Motion artifact reduction in PPG signals

Jaqub Ghairat & Hassan Mouhsen

2015

Lund University

Master's Thesis
Biomedical Engineering

Faculty of Engineering LTH
Department of Biomedical Engineering

Supervisor: Frida Sandberg

Examiner: Leif Sörnmo

Abstract

The aim of this thesis was to investigate methods for artifact removal in PPG signals and to implement and evaluate a few existing algorithms claiming that the amplitude information is recovered when removing motion artifacts from photoplethysmographic signals (PPG) captured from pulse oximeters. We developed a new proposed method that uses a two-stage based approach with singular value decomposition and fixed fast ICA algorithm in order to generate a PPG-correlated reference signal that is used in adaptive noise cancellation. The results were promising and our proposed method is easy to implement and converges quickly with good extraction performance. It has a few design parameters and only needs the estimated period of the PPG signal. Our method could be used in a clinical routine for prediction of intradialytic hypotension. However it should be mentioned that although our method has great potential the simulations were only conducted on two healthy males. Further studies on a larger dataset might be needed in order to establish a full value of the efficacy of our method.

Acknowledgments

First, we would like to express our sincere gratitude to our supervisor, Frida Sandberg, for her guidance and support. We appreciate the fact that she had always time to answer our questions, moreover, with such great enthusiasm. Last but definitely not least, we would like to express our deepest gratitude to our families and friends. It would most certainly not have been possible to conduct this thesis without you.

Contents

1 | Introduction

1.1 Background

In recent years biomedical technologies has increased rapidly for more accurate diagnoses and effective treatments. In order to obtain clinically reliable measurements medical monitoring devices should be optimized. Optimized devices will give a better utilization of health care and facilitate the clinician to make correct medical decisions.

Pulse oximetry is one such medical device regarded as standard in operating rooms and intensive care units (ICU). It is a non-invasive method for evaluating a patients arterial blood oxygen saturation or heart rate. [1]

The requirement when performing measurements is that the patient must remain still while the assessment is accomplished. However in reality and especially in ambulatory monitoring during patient transportation noise caused by motion is inevitable. The deterioration of the signal caused by these movements, with the clinician unaware of it, can lead to false medical decisions and unreliable readings. Reliable readings are thus of huge importance in the presence of motion artifacts. [2]

Results from previous studies suggest that the envelope of the PPG signal may be used to predict acute symptomatic hypotension which is the most common complication associated with hemodialysis treatment and occur in approximately 25% of all sessions. [3] [4] [5] The photopletysmography (PPG) signal obtained from the pulse oximeter is disturbed by noise and motion artifacts that may influence the accuracy of envelope estimation.

Hence the removal of motion artifact which is stochastic in nature and present as in-band noise of importance in clinical settings due to the above mentioned reasons.

The objective of this thesis is to implement and evaluate algorithms for noise and artifact removal in PPG signals with respect to maintained amplitude information. Few scientific papers claims that the amplitude information is reserved when the effect of motion artifact is reduced using adaptive filters and independent component analysis (ICA) [6] [7]. The goal is to investigate these algorithm implementations in Matlab for evaluation, validation and improvement for the pulse oximeter CardioHolter 6.2 used in this thesis work. The results from this study could be used in clinical routine for prediction of intradialytic hypotension where the PPG signal is analyzed with respect to changes in amplitude.

This report is organized by first introducing relevant background about pulse oximeters and artifacts in PPG signals, furthermore signal processing, optimization, adaptive filter and independent component analysis theory will be covered, followed by methods, experimentational setup and measurements. Finally the results are presented and discussed.

2 | Pulse Oximeters

A pulse oximeter is a device used to continuously monitor and measure the arterial blood oxygen saturation (SpO_2) or heart rate non-invasively. The simplicity and ability to provide fast and inexpensive measurements makes the pulse oximeters important in clinical settings. Patients risking respiratory failure, hypoxemia or cardiac problems can easily be detected by clinicians. Cardiopulmonary diseases and sleep disorders can also be screened [8].

2.1 Principles of pulse oximeters

Pulse oximeters utilizes a microprocessor unit and a peripheral probe consisting of a photodetector on one side of the probe and a pair of light-emitting diodes (LED) on the other side. The two LEDs emit lights at different wavelengths. One in the red spectrum at a wavelength of 660 nm and the other in the infrared spectrum at a wavelength of 940 nm. A translucent part of the body is used for measurements such as finger tips, earlobes, toes and foreheads. The transmitted light from the diodes through the tissue bed is determined by the photodetector (the amount of light not absorbed by the tissue). [9]

The different light-absorbing characteristics between oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) is the basis of the principle. The absorption is notably lower for oxyhemoglobin at 660 nm (red region of light spectrum) than deoxyhemoglobin. At 940 nm in the infrared region of light spectrum the absorption of deoxyhemoglobin is lower in relation to oxyhemoglobin.

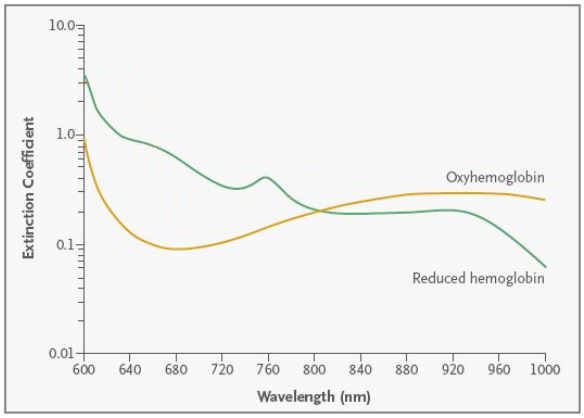


Figure 2.1: Light spectrum characteristics of HbO_2 (oxyhemoglobin) and Hb (reduced hemoglobin).

The arterial oxygen saturation estimate from the pulse oximeter is denoted as SpO_2 and is an estimate of SaO_2 which is defined as in the equation below.

$$SaO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb} + c_{COHb} + c_{MetHb}} \times 100\% \tag{2.1}$$

The total amount of hemoglobin in the denominator of equation 2.1 is not only Hb and HbO_2 but also other forms of hemoglobin as carboxyhemoglobin (COHb) and methemoglobin (MetHb). The latter two are referred as dysfunctional hemoglobin because of reduced oxygen transportation and the former functional hemoglobin. A pulse oximeter uses the definition of functional oxygen saturation defined in the equation below [10].

$$SpO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb}} \times 100\% \tag{2.2}$$

The signal retrieved from the pulse oximeter is called photoplethysmographic (PPG) and is produced as a result of the periodic heart contractions and relaxation. It is a volumetric measurement associated with arterial blood volume changes. The AC part is the pulsatile component of the PPG signal related to the arterial blood volume change by cause of systolic and diastolic phases of the cardiac cycle. The DC part is the non-pulsatile component of the PPG signal associated with light intensity baseline depending on tissue, skin, bone and venous blood [6]. The sudden drop in the systolic phase in the PPG signal is called dicrotic notch and is caused by aortic valve closure [11].

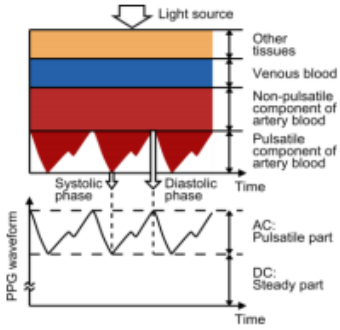


Figure 2.2: Light attenuation and the PPG waveform. [12]

2.1.1 Estimation of oxygen saturation

The estimation of the arterial oxygen saturation and the principle of pulse oximeters is based on Beer-Lambert law [10]. The law states that there exists an exponential relationship between the attenuation of light passing through a medium with respect to the properties of the material. The intensity of the transmitted light through the material is given by

$$\begin{aligned} &I = I_0 e^{-A} \\ &\text{where } A = -\ln I/I_0 = \epsilon(\lambda) cl \end{aligned} \tag{2.3}$$

and I_0 is the light intensity entering the volume, l is the length of the optical path, c the substance concentration of the light-absorbing material and $\epsilon(\lambda)$ the molar absorptivity or extinction coefficient as a function of wavelength λ. A is the absorbance amount. In the case of multiple absorbers the equations become as following.

$$
\begin{aligned}
I &= I_0 e^{-A} \\
A &= \sum_i A_i = \sum_i \epsilon_i(\lambda) c_i l
\end{aligned}
\tag{2.4}
$$

Taking into account the different concentrations and absorbance coefficients of the human body (bone, tissue, skin and hair), Beer-Lamberts law can be used to express the light intensity of the non-pulsatile components $I_{np} = I_0 e^{\sum_i -\epsilon_i(\lambda) c_i l}$. The ratio between the maximum light intensity at diastole (peak) and the minimum at systole at wavelength λ_1 is calculated as

$$
\begin{aligned}
\frac{I_{max}}{I_{min}} &= \frac{I_{np} e^{-(\epsilon_{HbO_2}(\lambda_1) c_{HbO_2} + \epsilon_{Hb}(\lambda_1) c_{Hb}) l}}{I_{np} e^{-(\epsilon_{HbO_2}(\lambda_1) c_{HbO_2} + \epsilon_{Hb}(\lambda_1) c_{Hb})(l + \Delta l)}} \\
&= e^{(\epsilon_{HbO_2}(\lambda_1) c_{HbO_2} + \epsilon_{Hb}(\lambda_1) c_{Hb}) \Delta l}
\end{aligned}
\tag{2.5}
$$

where Δl is the optical path length between diastole and systole. This ratio is independent of the effect of surrounding tissues as well as the incident light intensity. In order to eliminate the optical path length Δl also, the ratio of ratios gives the following relationship

$$
\begin{aligned}
\Lambda &= \frac{ln(\frac{I_{max}}{I_{min}}|_{\lambda_1})}{ln(\frac{I_{max}}{I_{min}}|_{\lambda_2})} \\
&= \frac{\epsilon_{HbO_2}(\lambda_1) c_{HbO_2} + \epsilon_{Hb}(\lambda_1) c_{Hb}}{\epsilon_{HbO_2}(\lambda_2) c_{HbO_2} + \epsilon_{Hb}(\lambda_2) c_{Hb}}
\end{aligned}
\tag{2.6}
$$

Solving for the arterial oxygen saturation level yields in an equation with a one-to-one relationship between the ratio and saturation. In the second evaluation of the equation below typical values are inserted.

$$
\begin{aligned}
SpO_2 &= \frac{\epsilon_{Hb}(\lambda_1) - \epsilon_{HbO_2}(\lambda_2)\Lambda}{\epsilon_{Hb}(\lambda_1) - \epsilon_{HbO_2}(\lambda_1) + [\epsilon_{HbO_2}(\lambda_2) - \epsilon_{Hb}(\lambda_2)]\Lambda} \\
&= \frac{0.81 - 0.18\Lambda}{0.63 + 0.11\Lambda}
\end{aligned}
\tag{2.7}
$$

However in most practical settings the above equation is rarely evaluated, instead a normalization technique is used by dividing the peak-to-peak AC component of the PPG signal with the DC component (equation 2.8). Manufacturers then relate the ratio of ratios to empirical calibration in order to obtain the arterial oxygen saturation. The Beer-Lambert model gives erroneous estimates of the true value below 85% due to scattering of light by red blood cells, hence the use of look-up tables based on empirical studies [13].

$$
\Lambda = \frac{AC_r}{DC_r} \Big/ \frac{AC_{ir}}{DC_{ir}} \tag{2.8}
$$

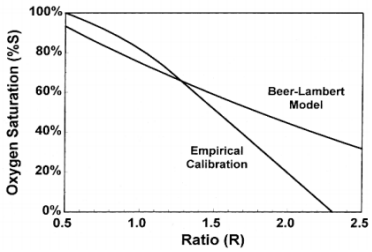


Figure 2.3: Calibration curves for pulse oximeters. [5]

2.2 Artifacts in PPG signal

There are numerous factors that influence the accuracy and limits the performance of a pulse oximeter causing a deteriorated PPG signal. The most critical limitations that can corrupt the signal to such an extent that it might not be used in clinical settings are the following.

2.2.1 Motion artifacts

This is a common name of the problematic noise that occurs due to displacement of the sensor probe which is present in real-world by movements during patient treatment such as waving, rubbing, seizures etc. The shape of the PPG signal is corrupted and leads to inefficient, unreliable treatment and an increased cost of care. One type of motion artifact is connected to a physically active individual moving the finger horizontally or vertically inducing low venous pressure. This perturbation can influence the AC part of the PPG signal resulting in lower SpO_2 readings [14]. It is often difficult to remove motion artifacts since the frequency band overlaps that of the PPG signal. Typically the frequency band caused by motion artifacts is 0.1 Hz and above. The PPG signal is in the range 0.5-4 Hz and hence the overlapping makes the use of classical signal processing techniques impractical in order to separate it [15]. It has been the subject of several biomedical research projects to remove motion artifacts completely.

2.2.2 Ambient light interference

Light sources other than the LED included within the pulse oximeter such as operating room lamp may result in ambient light artifacts leading to erroneous reading. Other types of interferences may also interfere such as electromagnetic radiation by magnetic resonance imaging MRI and the presence of intravascular dyes (methylene blue or indigo carmine). It may alter the red and infrared light-absorption properties of tissues. [8]

2.2.3 Anemia

Anemia is defined as the decrease of the amount of red blood cells and can effect the PPG signals resulting in large errors in the measured oxygen saturation [16]. Hypothermia, the potentially dangerous drop in body temperature leading to vasoconstriction is another complication [17].

2.3 Optical theory adopting Schuster's theory

The assumption when using Beer-Lamberts law is that the medium must not scatter the radiation. As mentioned before scattering of light leads to erroneous readings and therefore Aoyagi adopted Schuster's theory of *radiation through a foggy atmosphere* in his report in order to give more precise predictions [20]. The absorption in blood as a function of blood vessel thickness changes (Δl_b) is derived as following through experimentation by Aoyagi:

$$\begin{aligned} \Delta A_b &= \sqrt{\epsilon_h(\epsilon_h + F)} \cdot c_{totHb} \cdot \Delta l_b + Z_b \Delta l_b \\ \epsilon_h &= S \cdot \epsilon_{HbO_2} + (1 - S)\epsilon_{Hb} \end{aligned} \tag{2.9}$$

c_{totHb} is the hemoglobin concentration in the blood, F is a scattering constant, S the oxygen saturation and Z_b is a constant independent of wavelength which is zero when the optical receiver is wide enough. Modifying the above equation taking into consideration the tissue effect the expression become as

$$\Delta A_b = \sqrt{\epsilon_h(\epsilon_h + F)} \cdot c_{totHb} \cdot \Delta l_b + Z_b \Delta l_b + Z_t \Delta l_t \tag{2.10}$$

where Z_t is a tissue constant independent of the wavelength and Δl_t is the tissue thickness change [7]. The above equation holds for one blood vessel. The effect of both arterial and venous blood vessels expands the expression in the following form

$$\Delta A_b = \sqrt{\epsilon_{h_a}(\epsilon_{h_a} + F)} \cdot c_{totHb_a} \cdot \Delta l_{b_a} + \sqrt{\epsilon_{h_v}(\epsilon_{h_v} + F)} \cdot c_{totHb_v} \cdot \Delta l_{b_v} + \Delta A_s \tag{2.11}$$

where subscripts a and v refers to arterial blood and venous blood respectively and ΔA_s is the accumulated effect of both tissue and blood thickness changes that is wavelength independent.

3 | Signal Processing

In daily life noise and interference may be present in a signal (entity that carries information) and consequently the need for signal processing is of vital importance concerning representation and manipulation of signals. A signal may be a function of continuous variables or discrete variables. In this section the fundamental concepts of discrete time signals and systems will be presented as it is of importance in biomedical signal processing. The characteristics of the PPG signal is time-varying due to motion and local tissue changes and it is therefore natural to view it as an stochastic process rather than deterministic [21]. Stochastic processes will be described as well as topics including stationarity auto- and crosscorrelation. Frequency representation of the signal will also be introduced. Finally wiener filtering and linear prediction is explained.

3.1 Stochastic processes

A stochastic process is also called a random process and is a function of time on some observation interval. The name derives from the fact that it is not possible to predict or describe the exact waveform when conducting an experiment. Discrete time and uniformly spaced stochastic processes arise in practical applications such as radar, seismic, digital computer data and biomedical signals. [22] Before considering stochastic processes a brief review of basic terminology of the properties of random variables and distributions will be presented.

3.1.1 Random variables and distributions

Random variables are found in almost any practical application and it is therefore important to understand and manipulate them. There are two types of random variables, discrete random variables on a sample space Ω (set of all possible outcomes) consisting of a discrete set of events w_i (subsets of sample space) and continuous random variables assuming a continuum of values [23]. For a real-valued random variable x the cumulative distribution function (cdf) is defined as following for discrete type:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{x_i \leq x} p_X(x_i) \end{aligned} \tag{3.1}$$

In most cases the process is characterized by density function instead of cdf, the probability density function (pdf) of a discrete random variable is shown below, respectively.

$$p_X(x) = P(X = x) \tag{3.2}$$

The density function of a gaussian probability distribution is used frequently in signal processing operations, one conventional use is to describe additive white noise. The gaussian density function is given by (μ is the mean and σ is the standard deviation):

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.3}$$

3.1.2 Expectations, mean, covariances and correlation

The *expectation* of a quantity $g(\mathbf{x})$ derived from a random vector $\mathbf{x}$ is useful for processing and analyses in signal processing and is defined as:

$$\begin{aligned} E\{g(\mathbf{x})\} &= \int_{-\infty}^{\infty} g(\mathbf{x}) p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\ \mathbf{x} &= (x_1, x_2, \ldots, x_n)^T \end{aligned} \tag{3.4}$$

where $g(\mathbf{x})$ could be either a vector, scalar or a matrix. The *mean vector* or sometimes called first moment of a random vector is defined as the expectation of $\mathbf{x}$:

$$\begin{aligned} \mathbf{m}_{\mathbf{x}} &= E\{\mathbf{x}\} = \int_{-\infty}^{\infty} \mathbf{x} p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\ m_{x_i} &= E\{x_i\} = \int_{-\infty}^{\infty} x_i p_{x_i}(x_i) dx_i \end{aligned} \tag{3.5}$$

where the second equation is computed as the ith component of the n-dimensional vector $\mathbf{m}_{\mathbf{x}}$. The statistical relationship and dependence between random variables or sets of data using second-order statistics is another important issue, hence the use of *correlations*. The correlation between component i and j is denoted as r_{ij} and is given by the second moment. The $n \times n$ correlation matrix $\mathbf{R}_{\mathbf{x}}$ of the vector $\mathbf{x}$ forms all correlations r_{ij} where the element is found in row i and column j. The correlation matrix is symmetric, positive semidefinite and all its eigenvalues are real and nonnegative if the matrix is positive definite. [23]

$$\begin{aligned} r_{ij} &= E\{x_i x_j\} = \int_{-\infty}^{\infty} x_i x_j p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\ \mathbf{R}_{\mathbf{x}} &= E\{\mathbf{x}\mathbf{x}^T\} \end{aligned} \tag{3.6}$$

Central moments or also called *covariances* are calculated in a similar way by subtracting the mean vectors from the random vector. The elements of the covariance matrix $\mathbf{C}_{\mathbf{x}}$ are called covariances (c_{ij}) and can be found in the matrix in the same way as for the correlation matrix above.

$$\begin{aligned} c_{ij} &= E\{(x_i - m_i)(x_j - m_j)\} \\ \mathbf{C}_{\mathbf{x}} &= E\{(\mathbf{x} - \mathbf{m}_{\mathbf{x}})(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T\} \end{aligned} \tag{3.7}$$

The relationship between correlation and covariance matrix is easily shown using the properties of expectation operator, notice that if the mean vector is zero then the matrices become the same.

$$\mathbf{R_x} = \mathbf{C_x} + \mathbf{m_x}\mathbf{m_x}^T \tag{3.8}$$

The cross-correlation and cross-covariance matrices is also used when dealing with two different vectors x and y that could have different dimensions. In general they are not symmetric and not necessarily square matrices.

$$\begin{aligned} \mathbf{R_{xy}} &= E\{\mathbf{x}\mathbf{y}^T\} \\ \mathbf{C_{xy}} &= E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\} \end{aligned} \tag{3.9}$$

3.1.3 Estimating expectations

If the probability density function is unknown the expectation of $\mathbf{g}(\mathbf{x})$ could be estimated by averaging over a set of N available samples of the vector $\mathbf{x}$.

$$E\{\mathbf{g}(\mathbf{x})\} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}(\mathbf{x}_i) \tag{3.10}$$

In similar way the sample mean and the cross-correlation matrix can be estimated, this can be readily obtained also for $\hat{\mathbf{R}_{\mathbf{xx}}}$, $\hat{\mathbf{C}_{\mathbf{xx}}}$ and $\hat{\mathbf{C}_{\mathbf{xy}}}$.

$$\begin{aligned} \hat{\mathbf{m}}_\mathbf{x} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \\ \hat{\mathbf{R}}_\mathbf{xy} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i{\mathbf{y}_i}^T \end{aligned} \tag{3.11}$$

3.1.4 Uncorrelatedness and independence

The definition of uncorrelatedness is that two random vectors are said to be uncorrelated if their cross-covariance matrix is zero. The same holds for two scalar random variables, equivalently their covariance is zero. This means that there is no linear relationship between the variables. Zero covariance implies zero correlation.

$$\mathbf{C_{xy}} = E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\} = \mathbf{0} \tag{3.12}$$

Statistical independence is a key concept in many areas and two random variables or vectors are said to be independent if the occurrence of one does not give any information of the value of the other. In mathematical terms this can be described as a factorization of the joint density into the product of the marginal densities of two random variables x and y. [23]

$$p_{x,y}(x, y) = p_x(x)p_y(y) \tag{3.13}$$

The definition can be extended in a general way expressing the equation for more than two random variables, for random vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, ..., the independence condition is as following.

$$p_{\mathbf{x},\mathbf{y},\mathbf{z},\ldots}(\mathbf{x},\mathbf{y},\mathbf{z},\ldots) = p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{y}}(\mathbf{y})p_{\mathbf{z}}(\mathbf{z})\ldots \tag{3.14}$$

Two random variables that are independent are said to be uncorrelated, however the opposite, that uncorrelatedness implies independence does not always hold. [23]

3.1.5 Stochastic processes

A discrete-time stochastic process $x(t)$ is defined at a set of discrete times t_1, t_2, ..., t_k and is said to be *strictly stationary* if its joint probability distribution is time shift invariant. Consequently that implies that parameters such as mean and variance do not change over time. In order to characterize stochastic processes, the mean function, autocovariance and autocorrelation functions are defined. [24] The mean-value function is defined as

$$m_x(t) = E\{x(t)\} \tag{3.15}$$

When the stochastic process is strictly stationary the mean-value function is simply constant m_x [25]. The autocovariance function is defined below, especially for lag $\tau = 0$ it reduces to the variance σ_x^2 for a stationary process.

$$c_x(t, t-\tau) = E\{(x(t) - m_x(t))(x(t-\tau) - m_x(t-\tau))\}, \quad \tau = 0, \pm 1, \pm 2, \ldots \tag{3.16}$$

The autocorrelation function of the stochastic process is defined below and generally if lag $\tau = 0$ the function reduces to a constant $r_x(0) = E\{x(t)^2\}$ for a stationary process.

$$r_x(t, t-\tau) = E\{x(t)x(t-\tau)\}, \quad \tau = 0, \pm 1, \pm 2, \ldots \tag{3.17}$$

Moreover for strictly stationary processes the autocovariance and autocorrelation functions only depend on the difference between t and $t-\tau$, that is to say only depending on the lag τ.

$$\begin{aligned} r_x(t, t-\tau) &= r_x(\tau) \\ c_x(t, t-\tau) &= c_x(\tau) \end{aligned} \tag{3.18}$$

A weaker form of stationarity in signal processing is often called ***wide-sense stationary*** (WSS). Such processes are common in practice since many physical processes are at least mildly nonstationary. The following properties hold for a WSS process in general:

- $m_x(t) = m_x$ constant for all t
- Time shift invariance of autocorrelation function $r_x(t, t-\tau) = r_x(\tau)$
- Finite variance $r_x(0) = E\{x(t)^2\} < \infty$

Representing the stochastic process at time n using the random vector

$$\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-m)]^T \tag{3.19}$$

the *correlation matrix* can be defined as the outer product of the vector with itself using the expectation operator as below, where H denotes Hermitian transpose (transposition combined with complex conjugation).

$$\mathbf{R_x} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\} = \begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(m) \\ r_x(1) & r_x(0) & r_x(1) & \cdots & r_x(m-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x(m) & r_x(m-1) & r_x(m-2) & \cdots & r_x(0) \end{bmatrix} \tag{3.20}$$

The correlation matrix is Toeplitz, that is to say, all the main diagonal and subdiagonal elements are equal. The toeplitz property holds for wide sense stationary stochastic processes and is helpful when solving linear equations enabling faster algorithms.

3.1.6 Mean ergodic theorem and time averages estimation

A stochastic process is ergodic if the mean and autocorrelation (usually unknown) can be estimated to its ensemble average by replacing the expectation operators and calculate sufficiently long time averages from a single realization of the process [26]. Assuming that N samples are available, the estimates can be written as:

$$\begin{aligned} \hat{m}_x(N) &= \frac{1}{N}\sum_{n=1}^{N} x(n) \\ \hat{r}_x(l, N) &= \frac{1}{N-l}\sum_{n=1}^{N-l} x(n+l)x(n) \end{aligned} \tag{3.21}$$

3.1.7 Frequency domain representation

To gain more insight and understanding of a process it could be easier to represent it in the frequency domain instead of time domain. The statistical parameter is called *power spectral density*, assuming a WSS process, it is defined as the discrete-time Fourier transform of the autocorrelation series. [27]

$$S_x(w) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jwk} \tag{3.22}$$

The time domain representation can be obtained by applying the inverse discrete-time Fourier transform of the power spectral density.

$$r_x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(w)e^{jwk}dw \tag{3.23}$$

The power spectrum is a periodic, real-valued, even and continuous function of the angular frequency. If the process is complex valued then its power spectral density might not necessarily be even. Using the z-transform instead, the power spectral density can be written

$$S_x(z) = \sum_{k=-\infty}^{\infty} r_x(k) z^{-k} \tag{3.24}$$

3.1.8 Linear Prediction

As mentioned before linear prediction is the case when estimating or predicting future values of the desired signal based on linear combination of previous values of noisy observations. In linear prediction the operation is called one-step predictor and the Wiener filter theory may be used in order to estimate future values. Specifically if the desired signal $d(n) = u(n+1)$ then it is referred to as *forward linear prediction* and if $d(n) = u(n-1)$ it is called *backward linear prediction*. In the case of forward linear prediction the Wiener-hopf equations for the optimum linear prediction is as following, the same method can be applied in the case of backward linear prediction, the only difference is that the cross-correlation between the desired response and the input signal will be different. [28] [29]

$$r_{du}(k) = E\{d(n)u^*(n-k)\} = E\{u(n+1)u^*(n-k)\} = r_u(k+1)$$

$$\begin{bmatrix} r_u(0) & r_u^*(1) & \dots & r_u^*(p-1) \\ r_u(1) & r_u^*(0) & \dots & r_u^*(p-2) \\ \dots & \dots & \dots & \dots \\ r_u(p-1) & r_u^*(p-2) & \dots & r_u(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \dots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_u(1) \\ r_u(2) \\ \dots \\ r_u(p) \end{bmatrix} \tag{3.25}$$

The forward prediction error or the mean-square error is

$$J_{min} = r_d(0) - \sum_{k=0}^{p-1} w(k) r_u^*(k+1) \tag{3.26}$$

4 | Optimization

In this section the mathematical development of different optimization algorithms will be formulated. Optimization arise in many areas and the objective is to find an optimal solution to a given quantity with subject to possible restrictions. The goal may be to find a minimum or a maximum of a function. The *objective function* is the function to be optimized, decision variables are the parameters associated with the function and the restrictions on allowed parameter values is called *constraints*. Generally a constrained optimization problem can be formulated as

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2, \ldots, x_n)^T \\ \text{subject to} \quad & g_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, n \\ & h_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m \end{aligned} \tag{4.1}$$

where $g_i(\mathbf{x}) = 0$ is called *equality constraints* and $h_i(\mathbf{x}) \leq 0$ *inequality constraints*. The feasible solution is the vector x that satisfies all the constraints, all such points forms the feasible region [30]. The optimal point or feasible point $\mathbf{x}^*$ to the above problem must satisfy the condition $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x}$. Naturally an optimization problem can also be formulated as the maximization of a function in which case the constraints and the conditions will be of the reverse. An optimization problem with no constraints is called an unconstrained optimization problem. There are many classes, families and subfields of optimization, *convex optimization* is one where the objective function is convex (segment joining two points lies entirely above function graph) and hence there is only one optimal solution which is globally optimal. Other subfields are linear programming (LP) where the constraints and objective function is linear and nonlinear programming (NLP) where some objective function or constraints may be nonlinear. [31]

Some methods require only equality constraints in order to be solved (e.g. Lagrangian methods), thus a transformation of the inequality constraints into equality constraints is essential. This can be done introducing squared slack variables z_i transforming the inequality constraints $h_i(\mathbf{x}) \leq 0$ into equality constraints as following. [32]

$$h_i(\mathbf{x}) + z_i^2 = 0 \tag{4.2}$$

4.1 Gradient ascent

Gradient ascent is a first-order unconstrained optimization method also known as *steepest ascent*. It is an iterative method that finds local maximum of a continuously differentiable multivariate function [33]. The approach is to take steps in the direction proportional to the positive gradient. If moving in the direction of the negative gradient the approach is called *steepest descent* or gradient descent and the local minimum is searched for. Consider maximizing a function $f(\mathbf{x})$ starting with an initial guess $\mathbf{x}_0$ and moving in the direction of the positive gradient of the function ($\nabla f(\mathbf{x}_0)$) by a distance controlled by the positive scalar step-size parameter μ adjusting the length. The algorithm is generally described as following. [34]

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mu \nabla f(\mathbf{x}_n) \tag{4.3}$$

4.2 Newton's method

Newton's method can be used to solve unconstrained optimization problems when finding minimum or maximum of an objective function $f(\mathbf{x})$ [35]. Using the Taylor series expansion around $\mathbf{x}_n$ yields

$$f(\mathbf{x}) \approx f(\mathbf{x}_n) + \nabla f(\mathbf{x}_n)^T(\mathbf{x} - \mathbf{x}_n) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_n)^T \nabla^2 f(\mathbf{x}_n)^T(\mathbf{x} - \mathbf{x}_n) \tag{4.4}$$

where ∇f is the gradient vector and $\nabla^2 f$ is called the Hessian matrix and is defined as below.

$$\begin{aligned} H = &\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix} \\ &\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T \end{aligned} \tag{4.5}$$

The unique solution that minimizes $f(\mathbf{x})$ is obtained if the Hessian matrix is positive definite and it is negative definite if maximum is searched for [36]. Computing the gradient of the Taylor series expansion above with respect to $\mathbf{x} - \mathbf{x}_n$ and setting equal to zero gives

$$\nabla f(\mathbf{x}_n) + \nabla^2 f(\mathbf{x}_n)(\mathbf{x} - \mathbf{x}_n) = 0 \tag{4.6}$$

which yields in the symmetric linear system

$$(\mathbf{x} - \mathbf{x}_n) = -(\nabla^2 f(\mathbf{x}_n))^{-1} \nabla f(\mathbf{x}_n) \tag{4.7}$$

which results in the iterative equation below converging towards the minimizer $\mathbf{x}^*$.

$$\mathbf{x}_{n+1} = \mathbf{x}_n - (\nabla^2 f(\mathbf{x}_n))^{-1} \nabla f(\mathbf{x}_n) \tag{4.8}$$

4.3 Lagrange multiplier method

The Lagrange multiplier method solves constrained optimization problems with equality constraints [37]. The theory can also be extended in order to deal with inequalities. Consider the equality constraint problem

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2, \ldots, x_n)^T \\ &\text{subject to} \quad g_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, n \end{aligned} \tag{4.9}$$

where the function and constraints are continuously differentiable in $\mathbb{R}^n$. The Lagrange multiplier method is based on the *Lagrangian function* in the equation below. Solving for stationary points of the Lagrangian function yields in solution of the original constrained problem, however not all points guarantee this fact and hence optimality conditions are necessary [38].

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}), \quad \lambda = (\lambda_1, \ldots, \lambda_k) \tag{4.10}$$

The vector λ contains the *Lagrange multipliers* and if $\mathbf{x}^*$ is an optimal solution then the following necessary optimality conditions known as Karush-Kunn-Tucker (KKT) conditions will be satisfied by λ^*

$$\begin{aligned} \nabla_x L(\mathbf{x}^*, \lambda^*) = 0 &\iff \nabla f(\mathbf{x}^*) + \sum_{i=1}^{k} \lambda_i^* \nabla g_i(\mathbf{x}^*) = 0 \\ \nabla_\lambda L(\mathbf{x}^*, \lambda^*) = 0 &\iff g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, k \end{aligned} \tag{4.11}$$

Setting up the system of linear equations using the above conditions and solving for the stationary points will determine the optimal solutions. If the second order sufficient condition also known as local convexity assumption is satisfied ($\nabla^2_{xx} L(\mathbf{x}^*, \lambda^*) > 0$) or put in other words the Hessian is positive definite then a local minimum is found. The feasible point is a local maximizer if the Hessian is negative definite. [39]

$$\begin{cases} \nabla f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i \nabla g_i(\mathbf{x}) = 0 \\ g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, k \end{cases} \tag{4.12}$$

Extension of the Lagrange multipliers method can be done covering the case of both inequality and equality constraints. Consider the optimization problem stated in equation 2.48, then the Langrangian function is defined as

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{k} \mu_i h_i(\mathbf{x}) \tag{4.13}$$

where $\mu = (\mu_1, \ldots, \mu_k)$ are Lagrange multipliers for each constraint $h_i(\mathbf{x}) \leq 0$. Inequality constraints $h_i(\mathbf{x})$ can be active or inactive. If $h_i(\mathbf{x}) = 0$ at a feasible solution then the constraint is said to be active and inactive if $h_i(\mathbf{x}) \leq 0$. Inactive constraints have no influence in local properties satisfied by the optimal solution and hence if it were known a priori which constraints were active the solution could be determined by treating all constraints as equality constraints [40]. The necessary KKT conditions for determining the optimal solutions is now as following.

$$\begin{aligned}
&\nabla f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i \nabla g_i(\mathbf{x}) + \sum_{i=1}^{k} \mu_i \nabla h_i(\mathbf{x}) = 0 \\
&h_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, k \\
&g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, k \\
&\mu_i \geq 0, \quad i = 1, \ldots, k \\
&\mu_i g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, k
\end{aligned} \tag{4.14}$$

4.3.1 Dual problem

The dual problem separates the optimization problem into two parts, the *primal problem* which is the original objective and the *dual problem* that will provide a lower bound on the solution of the primal objective. The convergence properties may not be very satisfying, however the dual problem is often easy to implement and improves the approximation of the solution [40]. The *Lagrange dual function* is defined as infimum of the Lagrangian function

$$\phi(\lambda, \mu) = \inf_{x} L(\mathbf{x}, \lambda, \mu) \tag{4.15}$$

The infimum is defined as the largest number as a lower bound on a collection of points and it is easy to show that the dual function gives a lower bound on the optimal value $\phi^* \leq f^*$. Hence by maximizing the duality function will give the best approximation of the lower bound. The *dual problem* is therefore defined as

$$\text{maximize } \phi(\lambda, \mu) \tag{4.16}$$

4.3.2 Penalty methods

Constrained optimization problems can be replaced by unconstrained problems and this is accomplished by adding a term called *penalty function* to the objective function multiplied by a penalty parameter that violates the constraints [41]. The idea is that the penalty problem will converge to the solution of the original constrained problem when the penalty parameter tend towards infinity. Consider the optimization problem below with only equality constraints

$$\begin{aligned}
&\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x}) \\
&\text{subject to} \quad \mathbf{g}(\mathbf{x}) = 0
\end{aligned} \tag{4.17}$$

then a common and standard unconstrained penalty problem by adding a quadratic penalty term is

$$\text{minimize } f(\mathbf{x}) + \frac{1}{2} c \left\| \mathbf{g}(\mathbf{x}) \right\|^2 \tag{4.18}$$

where c is the scalar penalty parameter and $\|\cdot\|$ denotes the Euclidian norm [42].

4.3.3 Augmented Lagrangian Method

Combining the concepts of duality and penalty method one arrives at a concept called the augmented Lagrangian method or alternatively referred to as multiplier method. It is one of the most effective methods eliminating disadvantages associated with both methods alone [40].

Equality constraints
Consider the optimization problem in equation 2.64, as mentioned the equivalent problem is obtained using penalty functions as stated in equation 2.65. Combining this with duality problem by the following definition of the dual function

$$\phi(\lambda) = \min\{f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + \frac{1}{2} c \left\|\mathbf{g}(\mathbf{x})\right\|^2\} \tag{4.19}$$

and updating the multipliers with the following iterative process (based on steepest ascent) in order to maximize the dual function [40].

$$\lambda_{k+1} = \lambda_k + c\mathbf{g}(\mathbf{x}(\lambda_k)) \tag{4.20}$$

Inequality constraints
Inequality constraints can be incorporated and the derivation is straightforward but tedious. Consider the optimization problem with inequality constraints only

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) \leq 0 \end{aligned} \tag{4.21}$$

Using the concept of slack variables as introduced before, the above problem can be stated as an equality constrained version in which one can apply the dual function in equation 2.66.

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) + z_i^2 = 0, \quad i = 1, \ldots, m \end{aligned} \tag{4.22}$$

As to simplify the analytical derivation $v_i = z_i^2$ is defined for convenience. The dual function is hence

$$\begin{aligned} \phi(\mu) &= \min\{f(\mathbf{x}) + \sum_{i=1}^{m} (\mu_i[h_i(\mathbf{x}) + z_i^2] + \frac{1}{2} c \left\|h_i(\mathbf{x}) + z_i^2\right\|^2)\} \\ &= \min\{f(\mathbf{x}) + \sum_{i=1}^{m} (\mu_i[h_i(\mathbf{x}) + v_i] + \frac{1}{2} c \left\|h_i(\mathbf{x}) + v_i\right\|^2)\} \end{aligned} \tag{4.23}$$

Minimization of the introduced slack variable must be carried out in order to obtain a dual function that only depends on minimization with respect to $\mathbf{x}$. The expression below is defined as the part where the slack variable is involved in the dual function [40].

$$P_i = \mu_i[h_i(\mathbf{x}) + v_i] + \frac{1}{2}c\left\|h_i(\mathbf{x}) + v_i\right\|^2 \tag{4.24}$$

Taking the derivative with respect to the constraint function and solving for v_i and with the condition $v_i \geq 0$ in mind the following solution is obtained.

$$v_i = \begin{cases} -h_i(\mathbf{x}) - \frac{\mu_i}{c}, & \text{if} - h_i(\mathbf{x}) - \frac{\mu_i}{c} \geq 0 \\ 0 \quad \text{otherwise} \end{cases} \iff v_i = \max\left[0, -h_i(\mathbf{x}) - \frac{\mu_i}{c}\right] \tag{4.25}$$

Substituting the above solution in equation 2.71 yields

$$P_i = \begin{cases} -\frac{\mu_i^2}{2c}, & \text{for } v_i = -h_i(\mathbf{x}) - \frac{\mu_i}{c} \\ \\ \frac{1}{2c}\{[\mu_i + ch_i(\mathbf{x})]^2 - \mu_i^2\}, & \text{for } v_i = 0 \end{cases} \tag{4.26}$$

which is equivalent to

$$P_i = \frac{1}{2c}\{[\max(0, \mu_i + ch_i(\mathbf{x}))]^2 - \mu_i^2\} \tag{4.27}$$

The dual function for the the inequality constrained optimization problem is hence

$$\phi(\mu) = \min\{f(\mathbf{x}) + \sum_{i=1}^{m} P_i\} \tag{4.28}$$

where P_i is defined as in equation 2.74. The optimum multipliers μ can be obtained iteratively in the exact same way as for the constrained version as seen in equation 2.67.

Combined augmented lagrangian method
By combining the result from the inequality constrained and the equality constrained optimization problem the general augmented lagrangian method is obtained. The total dual function in vector form is hence

$$\phi(\lambda) = \min\{f(\mathbf{x}) + \frac{1}{2c}\{[\max(0, \mu + c\mathbf{h}(\mathbf{x}))]^2 - \mu^2\} + \lambda^T\mathbf{g}(\mathbf{x}) + \frac{1}{2}c\left\|\mathbf{g}(\mathbf{x})\right\|^2\} \tag{4.29}$$

or if one wishes to define the Lagrangian function instead

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \frac{1}{2c}\{[\max(0, \mu + c\mathbf{h}(\mathbf{x}))]^2 - \mu^2\} + \lambda^T\mathbf{g}(\mathbf{x}) + \frac{1}{2}c\left\|\mathbf{g}(\mathbf{x})\right\|^2 \tag{4.30}$$

with the optimum multipliers found iteratively based on steepest ascent as below.

$$\begin{aligned} \mu_{k+1} &= \max(0, \mu_k + c\mathbf{h}(\mathbf{x}(\mu_k))) \\ \lambda_{k+1} &= \lambda_k + c\mathbf{g}(\mathbf{x}(\lambda_k)) \end{aligned} \tag{4.31}$$

5 | Adaptive filters

An adaptive filter models the relationship between two signals in real time and relies for its operation on a recursive algorithm. The filter will be able to perform adequately in a setting where the characteristics of a relevant signal is unknown. Predetermined set of initial conditions is essential to start the algorithm. The block diagram below (figure 2.5) shows the structure of the adaptive filtering problem, where $u(n)$ is the input signal to the adaptive filter and $y(n)$ the output signal at time n. Comparison between the output signal and a desired response signal $d(n)$ is made. Subtracting the two signals a difference signal $e(n)$ is given which is known as the error signal.

$$e(n) = d(n) - y(n) \tag{5.1}$$

The error signal will alter the filter parameters till the output signal $y(n)$ of the adaptive filter hopefully becomes an improved match to the desired signal response $d(n)$ [28].

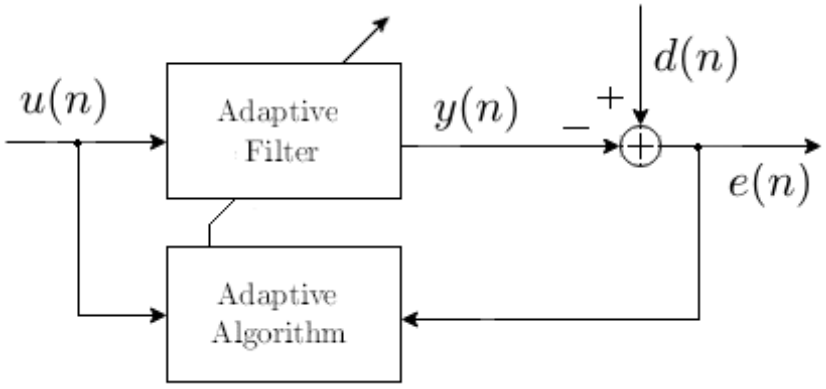


Figure 5.1: Block diagram representation of the adaptive filtering problem.

In this section a brief introduction of wiener filter and three different adaptive algorithms will be presented. Least mean squares (LMS), normalized least mean squares (NLMS) and recursive least squares (RLS). Also the method of adaptive noise cancellation will be explained.

5.1 Wiener Filter

In this section the design of linear optimum discrete-time filters, generally known as Wiener filters will be studied. In many applications the desired signal may be distorted and noisy. Wiener filter theory provides filters that are optimum in a sense that the best estimate of the desired signal will be produced. The assumption that is made is that the signal and the additive noise $v(n)$ are wide-sense stationary linear stochastic processes. The goal is to reproduce a desired signal $d(n)$ from noisy observations $u(n)$. Consider the block diagram in the figure below where an input $u(n)$ is filtered with $W(z)$ producing an estimate $\hat{d}(n)$

accompanied by error with statistical characteristics of its own. The estimation error $e(n)$ is the difference between the desired response $d(n)$ and the estimate. The objective is to minimize the estimation error in a statistical sense. To optimize the filter the mean-square error J as the cost function will be minimized [28].

$$\begin{aligned} u(n) &= d(n) + v(n) \\ e(n) &= d(n) - \hat{d}(n) \\ J &= E\{|e(n)|^2\} \end{aligned} \tag{5.2}$$

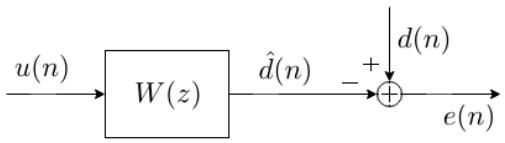


Figure 5.2: Block diagram representation of Wiener filtering problem.

For the development of Wiener filter theory in this section the finite impulse response (FIR) filter will be used since it is inherently stable, easy to implement, consists of only forward paths and are linear phase (no distortion of the phase). A choice of using infinite impulse response (IIR) filter could be made also, the choice depends on practical considerations. IIR filters include also feedback which can make the filter unstable and oscillate. Depending on how the desired signal is related to the noisy observation signal several problems could be solved [29]. These are some of the problems:

- **Filtering**
  The problem of estimating a desired signal $d(n)$ from noisy observations $x(n)$ using a causal filter. Current and past values of $d(n)$ is used.

- **Smoothing**
  This is the same as the filtering problem except for the fact that noncausal filter is used when estimating the desired signal.

- **Prediction**
  This is the problem of estimating future values of the desired signal based on linear combination of previous values of noisy observations $x(n)$. The Wiener filter in this case is called a linear predictor.

- **Deconvolution**
  If the noisy observation $x(n)$ and the desired signal $d(n)$ are related in the following way $x(n) = d(n) * g(n) + v(n)$ where $g(n)$ is the unit sample response of a linear shift-invariant filter the problem is called deconvolution. Estimating $g(n)$ is the main objective.

5.1.1 FIR Wiener Filter

Assume that the input and the desired signal are jointly WSS and denote the unit sample response of the Wiener filter $w(n)$, then the goal is to find the filter coefficients of the FIR

filter $w(k)$ that minimizes the mean square error. With $u(n)$ as the input to the filter $W(z)$ (*p-1* order filter) the estimate $\hat{d}(n)$ can be expressed as the convolution of $w(n)$ and $u(n)$. [28] [29]

$$\begin{aligned} \hat{d}(n) &= w(n) * u(n) = \sum_{l=0}^{p-1} w(l)u(n-l) \\ J &= E\{|e(n)|^2\} = E\{e(n)e^*(n)\} = E\{|d(n) - \hat{d}(n)|^2\} \end{aligned} \tag{5.3}$$

Taking the derivative of the cost function with respect to the filter coefficients equal to zero yields in

$$\frac{\partial J}{\partial w^*(k)} = \frac{\partial}{\partial w^*(k)} E\{e(n)e^*(n)\} = E\{e(n)\frac{\partial e^*(n)}{\partial w^*(k)}\} = 0 \tag{5.4}$$

Insertion of the estimate error in the derivate gives

$$E\{e(n)\frac{\partial}{\partial w^*(k)}[d^*(n) - \sum_{l=0}^{p-1} w^*(l)u^*(n-l)]\} = -E\{e(n)u^*(n-k)\} = 0 \tag{5.5}$$

and finally we have the following equation which is known as the ***orthogonality principle***.

$$E\{e(n)u^*(n-k)\} = 0 \quad k = 0, 1, \ldots, p-1 \tag{5.6}$$

Substituting the error e(n) in the above equation and using the fact that $r_u(k-l) = E\{u(n-l)u^*(n-k)\}$ and $r_{du}(k) = E\{d(n)u^*(n-k)\}$ gives the ***Wiener-Hopf equations*** below.

$$\sum_{l=0}^{p-1} w(l)r_u(k-l) = r_{du}(k) \quad k = 0, 1, \ldots, p-1 \tag{5.7}$$

Using matrix notation the Wiener-hopf equations can be expressed as

$$\mathbf{R}_u\mathbf{w} = \mathbf{r}_{du} \tag{5.8}$$

$$\begin{bmatrix} r_u(0) & r_u^*(1) & \ldots & r_u^*(p-1) \\ r_u(1) & r_u^*(0) & \ldots & r_u^*(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ r_u(p-1) & r_u^*(p-2) & \ldots & r_u(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \ldots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_{du}(0) \\ r_{du}(1) \\ \ldots \\ r_{du}(p-1) \end{bmatrix} \tag{5.9}$$

Applying the orthogonality principle in the same way as for the derivation of the Wiener-hopf equation one can arrive at the ***minimum mean square error*** as below.

$$J_{min} = r_d(0) - \sum_{l=0}^{p-1} w(l) r_{du}^*(l) = r_d(0) - \mathbf{r}_{du}^H \mathbf{R}_u^{-1} \mathbf{r}_{du} \tag{5.10}$$

The optimum tap-weight vector of the filter is computed by solving the Wiener-hopf equations.

$$\mathbf{w}_o = \mathbf{R}_u^{-1} \mathbf{r}_{du} \tag{5.11}$$

5.2 Least Mean Squares (LMS)

In the LMS algorithm the statistics of a signal is estimated continuously. This means that the correlation matrix $\mathbf{R}$ and the cross-correlation vector $\mathbf{p}$ is determined as below.

$$\begin{aligned} \hat{\mathbf{R}}(n) &= \mathbf{u}(n)\mathbf{u}^H(n) \\ \hat{\mathbf{p}} &= \mathbf{u}(n)d^*(n) \end{aligned} \tag{5.12}$$

The main principle is to minimize the mean-square error. It can be summarized in the following way where a desired output is computed by filtering the input with the tap-weight vector first. Then the estimation error is calculated as the difference between the desired signal and the filter output. Lastly the tap-weights are updated in accordance with the error.

$$\begin{aligned} \hat{d}(n) &= \hat{\mathbf{w}}^H(n)\mathbf{u}(n) \\ e(n) &= d(n) - \hat{d}(n) \\ \hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu \mathbf{u}(n) e^*(n) \end{aligned} \tag{5.13}$$

The mean square error can be expressed in the following by setting $\boldsymbol{\epsilon}(n) = \hat{\mathbf{w}}(n) - \mathbf{w_o}$ (where $\mathbf{w_o}$ is the optimal Wiener solution).

$$J(n) = E\{|e(n)|^2\} = J_{min} + J_{ex}(n) = E\{|e_o(n)|^2\} + E\{\boldsymbol{\epsilon}^H(n)\hat{\mathbf{R}}\boldsymbol{\epsilon}(n)\} \tag{5.14}$$

Convergence property of LMS is that the step-size parameter should be in the interval $0 < \mu < \frac{2}{\lambda_{max}}$, however since the eigenvalues are rarely known the tap-input power is used and the interval is rewritten as $0 < \mu < \frac{2}{Mr(0)}$ where M is the filter order and $r(0)$ the input variance. At best the LMS arrives at error $J_{min} + J_{ex}(\infty)$, so there will be a gradient noise. The misadjustment ($M = \frac{J_{ex}(\infty)}{J_{min}} = \sum_{i=1}^{M} \frac{\mu\lambda_i}{2-\mu\lambda_i}$) is a measure of how far away from the optimal solution the LMS is. This can also be approximated by $M = \frac{\mu}{2} Mr(0)$. [28]

5.3 Normalized Least Mean Squares (NLMS)

The difference between the standard LMS and the Normalized LMS is that the tap-weight vector is normalized by taking the squared Euclidian norm on the tap-input vector. This is

because the LMS suffers from gradient noise amplification when $u(n)$ is large, so to overcome this we normalize. A positive protective constant a is also added to the normalized tap-input vector in the denominator of the update equation in order to prevent the step size from being infinite. The method is the same as for the LMS, but the update equation now looks like this.

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\tilde{\mu}}{a + \|\mathbf{u}(n)\|^2}\mathbf{u}(n)e^*(n) \tag{5.15}$$

The step-size parameter varies now in the interval $0 < \tilde{\mu} < 2$. It can also be noted that the parameter is dimensionless whereas for the LMS algorithm it has the dimension of inverse power. [28]

5.4 Recursive Least Squares (RLS)

Recursive least square method is an adaptation of the least squares method that recursively finds the filter coefficient by minimizing a cost function. The input signals of the RLS algorithm are considered deterministic compared to the LMS which are stochastic. A benefit of the RLS compared to other algorithms is that it has extremely fast convergence. The recursive least square uses a value called the forgetting factor, λ. It is a decaying value making the algorithm forget about information in the past at a certain rate. If set to 1 the algorithm does not forget anything and thus uses all information to estimate its parameters. [28]

$$\begin{aligned} \hat{\mathbf{w}}(0) &= \mathbf{0} \\ \mathbf{P}(0) &= \delta^{-1}\mathbf{I} \end{aligned} \tag{5.16}$$

Initialization of the RLS algorithm is shown in equations (2.84) where δ is the regularization parameter. The parameter is assigned a small positive value for high SNR and large positive value for low SNR.

$$\mathbf{k}(n) = \frac{\mathbf{P}(n-1)\mathbf{u}(n)}{\lambda + \mathbf{u}^H(n)\mathbf{P}(n-1)\mathbf{u}(n)} \tag{5.17}$$

$$\xi(n) = d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n) \tag{5.18}$$

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\xi^*(n) \tag{5.19}$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{u}^H(n)\mathbf{P}(n-1) \tag{5.20}$$

The last equation above (2.88) is called the *Riccati equation* for the RLS algorithm, where $\mathbf{P}(n)$ is the inverse correlation matrix and $\mathbf{k}(n)$ is referred to as the *gain vector*. Equation (2.87) shows the desired recursive equation for updating the tap-weight vector, where ξ is the a priori estimation error.

5.5 Adaptive Noise Cancellation

The ANC consists of two inputs, the primary and a reference signal as shown in figure 2.6. The primary input is a mix of the signal from the source corrupted by noise (signal + noise). The reference input is either correlated with noise or the signal. If correlated with noise the output of the adaptive filter ($\hat{y}(n)$) are going to be a close estimate of the primary input noise. By subtracting the noise estimate from the corrupted signal the system output of the ANC will hence be an estimate of the signal. Similarly if the reference signal is correlated with the signal source the system output of the adaptive noise canceller is going to be an estimate of the noise. An adaptive process updates the filter coefficient based on the algorithm that is being used. [28]

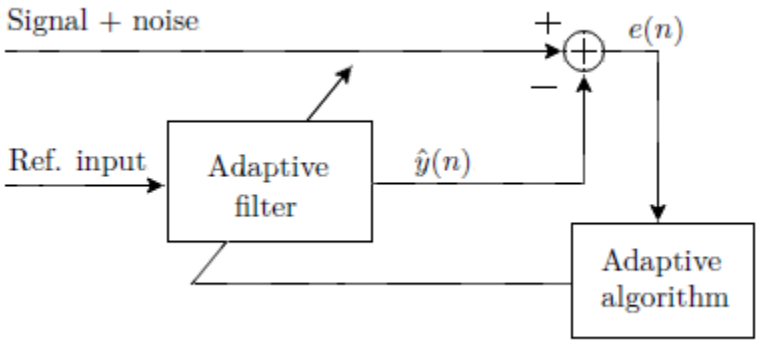


Figure 5.3: Block diagram representation of an Adaptive noise canceller.

6 | Independent Component Analysis (ICA)

ICA is the statistical method of which observed random multivariate data is linearly separated into a representation of independent non-gaussian subcomponents. This representation is very common in many applications such as feature extraction and signal separation. The general-purpose technique of ICA is used in many different areas of which some are biomedical signal processing, image processing, audio processing, telecommunication and econometrics [26]. A motivative example and illustration of ICA is the "cocktail party problem", where people are talking simultaneously in a room and the underlying speeches can be recovered. In the room microphones are placed at different locations and each time recording holds a different combination of the speakers' voices. In order to formalize the problem consider three people speaking in a room simultaneously and denote the emitted speech from each speaker as $s_1(t), s_2(t)$ and $s_3(t)$. Given three microphones the recorded signals ($x_1(t), x_2(t)$ and $x_3(t)$) are weighted summations of the emitted speech from each speaker as following

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \end{aligned} \tag{6.1}$$

where a_{ij} are parameters that depends on the microphone locations relative to the speaker. The assumption made is that neither the parameters a_{ij} nor the source signals $s_j(t)$ are known. The problem is to reconstruct the source signals by only knowing the mixes $x_i(t)$. This is done by using the statistical information of the source signals in order to estimate the mixing parameters a_{ij} and consequently simply inverting the linear system to obtain the source signal estimates. An adequate assumption is that the source signals $s_j(t)$ are statistically independent or at least mildly and this is realistic in most applications [45]. ICA can also be used to separate motion artifacts from biomedical signals such as the photoplethysmographic signal or electrocardiography (ECG) signal as research emphasizes [43] [44]. The mathematical framework of ICA algorithm depicted in the box of the figure below will be presented in this section. Firstly the definition and rigorous mathematical framework will be presented, thereafter why centering and whitening of the variables is necessary and lastly how the gradient algorithm for estimating the source signals is developed by measure of non-gaussanity.

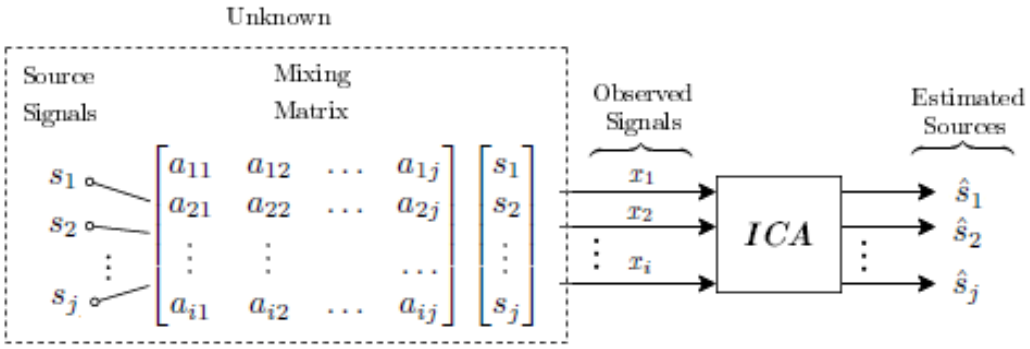


Figure 6.1: ICA model of estimating source signals from observed mixtures.

6.1 Definition

To define the ICA model a statistical latent variable model is used where a linear combination of random variables is present. The observed mixed signals is denoted as x_i and are linear combination of statistically independent components s_j as below [46].

$$x_i = \sum_{j=1}^{n} a_{ij} s_j, \quad i = 1, \ldots, n \tag{6.2}$$

Using vector-matrix notation which is convenient for ICA the mixing model is instead written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{6.3}$$

where $\mathbf{A}$ is a mixing matrix with the mixing elements as in figure 2.5 and $\mathbf{x}$ and $\mathbf{s}$ random vectors containing observed mixtures and independent components respectively. This basic model neglects time delays and noise for simplicity. In order to simplify calculations the mixing matrix is assumed to be square which means that the number of observed signals are equal to the number of independent components. As mentioned before the goal is to estimate the independent components when only the observed mixtures are known. Two main assumptions stated below are made to obtain the independent components estimates.

- **Statistically mutually independent components**
  A sufficient and important assumption that at each time instant the source signals must be statistically independent. From the basic definition independence can be described as a factorization of the joint density into the product of the marginal densities of the source signals.

- **Nongaussian independent components**
  ICA only works with nongaussian independent components. Components having gaussian probability distribution cannot be estimated. Beforehand the distribution of the independent components is often unknown and if some components are gaussian while others are

nongaussian ICA still works. The nongaussian components can be estimated, however the gaussian components will not be recovered. To see why gaussian components are prohibited consider first the joint distribution of two independent components s_1 and s_2 having nongaussian distribution (uniform distribution in this case) in the figure below.

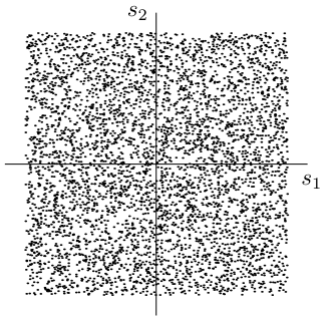


Figure 6.2: Joint distribution of nongaussian independent components s_1 and s_2.

Mixing the independent components with some arbitrary mixing matrix yields in two observed mixtures x_1 and x_2. It turns out that the source signals can be estimated by viewing the joint distribution of the mixtures. The edge of the parallelogram is in the directions of the columns of the mixing matrix. Hence the source signals can be extracted when having nongaussian independent components [47].

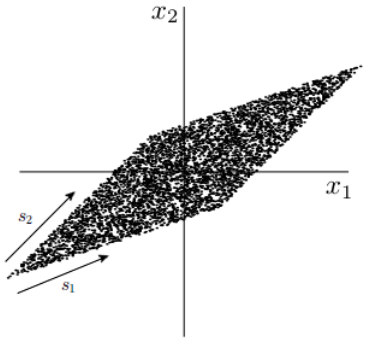


Figure 6.3: Joint distribution of observed mixtures x_1 and x_2.

Now let's consider two gaussian independent components instead, assume that the mixing matrix is orthogonal then it is easy to show that the joint distribution after mixing has the exact same distribution as the original one. The distributions in the picture below is symmetric and do not contain information of the directions of the columns of the mixing matrix.

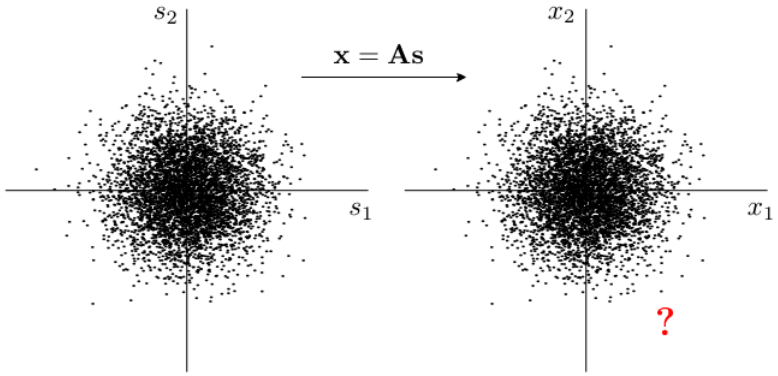


Figure 6.4: Joint distribution of gaussian independent components and of mixtures.

6.2 Centering

One of the preprocessing steps in ICA is to center the variables, this means that the sample mean $E\{\mathbf{x}\}$ is subtracted from the observed variables $\mathbf{x}$. This simplifies the calculations and the algorithms of ICA and both the independent components and the mixture variables have zero mean. [26]

6.3 Whitening

Whitening is also beneficial in the sense that it makes the components uncorrelated and of unit variance. Consequently the mixing matrix also becomes orthogonal which reduces the complexity and allows a more efficient and robust performance of the methods [26]. A well-known and common way to whiten the data is to use the eigenvalue decomposition (EVD) of the covariance matrix. The whitening matrix from EVD is

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \tag{6.4}$$

where $\mathbf{D}$ is the diagonal matrix containing eigenvalues of covariance matrix and $\mathbf{E}$ is the orthogonal eigenvector matrix. Multiplying this whitening matrix in the mixing model in equation 2.81 one obtains a new transformation $\mathbf{z}$ and a new mixing matrix $\tilde{\mathbf{A}}$ that is orthogonal.

$$\mathbf{z} = \mathbf{VAs} = \tilde{\mathbf{A}}\mathbf{s} \tag{6.5}$$

6.4 Non-gaussianity

As mentioned in the previous section nongaussanity is crucial when estimating independent components and can therefore be used in the ICA estimation. The independent components can be estimated by finding the inverse mixing matrix as following.

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \tag{6.6}$$

The central limit theorem can be applied in order to estimate ***one*** of the independent components. The basic concept is that one considers a linear combination of the mixture variables that results in one of the independent components. Given the following relationship

$$y = \mathbf{w}^T\mathbf{x} = \sum_i w_i x_i \tag{6.7}$$

where y is the one of the independent components and $\mathbf{w}$ is a vector that equals one of the rows of the inverse of $\mathbf{A}$. The equation above can be rewritten as a linear combination of the independent components instead in order to find such a vector.

$$y = \mathbf{w}^T\mathbf{x} = \mathbf{q}^T\mathbf{s} \tag{6.8}$$

Here $\mathbf{q}$ is equal to $\mathbf{A}^T\mathbf{w}$ and the vector contains only zeros except for one element which is equal to 1. The central limit theorem states that the sum of independent random variables is more gaussian than the original variables. In the case when one has the independent component $\mathbf{q}^T\mathbf{s}$ will be least gaussian. By maximizing the nongaussianity of $\mathbf{w}^T\mathbf{x}$ (since in practice the vector $\mathbf{q}$ and s_i is unknown) an independent component will be found [26]. The question is now how to define and measure nongaussianity, there are two common types of which one is kurtosis and the other negentropy. These will be described in the next section.

6.4.1 Kurtosis

Kurtosis is a measure of nongaussianity and is based on the fourth-order cumulant of a random variable. It is defined as following

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{6.9}$$

where y is some random variable. The fourth moment ($E\{y^4\}$) equals $3(E\{y^2\})^2$ for a gaussian variable, thus the kurtosis is zero for a gaussian random variable. For nongaussian variables the kurtosis is nonzero and greater than zero for most. As a measure of nongaussianity the absolute value is used and one can assume that the variables are normalized which means that variance equals one, hence the optimization landscape is restricted on the unit sphere [26]. In order to maximize the absolute value of the kurtosis lets assume that whitened data $\mathbf{z}$ has been used. The gradient algorithm obtained is as follows

$$\frac{\partial|\text{kurt}(\mathbf{w}^T\mathbf{z})|}{\partial\mathbf{w}} = 4\text{sign}(\text{kurt}(\mathbf{w}^T\mathbf{z}))[E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} - 3\mathbf{w}\,\|\mathbf{w}\|^2] \tag{6.10}$$

$$\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

where the vector w is projected on the unit sphere in each iteration. The vector w is moving in the direction that gives the maximum absolute value of kurtosis, since it is the direction

that is important then the change in norm is not significant and the latter term in equation 6.10 can be omitted. The gradient algorithm can then be written as

$$\begin{aligned} \Delta \mathbf{w} &\propto \text{sign}(\text{kurt}(\mathbf{w}^T\mathbf{z}))E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|} \end{aligned} \tag{6.11}$$

Kurtosis is sensitive to outliers and not very robust and that is why negentropy is introduced in the section below.

6.4.2 Negentropy

Negentropy is based on information theory and a measure called differential entropy. A random variable tending to be more unpredictable is said to have a larger entropy. Specifically gaussian variables are said to have the largest entropy, thus entropy can be used to define nongaussianity. Differential entropy is defined as following

$$H(\mathbf{y}) = -\int p_y(\eta) \log p_y(\eta) d\eta \tag{6.12}$$

where $\mathbf{y}$ is a random vector and $p_y(\eta)$ its probability density function. Negentropy is the normalized version of differential entropy which is nonnegative and zero only when $\mathbf{y}$ is of gaussian type [26]. It is defined below where $\mathbf{y}_{gauss}$ is a gaussian variable having the same statistical properties (correlation and covariance) as $\mathbf{y}$.

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{6.13}$$

In practice it is difficult to calculate negentropy because it would require an estimate of the pdf, therefore approximating negentropy is a good idea. The approximation is

$$J(\mathbf{y}) \approx \sum_{i=1}^{p} \rho_i [E\{G_i(y)\} - E\{G_i(\nu)\}]^2 \tag{6.14}$$

where ρ_i is positive constants, G_i nonquadratic functions and ν a zero mean gaussian variable with unit variance. For robust estimation it has been proven that the following nonquadratic functions should be used.

$$\begin{aligned} G_1(y) &= \frac{1}{a_1} \log \cosh a_1 y \quad \text{where} \quad 1 \leq a_1 \leq 2 \\ G_2(y) &= -\exp(-y^2/2) \end{aligned} \tag{6.15}$$

Considering only one nonquadratic function in the approximation the gradient algorithm can be derived by taking the gradient of the approximation with respect to $\mathbf{w}$.

$$\begin{aligned} \Delta \mathbf{w} &\propto \gamma E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|} \\ \Delta \gamma &\propto (E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\}) - \gamma \end{aligned} \tag{6.16}$$

In the above learning rule the normalization $\|\mathbf{w}\|^2 = 1$ has been taking into account and where $\gamma = E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\}$ and g is the derivative of function G.

6.4.3 Constrained Independent Component Analysis

There are ambiguities of ICA such as the indeterminacy of the order and variance of the resulting independent components. Constrained independent component analysis (cICA) is a method that combines ICA and constrained optimization in order to avoid arbitrary ordering on components [48]. By using reference signals and incorporating prior information desired signals can be extracted and ordered according to some statistical measure. The following constrained optimization problem with the objective function based on negentropy is used when searching for a desired signal.

$$\begin{aligned} \text{maximize} \quad & J(\mathbf{y}) \approx \rho[E\{G(\mathbf{w}^T \mathbf{x})\} - E\{G(\nu)\}]^2 \\ \text{subject to} \quad & g(\mathbf{w}) = \varepsilon(y, r) - \xi \leq 0 \\ & h(\mathbf{w}) = E\{y^2\} - 1 = 0 \end{aligned} \tag{6.17}$$

$\varepsilon(y, r)$ denotes the closeness measure between the reference signal and the estimated output, ρ is a positive constant. The closeness measure must be chosen in a suitable way, two common ways is correlation and mean square error by setting $\varepsilon(y, r) = 1/(E\{y \cdot r\})^2$ or $\varepsilon(y, r) = E\{(y - r)^2\}$ respectively. The threshold parameter ξ is in the region $[\varepsilon(\mathbf{w}^{*T}\mathbf{x}, r),\ \varepsilon(\mathbf{w}_1^T \mathbf{x}, r)]$ where $\mathbf{w}^*$ is the optimum vector when only one of the independent components (the desired source) is the closest to the reference $(\varepsilon(\mathbf{w}^{*T}\mathbf{x}, r) < \varepsilon(\mathbf{w}_1^T \mathbf{x}, r) \leq \ldots \leq \varepsilon(\mathbf{w}_{n-1}^T \mathbf{x}, r))$ [49]. A constraint $g(\mathbf{w}) = \varepsilon(y, r) - \xi \leq 0$ can therefore be constructed such that the closeness measure is less than or equal to the threshold value only when the independent component is the optimum ($y = \mathbf{w}^{*T}\mathbf{x}$). The constraint $h(\mathbf{w})$ makes the weight vector and $J(\mathbf{y})$ bounded. By introducing slack variables in order to transform inequality constraints to equality constraints and using the augmented lagrangian function theory the following lagrangian is defined [50].

$$L(\mathbf{w}, \mu, \lambda) = J(y) - \frac{1}{2\gamma}\{[max(0, \mu + \gamma g(\mathbf{w}))]^2 - \mu^2\} - \lambda h(\mathbf{w}) - \frac{1}{2}\gamma \|h(\mathbf{w})\|^2 \tag{6.18}$$

Newton's method in optimization can be used to maximize the Lagrangian function above, the learning rule is shown below

$$\begin{aligned} & \mathbf{w}_{k+1} = \mathbf{w}_k - \eta (L''_{\mathbf{w}_k^2})^{-1} L'_{\mathbf{w}_k} \\ & \text{where } L'_{\mathbf{w}_k} = \pm \rho E\{\mathbf{x} G'_y(y)\} - \frac{1}{2}\mu E\{\mathbf{x} g'_y(\mathbf{w})\} - \lambda E\{\mathbf{x} y\} \\ & \text{and } L''_{\mathbf{w}_k^2} = s(\mathbf{w}) \mathbf{R}_{\mathbf{xx}} \end{aligned} \tag{6.19}$$

where $s(\mathbf{w}) = \pm\rho E\{\mathbf{x}G^{''}_{y^2}(y)\} - \frac{1}{2}\mu E\{\mathbf{x}g^{''}_{y^2}(\mathbf{w})\} - \lambda$, $\mathbf{R_{xx}}$ is the covariance matrix and η a learning rate constant. The final algorithm in order to find the weight vector that will give the desired component together with the optimum multipliers that are updated iteratively using gradient ascent method is hence

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \eta\mathbf{R}^{-1}_{\mathbf{xx}}L^{'}_{\mathbf{w}_k}/s(\mathbf{w}_k) \\ \mu_{k+1} &= max\{0, \mu_k + \gamma g(\mathbf{w}_k)\} \\ \lambda_{k+1} &= \lambda_k + \gamma h(\mathbf{w}_k) \end{aligned} \tag{6.20}$$

7 | Methods

In this section three methods will be presented and described in a detailed way. The proposed methods uses adaptive noise cancellation in order to remove motion artifacts (MA). Adaptive noise cancellation consisting of adaptive filters that updates the weight vector based on an adaptive algorithm (LMS, NLMS or RLS) is an effective method of dealing with motion artifacts, however a reference signal must be provided. The reference signal could be correlated with the PPG signal or the MA based on the adaptive noise canceler setting. As seen in fig. 7.1 (a) if the reference signal is correlated with the noise an enhanced signal will be available at the output ($e(n)$) when minimizing error power in the mean square sense. On the other hand if the reference signal is correlated with the signal part as in figure (b) then the enhanced signal will be at the primary output ($\hat{s}(n)$) A reference signal that is correlated with the MA could also be provided with the access to accelerometers, that however would require the installation of extra hardware. Reference signals will be synthetically generated in the proposed methods in different ways. In Peng's method cICA is used, in Yousefi's method the optical density theory and in our proposed method extraction of periodic signals using eigenvalue decomposition and fast fixed-point ICA is used in order to obtain a reference signal.

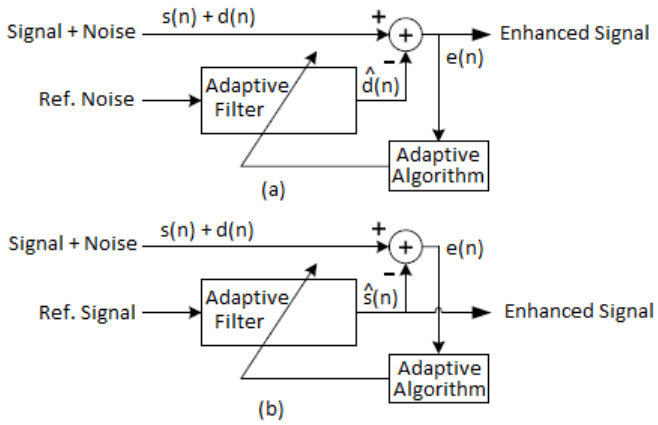


Figure 7.1: Picture showing different adaptive noise canceler settings.

7.1 Methods

7.1.1 Preprocessing

The PPG signal contain unwanted high frequency noise which is the result of ambient light, thermal noise or power interference/electromagnetic noise. In order to suppress the high frequency components low-pass filtering of the PPG signal is done using linear-phase FIR filter with hamming window in Matlab (MathWorks Inc.). The cutoff frequency was set to 8 Hz with 20 dB attenuation with a filter order of 13 to remove the bulk of noise. The impulse response of the FIR filter is shown below

$$h(n) = h_{lowpass}(n) \cdot w_{hamming}(n) = \frac{w_c}{\pi} \frac{\sin w_c(n - \frac{M-1}{2})}{w_c(n - \frac{M-1}{2})} \cdot (0.54 - 0.46 \cos \frac{2\pi n}{M-1}) \quad (7.1)$$

where w_c is the normalized cutoff frequency and M the filter order. In order to remove the baseline wander (DC part) an IIR-filter with 20 dB attenuation is used with transfer function below.

$$H(z) = \frac{1 - z^{-1}}{1 - 0.992z^{-1}} \quad (7.2)$$

In order to prevent phase shift due to the use of causal filters zero-phase digital filtering is implemented using the `filtfilt` function in Matlab. This function basically performs filtering in both forward and reverse directions and it can be shown that the spectrum of the output signal from the time-reversed second filter pass is given by

$$Y(e^{jw}) = |H(e^{jw})|^2 X(e^{jw}) \quad (7.3)$$

which shows that the signal is filtered with frequency response $|H(e^{jw})|^2$ that is real-valued and positive, hence no phase shift. In all three methods the same preprocessing step is made on the IR and R signals.

7.1.2 Yousefi's method

In the method proposed by Yousefi et al. [7] a two-stage adaptive noise cancellation setting is used where a reference signal is generated synthetically in each stage in order to extract a clean signal as possible. As seen from the block diagram in fig 7.8, the adaptive noise cancelers M1 and M2 at the first stage (1) uses a reference signal based on subtraction of IR and R PPG signals in order to remove unwanted tissue effect. In the second stage (4 in fig. 7.8), the adaptive noise cancelers N1 and N2 uses a synthetical noise reference signal associated with venous blood movement. The different parts of the algorithm will be described in detail in each section below where part 3 is about finding the venous noise reference using predictor filter and weighted subtraction based on the optical density theory that was described in Chapter 2.3 earlier.

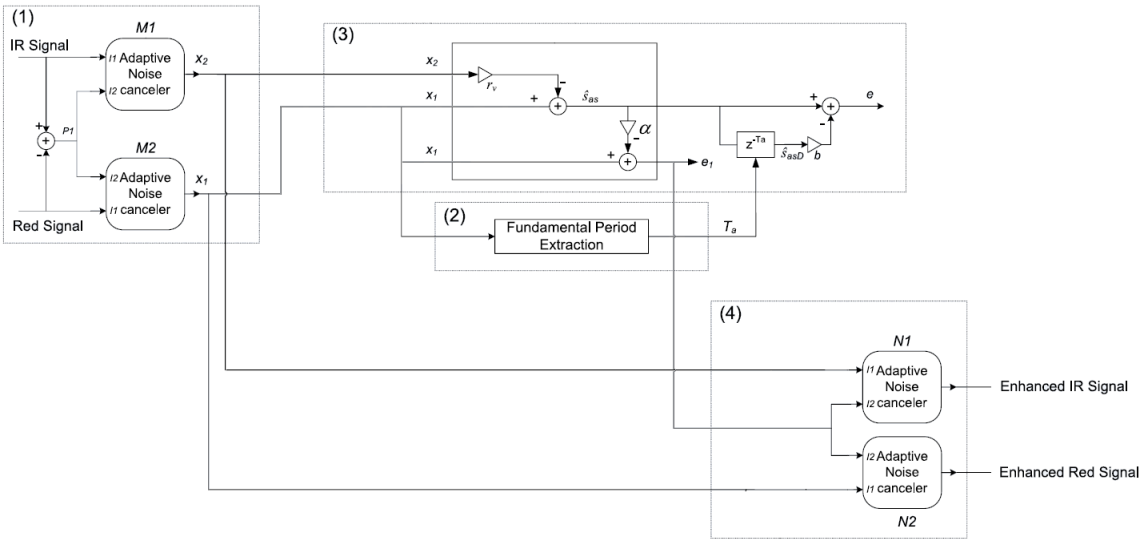


Figure 7.2: Block diagram of Yousefi's method.

(1) Reducing motion artifact due to the effect of tissue

During motion the effect of tissue is claimed to generate undesired fluctuations in the PPG signal. The tissue effect in itself is manifested as DC component in the PPG signal, however during motion the component is not constant anymore. In order to cancel out the effect of tissue (Z_t) and (Z_b) that are wavelength independent in equation 2.10 the IR signal is subtracted from the R signal resulting in a corresponding optical density change as follows

$$\begin{aligned}\Delta A_b = &(\sqrt{\epsilon_{h_{a_{IR}}}(\epsilon_{h_{a_{IR}}}+F)} - \sqrt{\epsilon_{h_{a_R}}(\epsilon_{h_{a_R}}+F)}) \cdot c_{totHb_a} \cdot \Delta l_{b_a} \\ &+ (\sqrt{\epsilon_{h_{v_{IR}}}(\epsilon_{h_{v_{IR}}}+F)} - \sqrt{\epsilon_{h_{v_R}}(\epsilon_{h_{v_R}}+F)}) \cdot c_{totHb_v} \cdot \Delta l_{b_v}\end{aligned} \tag{7.4}$$

where one can notice that the term ΔA_s is removed from equation 2.11. This reference signal that is now associated with blood pulsation is used in adaptive noise cancelers M1 and M2 (using NLMS as adaptive algorithm) which has the setting as in figure 7.1 b). Thus enhanced red and infrared signal after the removal of tissue effect and adaptive noise cancellation will be present in signals x_1 and x_2 in figure 7.8.

(2) and (3) - Design of venous noise reference signal

Extraction of the fundamental period in (2) is done in the same way as in Peng's method described in part 1 of the method. This fundamental period T_a will be used in a prediction filter in order to minimize the prediction error that will give a clean arterial source signal. During motion venous blood movement is a source of interference and should therefore be cancelled. Designing a venous noise reference signal that will be fed into the adaptive filters N1 and N2 will result in enhanced IR and R signals at the output with motion artifact due

to venous blood movement removed. The blood absorption or the optical density for IR and R signals can be described as

$$\begin{aligned}\Delta A_{IR} &= \sqrt{\epsilon_{h_{a_{IR}}}(\epsilon_{h_{a_{IR}}}+F)}\cdot c_{totHb_a}\cdot \Delta l_{b_a} \\ &+ \sqrt{\epsilon_{h_{v_{IR}}}(\epsilon_{h_{v_{IR}}}+F)}\cdot c_{totHb_v}\cdot \Delta l_{b_v} \\ \Delta A_{R} &= \sqrt{\epsilon_{h_{a_{R}}}(\epsilon_{h_{a_{R}}}+F)}\cdot c_{totHb_a}\cdot \Delta l_{b_a} \\ &+ \sqrt{\epsilon_{h_{v_{R}}}(\epsilon_{h_{v_{R}}}+F)}\cdot c_{totHb_v}\cdot \Delta l_{b_v}\end{aligned} \tag{7.5}$$

Weighted subtraction of the above optical densities will be used in order to design the venous noise reference, $\Delta A_v = \Delta A_{IR} - \beta \Delta A_R$. Each recording has an arterial source signal and a venous source signal as represented by subscripts a and v respectively in the equation above. The weighted subtraction can also be rewritten with the optical density ratios r_a and r_v as follows

$$\begin{aligned}\Delta A_v &= (r_a-\beta)\sqrt{\epsilon_{h_{a_{IR}}}(\epsilon_{h_{a_{IR}}}+F)}\cdot c_{totHb_a}\cdot \Delta l_{b_a} \\ &+ (r_v-\beta)\sqrt{\epsilon_{h_{v_{IR}}}(\epsilon_{h_{v_{IR}}}+F)}\cdot c_{totHb_v}\cdot \Delta l_{b_v} \\ r_a &= \sqrt{\epsilon_{h_{a_{IR}}}(\epsilon_{h_{a_{IR}}}+F)}/\sqrt{\epsilon_{h_{a_{R}}}(\epsilon_{h_{a_{R}}}+F)} \\ r_v &= \sqrt{\epsilon_{h_{v_{IR}}}(\epsilon_{h_{v_{IR}}}+F)}/\sqrt{\epsilon_{h_{v_{R}}}(\epsilon_{h_{v_{R}}}+F)}\end{aligned} \tag{7.6}$$

As suggested and developed by Masimo [54] the above equations can also be modelled in the alternative way using the optical density ratios for arterial and venous signal as follows

$$\begin{aligned}x_1 &= r_a s_a + r_v s_v \\ x_2 &= s_a + s_v\end{aligned} \tag{7.7}$$

where x_1 and x_2 corresponds to ΔA_R and ΔA_{IR} respectively, s_a and s_v are arterial source signal and venous source signal. The weighted subtraction can now be expressed in the following way $x_v = x_1 - \beta x_2 = (r_a - \beta)s_a + (r_v - \beta)s_v$. An interesting observation here is that if the term β is properly tuned then one can obtain either the arterial source or the venous source signal separately. There are several conditions now that can be examined, let's assume that there is motion present in the signal. When sweeping through range of β and arriving at $\beta = r_v$ the weighted subtraction x_v will contain the arterial source. If arriving at $\beta = r_a$ then x_v will have the venous source only. There exists a wider range of β where the weighted subtraction term contain only the arterial component than for the venous component. Extracting a signal that only consists of the venous component allows for the use as a reference signal in adaptive noise cancelers in step 4 of figure 7.8. In order to find the venous reference noise consider the structure below.

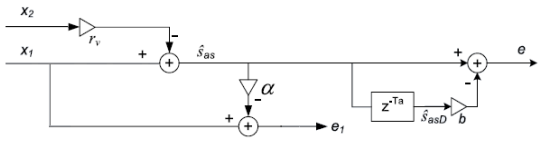


Figure 7.3: Block diagram of the structure when finding the venous noise reference.

A linear FIR predictor filter $B(z) = bz^{-T_a}$ with the fundamental period extracted is used in order to predict future values of the scaled arterial component $\hat{s}_{as}$. The goal is to minimize the the error $e(n) = \hat{s}_{as}(n) - b \cdot \hat{s}_{as}(n - T_a)$ in the mean square sense to receive a close estimate of the arterial signal. Optimization of r_v and b is done by taking the gradient of the cost function below with respect to each sought parameter.

$$J = E\{e(n)^2\} = E\{\hat{s}_{as}^2(n)\} - 2bE\{\hat{s}_{as}(n)\hat{s}_{as}(n - T_a)\} + b^2E\{\hat{s}_{as}^2(n - T_a)\} \tag{7.8}$$

When the prediction error is minimized by solving the system of equations the following optimized parameters are obtained.

$$r_v = \frac{-E\{x_1^2\}E\{\hat{s}_{asD}x_2\} + E\{x_1x_2\}E\{\hat{s}_{asD}x_1\}}{E\{x_2^2\}E\{\hat{s}_{asD}x_1\} - E\{x_1x_2\}E\{\hat{s}_{asD}x_2\}} \tag{7.9}$$

$$b = \frac{E\{\hat{s}_{as}(n)\hat{s}_{as}(n - T_a)\}}{E\{\hat{s}_{as}^2(n - T_a)\}} \tag{7.10}$$

Once the extraction of the scaled arterial signal estimate is done with the optimum parameters above the venous reference noise is obtained by removing this estimate (multiplying with α and subtracting from x_1) by minimizing the error e_1 in the mean square sense. The optimum value of α when the variance $E\{e_1^2\}$ is minimized is as follows

$$\alpha = \frac{E\{x_1^2\} - r_vE\{x_1x_2\}}{E\{\hat{s}_{as}^2\}} \tag{7.11}$$

(4) Reducing motion artifact due to venous blood change

The optimum values of r_v, b and α gives a proper estimate of the venous noise motion reference e_1 used as a reference signal in adaptive noise cancelers N1 and N2. The reference signal is correlated with MA and hence the output will be enhanced IR and R signals after minimizing the error power. NLMS is used as the adaptive algorithm for the noise cancelers.

7.1.3 Peng's method

This implementation is based on the algorithm developed by Peng et al. [6] with some slight modifications and improvements. Basically what differs is the fundamental period extraction where fast fourier transform is used instead of autocorrelation and the adaptive filter settings where RLS and NLMS is used also. The method combines constrained independent

component analysis (cICA) together with adaptive filters in order to remove MA. In this method the reference signal will be synthetically generated using cICA and the information from the two-channel PPG signal. The ICA model ($\mathbf{x} = \mathbf{As}$) where $\mathbf{x}$ is the observed signal samples of IR and R, $\mathbf{A}$ is the mixing matrix and $\mathbf{s}$ is the independent source signal samples can also be written as following.

$$\begin{aligned} x_{IR}(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_R(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{aligned} \tag{7.12}$$

$s_1(t)$ is the component representing the PPG signal and $s_2(t)$ is the independent noise component (motion artifact). The reason why cICA is used and not ICA is because of the ambiguities such as indeterminacy of the order and the variance of the independent components. By using cICA a desired independent component can be estimated according to some prior information. This prior information is basically a reference signal (that needs to be generated somehow) with the same periodic information as the PPG signal serving as input to the cICA algorithm.

Block diagram of the algorithm

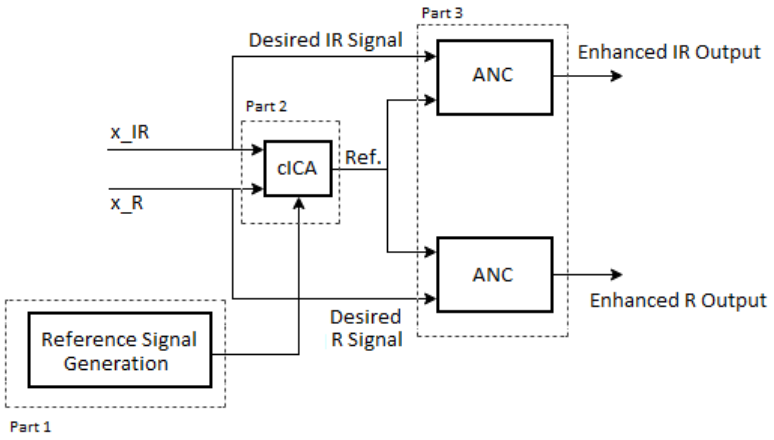


Figure 7.4: Picture showing the block diagram of the algorithm with the main parts.

As seen in fig. 7.2 there are three main parts of the algorithm, the reference signal generation, cICA algorithm and the adaptive noise cancellation part lastly. Assuming that each of the original IR and R signals recordings are mixed with motion artifact that is fed into the preprocessing step. The cICA output will be a PPG-correlated component, which will be discussed later on how to retrieve it, that serves as input to the adaptive filters. A more detailed explanation of the four parts of the algorithm is explained in the sections below.

Part 1 - Reference signal generation

Extracting a PPG-correlated component from cICA algorithm will require a reference signal with prior information about the interested component. The fundamental period is hence

needed in order to create a reference signal. Peng et al. [6] proposes a method of using rectangular pulse waves with the same period information obtained from the original PPG signals. Period extraction can be achieved using autocorrelation method suggested in [6][7]. However in this report Fast Fourier Transform (FFT) is used to estimate the period. FFT computes the discrete fourier transform in an efficient way. The approach is straightforward, by calculating the absolute value of the frequency representation (FFT) and finding the index of normalized frequency with maximum energy the period is found. Mathematically this can be described as following where x_n captured signal in time domain and X_k the N point fourier transform of the time series.

$$\max_k |X_k| = |\sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N}|, \quad k = 0, 1, \ldots, \frac{N-1}{2} \tag{7.13}$$

The fundamental frequency can then be calculated using the found index k, the sampling frequency F_s and the length N of the FFT as following, which will give the period T.

$$f = \frac{kF_s}{N} \leftrightarrow T = \frac{1}{f} \tag{7.14}$$

The generated rectangular pulse reference with the periodic information of the PPG signal is shown in fig. 7.3.

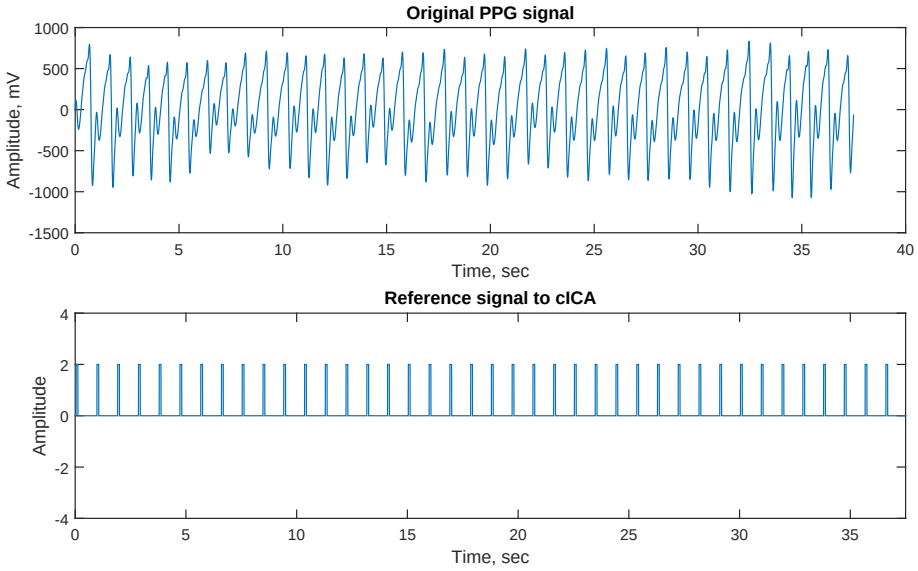


Figure 7.5: Rectangular reference signal used in cICA.

Part 2 - Constrained Independent Component Analysis (cICA)

The red and IR PPG signals are modelled as linear mixture $\mathbf{x}$ of motion artifact PPG signal source $\mathbf{s}$. The linear mixture is centered and whitened before applying the cICA algorithm.

The MA part is a combination of multiple sources and hence the use of cICA instead of ICA due to the fact that the former needs no assumption regarding underlying sources in order to extract a specific source. The goal is to find an optimum weight vector $\mathbf{w}$ that estimates the independent component, in this case the PPG-correlated component. As reference signal r the signal developed in part 1 will be used. The closeness measure is set to $\varepsilon(y, r) = E\{(y - r)^2\}$. The nonquadratic function G is chosen as $G(y) = \log\cosh y$. The iterative algorithm that is used when finding the optimum weight vector can be seen in eq. 6.20 in sec. 6.4.3.

The cICA output or the desired component that is generated is a PPG correlated component shown in figure below. This output will be fed into the adaptive filter for artifact cancellation. Constrained ICA manages to remove motion artifact but suffers in the way that the energy information of the original signal is lost. The amplitude information is hence not recovered. That is way the adaptive filters are used in order to recover the amplitude information.

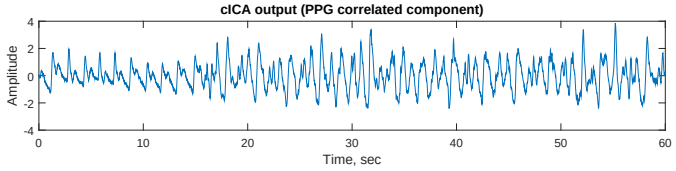


Figure 7.6: The output from the cICA part.

Part 3 - Adaptive noise cancellation

Adaptive noise cancellation is used in order to recover the amplitude information that is lost using cICA algorithm and as well as removing MA. The PPG correlated component (cICA output) is fed into the adaptive filters, minimizing error power yields in enhanced IR or R output from the adaptive filters. MA noise is the difference between the output from the adaptive filter and the desired signal that is fed back into the adaptive algorithm. As adaptive algorithms LMS, NLMS and RLS is used. Figure 7.5 shows a detailed view of part 3 in the implementation.

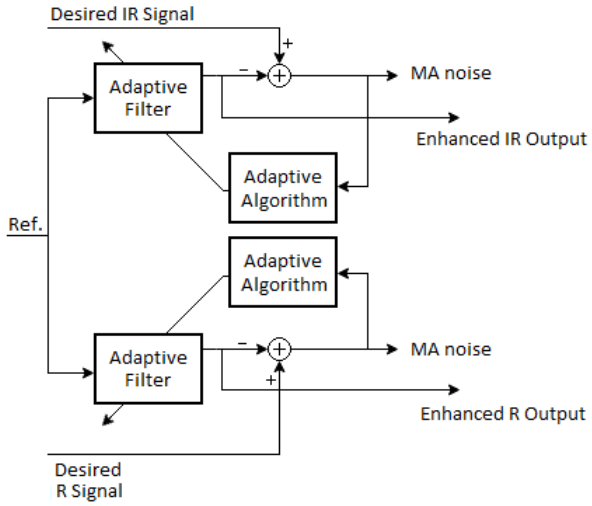


Figure 7.7: Detailed view of the adaptive noise cancellation of part 3.

7.1.4 Our proposed method

In this method we propose a reference signal generated using eigenvalue decomposition and fast fixed-point ICA. The PPG-correlated reference signal is then fed into adaptive noise cancelers and the output after minimization of the error results in enhanced signals.

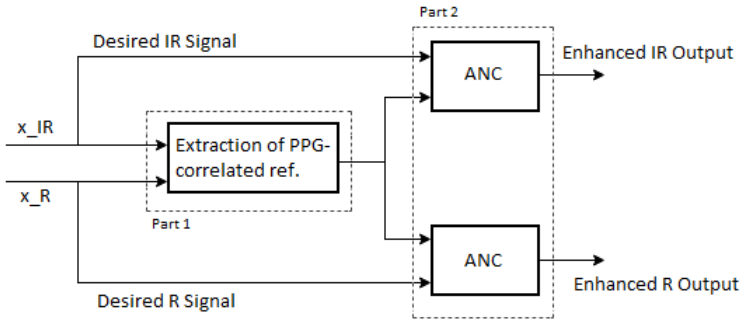


Figure 7.8: Block diagram of our method.

Reference signal generation

The reference signal generation used for this algorithm is proposed by Zhang et al. [51] and it is a two-stage based approach using eigenvalue decomposition in the first stage and fast fixed-point ICA in the second stage. Obtaining the PPG correlated component in this method only requires the estimated fundamental period unlike the cICA algorithm which needs a priori information such as the design of a reference signal closely related to the desired component. The PPG-correlated component (desired component) is fed into the adaptive filter serving as reference in order to suppress motion artifacts. Assuming that one

observes a vector $\mathbf{x} = \mathbf{As}$ with the IR and R recordings where $\mathbf{A}$ is unknow mixing matrix and $\mathbf{s}$ the independent sources (PPG and motion artifact). To find the PPG-correlated component one needs to find a vector $\mathbf{w}$ such that the scalar $y = \mathbf{w}^T\mathbf{x}$ is estimating the desired component. The observed vector is assumed to be whitened and centered before the algorithm.

First stage

The idea behind the proposal is to find P suitable time delays and use the definition of autocorrelation in order to roughly extract the desired signal. The time delays are estimated using the extracted period information and selecting peaks from autocorrelation. Assume that the desired signal s_1 is non-Gaussian and periodic with τ_0 with the following relationships satisfied

$$\begin{aligned} E\{\sum_{p=1}^{P} s_1(k)s_1(k-l_p\tau_0)\} > 0 \\ E\{\sum_{p=1}^{P} s_j(k)s_j(k-l_p\tau_0)\} = 0 \quad \forall j \neq 1 \end{aligned} \tag{7.15}$$

where l_p is positive integers. The relationship above tell us that the autocorrelation for the desired signal s_1 is large while it is zero (or very small) for other source signals. This information can be used in order to find the desired signal, by maximizing the autocorrelation of the output signal $y(k) = \mathbf{w}^T\mathbf{x}(k)$ as following

$$\max\ J(\mathbf{w}) = E\{y(k)y(k-l_p\tau_0)\} \tag{7.16}$$

under the constraint $\|w\| = 1$ then the output signal $y(k)$ estimates the desired signal. By rewriting the objective function above using $y(k) = \mathbf{w}^T\mathbf{x}(k)$, the following optimization problem is obtained with the mixing matrix involved.

$$\begin{aligned} \max\ J(\mathbf{w}) &= \frac{1}{2}J(\mathbf{w}) + \frac{1}{2}J(\mathbf{w})^T \\ &= \frac{1}{2}\mathbf{w}^T E\{\mathbf{x}(k)\mathbf{x}(k-l_p\tau_0)^T\}\mathbf{w} + \frac{1}{2}\mathbf{w}^T E\{\mathbf{x}(k-l_p\tau_0)\mathbf{x}(k)^T\}\mathbf{w} \\ &= \frac{1}{2}\mathbf{w}^T E\{\mathbf{R}_x(l_p\tau_0) + \mathbf{R}_x(l_p\tau_0)^T\}\mathbf{w} \end{aligned} \tag{7.17}$$

Under the constraint $\|w\| = 1$ the above optimization problem is equivalent of finding the eigenvector that belongs to the maximal eigenvalue of $\mathbf{R}_x(l_p\tau_0) + \mathbf{R}_x(l_p\tau_0)^T$. Hence the final algorithm of finding the reference signal is as follows.

$$\begin{aligned} &\mathbf{w} = EIG(\mathbf{R}_x(l_p\tau_0) + \mathbf{R}_x(l_p\tau_0)^T) \\ &\text{where } \mathbf{R}_x(l_p\tau_0) = E\{\mathbf{x}(k)\mathbf{x}(k-l_p\tau_0)^T\} \end{aligned} \tag{7.18}$$

EIG is the operator that calculates the eigenvector that corresponds to the maximal eigenvalue.

Second stage

In this second stage of the two-based approach the fast fixed-point one unit ICA algorithm is used in order to make the solution even closer to the optimum one by reducing noise and extracting a cleaner signal. The initial weight vector is the estimated $\mathbf{w}$ from the first stage.

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^T\mathbf{x})^3\} - 3\mathbf{w} \\ \mathbf{w} &= \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|} \end{aligned} \tag{7.19}$$

Running the algorithm on PPG signals the following extracted reference signal $y = \mathbf{w}^T\mathbf{x}$ is obtained as in the fig. 7.7.

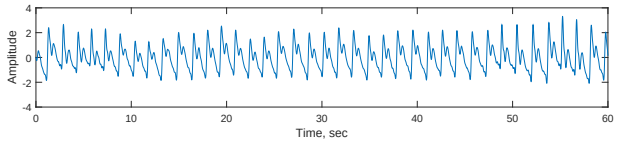


Figure 7.9: The generated reference signal in our method.

7.2 Material

7.2.1 CardioHolter

The CardioHolter6.2 is an ambulatory medical device and is only used in clinical purposes. During use in hemodialysis the Cardioholter can be on for 8 hours straight. The data that are recorded with this device is only meant for research, so it cannot be used for patient monitoring or in hospitals. The data that this device are capable of recording are ECG signal, PPG signal, temperature, movements, breathing and time. As shown in figure 3.1 below, the Cardioholter is designed with a main electronic unit and attached to this unit is four ECG cables with connector clips, three pulse oximeter sensors (PPG) where two of the sensors are meant to be placed on left and right finger and the third sensor on the forehead with a head belt. There is a power button on the front of the main unit and a status indicating LED.

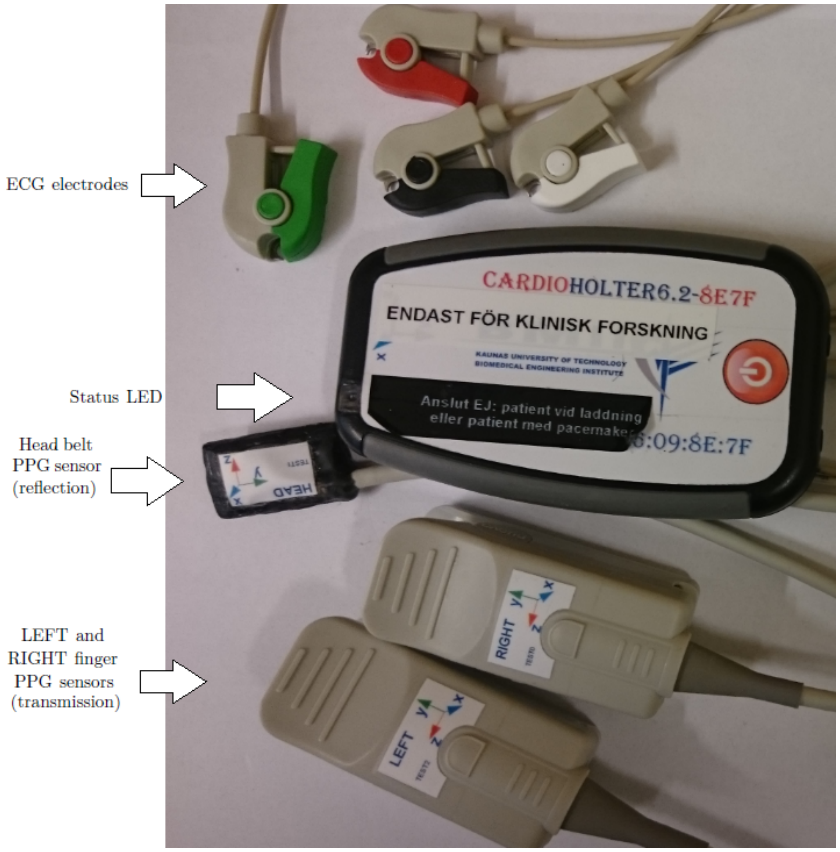


Figure 7.10: Picture showing the main parts of the CardioHolter 6.2.

As shown in figure 3.2 below, the CardioHolter have three input ports. One input port for charging the battery, one for isolated input of digital synchronization signal and a third input port for inserting of a microSD memory card. Next to the memory slot there is a charging LED indicator.

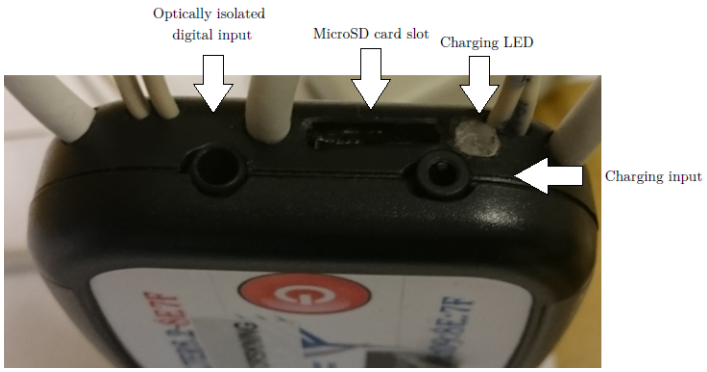


Figure 7.11: Picture showing the CardioHolters I/O.

The photopethysmography data recoded by the CardioHolter is collected from the index fingers and the forehead. When recording the system operates in pulsed mode by reason of saving power. It means that to lit the RED and IR LED's very short pulses (100ms) are used. The pulses have a sampling rate of 250 times/s. The CardioHolter records the PPG data at sampling frequency of 250 Hz.

7.2.2 CardioLogger

The application that is used for data collection is called CardioLogger (developed at Kaunas Technical University, Lithuania). In figure 3.3 below the interface of the CardioLogger application is shown. The program connects to the pulse oximeter (CardioHolter v6.2) using bluetooth. Raw signal data is saved locally when turning the system off in a .dat file. This file is converted to .mat file in Matlab for further processing.

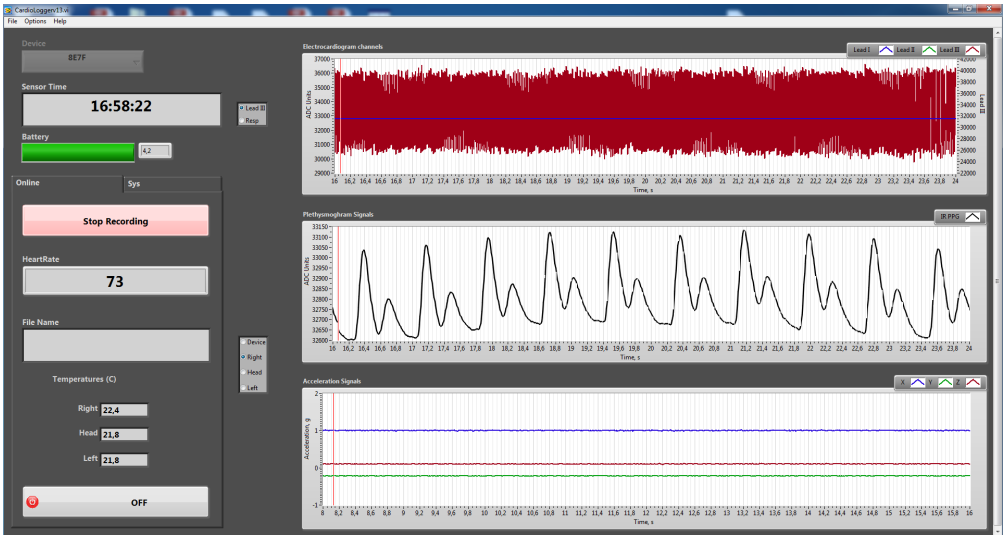


Figure 7.12: User interface of the CardioLogger program.

Data collection of motion and PPG signals

In order to evaluate the algorithms data was collected from two healthy males aged 24 using the CardioHolter device and the CardioLogger software. In order to make statistical performance evaluation motion was generated and added to a motionless PPG recording. Four types of motion was considered, bending the finger, horizontal movement of the finger, vertical movement of the finger and waving hand. Each total PPG recording is set to approximately 1 min. A target PPG signal without motion was captured wearing the probe on the index finger. Data acquisition of motion was generated using the finger probe to create the artifacts during the different types of movements mentioned. The mixed signal $x(k)$ is then used for performance evaluation where $s(k)$ is the target PPG signal and $n(k)$ the generated motion artifact that is added.

$$x(k) = s(k) + \sigma n(k) \tag{7.20}$$

The parameter σ is altered in order to define the portion of motion in the signal. The generated motion for the different types of movements was collected for different time periods. A 10-second, 20-second, 30-second and 40-second interval of motion was recorded for each type.

7.3 Performance evaluation

7.3.1 Signal-to-noise ratio (SNR)

In order to evaluate the quality and performance of the algorithms the signal-to-noise ratio (SNR) value is calculated. It is a common measure of the signal power to the amount of noise power occupied in the signal. The input SNR is defined as follows

$$SNR_{input} = 20 \log \frac{\text{RMS}(s(k))}{\text{RMS}(\sigma n(k))} \tag{7.21}$$

where RMS is the abbreviation of root mean square. After motion artifact reduction the residual noise can be obtained as the difference between the clean signal and the noise reduced signal $y(k)$. The output SNR can hence be calculated as follows

$$SNR_{output} = 20 \log \frac{\text{RMS}(s(k))}{\text{RMS}(s(k) - y(k))} \tag{7.22}$$

7.3.2 Relative root mean square error (RRMSE)

Another measure of how close the output signal $y(k)$ is to the clean wanted signal $s(k)$ is the relative root mean square error denoted as RRMSE. With this measure the efficiency of the algorithms can be calculated as

$$RRMSE = \frac{\text{RMS}(s(k) - y(k))}{\text{RMS}(s(k))} \times 100\% \tag{7.23}$$

where a smaller value indicates a better performance.

7.3.3 Relative magnitude of capillary pulse (RMCP)

Reflecting changes in the envelope of the signal can be computed using integrated PPG signal defined as

$$x(n) = \sum_{k=nK-L+1}^{nK} |p(k)| \tag{7.24}$$

where the above signal is a summation of the absolute values of PPG signal $p(k)$, K is the downsampling step and L is the moving average interval. The relative magnitude of capillary pulse (RMCP) estimates the PPG amplitude. In order to analyze amplitude difference for the methods the following measure is used

$$x_{diff}(n) = |x_o(n) - x_a(n)| \tag{7.25}$$

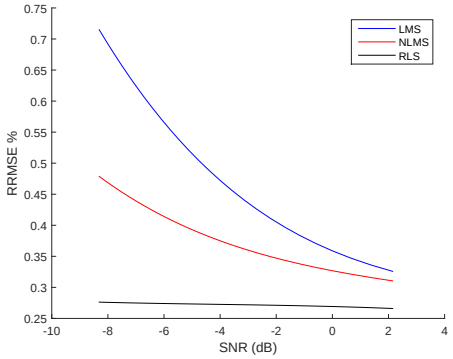
where $x_o(n)$ is the RMCP signal before the method has been used and $x_a(n)$ is the RMCP signal after the algorithm has been run.

8 | Results

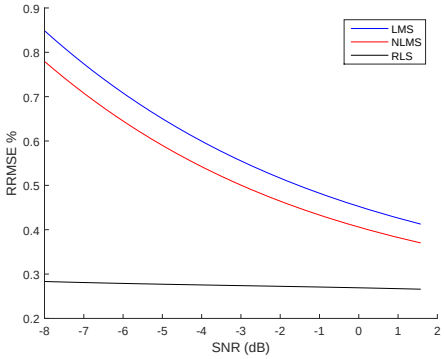
In the first section of this chapter a performance evaluation of our and Peng's method is presented. In the second section a comparison of all the results for different cases is made.

8.1 Performance of our method for different adaptive algorithms

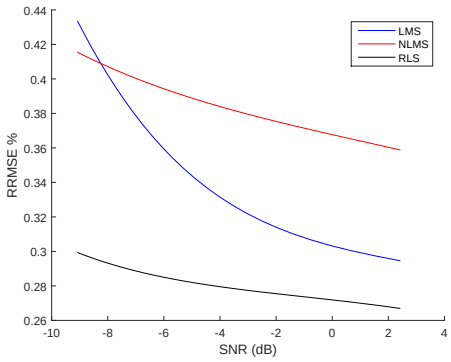
Relative root mean square error is plotted against SNR for our method with three different adaptive algorithms applied (LMS, NLMS and RLS) in fig. 8.1. In all cases (horizontal, vertical and waving movement) RLS performs the best. Normalized LMS yields in a better estimate of the PPG signal when horizontal and vertical motion is applied than for LMS as seen in fig. 8.1. For LMS a filter order of 30 was used and a step-size of 0.00001, for NLMS a step size of 0.00057 and filter order of 30 and lastly for RLS a forgetting factor equal to one and a filter order of 32. These parameters were chosen by series of trials and will be used for our method and Peng's method throughout the results section for the different movements.



(a) Horizontal movement



(b) Vertical movement



(c) Waving movement

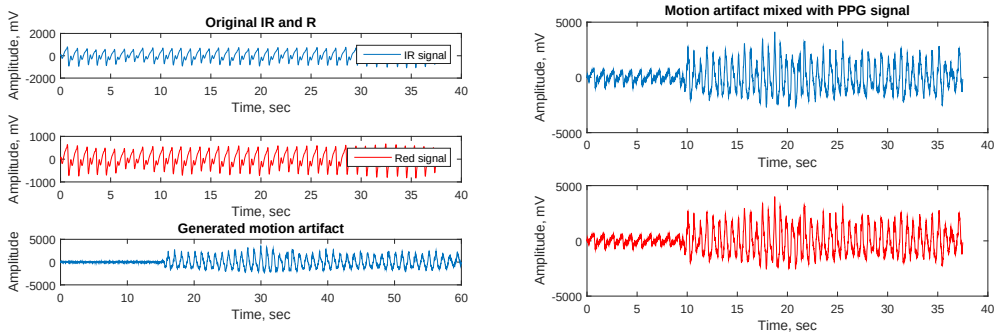
Figure 8.1: Plot of RRMSE against SNR for RLS, NLMS and LMS.

8.2 Comparison

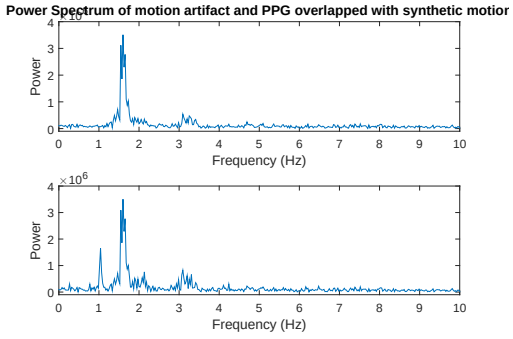
In this section the results of the three methods will be presented. First a performance evaluation of our method for different adaptive algorithms is displayed in three diagrams showing all cases (horizontal, vertical and waving). Then the results of the horizontal movement are presented followed by vertical and waving movement.

8.2.1 Horizontal movement

In fig. 8.2 (a) the original IR and R PPG signals as well as the generated horizontal motion artifact is shown. In fig. 8.2 (b) the mixed signals **x** are used for the three methods in this section. Throughout the results section RLS will be used as the adaptive algorithm for our method and Peng's method for all type of movements. In fig. 8.2 (c) the power spectrum is shown for the motion artifact (upper plot) and the power spectrum of the PPG signal mixed with the motion (lower plot).



(a) (b)



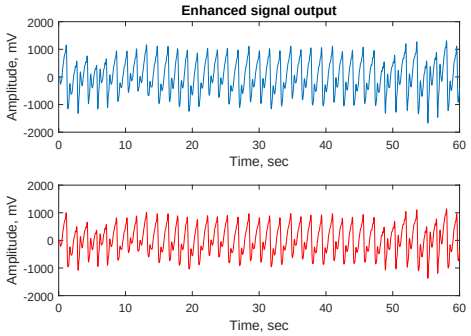
(c)

Figure 8.2: Characteristics of MA and the mixed PPG signals that will be used in the methods. SNR is set to -8.3 dB.

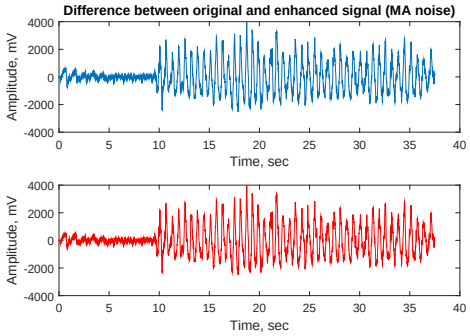
Yousefi's method

In fig 8.3 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.3 (c) the power spectrum of the enhanced signal and the original signal (non-mixed

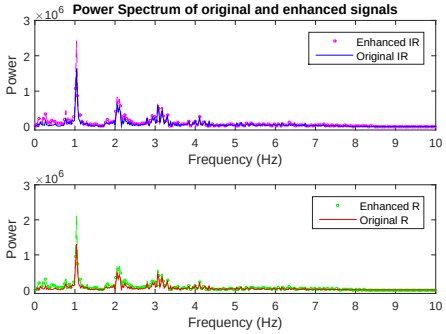
clean PPG-signal) is plotted against each other. Clearly the heart rate peak is visible at 1 Hz, one can notice also that the enhanced signal is overestimating the original signal which is seen in fig. 8.3 (c) and (d).



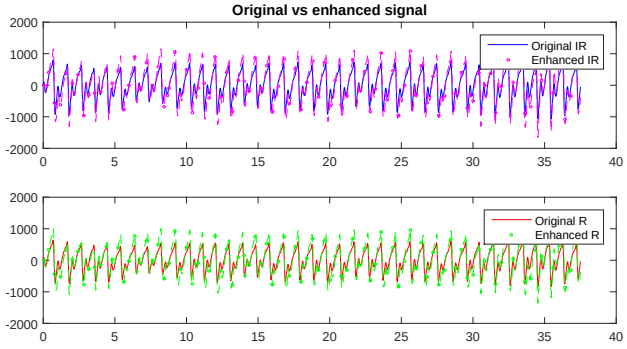
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

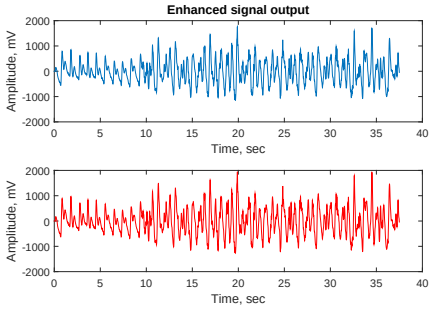


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

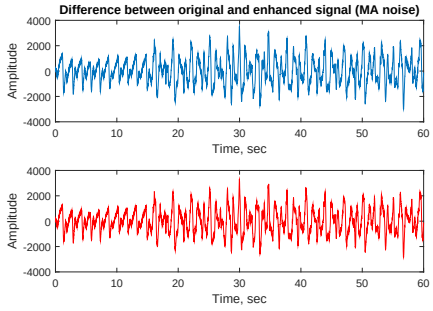
Figure 8.3

Peng's method

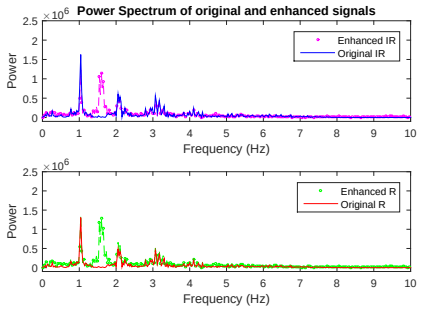
In fig 8.4 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.4 (c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. What we can notice is that the peak at approximately 1.8 Hz associated with motion is not reduced completely resulting in a output that has the shape as in fig. 8.4 (d).



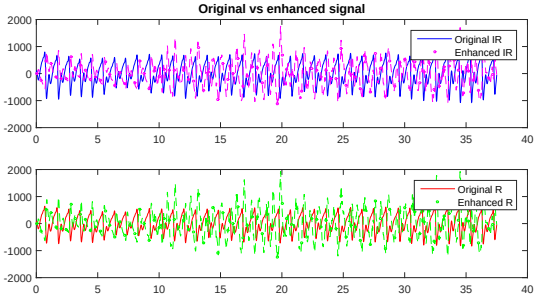
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.



(d) Plot of original IR/R signal and the enhanced output from the algorithm.

Figure 8.4

In these simulations the threshold value in cICA was set to $\xi = 0.05$, there exists no theory on choosing the optimal value of this parameter. The value is chosen by a series of trial. Altering the parameter and setting it to $\xi = 1.3$ gave a better estimate as seen in fig. 8.5. For different SNR this threshold parameter needs to be adjusted in order for the method to converge.

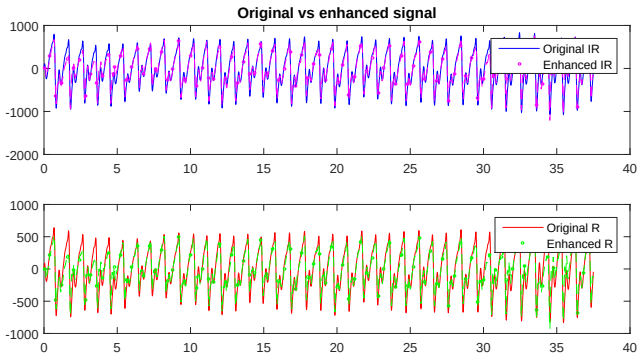


Figure 8.5: Enhanced signal output using the same mixed signal as input but with a different threshold value.

When $\xi = 0.05$ a reference signal as in fig. 8.6 is generated. This reference signal still contains motion and hence in the adaptive noise cancellation when minimizing the error it will result in a poor output as seen in fig. 8.4 (d). With the same threshold $\xi = 0.05$ for different signal fig. 8.14 (d) the result gives a good output.

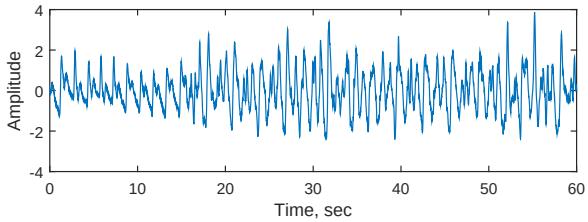


Figure 8.6: Reference signal when threshold parameter is set to $\xi = 0.05$

The reference signal in fig. 8.7 associated with $\xi = 1.3$ contain no motion and hence resulting in a better estimate of the PPG signal when used in adaptive noise cancellation. What we can notice from the reference signal is that the amplitude information is lost, but the MA part is removed.

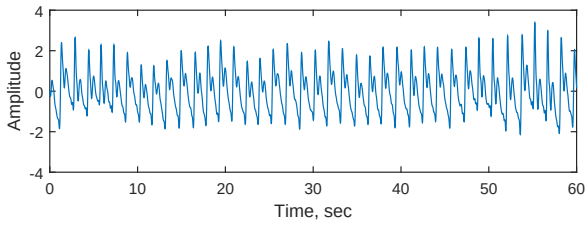


Figure 8.7: Reference signal when threshold parameter is set to $\xi = 1.3$

Figure 8.8 shows that when using $\xi = 1.3$ for a different signal in this case horizontal movement with 30 seconds motion interval the result gives a poor output.

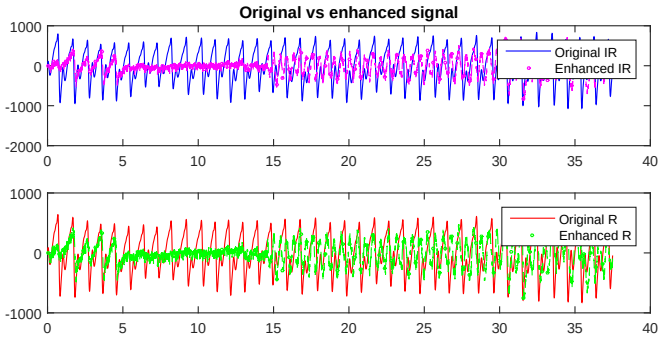


Figure 8.8: Enhanced signal output using the same mixed signal as input but with $\xi = 1.3$ giving a bad result.

The reference signal in fig. 8.9 associated with $\xi = 1.3$ for horizontal movement with 30 seconds motion interval still contains motion and hence in the adaptive noise cancellation when minimizing the error it will result in a poor output as seen in fig. 8.8.

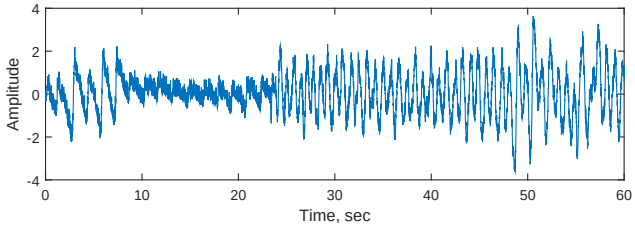
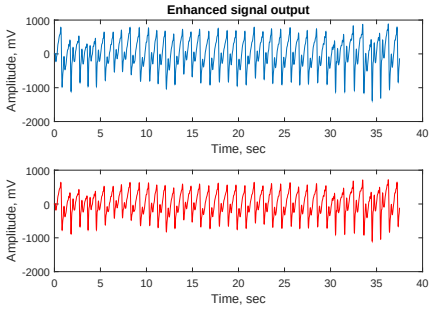


Figure 8.9: Reference signal when threshold parameter is set to $\xi = 1.3$

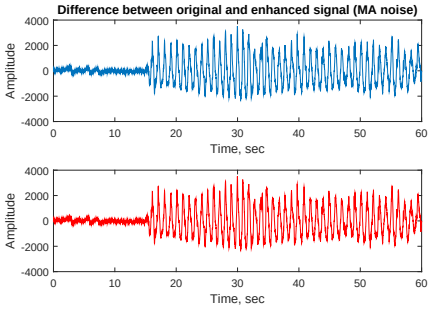
Our method

In fig 8.10 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.10 (c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. As seen from the plots the enhanced signal is estimating the original signal in a good way. Most of the MA part is

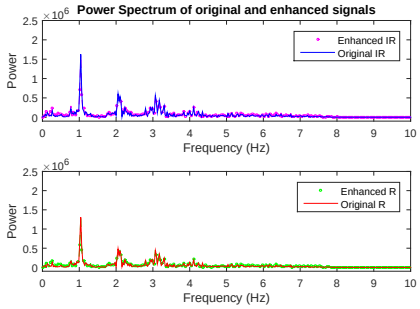
completely removed. In fig. 8.10 (d) the enhanced output follows the original signal slightly worse in the beginning between 0 and 5 seconds. That may have been due to the settling time of the algorithm.



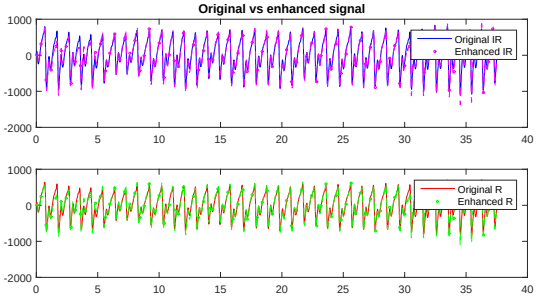
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

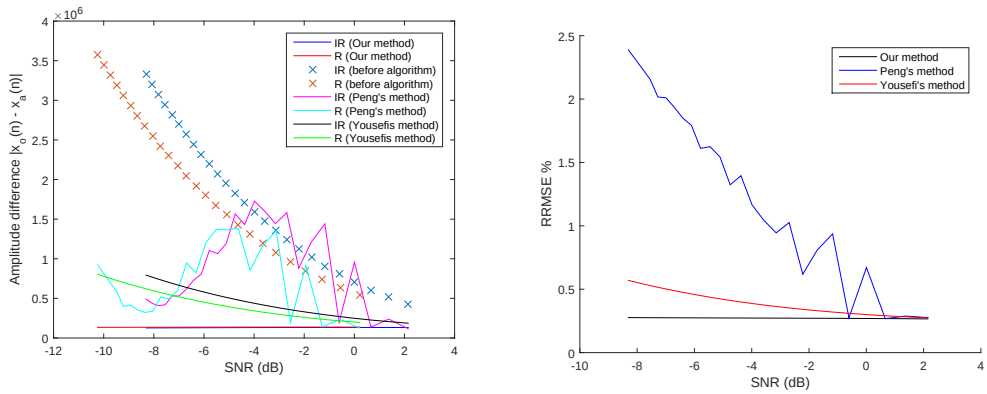


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

Figure 8.10

Comparison

In fig. 8.11 (a) the absolute value of the difference between the RMCP before and after the method is plotted against different SNR values. The dotted cross lines shows the IR and R signal before the algorithm is applied (the mixed signal containing PPG signal and MA).



(a) RMCP amplitude difference for the three methods.

(b) RRMSE plot for the three methods.

Figure 8.11: Comparison of the three methods.

The reason why the RMCP amplitude difference differs much for various SNR values in fig. 8.11 (a) is that the same threshold value was applied when doing the simulations. For horizontal movement with a motion interval of 40 seconds one can notice a SNR improvement of 24.64 dB compared to Peng's method which was 2.26 dB and 11.85 dB for Yousefi's method. It is seen that our method has the largest SNR improvement for all motion intervals generally.

Motion interval [s]	**SNR (before)** [dB]	**SNR (after)** [dB]			**RRMSE** [%]		
		Our	Peng	Yousefi	Our	Peng	Yousefi
10	-3.19	12.42	12.20	0.025	0.23	0.25	0.39
20	-5.75	12.89	-2.63	0.008	0.23	1.35	0.93
30	-9.87	12.40	4.27	0.03	0.24	0.61	0.41
40	-11.8	12.84	-9.54	0.05	0.23	3.00	0.82

Table 8.1: SNR value before the algorithm is applied and after as well as RRMSE for the three methods.

8.2.2 Vertical movement

In fig. 8.12 (a) the original IR and R PPG signals as well as the generated vertical motion artifact is shown. In fig. 8.12 (b) the mixed signals $\mathbf{x}$ are used for the three methods in this section. In fig. 8.12 (c) the power spectrum is shown for the motion artifact (upper plot) and the power spectrum of the PPG signal mixed with the motion (lower plot).

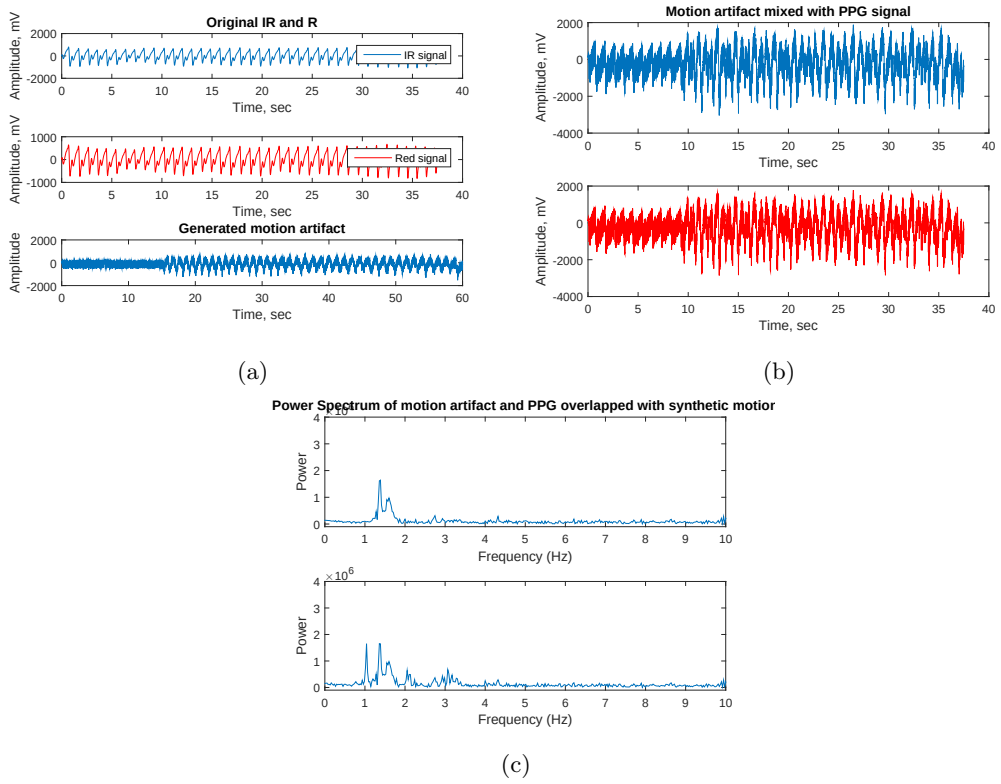
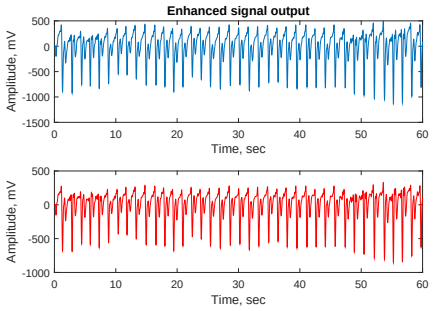


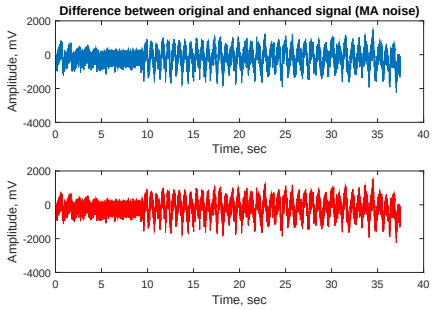
Figure 8.12: Characteristics of MA and the mixed PPG signals that will be used in the methods. SNR is set to -4.4 dB.

Yousefi's method

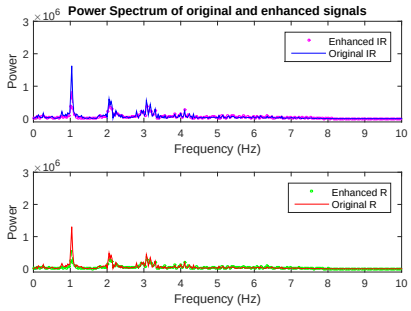
In fig 8.13 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.13(c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. The enhanced signal output is underestimating the original signal as seen in fig. 8.13 (d). This will further be explained in the discussion section later in the report.



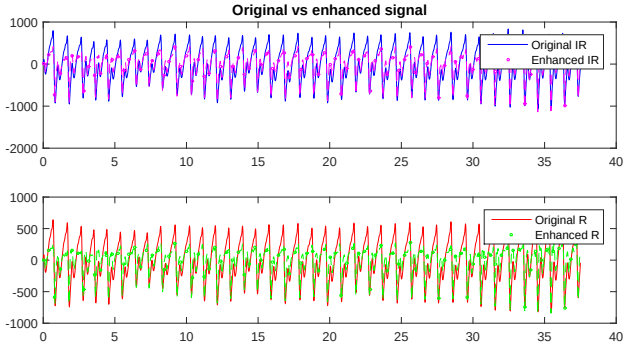
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

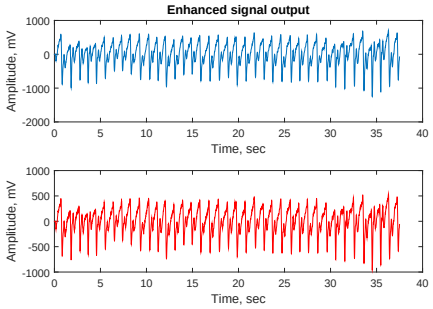


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

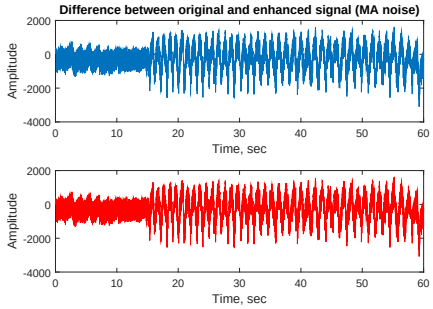
Figure 8.13

Peng's method

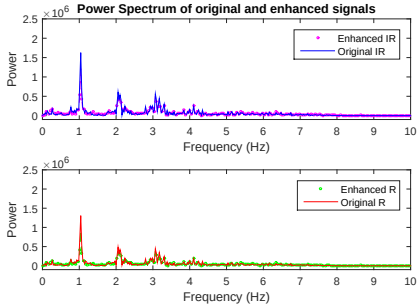
In fig 8.14 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.14(c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. A threshold parameter of $\xi = 0.05$ was used and it resulted in a good estimate as seen in fig. 8.14 (d).



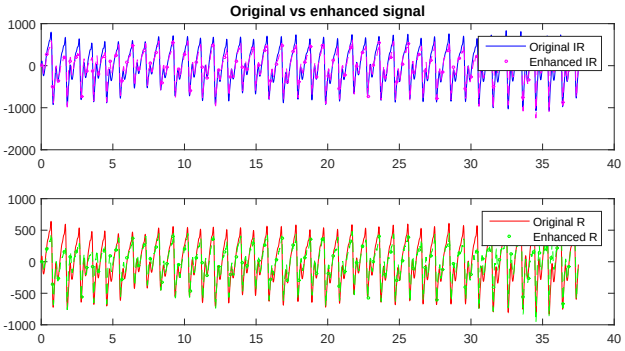
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

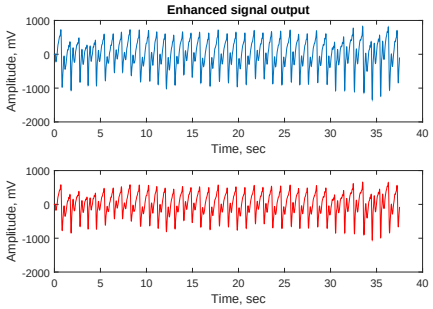


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

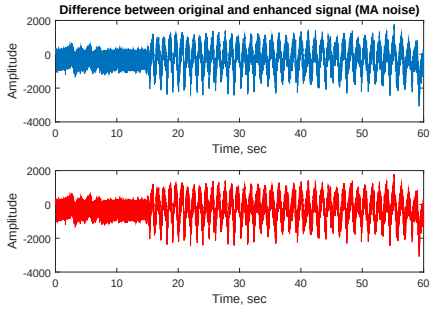
Figure 8.14

Our method

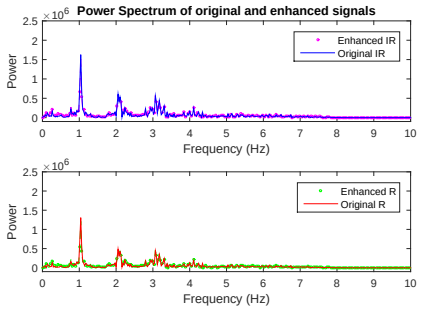
In fig 8.15 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.15(c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. Most of the MA part is reduced and also the some noise due to light resulting in a clean estimate in fig. 8.15 (d).



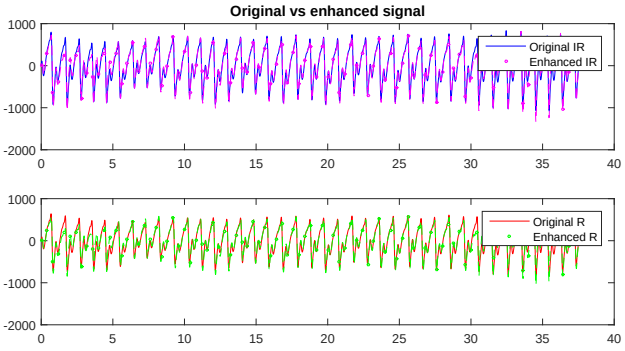
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

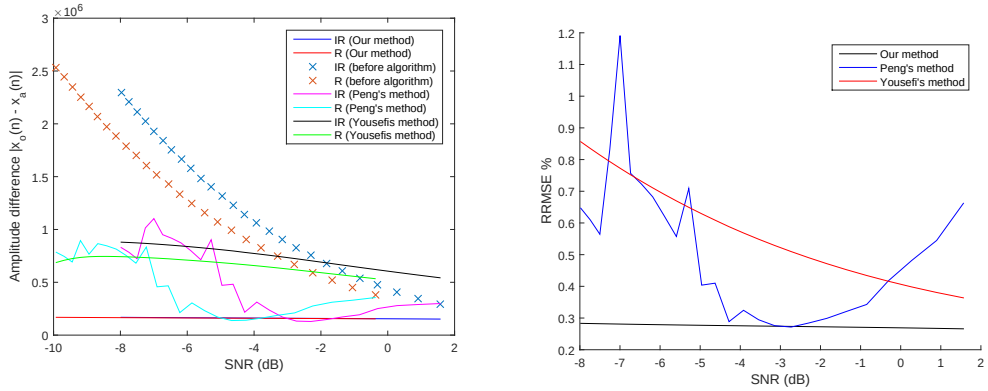


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

Figure 8.15

Comparison

In fig. 8.16 (a) the absolute value of the difference between the RMCP before and after the method is plotted against different SNR values. The dotted cross lines shows the IR and R signal before the algorithm is applied (the mixed signal containing PPG signal and MA).



(a) RMCP amplitude difference for the three methods.

(b) RRMSE plot for the three methods.

Figure 8.16: Comparison of the three methods.

The SNR improvement with the motion interval of 40 seconds for vertical movement was 24.24 dB for our method, 21.17 dB for Peng's method and 12 dB for Yousefi's as can be seen in table 8.2. It is seen that our method has the largest SNR improvement for all motion intervals generally.

Motion interval [s]	**SNR (before)** [dB]	**SNR (after)** [dB]			**RRMSE** [%]		
		Our	Peng	Yousefi	Our	Peng	Yousefi
10	-3.11	11.89	-2.80	0.034	0.25	1.38	0.29
20	-6.06	12.81	8.89	0.005	0.23	0.36	1.12
30	-8.92	11.66	-6.37	0.09	0.26	2.08	1.94
40	-12.06	12.18	9.11	0.006	0.24	0.35	1.33

Table 8.2: SNR value before the algorithm is applied and after as well as RRMSE for the three methods.

8.2.3 Waving movement

In fig. 8.17 (a) the original IR and R PPG signals as well as the generated vertical motion artifact is shown. In fig. 8.17 (b) the mixed signals $\mathbf{x}$ are used for the three methods in this section. In fig. 8.17 (c) the power spectrum is shown for the motion artifact (upper plot) and the power spectrum of the PPG signal mixed with the motion (lower plot).

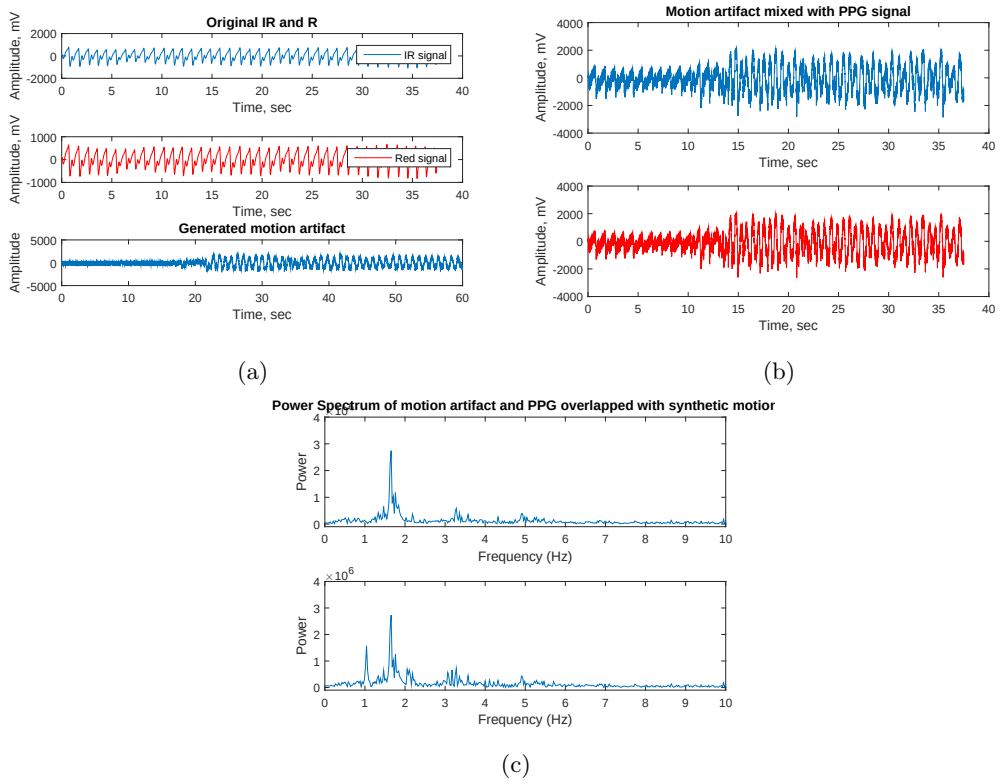
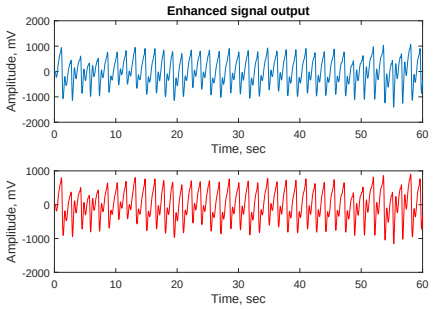


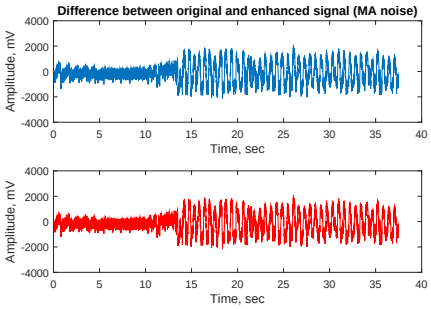
Figure 8.17: Characteristics of MA and the mixed PPG signals that will be used in the methods. SNR is set to -5.5 dB.

Yousefi's method

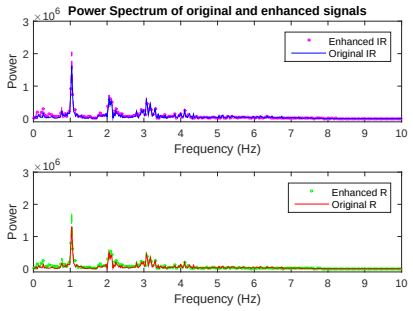
In fig 8.18 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.18 (c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. From the power spectrum in fig. 8.18 (c) we can notice that the enhanced spectrum for both IR and R channel overestimates the original spectrum resulting in an enhanced output with an amplitude information that is a bit larger than the original PPG signal.



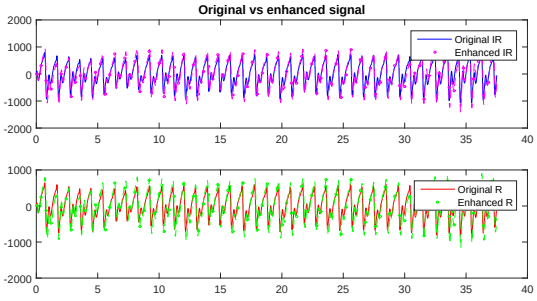
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

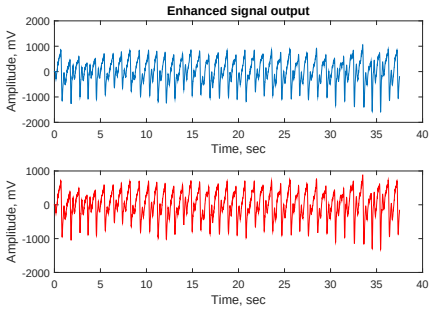


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

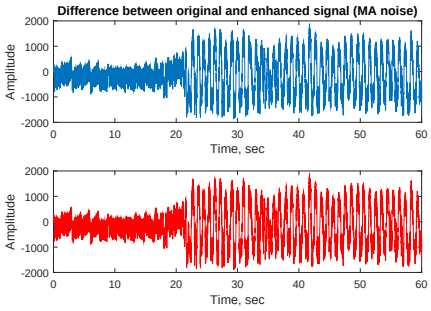
Figure 8.18

Peng's method

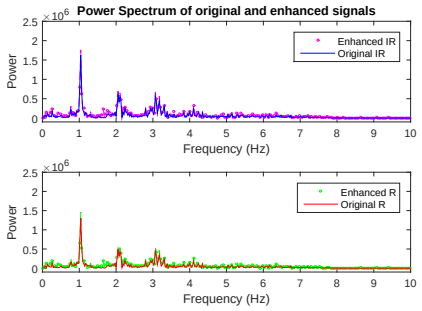
In fig 8.19 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.19 (c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. From fig. 8.19 (d) it is clear that most of the MA and the noise from the mixed signal in fig. 8.19 (b) is reduced, the enhanced output is following the original signal well but with a slight overshoot.



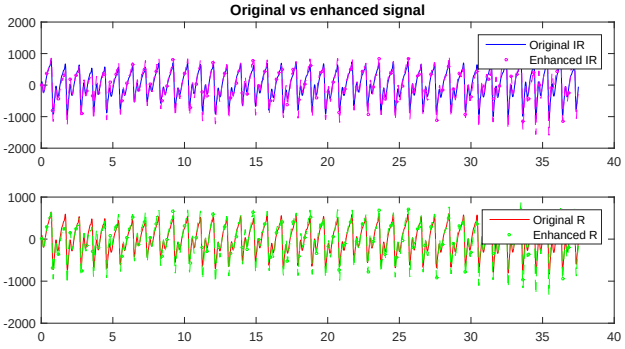
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

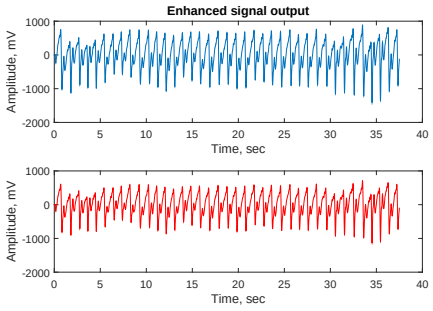


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

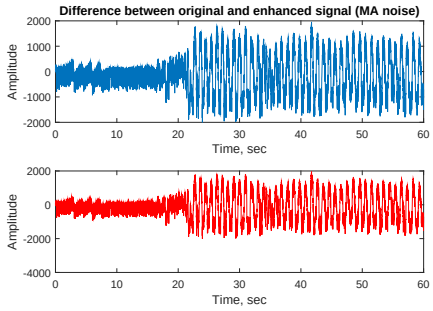
Figure 8.19

Our method

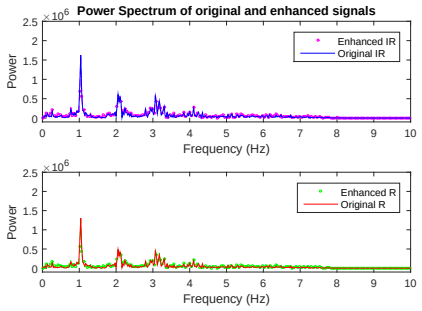
In fig 8.20 (b) the portion of MA that has been removed from the mixed PPG signal is shown. In fig. 8.20 (c) the power spectrum of the enhanced signal and the original signal (non-mixed clean PPG-signal) is plotted against each other. From fig. 8.20 (d) we can see the excellent performance of our method, the MA is removed and the enhanced output is estimating the original PPG signal with the amplitude information recovered.



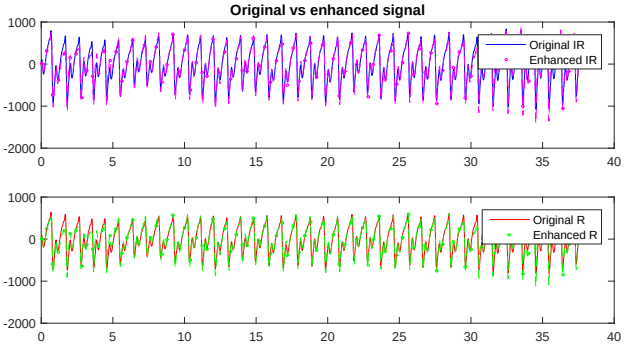
(a) The output from the algorithm (IR above and R below)



(b) MA removed from mixed IR and R PPG signals.



(c) FFT plot for original IR/R signals and enhanced output signals.

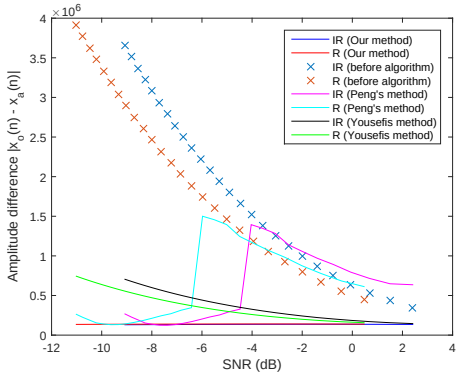


(d) Plot of original IR/R signal and the enhanced output from the algorithm.

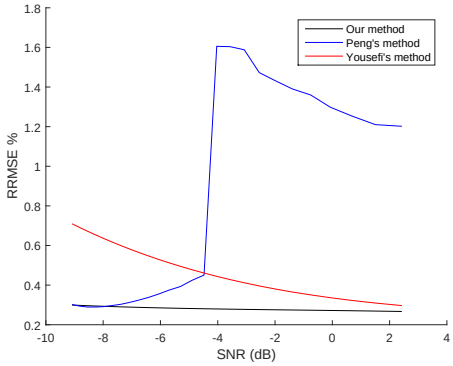
Figure 8.20

Comparison

In fig. 8.21 (a) the absolute value of the difference between the RMCP before and after the method is plotted against different SNR values. The dotted cross lines shows the IR and R signal before the algorithm is applied (the mixed signal containing PPG signal and MA).



(a) RMCP amplitude difference for the three methods.



(b) RRMSE plot for the three methods.

Figure 8.21: Comparison of the three methods.

For waving movement the SNR improvement can be read from table 8.3 for motion interval of 40 seconds to 23.67 dB for out method, 16 dB for Peng's method and 12 dB for Yousefi's method. It is seen that our method has the largest SNR improvement for all motion intervals generally.

Motion interval [s]	**SNR (before)** [dB]	**SNR (after)** [dB]			**RRMSE** [%]		
		Our	Peng	Yousefi	Our	Peng	Yousefi
10	-3.09	10.73	-4.37	0.046	0.29	1.65	0.53
20	-6.04	12.18	-0.47	0.06	0.245	1.05	0.66
30	-8.99	10.85	5.75	0.05	0.286	0.52	0.59
40	-11.99	11.68	4.04	0.04	0.26	0.63	0.96

Table 8.3: SNR value before the algorithm is applied and after as well as RRMSE for the three methods.

9 | Discussion

The results demonstrates that our proposed method outperforms Peng's and Yousefi's methods when extracting a clean estimate of the PPG signal with the MA component reduced. From the results we can see that this is the case for all the types of motion that was generated (horizontal, vertical and waving movement).

Yousefi's method claims to provide a clean PPG signal corresponding to the arterial blood flow. During motion it is assumed that the non-arterial component (venous blood flow) is associated with noise. The difficulty in Yousefi's method lies in finding the value of β that will give you either the arterial component or the venous component. If this parameter is slightly different than r_a the weighted subtraction term will contain both venous component but also some of the arterial component. Accordingly in the adaptive noise cancellation a small portion of the arterial signal component will also be cancelled, which is the case in fig. 8.13 (d) were the enhanced output is an underestimation than the original one. In the same way if β is slightly different than r_v the weighted subtraction will contain a small portion of venous noise also. This could be a problem when extracting the scaled arterial signal $\hat{s}_{as}$ that is supposed to only contain the arterial component.

Peng's method could perform better but depends strongly on the threshold parameter ξ of the cICA algorithm. If the parameter ξ is chosen in a suitable way the desired independent component will be obtained as optimum output. On the other hand if this parameter is selected beyond the upper bound of the range, a completely different component may be obtained. If ξ is chosen too small the method will not converge. The selection of this parameter therefore differs from one output to another and depends on the form of the reference signal. That is the drawback with Peng's method, you have to adjust a proper threshold value in order for the method to converge. It should also be noted that in Peng's method when using cICA the assumption made is that the PPG signal and MA are independent. In reality that is often the case, however if not the performance might be compromised. Constrained ICA algorithm may also produce an incorrect output if the MA and PPG signal has the same period. The reference signal in Peng's method plays a vital role and one needs a lot of a priori information in order to create such a signal which is closely related to the desired source signal. In this thesis we only needed the heart rate of the PPG signal in order to create the reference signal, but one could design a more advanced reference signal taking into consideration the shape and structure of the PPG signal. However that information will not be available in many cases.

Our method converges quickly with good extraction performance and only needs the estimated period information of the desired signal. It can also be mentioned that it is insensitive to errors in the estimation of the period. This proposed method was first a modified version of Peng's method with the extracted PPG-correlated component from the two-stage based approach as an advanced reference signal in the cICA algorithm. However we noticed that this extracted component was not much different than the output from the cICA algorithm. Hence we used the PPG-correlated component directly in the adaptive noise cancellation step and disregarded the cICA step resulting in our proposed method. For our method the best choice of adaptive algorithm was the conventional RLS algorithm. Generally the RLS algorithm has a faster convergence rate and gave a better signal estimate, the limitation

however is the computational complexity. Compared to LMS and NLMS the RLS algorithm requires more operations in general.

In all three methods syntethical reference signals has been generated and used in adaptive noise cancelers. Adaptive filters are effective tools when it comes to removal of the in-band noise given that the reference is correlated with eiter MA or PPG signal. Numerous reports and research has been dealing with the reduction of motion artifact [43] [44] [51] [52] [53] [54], however the amplitude information in these methods has not been taken into consideration. There are other methods of providing reference signals such as accelerometers or photoelectric devices, that however would require extra hardware, which is expensive. Another way is to generate MA from the PPG signal using singular value decomposition or independent component analysis. However the drawback with ICA is the arbitrary ordering on the independent component outputs. Especially when having a large number of channels it will require a highly subjective analysis on the large number of outputs.

It has to be mentioned that there may be limitations that has not been taking into consideration when one conducts the simulations. First and foremost the experiments was only tested on two healthy persons. In cases when we might have a person that has some kind of health problems such as abnormal heart rhythm or intradialytic hypotension it might temporarily reduce the quality of the algorithms. It should also be emphasized that only vertical, horizontal and waving movements were tested, there are of course other types of motion artifacts that could affect the amplitude of the PPG signal as well. That as well as testing the methods on a larger dataset might be improvements for better conclusions in future work.

10 | Conclusion

The objective of this thesis was to investigate methods for artifact removal in PPG signals and to implement and evaluate a few existing algorithms dealing with motion artifact in PPG signals with the amplitude information reserved. The results were promising and a new proposed method was developed dealing with the reduction of MA with the amplitude information recovered. Our proposed method is easy to implement and converges quickly with good extraction performance. It has a few design parameters and only needs the estimated period of the PPG signal. Our method could be used in a clinical routine for prediction of intradialytic hypotension. However it should be mentioned that although our method has great potential the simulations were only conducted on two healthy males. Further studies on a larger dataset might be needed in order to establish a full value of the efficacy of our method.

References

[1] G. Clarke. Signal Quality Analysis in Pulse Oximetry: Modelling and Detection of Motion Artifact. *Ottawa-Carleton Institute for Biomedical Engineering*, Carleton University, May 2015.

[2] A. R. Relente and L. G. Sison. Characterization and Adaptive Filtering of Motion Artifacts in Pulse Oximetry using Accelerometers. *Proceedings of the Second Joint EMBS/BMES Conference*, Houston, TX, USA, October 2002.

[3] L. Sörnmo, F. Sandberg, E. Gil, and K. Solem. Noninvasive techniques for prevention of intradialytic hypotension. *IEEE Reviews, Biomedical Engineering.*, vol 5 (p.45-59), 2012.

[4] K. Solem, B. Olde and L. Sörnmo. Prediction of intradialytic hypotension using photoplethysmography. *IEEE Transactions, Biomedical Engineering.*, vol 57 (p.1611-1619), 2010.

[5] F. Sandberg, R. Bailon, D. Hernando, P. Laguna, J. P. Martinez, K. Solem and L. Sörnmo. Prediction of hypotension in hemodialysis patients. *Physiological measurement, IOP Publishing.*, vol 35 (p.1885), 2014.

[6] F. Peng, Z. Zhang, X. Gou, H. Liu and W. Wang. Motion artifact removal from photoplethysmographic signals by combining temporally constrained independent component analysis and adaptive filter. *Peng et al. BioMedical Engineering OnLine*, vol 13, nbr 50, 2014.

[7] R. Yousefi, M. Nourani, S. Ostadabbas and I. Panahi. A Motion-Tolerant Adaptive Algorithm for Wearable Photoplethysmographic Biosensors. *IEEE Journal of Biomedical and Health Informatics.*, vol 18, nbr 2, March 2014.

[8] R. Ortega, C J. Hansen, K. Elterman and A. Woo. Pulse Oximetry. *The new England Journal of Medicine*, Massachusetts Medical Society 2011.

[9] J E. Sinex. Pulse Oximetry: Principles and Limitations. *American Journal of Emergency Medicine*, vol 17, nbr 1, January 1999.

[10] K N. Glaros. Low-power pulse oximetry and transimpedance amplifiers. *Imperial College London*, Department of Bioengineering, October 2011.

[11] N. Townsend. Pulse Oximetry. *Medical Electronics*, Michaelmas Term, 2001.

[12] T. Tamura, Y. Maeda, M. Sekine and M. Yoshida. Wearable Photoplethysmographic Sensors - Past and Present. *Electronics*, vol 3 (p.282-302), 2014.

[13] R. Dresher. Wearable Forehead Pulse Oximetry: Minimization of Motion and Pressure Artifacts. *Worcester Polytechnic Institute*, Department of Biomedical Engineering, May 3, 2006.

[14] A. Sola, L. Chow and M. Rogido. Pulse Oximetry in Neonatal Care. A Comprehensive State of the Art Review. *Anales of Pediatria*, vol 62, nbr 3, pp. 266-80, 2005

[15] H. Lee, J. Lee, W. Jung and G. Lee. The Periodic Moving Average Filter for Removing Motion Artifacts from PPG Signals. *International Journal of Control, Automation and Systems*, vol 5, nbr 6, pp. 701-706, December 2007

[16] J.W. Severinghaus and S.O. Koh. Effect of anemia on pulse oximeter accuracy at low saturation. *J. Clin. Monitor. Comp.*, vol 6, nbr 2, p. 85-88, 1990

[17] A.P. Lima, P. Beelen, and J. Bakker. Use of a peripheral perfusion index derived from the pulse oximetry signal as a noninvasive indicator of perfusion. *Crit. Care Med.*, vol 30, nbr 6, 2002

[18] KDOQI Clinical Practice Guidelines for Cardiovascular Disease in Dialysis Patients. `http://www2.kidney.org/professionals/KDOQI/guidelines_cvd/intradialytic.htm`, accessed June 2015.

[19] S. DeMeulenaere. Pulse Oximetry: Uses and Limitations. *The Journal for Nurse Practitioners.*, vol 3, Issue 5, (p.312-317), May 2007.

[20] T. Aoyagi, M. Fuse, N. Kobayashi, K. Machida, K. Miyasaka. Multiwavelength Pulse Oximetry: Theory for the Future. *International Anesthesia Research Society.*, vol 105, nbr 6, December 2007.

[21] L. Wood and H. Asada. Low Variance Adaptive Filter for Cancelling Motion Artifact in Wearable Photoplethysmogram Sensor Signals. *Proceedings of the 29th Annual International Conference of the IEEE EMBS.*, August 23-26, 2007.

[22] W. Gardner. Introduction to Random processes with Applications to Signal and Systems. *Macmillan*, 1986.

[23] A. Papoulis. Probability, Random Variables, and Stochastic Processes. *McGraw-Hill*, 3rd edition, 1991.

[24] R. Gray and L. Davisson. Random processes: A mathematical approach for engineers. *Prentice Hall*, 1986.

[25] C. Therrien. Discrete random signals and Statistical Signal Processing. *Prentice Hall*, 1992.

[26] A. Hyvärinen, J. Karhunen and E. Oja. Independent Component Analysis. *John Wiley & Sons, Inc*, 2001.

[27] L. Sörnmo and P. Laguna. Bioelectrical Signal Processing in Cardiac and Neurological Applications. *Elsevier Inc*, 2005.

[28] S. Haykin. Adaptive Filter Theory. *Prentice Hall*, 4th edition, 2002.

[29] M. H. Hayes. Statistical Digital Signal Processing and Modeling. *John Wiley & Sons, Inc*, 1996.

[30] B. C. Chachuat. Nonlinear and Dynamic Optimization. *Laboratoire d'Automatique, Ecole Polytechnique Federale de Lausanne*, IC-32: Winter Semester 2006/2007.

[31] Optimization. `https://en.wikipedia.org/wiki/Mathematical_optimization`, accessed July 2015.

[32] P. Armand and D. Orban. The Squared Slacks Transformation in Nonlinear Programming. *Cahier du GERAD G-2007-62*, August 29, 2007.

[33] M. Avriel. Nonlinear Programming: Analysis and Methods. *Dover Publishing*, 2003.

[34] J. A. Snyman. Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. *Springer Publishing*, 2005.

[35] C. T. Kelley. Iterative Methods for Optimization. *Society for Industrial and Applied Mathematics*, 1999.

[36] R. Fletcher. Practical methods of optimization. *John Wiley & Sons, Inc*, 2nd edition, 1987.

[37] Lagrange multiplier. `https://en.wikipedia.org/wiki/Lagrange_multiplier`, accessed July 2015.

[38] D. P. Bertsekas. Nonlinear programming. *Athena Scientific*, 2nd edition, 1999.

[39] H. Li. Lagrange Multipliers and their Applications. *Department of Electrical Engineering and Computer Science*, University of Tennessee, September 28, 2008.

[40] D. G. Luenberger and Y. Ye. Linear and Nonlinear Programming. *Springer Science*, 3rd edition, 2008.

[41] D. P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. *Athena Scientific*, 1996.

[42] B. W. Wah and T. Wang. Efficient and Adaptive Lagrange-Multiplier Methods for Nonlinear Continuous Global Optimization. *Journal of Global Optimization*, Kluwer Academic Publishers, vol 14 (p.1-25) 1999.

[43] J. Yao and S. Warren. A Novel Algorithm to Separate Motion Artifacts from Photoplethysmographic Signals Obtained with a Reflectance Pulse Oximeter. *Proceedings of the 26th Annual International Conference of the IEEE EMBS.*, September 1-5, 2004.

[44] M. Milanesi, N. Martini, N. Vanello, V. Positano, M. F. Santarelli and L. Landini. Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals. *Medical & Biological Engineering & Computing*, vol 46, Issue 3 (p. 251-261), March 2008.

[45] A. Hyvärinen and E. Oja. Independent Component Analysis: A Tutorial *Laboratory of Computer and Information Science*, Helsinki University of Technology, 1999.

[46] A. Hyvärinen. Survey on Independent Component Analysis *Neural Computing Surveys 2*, (p. 94-128), 1999.

[47] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications *Neural Networks*, 13(4-5): p. 411-430, 1999.

[48] W. Lu and J. C. Rajapakse. ICA with reference *School of Computer Engineering*, Nanyang Technological University of Singapore, 2006.

[49] W. Lu and J. C. Rajapakse. Constrained Independent Component Analysis *School of Computer Engineering*, Nanyang Technological University of Singapore, 2001.

[50] W. Lu and J. C. Rajapakse. Approach and Applications of Constrained ICA. *IEEE Transactions, Neural Networks.*, vol 16 (no. 1), 2005.

[51] Z. Zhang and L. Zhang. A Two-Stage Based Approach for Extracting Periodic Signals. *Proc. of the 6th International Conference on Independent Component Analysis and Blind Signal Separation.*, LNCS 3889, p. 303–310, ICA 2006.

[52] B. S. Kim and S. K. Yoo. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Transactions, Biomedical Engineering.*, vol 53 (issue 3), p. 566-568, 2006.

[53] K. W. Chan and Y. T. Zhang. Adaptive reduction of motion artifact from photoplethysmographic recordings using a variable step-size LMS filter. *Proceedings of IEEE, Sensors.*, vol 2 (issue 3), p. 1343-1346, 2002.

[54] C. M. Lee and Y. T. Zhang. Reduction of motion artifacts from photoplethysmographic recordings using a wavelet denoising approach. *IEEE EMBS Asian-Pacific Conference, Biomedical engineering.*, p. 194-195, 2003.

[55] J. M. Goldman, M. T. Petterson, R. J. Kopotic and S. J. Barker. Masimo Signal Extraction Pulse Oximetry. *Journal of Clinical Monitoring and Computing.*, vol 16 (issue 7), p. 475-483, 2000.