

# Notes on the MSE calculation

Jeffrey J. Early

October 16, 2018—updated January 31st, 2019

Using our definitions,

$$\text{MSE}(\lambda) = \frac{1}{N} \|(\mathbf{S}_\lambda - I) \mathbf{x}\|^2 + \frac{2\sigma^2}{N} \text{Tr} \mathbf{S}_\lambda - \sigma^2 \quad (1)$$

$$= \left(1 - \frac{1}{d_{\text{var}}}\right) \sigma^2 + \frac{2\sigma^2}{d_{\text{mean}}} - \sigma^2 \quad (2)$$

$$= \sigma^2 \left( \frac{2}{d_{\text{mean}}} - \frac{1}{d_{\text{var}}} \right) \quad (3)$$

If we can assume that the DOF are the same, then we find that,

$$\text{MSE} = \frac{\sigma^2}{d} \quad (4)$$

which is a fairly intuitive result.

Key to note here is that the tension parameter in terms of degrees of freedom, has a minimum of  $d = 1$ , and starts to asymptote much past  $d = 10$ . So the range of tension parameters to consider is actually quite small.

Ultimately we want to reduce the MSE of all order of the spline fit. If the higher order information is all garbage, then we should know that.

Roughly speaking, this means we simply compute the MSE for velocity, acceleration, etc.

$$\text{MSE}(\lambda) = \frac{1}{N} \|\mathbf{S}_\lambda \mathbf{x} - \mathbf{x}_{\text{true}}\|^2 \quad (5)$$

$$= \frac{1}{N} [\mathbf{S}_j^i x^j - g^i]^2 \quad (6)$$

$$= \frac{1}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j + \mathbf{S}_j^i \epsilon^j]^2 \quad (7)$$

$$= \frac{1}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j + \mathbf{S}_j^i \epsilon^j]^2 \quad (8)$$

$$= \frac{1}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 + \frac{1}{N} [\mathbf{S}_j^i \epsilon^j]^2 + \frac{2}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] [\mathbf{S}_j^i \epsilon^j] \quad (9)$$

Consider,

$$[\mathbf{S}_j^i \epsilon^j]^2 = \mathbf{S}_j^i \epsilon^j \mathbf{S}_k^i \epsilon^k \quad (10)$$

$$= \mathbf{S}_j^i \mathbf{S}_k^i \epsilon^j \epsilon^k \quad (11)$$

$$= \mathbf{S}_j^i \mathbf{S}_j^i \epsilon^j \epsilon^j \quad (12)$$

$$= \sigma^2 \mathbf{S}_j^i \mathbf{S}_j^i \quad (13)$$

$$= E \left[ \sum_i (\mathbf{S}_j^i \epsilon^j)^2 \right] \quad (14)$$

So this is the sum of the square of all the components in the matrix, which is the the trace of the of the matrix. Now, under expectation we have that,

$$\text{MSE}(\lambda) = \frac{1}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 + \frac{\sigma^2}{N} \mathbf{S}_j^i \mathbf{S}_j^i \quad (15)$$

$$= \frac{1}{N} [(\mathbf{S}_j^i - \mathbf{I}_j^i) (x^j - \epsilon^j)]^2 + \frac{\sigma^2}{N} \mathbf{S}_j^i \mathbf{S}_j^i \quad (16)$$

If we include a derivative in there,

$$\text{MSE}(\lambda) = \frac{1}{N} [D_i^k (\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 + \frac{1}{N} [D_i^k \mathbf{S}_j^i \epsilon^j]^2 \quad (17)$$

$$+ \frac{2}{N} [D_i^k (\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] [D_i^k \mathbf{S}_j^i \epsilon^j] \quad (18)$$

The last term vanishes under expectation, as usual. The second to last term looks the same as before, in the sense that the matrix operation and the derivative can be combined,

and so nothing really changed.

$$\text{MSE}(\lambda) = \frac{1}{N} \left[ D^k_i (\mathbf{S}^i_j - \mathbf{I}^i_j) g^j \right]^2 + \frac{1}{N} \left[ D^k_i \mathbf{S}^i_j \epsilon^j \right]^2 \quad (19)$$

$$= \frac{1}{N} \left[ D^k_i (\mathbf{S}^i_j - \mathbf{I}^i_j) (x^j - \epsilon^j) \right]^2 + \frac{1}{N} \left[ D^k_i \mathbf{S}^i_j \epsilon^j \right]^2$$

$$= \frac{1}{N} \left[ A^k_j (x^j - \epsilon^j) \right]^2 + \frac{1}{N} \left[ D^k_i \mathbf{S}^i_j \epsilon^j \right]^2 \quad (20)$$

$$= \frac{1}{N} \left[ A^k_j x^j - A^k_j \epsilon^j \right]^2 + \frac{1}{N} \left[ D^k_i \mathbf{S}^i_j \epsilon^j \right]^2 \quad (21)$$

$$= \left[ A^k_j x^j \right]^2 + \left[ A^k_j \epsilon^j \right]^2 - 2 \left[ A^k_j x^j \right] \left[ A^k_j \epsilon^j \right]$$

$$+ \left[ D^k_i \mathbf{S}^i_j \epsilon^j \right]^2 \quad (22)$$

Ultimately I get that,

$$MSE(\lambda) = \|D(I - S)x\|^2 - \sigma^2 \text{Tr}(D - DS)^2 + \sigma^2 \text{Tr}(DS)^2 \quad (23)$$

## 1 Unknown error

Just eliminate  $d$  from the sample variance and sample mean equations, so that,

$$\left( 1 - \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda) \right) \sigma^2 = \frac{1}{N} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{x}\|^2 \quad (24)$$

or

$$\sigma^2 = \frac{\frac{1}{N} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{x}\|^2}{1 - \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda)} \quad (25)$$

Now the MSE becomes,

$$\text{MSE}(\lambda) = \frac{1}{N} \|(\mathbf{S}_\lambda - I) \mathbf{x}\|^2$$

$$+ \left( \frac{\frac{1}{N} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{x}\|^2}{1 - \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda)} \right) \left( \frac{2}{N} \text{Tr} \mathbf{S}_\lambda - 1 \right) \quad (26)$$

$$= \frac{1}{N} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{x}\|^2 \left( \frac{\frac{1}{N} \text{Tr} \mathbf{S}_\lambda}{1 - \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda)} \right) \quad (27)$$

Alternatively, if we normalize the MSE by  $\sigma^2$ , then

$$\frac{1}{\sigma^2} \text{MSE}(\lambda) = \frac{1}{N} \text{Tr}(\mathbf{S}_\lambda) \quad (28)$$

which, duh, should have been obvious from my earlier calculation that the MSE is  $\sigma^2/d$ .

If the error  $\sigma$  is not known, we can estimate it. For each value of sigma.

At one extreme the signal could be all noise

$$\sigma_{\max} = \sqrt{\frac{1}{N} \sum (x_i - \bar{x})^2} \quad (29)$$

and thus our total degrees of freedom is  $d = N$ . At the other extreme, the noise is effectively zero, in that it is much less than the mean distance traversed,

$$\sigma_{\min} = \frac{1}{50} \sqrt{\frac{1}{N-1} \sum (x_i - x_{i-1})^2} \quad (30)$$

which is equivalent to setting  $\Gamma = 0.1$

## 2 MSE calculation revisited

I want to revisit the MSE calculation with the idea that the noise has a known and an unknown part.

$$N \cdot \text{MSE}(\lambda) = \|\mathbf{S}_\lambda \mathbf{x} - \mathbf{x}_{\text{true}}\|^2 \quad (31)$$

$$= [\mathbf{S}_j^i x^j - g^i]^2 \quad (32)$$

$$= [\mathbf{S}_j^i (g^j + \epsilon^j) - g^i]^2 \quad (33)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j + \mathbf{S}_j^i \epsilon^j]^2 \quad (34)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 + 2 [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] [\mathbf{S}_j^i \epsilon^j] + [\mathbf{S}_j^i \epsilon^j]^2 \quad (35)$$

That middle term will vanish under expectation. But, before we do that, let's expand the first term, that is,

$$[(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 = [(\mathbf{S}_j^i - \mathbf{I}_j^i) (x^j - \epsilon^j)]^2 \quad (36)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j]^2 - 2 [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j] [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j] + [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j]^2 \quad (37)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j]^2 + [(\mathbf{S}_j^i - \mathbf{I}_j^i) (\epsilon^j - 2x^j)] [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j] \quad (38)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j]^2 + [(\mathbf{S}_j^i - \mathbf{I}_j^i) (-\epsilon^j - 2g^j)] [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j] \quad (39)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j]^2 - [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j]^2 - 2 [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j] \quad (40)$$

Adding this back in to the original calculation, we end up with

$$N \cdot \text{MSE}(\lambda) = [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j]^2 + 2 [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] [\mathbf{S}_j^i \epsilon^j] + [\mathbf{S}_j^i \epsilon^j]^2 \quad (41)$$

$$= [(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j]^2 - [(\mathbf{S}_j^i - \mathbf{I}_j^i) \epsilon^j]^2 + [\mathbf{S}_j^i \epsilon^j]^2 - 2 [(\mathbf{S}_j^i - \mathbf{I}_j^i) g^j] \epsilon^i \quad (42)$$

Now, under expectation, two things happen: the squared operator on random variable terms become the trace of that operator, and the random variable on deterministic term becomes zero. So, in total we have that

$$N \cdot \text{MSE}(\lambda) = \|(\mathbf{S}_j^i - \mathbf{I}_j^i) x^j\|^2 - \sigma^2 \text{Tr}(S - I)^2 + \sigma^2 \text{Tr} S^2 \quad (43)$$

$$\text{MSE}(\lambda) = \frac{1}{N} \|(\mathbf{S}_\lambda - I) \mathbf{x}\|^2 - \sigma^2 + \frac{2\sigma^2}{N} \text{Tr} \mathbf{S}_\lambda \quad (44)$$

Now let's assume that our noise looks like this,

$$p_{\text{robust}}(\epsilon) = (1 - \alpha) \cdot p_{\text{noise}}(\epsilon) + \alpha \cdot p_{\text{outlier}}(\epsilon). \quad (45)$$

where it's divided into two added components So now let's say we magically know which observations

$$(N_o + N_n) \cdot \text{MSE}(\lambda) = \|(\mathbf{S}_\lambda^n - I) \mathbf{x}\|^2 - N_n \sigma_n^2 + 2\sigma_n^2 \text{Tr} \mathbf{S}_\lambda^n \quad (46)$$

$$\|(\mathbf{S}_\lambda^o - I) \mathbf{x}\|^2 - N_o \sigma_o^2 + 2\sigma_o^2 \text{Tr} \mathbf{S}_\lambda^o \quad (47)$$