



Polynomial least squares multiple-model estimation: simple, optimal, adaptive, practical

J. W. Bell¹ 

Received: 2 December 2019 / Accepted: 26 August 2020 / Published online: 9 November 2020
© Springer Nature Switzerland AG 2020

Abstract

This paper presents a polynomial least squares (LS) method of overcoming numerous weaknesses and deficiencies in the interacting multiple model (IMM), which comprises multiple Kalman filter models as the state-of-the-art algorithm for tracking maneuvering targets. The paper also addresses several polynomial LS misunderstandings and flaws applied in econometrics. These aims are achieved by first uniquely deriving a very simple version of conventional LS, which fits discrete deterministic data by minimizing the sum of squared deviations of data from an assumed polynomial (often called data-fitting). The contrasting contemporary polynomial LS method filters out corrupting statistical noise from already existing known polynomials, as in target tracking. Contemporary polynomial LS is developed from the conventional LS method derived by Gauss. Polynomial LS conventionally applied to data-fitting and contemporarily applied to noise filtering are shown to be different problems that are not to be conflated as done in econometrics, where data-fitting is described as regression using a 1st degree polynomial called ordinary LS. Most significantly, analogous to the IMM, the polynomial LS multiple model (LSMM) is derived. Contrary to the IMM, which is not optimized; the LSMM is optimized with a trade-off between the statistical variance and the deterministic bias-squared to minimize the mean-square-error (variance plus bias-squared). A sequence of optimal LSMMs matched to accelerations covering the spectrum of acceleration between zero and the assumed maximum are derived and applied in designing an adaptive algorithm. Results demonstrate improved target tracking and accuracy by the LSMM of maneuvering targets compared with published IMM performances.

Keywords Polynomial least squares · Multiple-model · IMM · Adaptivity · Target tracking · Kalman filter · Estimation

1 Introduction

Original least squares (LS) methods approximate an existing set of data with a mathematical function or a series by minimizing the sum of the squared deviations of the data from the function or series [1–5]. Often called data-fitting, conventional LS constructs coefficients to approximate deterministic data with an assumed polynomial [6–8].

However, in contrasting contemporary problems (e.g., target tracking), LS estimates already existing polynomial coefficients by filtering out corrupting statistical errors [8–10]. (More references in [11]). Contemporary LS is

associated with maximum likelihood estimation [12, 13], general method of moments [14–16], and Bayesian estimation [17].

Polynomial LS applies in tracking [9, 10], signal processing [18, 19], and statistical estimation [19, 20]. It is also applied in statistics and econometrics where a 1st degree polynomial is assumed, called ordinary least squares (OLS), and described as regression [21, 22].

Polynomial LS can be derived and applied using the orthogonality principle [6, 7, 23], matrix theory [1], and the basic approach of determining the solution to an over-determined linear system of equations [6, 7].

✉ J. W. Bell, jwbell@infowest.com | ¹Santa Clara, UT, USA



As the solution to an overdetermined linear system of equations, conventional polynomial LS is deterministic. It says nothing about statistics. In fact, statistics did not exist when Gauss invented and used LS on Ceres two and a quarter centuries ago [24]. It was not considered an organized mathematical science until Fisher established the foundation of statistics and estimation theory more than a century later—in the 1920s [25, 26].

Because of contemporary polynomial LS resource intensive matrix inversion during computer infancy, it gave way to recursive tracking in the form of 2nd order (constant velocity) $\alpha - \beta$ as well as 3rd order (constant acceleration) $\alpha - \beta - \gamma$ estimators [9, 27, 28].

As a state space estimator; the Kalman filter (KF) [11] makes improvements: (a) providing a running measure of prediction accuracy (variance) and (b) adding to the filtered error variance prior to prediction the variance of state noise, which prevents the filtered error variance from going to zero and causing the KF to become unstable and blow up [8, 9].

The advantage of 2nd order estimators is that the variance is smaller than from 3rd order estimators. The disadvantage is that a bias is created from target acceleration. The bias increases as more samples are processed, causing the tracking filter mean-square-error (MSE), comprising variance plus bias-squared, to rapidly exceed the 3rd order variance (MSE).

The disadvantage of a 3rd order estimator is that the variance remains the same no matter the size of the acceleration. Thus, in the absence of acceleration, the 3rd order variance remains larger than the 2nd order variance. Thus, a trade-off is needed.

To deal more effectively with acceleration jumps from maneuvering targets, various combinations of 2nd and 3rd order estimators described as multiple-models (MMs) have become available. 2nd and 3rd order KFs are selected and fused based on thresholds and computed probabilities [29–31]. The interacting multiple-model (IMM) adaptively and interactively makes adjustments between 2nd and 3rd order KFs based on likelihoods from KF residuals and ad hoc transition probabilities of assumed acceleration jumps [32–34]. The IMM is generally considered the best MM combination of performance and computational complexity.

Presumably as a holdover from KF use in control theory, steady-state (Oxymoronic, given continually maneuvering targets?) functions as the IMM default [35]. (The KF does not control targets. It filters out noise, which is actually a signal processing problem, not a control problem.) Thus, when an acceleration jump occurs as the IMM

approaches “steady-state”, a large bias-squared creates high MSE transition spikes at the maneuver beginning and end. Yet, the IMM does not address the bias, perform a variance/bias-squared trade-off, or minimize the MSE.

There have been no significant theoretical improvements to the IMM since its inception in the 1984 [32]. The thrust of IMM has been focused instead on applications [36–38]. This lack of recent IMM improvements opens the favorable opportunity for introducing the new LS multiple-model (LSMM) for overcoming adverse effects of the IMM addressed below and improving tracking and accuracy of maneuvering targets.

This paper addresses the problem by first formulating a weighting function from conventional polynomial LS and expanding it to form unique orthogonal polynomial coefficient estimators. This has the advantage of independently estimating each coefficient and its variance, which makes estimating the tracking variance easier. It offers a simple method of processing existing polynomials corrupted with noise to yield the empirical statistical estimate, expected value, and variance of any arbitrary point on the estimate of the existing polynomial.

Based on the orthogonal coefficient estimators; the LSMM is created by establishing linear interpolation between 2nd and 3rd order polynomial LS estimators, the 2nd order acceleration bias is defined, and interpolation is optimized (LSMM MSE minimized) in a variance/bias-squared trade-off. A sequence of optimal LSMMs matched to accelerations covering the spectrum of acceleration between zero and assumed maximum is derived and an adaptive algorithm is designed.

Optimized at the worst case acceleration (worst case is normally assumed in tracking, not left open-ended) as a function of sliding window size (sample number) and noise variance; the LSMM offers improved tracking accuracy, smooth RMSE transitions (no MSE transition spikes from acceleration jumps), and a stable RMSE.

After the two model IMM is introduced and shown to analogously linearly interpolate between 2nd and 3rd order KFs, the optimal LSMM is compared with published IMM performance and shown to yield improved accuracy over the IMM.

The paper is organized as follows: Sect. 2 derives the contemporary polynomial LS estimator from conventional polynomial LS data-fitting. Section 3 establishes orthogonal polynomial coefficient estimators and formulates the LSMM. The IMM is introduced and described in Sect. 4. In Sect. 5 the LSMM is applied to tracking examples in the literature, adaptivity is designed, and comparison is made with the IMM. Section 6 is a review and discussion. Section 7 is the conclusion.

2 LS estimate of polynomials corrupted by noise

2.1 Preliminaries and definitions: statistics, sample statistics, deterministic modeling

Deterministic modeling applies in the absence of statistics. For example, a set of data that may look random, but is not described statistically, is therefore deterministic. Conventional polynomial LS fits deterministic data, such as scatter grams or a mathematical function. It constructs coefficients in an assumed polynomial to fit the deterministic data.

Mechanical moments in deterministic polynomial LS data-fitting are commonly conflated with statistical moments in discussions and treatments of polynomial LS and regression (OLS), as in [20, 21, 39]. In reality, the average of deterministic data is neither the expected value (mean) nor the sample average. It is the mechanical 1st order moment, commonly called the centroid. The expected value (mean) is the statistical centroid analogous to the mechanical centroid. Likewise, the average of the squared differences between the data and the centroid is neither the variance nor the sample variance. It is the deterministic mechanical 2nd order central moment of the data. The statistical 2nd order central moment called the variance is analogous to the mechanical 2nd order central moment.

Stochastic processes and random variables are defined only in terms of statistics, not in sample statistics or the absence of statistics [23, 40]. Expected value (mean) and variance are defined only in terms of statistics, not sample statistics [23, 40]. Whereas, sample average and sample variance are defined in terms of sample statistics [40]. Thus, it is important to not confuse sample statistics with statistics as done in [20, 21, 39].

Consider the issue from the following perspective: adding, removing, or changing a single sample in conventional polynomial LS or sample statistics changes parameters in polynomial LS or sample statistics predictably and deterministically. Whereas; adding, removing, or changing samples in infinite statistics has no effect on statistical parameters.

Thus, it is crucial to understand and keep in mind differences among statistics, sample statistics, and deterministic modeling; and to not conflate them. Conventional polynomial LS is not defined in terms of statistics, it is deterministic. Nevertheless, if samples of an existing polynomial include statistical errors, polynomial LS may be viewed in terms of sample statistics [41].

For clarity, the term “estimation” from statistical estimation theory is used in this paper when statistically

described errors are assumed; whereas, the term “approximation” is used in the absence of statistical errors.

2.2 Conventional polynomial LS data-fitting

To apply conventional LS for constructing 1st degree polynomial coefficients that fit existing deterministic data $y_n \triangleq y(t_n)$ at points t_n , set $\frac{\partial S}{\partial \alpha} = \frac{\partial S}{\partial \beta} = 0$ and solve, where

$$S = \sum_{n=1}^N (y_n - \alpha - \beta t_n)^2 \tag{1}$$

This yields $\tilde{\alpha} = \frac{(\overline{y t^2} - \overline{y t} \bar{t})}{(\overline{t^2} - \bar{t}^2)}$ and $\tilde{\beta} = \frac{(\overline{y t} - \overline{y} \bar{t})}{(\overline{t^2} - \bar{t}^2)}$, where the tilde (\sim) denotes deterministic construction and the overline ($\bar{\cdot}$) denotes deterministic construction and the overline $\bar{v} \triangleq \frac{1}{N} \sum_{n=1}^N v_n$ is defined here as the arithmetic average.

Approximation $\tilde{\alpha}$ can be written as $y_n \frac{(\overline{t^2} - t_n \bar{t})}{(\overline{t^2} - \bar{t}^2)} = \overline{y_n a_n}$, where $a_n = \frac{(\overline{t^2} - t_n \bar{t})}{(\overline{t^2} - \bar{t}^2)}$ comprises weights for constructing $\tilde{\alpha}$.

Likewise, $\tilde{\beta}$ can be written as $y_n \frac{(t_n - \bar{t})}{(\overline{t^2} - \bar{t}^2)} = \overline{y_n b_n}$, where $b_n = \frac{(t_n - \bar{t})}{(\overline{t^2} - \bar{t}^2)} = (t_n - \bar{t}) / \sqrt{t_n(t_n - \bar{t})}$ comprises weights for constructing $\tilde{\beta}$. Multiplying b_n by τ and adding to a_n yields the deterministic weighting function

$$w_n(\tau) \equiv \frac{(\overline{t^2} - t_n \bar{t}) + (t_n - \bar{t}) \tau}{(\overline{t^2} - \bar{t}^2)} \tag{2}$$

for approximating y at any arbitrary point $t = \tau$ on the constructed polynomial:

$$\tilde{y}(\tau) = \overline{y_n w_n(\tau)} = \tilde{\alpha} + \tilde{\beta} \tau \tag{3}$$

In effect, Gauss used this concept on Ceres [8].

2.3 From conventional polynomial LS data-fitting to contemporary polynomial LS estimation

Now consider the contemporary and contrasting problem of estimating already existing 1st degree polynomial coefficients in $y(t) = \alpha + \beta t$ corrupted by the additive measurement stochastic process noise η :

$$z(t) = y(t) + \eta = \alpha + \beta t + \eta \tag{4}$$

samples η_n of which are zero mean, stationary, and white noise random variables (not necessarily Gaussian) with variance σ_η^2 .

Applying deterministic $w_n(\tau)$ to estimate existing constants α and β from samples of (4) yields

$$\hat{y}(\tau) = \overline{z_n w_n(\tau)} = \hat{\alpha} + \hat{\beta} \tau = \left[\alpha + \frac{(\overline{\eta t^2} - \overline{\eta t} \bar{t})}{(\bar{t}^2 - \bar{t}^2)} \right] + \left[\beta + \frac{(\overline{\eta t} - \overline{\eta} \bar{t})}{(\bar{t}^2 - \bar{t}^2)} \right] \tau \tag{5}$$

because $\overline{t_n a_n} = \overline{t_n \frac{(\bar{t}^2 - t_n \bar{t})}{(\bar{t}^2 - \bar{t}^2)}} = \overline{b_n} = \overline{\frac{(t_n - \bar{t})}{(\bar{t}^2 - \bar{t}^2)}} = 0$. The hat (\wedge) represents the statistical estimate from noise samples η_n and (5) filters out noise (a deterministic noise filter) to reduce the variance as a function of the sliding window of size N . For the boundary condition of $\eta = 0$, $\hat{y}(\tau) = y(\tau) = \alpha + \beta \tau$.

(Note: In effect, $w_n(\tau)$ establishes an average curve. When applied to deterministic data for constructing polynomial coefficients, it is described as data-fitting. When applied to noisy data for estimating already existing polynomial coefficients, it is described as noise filtering. The fact that $w_n(\tau)$ is the same in both cases seems to be a source of confusion and conflation.)

2.4 Difference between (3) and (5)

The fundamental difference between (3) and (5) is that (3) constructs polynomial coefficients to fit deterministic data from (2); whereas, (5) estimates already existing underlying polynomial coefficients by using (2) to filter out corrupting noise. These are two different problems that are not to be conflated.

Nevertheless, (3) is often conflated with (5), as demonstrated in [20, 21, 39]. Since samples y_n are not described statistically in (1), it is a deterministic problem. However, conflation occurs when the deterministic deviation $e_n \triangleq y_n - a - \beta t_n$ as a component of (1) is rewritten as $y_n = a + \beta t_n + e_n$, which mysteriously transforms deterministic e_n into a statistical error and then adds it to the non-existent deterministic $a + \beta t_n$. Of course, (1) does not include statistics; and since a and β do not exist, they are deterministically constructed by minimizing (1), which results in deterministic (2).

Another universal mistake is describing $\tilde{\beta} = \frac{(y - \bar{y})(t - \bar{t})}{(t - \bar{t})^2} = \frac{(\overline{yt} - \bar{y}\bar{t})}{(\bar{t}^2 - \bar{t}^2)}$ as the covariance of y and t divided by the variance of t , as in [20, 21]. In reality, $\frac{cov(y, t)}{var(t)} \triangleq$

$\frac{E[(y - E[y])(t - E[t])]}{E[(t - E[t])^2]} = \frac{E[yt] - E[y]E[t]}{E[t^2] - (E[t])^2}$, where $E[\cdot]$ is the statistical expected value operator. Note: *Same form \neq same function*. The independent variable t is not described statistically (often equally spaced samples). It is deterministic, although arbitrary samples may look random. Thus, \bar{t} is not

the expected value or even the sample average of t ; it is the deterministic mechanical 1st order moment commonly called the centroid. Likewise, $(\bar{t}^2 - \bar{t}^2)$ is the deterministic mechanical 2nd order central moment of t , not the statistical variance or even the sample variance. Thus, $var(t) = 0$. Moreover, $cov(y, t) = 0$, even if y is statistical as in [42].

Because η is statistical, (5) is the statistical estimate of existing coefficients α and β , where (2) is in essence a deterministic noise filter. Whereas, (3) does not filter out noise to estimate α and β ; it constructs them to minimize the sum of squared deviations of samples from the assumed polynomial.

However, the average of the squared deviations of the deterministic data from the constructed polynomial is minimized (1) divided by N . Thus, if the fictitious assumption is then made that the data were statistical deviations from an underlying polynomial that the constructed polynomial approximates, this average can be fictitiously assumed to be a sample average of the statistical noise variance and substituted into (9) below, which describes the filtered variance at any point on the estimated polynomial as is common in tracking.

Nevertheless, this includes two major fictitious assumptions: (a) The data are statistical deviations from an underlying polynomial and (b) the average of the squared deviations represents a sample variance. However, the crucial point is that this average is of squared deterministic deviations from the constructed polynomial, not a true sample average of squared statistical deviations from an actual underlying existing polynomial. Therefore, this constitutes reframing conventional deterministic LS in terms of contemporary LS estimating a fictitious underlying polynomial corrupted by fictitious statistical noise in order to estimate the fictitious noise variance. Although this mathematical model is fictitious, it may be useful if it produces effective and reasonably valid results. [43, 44]

References [45–47] address applying OLS in econometrics and polynomial LS. The remainder of this paper focuses on the main aim of deriving and using the LSMM to improve tracking and accuracy of maneuvering targets over that from the IMM.

2.5 Variance of the contemporary polynomial LS estimate

Because the statistical expectation operator $E[\cdot]$ and arithmetic average operator $\frac{1}{N} \sum_{n=1}^N (\cdot)$ are linear, their order is interchangeable [23]. Since noise samples η_n are zero mean, i.e., $E[\eta_n] = 0$; the expected value of estimate $\hat{y}(\tau)$ in

(5) at any τ is the 1st order statistical moment from statistical estimation theory [23]:

$$E[\hat{y}(\tau)] = E\left[\frac{1}{N} \sum_{n=1}^N z_n w_n(\tau)\right] = E\left[\frac{1}{N} \sum_{n=1}^N (\alpha + \beta t_n + \eta_n) w_n(\tau)\right]$$

$$= \alpha + \beta \tau + \frac{1}{N} \sum_{n=1}^N E[\eta_n] w_n(\tau) = \alpha + \beta \tau \tag{6}$$

which naturally matches the noise free case.

The variance of $\hat{y}(\tau)$ is the 2nd order statistical central moment from estimation theory:

$$\sigma_{\hat{y}}^2 = E\left[\{\hat{y}(\tau) - E[\hat{y}(\tau)]\}^2\right]$$

$$= \frac{1}{N} \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^N w_n(\tau) E[\eta_n \eta_i] w_i(\tau) \tag{7}$$

$$= \sigma_{\eta}^2 \frac{1}{N} \sum_{n=1}^N w_n^2(\tau)$$

where $\sum_{i=1}^N E[\eta_n \eta_i] w_i(\tau) = \sigma_{\eta}^2 w_n(\tau)$ from the white noise:

$$E[\eta_n \eta_i] = \begin{cases} \sigma_{\eta}^2 & \text{for } i = n \\ 0 & \text{for } i \neq n \end{cases} \tag{8}$$

Carrying out multiplications and summations in (7) yields the variance of the estimate at time τ :

$$\sigma_{\hat{y}}^2(\tau) = \sigma_{\eta}^2 \frac{\tau^2 - 2\bar{t}\tau + \bar{t}^2}{N(\bar{t}^2 - \bar{t}^2)} \tag{9}$$

where $\frac{\tau^2 - 2\bar{t}\tau + \bar{t}^2}{N(\bar{t}^2 - \bar{t}^2)}$ is also deterministic. Equivalent to the variance used in statistics and econometrics as described in [39, 48], (9) is more conducive to tracking. (Note that setting the derivative of (9) with respect to τ equal to zero and solving yields $\frac{\sigma_{\hat{y}}^2}{N}$, which as the minimum of (9) is also the variance of the estimate of a noisy zero degree polynomial—a constant.)

Missing from original LS invented and used by Gauss on Ceres two and a quarter centuries ago, (9) is the empirically determined statistical tracking variance of $\hat{y}(\tau)$ from additive noise samples η_n by deterministic estimator $w_n(\tau)$ (noise filter) and statistical estimation theory at estimation point τ on the estimated polynomial. Since statistics did not exist when Gauss invented LS [24], a statistical variance of the estimate was not available.

3 The LS multiple-model (LSMM)

3.1 Orthogonal polynomial coefficient estimators for equally spaced samples

Consider generalizing estimator $w_n(\tau)$ for equally spaced samples to $w_{mn}(\tau)$ where the new subscript m represents estimator order, which is 2 in (2). (Note: Estimator order from KF theory is degree of the polynomial estimated plus one.) For $m = 1$ the 1st order estimator $w_{1n}(\tau)$ is simply $\frac{1}{N}$. Estimating a 1st degree polynomial with $w_{1n}(\tau)$ from very simplified equally spaced samples yields $\sum_{n=1}^N (\alpha + \beta n) w_{1n}(\tau) = \alpha + \beta \frac{1}{N} \sum_{n=1}^N n = \alpha + \beta \left(\frac{N+1}{2}\right)$, where $\beta \left(\frac{N+1}{2}\right)$ is the bias. Furthermore, $b_n = (n - \bar{n}) / (n^2 - \bar{n}^2)$ yields $\sum_{n=1}^N (\alpha + \beta n) b_n \tau = \beta \tau$. Thus, $\alpha + \beta \tau = \sum_{n=1}^N (\alpha + \beta n) \left[w_{1n}(\tau) + b_n \left(\tau - \frac{N+1}{2}\right) \right] = \sum_{n=1}^N (\alpha + \beta n) w_{2n}(\tau)$, where $w_{2n}(\tau) \triangleq w_{1n}(\tau) + b_n \left(\tau - \frac{N+1}{2}\right)$ (10)

Continuing with induction, the m th order estimator can be written recursively for equally spaced samples in the unique form

$$w_{mn}(\tau) = w_{(m-1)n}(\tau) + U_{mn} \left[\tau^{m-1} - \sum_{n=1}^N n^{m-1} w_{(m-1)n}(\tau) \right] \tag{11}$$

where

$$\sum_{n=1}^N U_{mn} U_{kn} \begin{cases} \neq 0 & \text{for } m = k \\ = 0 & \text{for } m \neq k \end{cases} \tag{12}$$

starting with $w_{1n}(\tau) = U_{1n} = 1/N$ and $U_{2n} = b_n$. Orthogonality of U_{mn} s is exemplified by $\sum_{n=1}^N U_{1n} U_{2n} = 0$.

Note: each U_{mn} is fortuitously the estimator of the highest polynomial coefficient in the contemporary LS m th order estimator for equally spaced samples, analogous to $U_{2n} = b_n$. That is, U_{1n} from a 1st order estimator $\left[U_{1n} = \frac{1}{N} \right]$, U_{2n} from a 2nd order estimator $\left[U_{2n} = b_n = \frac{n-\bar{n}}{n^2-\bar{n}^2} = \frac{\text{numerator}}{n(\text{numerator})} \right]$, and U_{3n} from a 3rd order estimator $\left[U_{3n} = \frac{n^2(\bar{n}^2-\bar{n}^2)+n(\bar{n}^2\bar{n}-\bar{n}^3)+(\bar{n}^3\bar{n}-\bar{n}^2)}{n^4(\bar{n}^2-\bar{n}^2)+n^3(\bar{n}^2\bar{n}-\bar{n}^3)+n^2(\bar{n}^3\bar{n}-\bar{n}^2)} = \frac{\text{numerator}}{n^2(\text{numerator})} \right]$ ad infinitum are all orthogonal to each other over n . The U_{mn} s can also be defined more easily in terms of unique orthogonal polynomials specifically used in polynomial LS as in (15) below [6, 7, 49]. The $w_{mn}(\tau)$ in (11) is reminiscent of and equivalent to equally spaced Gram-Schmidt

orthogonalization and seems suggestive in (5). Orthogonalization could be used in the absence of equally spaced samples.

Advantage of orthogonality is that it allows each polynomial coefficient and its variance to be independently estimated, simplifying LSMM tracking variance estimation.

3.2 Forming the LSMM from orthogonal coefficient weights

Consider approximating deterministic $x(t)$ at time τ with a polynomial weighting function $w_{Mn}(\tau)$ of order M from equally spaced samples x_n :

$$\tilde{x}(\tau) = \sum_{n=1}^N x_n w_{Mn}(\tau) = \sum_{n=1}^N \sum_{m=1}^M x_n U_{mn} T_m(\tau) \tag{13}$$

where $T_m(\tau) \triangleq \left[\tau^{m-1} - \sum_{n=1}^N n^{m-1} w_{(m-1)n}(\tau) \right]$ for brevity from (11), which projects the m th coefficient estimate to time τ .

For 3rd order processing, $M = 3$ and $N \geq M$ yield

$$\begin{aligned} w_{3n}(\tau) &= U_{1n} + U_{2n} \left[\tau - \frac{N+1}{2} \right] \\ &\quad + U_{3n} \left[\tau^2 - (N+1)\tau + \frac{(N+1)(N+2)}{6} \right] \\ &= w_{2n}(\tau) + U_{3n} \left[\tau^2 - (N+1)\tau + \frac{(N+1)(N+2)}{6} \right] \end{aligned} \tag{14}$$

where

$$\begin{aligned} U_{1n} &= \frac{1}{N}, \quad U_{2n} = \left[\frac{-6}{N(N+1)} \right] \left[1 - 2 \frac{(n-1)}{(N-1)} \right], \text{ and} \\ U_{3n} &= \left[\frac{30}{N(N+1)(N+2)} \right] \left[1 - 6 \frac{(n-1)}{(N-1)} + 6 \frac{(n-1)(n-2)}{(N-1)(N-2)} \right] \end{aligned} \tag{15}$$

are derived from orthogonal polynomials [6, 7, 49]. Terms in brackets of (14) are manifestations of $T_m(\tau)$ in (13).

Consider now linearly interpolating between the 2nd and 3rd order estimators $w_{2n}(\tau)$ and $w_{3n}(\tau)$ with $0 \leq f_3 \leq 1$ to formulate the LSMM weighting function, the order of which can be described as $2 + f_3$ [10]. It is described in various forms simply as

$$\begin{aligned} w_{f_3n}(\tau) &= w_{2n}(\tau) [1 - f_3] + f_3 w_{3n}(\tau) \\ &= w_{2n}(\tau) + f_3 [w_{3n}(\tau) - w_{2n}(\tau)] \\ &= U_{1n} + U_{2n} T_2(\tau) + f_3 U_{3n} T_3(\tau) \\ &= w_{2n}(\tau) + f_3 U_{3n} T_3(\tau) \end{aligned} \tag{16}$$

As the estimator of fractional order $2 + f_3$, this is the fraction f_3 times the orthogonal acceleration estimator U_{3n} multiplied by $T_3(\tau)$ and added to $w_{2n}(\tau)$, where the

fraction subscript matches 3 of U_{3n} . This is a special case of the multi-fractional order estimator [10]. As shown later, it is analogous to the IMM.

Now assume a tracking problem with equally time spaced samples corrupted with measurement noise:

$$y_n = \sum_{j=1}^3 c_j n^{j-1} + \eta_n = x_n + \eta_n \tag{17}$$

where the η_n are equivalent to noise samples from the stochastic process noise η in (4). Note that U_{mn} s in (14) do not produce coefficients c_j from equally spaced samples of $x(t) = \sum_{j=1}^3 c_j t^{j-1}$ because of orthogonalization; but (14) does produce the same value at any $x(t = \tau)$.

The estimate of $x(t)$ at time τ from noisy measurements y_n by the LSMM weighting function is

$$\begin{aligned} \hat{x}(\tau) &= \sum_{n=1}^N y_n w_{f_3n}(\tau) = \sum_{n=1}^N y_n \left\{ U_{1n} + U_{2n} \left[\tau - \frac{N+1}{2} \right] \right. \\ &\quad \left. + f_3 U_{3n} \left[\tau^2 - (N+1)\tau + \frac{(N+1)(N+2)}{6} \right] \right\} \end{aligned} \tag{18}$$

For the filtered position corresponding to the last measurement, set $\tau = N$, which yields

$$w_{f_3n}(\tau = N) = U_{1n} + U_{2n} \left[\frac{N-1}{2} \right] + f_3 U_{3n} \left[\frac{(N-1)(N-2)}{6} \right] \tag{19}$$

Because the variance of the sum of weighted uncorrelated random variables equals the sum of the variance of the random variables times the square of their weights [40, 50], and because of coefficient estimator orthogonality

$$\sigma_{f_3}^2 \triangleq \sum_{n=1}^N w_{f_3n}^2 = \frac{1}{N} + \frac{3(N-1)}{N(N+1)} + f_3^2 \frac{5(N-1)(N-2)}{N(N+1)(N+2)} \tag{20}$$

which is the variance of the estimate $\hat{x}(\tau)$ normalized with respect to noise variance σ_η^2 . The first term on the right of the equal sign is the variance of the position estimate; the second term, the variance of the velocity estimate; and the third term for $f_3 = 1$, the variance of the acceleration estimate. The first two terms sum to the 2nd order variance $\frac{4N-2}{N(N+1)}$. Adding the variance from the third term for $f_3 = 1$ yields the 3rd order variance $\frac{9N^2-9N+6}{N(N+1)(N+2)}$. Absent state noise, 2nd and 3rd order KF variances are equivalent to these two variances [9] as a result of the KF first being equivalent to polynomial LS (as Sorenson maintains [8] and Brookner shows [9]).

For $0 \leq f_3 < 1$ a bias is created by the 2nd order estimator applied to a 2nd degree polynomial describing an accelerating target. Normalized with respect to noise standard deviation (SD), the bias is easily discovered to be

$$\begin{aligned}
 B(\tau) &= \rho_3(1 - f_3)T_3(\tau) \\
 &= \rho_3(1 - f_3) \left[\tau^2 - (N + 1)\tau + \frac{(N + 1)(N + 2)}{6} \right] \quad (21)
 \end{aligned}$$

where $\rho_3 \triangleq \frac{c_3}{\sigma_\eta} = \frac{a\Delta^2}{2\sigma_\eta}$ and subscript 3 matches that of c_3 in (17); a is acceleration and Δ is the sample period. For $\tau = N$ this reduces to

$$B(\tau = N) = \rho_3(1 - f_3) \left[\frac{(N - 1)(N - 2)}{6} \right] \quad (22)$$

Normalized with respect to the noise variance, the LSMM MSE is $\sigma_f^2 + B^2$:

$$\begin{aligned}
 MSE &= \frac{4N - 2}{N(N + 1)} + f_3^2 \left[\frac{5(N - 1)(N - 2)}{N(N + 1)(N + 2)} \right] \\
 &+ \rho_3^2(1 - f_3)^2 \left[\frac{(N - 1)(N - 2)}{6} \right]^2 \quad (23)
 \end{aligned}$$

The variance decreases roughly as only N^{-1} ; whereas, the bias-squared increases roughly as the staggering N^4 (deviation between the two is asymptotically proportional to stunning N^5), which quickly exceeds the variance, demanding a trade-off.

3.3 Minimum LSMM MSE

Minimizing the LSMM MSE with variance/bias-squared trade-off by setting the derivative of (23) with respect to f_3 equal to zero and solving yields

$$f_{3opt} = \frac{\rho_3^2}{\rho_3^2 + \frac{180}{N(N^2 - 1)(N^2 - 4)}} \quad (24)$$

Substituting f_{3opt} into (23) reduces the minimum LSMM MSE to

$$MSE_{min} = \frac{4N - 2}{N(N + 1)} + \frac{\rho_3^2}{\rho_3^2 + \frac{180}{N(N^2 - 1)(N^2 - 4)}} \left[\frac{5(N - 1)(N - 2)}{N(N + 1)(N + 2)} \right] \quad (25)$$

Minimized as a function of f_3 , MSE_{min} is smaller than the 3rd order variance, which occurs only when $\rho_3^2 \rightarrow \infty \Rightarrow f_{3opt} \rightarrow 1$ (i.e., $a\Delta^2 \rightarrow \infty$ or $\sigma_\eta = 0$. This can happen only when noise is zero, i.e., $\sigma_\eta = 0$, rendering variance and MSE meaningless.) Thus, 3rd order estimators, including the 3rd order KF, are not minimum MSE estimators. This is a LSMM advantage as a function f_3 and separable functions $U_{1n}T_1(\tau)$, $U_{2n}T_2(\tau)$, and $U_{3n}T_3(\tau)$.

MSE_{min} is a complex non-linear function of two simple parameters, $\rho_3 = \rho_{3opt}$ (effectively, intuitive acceleration-to-noise ratio) and N . Although not obvious, MSE_{min} includes the bias-squared and is the linear interpolation

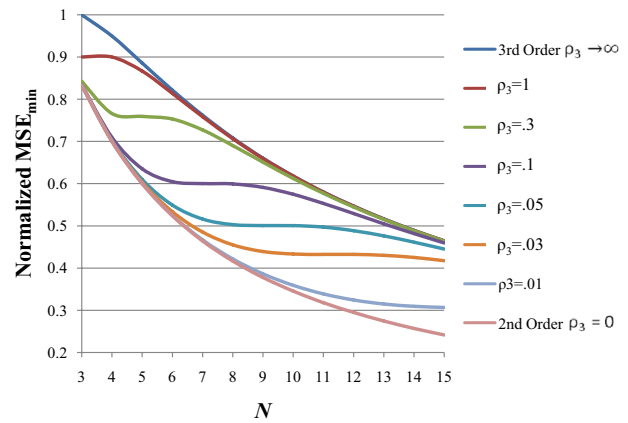


Fig. 1 Normalized MSE_{min} versus N

between the 2nd and 3rd order variances as a function of $f_3 = f_{3opt}$. Whereas, the variance in (23) is the quadratic interpolation between these same 2nd and the 3rd order variances as a function of f_3^2 .

MSE_{min} is similar to KF tuning; the target acceleration being the numerator of ρ_{3opt} , the denominator of ρ_{3opt} being the measurement noise SD, and N corresponding to the state noise variance.

MSE_{min} is uniquely plotted as a function of N for various ρ_3 values in Fig. 1. Several important points are noteworthy in Fig. 1: (a) MSE_{min} curves drop rapidly to a knee, then (b) flatten out beyond the knee yielding virtually no increase in accuracy as N increases until they begin to approach the 3rd order variance. (c) Increasing parameters ρ_3 and N each drive the MSE_{min} toward the 3rd order variance. (d) Parameter $\rho_3 > 1$ almost yields a 3rd order estimator.

Choosing a data window of N samples near the knee is advantageous for two reasons: First, there is virtually no MSE reduction between the knee and the point where the MSE begins to approach the 3rd order variance. Second, windows beyond the knee create RMSE transition spikes from acceleration jumps. The longer the window, the higher the spikes. Shown later.

3.4 False assertion of Kalman filter optimality

Prior to the IMM, the KF was ubiquitously declared to be optimal. In fact, Maybeck audaciously asserts "...the Kalman Filter is optimal with respect to virtually any criterion that makes sense" [51]. Criteria generally cited are minimum MSE and minimum variance unbiased estimator, both of which the KF is claimed to achieve.

However, Brookner points out that state noise variance limits KF measurement noise reduction [9]. This can be easily seen by comparing filtered numerical KF measurement noise variances at each iteration with and without state noise, which demonstrates that the variance from the KF

with state noise is larger than without state noise. Since the KF without state noise is equivalent to contemporary polynomial LS [8,9], polynomial LS is actually the minimum variance unbiased estimator, not the KF.

Furthermore, the KF is not the minimum MSE estimator either. Recall that the MSE is defined as the variance plus the bias-squared. The minimum LSMM MSE in (25) is less than (23) for $f_{3opt} = 1$, which is equal to the KF MSE absent state noise and less than the KF MSE with state noise.

4 Enter The IMM for comparison with the optimal LSMM

Several tracking surveys and comparisons generally concur that the IMM is the best overall maneuvering target tracking estimator [29–31]. This paper offers a comparison of the fundamental differences in the two model—constant velocity and constant acceleration—IMM and optimal LSMM approaches.

4.1 The two model IMM

The estimated IMM state equation is the sum of 2nd order KF times the model probability $\mu_1(k)$ plus 3rd order KF times probability $\mu_2(k)$:

$$X(k|k) = X_1(k|k)\mu_1(k) + X_2(k|k)\mu_2(k) \tag{26}$$

where $X_1(k|k)$ represents 2nd order KF, $X_2(k|k)$ represents 3rd order KF, and k is the time increment [31, 52, 53]. Since model probabilities must necessarily sum to one, i.e., $\mu_1(k) + \mu_2(k) = 1$ [31]; this constitutes linear interpolation (a straight line) between 2nd and 3rd order KFs, analogous to LSMM in (16). IMM interpolation is formed during each recursive cycle as model probabilities are interactively produced and adaptively applied.

The recursive calculation of the model probability $\mu_i(k)$ as a function of time increment k is as follows: $\bar{c}_j = \sum_i \mu_i(k-1)p_{ij}$, p_{ij} is the ad hoc probability of transition from model i to j , $c = \sum_i \lambda_i(k)\bar{c}_i$, $\lambda_i(k)$ is the model likelihood, and $\mu_j(k) = \frac{\lambda_j(k)\bar{c}_j}{c}$ [31, 52, 53].

From $\mu_1(k) = 1 - \mu_2(k)$, the IMM estimate is equal to the sum of the 2nd order KF plus the difference between the 3rd and 2nd order KFs times $\mu_2(k)$, analogous to LSMM:

$$X(k|k) = X_1(k|k) + [X_2(k|k) - X_1(k|k)]\mu_2(k) \tag{27}$$

The difference between $X_2(k|k)$ and $X_1(k|k)$ effectively augments the 2nd order KF with a fraction of the estimated target acceleration as a function of the single model probability $\mu_2(k)$, analogous to the LSMM. This form of IMM is reminiscent of augmenting a 2nd order KF with an acceleration bias based on maneuver detection in [52].

4.2 IMM weakness and deficiencies

Both the LSMM and IMM linearly interpolate between 2nd and 3rd order estimators. However, there are significant differences that expose IMM weaknesses and deficiencies.

(1) State noise is added to the KF state equation prior to prediction. This results in state noise variance being added to the filtered measurement noise variance. Absent state noise; filtered measurement noise variance approaches zero as more data are processed, causing the KF to become unstable and blow up. State noise artificially prevents this by establishing a non-zero variance floor the sum of the two variances asymptotically approach [8, 9].

(2) As noted in Sect. 3.4, establishing a variance floor with state noise simultaneously renders the KF less accurate than absence of state noise (and a variance floor). Moreover, the KF absent state noise (and a variance floor) is equivalent to the truly minimum variance unbiased polynomial LS [8, 9].

(3) In effect, the KF state noise variance creates recursive sliding windows, analogous to non-recursive sliding windows. By hijacking and controlling state noise variance, designers can formulate KF recursive windows of desired length, analogous to non-recursive windows of N samples. When targets accelerate, 2nd order KFs create biases. Increasing state noise variance decreases recursive sliding window sizes—like reducing N in non-recursive windows, which reduces biases. As a specialized KF design parameter, the purpose of which is to size KF recursive sliding windows and establish filtering limits; state noise variance is actually fictitious [52]. Biases from jumps in target acceleration are reduced by designer introduced and controlled fictitious state noise variances.

The purpose of KF based complex IMM filtering is to filter out measurement noise. Is it not utterly obtuse to add even more noise and complexity to force the KF to artificially reduce a bias from a jump in acceleration, but only after discovering that the bias has already created an RMSE transition spike because the KF and IMM obscure the bias itself? The IMM does absolutely nothing to prevent RMSE transition spikes; instead, it continually plays after-the-fact “catch-up”. This paper offers estimators that preclude troublesome transition spikes.

(4) Although KF theory has a random variable representing state noise added to the state equation, it is not actually added in Monte Carlo simulations. The reason is that the sole purpose of fictitious state noise is to add the state noise variance to the random measurement noise variance to prevent the sum of the variances from going to zero, causing the KF from becoming unstable and blowing up. Therefore, state noise variance is not in reality reduced in simulations along with measurement noise variance. Since in KF theory state and measurement noises

are assumed to be independent zero mean, stationary, and white; their variances add. The sum of the two should be reduced in KF simulations, but isn't. Whereas, the sum of the two noises added to a polynomial representing a target trajectory in contemporary polynomial LS estimation does yield the reduced variance of the sum, as it should.

(5) The IMM does not directly address, describe, or even explicitly acknowledge the 2nd order KF acceleration bias per se—the “elephant in the room”, as it were. Hence, the IMM does not perform any form of variance/bias-squared trade-off to actually minimize the MSE; even though the 2nd order KF bias-squared is analogous to the LSMM bias-squared in (22), which quickly exceeds the variance as the bias-squared to variance ratio increases asymptotically proportional to N^5 in (23). This will be shown in examples later.

(6) Furthermore, μ_2 is a combination of likelihoods from KF residuals and ad hoc transition probabilities. The transition probabilities are control parameters, the sole purpose of which is to match specifically assumed scenarios, like KF tuning matches proposed scenarios.

(7) Since the 3rd order KF estimates acceleration from noisy data, multiplying it by μ_2 is noisily incestuous because it too is a function of noisy acceleration data through residuals. This is remindful of positive noise feedback.

(8) In addition, likelihoods in the IMM from KF residuals prevent μ_2 from meeting the required boundary condition of zero absent acceleration to correctly execute the IMM 2nd order KF alone. Instead, $\mu_2 > 0$ augments the 2nd order KF estimate with part of the difference between the 3rd and 2nd order KF estimates (effectively, noise only acceleration estimate), thereby counteracting the IMM oxymoronic default KF 2nd order steady-state measurement noise variance by adding part of the acceleration measurement noise variance to it.

In contradistinction, the LSMM produces the minimum possible two model MSE by performing a variance/bias-squared trade-off while avoiding adverse IMM effects.

4.3 Comparison: f_{3opt} versus μ_2

Functions f_{3opt} and μ_2 have the commonality of linearly interpolating between 2nd and 3rd order estimators. The KF being equivalent to contemporary polynomial LS in the absence of state noise [8, 9], the main substantive differences between optimal LSMM and IMM are differences between f_{3opt} and μ_2 . Table 1 compares them.

Given differences between μ_2 and f_{3opt} , consider differences in performance of IMM and optimal LSMM.

Table 1 Detailed dissection and comparison of f_{3opt} and μ_2

f_{3opt}	μ_2
1. Direct and deterministic function of measurement noise variance	1. Function of measurement noise variance only through multiple likelihoods
2. Direct and deterministic function of noise reduction through N	2. Function of noise reduction only through multiple likelihoods
3. Direct and deterministic function of worst case (high) acceleration, the only unknown to be assumed. (Also commonly assumed in IMM and KF tuning.)	3. Function of acceleration only through noisy estimates of acceleration biases from residuals in multiple likelihoods
4. Direct function of sample rate	4. No direct function of sample rate
5. Optimized with variance/bias-squared trade-off from MSE derivative	5. Not optimized with variance/bias-squared trade-off from MSE derivative, or otherwise
6. No transition probabilities or other ad hoc design parameters	6. Function of transition probabilities as ad hoc design parameters to improve accuracy based on assumed frequency of acceleration changes
7. No recursivity, interactivity, updating, or adaptivity	7. Function of recursivity, interactivity, updating, and adaptivity
8. Very simple one time off-line deterministic computation from ρ_{3opt} and N	8. Complicated with recursive KF, residual, and likelihood updating, interactivity, adaptivity—infused with ad hoc transition probabilities
9. Correctly executes the 2nd order LSMM from the absence of acceleration (i.e., $f_{3opt} = 0$ from $\rho_{3opt} = 0$.)	9. Prevented by likelihoods from matching the boundary condition zero absent acceleration for correctly executing 2nd order KF alone

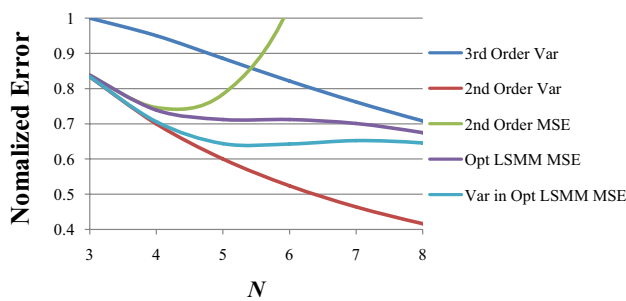


Fig. 2 Variances and MSEs for $\rho_{3opt} = -0.214$ (Pitre)

5 Examples

Generally concurred to be the best overall MM [29–31], the IMM continues to be addressed in the literature. However, a search revealed only the following two particularly well suited examples [31, 53] for comparing the IMM and optimal LSMM. Both examples include comparisons themselves. Ref. [31] compares seven MMs, including the IMM. Ref. [53] compares IMM and KF. Documented by well respected authors whose work is generally included in the KF and IMM gold-standard, these examples facilitate ideal comparisons.

5.1 Example 1 (Pitre scenario)

Consider Pitre’s Scenario 1 in [31, 54] where seven MM tracking algorithms are compared, including the IMM. Pitre compares performance and computational complexity. He states: “IMM and the rest of the algorithms had comparable RMSE values ... IMM, however, achieved it much more efficiently.”

Pitre’s Scenario 1 target constantly accelerates at 20 m/s² for 30 s, travels at constant velocity for 30 s, then constantly decelerates -60 m/s^2 ($> 6 \text{ g}$) for only 10 s, and finally continues at a constant velocity. Sample rate is 1 Hz. The two simulated Pitre accelerations are matched here, but time periods are changed slightly for expediency: Each acceleration lasts 10 s, as does their separation. The measurement SD is 140 m. For worst case acceleration this yields $\rho_3 = \rho_{3opt} = -0.214$ in (25).

From (24) $f_{3opt} \approx 0.39$ weights the 3rd order estimator in (16). The 2nd order estimator is weighted with $(1 - f_{3opt}) \approx 0.61$. Accordingly, the LSMM MSE_{min} of (25) is plotted in Fig. 2 (Opt LSMM MSE). Other curves are shown from (23): variance portion of Opt LSMM MSE (Var in Opt LSMM MSE: $f_3 = f_{3opt}$ and $\rho_3 = 0$), 2nd order MSE ($f_3 = 0$), 2nd order variance ($f_3 = \rho_3 = 0$), and 3rd order variance ($f_3 = 1$).

The knee of the Opt LSMM MSE curve occurs near $N = 5$. Although designed for $\rho_{3opt} = -0.214$ in f_{3opt} , the Opt LSMM MSE is reduced to only the variance—first two terms right of equal sign in (23)—when the acceleration bias is zero, i.e., $\rho_3 = 0$. It also levels off for several increments, as does Opt LSMM MSE. Both Opt LSMM MSE and Var in Opt LSMM MSE approach the 3rd order variance as $N \rightarrow \infty$.

Because of the bias-squared, the 2nd order MSE exceeds the 3rd order variance also near $N = 5$ as pointed out in item 5) of the IMM deficiencies in Sect. 4.2. Particularly troubling is the resulting enormous asymptotic N^5 deviation between the 2nd order MSE and the 3rd order variance for $N > 6$. Since μ_2 is analogous to f_{3opt} , this same general concept applies to the IMM.

Feeding Pitre’s scenario into a 5-point LSMM optimized for $\rho_{3opt} = -0.214$ yields the RMSE shown in Fig. 3 as ρ_3 changes from acceleration jumps between minimum of $\rho_3 = 0$ and worst case of $\rho_{3opt} = -0.214$.

Of particular significance: LSMM linearity obviates Monte Carlo simulation. This makes LSMM plots much simpler. It also eliminates IMM RMSE volatility seen in Pitre’s plots.

The Opt LSMM MSE is the variance plus the bias-squared, the bias-squared being in the form of the square of the difference between the true position and noiseless LSMM estimate. Track initiation is ignored for expediency.

In effect, the optimal LSMM with no RMSE transition spikes reaches stability faster from acceleration jumps than does recursive IMM steady-state.

Choosing $N = 8$ on the Opt LSMM MSE curve in Fig. 2 (beyond the knee) yields the second curve with transition spikes in Fig. 3. In this case $f_{3opt} = 0.89$ yields nearly a 3rd order estimator as Fig. 2 indicates.

An interesting feature in Fig. 3 is that for no acceleration—no bias in (23)—the 5-point optimal LSMM RMSE (i.e., SD) is smaller than from the 8-point optimal LSMM. This seems to violate the notion that larger windows always yield smaller variances (and SDs). The reason: The

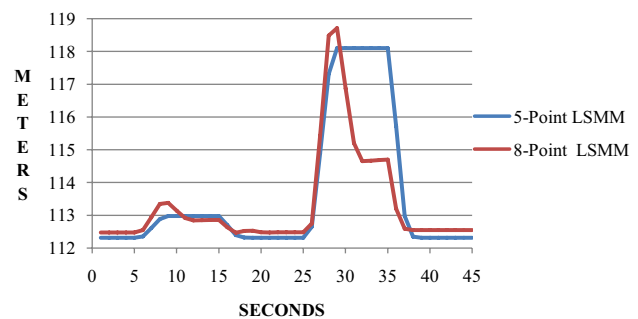


Fig. 3 Optimal LSMM RMSEs for $\rho_{3opt} = -0.214$ (Pitre)

MSE is minimized, not the variance, which in Fig. 2 for $N = 8$ is larger than for $N = 5$.

Obviously, increasing from 5 to 8 data points does not improve the LSMM accuracy much, if at all. Accuracy is arguably worsened by transition spikes before settling to correct RMSEs. There is a small spike at the onset of the first maneuver and tiny ripple at the end from the 8-point LSMM. The second maneuver creates a larger spike at the beginning and just a little uptick at the end. Since the LSMM is optimized for worst case acceleration, the lesser and no accelerations automatically produce corresponding smaller RMSEs from smaller biases.

5.1.1 Comparing LSMM RMSEs with Pitre's Fig. 6

Notice the extremely high RMSE transition spikes above steady-state ($\approx 3\times$ higher) in Pitre's Fig. 6 [31], a transformative use of which is copied in Fig. 4 of this paper. The spikes result from biases caused by jumps in acceleration, which the IMM does nothing to prevent because it fails to explicitly acknowledge existence of biases. The IMM simply adjusts state noise only after recognizing that spikes have occurred.

Within the purpose and scope of this paper; an easy and effective comparison initiated by Pitre himself in Fig. 8 of [31] is applied here. Pitre plots horizontal lines of "credibility measured in terms of log average normalized estimation error squared (LNEES)".

Similar horizontal lines representing the LSMM RMSEs are added here to Pitre's Monte Carlo simulated 7 MM RMSEs in Fig. 4 (Pitre's Fig. 6). There are two sets of horizontal lines representing LSMM RMSEs applied in Fig. 4 that serve different purposes: One set red, the other, brown.

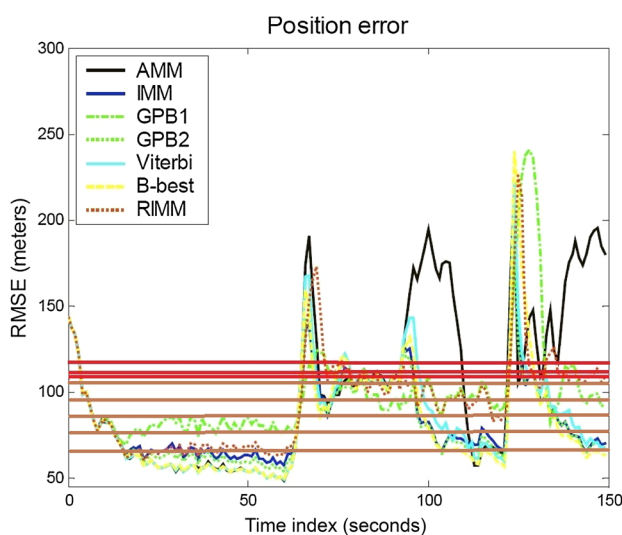


Fig. 4 Copy of Pitre's Fig. 6 with added horizontal lines representing LSMM RMSEs [31]

5.1.2 Red lines in Fig. 4

The red lines represent the 5-point LSMM RMSEs of 118.1 m, 113.0 m, and 112.3 m at the maximum acceleration, the lesser acceleration, and no acceleration; respectively. The lines were located manually (with Windows Paint, mouse, and computer screen) to Pitre's computer created Fig. 6. They are unquestionably imprecise: they are approximations for illustration.

Because of Monte Carlo simulation; the transition spikes and volatile MM RMSEs are due to biases and the algorithms themselves, not the noise. Absent quantitative data in [31], qualitative visible inspection clearly indicates that the IMM RMSE average (including transition spikes) is higher than LSMM RMSEs during both accelerations.

Pitre's IMM curve has a large transition spike at the beginning of the first maneuver and a smaller spike at the end. Thus, both initial first maneuver spikes of the 8-point LSMM and IMM are consistent, although the second LSMM spike is very small because of the relatively small drop in acceleration. However, Pitre's IMM curve shows a very large transition spike at the beginning of the second maneuver and no spike at the end, which is consistent with the 8-point LSMM.

Presumably because of Pitre's higher acceleration short 10 s duration; neither the 8-point LSMM nor the IMM is able to settle down before the end of the second maneuver, which appears to be undetected by both and causes the gradual RMSE reduction with no spikes.

Although Pitre's Fig. 6 provides a variety of colors showing curves well, much of the IMM is obscured by plots of other MMs. The same plots are given in Fig. 2 of [54], where vivid colors are not displayed, but the IMM plot is more distinguishable.

Zooming in on the peaks of plots in Fig. 2 [54] on a computer screen allows slightly better approximations. This shows the first maneuver IMM onset spike near 145 m (28% larger than the LSMM 113 m RMSE); the end spike, about 128 m (13% larger). The second maneuver onset spike is near 210 m (a whopping 78% larger than the LSMM 118 m RMSE).

This suggests that the optimal 5-point LSMM yields significantly lower RMSEs during both accelerations than the IMM. Furthermore, the first LSMM RMSE always remains 19% below the measurement SD of 140 m; the second, 16% below. Whereas, the first IMM onset spike is about the same; the second, 50% larger.

Although Pitre's curves display transition spikes at both the beginning and end of the first maneuver (the spike at the end being smaller), the LSMM 8-point shows a spike only at the beginning. This results from the very small $\rho_{3opt} = -0.214$. Larger values of ρ_{3opt} yield small spikes at maneuver ends, as the next example shows.

This exercise demonstrates a different perspective on the trade-off between worst case and steady-state RMSEs. This paper emphasizes worst case acceleration accuracy. Pitre seems to prefer steady-state over worst case accuracy in the statement: “acceptable RMSE during steady-state with a tradeoff in peak error and quick transition between states”. However, Pitre does not define “acceptable”. Therefore, it must be ad hoc and subjective.

Pitre recommends the IMM based on less complexity, but does not give quantitative complexity results. Pitre states instead: “The computational complexity was calculated using MATLAB’s stopwatch timer” [31]. Fortunately, the optimal 5-point LSMM requires only 5 multiplies and 4 adds.

Figure 4 shows the 5-point LSMM RMSE to be nearly constant across the assumed acceleration spectrum and slightly less than IMM steady-state worst case RMSE (including spikes and volatility), and to be near the average of IMM RMSE including spikes and steady-state volatility.

The optimal LSMM offers significant advantages of lower RMSEs during worst case acceleration, smooth RMSE transitions (no spikes), stable RMSEs, and much less processing.

5.1.3 Brown lines in Fig. 4

Visual inspection of Fig. 4 (copy of Pitre’s Fig. 6) shows 2nd order steady-state position RMSE to be near 68 m, about 50% reduction of 140 m measurement SD. This corresponds to a non-recursive 2nd order LS sliding window of about 15 samples. The optimal LSMM at worst case acceleration is about 118 m, 16% reduction from only 5 samples. This represents a 34% difference in RMSE reduction between 5 and 15 samples.

Consider an adaptive LSMM concept with the following general outline: 4 additional position LSMM estimators ($\tau = N$) covering the two extremes represented by the brown lines in Fig. 4—also drawn manually for illustration and undeniably imprecise. Each is optimally matched to a specific acceleration with a corresponding RMSE shown in Table 2. The RMSEs are purposely designed to be about 10 m apart while keeping ρ_{3opt} at the knee to specifically avoid RMSE transitions spikes. Again, the lines were drawn

Table 2 LSMM RMSEs ($\tau = N$) (Pitre)

N	ρ_{3opt}	Acceleration (m/s ²)	RMSE (m)
5	0.214	60	≈ 118
6	≈ 0.09	≈ 25	≈ 108
7	≈ 0.034	≈ 9.5	≈ 98
9	≈ 0.015	≈ 4.2	≈ 88
12	≈ 0.007	≈ 2	≈ 78
15	≈ 0	≈ 0	≈ 68

manually and are only approximations. Although not shown to avoid cluttering up the figure even more, each of these LSMMs would also yield smaller RMSEs at lower accelerations, like the two lower red lines.

This concept is somewhat analogous to recursive MMs where multiple acceleration models are used, as described by Pitre: “... , a multiple-model target-tracking algorithm runs a set of filters that model several possible maneuvers and non-maneuver target motion.” However, no ad hoc transition probabilities are included in the LSMM for matching specific maneuver scenarios as in the IMM.

The separation of these RMSEs is roughly the same as the variability due to IMM RMSE volatility of IMM steady-state during all three accelerations of zero, 20 m/s², and -60 m/s² in Fig. 4. Most obvious is the deviation at zero acceleration. This suggests that the adaptive LSMMs with no RMSE transition spikes yields better accuracy than the IMM across the spectrum of accelerations between 20 m/s² and -60 m/s².

Multiple LSMMs, each with its own theoretical RMSE threshold in Table 2, against which a corresponding sample RMSE could be compared, would allow rapid detection of threshold crossings. Shorter windows would detect crossings faster than longer ones.

The 2nd order KF steady-state functions as the IMM default. However, the optimal LSMM at maximum acceleration should be default: (a) Because as the shortest window, it will be established first. (b) Because it is likely that in some scenarios large transition spikes from longer windows could disrupt tracking and be more troublesome than no spikes in larger MSEs from shorter LSMM windows.

5.1.4 Adaptive algorithm: one simple example of several possible

Begin each update with the shortest LSMM window (default). Check sample RMSE against the theoretical RMSE threshold. If crossed, predict target location and MSE_{min} with it. Otherwise, sequentially check next longer window sample RMSE against theoretical RMSE threshold until first crossing. Predict target location and MSE_{min} with previous LSMM. No crossings: Predict with longest window LSMM. Update.

Other possibilities: (a) Set a probability threshold for the default window to reduce crossings by large random noise spikes. (b) Weight RMSEs from a sample threshold crossing and the previous theoretical LSMM.

For one-step prediction change τ in (18) to $N + 1$, which is the same as changing all the negative signs on the right side of (19) to positive. Similarly, $\tau = N + 1$ in (the normalized) MSE_{min} —i.e., in $w_{f_3,n}(\tau)$ from (18) and (21) preceding (25)—is equivalent for one-step prediction to switching all signs (+ and -) on the right of (25) in the first term and in

the brackets (not the denominator of f_{3opt}) of the second term [49].

The weights can be stored and applied as needed, requiring no more than N multiplies and $(N - 1)$ adds. Example weights and MSEs for 7-point one-step predictors are given in Table 3.

Notice the characteristic polynomial LS pattern of predictor weights and MSEs, including $w_3 = 0$ in the 2nd order predictor and division of each weight and MSE in the integer order predictors by the number of weights ($N = 7$). Numerators of order 2.0 differ by 1; of order 3.0, by 4, 2, 0, 2, 4, 6, and 8 in sequence.

An alternative method can be adapted from (16) for calculating each optimal LSMM predictor weight in Table 3 from the 2nd and 3rd order estimator weights: $w_{f_{3opt}n} = w_{2n} + f_{3opt}(w_{3n} - w_{2n})$. For example, converting fractions to decimals yields: $w_1 = \{-0.286 + 0.089[0.429 - (-0.286)] = -0.222\}$. Of course, $w_2 = \frac{1}{7}$ and $w_6 = \frac{3}{7}$ because of equal 2nd and 3rd order weights, the difference of which multiplied by f_{3opt} is zero leaving only the 2nd order weight.

Because the normalized variance of non-recursive estimators is the sum of the squares of the weights, which is equivalent to the MSE of the 2nd and 3rd order predictors in Table 3 based on the assumption that each polynomial degree does not exceed its matching predictor order (i.e., no bias); the optimal 7-point LSMM predictor MSE can be calculated analogously from (25): $MSE_{min} = MSE_{2+f_{3opt}} = MSE_2 + f_{3opt}(MSE_3 - MSE_2)$.

However, the sum of the squares of the LSMM weights do not match its MSE. The bias-squared is the difference between the MSE and variance. The normalized variance (sum of the squared weights) of the optimal 7-point 2.089 order LSMM predictor is 0.728, which when subtracted from the MSE of 0.867 yields a normalized bias-squared of 0.139. As described in (21), the bias comes from the 2nd order predictor applied to the maximum acceleration in Pitre’s scenario.

Another advantage of the LSMM over the IMM: It is easy to see what actually transpires in non-recursive estimators, especially regarding biases. *The IMM does not directly address biases. Instead, biases are hidden in mathematical gymnastics, “Kalmanisthenics”, and ever expanding “Complication Theory”: Out of sight, out of mind.*

It is easy to show from Table 3 that the 7-point LSMM predictor RMSE differs very little from the 2nd order RMSE in comparison with the 3rd order RMSE (RMSE ratio of 7-point LSMM to the 2nd order is only 1.102). In practical real world scenarios it probably makes little sense to apply a predictor beyond 7-points (or corresponding KF state noise variance). Example from Table 2: $N = 9$ and $\rho_{3opt} \approx 0.015 \Rightarrow f_{3opt} \approx 0.065 \Rightarrow$ a 2.065 order LSMM predictor and RMSE ratio ≈ 1.065 —virtually no real advantage. In fact, the trade-off from IMM conventionally processing more data samples than the optimal LSMMs at knees amounts to diminishing 2nd order variance reduction (like diminishing returns in economics) versus ever larger MSE transition spikes from target acceleration changes. Thus, 15-point and analogous MM predictors may be great in theory; but are way beyond practicality against maneuvering targets—the original and still main purpose of MMs.

(For Gaussian statistics, a correlation gate based on [55] may apply.)

Since IMM residuals used for adaptation include some, if not all, of the same noise samples in the estimates, there exists some correlation between the two. Thus, distortion from noise in one is shared in the other. Conversely, the LSMM maneuver detection would be based on theoretical RMSEs, not on random noisy residuals.

5.2 Example 2 (Blair scenario)

The main purpose of this scenario is to show a transition spike at the maneuver end from a larger ρ_{3opt} like those in response to the first maneuver in Pitre’s comparison.

Consider Blair’s scenario of [53], which compares the IMM with the KF. After traveling at constant velocity, the target constantly accelerates at 20 m/s² for 20 s and then continues at constant velocity. Blair’s acceleration, 1 Hz sample rate, and $\sigma_\eta = 25$ m all yield $\rho_3 = \rho_{3opt} = 0.4$; which is larger value than Pitre’s.

LSMM MSE_{min} and Opt LSMM Var of (23) are plotted in Fig. 5. Also shown are 2nd order MSE as well as 2nd and 3rd order variances. For Blair’s $\rho_3 = \rho_{3opt} = 0.4$, the knee of the MSE_{min} curve occurs near $N = 4$. These two parameters yield $f_{3opt} \approx 0.39$ —interestingly, the same as Pitre’s for $N = 5$ and $\rho_{3opt} = -0.214$. Again, the 2nd order MSE crosses the 3rd order variance near the MSE_{min} knee, in his case near $N = 4$.

Table 3 7-point predictors. (Subscripts of w_n are n from w_{mn} , which are of order m .)

Order	w_1	w_2	w_3	w_4	w_5	w_6	w_7	MSE
2.0	$-\frac{2}{7}$	$-\frac{1}{7}$	0	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{5}{7}$
3.0	$\frac{3}{7}$	$-\frac{1}{7}$	$-\frac{3}{7}$	$-\frac{3}{7}$	$-\frac{1}{7}$	$\frac{3}{7}$	$\frac{9}{7}$	$\frac{17}{7}$
2.089 (LSMM)	-0.222	$-\frac{1}{7}$	-0.038	0.092	0.248	$\frac{3}{7}$	0.635	0.867

N and ρ_{3opt} from Table 2 in (24) $\Rightarrow f_{3opt} \approx 0.089 \Rightarrow$ optimal LSMM predictor order of 2.089 from (16)

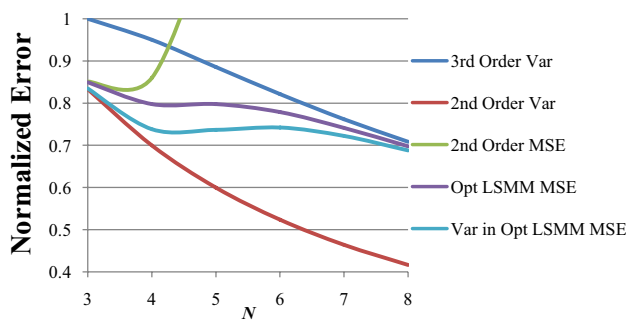


Fig. 5 Variances and MSEs for $\rho_{3, opt} = 0.4$ (Blair)

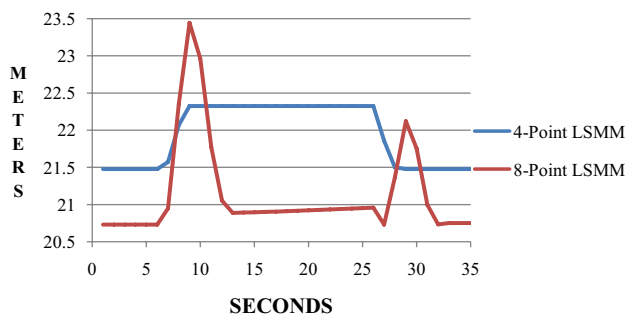


Fig. 6 Optimal LSMM RMSEs for $\rho_{3, opt} = 0.4$ (Blair)

Feeding Blair’s scenario into 4-point optimal LSMM optimized for $\rho_3 = \rho_{3, opt} = 0.4$ yields RMSE shown in Fig. 6. Choosing $N = 8$ (beyond the knee) on Opt LSMM MSE curve in Fig. 5 yields the second curve in Fig. 6, in which case $f_{3, opt} = 0.96$ yields nearly a 3rd order estimator as indicated in Fig. 5. The 8-point RMSE clearly exhibits transition spikes at both the beginning and end of the maneuver, as does Blair’s Fig. 4. However, because of Monte Carlo simulation; Blair’s spike at the end is less pronounced resulting from the RMSE volatility due to the algorithm, not noise. Both the LSMM and Blair’s IMM spike at the end is smaller, as is Pitre’s. However, ignoring Blair’s transition spikes in a visual inspection of his Fig. 4 compared to this Fig. 6 suggests that the optimal 4-point LSMM yields roughly the same steady-state RMSE during acceleration as Blair’s IMM.

Table 4 LSMM RMSEs ($\tau = N$) (Blair)

N	$\rho_{3, opt}$	Acceleration (m/s^2)	RMSE (m)
4	≈ 0.4	≈ 20	≈ 22.5
5	≈ 0.175	≈ 8.75	≈ 20.7
6	≈ 0.065	≈ 3.25	≈ 18.8
7	≈ 0.0	≈ 0	≈ 17

However, including Blair’s spikes suggests superior accuracy from the LSMM RMSE.

Adaptive LSMMs for Blair’s scenario are shown in Table 4, where only 4 are needed. The RMSEs are separated by about a mere 1.75 m while again keeping $\rho_{3, opt}$ at the knee to avoid RMSE transition spikes. Note that this separation is much smaller than the variation of Blair’s IMM during steady-state, which appears to be greater than 2 m—possibly up to 4 m.

6 Review and discussion

The optimal LSMM produces several results for improved tracking: (a) Defining $\rho_3 \triangleq \frac{a\Delta^2}{2\sigma_n}$ (proportional to the acceleration-to-noise ratio) yields this single parameter for tracker design, analysis, and comparison. (b) For comparison, the optimal LSMM MSE_{min} for tracking maneuvering targets can be plotted as a function of time for given ρ_3 and N . (c) Plotting optimal LSMM MSEs (i.e., MSE_{min} s) as a function of N for given ρ_3 yields accuracy curves with knees. (d) The larger is ρ_3 , the smaller is N at the knee and (e) the larger are transition spikes for N beyond the knee. (f) For many—if not most—practical scenario accelerations, the knee is probably in the range $3 \leq N \lesssim 6$ (where $N \geq 3$ for 3rd order acceleration estimation).

Comparisons reveal that the optimal LSMMs designed at MSE_{min} knees produce improved accuracy over the IMM with no volatility or MSE transition spikes during maneuvers at worst case (high) acceleration. It yields smaller MSEs at all lower accelerations automatically (no external adaptivity), even though still optimized for worst case acceleration. This results from smaller biases at lower accelerations in ρ_3 , while the variance remains constant in (23) for $f_{3, opt}$.

The LSMM does not try to capture target acceleration changes when new measurements are received and then decide how to respond. It processes all measurements the same. Maneuvers pass through optimal LSMMs designed at knees of MSE_{min} curves as smooth RMSE bulges with no transition spikes, somewhat analogous to “a pig in a python”.

The LSMM is not limited to a fraction of acceleration (i.e., interpolating only between 2nd and 3rd order estimators). For example, if acceleration is negligible, a fraction of the velocity analogous to (16) may be applicable: $w_{f_2, n}(\tau) = U_{1n} + f_2 U_{2n} T_2(\tau)$.

Based on unique orthogonal polynomial coefficient estimators combined with variance/bias-squared trade-offs; the LSMM designed at MSE_{min} knees is practical and yields optimal accuracy (minimum MSE), RMSE stability, and no transition spikes with only handfuls of multiplies

and adds each during high acceleration maneuvers (when needed most) by giving up minimal accuracy during absence of acceleration (when needed least).

By introducing innovative concepts and new perspectives on tracking; the LSMM offers insight, understanding, flexibility, simplicity, optimality, adaptivity, and practicality. Strictly adhering to straightforward and essential fundamentals while eschewing inimical complication is the key.

Glaring central questions about the IMM: (1) Theoretical: Does IMM steady-state applied to a recurrently maneuvering target with acceleration jumps actually make sense; or is it obviously self-contradictory? (2) Practical: Is the specific benefit of short term IMM RMSE steady-state from maneuvering targets at the expense of volatility and large transition spikes more important than the nearly constant LSMM RMSE?

7 Conclusion

The main purpose of this paper is to derive and apply the novel LSMM based on orthogonal coefficient estimators and a variance/bias-squared trade-off for overcoming weaknesses and deficiencies in the KF based state-of-art IMM target tracking algorithm. In addition, the paper sought to address and clarify several polynomial LS misunderstandings and modeling flaws applied mostly in econometrics.

Thru analysis and examples the paper demonstrates that the LSMM offers improved tracking accuracy and steady-state of continually maneuvering targets with less processing. The paper also demonstrates that the polynomial LS has two main applications that are often confused: First, it fits deterministic data to create polynomial coefficients. Second, it filters out statistical noise to estimate existing polynomial coefficients. They are disparate functions not to be conflated. Additionally, the paper clarifies crucial differences among often conflated deterministic, statistical, and sample statistical modeling.

The state space recursive KF was introduced in 1960 during the computer infancy to overcome the intensive processing required to invert polynomial LS matrices, and to include state noise to keep filtered measurement noise from approaching zero, causing the algorithm from blowing up. The KF has been used for tracking targets with constant velocity or acceleration. The IMM was introduced in 1984 to deal with continually maneuvering targets. These algorithms became state-of-the-art for tracking and largely obviated the incentive to seek alternatives. Fortunately, this author stumbled onto the embryonic form of the LSMM in the mid '80 s and has continued research and

application since. Advantages of polynomial LS have been found and exploited, culminating in this paper.

The main limitation of this study is the development and simulation of specific adaptive LSMM algorithms, some suggested herein. This paper addresses only the fraction of acceleration with adaptivity. However, fractions of velocity, acceleration, jerk, and jink could also possibly be useful. They were used in simulations for the US NAVY on real IRST data and found to yield accuracy superior to that of the extended KF [56].

Three directions for future research come to mind. First, pursue adaptivity. Second, consider the concept in areas other than polynomials, orthogonal series expansions, and tracking. For example, tracking with fractions of derivatives in addition to acceleration is not unlike amplitude weighting in signal processing, antenna beam shaping, and apodizing in optics. Other applications surely await discovery. Third, optimized at critical points on the estimated target trajectory, the LSMM might serve as a benchmark for comparing and improving IMM performance.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

1. Boyd S, Vandenberghe L (2018) Introduction to applied linear algebra vectors, matrices, and least squares. Cambridge University Press, Cambridge
2. Hansen PC (2013) Least squares data fitting with applications. Johns Hopkins University Press, Baltimore, MD
3. Wolberg J (2006) Data analysis using the method of least squares: extracting the most information from experiments, 2006th edn. Springer, Berlin
4. Kariya T, Kurata H (2004) Generalized least squares. Wiley, New York
5. Huffel S, Lemmerling P (eds) (2002) Total least squares and errors-in-variables modeling: analysis, algorithms and applications. Kluwer Academic, Dordrecht
6. Schied F (1968) Numerical analysis, Schaum's Outline Series. McGraw-Hill, New York
7. Wylie CR Jr (1960) Advanced engineering mathematics. McGraw-Hill, New York
8. Sorenson HW (1970) Least-squares estimation: Gauss to Kalman. IEEE Spectrum 7:63–68
9. Brookner E (1998) Tracking and Kalman filtering made easy. Wiley, New York
10. Bell JW (2013) A simple Kalman filter alternative: the multi-fractional order estimator. IET-RSN 7(8):827–835
11. Kalman RE (1960) A new approach to linear filtering and prediction problem. J Basic Eng 82D:35–45
12. Eliason SR (1993) Maximum likelihood estimation: logic and practice. Sage, Newbury Park, CA
13. Nagelkerke NJD (1992) Maximum likelihood estimation of functional relationships. Springer, Berlin

14. Bowman KO, Shenton LR (1998) Estimator: method of moments. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL (eds) Encyclopedia of statistical sciences. Wiley, New York, pp 2092–2098
15. Avdonin SA (1995) Families of exponentials: the method of moments in controllability problems for distributed parameters. Cambridge University Press, Cambridge
16. Harrington R (1993) Field computation by method of moments. Wiley-IEEE Press, New York
17. Haug AJ (2012) Bayesian estimation and tracking: a practical guide. Wiley, Hoboken, NJ
18. Willsky AS (1979) Digital signal processing and control and estimation theory: points of tangency, areas of intersect. MIT Press, Cambridge
19. Sage AP (1971) Estimation theory with applications to communications and control. McGraw-Hill, New York
20. Davidson R (1993) Estimation and inference in econometrics. Oxford University Press, New York
21. Copland TE, Weston JF (1988) Financial theory and corporate policy, 3rd edn. Addison-Wesley, New York
22. Campbell J, Lo A, MacKinlay C (1997) The econometrics of financial markets. Princeton University Press, Princeton
23. Papoulis A (1965) Probability, RVs, and stochastic processes. McGraw-Hill, New York
24. Sorenson HW (1965) Theoretical and computational considerations of linear estimation theory in navigation and guidance, memorandum-LA 3005, June 1965, revised November, 1965. AC Electronics Division, General Motors Corporation, El Segundo, California. (Said in the abstract to “appear as Chapter 5 in Volume III of the Academic Press series advances in control systems, edited by C. T. Leondes of UCLA.”)
25. Fisher RA (1921) The mathematical foundations of theoretical statistics. Philos Trans A 222:309–368
26. Fisher RA (1925) Theory of statistical estimation. Proc Camb Philos Soc 22:700–725
27. Kingsley S, Quegan S (1992) Understanding radar systems. McGraw-Hill, New York
28. Kalata PR (1984) The tracking index: a generalized parameter for α - β and α - β - γ target trackers. IEEE Trans Aerosp Electron Syst 20(2):174–182
29. Li XR, Jilkov VP (2005) Survey of maneuvering target tracking. Part V: multiple-model methods. IEEE Trans Aerosp Electron Syst 41(4):1255–1321
30. Gomes J (2008) An overview on target tracking using multiple model methods. Masters thesis. <https://fenix.tecnico.ulisboa.pt/downloadFile/395137804053/thesis.pdf>. Accessed 2 June 2013
31. Pitre R (2004) A comparison of multiple-model target tracking algorithms. University of New Orleans theses and dissertation. <http://scholarworks.uno.edu/cgi/viewcontent.cgi?article=1197&context=td>. Accessed 9 Apr 2014
32. Blom HAP (1984) An efficient filter for abruptly changing systems. In: Proceedings of the 23rd IEEE conference on decision and control Las Vegas, NV, pp 656–658
33. Blom HAP, Bar-Shalom Y (1988) The interacting multiple model algorithm for systems with Markovian switching coefficients. IEEE Trans Autom Control 33:780–783
34. Mazor E, Averbuch A, Bar-Shalom Y, Dayan J (1998) Interacting multiple model methods in target tracking: a survey. IEEE Trans Aerosp Electron Syst 34(1):103–123
35. Yang C, Blasch E (2008) Characteristic errors of the IMM algorithm under three maneuver models for an accelerating target. In: IEEE 2008 11th international conference on information fusion
36. Liu H, Wen W (2017) Interacting multiple model (IMM) fifth-degree spherical simplex-radial cubature Kalman filter for maneuvering target tracking. Sensors (Basel) 17(6):1374. <https://doi.org/10.3390/s17061374>
37. Liu H, Zhou Z, Yang H (2017) Tracking maneuver target using interacting multiple model-square root cubature Kalman filter based on range rate measurement. Int J Distrib Sens Netw. <https://doi.org/10.1177/1550147717747848>
38. Jan S-S, Kao Y-C (2013) Radar tracking with an interacting multiple model and probabilistic data association filter for civil aviation applications. Sensors 13:6636–6650. <https://doi.org/10.3390/s130506636>
39. Quantitative methods for economics and finance. <http://econ.boun.edu.tr/ozertan/ef507/ch12.ppt>
40. Mosteller F et al (1961) Probability and statistics. Addison-Wesley, Reading, MA
41. Estimating tracking variance. <https://ssrn.com/abstract=3319524>
42. https://en.wikipedia.org/wiki/Ordinary_least_squares
43. A better method of applying OLS to home prices vs. square footage. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3193741
44. Global temperature prediction. <https://ssrn.com/abstract=3319526>
45. Yes, the CAPM is absurd. https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2556958
46. Ordinary least squares (OLS) revolutionized. https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2573840
47. Applying OLS to the CAPM. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3162767
48. https://en.wikipedia.org/wiki/Mean_and_predicted_response
49. MFOE. <https://ssrn.com/abstract=3319519>
50. Selby SM (ed) (1965) Standard mathematical tables, 14th edn. The Chemical Rubber Co., Boca Raton
51. Maybeck PS (1979) Statistic models, estimation, and control, vol 1. Academic Press, New York
52. Watson GA, Blair WD (1995) Interacting acceleration compensation algorithm for tracking maneuvering targets. IEEE Trans Aerosp Electron Syst 31(3):1152–1159
53. Blair WD, Bar-Shalom Y (1996) Tracking maneuvering targets with multiple sensors: does more data always mean better estimates? IEEE Trans Aerosp Electron Syst 32(1):450–456
54. Pitre RA, Jilkov VP, Li XR. A comparative study of multiple-model algorithms for maneuvering target tracking. <https://pdfs.semanticscholar.org/1f18/330549fbc43ce75d9363006ed3dd5b5c636a.pdf>
55. Approximate Gauss PDF. <https://ssrn.com/abstract=2579686>
56. Burkhardt R et al (200) Titan Systems Corporation Atlantic Aerospace Division; Shipboard IRST processing with enhanced discrimination capability; sponsor: Naval Surface Warfare Center, Dahlgren, VA; Contract #: N00178-98-C-3020; September 19, 2000

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.