Horne (1986) Optimal Spectrum Extraction

Introduction

Horne's (1986) optimal extraction method treats each pixel of a two-dimensional spectrum as a noisy measurement of the same underlying flux at a given wavelength. Instead of simply summing all pixels in a rectangular aperture, it uses the measured spatial profile and per-pixel variances to produce a minimum-variance, unbiased estimate of the flux. The method follows directly from a maximum-likelihood or weighted least-squares approach. We derive the core equations and explain why the numerator and denominator of the extraction formula take their particular form.

1. Modelling the Data

Consider a single wavelength bin in the two-dimensional spectrogram. Along the cross-dispersion direction we have a spatial profile describing how the light falls across the detector. We assume this profile is known from collapsing or calibrating the data. We model each pixel at position x as

$$d(x) = F\,P(x) + \epsilon(x), \tag{1}$$

where F is the unknown total flux at that wavelength, $P(x)$ is the normalized spatial profile such that $\sum_x P(x) = 1$, and $\epsilon(x)$ represents random noise with variance $v(x)$. In this model each pixel contains the same underlying flux but weighted by its fraction of the profile and corrupted by noise.

2. Weighted Least Squares under Gaussian Noise

If the noise is approximately Gaussian, the log-likelihood of the data is equivalent to minimizing the weighted chi-square

$$\chi^2 = \sum_x \frac{\left[d(x) - F\,P(x)\right]^2}{v(x)}. \tag{2}$$

This expression encodes the principle that pixels with lower variance contribute more strongly to the fit, while pixels with large variance or cosmic rays contribute little.

3. Differentiating the Chi-Square

Taking the derivative of χ^2 with respect to F and setting it to zero yields

$$\frac{\partial \chi^2}{\partial F} = -2 \sum_x \frac{P(x)\left[d(x) - F\,P(x)\right]}{v(x)} = 0. \tag{3}$$

Rearranging gives

$$\sum_x \frac{P(x)\,d(x)}{v(x)} = F \sum_x \frac{P(x)^2}{v(x)}. \tag{4}$$

4. The Optimal Flux Estimate

Solving for F leads directly to the optimal flux estimate,

$$\boxed{F_{\mathrm{opt}} = \frac{\displaystyle\sum_x \frac{P(x)\,d(x)}{v(x)}}{\displaystyle\sum_x \frac{P(x)^2}{v(x)}}}. \tag{5}$$

The numerator is effectively a cross-correlation of the data with the expected spatial profile, weighted by the inverse variance of each pixel. The denominator is the total effective weight of the profile squared, also variance-weighted. This expression is the maximum-likelihood or weighted least-squares solution for one parameter, the flux.

5. Variance of the Estimate

From standard weighted least-squares theory, the variance of the optimal flux estimate is simply the reciprocal of the denominator:

$$\boxed{\mathrm{Var}[F_{\mathrm{opt}}] = \frac{1}{\displaystyle\sum_x \frac{P(x)^2}{v(x)}}}. \tag{6}$$

This variance naturally decreases when many low-noise pixels contribute and increases if only a few noisy pixels remain.

6. Beyond the Gaussian Assumption: Non-Gaussian Noise

The derivation above assumes Gaussian noise, which makes the negative log-likelihood quadratic in F and leads to a closed-form solution. In practice, detector noise can be Poissonian, heavy-tailed, or a mixture of different sources. When the noise is Poisson-dominated at low counts, the appropriate likelihood for each pixel is

$$\mathcal{L}(d|F) = \frac{\lambda(x)^{d(x)}\, e^{-\lambda(x)}}{d(x)!}, \quad \text{with} \quad \lambda(x) = F\,P(x) + b(x),$$

where $b(x)$ is the background. Maximizing the total likelihood over F no longer yields a simple ratio but requires solving

$$\frac{\partial}{\partial F} \sum_x \ln \mathcal{L}(d|F) = 0$$

numerically. For large counts, the Poisson distribution tends toward Gaussian and the standard Horne formula becomes an excellent approximation.

When noise distributions are heavy-tailed or contaminated by outliers, one can adopt robust loss functions such as absolute deviations or M-estimators. These methods continue the same philosophy — weighting pixels according to their reliability — but adaptively reduce the influence of outliers and deviate from the simple numerator/denominator expression.

The essential idea remains the same: write down a likelihood model appropriate to the noise statistics, then maximize it for F. Under Gaussian noise this produces Horne's analytic formula; under other noise models it produces a similar but generally iterative or non-linear estimator. In all cases, the goal is to assign each pixel a weight proportional to how much it can be trusted, combining spatial profile information with a realistic noise model.

7. Interpretation and Philosophy

The form of the numerator and denominator follows directly from the principle of combining multiple noisy measurements of the same quantity in the statistically optimal way. Each pixel is treated as an independent estimate of the flux, scaled by how much light is expected there and weighted by how reliable it is. Pixels near the center of the spatial profile, where $P(x)$ is large, contribute more strongly, while pixels with large variance are down-weighted automatically. In the extreme case of a cosmic ray or a flagged bad pixel, the variance becomes large and the pixel's contribution vanishes. The formula is mathematically equivalent to computing a weighted mean, but with the spatial profile incorporated.

This approach beats simple summation because real spectra are not uniform across the slit and noise is not the same in every pixel. A boxcar extraction gives equal weight to all pixels regardless of profile or noise, which is only optimal in the trivial case of a perfectly flat profile and identical noise everywhere. In reality the profile is peaked and the noise varies, so optimal extraction extracts every bit of usable information from the data. Even at high signal-to-noise, the method provides a transparent way to propagate uncertainties.

In essence, Horne's method is an application of maximum-likelihood estimation to spectrum extraction: use the known shape of the light on the detector together with the noise properties to estimate the flux that is most likely given the data. Under Gaussian noise this leads to a simple analytic expression; under non-Gaussian noise the same principle applies but the solution must be adapted to the actual noise distribution.