Optimized Spectroscopic Signal Estimates Using Integration and Matched Filters

Stephen E. Bialkowski, Utah State University

Available at: https://works.bepress.com/stephen_bialkowski/6/
Optimized Spectroscopic Signal Estimates Using Integration and Matched Filters

STEPHEN E. BIALKOWSKI
Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322-0300

This paper examines theories of signal processing as applied to peak magnitude estimation in absorption and emission spectroscopy. Signals obtained from Fourier transform, fixed wavelength, and scanning dispersive instruments are modeled in terms of the time required to obtain a spectrum. The differences between these techniques and the signal processing procedures that should be used for each technique are characterized for a Lorentzian spectral feature. Including the time required to scan over a range of optical frequencies results in optimal signal magnitudes. The differences between these techniques and the signal obtained from Fourier transform, fixed wavelength, and scanning dispersive instruments are modeled in terms of the time required to obtain the maximum theoretical SNR in the signal magnitude estimate. Using a Lorentzian absorption model, T. Hirschfeld showed that the matched filter would result in a SNR increase over that for simple peak magnitude measurements. More recently K. Sasaki et al. performed experiments which indicate that a maximum SNR is obtained when replicate measurements of the peak magnitude are taken. They argued that this result was to be expected with finite measurement times, since time spent measuring that portion of the peak with a maximum signal was more useful than that spent measuring less than optimum information.

The matched filter, born from a theoretical basis, guarantees the maximum SNR, while the intuitively appealing arguments of K. Sasaki et al. teach that replicate measurements of the peak maximum amplitude measurements must result in the maximum SNR. The question naturally arises concerning which method one should use to be assured the maximum theoretical SNR in the signal magnitude estimate. It is shown below that both methods yield the maximum theoretical SNR, but under different conditions. The important clue to the resolution of this dilemma is the availability of spectral information. T. Hirschfeld's arguments were based on spectra in Fourier transform spectroscopy (FTS). K. Sasaki et al. used a dispersive instrument. With FTS, the entire spectrum is available with no additional time cost over that for a single resolution element. This is not the case with dispersive instruments, which have been categorized as time sequential. In the case of these instruments, there is a time cost for each measurement obtained at a different optical frequency. It is shown below that scanning dispersive instruments will yield the optimum SNR when they are not scanned across the peak. This result is independent of any considerations about scan rate, resolution, etc.

BACKGROUND

Consider a spectroscopic signal that is a single line defined by the normalized Lorentzian function

$$S(v) = A(1/v)/v^2 + v^2$$

(1)

where $v$ is the optical frequency, $A$ is the peak magnitude, and $v$ is the halfwidth at half-maximum. The measurement value is a sum of signal plus zero mean noise,$$

$$X(v) = S(v) + V(v)$$

(2)

where $X(v)$ is the measurement, and $V(v)$ the noise. A generalized filter is formulated on the basis of the signal transmitted through a transfer function, $H(v)$:

$$Y(v) = H(v)X(v)$$

(3)

where $Y(v)$ is the filtered measurement. The transfer function is purposely undefined at this point. It could be that of the digital filter used in the data processing step of the peak magnitude analysis, or that due to the effect of slits on the optical transfer in a dispersive instrument. In any event, $H(v)$ is chosen so that it yields an estimate of the signal. The signal estimate is obtained by integration over $v$:

$$S = \int_{-\infty}^{\infty} Y(v) \, dv$$

(4)

where $S$ is the estimated signal magnitude. This estimate is to be optimized by the choice of $H(v)$.

The signal estimate is in turn optimized by optimization of the SNR. Signal and noise powers are first predicted separately. The signal power is obtained as the square of the signal passing through the transfer function

$$P_s = \left[ \int_{-\infty}^{\infty} H(v)S(v) \, dv \right]^2$$

(5)

where $P_s$ is the signal power, while the noise power, $P_n$, is obtained as the squared noise transfer.
\[ P_n = \int_{-\infty}^{\infty} [H(\nu) V(\nu)]^2 \, d\nu \]  

(6)

for real (vs. complex) frequency domain functions. These two power definitions differ because, while the signal is constant and thus additive, the noise is a zero mean random variable which must be accumulated as the square. Combining Eqs. 5 and 6 results in the signal-to-noise power ratio,

\[ \text{SNR}^2 = P_s/P_n, \]  

(7)

and the SNR is the square root of the signal-to-noise power ratio.

There is a subtle problem here. The noise has instantaneous values that change with time, not with optical frequency. Thus \( V(\nu) \) is not really a random variable. The time-averaged noise must be zero since, for it to be zero mean in the optical frequency domain, it must be zero mean in the time domain as well. It is possible to model this noise as a time-averaged expectation value for the purposes of SNR estimation in the optical experiment. Noise is an apparent random variable of optical frequency only because it takes a finite time to obtain the optical spectrum. Thus the random variable of time appears to be a random variable of optical frequency, but is characterized by the expectation value of the square. This time-averaged square noise is used to simplify the time-dependent problem by making it a frequency dependent one.

There are two limiting cases of the time-averaged square noise magnitude for the two noise processes: white noise and shot noise. White noise is characterized by a constant noise variance, \( \sigma^2 \), which is independent of frequency. Although all spectroscopy is ultimately shot-noise limited, in absorption spectrophotometry of optically thin samples, the signal is relatively constant across a spectrum. Thus shot noise generated by the detection process is apparently constant and characteristically white. The noise manifested in the optical spectrum is independent of \( \nu \) if the intensity is independent of \( \nu \). Another instance where the white noise model is appropriate occurs when the optical detector has a dominant thermal noise component. In contrast to the white noise process, a shot-noise process with a variable signal magnitude is characterized by the optical-frequency-dependent noise power spectrum, \( V(\nu)^2 \). This noise power spectrum is not obtained from the transform of the time-dependent noise. It is that which describes how the noise varies with optical frequency. Emission spectroscopy can be modeled as having shot-noise limited signals when quantum detectors are used. The emission itself results in noise, and the average noise power spectrum, \( V(\nu)^2 \), is equal to the signal, \( S(\nu) \).

Replicate measurements increase the SNR. The signal is not a random variable and will accumulate in direct proportion to the number of replicate measurements. Replicate measurements of the noise will tend to zero, while replicate measurements of the noise power will be proportional to the number of measurements. Thus the average power ratio is proportional to the number of measurements:

\[ \langle \text{SNR}^2 \rangle = n \times \text{SNR}^2 \]  

(8)

where \( \langle \text{SNR}^2 \rangle \) is used to indicate the accumulated \( \text{SNR}^2 \) for \( n \) replicate measurements and \( \langle \text{SNR} \rangle \) is \( \sqrt{\langle \text{SNR}^2 \rangle} \). The latter results in the familiar square root relationship between SNR and number of replicate measurements. Signal averaging by replicate measurements takes time, and so there is a time cost for this signal averaging. Only a finite number of measurements can be taken during a given analysis time. The maximum number of replicate measurements is approximated by the ratio of the measurement time, \( T_m \), to the time required to obtain the spectral information, \( T_s \). Thus Eq. 8 can be recast in terms of these times.

\[ \langle \text{SNR}^2 \rangle = \left( \frac{T_m}{T_s} \right) \times \text{SNR}^2. \]  

(9)

The time cost to obtain the optical signal integrated over a frequency interval is different for different optical instruments. The use of a sequential scanning dispersive instrument, the time cost is proportional to the integration bandwidth,

\[ T_s = k \Delta \nu \]  

(10)

where \( k \) is the proportionality constant and \( \Delta \nu \) is the band. For the case of a fixed frequency instrument where the integration bandwidth is changed by adjustments to the instrument, e.g., changes in slit widths of a monochromator, we may let \( T_s = 1 \). In fact, there is no additional time cost for increasing the optical bandwidth in this case, and the maximum SNR improvement is equal to \( \sqrt{T_m} \). For FTS, \( T_s \) is finite and not proportional to the integration band. Many time-dependent factors could be taken into account in assignment of a scan time for FTS. Nonetheless, a constant \( T_s \) is all that is required to model the FTS instrument for the purpose at hand. These simple approximations do not allow a comparison between different types of instruments, only different signal processing methods with a given instrument. Different instruments are cross compared in Ref. 4.

**ABSORPTION SPECTROSCOPY**

Absorption spectroscopy of optically thin samples can be modeled as having zero mean additive white noise only when the light intensity observed by the detector does not vary substantially. Two types of spectral filtering are considered: integration and matched filter. For the integration filter \( H(\nu) \) is unity over a range \( -\Delta \nu/2 \leq \nu \leq \Delta \nu/2 \) and zero outside that range. Thus the infinite integration limits in Eqs. 5 and 6 are replaced with finite integration limits, and \( H(\nu) \) is eliminated:

\[ P_s = \int_{-\Delta \nu/2}^{\Delta \nu/2} S(\nu) \, d\nu \]  

(11)

and

\[ P_n = \int_{-\Delta \nu/2}^{\Delta \nu/2} V(\nu)^2 \, d\nu = \Delta \nu \sigma^2. \]  

(12)

The simplest case is that of the limit where \( \Delta \nu/T \to 0 \). With zero mean white noise, this case represents a sampling of the peak maximum, and the resulting SNR power is

\[ \text{SNR}_u^2 = \Delta \nu S(0)^2/\sigma^2 \]  

(13)
which for the normalized Lorentzian yields
\[
\text{SNR}_t = 2A x \tan^{-1}(\Delta \nu/2\Gamma)/(\pi \sigma \sqrt{\Delta \nu}).
\]  
(15)
With a fixed time cost for spectral integration, the maximum is obtained by differentiating with respect to \(\Delta \nu\) and setting this result to zero. This results in the equality
\[
\Delta \nu = 2\Gamma \times \tan\left[(\Delta \nu/\Gamma)/(1 + (\Delta \nu/2\Gamma)^2)\right],
\]  
(16)
which is satisfied when \(\Delta \nu = 2.7834\Gamma\). Thus the optimum SNR is
\[
\text{SNR}_t \approx 0.36A/(\sigma \sqrt{\Gamma}).
\]  
(17)
As in the limited sample case, the SNR decreases with increasing \(\Gamma\). But here, the SNR decreases as \(1/\sqrt{\Gamma}\) versus \(1/\Gamma\). The ratio, \(\text{SNR}_t/\text{SNR}_L\), is proportional to \(\sqrt{\Gamma}/\sqrt{\Delta \nu}\). The reason for the difference is that the bandpass of the integration filter is fixed relative to \(\Gamma\). There is no reason why the limited sample filter could not satisfy the condition leading to Eq. 17. This will occur when the Lorentzian halfwidths are equal to roughly \(1/3\) the limiting resolution of the dispersion device.

Equation 17 is easily corrected for FTS and fixed-frequency instruments. For these instruments,
\[
\langle \text{SNR} \rangle = \sqrt{(T_m/T_\infty)} \times \text{SNR}. 
\]  
(18)
When the time cost is proportional to the integration bandwidth, then Eq. 18 no longer applies. Starting again from Eq. 15, the optimum \(\Delta \nu\) is obtained from
\[
\langle \text{SNR}_t \rangle = 2(2A/\sqrt{T_m/\hbar}) \tan^{-1}(\Delta \nu/2\Gamma)/(\pi \Delta \nu \sigma).
\]  
(19)
In this instance the optimum \(\Delta \nu\) occurs when
\[
\Delta \nu = 2\Gamma \times \tan\left[(\Delta \nu/2\Gamma)/(1 + (\Delta \nu/2\Gamma)^2)\right].
\]  
(20)
This condition is met only as \(\Delta \nu \to 0\). Thus, including the time cost for scanning results in a maximum SNR when the dispersion device is not scanned but is set to transmit the peak maximum. The resulting SNR is the same as that for the limited sample filter in Eq. 14.

The matched filter is found by maximizing the signal power in Eq. 5. From Schwartz’ inequality,
\[
\left[\int_{-\infty}^{\infty} H(\nu) S(\nu) \, d\nu\right]^2 
\leq \int_{-\infty}^{\infty} H(\nu)^2 \, d\nu \times \int_{-\infty}^{\infty} S(\nu)^2 \, d\nu.
\]  
(21)
Since the left side of this equation is less than or equal to the right, the maximum must occur when both sides are equal. This equality condition is met when \(H(\nu)\) is a constant multiple of \(S(\nu)\). For the Lorentzian we choose \(H(\nu)\) to be
\[
H(\nu) = \Gamma^2/(\Gamma^2 + \nu^2)
\]  
(22)
for the simple reason that the maximum throughput of a spectrometer is one. This \(H(\nu)\) results in the optimum estimate of the peak magnitude for additive white noise. Since the noise is a constant, the SNR2 for the matched filter is simple to calculate from Eqs. 5–7 and the equality in Eq. 21:
\[
\text{SNR}_M^2 = \int_{-\infty}^{\infty} S(\nu)^2 \, d\nu/\sigma^2.
\]  
(23)
The SNR for infinite integration limits is
\[
\text{SNR}_M = A/(\sigma \sqrt{2\pi \Gamma}) \approx 0.40A/(\sigma \sqrt{\Gamma}).
\]  
(24)
And the ratio of the SNR for the matched filter vs. that for integration is
\[
\text{SNR}_M/\text{SNR}_t = 0.4/0.36 \approx 1.1,
\]  
(25)
which is a factor of 7 less than that reported in Ref. 1. This error in the latter reference is probably due to the failure to see that Eq. 23 describes the power ratio when Eq. 22 is used for \(H(\nu)\). Nonetheless, the matched filter does give the optimum SNR for processing spectral data. The result in Eq. 25 is then a limiting SNR by which other, less than optimal, filters can be gauged.

The infinite integration limits required for the optimum matched filter are neither practical nor required. Integration limits encompassing 5 times the halfwidth will result in a SNR that is 99.97% of the theoretical maximum obtained with infinite limits. The analytical result for finite integration limits is
\[
\text{SNR}_M^2 = (A^2/\pi \sigma^4 \Gamma) \times [\tau/(1 + \tau^2) + \tan^{-1}(\tau)]
\]  
(26)
where \(\tau = \Delta \nu/2\Gamma\). Of course, maximizing this SNR with respect to \(\Delta \nu\) results in infinite limits. But, if there is an integration-band-dependent time cost, then,
\[
\langle \text{SNR}_M^2 \rangle = (A^2/\pi \sigma^4 \Gamma) \times [1/(1 + \tau^2) + (1/\tau)\tan^{-1}(\tau)].
\]  
(27)
For one to find the maximum SNR as a function of \(\Delta \nu\), it is sufficient to set the differential of Eq. 27 with respect to \(\tau\) equal to zero. This leads to the condition that
\[
\tau = \tan[(\tau - \tau^3)/(1 + \tau^2)^2],
\]  
(28)
which is again only satisfied if \(\tau \to 0\). Since this limit was obtained for the best possible signal processing case, i.e., the matched filter, it positively shows that the maximum SNR is obtained when spectral scanning is not utilized, thus providing theoretical evidence for the claims of K. Sasaki et al.3

**EMISSION SPECTROSCOPY**

Emission spectroscopy is shot-noise limited. The noise is correlated to the signal. In fact, in the photon-counting Poisson statistics limit, \(V(\nu)^2 = S(\nu)\). The subsequent SNR2 is easily calculated since
\[
\text{SNR}_M^2 = \int_{-\infty}^{\infty} S(\nu) \, d\nu,
\]  
(29)
which, for the Lorentzian, results in
\[
\text{SNR}_M^2 = (2A/\pi) \times \tan^{-1}(\Delta \nu/2\Gamma)
\]  
(30)
being maximized when \(\Delta \nu = \infty\), and with a maximum SNR of \(\sqrt{A}\). This is an appealing result since, in the
absence of interferences, one can do no better than to collect all of the light emitted by the analyte. But, if the time cost for integration is finite, then

\[ \text{SNR}^2 = (T_w A/k \Gamma \pi) \times \left( \tan^{-1}(\tau)/\tau \right) \]

where \( \tau = \Delta \nu / 2 \Gamma \). Maximizing with respect to \( \tau \) results in the condition

\[ \tau = \tan\left[\frac{\pi}{1 + \tau^2}\right], \]

which is satisfied only as \( \tau \to 0 \).

The matched filter does not really apply in this instance, since the provision for using it is that the noise be white. However, A. Papoulis gives an extension of this filter in which the signal is “pre-whitened” by dividing the matched \( H(\nu) \) by the noise power. Since the noise power is equal to the signal in this case, the resulting \( H(\nu) \) is a constant. That is, all the light is to be collected for a maximum SNR, in agreement with the results of Eq. 30.

**DISCUSSION**

In all cases examined in which there is a time cost proportional to the integration bandwidth, i.e., for scanning dispersive instruments, the maximum (SNR) occurs when the peak maximum is repetitively sampled. It is almost ironic that including the time cost for scanning results in a prediction that peak scanning should not be used. This result was shown to hold for the two limiting noise cases, i.e., both Gaussian and Poisson statistics, and so should hold for all types of spectroscopy.

The critical assumption in the above derivations was that the absorption or emission was Lorentzian. However, these results may hold for any peaked spectral feature. Consider an arbitrary, asymmetric peak specified only by \( S(\nu) \), and with a maximum at \( \nu = 0 \). Since the peak is asymmetric, the optimal integration limits must be found for both negative and positive \( \nu \) sides of the peak. The two different limits may be found independently. For the positive side, the optimum limit for the integration filter can be found from the differential equation,

\[ \text{SNR}_1 = \left(1/\sigma \sqrt{\Delta \nu}\right) \int_0^{\Delta \nu} S(\nu) d\nu \]

with respect to \( \Delta \nu \). Setting the derivative equal to zero results in the integral equation

\[ 2S(\Delta \nu) \Delta \nu = \int_0^{\Delta \nu} S(\nu) d\nu \]

where \( S(\Delta \nu) \) is the signal evaluated at the point \( \Delta \nu \). This is the condition that must be satisfied to obtain the optimal integration bandwidth on one side of the peak. So two such equations are required for the asymmetric peak. Similarly, if there is a time cost for scanning, then the appropriate condition is

\[ S(\Delta \nu) \Delta \nu = \int_0^{\Delta \nu} S(\nu) d\nu \]

where the time cost for scanning was taken as being \( k \Delta \nu \). The factor of two in the left-side terms of these equations is the only difference.

Now, the very concept of a peak implies that \( S(\nu) \) will be less than the maximum for all values other than \( \nu = 0 \). Subsequently, the simplest model of such a function is the un-normalized line equation,

\[ S(\nu) = A \times (1 - av) \quad av \leq 1 \]

\[ S(\nu) = 0 \quad av > 1 \]

where \( a \) is the slope in the immediate region of the peak maximum. With no time cost for scanning, Eq. 36 can be solved with the use of this signal definition. The solution results in the one-sided integration bandwidth, \( \Delta \nu = 2/(3a) \), which will in general result in a finite bandwidth for a finite slope local to the peak maximum. However, with a time cost, Eqs. 35 and 36 yield the condition \( 2\Delta \nu = \Delta \nu \). Clearly, this condition can only be satisfied if \( \Delta \nu = 0 \). Similar results can be obtained for the matched filter as well. These results indicate that the results obtained for the Lorentzian can be generalized to any peak.

Some important experimental guidelines may be gleaned from these theoretical results. The most obvious one is that scanning a dispersive instrument cannot result in optimum SNR for quantitative analysis, even when optimal signal processing filters are used. Clearly, it is a valuable qualitative method, but not a very efficient quantitative one. Even the SNR improvement gained from the optimal matched filter is not enough to overcome the cost for scanning. This guideline may be applied to other types of signal processing used in spectroscopy. For example, the mathematical procedures used in correlation are similar to the matched filter. In fact, the matched filter is the optimum correlation filter. On the basis of the fact that scanning is unfavorable for the matched filter, correlation spectroscopy based on these instruments cannot be efficient in terms of SNR improvement.

The second guideline implied by these results is that fixed-wavelength bandpass methods are optimal for absorption spectroscopy. Although the results for both fixed-wavelength and FTS cases were based on a constant \( T_w \), the FTS case is less efficient in terms of SNR improvement because of the finite measurement time. Limited time economy may be realized by using discrete Fourier transforms to calculate only those portions of the spectrum that would be used in either the integration or finite matched filter. Nonetheless, most modern FTS instruments perform interferometer scanning and numerical computations independent of one another. By allowing more replicate measurements, the savings in computation time may not result in an increased SNR.

Third, the SNR obtained from the matched filter was only 11% greater than that obtained from the optimal integration filter. Given the fact that the matched filter uses infinite integration limits and that a complicated transfer function, \( H(\nu) \), is required, this extra 11% may not seem worth the extra computation time. Further, the matched filter would be difficult to realize experimentally for the scanning dispersive instrument. For a monochromator-based system, a graded neutral-density filter would have to be used in place of the exit slit. Manufacturing this neutral-density filter would no doubt be difficult. On the other hand, the integration filter would be easily implemented on a monochromometer-based instrument.
INTRODUCTION

Hollow cathode discharges have been applied in lasers, ion thrusters for space propulsion, spectral light sources, radical and ion sources for molecular spectroscopy, and ordinary lighting devices. The physics of the processes inside such discharges has been studied in various approximations, ranging from fluid models to particle kinetic models, invoking sheath approximations and the role of photo-emission. A simplified theory which describes the radial distributions of electric field and ion density in the cathode fall region of a cylindrical hollow cathode discharge is presented. The continuity equation for ion density and Poisson's equation for space-charge controlled field are used. An analytical form is derived on the assumptions that the drift velocity of ions obeys the high-field approximation, that the rate constant of ionization is constant over the cathode fall region, and that the radial distribution of ionizing particles (electrons or photons) is in the form which obeys Beer's law as applied to cylindrical symmetry. The application of the derived formula to the analysis of observed data of Ar hollow cathode discharges is briefly discussed.

Index Headings: Hollow cathode lamps; Visible; UV Spectroscopy; Hollow cathode excitation; Optogalvanic spectroscopy.

A Theory of Electric Field and Charge Density Distributions Inside a Cylindrical Hollow Cathode

CHIAKI HIROSE* and TAKASHI MASAKI

Research Laboratory of Resources Utilization, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 227, Japan

A simplified theory which describes the radial distributions of electric field and ion density in the cathode fall region of a cylindrical hollow cathode discharge is presented. The continuity equation for ion density and Poisson's equation for space-charge controlled field are used. An analytical form is derived on the assumptions that the drift velocity of ions obeys the high-field approximation, that the rate constant of ionization is constant over the cathode fall region, and that the radial distribution of ionizing particles (electrons or photons) is in the form which obeys Beer's law as applied to cylindrical symmetry. The application of the derived formula to the analysis of observed data of Ar hollow cathode discharges is briefly discussed.

Index Headings: Hollow cathode lamps; Visible; UV Spectroscopy; Hollow cathode excitation; Optogalvanic spectroscopy.

Received 22 December 1987.
* Author to whom correspondence should be sent.

A number of investigations of hollow cathode discharges have been carried out in the past, but most of the measurements have been performed on a simulated configuration with plane-parallel cathodes placed face-to-face, and the accuracy was not very high in comparison with that attainable by the use of lasers. Various theories which are related to the distribution of the electric field in the cathode fall region of planar symmetric glow discharges have been developed, but it turns out that a simple translation into a cylindrical coordinate system does not lead to practical solutions. The geometry and symmetry of the cathode fall region of a cylindrical hollow cathode are not the same as those used in glow discharges, and a different distribution of ions is expected since the cross sections for ion and electron currents do not stay constant in the direction of the discharge. As a matter of fact, the experimental result for the Kr discharges revealed that the electric field decreases in a manner very close to being linear, according to the distance from cathode toward the negative glow. This cannot be the case for the field expected from a uniform ion distribution; instead, the ion density is proportional to the distance from cathode toward the negative glow. Consequently, a new optical instrument was developed which allows the detection of the Stark shift of atomic lines by optogalvanic spectroscopy (OGS) in Kr and Ar hollow cathode discharges. The same technique was used by Doughty et al. to measure the cathode fall region of Ne glow discharge, and they showed that the electric field in the cathode fall region decreases linearly according to the distance from the planar cathode surface. Their result confirms the historic results and conclusions by Little and von Engel, which show that the ion density in the cathode fall region is uniformly distributed in the discharges with plane-parallel electrodes.

Acknowledgment

The author is pleased to acknowledge support by CHE-8520050 awarded by the National Science Foundation.