APPENDIX D

Offline guarantees

We show here how the guarantee given by Corollary 4 can be immediately used to derive a fully *offline* guarantee for estimation. Recalling Corollary 4 here for clarity:

Corollary 5. *Assume that $|y^{(d+1)}(\xi)| \leq M$ for all $\xi \in [s_0, s_d]$, and that the measurement noise is uniformly bounded by $|e(s_i)| \leq E$ for all $s_i \in \mathcal{D}$. If the subset $\mathcal{D}$ has maximal inter-sample spacing $s_{i+1} - s_i \leq \delta$, then:*

$$\begin{aligned} |y^{(k)}(t) - p^{(k)}(t)| \leq \sum_{s_i \in \mathcal{D}} \left| l_i^{(k)}(t) \left(y(s_i) + e(s_i) - p(s_i) \right) \right| \\ + E \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| + M\delta^{d-k+1}, \end{aligned} \tag{D.0.1}$$

where $l_i : \mathbb{R} \to \mathbb{R}$ with $i = 0, 1, ..., d$ are the Lagrange basis polynomials for $\mathcal{D}$:

$$l_i(t) = \prod_{s_j \in \mathcal{D} \setminus \{s_i\}} \frac{t - s_j}{s_i - s_j}. \tag{D.0.2}$$

D.1 Explicitly bounding residuals

In order to transition the guarantee given by (D.0.1) from one that depends on the online residuals to a fully offline bound, we simply need to bound the worst-case values of these residuals. By assuming a global limit $||y^{(d+1)}(t)| \leq M$ for all $t \in \mathbb{R}_{\geq 0}$, we can immediately derive this bound with a simple application of the Taylor Remainder Theorem.

Corollary 6. *Assume $|y^{(d+1)}| \leq M$ for all $t \in \mathbb{R}$. Then for any time $t \in \mathbb{R}$, the least-squares* *polynomial* $p_{LS} : \mathbb{R} \to \mathbb{R}$ *fit to the window of* $N+1$ *data points satisfies the bound:*

$$|y^{(k)}(t) - p_{LS}^{(k)}(t)| \leq M \left(\frac{L^{(k)}(t)\delta^{d+1}}{(d+1)!} \sqrt{2S_{2(d+1)}\left(\left\lceil \frac{N+1}{2} \right\rceil\right)} + \delta^{d-k+1} \right)$$
$$+ EL^{(k)}(t)\left(\sqrt{N+1}+1\right),$$

where we have defined:

$$L^{(k)}(t) = \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)|.$$

Proof. First, consider any sequence of $N+1$ measurements, from which we have constructed the least-squares polynomial p_{LS}. Then (D.0.1) states:

$$\begin{aligned}|y^{(k)}(t) - p_{LS}^{(k)}(t)| &\leq \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)\left(y(s_i) + e(s_i) - p_{LS}(s_i)\right)| + E \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| + M\delta^{d-k+1} \\ &\leq \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| \cdot |(y(s_i) + e(s_i) - p_{LS}(s_i)| + E \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| + M\delta^{d-k+1}.\end{aligned} \tag{D.1.1}$$

Now we also consider the degree d Taylor approximation to y, denoted $p_T : \mathbb{R} \to \mathbb{R}$, expanded about some time $t_0 \in [s_0, s_d]$. By construction, the Taylor approximation satisfies:

$$y(t) - P_T(t) = \frac{y^{(d+1)}(c)}{(d+1)!}(t - t_0)^{d+1}, \tag{D.1.2}$$

for some $c \in [s_0, s_d]$.

To finish our conversion to an offline bound, we need to characterize the weighted sum of the residual errors in (D.1.1). Consider any single residual from this summation and note:

$$|y(s_i) + e(s_i) - p_{LS}(s_i)| \leq \max_{s_i \in \mathcal{D}} |y(s_i) + e(s_i) - p_{LS}(s_i)| \tag{D.1.3}$$

$$\leq \max_{i=0,1,\ldots,N} |y(t_i) + e(t_i) - p_{LS}(t_i)| \tag{D.1.4}$$

$$\leq \sqrt{\sum_{i=0}^{N} |y(t_i) + e(t_i) - p_{LS}(t_i)|^2}, \tag{D.1.5}$$

where the final inequality arises by the fact that the ℓ_2 norm always upper bounds the ℓ_∞ norm. Next we recall that the least-squares polynomial achieves minimal ℓ_2 norm of its residuals, and further that the degree d Taylor polynomial p_T is *feasible* in the associated optimization problem. Therefore, we have the bound:

$$|y(s_i) + e(s_i) - p_{LS}(s_i)| \leq \sqrt{\sum_{i=0}^{N} |y(t_i) + e(t_i) - p_T(t_i)|^2}. \tag{D.1.6}$$

Next, by directly applying (D.1.2) and the triangle inequality for the ℓ_2 norm, we have:

$$\begin{aligned}
|y(s_i) + e(s_i) - p_{LS}(s_i)| &\leq \sqrt{\sum_{i=0}^{N} \left| \frac{y^{(d+1)}(c)}{(d+1)!}(t_i - t_0)^{d+1} + e(t_i) \right|^2} && \text{(D.1.7)} \\
&\leq \sqrt{\sum_{i=0}^{N} \left| \frac{y^{(d+1)}(c)}{(d+1)!}(t_i - t_0)^{d+1} \right|^2} + \sqrt{\sum_{i=0}^{N} |e(t_i)|^2} && \text{(D.1.8)} \\
&\leq \sqrt{\left(\frac{|y^{(d+1)}(c)|}{(d+1)!} \right)^2 \sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} + \sqrt{\sum_{i=0}^{N} |e(t_i)|^2} && \text{(D.1.9)} \\
&= \frac{|y^{(d+1)}(c)|}{(d+1)!} \sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} + \sqrt{\sum_{i=0}^{N} |e(t_i)|^2}. && \text{(D.1.10)}
\end{aligned}$$

Next, we recall our global bounds, in particular that $|y^{(d+1)}(t)| \leq M$ for all $t \in \mathbb{R}$, and also that the measurement errors are bounded, $|e(t_i)| \leq E$ for all t_i. We immediately recover the following:

$$\begin{aligned}
|y(s_i) + e(s_i) - p_{LS}(s_i)| &\leq \frac{M}{(d+1)!} \sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} + \sqrt{(N+1)E^2} && \text{(D.1.11)} \\
&= \frac{M}{(d+1)!} \sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} + E\sqrt{N+1}. && \text{(D.1.12)}
\end{aligned}$$

Lastly, we note that we may only claim the inequality given by Corollary 4 when considering a $t \in [s_0, s_d]$. We may place the expansion point for our Taylor approximation t_0 anywhere in our window of interest, and by lazily selecting t_0 to be the median in the window of $N+1$

points, we can bound:

$$\sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} \leq \sqrt{\sum_{i=1}^{\lceil \frac{N+1}{2} \rceil} 2\left(i \cdot \delta\right)^{2(d+1)}} \tag{D.1.13}$$

$$\leq \delta^{d+1} \sqrt{2 \sum_{i=1}^{\lceil \frac{N+1}{2} \rceil} i^{2(d+1)}}. \tag{D.1.14}$$

since the maximal inter-sample spacing is δ. We note here that the final summation is a *power sum*, typically denoted by:

$$S_{2(d+1)}\left(\left\lceil \frac{N+1}{2} \right\rceil\right) = \sum_{i=1}^{\lceil \frac{N+1}{2} \rceil} i^{2(d+1)}, \tag{D.1.15}$$

for which various closed-form expressions exist. A more naive (and definitively looser) bound for these terms could easily be derived by bounding every term $|t_i - t_0|$ by the maximal window length, instead deriving the bound:

$$\sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} \leq \sqrt{\sum_{i=0}^{N} (N\delta)^{2(d+1)}} \tag{D.1.16}$$

$$= \sqrt{N+1}(N\delta)^{(d+1)}. \tag{D.1.17}$$

Finally, we return to our bound for a single term (D.1.12), applying the power sum bound (D.1.15) and find:

$$|y(s_i) + e(s_i) - p_{LS}(s_i)| \leq \frac{M}{(d+1)!} \sqrt{\sum_{i=0}^{N} |t_i - t_0|^{2(d+1)}} + E\sqrt{N+1} \tag{D.1.18}$$

$$\leq \frac{M\delta^{d+1}}{(d+1)!} \sqrt{2S_{2(d+1)}\left(\left\lceil \frac{N+1}{2} \right\rceil\right)} + E\sqrt{N+1}. \tag{D.1.19}$$

Or alternatively, choosing the looser bound (D.1.17):

$$|y(s_i) + e(s_i) - p_{LS}(s_i)| \leq \frac{M\sqrt{N+1}(N\delta)^{d+1}}{(d+1)!} + E\sqrt{N+1}. \tag{D.1.20}$$

Finally, we can return to the full weighted sum (D.1.1) to find:

$$|y^{(k)}(t) - p_{LS}^{(k)}(t)| \leq \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| \cdot \left[\frac{M\delta^{d+1}}{(d+1)!} \sqrt{2S_{2(d+1)}\left(\left\lceil \frac{N+1}{2} \right\rceil\right)} + E\sqrt{N+1} \right] \tag{D.1.21}$$

$$+ E \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)| + M\delta^{d-k+1}. \tag{D.1.22}$$

For notational cleanliness, we define $L^{(k)}(t) = \sum_{s_i \in \mathcal{D}} |l_i^{(k)}(t)|$, and re-arrange terms to see:

$$|y^{(k)}(t) - p_{LS}^{(k)}(t)| \leq M \left(\frac{L^{(k)}(t)\delta^{d+1}}{(d+1)!} \sqrt{2S_{2(d+1)}\left(\left\lceil \frac{N+1}{2} \right\rceil\right)} + \delta^{d-k+1} \right) \tag{D.1.23}$$

$$+ EL^{(k)}(t)\left(\sqrt{N+1} + 1\right). \tag{D.1.24}$$

□

The exact algebraic form of the bound in Corollary 6 is complicated, but it does exhibit some key behaviors we would expect. For example, hgiher derivative orders have weaker guarantees. In particular, the value of δ^{d-k+1} increases and the values of $L^{(k)}(t)$ increase monotonically with increasing k. Additionally, as the sampling time decreases (δ shrinks), the guarantees become tighter.

One key behavior exhibited in the bound from Corollary 6 is that it consists of two terms, one which is proportional to the measurement error bound E, and one which is proportional to the "ill-conditioning" of the target function M. This type of bound not only makes sense, but has its theoretical roots in ill-posed inverse problem theory [Dio07, Kir11].

D.2 Directly using least-squares

A final observation is that we could have derived the offline guarantee in Corollary 6 through the lens of a pure least-squares analysis.

In particular, Savitzky-Golay filtering solves the least squares problem:

$$\underset{a \in \mathbb{R}^{d+1}}{\text{minimize}} \quad \|Y + Z - Fa\|_2^2, \tag{D.2.1}$$

where $Y \in \mathbb{R}^{N+1}$ and $Z \in \mathbb{R}^{N+1}$ denote the measurement and noise vectors, and $F \in \mathbb{R}^{N+1 \times d+1}$ is the relevant *Vandermonde* matrix of polynomial regression coefficients. Explicitly, these matrices are:

$$Y = \begin{pmatrix} y(t_0) \\ y(t_1) \\ \cdots \\ y(t_N) \end{pmatrix}, \quad Z = \begin{pmatrix} e(t_0) \\ e(t_1) \\ \cdots \\ e(t_N) \end{pmatrix}, \quad F = \begin{pmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^d \\ 1 & t_1 & t_1^2 & \cdots & t_1^d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_N & t_N^2 & \cdots & t_N^d \end{pmatrix}. \tag{D.2.2}$$

The solution of this optimization problem is a set of coefficients $a_{LS} \in \mathbb{R}^{d+1}$ for the least-squares polynomial. To evaluate it and its derivatives at a point, we may consider the "evaluation matrix" $B(t) \in \mathbb{R}^{d \times d+1}$:

$$B(t) = \begin{pmatrix} 1 & t & t^2 & \cdots & t^d \\ 0 & 1 & 2t & \cdots & dt^{d-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d! & \end{pmatrix}, \tag{D.2.3}$$

which is built such that $B(t)a_{LS} \in \mathbb{R}^d$ produces a vector containing the Savitzky-Golay filter estimates of the d derivatives of y at the time t.

As we did in the proof of Corollary 6, we note that the degree d Taylor approximating polynomial expanded about some t_0 is also feasible for the optimization problem (D.2.1). The Taylor approximation corresponds to a set of coefficients $a_T \in \mathbb{R}^{d+1}$, and in particular we know (via the Taylor Remainder Theorem) that:

$$Y = Fa_T + R_T, \tag{D.2.4}$$

where the errors R_T are derived from (D.1.2). We actually know from our work in Corollary 6 that these bounds will consist of two terms, one relating to the $d+1$ derivative bound M and one relating to the measurement noise bound E, with powers of the inter-sample spacing δ.

Using this simple observation, we can directly derive a simple offline bound.

Lemma 12. *Let the vector of remainders for a degree d Taylor approximation of y about $t \in \mathbb{R}$ be defined as R_T (as we did in* (D.2.4)*). Then the least-squares estimator has error bounded by:*

$$|y^{(k)}(t) - \hat{y}^{(k)}(t)| = |y^{(k)}(t) - [B(t)a_{LS}]_k| \leq |B_k(t)F^{\dagger}|_{\infty} \left(|R_T|_{\infty} + E\right), \tag{D.2.5}$$

where $F^{\dagger} \in \mathbb{R}^{d+1 \times N+1}$ denotes the Moore-Penrose pseudo-inverse of the matrix F.

Proof. Directly computing:

$$|y^{(k)}(t) - [B(t)a_{LS}]_k| = |y^{(k)}(t) - [B(t)a_T]_k + [B(t)a_T]_k - [B(t)a_{LS}]_k|. \tag{D.2.6}$$

Because we have considered the Taylor series expanded about the time t, the error in its approximation of $y^{(k)}(t)$ is exactly zero for all orders $k \leq d$. Then, using our definition of the residual vector in (D.2.4) and the closed form expression for the least-squares solution to the optimization problem (D.2.1), we have:

$$\begin{aligned}
|y^{(k)}(t) - [B(t)a_{LS}]_k| &= |[B(t)F^{\dagger}(Y + R_T)]_k - [B(t)F^{\dagger}(Y + Z)]_k| && \text{(D.2.7)}\\
&\leq |[B(t)F^{\dagger}(R_T - Z)]_k| && \text{(D.2.8)}\\
&\leq \|B(t)F^{\dagger}(R_T - Z)\|_{\infty} && \text{(D.2.9)}\\
&\leq \|B(t)F^{\dagger}\|_{\infty}\|R_T - Z\|_{\infty} && \text{(D.2.10)}\\
&\leq \|B(t)F^{\dagger}\|_{\infty}\left(\|R_T\|_{\infty} + \|Z\|_{\infty}\right) && \text{(D.2.11)}\\
&\leq \|B(t)F^{\dagger}\|_{\infty}\left(\|R_T\|_{\infty} + E\right), && \text{(D.2.12)}
\end{aligned}$$

where we slightly abused notation to use $\|\cdot\|_{\infty}$ to denote the ℓ_{∞} vector norm and also the operator infinity norm of a matrix. □

These guarantees, while very direct, obfuscate the roles of the relevant design parameters. In particular, the smoothing properties of least-squares and the roles of the residual are all hidden within the complicated pseudo-inverse of the Vandermonde matrix F. We opted for the form o guarantees in the main body of this work because they provide a more explicit characterization of the tradeoffs between various design parameters in the estimator.