Neural Filtering for State Space Models

Randolf Scholz

July 20, 2023

A filter is a function $F\colon \mathcal{X} \oplus \mathcal{Y} \longrightarrow \mathcal{X}$, that, given an observation $y \in \mathcal{Y}$ and a state $x \in \mathcal{X}$ returns an updated state estimate $x' \in \mathcal{X}$

$$x' = F(x, y)$$

We say a filter is *autoregressive*, if $\mathcal{Y} = \mathcal{X}$ ($\mathcal{Y} = {}^{\bullet}\mathcal{X}$ including nan values), i.e. the observation space $\mathcal{Y}$ is equal to the state space $\mathcal{X}$. In this case we usually write x^{obs} instead of y. We say an autoregressive filter is *idempotent*, if and only if $x \equiv x^{\text{obs}} \implies x' = x$.[1] More specifically, we say a filter is continuously (differentiably) idempotent, if $F(x, x^{\text{obs}}) = G(x - x^{\text{obs}})$ for some continuous (differentiable) function G with $z \equiv 0 \implies G(z) = 0$. We also consider the problem of dealing with multiple simultaneous measurements. In this case, we ask that the filter should be order-independent, for which there are 2 options:

- satisfy $F(F(x, y_1), y_2) = F(F(x, y_2), y_1)$
- change the filter to accept multiple y-values: $F\colon \mathcal{X} \oplus \bigcup_{n\in\mathbb{N}} \mathcal{Y}^n \longrightarrow \mathcal{X}$ such that $F(x, Y) = F(x, P \cdot Y)$ for any permutation matrix P.

Additionally, it should satisfy a certain scale invariance:

Example 1 (Linear Filter)**.** We say a filter is *linear*, if it is of the form $F(x, y) = Ax + By$. When is such a filter order-independent?

$$\begin{aligned} F(F(x, y_1), y_2) = F(F(x, y_1), y_2) &\iff A(Ax + By_1) + By_2 = A(Ax + By_2) + By_1 \\ &\iff ABy_1 + By_2 = ABy_2 + By_1 \\ &\iff AB(y_1 - y_2) = B(y_1 - y_2) \\ &\iff (\mathbb{I} - A)B(y_1 - y_2) = 0 \end{aligned}$$

Since the equality must hold all combinations of y_1 and y_2, it follows that $(\mathbb{I} - A)B = 0$, i.e. $B = AB$. Alternatively, consider the multi-value linear filter $f(x, Y) = Ax + g(Y)$, where $g\colon \mathcal{Y}^{\oplus\mathbb{N}} \longrightarrow \mathcal{X}$ is linear. We can show such g are of the form $g(Y) = \mathbf{1}_n^{\top} Y B$.

[1] $u \equiv v$ if $u_i = v_i$ for all non-nan components.

Example 2 (conditional filtering)**.** Consider a linear control system

$$\dot{x} = Ax + Bu + \nu$$
$$y = Cx + Du + \omega$$

In this situation, we consider *conditional filters* that have an additional input:

$$F\colon \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{U} \longrightarrow \mathcal{X}, (x, y, u) \longmapsto x'$$