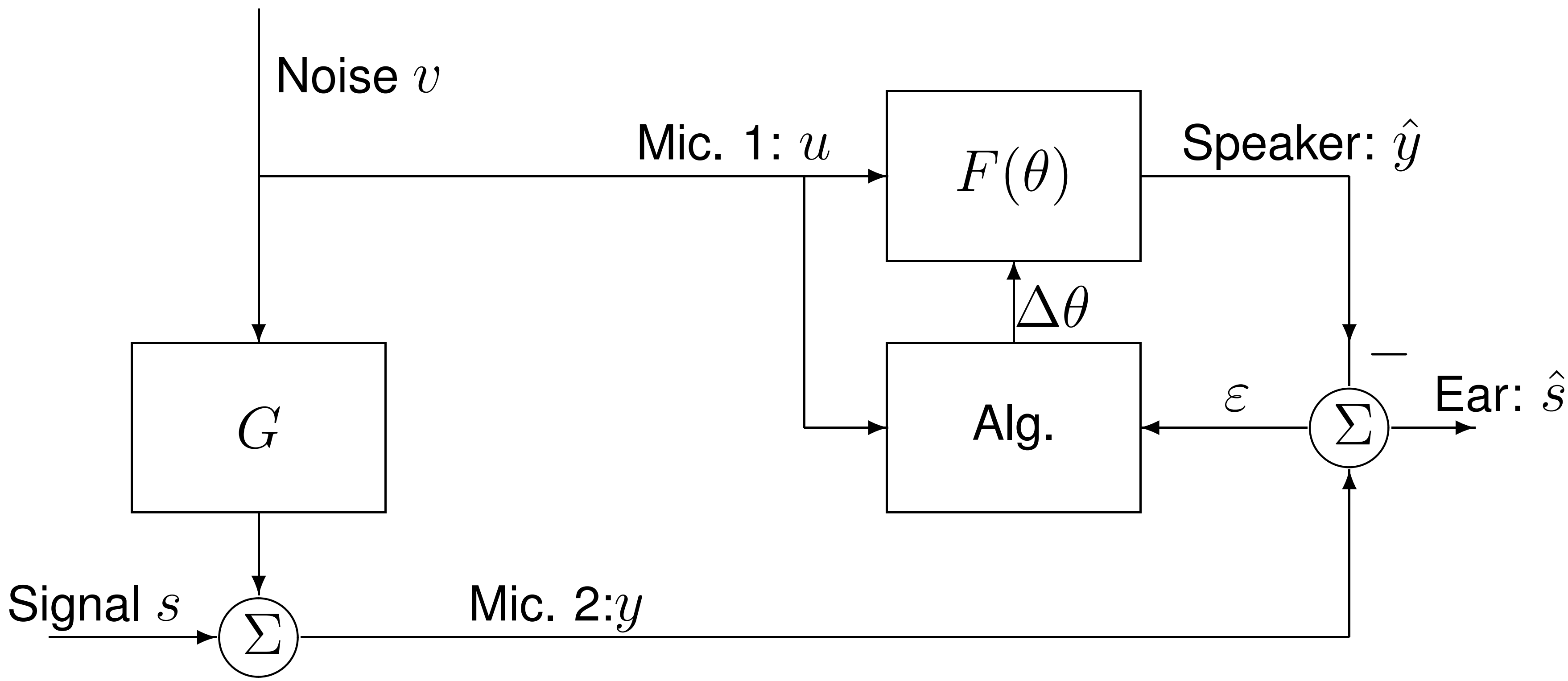


GSM transmits bursts of 148 bits. Of these, 16 bits located in the middle are the training sequence.

Optimization of all 2^{16} sequences gives the best pulse approximation, $g * a(t) \approx \delta(0)$.

The equalizer assumes channel FIR(4). This is sufficient to equalize Rayleigh fading for velocities up to 250 km/h.

Noise Cancellation



Undergraduate Lab in Linköping

Two radio stations, two microphones and headphones.

Task: use LMS on FIR(50) channel model, move mic or radio speakers and listen to the convergence!

SAAB 2000

SAAB 2000 is the most quite propeller airplane on the market partly due to active noise cancellation.

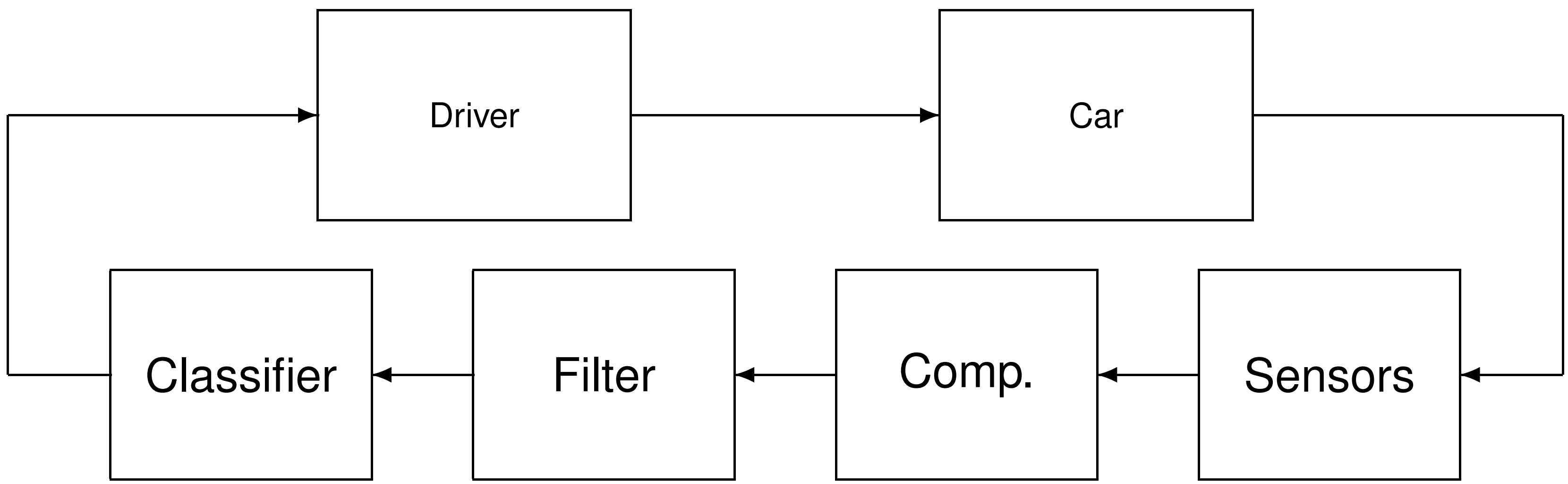
System consists of 64 microphones and 32 speakers.

Overall performance is 10 dB less engine noise for all passengers (only for normal positions in the seat).

Filtered X-LMS algorithm used. The transfer function from speakers to mic's are estimated off-line on the ground, and the regressor elements u are prefiltered with this filter.

Spin-off Company: Caran A2 Acoustics ”The singing window”

Friction Estimation



Description

Sensors	wheel velocities (ABS) load signal from engine braking light signal
State computation	velocity, yaw rate accelerations, gearing level, torques, slip, relative tire radii
Filtering	parameters in a friction model
Classification	friction/surface evaluation based on test drives.

Tire–Road Friction

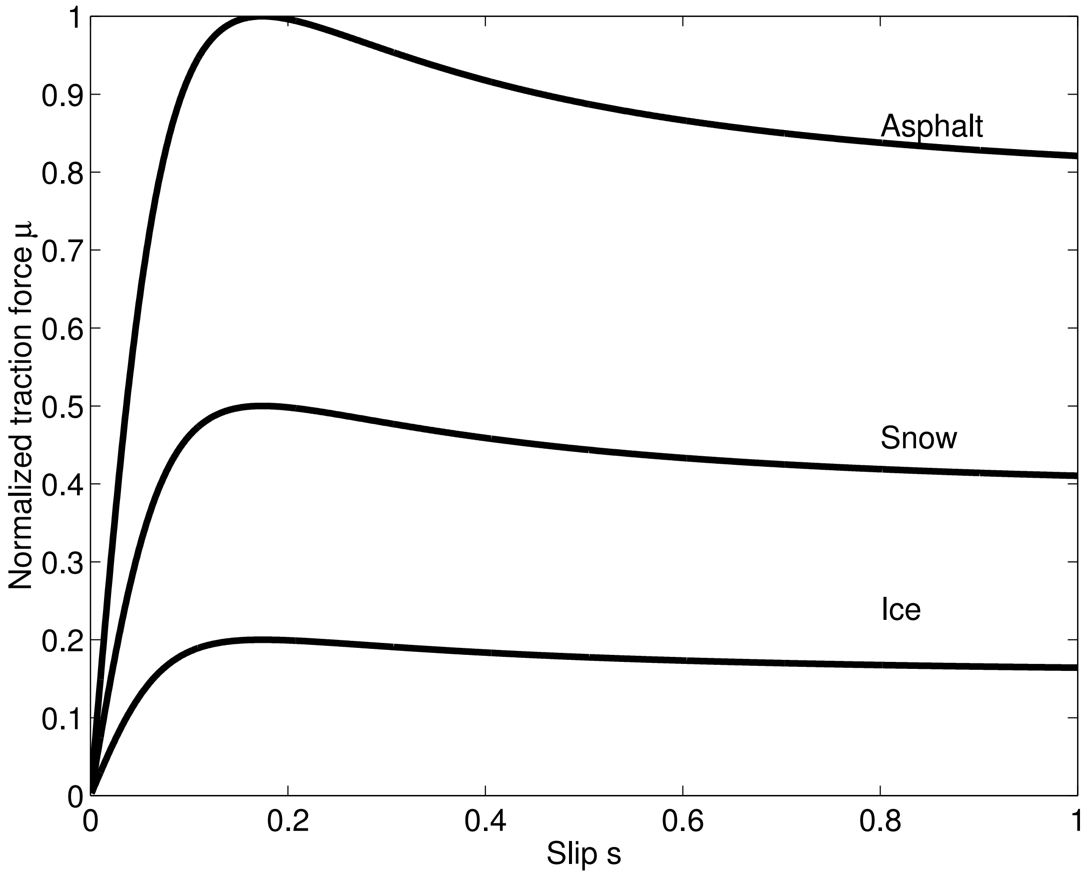
Slip: $s \triangleq \frac{\text{A driven wheel's circumferential velocity}}{\text{Its absolute velocity}} - 1$

Measurement: $s_m = \frac{\omega_{front}}{\omega_{rear}} - 1$

"Friction coefficient:" $\mu \triangleq \frac{F_f}{N}$

Measurement: μ_m from static engine model

Tire-road friction characterized by the slip curve, that is μ versus s.



Slip Slope

Question: Can μ_{max} be estimated from observations on small values of μ?

Definition Slip slope: $k \triangleq \left.\frac{d\mu}{ds}\right|_{\mu=0}$

In the literature, the slip slope is refereed to as the longitudinal stiffness, which is a pure tire characteristics.

Hypothesis: The slip slope is a (one-to-one?) function of the friction μ_{max}.

Filtering

Relations:

$$
\begin{aligned}
k &= \frac{F_f}{sN} &&= \frac{\mu}{s} \\
s_m &= \frac{\omega_d}{\omega_n} - 1 &&= s - \delta_w - \delta_R + e_s \\
\mu_m &= \mu + e_\mu
\end{aligned}
$$

N computable normal force
δ_w offset caused by tire radii differences
δ_R computable offset caused by cornering

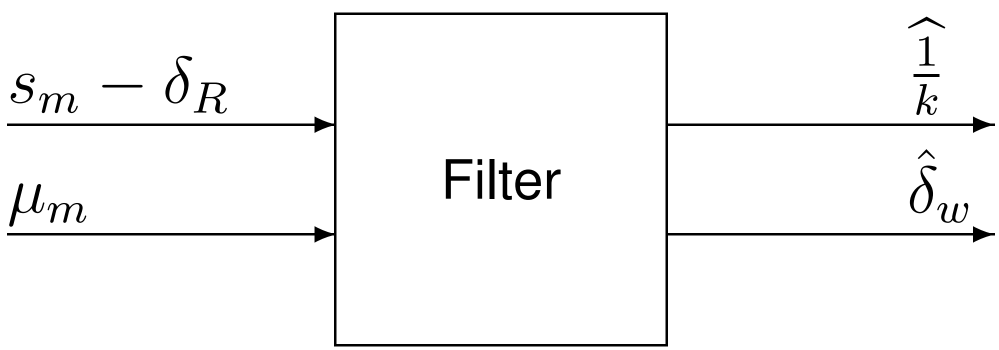
Variations in N and δ_R are very important to incorporate in curves.

Parameters: k time-varying and abruptly changing and δ_w slowly time-varying due to tire radii changes.

Regression Model:

$$s_m - \delta_R = \frac{1}{k}\mu_m + \delta_w$$

Parameter estimation filter:



Time-variations in slip slope k much faster than in the offset δ. RLS and LMS give the same tracking speed of all parameters (only one degree of freedom).

Kalman Filter

$\hat{k}(t)$ and $\hat{\delta}_w(t)$ are computed by a Kalman filter for the model

$$\begin{aligned} x &= (1/k,\ \delta_w)^T \\ y &= s - \delta_R \\ H &= (F_f/N,\ 1) \\ x(t+1) &= x(t) + w(t) \\ y(t) &= H(t)x(t) + e(t) \end{aligned}$$

Design:

$$\begin{aligned} R_1(t) &= \mathrm{diag}(r_k, r_\delta) \\ R_2 &= 1 \end{aligned}$$

Choose $r_k >> r_\delta$ which implies that k is tracked faster than δ.

Potential Problems:

- Outliers, some measurements don't make sense.
- Lack of excitation if Var(μ) is small $\Rightarrow$ large *parameter uncertainty* in estimates.
- Measurement error in μ introduces a *parameter bias*.

Parameter Uncertainty

Assume constant parameters for N samples.

$$P_N = \mathrm{Cov}(\hat{x}_N) = \sigma^2 \left(\sum_{t=1}^{N} H^T(t) H(t) \right)^{-1}$$

We can check how close this matrix is to singularity by computing its determinant

$$\det \left(\sum_{t=1}^{N} H^T(t) H(t) \right) = N^2 \overline{\mathrm{Var}}(\mu(t)).$$

Conclusions: Excitation of traction force important for parameter accuracy.

Noisy Regressors

Assume that

$$\mu_m(t) = \mu(t) + v_\mu(t), \quad \mathrm{Var}(v_\mu(t)) = \lambda_\mu$$

is used in the regressor $H(t)$. Straightforward calculations give

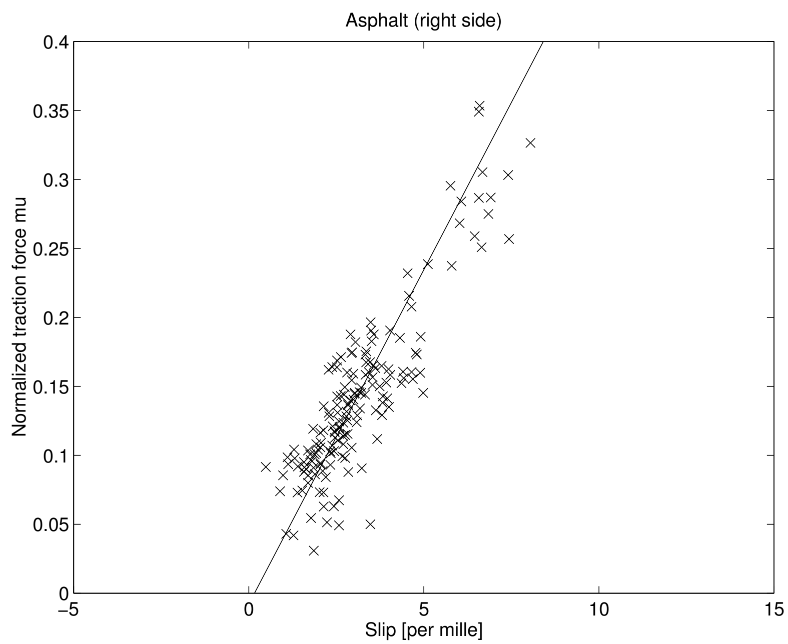
$$\hat{k} \approx k \frac{\overline{\mathrm{Var}}(\mu) + \lambda_\mu}{\overline{\mathrm{Var}}(\mu)} > k$$

Conclusions: Excitation of traction force important for parameter bias.

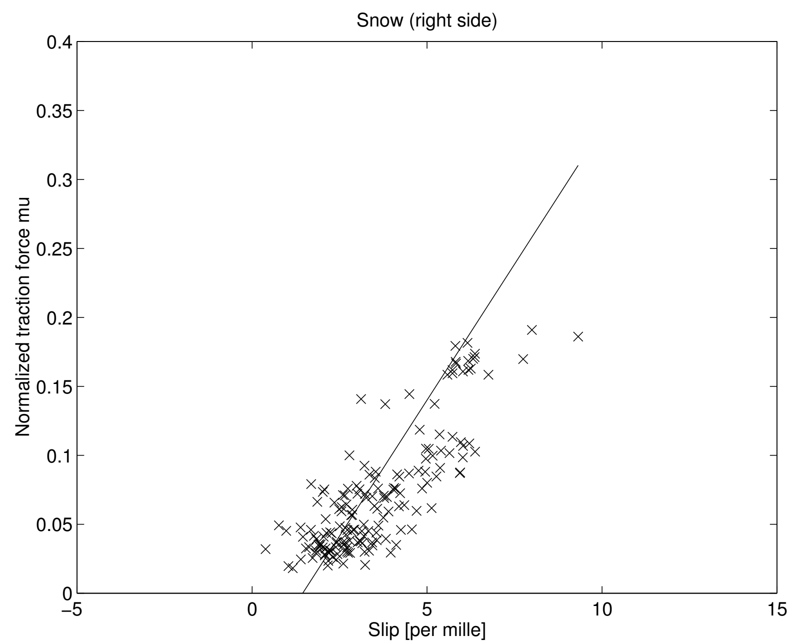
Generally: Noise in the regressor causes parameter bias

Example Modeling

Asphalt



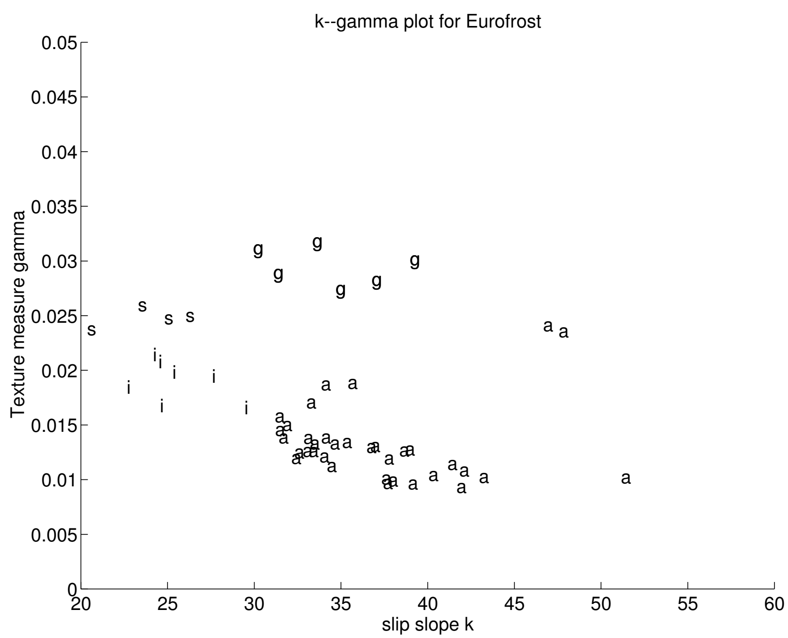
snow

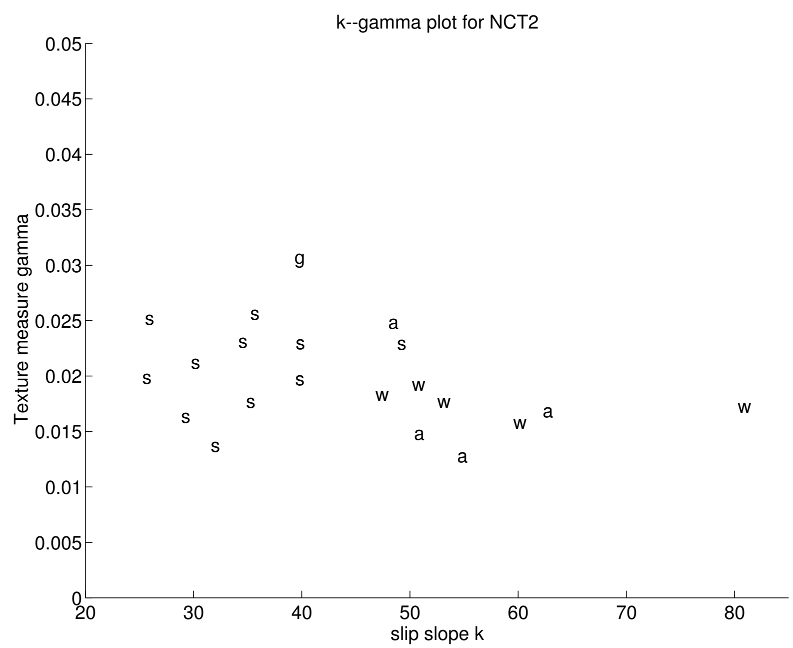


The lines correspond to the estimated slope and offset.

Classification

Test drives on asphalt (a), wet asphalt (w), gravel (g), ice (i) and snow (s). Plot of a texture measure versus slip slope for winter (left) and summer (right) tires.





A classifier is easy to construct for a given tire. Gravel is classified separately by the surface roughness.

Tests have shown fair robustness to tire pressure, weather and load.
Problems after tire change and cold start.

Exercises for Lectures 4

Exercise:19, 20, 23, 24, (25, 26, 27, 28, 29, 30, 31)

Next Time: Change Detection based on Parallel or Multiple Filters.

- Parallel filters as a consistency check or model validation
- Distance functions: constant variance
- Distance functions: the general case
- Diagnosis
- Multiple filter approaches
- Loss functions
- Minimization strategies