Adaptive Linear Filtering Using Interior Point Optimization Techniques

Lecturer: Tom Luo

Overview

A. Interior Point Least Squares (IPLS) Filtering

- Introduction to IPLS
- Recursive update of IPLS
- Convergence/transient analysis of IPLS

B. Applications

- System identification
- Beamforming
- Channel equalization in a CDMA forward link

Interior Point Optimization for Optimal Linear Filtering

- A discrete-time linear system can be described by

$$y_i = \mathbf{x}_i^T \mathbf{w}_* + v_i, \qquad i = 1, 2, \ldots$$

- Using input output pairs $\{\mathbf{x}_i, y_i\}$ the linear least-squares problem is then to estimate a filter $\mathbf{w}$ that minimizes the mean-squared error

$$\mathcal{F}_n(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i^T\mathbf{w}\right)^2 = \frac{1}{n}\mathbf{y}_n^T\mathbf{y}_n - 2\mathbf{w}_n^T\mathbf{p}_{xy}(n) + \mathbf{w}^T\mathbf{R}_{xx}(n)\mathbf{w}, \quad (1)$$

where $\mathbf{y}_n = [y_1, y_2, \ldots, y_n]^T$, $\mathbf{p}_{xy}(n) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i y_i$, $\mathbf{R}_{xx}(n) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$. Note: $\mathbf{p}_{xy}(n)$ and $\mathbf{R}_{xx}(n)$ are both recursively updatable with per-sample complexity of $O(M^2)$.

- The optimum linear filter then satisfies $\nabla\mathcal{F}_n(\mathbf{w}) = 0$, or $\mathbf{R}_{xx}(n)\mathbf{w} - \mathbf{p}_{xy}(n) = 0$.

One Motivation: Transient Convergence

- The RLS algorithm estimates (see *e.g.*, Sayed and Kailath '96)

$$\mathbf{w}_n^{rls} := \left[\frac{\delta}{n}\mathbf{I} + \mathbf{R}_{xx}(n)\right]^{-1} \mathbf{p}_{xy}(n),$$

  where $\frac{\delta}{n}\mathbf{I}$ is a *regularization* term to improve conditioning.

- **Problem:** The regularization depends entirely on the constant δ. If, for example,
  - SNR is underestimated $\Longrightarrow$ slower asymptotic convergence
  - SNR is overestimated $\Longrightarrow$ bad transient behaviour
- **Remedy:**

$$\boxed{\mathbf{w}_n := [\alpha_n \mathbf{I} + \mathbf{R}_{xx}(n)]^{-1}\mathbf{p}_{xy}(n), \qquad \alpha_n \ \text{adjusted adaptively}}$$

The Analytic Center Approach

- Formulate a *convex feasibility problem* at each iteration. $\mathbf{w}$ is a feasible filter only if it is contained in

$$\Omega_n = \{\mathbf{w} \in \mathbb{R}^M \mid \mathcal{F}_n(\mathbf{w}) \leq \tau_n,\ \|\mathbf{w}\|^2 \leq R^2\}, \tag{2}$$

1^{st} constraint: minimize the mean-squared error $\mathcal{F}_n(\mathbf{w})$.
2^{nd} constraint: make Ω_n a bounded region.

- The analytic center $\mathbf{w}_n^a$ of Ω_n is the minimizer of

$$\phi_n(\mathbf{w}) = -\log(\tau_n - \mathcal{F}_n(\mathbf{w})) - \log(R^2 - \|\mathbf{w}\|^2)$$

which can be found by solving $\nabla \phi_n(\mathbf{w}) = 0$

$$\frac{\nabla \mathcal{F}_n}{s_n(\mathbf{w}_n^a)} + \frac{2\mathbf{w}_n^a}{t_n(\mathbf{w}_n^a)} = 0, \quad \text{therefore} \quad \mathbf{w}_n^a = \left(\frac{s_n(\mathbf{w}_n^a)}{t_n(\mathbf{w}_n^a)}\mathbf{I} + \mathbf{R}_{xx}(n)\right)^{-1} \mathbf{p}_{xy}(n).$$

where $s_n(\mathbf{w}) := \tau_n - \mathcal{F}_n(\mathbf{w})$ and $t_n(\mathbf{w}) := R^2 - \|\mathbf{w}_n^a\|^2$.

Definition of τ_n

We are concerned with the behaviour of $\alpha_n = s_n/t_n$.

1. The goal is to make $\alpha_n \sim \|\nabla \mathcal{F}_n(\mathbf{w})\|$.

$$\nabla \mathcal{F}_n(\mathbf{w}) = 2(\mathbf{R}_{xx}(n)\mathbf{w} - \mathbf{p}_{xy}(n))$$

   Thus, if

   - $\|\nabla \mathcal{F}_n(\mathbf{w})\|$ is large $\Longrightarrow$ need α_n large for regularization.
   - $\|\nabla \mathcal{F}_n(\mathbf{w})\|$ is small $\Longrightarrow$ need only a small α_n.

2. Define

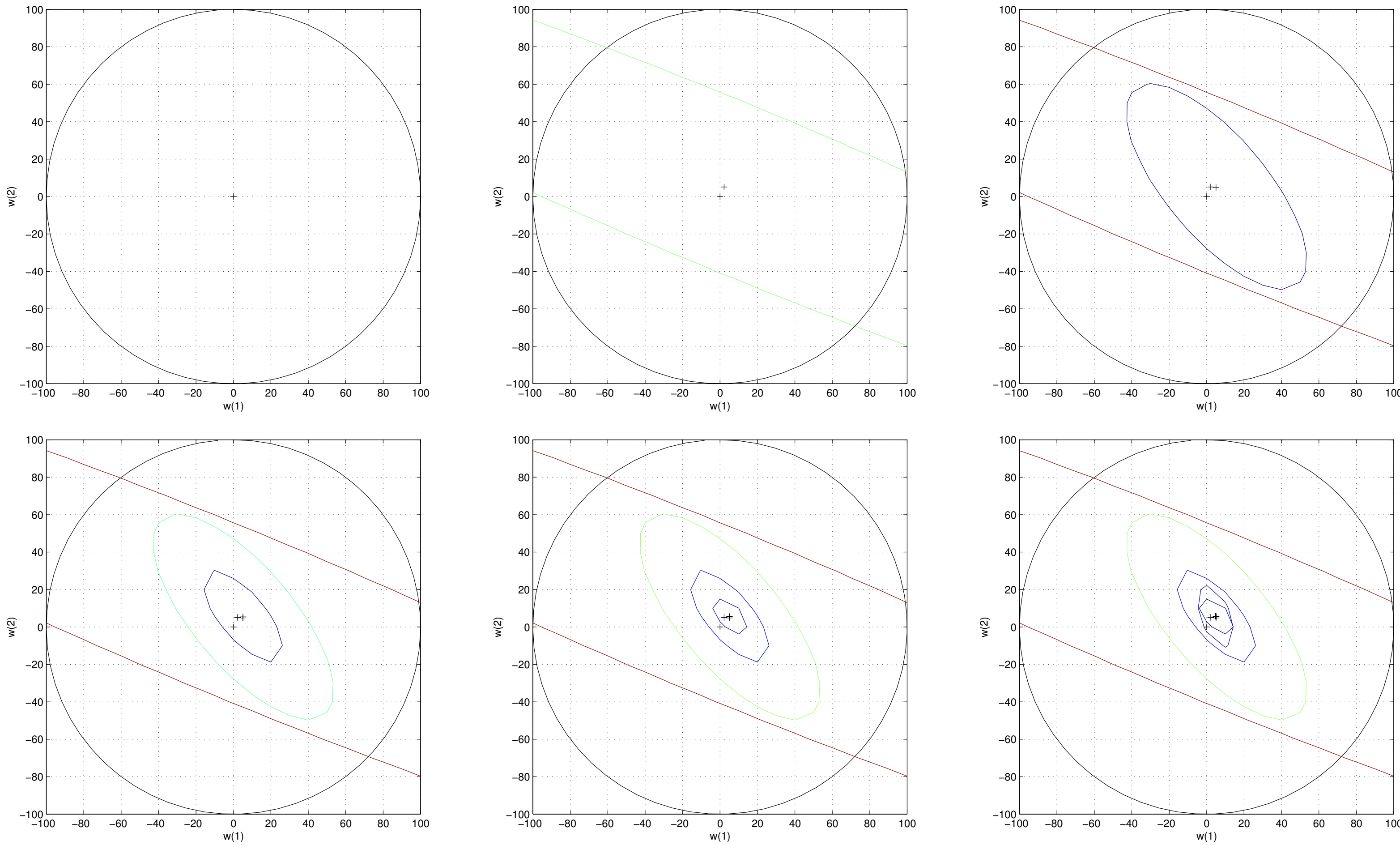
$$s_n := \beta \frac{R}{\sqrt{2}} \|\nabla \mathcal{F}_n(\mathbf{w}^a_{n-1})\|$$

   $R/\sqrt{2}$: a normalization constant required to show asymptotic convergence

   β: also required to show convergence

3. The definition of τ_n follows simply from $\tau_n = \mathcal{F}_n(\mathbf{w}^a_{n-1}) + s_n$

An Example



Asymptotic Convergence Analysis

Condition 1. (*Bounded Autocorrelation matrix*) *There exist* $n_0 > 0, \sigma_1 > 0, \sigma_2 > 0$ *such that*

$$\sigma_1 \mathbf{I} \leq \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \leq \sigma_2 \mathbf{I}, \qquad \forall n \geq n_0.$$

Condition 2. (*Bounded Outputs*) *There exists a fixed* ρ_y *such that for all* $n > n_0$ *there holds*

$$\frac{1}{n} \sum_{i=1}^{n} y_i^2 \leq \rho_y.$$

The left inequality in Condition 1 is known as *weak persistent excitation* condition.

Theorem 1. *Let the sequence of estimates* $\{\mathbf{w}_n, n = 1, 2, 3, \ldots\}$ *be generated by the IPLS algorithm. Then*

$$\|\nabla \mathcal{F}_n(\mathbf{w}_n)\| = 2\|\mathbf{R}_{xx}(n)\mathbf{w}_n - \mathbf{p}_{xy}(n)\| = O\left(1/n\right).$$

Transient Convergence Analysis

Condition 1. (*Bounded Autocorrelation matrix*) *There exist* $n_0 > 0, \sigma_1 > 0, \sigma_2 > 0$ *such that*

$$\sigma_1 \mathbf{I} \leq \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \leq \sigma_2 \mathbf{I}, \qquad \forall n \geq n_0.$$

Condition 2. (*No Noise*) *The system is free from measurement noise,* i.e.,

$$y_i = \mathbf{x}_i^T \mathbf{w}_*, \qquad i = 1, 2, \ldots .$$

We assume that the data has no statistical fluctuations. Convergence then implies the phasing out of effects of initialization and thus is dictated entirely by the transient behaviour of the algorithm.

Theorem 2. *Let the sequence of estimates* $\{\mathbf{w}_n, n \geq M\}$ *be generated by the IPLS algorithm. If the observations are free of noise, then*

$$\|\mathbf{w}_n - \mathbf{w}_*\| = O(R^{-1}) \|\mathbf{w}_{n-1} - \mathbf{w}_*\|.$$

Transient Analysis – a simple example

Example. $y_i = w_* x_i + v_i, \quad$ with $x_i = w_* = 1, \qquad i = 1, 2, \ldots$

- **RLS** Assuming no statistical averaging (*e.g.*, no noise),

$$\mathbf{R}_{xx}(n) = E(\mathbf{x}\mathbf{x}^T) = 1, \quad \mathbf{p}_{xy}(n) = E(\mathbf{x}y) = 1, \quad \forall n$$

  The RLS estimator then reduces to $\mathbf{w}_n^{rls} = (1 + \delta/n)^{-1}$.

- **IPLS** Now, $\mathcal{F}_n(\mathbf{w}) = (\mathbf{w} - 1)^2$, and $\nabla \mathcal{F}_n(\mathbf{w}) = 2(\mathbf{w} - 1)$. Evaluating τ_n, the condition $\nabla \phi_n(\mathbf{w}_n^a) = 0$ for the analytic center becomes,

$$\frac{2(w_n^a - 1)}{\beta R\sqrt{2}|w_{n-1}^a - 1| + (w_{n-1}^a - 1)^2 - (w_n^a - 1)^2} + \frac{2w_n^a}{R^2 - (w_n^a)^2} = 0$$

  which implies

$$|w_n^a - 1| = O(R^{-1})|w_{n-1}^a - 1|,$$

  i.e., exponential decay of the transient error.

Direct Comparison of RLS, IPLS

Property	RLS	IPLS
Asymptotic Convergence	$O(1/n)$	$O(1/n)$
Computational Complexity	$O(M^2)$	$O(M^{2.2})$
Transient Convergence	$O(1/n)$	$O(R^{-n})$
Robustness to Initialization	no	yes
Additional constraints	req. new algorithm	easily accommodated
Numerical Stability (in constrained case)	problems occur when λ is small (λ : forgetting factor)	stable even at small values of λ, and in cases of limited precision calculations
Sliding window implementation	can be accommodated	easily accommodated

The Interior Point Least Squares (IPLS) algorithm

1. We don't need the *exact* analytic center of Ω_n, an approximate center is sufficient.

2. Such an approximate center is found by taking just a *single* Newton iteration in the minimization of $\phi_n(\mathbf{w})$.

$$\mathbf{w}_n := \mathbf{w}_{n-1} - (\nabla^2 \phi_n(\mathbf{w}_{n-1}))^{-1} \nabla \phi_n(\mathbf{w}_{n-1}), \tag{3}$$

   To compute (3) we need

$$\nabla \phi_n(\mathbf{w}) = \frac{\nabla \mathcal{F}_n}{s_n(\mathbf{w})} + \frac{2\mathbf{w}}{t_n(\mathbf{w})}, \tag{4}$$

$$\nabla^2 \phi_n(\mathbf{w}) = \frac{(\nabla \mathcal{F}_n)^T \nabla \mathcal{F}_n}{s_n^2(\mathbf{w})} + \frac{\nabla^2 \mathcal{F}_n}{s_n(\mathbf{w})} + \frac{4\mathbf{w}\mathbf{w}^T}{t_n^2(\mathbf{w})} + \frac{2\mathbf{I}}{t_n(\mathbf{w})} \tag{5}$$

   where $\nabla \mathcal{F}_n = -2\mathbf{p}_{xy}(n) + 2\mathbf{R}_{xx}(n)\mathbf{w}$ and $\nabla^2 \mathcal{F}_n = 2\mathbf{R}_{xx}(n)$.

3. To compute the Newton direction, an $O(M^{2.2})$ recursive update procedure has been devised (using the work of Powell, 1997).

Interior Point Least Squares (IPLS) Algorithm

Step 1: Initialization. Let β, R be given. Set $\mathbf{w}_0 = \mathbf{0}$, $\mathbf{p}_{xy}(0) = \mathbf{0}$, $\mathbf{R}_{xx}(0) = \mathbf{0}$, $\nabla \mathcal{F}_0(\mathbf{0}) = \mathbf{0}$.

Step 2: Updating. For $n \geq 1$, acquire new data $\mathbf{x}_n, y_n$. Then recursively update

$$\mathbf{p}_{xy}(n) = \frac{n-1}{n}\mathbf{p}_{xy}(n-1) + \frac{1}{n}\mathbf{x}_n y_n, \quad \mathbf{R}_{xx}(n) = \frac{n-1}{n}\mathbf{R}_{xx}(n-1) + \frac{1}{n}\mathbf{x}_n \mathbf{x}_n^T.$$

Update

- $\nabla^2 \mathcal{F}_n(\mathbf{w}_{n-1})$ and $\nabla \mathcal{F}_n(\mathbf{w}_{n-1})$
- $s_n(\mathbf{w}_{n-1})$ and $t_n(\mathbf{w}_{n-1})$
- $(\nabla^2 \phi_n(\mathbf{w}_{n-1}))^{-1} \nabla \phi_n(\mathbf{w}_{n-1})$ using the update procedure (or using (4), (5))

Step 3: Recentering. The new center of Ω_n is obtained by taking just one Newton iteration starting at $\mathbf{w}_{n-1}$:

$$\mathbf{w}_n := \mathbf{w}_{n-1} - (\nabla^2 \phi_n(\mathbf{w}_{n-1}))^{-1} \nabla \phi_n(\mathbf{w}_{n-1})$$

Set $n := n + 1$, and return to Step 2.

Summary

Contributions

- provided a new look at (recursive) adaptive filtering
- first application of interior point optimization to a dynamic problem

Features of IPLS

- converges asymptotically at the rate $O(1/n)$
- exhibits **fast transient** convergence, and is **robust to initialization**
- easily accommodates additional **linear or convex quadratic constraints**, and is **numerically stable**
- $O(M^{2.2})$ complexity

Application: System Identification

- **Performance Measure** $\varepsilon_{ip}(n) = \|\mathbf{w}_n - \mathbf{w}_*\|^2$ and $\varepsilon_{rls}(n) = \|\mathbf{w}_n^{rls} - \mathbf{w}_*\|^2$
- **Sources** (i) White Gaussian noise, (ii) White Gaussian noise filtered through

$$H(z) = \frac{1 + 2z^{-1} + 3z^{-2}}{(1 - 1.1314z^{-1} + 0.64z^{-2})(1 + 0.9z^{-1})}.$$

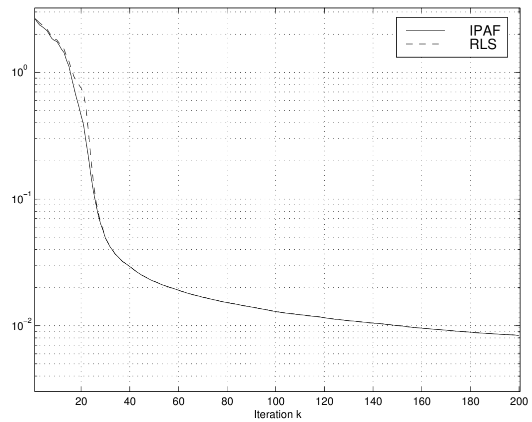
- **SNRs** (i) SNR$_1$ = 40dB, (ii) SNR$_2$ = 10dB
- **Nominal Parameter Settings**

RLS	$\lambda = 1, \quad \delta = 10^{-4}$
IPLS	$\beta = 2, \quad R = 1000$

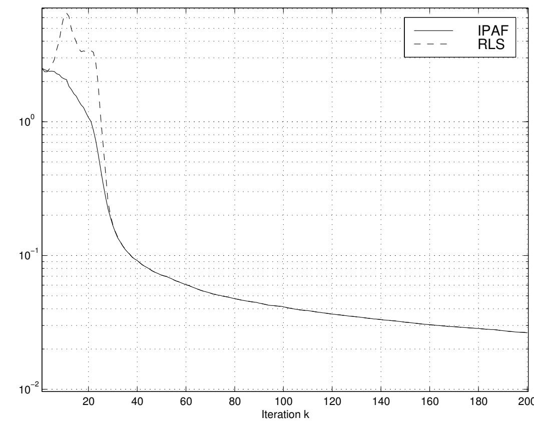
- **Experiment 1** $\mathbf{w} \in \mathbb{R}^{20}$, $\mathbf{w}(i) \in [-1, +1]$, 500 independent Monte Carlo trials
- **Experiment 2** Comparing sliding window versions of RLS (Liu & He '95) and IPLS:

$$\mathbf{w} \in \mathbb{R}^{10}, \quad T_l = 15$$

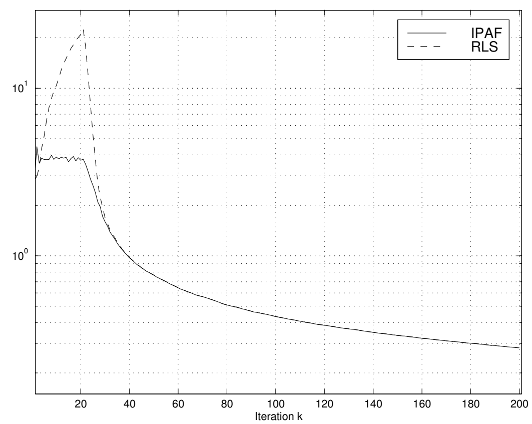
System Identification: Experiment 1



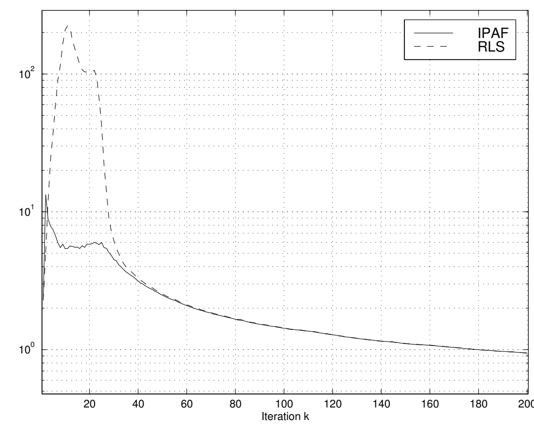
(a) SNR = 40dB, Source: White Gaussian Noise



(b) SNR = 40dB, Source: Correlated



(c) SNR = 10dB, Source: White Gaussian Noise



(d) SNR = 10dB, Source: Correlated

System Identification: Experiment 2

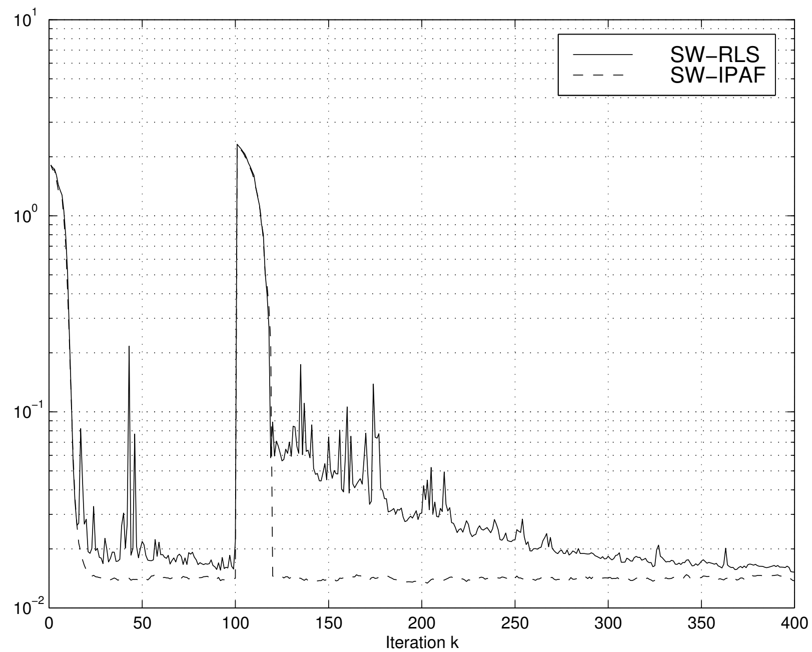
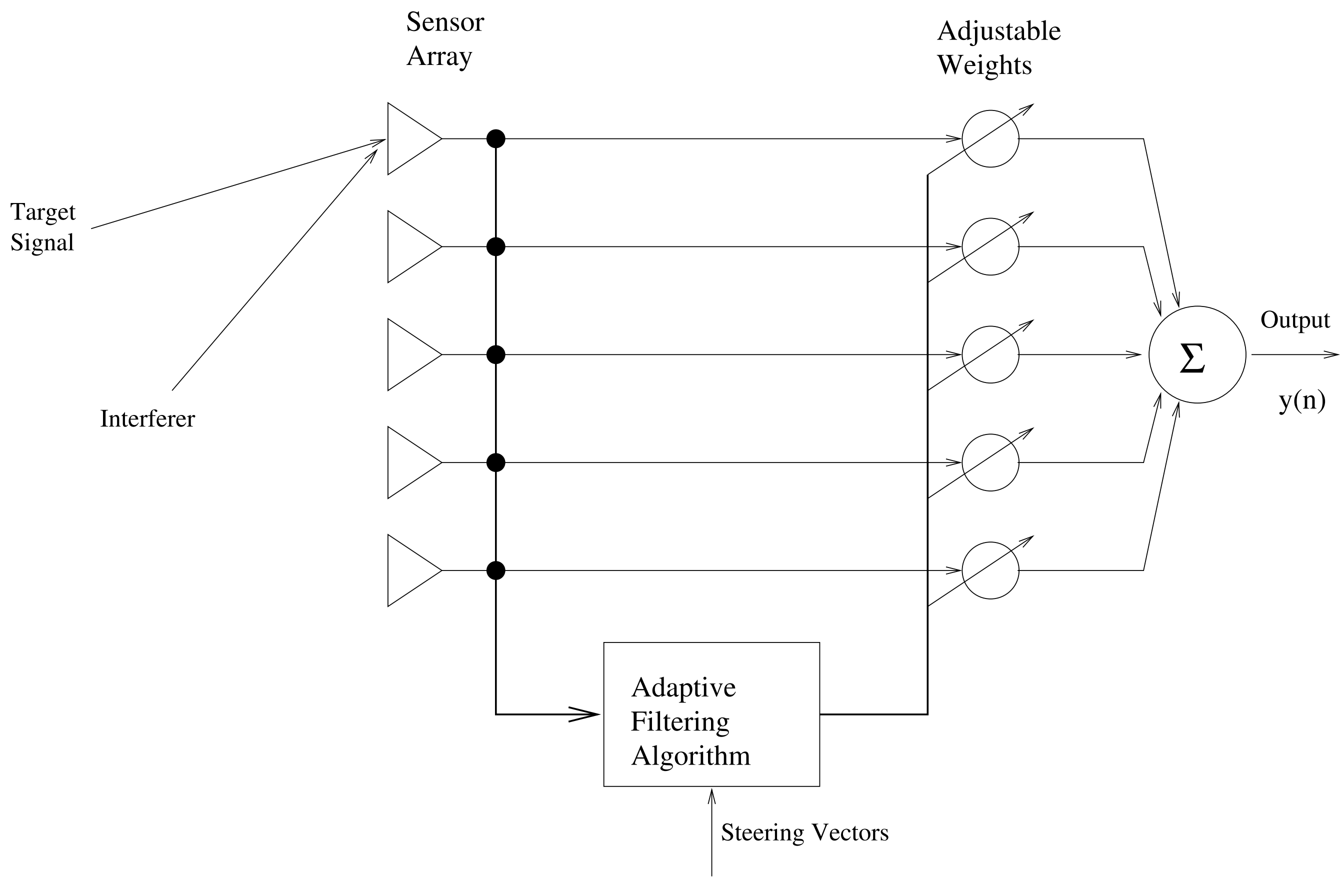


Figure 1: Comparison of sliding-window versions of IPLS and RLS when channel characteristics change abruptly (at iteration 100).

Application: Minimum Variance Beamforming



Minimum Variance Beamforming

By adaptively adjusting the tap weights $h_i(n)$ the beamformer must

1. Steering Capability: protect the target signal

$$\mathbf{c}^H(\theta)\mathbf{h}(n) = 1, \qquad \forall n,\ \theta = \theta_1, \theta_2, \ldots$$
$$\mathbf{c}^H(\theta) = [1, e^{-j\theta}, \ldots, e^{-j(M-1)\theta}],$$

   where
   M : number of tap weights h_i,
   θ_i : *Electrical Angle* determined by the direction of the target i with respect to the first sensor

2. Minimize the effects of the interferers
   i.e., minimize the Output Power $E(|y|^2)$ of the beamformer

This beamforming problem can be cast in the framework of a **Constrained Adaptive Estimation Problem**.

Beamforming: Constrained Adaptive Estimation

$$\begin{aligned} \text{minimize} \quad & \mathcal{F}_n := \frac{1}{n}\sum_{i=1}^{n} \lambda^{n-i}|d(i) - \mathbf{x}_i^T\mathbf{h}|^2, \\ \text{subject to} \quad & \mathbf{C}^T\mathbf{h} = \mathbf{f},\ \mathbf{h} \in \mathbb{R}^M, \end{aligned} \tag{6}$$

λ:	forgetting factor
$d(\cdot)$:	desired response
$\mathbf{x}_i^T$:	vector input sequence
$\mathbf{h}$:	vector of tap weights
$\mathbf{C}, \mathbf{f}$:	define the linear constraints on $\mathbf{h}$

- In the Minimum Variance Beamforming problem the reference signal $d(\cdot)$ is zero. During the adaption process we assume that *no target* is present.
- The rows of $\mathbf{C}$ correspond to steering vector constraints.

Beamforming: Numerical Simulation

Input: (interference at 0.3, 0.325 and 0.7)

$$x(n) = \sin(0.3n\pi) + \sin(0.325n\pi) + \sin(0.7n\pi) + b(n).$$

$b(n)$: white Gaussian noise at $40dB$.

Constraints: (desired response at freq. 0.2 and 0.5)

$$\mathbf{C}^T = \begin{bmatrix} 1 & \cos(0.2\pi) & \ldots & \cos((M-1)0.2\pi) \\ 1 & \cos(0.5\pi) & \ldots & \cos((M-1)0.5\pi) \\ 0 & \sin(0.2\pi) & \ldots & \sin((M-1)0.2\pi) \\ 0 & \sin(0.5\pi) & \ldots & \sin((M-1)0.5\pi) \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Beamforming: Numerical Simulation

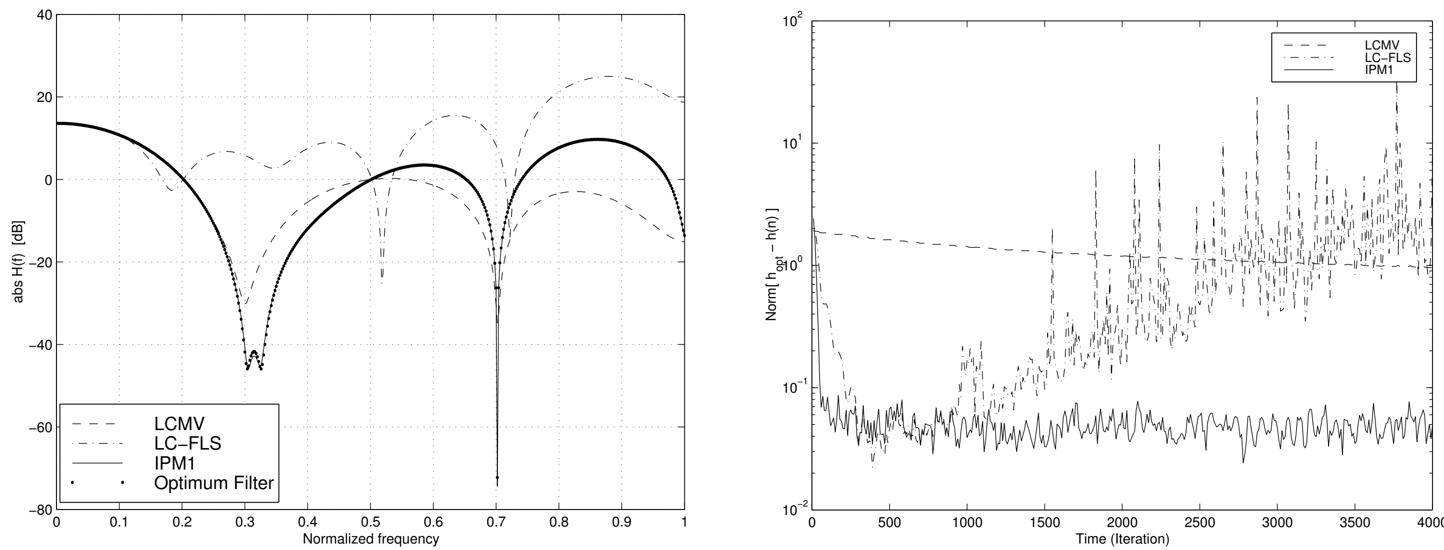


Figure 2: (a) Freq. response at Iteration 4000, (b) Mean-squared error in $\mathbf{h}(n)$

LCMV	$\mu = 0.1$
LCFLS	$\lambda = 0.99, E_o = 0.1$
IPLS	$\lambda = 0.99, \varepsilon = 0.01, R = 100, \beta = 2$
Precision	4 digits for LCFLS and IPLS

Application: Channel Equalization in a CDMA Downlink

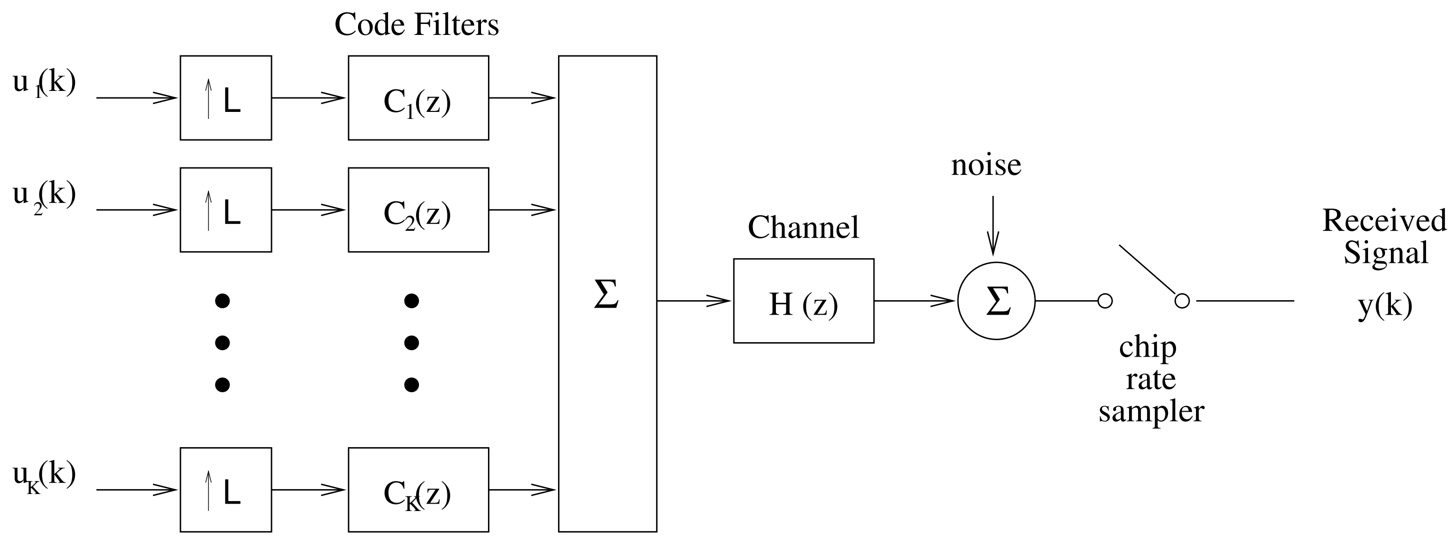


Figure 3: Discrete-time model of CDMA downlink

Equalizer/Decoder Structure

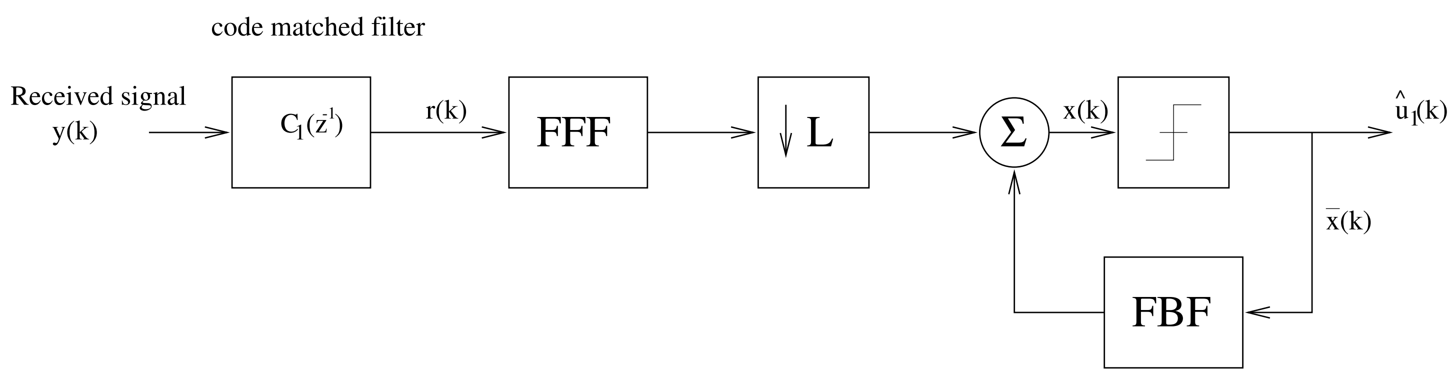
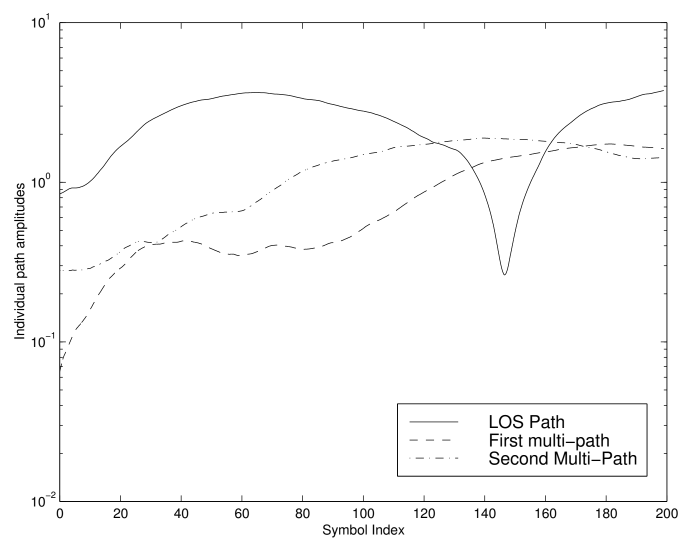


Figure 4: Code Matched Filter – Chip rate DFE

CDMA Downlink: System Description

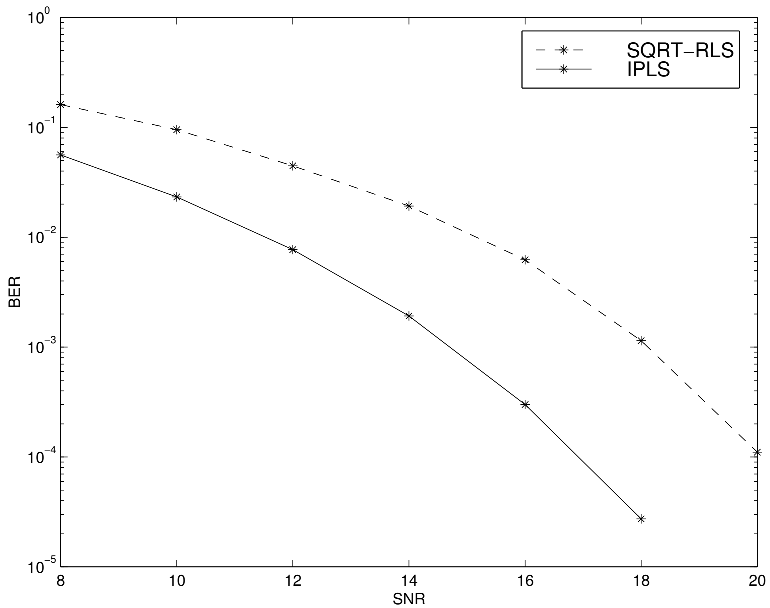
Sources QPSK with uniform probabilities for each symbol (i.i.d.)

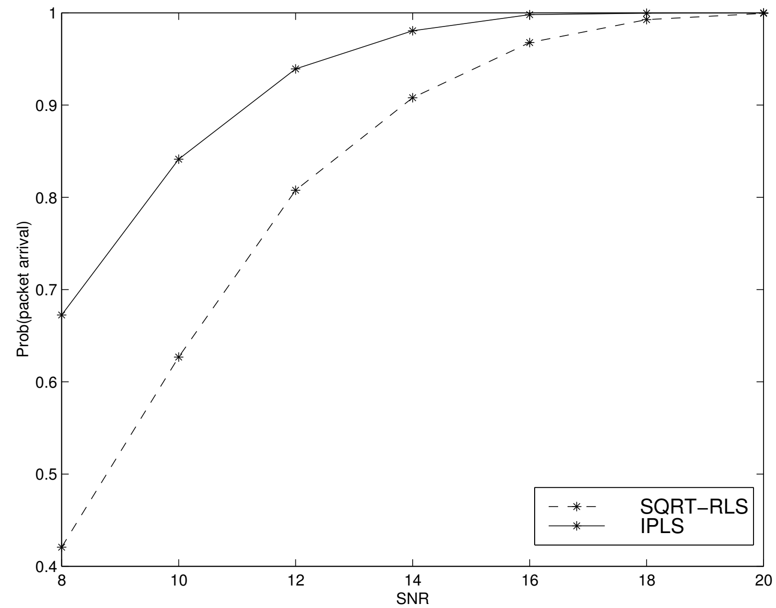
Fading Channel LOS component is 5 dB higher than 2 multipath components, fading rate $f_D = 0.005$, delay spread: $\leq 6T_c$



Static Channel Fading Channel sampled at a random instant.

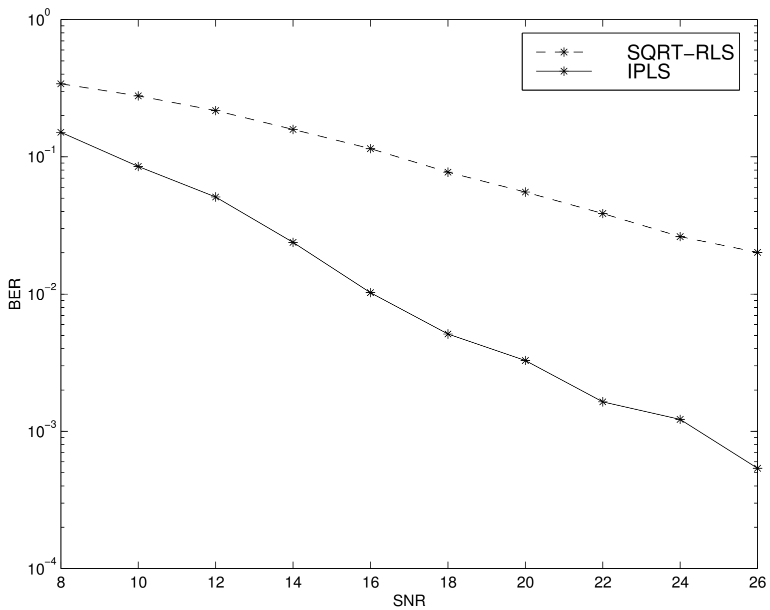
Experiment 1: Static Channel, Single user

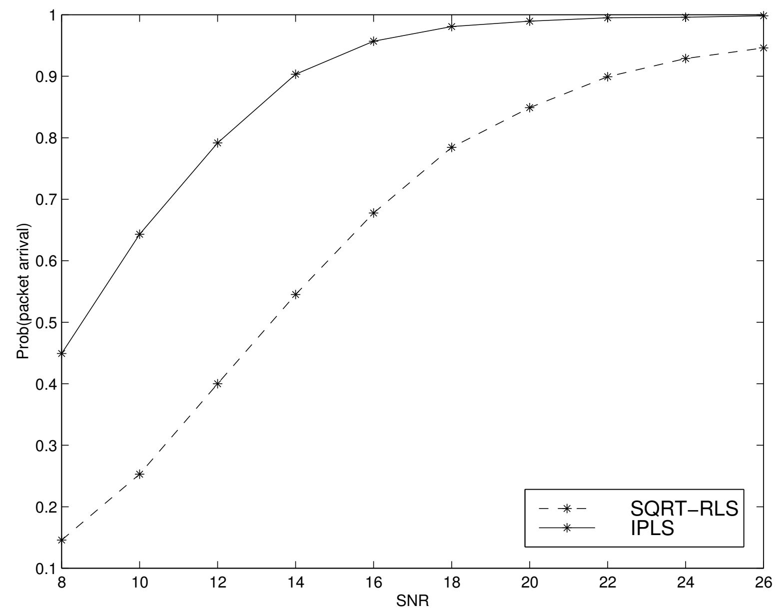




Message Signal	$N = 200, N_T = 10, C_L = 16, Users = 1$
Algorithms	$\lambda = 1.0, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2$, delay $= 1, pfr = 10^{-2}$

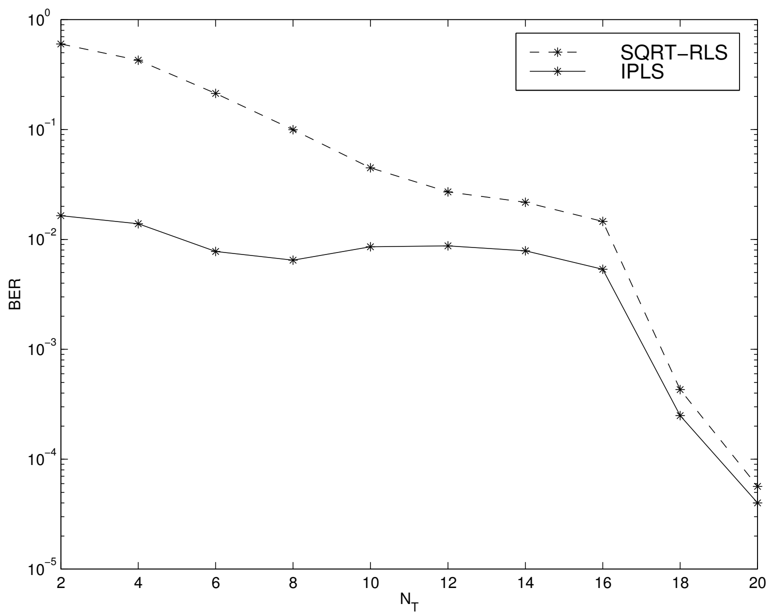
Experiment 1: Static Channel, 4 Users

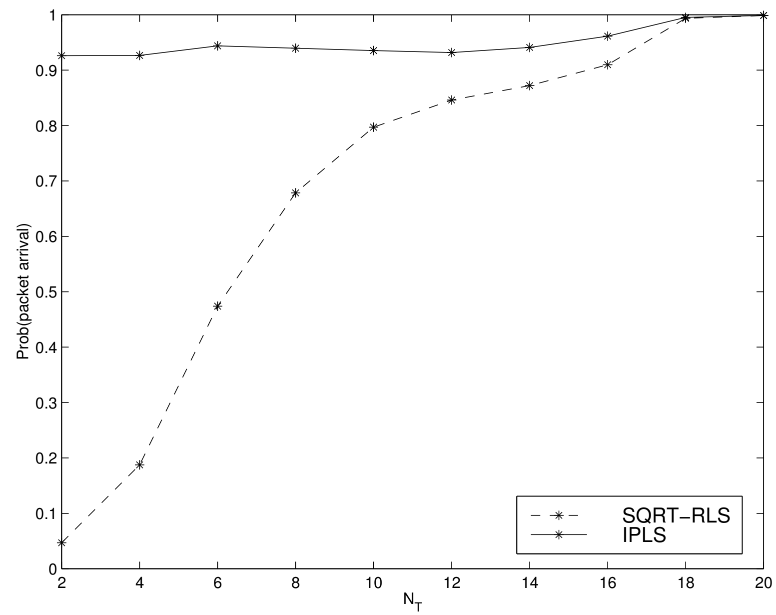




Message Signal	$N = 200, N_T = 10, C_L = 16, Users = 4$
Algorithms	$\lambda = 1.0, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2$, delay $= 1, pfr = 10^{-2}$

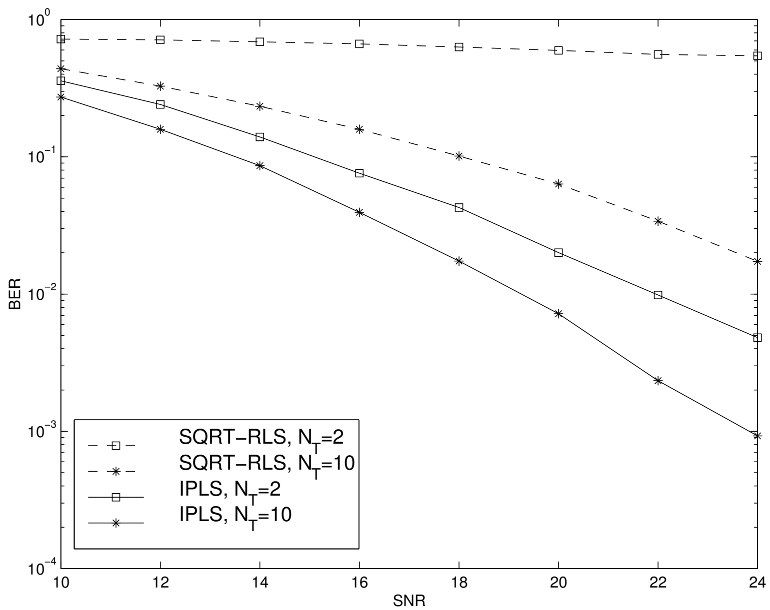
Experiment 1: Dependence on Training Length

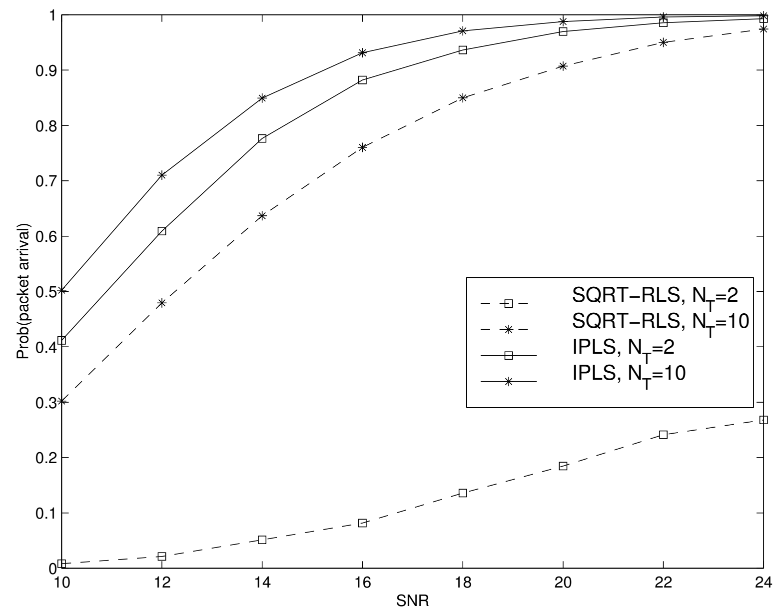




Message Signal	$N = 200, SNR = 12dB, C_L = 16, Users = 1$
Algorithms	$\lambda = 1.0, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2,$ delay $= 1, pfr = 10^{-2}$

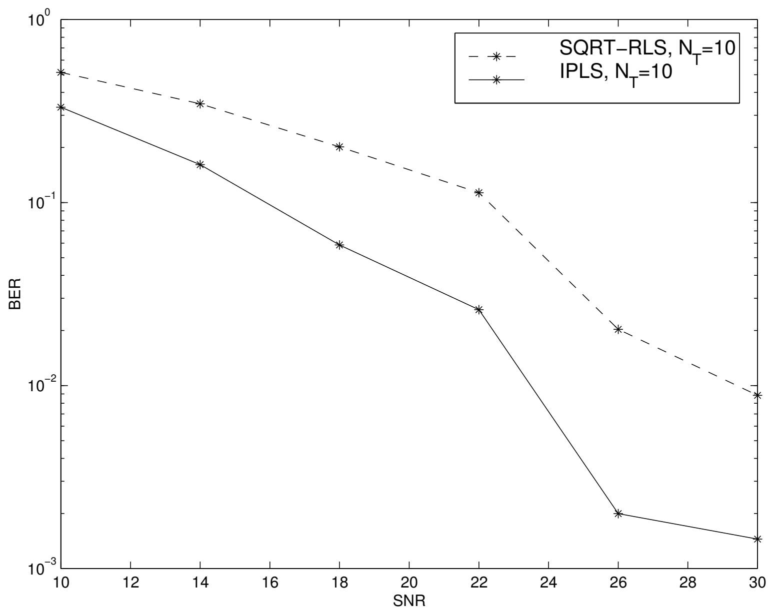
Experiment 2: Time-Varying Channel, Single user

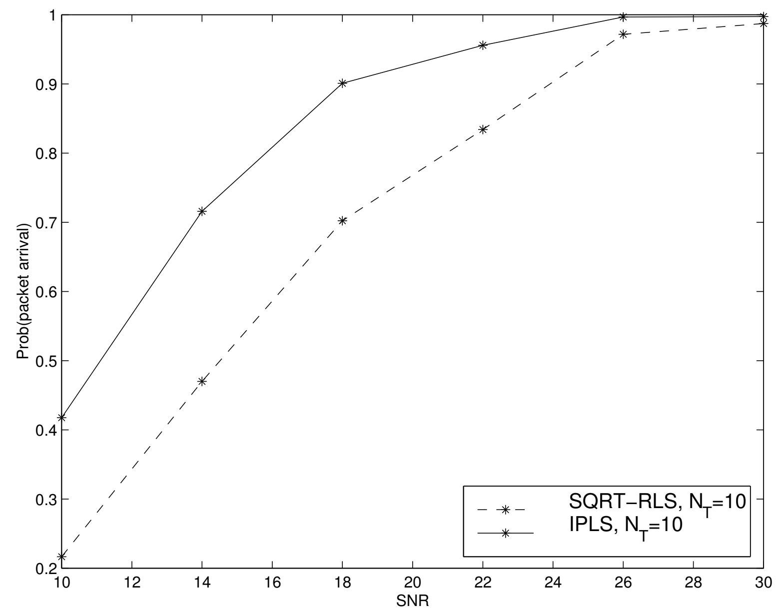




Message Signal	$N = 200, N_T = 2/10, C_L = 16, Users = 1$
Algorithms	$\lambda = 0.85, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2,$ delay $= 1, pfr = 10^{-2}$

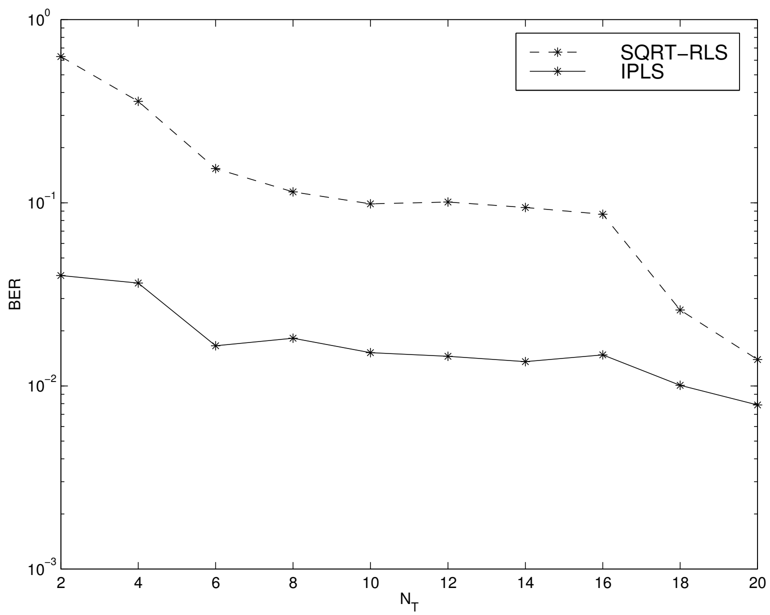
Experiment 2: Time-Varying Channel, 4 users

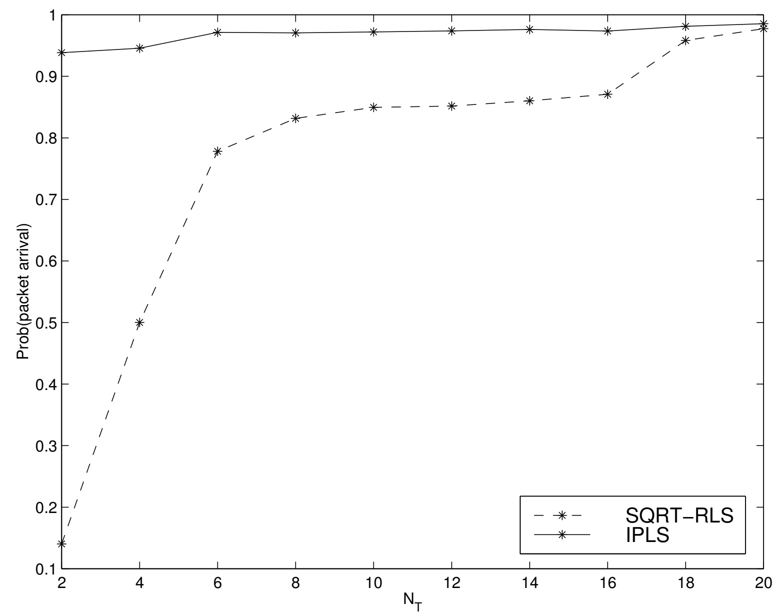




Message Signal	$N = 200, N_T = 2/10, C_L = 16, Users = 4$
Algorithms	$\lambda = 0.85, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2,$ delay $= 1, pfr = 10^{-2}$

Experiment 2: Dependence on Training Length





Message Signal	$N = 200, SNR = 16dB, C_L = 16, Users = 1$
Algorithms	$\lambda = 0.85, \delta = R = 10^4, \beta = 2$
Equalizer	$M_{ff} = 14, M_{fb} = 2,$ delay $= 1, pfr = 10^{-2}$

Conclusions

- Transient convergence of IPLS is $O(1/R^n), \quad n \geq M$
- Superior transient convergence to RLS even when $n < M$
- Gain of using IPLS over the RLS algorithm can range from 5-6 dB to well over 10 dB