

Twice Universal Linear Prediction of Individual Sequences

Andrew C. Singer
Advanced Systems Directorate
Sanders, Nashua, NH 03061 USA
e-mail: andrew.c.singer@lmco.com

Meir Feder
Dept. of EE-Systems,
Tel-Aviv University, Israel
e-mail: meir@eng.tau.ac.il

Abstract — We present a “twice universal” linear prediction algorithm over the unknown parameters and model orders, in which the sequentially accumulated square prediction error is as good as any linear predictor of order up to some M , for any individual sequence. The extra loss comprises of a parameter “redundancy” term proportional to $(p/2)n^{-1} \ln(n)$, and a model order “redundancy” term proportional to $n^{-1} \ln(p)$, where p is the model order we compare with, and n is the data length. The computational complexity of the algorithm is about the complexity of a recursive least squares (RLS) linear predictor of order M .

I. INTRODUCTION

Linear prediction is an important component of a wide variety of signal processing and communication systems. Given the data sequence $x_1^n = \{x[t]\}_{t=1}^n$ the optimal set of p linear prediction coefficients $\underline{a}[n] = [a_1[n], \dots, a_p[n]]^T$ minimizes the total squared prediction error,

$$\underline{a}[n] = \arg \min_{\underline{a}} l(x_1^n, \hat{x}_{\underline{a},1}^n) = \arg \min_{\underline{a}} \sum_{t=1}^n (x[t] - \hat{x}_{\underline{a}}[t])^2$$

where $\hat{x}_{\underline{a}}[t] = \sum_{j=1}^p a_j x[t-j]$ is the predicted value of $\hat{x}[t]$. The resulting coefficients depend, of course, on the input sequence. In this work we provide a “twice universal” computationally efficient algorithm, that is not tuned to the data in advance and does not assume a model order, yet for any bounded sequence x_1^n asymptotically attains the minimal average square error of any linear predictor up to some order M , at a rate achievable given the model order.

II. UNIVERSAL FIXED-ORDER LINEAR PREDICTION

We assume that the sequence to be predicted x_1^n is a real-valued arbitrary sequence, such that $|x[t]| < A$ for all t . Let $\hat{x}_{\underline{a}}[t] = \underline{a}^T \underline{x}[n]$, where $\underline{x}[n] = [x[t-1], \dots, x[t-p]]$, be the output of a p -th-order linear predictor with parameter vector \underline{a} . The p -th order universal predictor generates predictions $\hat{x}_{u,p}[n] = \underline{a}_u[n-1]^T \underline{x}[n]$ where,

$$\underline{a}_u[n] = \left[R_{xx}^n + \frac{4\lambda_{\infty}^2}{A^2} I \right]^{-1} r_x^n,$$

$r_x^n = \sum_{k=1}^n x[k] \underline{x}[k]$, $R_{xx}^n = \sum_{k=1}^n \underline{x}[k] \underline{x}[k]^T$ and $\frac{1}{n} R_{xx}^n$ has a unique minimum eigenvalue $\lambda \geq \lambda_{\infty} > 0$. Its performance is given in the following Theorem:

Theorem 1. $l(x_1^n, \hat{x}_{u,p,1}^n) = \sum_{t=1}^n (x[t] - \hat{x}_u[t])^2$, satisfies

$$\begin{aligned} \frac{1}{n} l(x_1^n, \hat{x}_{u,p,1}^n) &\leq \min_{\underline{a}} \frac{1}{n} l(x_1^n, \hat{x}_{\underline{a},1}^n) \\ &+ \frac{4A^2 p}{n} \ln \left(\frac{A^4(p+1)n}{8\lambda^2} + 1 \right) + O(n^{-2}). \end{aligned}$$

The idea behind the universal predictor and the proof of the Theorem is as follows. We define a “probability” assignment of each of the continuum of predictors to the data sequence x_1^n such that the probability will be an exponentially decreasing function of the total squared-error for that predictor. By defining a universal probability as an a priori average of the assigned probabilities, then to first order in the exponent, the universal probability will be dominated by the largest exponential, i.e., the probability assignment of the model with the smallest total squared error. Specifically, by using a conjugate prior, which in this case is Gaussian, the mixture over the parameters is obtained in closed form, and we calculate explicitly the “redundancy” with respect to best model. We then relate the universal probability assignment to the accumulated squared error of the universal predictor and get the desired result.

III. UNKNOWN MODEL ORDER

We now have universal predictors for each model order $i = 1, \dots, M$. Each of these predictors generate predictions $\hat{x}_i[n]$ at each time t . The suggested twice-universal predictor is $\hat{x}_{tu}[t] = \sum_{i=1}^M \mu_i[t] \hat{x}_i[t]$, where

$$\mu_i[t] = \frac{\exp(-\frac{1}{2c} l(x_1^{t-1}, \hat{x}_{i,1}^{t-1}))}{\sum_{k=1}^M \exp(-\frac{1}{2c} l(x_1^{t-1}, \hat{x}_{k,1}^{t-1}))}.$$

In other words, this predictor at each time point is a weighted linear combination of all the predictors of various orders where the weights depend exponentially on the performance of each predictor so far. For this predictor we can state the following:

Theorem 2. The total squared prediction error of the twice-universal predictor, $l(x_1^n, \hat{x}_{tu,1}^n) = \sum_{t=1}^n (x[t] - \hat{x}_{tu}[t])^2$, satisfies

$$\frac{1}{n} l(x_1^n, \hat{x}_{tu,1}^n) \leq \min_p \frac{1}{n} l(x_1^n, \hat{x}_{u,p,1}^n) + \frac{8A^2}{n} \ln(p) + O(n^{-1}).$$

IV. COMPUTATIONAL COMPLEXITY

The universal predictor requires a mixture over the parameters and over the model orders. As discussed above, the mixture over the parameters can be accomplished through a properly initialized RLS algorithm. Thus, it only remains to solve for each of the RLS predictors of orders $i = 1, \dots, M$. This can be still complex, however the RLS algorithm for a given model order can be written as a time-recursion and order-recursive algorithm. Specifically, for the M -th order problem, at each time point the predictor is constructed by recursively solving for each of the predictors of lower order. This is the lattice filter formulation of the RLS algorithm. The resulting complexity of these algorithms can be made to have $O(M)$ operations per time sample which results in a total complexity of $O(Mn)$ for the universal predictor.