Polynomial Least-Squares

Jonathan Bunton

Problem setup

Consider an output trajectory $y : \mathbb{R}_{>0} \to \mathcal{Y}$. We are given $N+1$ samples of the trajectory in a time window $[t_0, t_0 + (N+1)\delta] \subseteq \mathbb{R}_{>0}$. More explicitly, we are given samples $\{(t_i, y(t_i))\}$ with $t_i = t_0 + \delta \cdot i$ and $i = 0, 1, ..., N$. We then want to use these samples to estimate the trajectory y and its first d derivatives within the entire time window.

In this note, we propose estimating y within the window by fitting (within each window) a polynomial of degree d, call it $p_d : \mathbb{R} \to \mathcal{Y}$, to the output trajectory samples via least-squares. We then differentiate this polynomial as an approximation to the true derivative of y. We then analyze the error in using this approximation.

Mathematical preliminaries

We assume that the output trajectory y is continuous and $d+1$ times differentiable. We will approximate it using a polynomial (in time) of degree d, denoted by $p_d : \mathbb{R} \to \mathcal{Y}$. For full clarity, p_d is a polynomial of the form:

$$p_{d(t)} = t^d + a_{d-1} t^{d-1} + a_{d-2} t^{d-2} + \cdots + a_1 t + a_0. \tag{1}$$

We fit the polynomial p_d in Equation 1 to the $N+1$ measurements of y, the samples $y(t_i)$, with $i = 0, 1, ..., N$.

Analysis

Let $\mathcal{D} \subseteq \{0, 1, ..., n\}$ be an arbitrary subset of time indices with cardinality $|\mathcal{D}| = d + 1 \leq N + 1$, and then define an associated *residual interpolant* $e_{\mathcal{D}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ as the degree-d polynomial that interpolates the residual errors at the times indexed by $\mathcal{D}$, i.e., such that:

$$y(t_i) - p(t_i) - e_{\mathcal{D}}(t_i) = 0, \quad \forall i \in \mathcal{D}, \tag{2}$$

and the time subset of times are such that $t_{\mathcal{D}_i} < t_{\mathcal{D}_{i+1}}$.

Theorem 01: For any $\mathcal{D} \subseteq \{0, 1, ..., N\}$ and associated residual interpolant $e_{\mathcal{D}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, it holds that:

$$y^{(k)}(t) - p_d^{(k)}(t) = e_{\mathcal{D}}^{(k)}(t) + \frac{y^{(d+1)}(\xi)}{(d-k+1)!} \prod_{i=0}^{d-k} (t - s_i). \tag{3}$$

where $t_{\mathcal{D}_i} < s_i < t_{\mathcal{D}_{i+k}}$ for $i = 0, 1, ..., d-k$ and $\xi \in \left[t_{\mathcal{D}_0}, t_{\mathcal{D}_d}\right]$ is a point inside the interval spanned by the samples in $\mathcal{D}$.

Proof: Define the function $R : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as:

$$R(t) = y(t) - p_{d(t)} - e_{\mathcal{D}}(t). \tag{4}$$

Then by construction, R is continuous, $d+1$ times differentiable, and has $d+1$ zeros, namely:

$$R\left(t_{\mathcal{D}_i}\right) = y\left(t_{\mathcal{D}_i}\right) - p_{d\left(t_{\mathcal{D}_i}\right)} - e_{\mathcal{D}}\left(t_{\mathcal{D}_i}\right) = 0. \tag{5}$$

By Rolle's Theorem, the derivative of R, $R^{(1)}$, has at least d zeros, each located in the interval $\left[t_{\mathcal{D}_i}, t_{\mathcal{D}_{i+1}}\right]$ for $i = 0, 1, ..., d$. Similarly, the second derivative of R, $R^{(2)}$, has at least $d-1$ zeros, each in $\left[t_{\mathcal{D}_i}, t_{\mathcal{D}_{i+2}}\right]$ for $i = 0, 1, ..., d-2$. More generally, the function $R^{(k)}$ has $d-k+1$ zeros, call them s_i, such that $t_{\mathcal{D}_i} < s_i < t_{\mathcal{D}_{i+k}}$ for $i = 0, 1, ..., d-k$.

Then define the auxiliary function $H : \mathbb{R}_{\geq 0} \to \mathbb{R}$ as:

$$H(z) = R^{(k)}(z) - \alpha \prod_{i=0}^{d-k} (z - s_i), \tag{6}$$

for an $\alpha \in \mathbb{R}$. Assume that there exists a value of α such that for some chosen $t \in \mathbb{R}$, $H(t) = 0$. We will show that such an α exists, and its value is independent of $t \in \mathbb{R}$.

If such an α exists, then the auxiliary function H is $d-k+1$ times differentiable and has $d-k+2$ zeros, $z = s_i$, one for each $i = 0, 1, ..., d-k$, but also one at $z = t$. Moreover, using Rolle's Theorem, we know that the $d-k+1$-th derivative of H also has at least one zero, and therefore there exists some $\xi \in \mathbb{R}$ such that:

$$\begin{aligned} H^{(d-k+1)}(\xi) &= R^{k+(d-k+1)}(\xi) - \alpha(d-k+1)! \\ &= y^{(d+1)}(\xi) - \underbrace{p_d^{(d+1)}(\xi)}_{0} - \underbrace{e_{\mathcal{D}}^{(d+1)}(\xi)}_{0} - \alpha(d-k+1)! = 0 \\ \Rightarrow \quad \alpha &= \frac{y^{d+1}(\xi)}{(d-k+1)!}. \end{aligned} \tag{7}$$

By the definition of H in Equation 6 and our construction that requires $H(t) = 0$, plugging in this choice of α we have the equality:

$$\begin{aligned} H(t) &= R^{(k)}(t) - \alpha \prod_{i=0}^{d-k} (t - s_i) \\ &= y^{(k)}(t) - p_d^{(k)}(t) - e_{\mathcal{D}}^{(k)}(t) - \frac{y^{d+1}(\xi)}{(d-k+1)!} \prod_{i=0}^{d-k} (t - s_i) \\ &= 0 \\ \Rightarrow \quad y^{(k)} - p_d^{(k)} &= e_{\mathcal{D}}^{(k)}(t) + \frac{y^{d+1}(\xi)}{(d-k+1)!} \prod_{i=0}^{d-k} (t - s_i), \end{aligned} \tag{8}$$

and our theorem is proved. ∎

Now, some remarks are in order.

Remark 01: Theorem 01 is parameterized by a subset $\mathcal{D}$ with $|\mathcal{D}| = d + 1$. Choosing $\mathcal{D}$ determines the residual interpolant $e_{\mathcal{D}}$ and also the terms $(t - s_i)$ in the product on the right-hand side of Equation 3.

Remark 02: Theorem 01 is an exact equality, meaning there was no "bound loosening" acquired in the process. The result is tight.

Remark 03: In the degenerate case, where we fit a polnomial of degree N to the $N + 1$ data points, this guarantee deteriorates to the one given for perfectly interpolating the data points with a polynomial (see, e.g., Lagrange interpolation).

Remark 04: In the other degenerate case, when the output trajectory y is *actually a polynomial* of degree less than or equal to d, the residual $e_{\mathcal{D}}(t)$ is uniformly zero for all t and choices of $\mathcal{D}$, and $y^{d+1}(\xi) = 0$ for all $\xi \in \mathbb{R}$, thus Theorem 01 guarantees perfect recovery, as expected.

Remark 05: The expression in Equation 3 from Theorem 01 still involves $e_{\mathcal{D}}^{(k)}(t)$, which is the k^{th} derivative of the residual interpolant for the subset $\mathcal{D}$. In principle, this expression could be further unpacked algebraically, since the interpolant of $d + 1$ points by a degree d polynomial has a closed-form expression in terms of its input data (the residuals, in this case). However, the residual error's dependence on this expression alone should be enough to prove the existence of *some* class $\mathcal{K}_\infty$ function from the polynomial regression's residuals to the derivative estimation error for $p_{d(t)}$.

Remark 06: Choosing to use a least-squares polynomial rather than an interpolating one ought to provide some "smoothing" effects. At first glance, these seem absent from the equality in Theorem 01, but they are implicit. The equality consists of two terms. The first term, $e_{\mathcal{D}}^{(k)}$, is the k-th derivative of the error interpolant. Note that as the time between the samples in $\mathcal{D}$ becomes smaller, *these derivatives become smaller*. Conversely, the second term $\frac{y^{(d+1)}(\xi)}{(d-k+1)!} \prod_{i=0}^{d-k} (t - s_i)$ characterizes how "badly behaved" the function y is in the window of time indexed by $\mathcal{D}$. As we shrink this window of times, the price we pay for the function y's "bad behavior", and thus this second term *also decrease*.

Critically, we can choose any subset $\mathcal{D} \subseteq \{t_0, t_1, ..., t_{N-1}\}$, since we solved a *least-squares* problem over all of these times. We should, then, choose the subset $\mathcal{D}$ that strikes the optimal balance between the "smoothing benefits" and the "bad behavior penalty". Doing this explicitly in its most general form would require a lot of algebraic manipulations. For specific instantiations of the problem parameters, we can effectively do this.

From equality to worst-case bound

Theorem 01 gives an expression with *equality*, but relies on computing the zeros of $R^{(k)}$ (the s_i's in the expression for Equation 3) which is not feasible in practice. Some naive upper bounds can help us make the jump from this equality-bound statement to a computable inequality.

Proposition 0.1.1. (Lazy bounds): Assume that the output trajectory has a bounded $d+1$ derivative, i.e., $\left|y^{(d+1)}\right| \leq M$. The error between the k^{th} derivative of the degree-d polynomial $p_d^{(k)}$ and the k^{th} function derivative $y^{(k)}$ is bounded as:

$$\left|y^{(k)} - p_d^{(k)}\right| \leq \alpha\left(\sum_{i=0}^{N} |e(t_i)|\right) + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1}, \tag{9}$$

where $\alpha : \mathbb{R} \to \mathbb{R}$ is a class $\mathcal{K}_\infty$ function.

Proof: Theorem 01 directly implies:

$$y^{(k)}(t) - p_d^{(k)}(t) = e_{\mathcal{D}}^{(k)}(t) + \frac{y^{(d+1)}(\xi)}{(d-k+1)!}\prod_{i=0}^{d-k}(t - s_i) \tag{10}$$

for any choice of $\mathcal{D} \subset \{0, 1, ..., N\}$. We naively choose a subset $\mathcal{D}$ of sequential times that contain the evaluation point $t \in [t_0, t_N]$. Then some simple algebraic steps and the fact that $s_i \in \left[t_{D_i}, t_{D_{i+k}}\right]$ yield:

$$\begin{aligned}
\left|y^{(k)}(t) - p_d^{(k)}(t)\right| &= \left|e_{\mathcal{D}}^{(k)}(t) + \frac{y^{(d+1)}(\xi)}{(d-k+1)!}\prod_{i=0}^{d-k}(t - s_i)\right| \\
&\leq \left|e_{\mathcal{D}}^{(k)}(t)\right| + \frac{M}{(d-k+1)!}\prod_{i=0}^{d-k}|t - s_i| \\
&\leq \left|e_{\mathcal{D}}^{(k)}(t)\right| + \frac{M}{(d-k+1)!}\prod_{i=0}^{d-k}(k+1)\delta \\
&= \left|e_{\mathcal{D}}^{(k)}(t)\right| + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1}.
\end{aligned} \tag{11}$$

Now we turn to the residual interpolant $e_{\mathcal{D}}^{(k)}(t)$. Note that in its Lagrange form, the function $e_{\mathcal{D}}$ can be explicitly written as:

$$\begin{aligned}
e_{\mathcal{D}}(t) &= \sum_{i \in \mathcal{D}} \left(y(t_i) - p_{d(t_i)}\right) l_{i(t)}, \\
l_{i(t)} &= \prod_{j \in \mathcal{D}\setminus\{i\}} \frac{t - t_j}{t_i - t_j}.
\end{aligned} \tag{12}$$

Then for any derivative degree k, the derivative of $e_{\mathcal{D}}(t)$ can be written as:

$$e_{\mathcal{D}}^{(k)}(t) = \sum_{i \in \mathcal{D}} \left(y(t_i) - p_{d(t_i)}\right) l_i^{(k)}(t). \tag{13}$$

Returning to our bound, we find:

$$
\begin{aligned}
\left|y^{(k)} - p_d^{(k)}\right| &\leq \left|e_{\mathcal{D}}^{(k)}\right| + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1} \\
&\leq \sum_{i \in \mathcal{D}} \left|y(t_i) - p_{d(t_i)}\right| \left|l_i^{(k)}(t)\right| + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1} \\
&\leq \sum_{i=0}^{N} \left|y(t_i) - p_{d(t_i)}\right| \left|l_i^{(k)}(t)\right| + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1}.
\end{aligned}
\tag{14}
$$

Note that we have artificially inflated the bound by adding more terms in the summation to accommodate the missing residuals, and we could alternatively retain dependence on the subset $\mathcal{D} \subset \{0, 1, ..., N-1\}$ with $|\mathcal{D}| = d$. Letting $\alpha : \mathbb{R}^{N+1} \to \mathbb{R}$ be the function:

$$
\alpha(a) = \sum_{i=0}^{N} |a_i| \left|l_i^{(k)}(t)\right|, \tag{15}
$$

we have proved exactly what we wanted:

$$
\left|y^{(k)} - p_d^{(k)}\right| \leq \alpha\left(\sum_{i=0}^{N} |e(t_i)|\right) + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1} \tag{16}
$$

■

Slack?

We have explicitly bounded the estimation error as a function of the residuals which we can compute online. What remains is to see how much extra information this provides. In particular, we know that if we assume a *global* bound $\left|y^{(d+1)}\right| \leq M$, this guarantees some *global maximum least-squares residual* within any window of N data points. Combined with Theorem 01 gives us a *global* bound on the worst-case estimation error incurred by our algorithm.

Proposition 0.1.2. (Lazy global bound): Assume $\left|y^{(d+1)}(t)\right| \leq M$ for all $t \in \mathbb{R}$, and let $\alpha : \mathbb{R} \to \mathbb{R}$ be as in Theorem 01. Then at any time $t \in \mathbb{R}$, the least-squares polynomial p_d fit to the window of $N + 1$ data points has the estimation error bound:

$$
\left|y^{(k)} - p_d^{(k)}\right| \leq \alpha\left(\frac{M}{(d+1)!}\sqrt{N+1}(N\delta)^{d+1}\right) + \frac{M}{(d-k+1)!}(k\delta)^{d-k+1}. \tag{17}
$$

Proof: Note that to prove a global bound, we simply need to upper bound $\sum_{i=0}^{N} |e(t_i)|$ and then apply Equation 3 to the result, since α is an increasing function.

Let $p_T : \mathbb{R} \to \mathbb{R}$ be the degree d Taylor approximation to the output map about some time $t_0 \in \mathbb{R}$. Then by construction, its residual at any point satisfies the expression:

$$
y(t) - p_{T(t)} = \underbrace{\frac{y^{(d+1)}(c)}{(d+1)!}(t - t_0)^{d+1}}_{R_{T(t)}} \tag{18}
$$

for some $c \in \mathbb{R}$.

Then, consider that the least-squares polynomial p_d is optimal for the least-squares cost in any window of $N+1$ output measurements. Moreover, the degree d Taylor approximation about t_0 is *feasible* for this cost, which implies:

$$\begin{aligned}\sum_{i=0}^{N} \left|p_{d(t_i)} - y(t_i)\right|^2 &\leq \sum_{i=0}^{N} \left|p_{T(t_i)} - y(t_i)\right|^2 \\ &= \sum_{i=0}^{N} \left|p_{T(t_i)} - \left(p_{T(t_i)} + R_{T(t_i)}\right)\right|^2 \\ &= \sum_{i=0}^{N} \left|R_{T(t_i)}\right|^2 \\ &\leq \sum_{i=0}^{N} \left[\frac{\left|y^{(d+1)}(c)\right|}{(d+1)!}\left|t_i - t_0\right|^{d+1}\right]^2 \\ &\leq \sum_{i=0}^{N} \left[\frac{M}{(d+1)!}\left|t_i - t_0\right|^{d+1}\right]^2.\end{aligned} \tag{19}$$

Now, we note that if we place t_0 in our window, then we can loosely bound:

$$\begin{aligned} \sum_{i=0}^{N} |e(t_i)|^2 &\leq \sum_{i=0}^{N} \left[\frac{M}{(d+1)!}\left|t_i - t_0\right|^{d+1}\right]^2 \\ \Rightarrow \quad \sqrt{\sum_{i=0}^{N} |e(t_i)|^2} &\leq \frac{M}{(d+1)!}\sqrt{N+1}(N\delta)^{d+1}. \end{aligned} \tag{20}$$

Finally, since $\|e\|_1 \leq \sqrt{N}\|e\|_2$, we have:

$$\sum_{i=0}^{N} |e(t_i)| \leq \frac{M\sqrt{N^2+N}}{(d+1)!}(N\delta)^{d+1}. \tag{21}$$

We can then plug this bound into Equation 3 and recover our desired result. ■