An interpolation method for cross-correlation and cross-spectral density estimation for two-channel laser Doppler anemometry

Holger Nobach

April 25, 2016

Abstract

The interpolation method for calculating the cross-correlation function and the cross-spectra from two-channel data is revisited and extended by recently developed processing steps.

1 Introduction

For calculating the correlation function or the power spectral density from randomly sampled data from laser Doppler velocity measurements, estimation procedures are required, which consider the specific characteristics of LDV data, namely the sampling of the flow velocity at random arrival times, the data noise and the correlation of the sampling rate and the instantaneous velocity. Much effort has been put onto autocorrelation and auto-spectral estimators following thee different estimator classes, slot correlation, estimating a correlation function (correlogram) from the data [3, 6, 29, 30, 10, 11, 15, 17, 21, 23, 24, 25, 26, 28], direct spectral estimators, estimating a spectrum (periodogram) directly from the randomly sampled data [3, 4, 7, 8, 9, 18, 19, 31] and interpolation methods of the randomly sampled LDV data set yielding a continuous velocity over time, which then is re-sampled equidistantly with a given sampling frequency and processed by usual signal processing tools for equidistantly sampled data, including corrections of systematic errors [2, 12, 22, 27] and noise removal [20, 22].

Much less details are known about respective estimation procedures for two-channel data yielding the cross-correlation function and the cross-spectral density. So far detailed investigations exist about the following algorithms and applications.

- In [13] the possibility to use the slotting technique for the estimation of the cross-correlation function and the cross-spectrum is mentioned. There, no weighting mechanism has been realized, no local normalization, no fuzzy slotting, and no investigation has been made about independent and dependent measurements between the channels. For autocorrelation, weighting schemes have been implemented [3], including the forward-backward

inter-arrival-time weighting if transit times for individual weighting are not available [15, 17], local normalization and fuzzy slotting [29, 30, 21, 28] as well as Bessel's correction, if the data sets or data blocks are short and systematic errors due to the under-estimation of the block variance occur if the empirical block mean value is removed from the data blocks [18, 19].

- The application of the interpolation method to LDV cross-correlation and cross-spectra estimation has been investigated in [13] and [5]. Unfortunately, in both publications only special cases of two-channel measurements have been studied, namely either strictly independent or strictly coincident measurements in [5] or a mixture of only these two cases of measurements in [13]. The possibility of having a certain time delay of dependent measurements between the measurement channels has been mentioned in [13]. However, the there given procedures are valid only for a mixture of independent measurements and coincident dependent measurements between the channels, which is the case only if the respective measurement volumes of the two channels overlap. The method inherent weighting by holding the values longer if the data rate is lower can reduce the statistical bias due to the correlation between the instantaneous data rate and the velocity. At least at high data rates the suppression works efficient. Other, individual weighting schemes have not been realized for the interpolation method yet. Neither local normalization nor fuzzy slotting, originally developed for the slot correlation, have been adapted to the interpolation method so far. Bessel's correction, to suppress systematic errors due to the under-estimation of the velocity variances and velocity co-variance for short data sets or data blocks was not available at that time.

- The direct estimation has been used for the estimation of autocorrelation and auto-spectra only, including individual weighting [4, 8, 9, 31] or forward-backward inter-arrival-time weighting and Bessel's correction [18], local normalization and fuzzy time quantization [19]. Cross-correlation or cross-spectra have not been calculated with the direct estimation procedure so far.

- The direct estimation has also been used with quantized arrival times [4]. Quantized arrival times yield a quasi-equidistant data set with gaps with no data between the original samples. Filling these gaps with zeros yields an equidistant data set, which can be processed with common signal processing tools. This way either the correlation function can be calculated directly or the spectrum utilizing the fast Fourier transform. Both, the correlation function and the spectrum then are related through the Wiener-Khinchin theorem. This way the calculations for the direct estimation can be accelerated significantly. Since the time quantization changes the results obtained, this method is counted as a fourth estimation type. It has not been used previously for the calculation of the cross-correlation or the cross-spectrum.
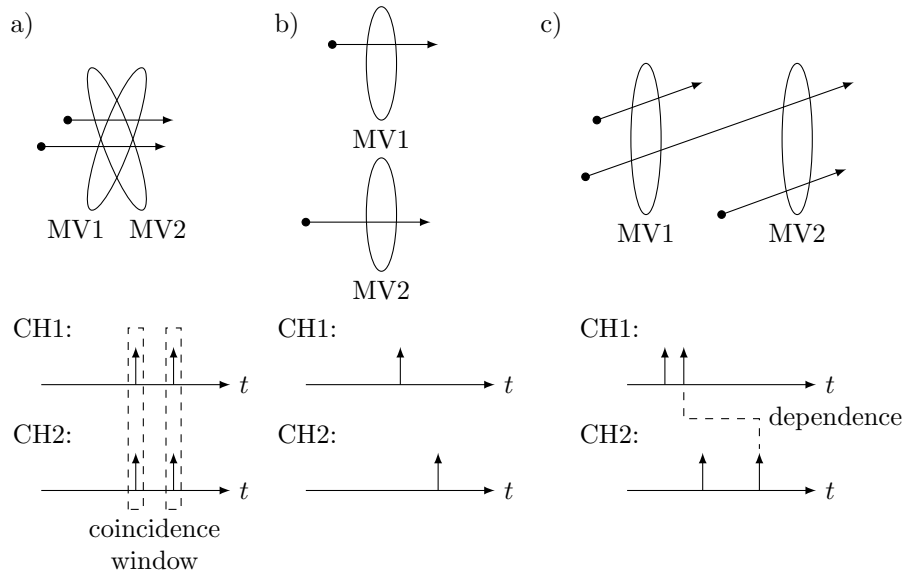
Figure 1: Fundamental sampling cases of two-channel laser Doppler data (CH1 and CH2) depending on the arrangement of the two measurement volumes MV1 and MV2: a) coincident measurements; b) independent measurements; c) mixed independent and dependent measurements with variation of the delay time between the two channels

Unfortunately, the adaption of autocorrelation and auto-spectrum estimators to the two-channel case for randomly sampled LDV data is not as straight forward as for equidistantly sampled data. While for equidistant sampling, only one of two identical data sets in the autocorrelation/auto-spectrum calculation is replaced by a second data set, besides the always present irregular sampling and the correlation between the velocity and the data rate, additionally dependent and independent samples in the two channels must be considered [13] in the LDV case. Therefore, a detailed view into adequate estimation procedures is necessary for the three classes of estimation procedures given above.

2   Dependent and independent measurements

Two-channel data from multi-component or multi-point LDV systems can produce different sampling cases depending on the configuration of the system. These sampling characteristics may lead to different systematic errors. Therefore, the following fundamental cases must be considered (Fig. 1).

- Coincident measurements: e.g. from two-component arrangements. The two channels are sampled together. This scheme yields a data set with

identical sampling times $t_{1,i} = t_{2,i} = t_i$ and with identical number of samples $N_1 = N_2 = N$. This is most similar to the autocorrelation case. Estimation routines, errors and corrections are similar.

- Independent measurements: e.g. from transversal two-point measurements. The sampling of the two channels (number of samples and sampling times) is completely independent. This scheme yields a data set with no dependence between the two channels. Both, the number of samples and the sampling times of the two channels are independent of each other. The errors and corrections are different from the coincidence case.

- Mixed measurements: e.g. from two-component arrangements in free-running mode or from longitudinal two-point arrangements. There are both, independent measurements, from particles that are measured by only one of the two laser Doppler systems, and N_{dep} dependent measurements, from particles that cross both measurement volumes. In the latter case, the respective dependent samples $u_{\mathrm{d},1,i} = u(t_{\mathrm{d},1,i})$ and $u_{\mathrm{d},2,i} = u(t_{\mathrm{d},2,i})$ at arrival times $t_{\mathrm{d},1,i}$ and $t_{\mathrm{d},2,i}$ respectively with $i = 0 \ldots N_{\mathrm{dep}} - 1$ are subsets of the measured data $u_{1,i} = u(t_{1,i}), i = 0 \ldots N_1 - 1$ and $u_{2,j} = u(t_{2,j}), j = 0 \ldots N_2 - 1$. The samples of the two channels are time delayed $(t_{\mathrm{d},1,i} \neq t_{\mathrm{d},2,i})$, where the delay t_{d} may vary with the instantaneous velocity.

3   The data sets

In the most general case, the data sets are assumed to be a mixture of both, independent measurements on the two channels as well as dependent measurements, which yield samples in both channels, where the time delay between the accurence of dependent measurements in the two channels varies for different measurements depending on the instantaneous velocity.

In this, general case, two sets of irregularly sampled velocity data $u_{1,i} = u_1(t_{1,i})$ and $u_{2,j} = u_2(t_{2,j})$ at sampling times $t_{1,i}, i = 0 \ldots N_1 - 1$ and $t_{2,j}, j = 0 \ldots N_2 - 1$ are assumed. Individual weight, e.g. the particle's transit times, cannot be considered by the present interpolation method.

Two other, special cases, are considered, namely the coincident measurements, where the two channels share a common sampling, and the independent measurements, where the sampling of the two channels is completely independent, and no dependent measurements occur. In the case of completely independent measurement, the notation stays unchanged, while the following correction procedure in this case can be simplified. In the case of only coincident measurements, the two data sets have identical sampling, leading to the velocity data $u_{1,i} = u_1(t_i)$ and $u_{2,i} = u_2(t_i)$ at sampling times $t_i, i = 0 \ldots N - 1$.
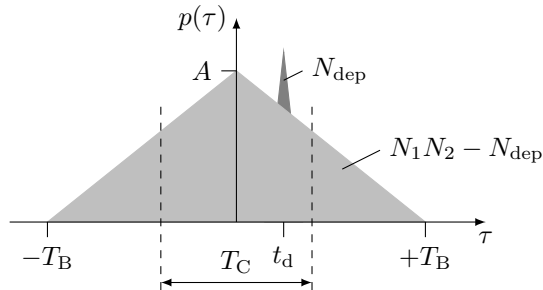
Figure 2: Dependent and independent measurements

4 Determination of the rates of dependent and independent measurements

Some of the following procedures of correlation and spectrum estimators require rates of dependent and independent measurements of the two channels and/or the preffered time delay between the two channels. A similar requirement has been investigated in [13]. The procedure given here follows the same principals, however, the parameter identification has been modified, allowing also dependent measurements with a preferred time delay $t_d \neq 0$.

Assuming a number N_1 of measurements in the first channel and N_2 in the second channel, and a number of N_{dep} dependent measurements, which occur in both channels, then the number of independent measurements in channel 1 is $N_1 - N_{dep}$ and in channel 2 it is $N_2 - N_{dep}$. Counting the number of cross-products one obtains a total number of $N_1 N_2$ cross-products, where the N_{dep} dependent measurements concentrate around t_d (Fig. 2) leaving $N_1 N_2 - N_{dep}$ cross-products with a random triangular distribution $p(\tau)$ between $-T_B$ and $+T_B$ following the model distribution

$$ p(\tau) = A \left(1 - \frac{|\tau|}{T_B} \right). $$

Defining a total length of the final correlation function $T_C < 2T_B$ (between $-T_C/2$ and $+T_C/2$), which includes all dependent measurements, one can expect to have only independent cross-products outside this interval. Note that T_C is usually chosen much smaller than the total length of the data set, if a subdivision of the data set into blocks is made, the block length should still be significantly larger then the chosen length of the correlation function. By counting the number N_C of cross-products falling into the interval T_C, assuming that the data sets of the two channels have the same duration T, the remaining $N_1 N_2 - N_C$ cross-products distribute following the triangular shape between $-T_B$ and $-T_C/2$ and between $+T_C/2$ and $+T_B$. This yields an expected height of the triangular

5

shape of independent measurements of

$$p\left(\pm\frac{T_{\mathrm{C}}}{2}\right) = \frac{N_1 N_2 - N_{\mathrm{C}}}{T_{\mathrm{B}} - \frac{T_{\mathrm{C}}}{2}}$$

at the time delay of $\pm T_{\mathrm{C}}/2$, which then leads to the amplitude A of the triangular shape of independent measurements

$$A = \frac{T_{\mathrm{B}}\left(N_1 N_2 - N_{\mathrm{C}}\right)}{\left(T_{\mathrm{B}} - \frac{T_{\mathrm{C}}}{2}\right)^2}.$$

All deviations from the triangular model distribution of independent measurements are interpreted as dependent measurements by the following processing methods.

The advantage of the method is that it is independent of the parameters of the following estimation of the correlation function and the spectrum, except for the time T_{C} which is re-used by all following methods as the total length of the estimated correlation function.

The determination of the rates of dependent measurements with different delay times follows as a results of the various interpolated data series as part of the following interpolation steps.

5   Interpolation Method

The interpolation method principally follows the procedure in [13], except for a few extensions/modifications, which are

1. Dependent measurements with a prefered time delay between the measurement channels larger than zero are allowed, which is expected in the case of real two-point measurements, including a possible distribution of the time delay of dependent measurements between the channels due to variations of the convection velocity.

2. An additional weighting factor is added, which can be set zero, if a large gap between measurements occurs, as has been seen in experimental data. The removal of data gaps by weighting is similar to the procedure introduced below for the inter-arrival time or forwar-backward inter-arrival time weighting for the other estimation procedures. Note that this particular weighting factor for the interpolation method cannot be used as freely as for the other two estimation methods, only factors of one or zero are allowed here.

3. An explicite formulation for the sample-and-hold interpolation is used for the correction of the low-pass filter due to the interpolation considering a mixture of dependent and independent pairs of measurements.

6

4. Bessel's correction of the correlation estimate is added, which suppresses systematic deviations due the under-estimation of the velocity variance for short data sets, if the mean is estimated and removed from the data sets following [18].

Since these little modifications influence the entire estimation procedure, it is summarized here including the appropriate changes.

5.1 Data pre-processing

The available data may be subdivided into blocks of a certain time duration T_B or the data may be obtained in blocks of a given record length. Due to the combination of Bessel's correction and the temporal limitation of the correlation function, both given below, the block duration can be chosen very flexible (compare [18]). It should be larger than the expected correlation interval of the flow and can be as large as the full data set. Since for the interpolation method, the computational costs increase with the square of the block length, too large a block duration will be computational costly.

The two assumed data sets $u_{1,i} = u_1(t_{1,i})$ and $u_{2,j} = u_2(t_{2,j})$ of the block duration T_B are interpolated using the sample-and-hold interpolation and re-sampled equidistantly with the frequency $F = {}^1\!/\!\Delta\tau$, which defines the fundamental frequency of all derived statistical functions hereof.

To avoid the wrap-around error of the derived statistical functions, the interpolation is done for the duration of $2T_\mathrm{B}$, where only the duration $1T_\mathrm{B}$ gets measured data, where the above mentioned weighting factor is set to one. For the other duration of $1T_\mathrm{B}$ the above weighting factor is set to zero to identify the interpolated and re-sampled data as invalid or unknown. This is the pendent to zero padding of equidistantly sampled data for the case of randomly sampled data. An important detail to avoid systematic deviations is to interpolate the valid data for exactly the duration of $1T_\mathrm{B}$. For this purpose, the arrival time of the first data point $t_{1,0}$ for the first channel and $t_{2,0}$ for the second channel are translated to $T_\mathrm{B} + t_{1,0}$ and $T_\mathrm{B} + t_{2,0}$ respectively and the values u_{1,N_1-1} and u_{2,N_2-1} of the last samples in the respective data records are hold between their occurence at t_{1,N_1-1} and t_{2,N_2-1} until these points in time. This yields two interpolated data sets of exactly the time duration of $1T_\mathrm{B}$, however if $t_{1,0} \neq t_{2,0}$, the two interpolated data sets may have a time shift, which is acceptable and correct for the following processing steps.

Formally, the interpolation looks like

$$
u'_{1,i} = u'_1(t_i) = u'_1(i\Delta\tau) = u_{1,k} \left\{ \begin{array}{l} \forall i : t_{1,k} \leq i\Delta\tau < t_{1,k+1} \text{ for } k = 0\ldots N_1 - 2 \\ \forall i : t_{1,N_1-1} \leq i\Delta\tau < T_\mathrm{B} + t_{1,0} \text{ for } k = N_1 - 1 \end{array} \right.
$$

$$
u'_{2,i} = u'_2(t_i) = u'_2(i\Delta\tau) = u_{2,k} \left\{ \begin{array}{l} \forall i : t_{2,k} \leq i\Delta\tau < t_{2,k+1} \text{ for } k = 0\ldots N_2 - 2 \\ \forall i : t_{2,N_2-1} \leq i\Delta\tau < T_\mathrm{B} + t_{2,0} \text{ for } k = N_2 - 1 \end{array} \right.
$$

with $\Delta\tau = {}^1\!/\!F$. Outside the interval $t_{1,0} \leq i\Delta\tau < T_\mathrm{B} + t_{1,0}$ or $t_{2,0} \leq i\Delta\tau < T_\mathrm{B} + t_{2,0}$ respectively, all values are zero. With experimentally obtained data,

gaps in the data stream have been identified, which significantly affect the derived statistical functions. Therefore, interpolated weighting factors are defined similar to the interpolated velocities as

$$w'_{1,i} \quad = \quad w'_1(t_i) = w'_1(i\Delta\tau) = w_{1,k} \begin{cases} \forall i : t_{1,k} \le i\Delta\tau < t_{1,k+1} \text{ for } k = 0 \dots N_1 - 2 \\ \forall i : t_{1,N_1-1} \le i\Delta\tau < T_{\mathrm{B}} + t_{1,0} \text{ for } k = N_1 - 1 \end{cases}$$

$$w'_{2,i} \quad = \quad w'_2(t_i) = w'_2(i\Delta\tau) = w_{2,k} \begin{cases} \forall i : t_{2,k} \le i\Delta\tau < t_{2,k+1} \text{ for } k = 0 \dots N_2 - 2 \\ \forall i : t_{2,N_2-1} \le i\Delta\tau < T_{\mathrm{B}} + t_{2,0} \text{ for } k = N_2 - 1 \end{cases}$$

For the interpolation method, the weights $w_{1,k}$ and $w_{2,k}$ are usually set to one. However, these weights can be used to suppress the gaps in the data stream by setting the weights to zero, if the inter-arrival time between two samples exceeds a certain limit. Good experience has been obtained with a maximum value of five times the mean inter-arrival time. Due to this one looses about 0.7% of useful data, while the outliers of large inter-arrival times are suppressed effectively. The weights then read

$$w_{1,k} \quad = \quad \begin{cases} \begin{cases} 1 & \text{for } t_{1,k+1} - t_{1,k} < 5n_1 \\ 0 & \text{otherwise} \end{cases} & \text{for } k = 0 \dots N_1 - 2 \\ \begin{cases} 1 & \text{for } T_{\mathrm{B}} + t_{1,0} - t_{1,N_1-1} < 5n_1 \\ 0 & \text{otherwise} \end{cases} & \text{for } k = N_1 - 1 \end{cases}$$

$$w_{2,k} \quad = \quad \begin{cases} \begin{cases} 1 & \text{for } t_{2,k+1} - t_{2,k} < 5n_2 \\ 0 & \text{otherwise} \end{cases} & \text{for } k = 0 \dots N_2 - 2 \\ \begin{cases} 1 & \text{for } T_{\mathrm{B}} + t_{2,0} - t_{2,N_2-1} < 5n_2 \\ 0 & \text{otherwise} \end{cases} & \text{for } k = N_2 - 1 \end{cases}$$

with the mean data rates

$$n_1 \quad = \quad \frac{N_1}{T_{\mathrm{B}}}$$

$$n_2 \quad = \quad \frac{N_2}{T_{\mathrm{B}}}$$

Other individual weighting, e.g. transit-time weighting, has not been realized for the interpolation method yet. Since the interpolation holds the values between samples, inherently a kind of an inter-arrival time weighting is realized. The introduction of another weighting scheme then would over-weight the samples. However, the weights introduced here, for the interpolation method, cannot be used as freely as for the other processing methods, while the use of the weighting values to suppress parts of the data stream is possible.

A third kind of interpolated signals is derived counting the numbers of measurements of the two channes between the time instances of the re-sampled data streams, $c'_{1,i} = c'_1(t_i) = c'_1(i\Delta\tau)$, the number of measurements in channel 1 between $t_{i-1} = (i-1)\Delta\tau$ and $t_i = i\Delta\tau$ and $c'_{2,i} = c'_2(t_i) = c'_2(i\Delta\tau)$, the number measurements in channel 2 in the same interval.

From the interpolated and equidistantly re-sampled data series $u'_{1,i}$ and $u'_{2,i}, i = 0 \dots J$ with $J = \lceil 2T_{\mathrm{B}}F \rceil$ (Note that J must be chosen odd for the

8

unsymmetric cross-correlation functions.) and the appropriate interpolated weigths $w'_{1,i}$ and $w'_{2,i}$ one can calculate the block mean values as

$$\bar{u}_1 \quad = \quad \frac{\displaystyle\sum_{i=0}^{J-1} w'_{1,i} u'_{1,i}}{\displaystyle\sum_{i=0}^{J-1} w'_{1,i}}$$

$$\bar{u}_2 \quad = \quad \frac{\displaystyle\sum_{i=0}^{J-1} w'_{2,i} u'_{2,i}}{\displaystyle\sum_{i=0}^{J-1} w'_{2,i}}$$

and remove the mean from the interpolated data to generate mean free data blocks for the following calculations of the cross-correlation function and the appropriate power spectral density.

5.2 Estimation of the initial correlation functions

From the interpolated data one can obtain the cross-correlation functions of the weighted velocity, that of the weights and that of the measurement counts either directly as

$$R'_{u,12}(\tau_k) \quad = \quad \frac{1}{T_\mathrm{B} F} \sum_{i=0}^{J-1} \sum_{j=0}^{J-1} w'_{1,i} u'_{1,i} w'_{2,j} u'_{2,j}$$

$$R'_{w,12}(\tau_k) \quad = \quad \frac{1}{T_\mathrm{B} F} \sum_{i=0}^{J-1} \sum_{j=0}^{J-1} w'_{1,i} w'_{2,j}$$

$$R'_{c,12}(\tau_k) \quad = \quad \frac{1}{T_\mathrm{B} F} \sum_{i=0}^{J-1} \sum_{j=0}^{J-1} c'_{1,i} c'_{2,j}$$

or, with less computational costs, via the spectrum using the discrete Fouier transform (DFT)

$$U_1'(f_j) \;=\; \mathrm{DFT}\left\{w_{1,i}' u_{1,i}'\right\} = \sum_{i=0}^{J-1} w_{1,i}' u_{1,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

$$U_2'(f_j) \;=\; \mathrm{DFT}\left\{w_{2,i}' u_{2,i}'\right\} = \sum_{i=0}^{J-1} w_{2,i}' u_{2,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

$$W_1'(f_j) \;=\; \mathrm{DFT}\left\{w_{1,i}'\right\} = \sum_{i=0}^{J-1} w_{1,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

$$W_2'(f_j) \;=\; \mathrm{DFT}\left\{w_{2,i}'\right\} = \sum_{i=0}^{J-1} w_{2,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

$$C_1'(f_j) \;=\; \mathrm{DFT}\left\{c_{1,i}'\right\} = \sum_{i=0}^{J-1} c_{1,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

$$C_2'(f_j) \;=\; \mathrm{DFT}\left\{c_{2,i}'\right\} = \sum_{i=0}^{J-1} c_{2,i}' \mathrm{e}^{-2\pi\mathrm{i}f_j i \Delta\tau}$$

with the imaginary unit i yielding the complex Fourier tranforms $U_1'(f_j)$, $U_2'(f_j)$, $W_1'(f_j)$, $W_2'(f_j)$, $C_1'(f_j)$ and $C_2'(f_j)$, $f_j = j/2T_\mathrm{B}, j = 0 \ldots J-1$. The cross-energy spectra of the interpolated signals then are

$$E_{u,12}'(f_j) \;=\; \frac{1}{F^2} U'^*{}_1(f_j) U_2'(f_j)$$

$$E_{w,12}'(f_j) \;=\; \frac{1}{F^2} W'^*{}_1(f_j) W_2'(f_j)$$

$$E_{c,12}'(f_j) \;=\; \frac{1}{F^2} C'^*{}_1(f_j) C_2'(f_j)$$

with the conjugate complex * and the cross-correlation functions of the interpolated signals can be derived using the inverse DFT (IDFT)

$$R_{u,12}'(\tau_k) \;=\; \frac{F}{T_\mathrm{B}}\mathrm{IDFT}\left\{E_{u,12}'(f_j)\right\} = \frac{1}{2T_\mathrm{B}^2} \sum_{j=0}^{J-1} E_{u,12}'(f_j) \mathrm{e}^{2\pi\mathrm{i}\tau_k f_j}$$

$$R_{w,12}'(\tau_k) \;=\; \frac{F}{T_\mathrm{B}}\mathrm{IDFT}\left\{E_{w,12}'(f_j)\right\} = \frac{1}{2T_\mathrm{B}^2} \sum_{j=0}^{J-1} E_{w,12}'(f_j) \mathrm{e}^{2\pi\mathrm{i}\tau_k f_j}$$

$$R_{c,12}'(\tau_k) \;=\; \frac{F}{T_\mathrm{B}}\mathrm{IDFT}\left\{E_{c,12}'(f_j)\right\} = \frac{1}{2T_\mathrm{B}^2} \sum_{j=0}^{J-1} E_{c,12}'(f_j) \mathrm{e}^{2\pi\mathrm{i}\tau_k f_j}$$

At this point the cross-correlation functions are $2T_\mathrm{B}$ long. The maximum time lag of the correlation function is typically chosen much smaller than the duration of the measurement. This reduces the estimation variance of the final
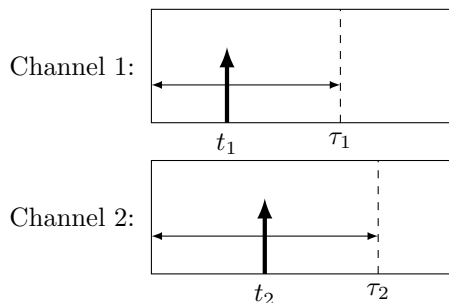
Figure 3: Fundamental sequence of particle arrivals and resampling in two LDV channels: t_1 and t_2 particle arrival times; τ_1 and τ_2 resample times

spectral estimate [16, 18]. With a given temporal resolution of the correlation function of $\Delta\tau$ the length of the correlation functions can be reduced by choosing a number of samples K with $K\Delta\tau < 2T_{\mathrm{B}}$. By rearranging the values obtained by the IDFT, the correlation function can be estimated for $\tau_k = k\Delta\tau, k = -\lfloor K/2 \rfloor \ldots \lfloor (K-1)/2 \rfloor$.

5.3 Correction of the interpolation filter

At the same time the low-pass filter of the interpolation can be corrected following the procedure given in [13], however, here considering also dependent measurements with a preferred time delay different than zero and also a distribution of dependent measurements over various time lags. If only independent measurements occur, the correlation function can be corrected using

$$
\begin{aligned}
R''_{u,12}(\tau_k) &= [1 + a + b]R'_{u,12}(\tau_k) - aR'_{u,12}(\tau_{k+1}) - bR'_{u,12}(\tau_{k-1}) \\
R''_{w,12}(\tau_k) &= [1 + a + b]R'_{w,12}(\tau_k) - aR'_{w,12}(\tau_{k+1}) - bR'_{w,12}(\tau_{k-1})
\end{aligned}
$$

with the two constants

$$
\begin{aligned}
a &= \frac{\mathrm{e}^{-n_1\Delta\tau}}{(1 - \mathrm{e}^{-n_1\Delta\tau})(1 - \mathrm{e}^{-n_2\Delta\tau})} \\
b &= \frac{\mathrm{e}^{-n_2\Delta\tau}}{(1 - \mathrm{e}^{-n_1\Delta\tau})(1 - \mathrm{e}^{-n_2\Delta\tau})}
\end{aligned}
$$

for all τ_k, with the mean data rates n_1 and n_2 as given above.

If only dependent measurements with the preferred delay time t_{d} between the channels occur, the estimated correlation value $R'_{u,12}(\tau_k)$ at $\tau_k = t_{\mathrm{d}}$ is correct and needs no further correction, while all other time lags need the above correction with exactly the same formulae as for independent measurements, except that n_1 and n_2 are replaced by the common mean data rate of the dependent measurements.

11

For a mixture of dependent and independent measurements and a distribution of dependent measurements over various τ_k an inverse filter is required, which consideres the distribution of dependent pairs of samples among all pairs of samples. The derivation and solution follows the procedure given in [13]. There, the mapping between the true correlation function and the expected correlation estimates after interpolation and resampling is given by a matrix, derived numerically from probability densities of various combinations of independent and dependent samples occuring in the data channels. In contrast to that derivation, here time delays of dependent measurements between the two channels are considered including variations of these time delays. Because of this, a proof by exhaustion by discrimination of time intervals like in [13] is not possible here. Instead, here only one case is considered, where measurements occure at t_1 in channel 1 prior to the resampling time τ_1 and at t_2 in channel 2 prior to the resampling time τ_2 and no further measurements in channel 1 between t_1 and τ_1 and no further measurements in channel 2 between t_2 and τ_2 (Fig. 3). The measurement times t_1 and t_2 may vary between $-\infty$ and τ_1 and τ_2 respectively. For all combinations of t_1, t_2 and $\tau_k = \tau_2 - \tau_1$, the probabilities of having measurements in channel 1 at t_1 and in channel 2 at t_2 must be derived, considering the data rates of independent measurements $n_{1,\text{indep}}$ in channel 1 and $n_{2,\text{indep}}$ in channel 2 and dependent measurements $n_{\text{dep}}(\tau_k)$ as a function of the various delay times. This gives the data rate of the two channels as the sum of all possible independent and dependent measurements as

$$
\begin{aligned}
n_1 &= n_{1,\text{indep}} + \sum_{k=-\infty}^{\infty} n_{\text{dep}}(\tau_k) \\
n_2 &= n_{2,\text{indep}} + \sum_{k=-\infty}^{\infty} n_{\text{dep}}(\tau_k)
\end{aligned}
$$

Based on the numbers estimated in section 4 and 5.1, the rates of dependent measurements as a function of the lag time at $\tau_k = k\Delta\tau, k = -\lfloor K/2 \rfloor \ldots \lfloor (K-1)/2 \rfloor$ are obtained as

$$
n_{\text{dep}}(\tau_k) = \frac{\frac{J}{2} R'_{c,12}(\tau_k) - A\Delta\tau \left(1 - \frac{|\tau_k|}{T_B} \right)}{T_B}
$$

This estimate is a crucial step for the following filter correction. Unfortunately, It has been found to be not sufficiently reliable. Therefore, the search for alternative methods could be rewarding.

The rates of independent measurements can be obtained as

$$
\begin{aligned}
n_{1,\text{indep}} &= \frac{N_1 - N_{\text{dep}}}{T_B} \\
n_{2,\text{indep}} &= \frac{N_2 - N_{\text{dep}}}{T_B}
\end{aligned}
$$

12

or, alternatively as

$$
n_{1,\text{indep}} = n_1 - \sum_{k=-\lfloor K/2\rfloor}^{\lfloor (K-1)/2\rfloor} n_{\text{dep}}(\tau_k)
$$

$$
n_{2,\text{indep}} = n_2 - \sum_{k=-\lfloor K/2\rfloor}^{\lfloor (K-1)/2\rfloor} n_{\text{dep}}(\tau_k)
$$

assuming that the delay of all dependent measurements are within $-T_{\text{C}}/2$ and $+T_{\text{C}}/2$. However, the rates of independent measurements are not required for the following procedure.

Considering the resampling times τ_1 and τ_2, at these times the estimated correlation after the interpolation is the real correlation between the original samples at t_1 and t_2, provided that there are appropriate samples at t_1 and t_2 and there are no further samples between t_1 and τ_1 and between t_2 and τ_2. Assuming a random sampling of the two channels with the data rates $n_{1,\text{indep}}$ and $n_{2,\text{indep}}$ of the independent measurments and $n_{\text{dep}}(\tau_k)$ for all possible dependent measurements, this yields an expected correlation estimate of

$$
\hat{R}_{12}(\tau_k) = \sum_{i=-\infty}^{0} \sum_{j=-\infty}^{k} P_1(i,j,k) P_0(i,j,k) R_{12}(\tau_{j-i})
$$

as a function of the true correlation function $R_{12}(\tau_k)$ and the respective data rates.

$P_1(i,j,k)$ is the probability of a measurement in an interval $\Delta\tau$ before $t_1 = i\Delta\tau$ and a measurement in an interval $\Delta\tau$ before $t_2 = j\Delta\tau$. It includes all combinations of independent and dependent measurements in channels 1 and 2 and reads

$$
P_1(i,j,k) = P_{1a}(i,j,k) P_{1b}(i,j,k) + P_{1c}(i,j,k) - P_{1a}(i,j,k) P_{1b}(i,j,k) P_{1c}(i,j,k)
$$

where $P_{1a}(i,j,k)$ is the probability of a measurement in an interval $\Delta\tau$ before $t_1 = i\Delta\tau$. It includes independent measurements in channel 1 and all dependent measurements with the counterpart in channel 2 outside the interval $t_2 : \tau_k$ and reads as

$$
P_{1a}(i,j,k) = 1 - e^{-n_{1,\text{indep}}\Delta\tau} \prod_{l=-\infty}^{j-i-1} e^{-n_{\text{dep}}(\tau_l)\Delta\tau} \prod_{l=k-i+1}^{\infty} e^{-n_{\text{dep}}(\tau_l)\Delta\tau}
$$

$P_{1b}(i,j,k)$ is the probability of a measurement in an interval $\Delta\tau$ before $t_2 = j\Delta\tau$. It includes independent measurements in channel 2 and all dependent measurements with the counterpart in channel 1 outside the interval $t_1 : 0$ and reads as

$$
P_{1b}(i,j,k) = 1 - e^{-n_{2,\text{indep}}\Delta\tau} \prod_{l=-\infty}^{i+j-1} e^{-n_{\text{dep}}(\tau_l)\Delta\tau} \prod_{l=j+1}^{\infty} e^{-n_{\text{dep}}(\tau_l)\Delta\tau}
$$

13

Note that τ_1 is chosen to be zero without restriction of generality.

$P_{1c}(i, j, k)$ is the probability of a common dependent measurement in channel 1 and channel 2 in an interval $\Delta\tau$ before t_1 and t_2. It reads as

$$P_{1c}(i, j, k) = 1 - \mathrm{e}^{-n_\mathrm{dep}(\tau_{j-i})\Delta\tau}$$

$P_0(i, j, k)$ is the probability that no further measurements occure in channel 1 between t_1 and 0 and in channel 2 between t_2 and τ_k. It includes all independent measurements in channels 1 and 2, all dependent measurements in channels 1 and 2, where the respective counterparts in the other channel lay outside the respective interval $t_2 : \tau_k$ or $t_1 : 0$ and all dependent measurements where the couterparts in the other channel lay within the respective intervals. The probability reads as

$$P_0(i, j, k) = P_{0a}(i, j, k)P_{0b}(i, j, k)P_{0c}(i, j, k)$$

where $P_{0a}(i, j, k)$ is the probability of no further independent measurements in channel 1 between t_1 and 0 or dependent measurements in channel 1 between t_1 and 0, where its counterpart in channel 2 is outside the interval $t_2 : \tau_k$. It reads

$$P_{0a}(i, j, k) = \mathrm{e}^{n_{1,\mathrm{indep}}i\Delta\tau} \prod_{l=-\infty}^{j-i} \mathrm{e}^{n_\mathrm{dep}(\tau_l)\max(i, l+i-j)\Delta\tau} \prod_{l=k}^{\infty} \mathrm{e}^{n_\mathrm{dep}(\tau_l)\max(i, k-l)\Delta\tau}$$

$P_{0b}(i, j, k)$ is the probability of no further independent measurements in channel 2 between t_2 and τ_k or dependent measurements in channel 2 between t_2 and τ_k, where its counterpart in channel 1 is outside the interval $t_1 : 0$. It reads

$$P_{0b}(i, j, k) = \mathrm{e}^{n_{2,\mathrm{indep}}(j-k)\Delta\tau} \prod_{l=-\infty}^{k} \mathrm{e}^{n_\mathrm{dep}(\tau_l)\max(j-k, l-k)\Delta\tau} \prod_{l=j-i}^{\infty} \mathrm{e}^{n_\mathrm{dep}(\tau_l)\max(j-k, j-i-l)\Delta\tau}$$

$P_{0c}(i, j, k)$ is the probability of no further dependent measurements in channels 1 and 2 between t_1 and 0 and between t_2 and τ_k respectively. It reads

$$P_{0c}(i, j, k) = \prod_{l=j}^{k-i} \mathrm{e}^{n_\mathrm{dep}(\tau_l)\max(i, j-k, j-l, l+i-k)\Delta\tau}$$

By rearranging the above sums and products, it is possible to first include all independent and dependent measurements, replacing $n_{1,\mathrm{indep}}$ and all $n_\mathrm{dep}(\tau_k)$ by n_1 and $n_{2,\mathrm{indep}}$ and all $n_\mathrm{dep}(\tau_k)$ by n_2 and reduce the respective sums afterwards by the contributions, which have been counted twice. Then the following probabilities can be written as

$$P_{1a}(i, j, k) = 1 - \mathrm{e}^{-n_1\Delta\tau + \sum_{l=j-i}^{k-i} n_\mathrm{dep}(\tau_l)\Delta\tau}$$

$$P_{1b}(i, j, k) = 1 - \mathrm{e}^{-n_2\Delta\tau + \sum_{l=j}^{j-i} n_\mathrm{dep}(\tau_l)\Delta\tau}$$

$$P_0(i, j, k) = \mathrm{e}^{n_1 i\Delta\tau + n_2(j-k)\Delta\tau + \sum_{l=j}^{k-i} n_\mathrm{dep}(\tau_l)[\min(l,k) - \max(i+l,j)]\Delta\tau}$$

14

The above sums yield a matrix $\mathbf{M}$, mapping the true correlation function R_{12} onto the expected estimated correlation function $\hat{R}_{12}$ after interpolation and resampling as

$$\hat{R}_{12} = \mathbf{M} R_{12}$$

This prediction filter matrix then can be inverted, yielding the correction filter matrix, which applied to the estimated correlation functions $R'_{u,12}$ and $R'_{w,12}$ from the interpolated data finally yields corrected estimations of the correlation functions

$$
\begin{aligned}
R''_{u,12} &= \mathbf{M}^{-1} R'_{u,12} \\
R''_{w,12} &= \mathbf{M}^{-1} R'_{w,12}
\end{aligned}
$$

5.4 Normalization, Bessel's correction and final transformation

The final estimate of the correlation function (reduced to the total length of T_{C}) is obtained by normalization

$$R_{12}(\tau_k) = \frac{R''_{u,12}(\tau_k)}{R''_{w,12}(\tau_k)} + c_B$$

including Bessel's correction, where the correction c_B is related to the estimated variances of the mean estimators above. In the case of the interpolation method c_B is obtained similar to the procedure given in [18] for the direct spectral estimator and for the slotting technique adpated to the interpolation method and to the cross-correlation case as

$$c_B = \frac{\displaystyle\sum_{k=-\lfloor K/2 \rfloor}^{\lfloor (K-1)/2 \rfloor} R''_{u,12}(\tau_k)}{T_{\mathrm{B}} F - \displaystyle\sum_{k=-\lfloor K/2 \rfloor}^{\lfloor (K-1)/2 \rfloor} R''_{w,12}(\tau_k)}$$

The final correlation estimate is then transformed by means of the discrete Fourier transform (DFT) to a power spectral density

$$S_{12}(f_j) = \Delta\tau \mathrm{DFT}\left\{R_{12}(\tau_k)\right\} = \Delta\tau \sum_{k=-\lfloor K/2 \rfloor}^{\lfloor (K-1)/2 \rfloor} R_{12}(\tau_k) \mathrm{e}^{-2\pi\mathrm{i} f_j \tau_k}$$

with $f_j = j\Delta f, j = -\lfloor K/2 \rfloor \ldots \lfloor (K-1)/2 \rfloor$ giving a frequency resolution of $\Delta f = 1/K\Delta\tau$.

5.5 Remarks

Local normalization and fuzzy slotting [29, 30, 21, 28, 19] have not yet been adapted to the interpolation method. In contrast to the autocorrelation case,

the noise of the two channels can be assumed to be independent. Therefore, the interpolation method doesn't have any systematic error due to a noise component in the measured data. The statistical bias is suppressed due to the inherent weighting of the sample-and-hold interpolation, which holds values longer if the local data rate decreases. An example program can be found at [1]. The large computational effort to derive the filter correction matrix is a strong argument against the use of the interpolation method, at least in the general case of mixed independent and dependent measurements.

References

[1] URL `http://ldvproc.nambis.de/programs/pyLDV.html`.

[2] R J Adrian and C S Yao. Power spectra of fluid velocities measured by laser Doppler velocimetry. *Exp. in Fluids*, 5:17–28, 1987.

[3] P Buchhave, W K George, Jr, and J L Lumley. The measurement of turbulence with the laser Doppler anemometer. In *Annual Review of Fluid Mechanics*, volume 11, pages 442–503. Annual Reviews, Inc., Palo Alto, CA, 1979.

[4] P Buchhave and C M Velte. Reduction of noise and bias in randomly sampled power spectra. *Experiments in Fluids*, 56:79, 2015.

[5] J Fitzpatrick and L Simon. Estimation of cross-power spectra using sample-and-hold reconstruction of laser Doppler anemometry data. *Exp. in Fluids*, 39:954–965, 2005.

[6] M Gaster and J B Roberts. Spectral analysis of randomly sampled signals. *J. Inst. Maths. Applics.*, 15:195–216, 1975.

[7] M Gaster and J B Roberts. The spectral analysis of randomly sampled records by a direct transform. *Proc. R. Soc. Lond. A.*, 354:27–58, 1977.

[8] W K George. Quantitative measurement with the burst-mode laser Doppler anemometer. *Exp. Therm. and Fluid Sci.*, 1:29–40, 1988.

[9] W K George, P D Beuther, and J L Lumley. Processing of random signals. pages 757–800, 1978. Proceedings of the dynamic flow conference, Skovlunde, Denmark, (PO Box 121, DK-2740 Skovlunde).

[10] W T Mayo, Jr. A discussion of the limitations and extensions of power spectrum estimation with burst counter LDV systems. pages 90–104, 1974. Proceedings of the second international workshop on laser velocimetry, (Bull. No. 144), Purdue Univ., West Lafayette, Ind.

[11] W T Mayo, Jr, M T Shay, and S Ritter. The development of new digital data processing techniques for turbulence measurements with a laser velocimeter, 1974.

[12] S Moreau, G Plantier, J-C Valière, H Bailliet, and L Simon. Estimation of power spectral density from laser Doppler data via linear interpolation and deconvolution. *Experiments in Fluids*, 50:179–188, 2011.

[13] E Müller, H Nobach, and C Tropea. A refined reconstruction-based correlation estimator for two-channel, non-coincidence laser Doppler anemometry. *Meas. Sci. and Technology*, 9(3):442–451, 1998.

[14] H Nobach. *Verarbeitung stochastisch abgetasteter Signale: Anwendung in der Laser-Doppler-Anemometrie*. Shaker, Aachen, 1998. ISBN 3-8265-3332-1, Zugl.: Rostock, Univ., Diss., 1997.

[15] H Nobach. Processing of stochastic sampled data in laser Doppler anemometry. In *Proc. 3rd International Workshop on Sampling Theory and Applications, Loen, Norway*, pages 149–154, 1999.

[16] H Nobach. A global concept of autocorrelation and power spectral density estimation from LDA data sets. In *Proc. 10th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*, Lisbon, Portugal, 2000. paper 3.5.

[17] H Nobach. Local time estimation for the slotted correlation function of randomly sampled LDA data. *Exp. in Fluids*, 32:337–345, 2002.

[18] H Nobach. Corrections to the direct spectral estimation for laser Doppler data. *Exp. in Fluids*, 56:109, 2015.

[19] H Nobach. Fuzzy time quantization and local normalization for the direct spectral estimation from laser Doppler velocimetry data. *Exp. in Fluids*, 56:182, 2015.

[20] H Nobach. A model-free noise removal for the interpolation method of correlation and spectral estimation from laser Doppler data. *Exp. in Fluids*, 56:100, 2015.

[21] H Nobach, E Müller, and C Tropea. Correlation estimator for two-channel, non-coincidence laser-Doppler-anemometer. In *Proc. 9th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*, Lisbon, Portugal, 1998. paper 32.1.

[22] H Nobach, E Müller, and C Tropea. Efficient estimation of power spectral density from laser Doppler anemometer data. *Experiments in Fluids*, 24:499–509, 1998.

[23] J B Roberts and D B S Ajmani. Spectral analysis of randomly sampled signals using a correlation-based slotting technique. *IEE Proc.*, 133(2):153–162, 1986.

[24] P F Scott. Random sampling theory and its application to laser velocimeter turbulent spectral measurements. Technical report, General Electric Co., Corporate Res. and Development, 1974. Report No. 74CRD216, Tech, Info Series.

17

[25] P F Scott. *Distortion and Estimation of the Autocorrelation Function and Spectrum of a Randomly Sampled Signal.* PhD thesis, Rensselear Polytechnic Institute, Troy, NY, 1976.

[26] M T Shay. *Digital Estimation of Autocovariance Functions and Power Spectra from Randomly Sampled Data Using a Lag Product Technique.* PhD thesis, Texas A&M University, College Station, TX, 1976.

[27] L Simon and J Fitzpatrick. An improved sample-and-hold reconstruction procedure for estimation of power spectra from LDA data. *Exp. in Fluids*, 37:272–280, 2004.

[28] M J Tummers, D M Passchier, and P A Aswatha Narayana. LDA measurements of time- and spatial correlation functions in an adverse pressure gradient wake. In *Proc. ASME/JSME Fluids Eng. and Laser Anemometry Conf.*, pages 347–355, Hilton Head Island, South Carolina, USA, 1995. FED-Vol. 229.

[29] H R E van Maanen, H Nobach, and L H Benedict. Improved estimator for the slotted autocorrelation function of randomly sampled LDA data. *Meas. Sci. Technol.*, 10(1):L4–L7, 1999.

[30] H R E van Maanen and M J Tummers. Estimation of the autocorrelation function of turbulent velocity fluctuations using the slotting technique with local normalization. In *Proc. 8th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*, Lisbon, Portugal, 1996. paper 36.4.

[31] C M Velte, W K George, and P Buchhave. Estimation of burst-mode LDA power spectra. *Exp. in Fluids*, 55:1674, 2014.