Time varying parameters

Let us first consider, for simplicity of presentation, the problem of estimation of constant value θ^*, observed in the presence of additive, zero-mean random noise z_k with finite variance σ_z^2

$$y_k = \theta^* + z_k,$$

i.e., we recover θ^* from the observations $\{y_k\}_{k=1}^N$. In general, the measurements $\{y_k\}_{k=1}^N$ can be taken into account with different levels of significancy, i.e., we want to find the weighted least squares minimum

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} \alpha_k \left(y_k - \theta\right)^2, \tag{1}$$

where α_k's are some weights representing the priorities of respective measurements y_k's. Since

$$\frac{\partial}{\partial\theta} \sum_{k=1}^{N} \alpha_k \left(y_k - \theta\right)^2 = 2\sum_{k=1}^{N} \left(\theta\alpha_k - y_k\alpha_k\right),$$

the minimum in (1) is the solution of the equation

$$\theta \sum_{k=1}^{N} \alpha_k = \sum_{k=1}^{N} y_k\alpha_k$$

which leads to

$$\widehat{\theta} = \frac{\sum_{k=1}^{N} y_k\alpha_k}{\sum_{k=1}^{N} \alpha_k}, \tag{2}$$

where $\sum_{k=1}^{N} \alpha_k$ can be understand as effective number of measurements. In particular, if α_k's are the same for each k, i.e. $\alpha_k = \alpha = \text{const}$, then we obtain the standard least squares solution

$$\widehat{\theta} = \frac{\alpha\sum_{k=1}^{N} y_k}{N\alpha} = \frac{1}{N}\sum_{k=1}^{N} y_k.$$

Non-uniform α_k's are commonly applied due to two main reasons: (i) time-varying estimated value $\theta^*(N) = \theta^*_N$, and (ii) error-in variable problem in local regression estimation.
(i) If the estimated value θ^*_N is changing, the popular choice is to take

$$\alpha_k = \lambda^{N-k}, \text{ with } 0 < \lambda < 1 \tag{3}$$

which damps old measurements and increase the influence of recent observations on the resulting estimate. Moreover, such a choice is very convenient for designing the recursive versions of identification algorithms, since

$$\alpha_{k+1} = \frac{1}{\lambda}\alpha_k.$$

On the other hand, since the effective number of measurements is finite, i.e. $\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} \lambda^{N-k} < \infty$, the variance of the estimate (2) does not tend to zero. It is a cost paid for good tracking abilities of the algorithm. The most popular models of θ^*_N variations are random walk, random walk with trends, jump changes, Markov chains and knwoledge-based descriptions (for details see [**?**] and the references cited therein).
Also the restrictive (hard) selection of the fixed number n of the last measurements can be applied, i.e.,

$$\alpha_k = \begin{cases} 1, \text{ as } N - n < k \leq N \\ 0, \text{ as } k \leq N - n \end{cases}.$$

The estimate $\widehat{\theta}_n^{(N)}$, selecting n last observations from the N element data set, is then as follows

$$\widehat{\theta}_n^{(N)} = \frac{1}{n} \sum_{k=N-n+1}^{N} y_k$$

Let us assume that the true value of estimated parameter jumps, at the time instant N, from θ^* to $\theta^* + \Delta$, and the moments of the measurement noise z_k remain the same. Let us also introduce the following quality index of the tracking procedure

$$Q(n) = \sum_{i=N+1}^{N+H} \left(\operatorname{var} \widehat{\theta}_n^{(i)} + \operatorname{bias}^2 \widehat{\theta}_n^{(i)} \right),$$

where H denotes horizon of the tracking. We simply get

$$\begin{aligned} Q(n) &= \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{n^2} \sum_{j=1}^{n-1} j^2 = \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{n^2} \left(\frac{n(n+1)(2n+1)}{6} - n^2 \right) \\ &\simeq \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{3} n, \end{aligned}$$

where the cummulated variance component $\frac{H\sigma_z^2}{n}$ dominates for small n, while the cummulated bias error $\frac{\Delta^2}{3} n$ inreases for large n (see Fig. 1). The optimal value of n

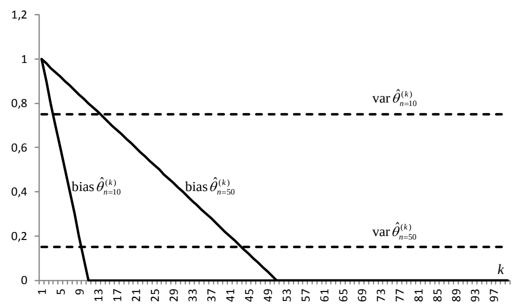

Figure 1: Sample behaviour of bias and variance after parameter jump for $n = 10$ and $n = 50$.

$$n_{opt.} = \arg\min_n Q(n)$$

obviously depends on the relation between horizon H, height of jump Δ, and the variance of the noise σ_z^2, i.e. $n_{opt.} = n_{opt.}(H, \Delta, \sigma_z^2)$. Nevertheless, this dependence is weak, in the sense that in typical cases (Δ

comparable with σ_z^2) n_{opt} usually lays between 10 and 20 (see Example in Fig. 2). Let us also emphasize that since Δ and σ_z^2 are unknown, computing $n_{opt.}$ is not possible. Also the asymptotic convergence of the estimate cannot be achieved, because of finite number of effective measurements.

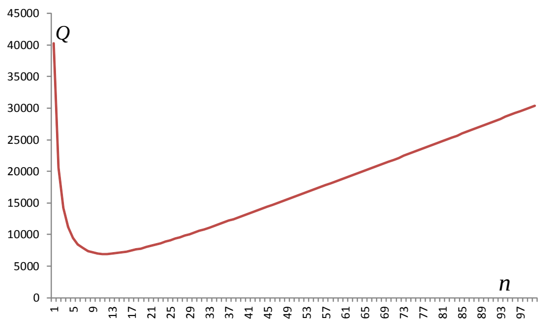

Figure 2: Dependence between tracking error Q and the window length n, for $H = 100$, $\sigma_z = 20$, and $\Delta = 30$.