Independent Component Analysis for Beam Measurement

Xiaoying Pang

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirement
for the degree
Doctor of Philosophy
in the Department of Physics,
Indiana University

August, 2009
Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

Shyh-Yuan Lee, Ph.D.

Steven A. Gottlieb, Ph.D.

Mark H. Hess, Ph.D.

August, 2009

Paul E. Sokol, Ph.D.
To my parents.
Acknowledgments
Abstract

Xiaoying Pang

Independent Component Analysis for Beam Measurement

Independent component analysis (ICA) has been used for data mining in many branches of science. In beam physics, ICA can be applied to analyze the turn-by-turn beam position monitor (BPM) data of the transverse beam motion. Transverse beam motion consists of both linear betatron motion and the nonlinear motions which can be excited by higher order nonlinear magnetic elements. The diagnosis of the linear betatron motion has been developed by other methods like the model independent analysis (MIA) which is essentially built upon the principle component analysis (PCA). However, the detection and analysis of the nonlinear optics of an accelerator has never been effectively addressed. With the introduction of the ICA method which is inherently superior over the other traditional multivariate statistical methods like PCA or factor analysis, one can gain better accuracy in the linear betatron motion analysis and make the nonlinear optics diagnosis feasible.

The narrow-band filtering of ICA provides us with accurate interpretation of both the linear and nonlinear betatron motions, especially the $x_{2\nu}$ motion as we show in this thesis. For linear betatron oscillation, ICA can extract the betatron amplitude function and phase advance with high precision. For the nonlinear motion, the $x_{2\nu}$ motion uncovered by ICA agrees very well the analytical result derived from the Hill’s equation. Based upon the study of both linear and nonlinear optics using ICA method, we finally propose two methods of beam-based measurement of the higher order nonlinear magnetic elements.
Contents

Acceptance iii

Acknowledgments ix

Abstract xi

1 Introduction 1

1.1 Motivation of Study .......................... 1

1.2 Basics of Accelerator Physics .................. 4

1.2.1 Accelerator Coordinate System ................ 4

1.2.2 Beam Steering Magnets ....................... 5

1.2.3 Transverse Motion ............................ 7

2 Introduction to Independent Component Analysis (ICA) 13

2.1 The Basics of ICA .............................. 14

2.1.1 ICA Model .................................. 15

2.1.2 The basic Assumption of ICA ................. 15

2.1.3 Application and History of ICA ............... 16

2.2 ICA by Maximization of Non-gaussianity .......... 17

2.2.1 Measuring Non-gaussianity by Kurtosis ........ 18

2.2.2 Measuring non-gaussianity by Neg-entropy .... 19
### 2.3 ICA Using Time Structure
- 2.3.1 Using One Time Lag and The AMUSE Algorithm
- 2.3.2 Using Several Time Lags and the SOBI algorithm

### 2.4 Summary

### 3 Independent Component Analysis (ICA) for Linear Motion
- 3.1 BPM Data Arrangement and ICA
- 3.2 Linear Beam Motion Analysis
  - 3.2.1 PCA on Simple Sinusoidal Motion
  - 3.2.2 Betatron Motion
- 3.3 Error Analysis
  - 3.3.1 Effect of Sampling Error
  - 3.3.2 Effect of the cut-off threshold $\lambda_c$ and the number of time lags $\tau_k$
  - 3.3.3 Effect of BPM noise
- 3.4 Summary

### 4 Independent Component Analysis (ICA) for Nonlinear Motion
- 4.1 Equation of Motion for $2\nu_x$ mode
- 4.2 Singular Values of the Nonlinear Modes
  - 4.2.1 Analysis of Singular Value with One Sextupole
  - 4.2.2 Analysis of Singular Values with two Sextupoles
- 4.3 Compare ICA with the analytical calculation
- 4.4 Beam-based measurement of sextupole strength
  - 4.4.1 Action change of $x_{2\nu}$ and sextupole strength
  - 4.4.2 Error Analysis
  - 4.4.3 Determine sextupole strength using Less BPMs
- 4.5 Summary
CONTENTS

5 Conclusions 109

Appendix 112

A Principle component analysis (PCA) 113
A.1 Principle Component . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
A.2 Singular Value Decomposition . . . . . . . . . . . . . . . . . . . . . 116

Bibliography 116
List of Tables
List of Figures

1.1 Frenet-Serret coordinate system. ............................................ 4

1.2 Schematic drawing of dipole, quadrupole and sextupole. ................. 6

1.3 Schematic drawing of strip beam position monitor [9] .................... 11

3.1 The 7th superperiod of AGS lattice, which contains one sextupole at location 420.37m. The AGS lattice consists of bending dipoles with various gradients. ................................................................. 37

3.2 Compare the sum of squares of the first two columns of the U matrix (Eq. (3.19)) with the real $\beta_x$ function from MAD. The initial tracking position is $x_0 = 1\text{mm}$. ................................................................. 38

3.3 Compare the phase obtained by Eq. (3.20) with the real phase advance $\phi$ from MAD. The initial tracking position is $x_0 = 1\text{mm}$. ................................. 39
3.4 The fast Fourier transform (FFT) of source signals for 6 modes. The first two modes correspond to betatron motion. The 3rd and 4th modes are $2\nu_x$ modes which are caused by sextupole kicks. The 5th and 6th are $3\nu_x$ modes. They are the result of the combination of betatron motion and $2\nu_x$ motions. With 1 mm initial tracking displacement, the singular values (SVs or eigenvalues) of the first two modes are about 600 times larger than the SV of the 3rd mode, $2 \times 10^5$ times of the 4th SV, $3 \times 10^7$ times of the 5th SV and $10^8$ times of the SV of the 6th mode. Notice that PCA or SVD ranks its modes by the magnitude of the corresponding eigenvalues. 41

3.5 Left (from top to bottom): normalized spatial function of the first mode $A_1$ (first column of the mixing matrix $A$), temporal function of the first mode $s_1$ (first source signal), FFT of the temporal function of the first mode. Right (from top to bottom): normalized spatial function of the second mode $A_2$ (second column of the mixing matrix $A$), temporal function of the second mode $s_1$ (second source signal), FFT of the temporal function of the second mode. 43

3.6 Compare $\beta_x$ from ICA with the real $\beta_x$ at every BPM from MAD. The initial tracking position is $x_0 = 1$ mm. 44

3.7 Compare $\phi$ from ICA with the phase advance at every BPM from MAD. The initial tracking position is $x_0 = 1$ mm. 45

3.8 $\Delta \beta/\beta$ of ICA calculation for the first two betatron modes vs. different tracking turn numbers. $\nu_x = 8.825$. 47

3.9 $\Delta \beta/\beta$ of PCA calculation for the first two betatron modes vs. different tracking turn numbers. $\nu_x = 8.825$. 48

3.10 Compare $\Delta \beta/\beta$ of ICA with that of PCA for different numbers of tracking turns. $\nu_x = 8.825$. 49
3.11 Top: $(\Delta \beta/\beta)_{rms}$ vs. number of time lags for different number of eigenmode kept in whitening process. Tracking initial position is $x_0 = 1 mm$. Bottom: $(\Delta \phi)_{rms}$ vs. number of time lags for different number of eigenmode kept in whitening process. .......................... 50

3.12 Top: $(\Delta \beta/\beta)_{rms}$ vs. number of time lags for different BPM noise levels. Tracking initial position is $x_0 = 1 mm$. six eigenmodes are kept. in the whitening process. Bottom: $(\Delta \phi)_{rms}$ vs. number of time lags for different BPM noise levels. Tracking initial position is $x_0 = 1 mm$. six eigenmodes are kept. .......................... 51

3.13 In electron machine, the BPM noise level is about 100$\mu m$. The estimation error for $\beta$ function and phase using one random seed. The large error in the phase calculation at small phase angle is derived from the numerical error of dividing by a very small number. .......................... 53

3.14 In proton machine, the BPM noise level is about 200$\mu m$. The estimation error for $\beta$ function and phase using one random seed. .......................... 54

3.15 The ICA and PCA errors of $\beta$ function and phase advance vs. different BPM noise levels for various initial positions of tracking. The solid lines correspond to ICA results. Dashed lines are for the PCA results. The blue, red, green and black lines correspond to initial tracking position $x_0 = 1 mm, 2 mm, 5 mm, 10 mm$ respectively. .......................... 55

3.16 The scaled errors of $\beta$ function and phase advance vs. different BPM noise levels for various initial positions of tracking. When $x_0$ is too large the scaling property would no longer be valid. .......................... 56
4.1 The fast fourier transform (FFT) of the first 6 columns of temporal matrix \( V \). All other modes due to truncation errors, noises are small. The first two modes are betatron modes. The 3rd and 4th are \( 2\nu_x \) modes. The 5th and 6th are \( 3\nu_x \) modes. Only 6 modes are relevant in this example. 64

4.2 The SVs of the 3rd and 4th modes vs the sextupole strength \( K_2L \) for a single sextupole in an ideal AGS lattice with horizontal tune 8.825. The sextupole is located at the 7th superperiod of the ring. The \( \beta \) function at that location is 15.217m. Note that the SV of the 3rd mode is much larger than that of the 4th mode. However, if there are more than one sextupoles with alternating signs of field strengths. The 4th mode will also be proportional to sextupole strength. 66

4.3 The SVs of the 3rd and 4th modes vs the sextupole strength \( K_2L \) for a single sextupole in an ideal AGS lattice with horizontal tune 8.501. The sextupole is located at the 7th superperiod of the ring. The \( \beta \) function at that location is 15.804m. Because the tune is close to half integer, higher sextupole strength may result in larger SV, thereby distorts the linear dependence of the SV on sextupole strength \( K_2L \). This is more apparent for the 4th mode whose SV is much smaller therefore easier to see. 67

4.4 The SVs of the 3rd and 4th modes vs. the sextupole strength \( K_2L \) when two identical sextupoles are placed in an ideal AGS lattice with horizontal tune 8.825. The sextupoles are located respectively at the 3rd and 7th superperiod of the ring. The \( \beta \) function at the sextupole locations are 15.217m. Compare to single sextupole cases, the 4th mode SV is relatively larger. 68
4.5 The SVs of the 3rd and 4th modes vs the sextupole strength \( K_2L \) when two sextupoles with opposite signs are placed in an ideal AGS lattice with horizontal tune 8.825. The sextupoles are located respectively at the 3rd and 7th superperiod of the ring. The \( \beta \) function at the sextupole locations are 15.271m. The alternating signs of the sextupoles enhance SVs of the 4th mode. .................... 69

4.6 AGS lattice with single sextupole at 420.37m and strength \( K_2L = 1 \). From top to bottom, left to right: 3rd column of the mixing matrix A divided by \( \sqrt{\beta_x} \), 4th column of the mixing matrix A divided by \( \sqrt{\beta_x} \), 3rd row of temporal function \( s \), 4th row of temporal function \( s \), FFT of \( s_3 \), FFT of \( s_4 \), \( x_{2\nu} \) divided by \( \sqrt{\beta_x} \) which follows Eq. (4.9). The sextupole location was marked by a red line. .................... 72

4.7 AGS lattice with single sextupole located at 420.37m with sextupole strength \( K_2L = 1 \). The vertical line marks the location of the sextupole. .............. 73

4.8 AGS lattice with single sextupole located at 420.37m with sextupole strength \( K_2L = 1 \). Results for two tracking turns are compared. .................... 74

4.9 AGS lattice with two sextupoles located at 185m and 420.37m with strengths \( K_2L = 1m^{-2} \) and \( K_2L = -1.5m^{-2} \) respectively. Green lines indicate the locations of the two sextupoles. .................... 75

4.10 AGS lattice with ten sextupoles located at various locations around the ring. Green lines indicate the locations of the sextupoles. .................... 76

4.12 AGS lattice with single sextupole at 420.37m and strength \( K_2L = 1 \). The sextupole location is indicated by a red dashed line. PCA can not extract proper \( x_{2\nu} \) while ICA can. Notice that \( u_4 \) does not display any betatron motion like feature outside of the sextupole location. .................... 78
4.13 AGS lattice with one sextupole. The initial tracking position is 1mm. Top: differences between the $x_{2\nu}$ from ICA with the first term of Eq. (4.7) for different tunes. Middle: differences between the closed orbit from ICA with the second term of Eq. (4.7) for different tunes. Bottom: ratio of the singular value for the third mode (the mode corresponds to $x_{2\nu}$ motion) over the average of the first two singular values (correspond to linear betatron motion).

4.14 Zoom into the tune range around the third order resonance for Fig. 4.13.

4.15 Same as Fig. 4.13 except that the initial tracking position is 10mm.

4.16 Zoom into the tune range around the third order resonance for Fig. 4.15.

4.35 Standard cell of SPEAR3.
Chapter 1

Introduction

1.1 Motivation of Study

The performance of storage ring light sources or circular colliders depends crucially on correct beam optics. Both the linear and nonlinear optics require lengthy and complicated beam dynamics optimization processes in order to achieve high brilliance, small emittance, long beam lifetime, large dynamic aperture and good injection efficiency. Therefore, good beam diagnostics technique is in great need. Data mining and analysis play an important role in uncovering underlying physics principles in complex dynamical systems. Using high precision closed orbit data under a small perturbation, one can attain a vast data array of the orbit response matrix (ORM) and successfully model and uncover problems associated with high brightness storage rings [1]. The model independent analysis (MIA), an analysis of a massive data-set obtained from a large number of pulses and beam position monitors (BPM), can be used to untangle eigenmodes (spatial and temporal information) of perturbations in linac transport systems and in high brightness storage rings with coherent rf dipole field excitations [2].
The data sampled by BPMs reflect the beam transverse motion and characteristics of the BPM system, i.e. the BPM noise level. The transverse motion is a combination of betatron motion, synchrotron motion (coupled through dispersion function) and perturbations from other sources, such as noise, ground motion, mechanical vibration, wake field, etc. Assuming a linear BPM system, the samples can be considered as a linear mixture of a few physical source signals. The ultimate goal of data analysis is to uncover these independent physical signals.

In fact, MIA is essentially a Principal Component Analysis (PCA) method, which tries to find a linear transformation of the samples to express the maximum amount of variance in the least number of uncorrelated components (See Appendix A). However, when these source modes are not weakly coupled, MIA may encounter difficulty in isolating the independent modes. In many applications in accelerators, principle component analysis (PCA) can not extract independent components when eigenvalues are close and eigenmodes are coupled.

A multivariate statistical technique called independent component analysis (ICA) was developed in the past decades to solve many engineering problems relevant to communication and medicine. ICA has also been used in neuro-science research and beam measurements. In beam physics, ICA methods can provide a remedy for MIA’s limitations by identifying the independent source signals from the samples using unequal time correlations. The source signals, once identified, can provide information about the betatron, synchrotron, and other modes according to their spatial and temporal properties. This new method has been shown to be more robust and more accurate than PCA since the latter is less immune to mode-mixing and noise signals. Because of the de-mixing, it is possible to identify and study the beam motions arising from other perturbing sources.

The time-correlation based ICA method is particularly efficient for the BPM turn-by-turn data. In our previous study, X.Huang and F.Wang have shown that ICA
can be used to determine the accelerator properties of a rapid cycling accelerator [3]; and determine the vertical dispersion function in high brilliance storage ring for making effective stopband correction [4]. However, the detection and modeling of the nonlinear optics of storage rings has never been addressed by using ICA methods. Since higher order magnetic multipoles are important to the performance of high brilliance accelerators, many attempts have already been made to measure these nonlinear magnetic fields [5, 6, 7]. It would be important for us to apply ICA to facilitate the measurements of multipoles. Since ICA can provide many fold increase in precision compared with PCA, ICA method is a natural candidate for this effort.

This thesis studies the feasibility of using ICA to disentangle the source of non-linearity in accelerators. Sources of non-linearity include sextupole and higher magnetic multipoles. Our study will focus on non-linearity introduced by sextupoles which are commonly used in accelerators for chromatic aberration correction. Four chapters are contained in this thesis. The first chapter describes the motivation of this study and briefly introduce some basic accelerator physics concepts. In chapter two, the fundamental principles and different algorithms of independent component analysis (ICA) are laid out. In our research, a time-related ICA algorithm, second order blind identification (SOBI) is employed. Chapter three is devoted to the ICA analysis of linear betatron motion. The details of the application of the ICA method to the BPM data analysis are presented. A simple lattice with 12 super-periods of FODO cells is used in our simulation. The linear betatron amplitude function and phase are extracted precisely from the measured BPM data by using ICA method. Based upon the analysis of the linear motion, we extend the ICA method onto the nonlinear motion study which is thoroughly discussed in chapter four. A few sextupoles are added into the simple lattice to excite the nonlinear motions. ICA separates out the nonlinear motion which is $2\nu_x$ motion in this case, with good accuracy. Inspired by the $2\nu_x$ motion study, later in chapter four, two methods of beam-based measurement
of the sextupole strength are proposed.

1.2 Basics of Accelerator Physics

1.2.1 Accelerator Coordinate System

When an accelerator is constructed, the nominal trajectory for the ideal particles is fixed. For ideal circular accelerators, like storage rings, this trajectory is a circle. Ideally, particles circulate around the circle indefinitely. However, in a realistic accelerator, most of the particles circulate around the circle with slight deviations. Therefore, a well-designed accelerator must have a finite acceptance around the ideal trajectory. To describe the motion of a particle in the vicinity of the nominal trajectory, we introduce a curvilinear coordinate system for circular machines. In Fig. 1.1,

\[ \mathbf{r}_0 \]

is the reference orbit. It is the ideal nominal trajectory of particles without the presence of any magnets imperfections. \( \mathbf{x}, \mathbf{s} \) and \( \hat{z} \) form the basis of the curvilinear coordinate system. \( s \) is the path length along the designed trajectory. It is designated

\[ \hat{x}, \hat{y}, \hat{z} \]

Figure 1.1: Frenet-Serret coordinate system.
as an independent variable in accelerator physics. Any point in the phase space can be expressed as:

\[ \vec{r} = \vec{r}_0 + x\hat{x} + z\hat{z} \]  

(1.1)

\( \hat{s} \) is the longitudinal coordinate along the reference orbit. The tangent and the transverse unit vectors can be defined as follows:

\[ \hat{s} = \frac{d\hat{r}_0}{ds} \]  

(1.2)

\[ \hat{x} = -\rho(s) \frac{d\hat{s}}{ds} \]  

(1.3)

\[ \hat{z} = \hat{x}(s) \times \hat{s}(s) \]  

(1.4)

where \( \rho(s) \) is the radius of the curvature. The transverse derivatives are given by:

\[ x' = \frac{dx}{ds}, \quad z' = \frac{dz}{ds} \]  

(1.5)

### 1.2.2 Beam Steering Magnets

Speaking of the field strength, steering magnets fall into two categories, conventional ferromagnets and superconducting magnets. Most of the accelerator magnets are the conventional iron magnets which can provide a maximum field strength up to 2 Tesla. However, high energy accelerators, in particular, proton accelerators require higher fields which can only be produced by superconductors.

No matter how strong the field strength, when it comes to the functionality, magnets can be classified as dipole, quadrupole, sextupole, octupole, etc. Figure 1.2 shows schematic drawings of dipole, quadrupole and sextupole magnets. To bend charged beam around a circular path, dipole magnets are used. They generate constant fields along the beam path. The dipole length \( l \) and the bending radius of the curvature \( \rho \) are related by \( l = \rho\theta \), where \( \theta \) is the bending angle. The bending radius of the dipole
can be determined by both the magnetic field strength and the particle momentum.

\[ \rho = \frac{p}{eB_0} \]  

(1.6)

where \( p \) is the momentum of the particle with charge \( e \), \( B_0 \) is the dipole field.

Quadrupole magnets are used to focus or defocus beams. The field strength of a quadrupole disappear along the beam axis and increase linearly with transverse distance. The quadrupole strength, \( B_1 \) is usually defined by:

\[ B_1 = \frac{\partial B_x}{\partial z} = \frac{\partial B_z}{\partial x} \]  

(1.7)

Particles passing through the center of the a quadrupole would not feel any magnetic field. However, with a small deviation from the center, particles will experience an angular kick that is proportional to its transverse position. Based upon the thin lens approximation, the angular kicks are:

\[ \Delta x' = K_1 lx \]  

(1.8)

\[ \Delta z' = -K_1 lz \]  

(1.9)

where \( K_1 = \frac{B_1}{\rho} \). A horizontally focusing quadrupole defocuses vertically and vice versa.
Aside from dipole and quadrupole, another most commonly used magnet is sextupole. Sextupoles are primarily used to compensate the chromatic aberration in strongly focusing magnetic structures. The magnetic field of a sextupole can be expressed as:

\[ B_x = B_2xz \]  \hspace{1cm} (1.10)
\[ B_z = B_2 \frac{(x^2 - z^2)}{2} \]  \hspace{1cm} (1.11)

where the sextupole coefficient \( B_2 \) is defined as \( B_2 = \frac{\partial^2 B_z}{\partial x^2} \). From Eq. (1.10) and (1.11), one can see that sextupoles provide position-dependent focusing/defocusing field.

### 1.2.3 Transverse Motion

#### I. Equation of Motion

In an accelerator, as beam travels along the circular trajectory, it also oscillates around the reference orbit in transverse direction. This kind of transverse motion is called betatron motion.

In Frenet-Serret coordinate system, the betatron motion can be described by the following equations:

\[ x'' - \frac{\rho + x}{\rho^2} = \pm \frac{B_x p_0}{B \rho p}(1 + \frac{x}{\rho})^2 \]  \hspace{1cm} (1.12)
\[ z'' = \mp \frac{B_z p_0}{B \rho p}(1 + \frac{x}{\rho})^2 \]  \hspace{1cm} (1.13)

where \( \rho \) is the momentum of the particle, \( p_0 \) is the momentum of the reference particle. \( \rho \) represents the bending radius of the dipole. \( B \rho = p_0/e \) is called the magnetic rigidity. The plus minus signs correspond to positive and negative charged particles.

With solely dipole and quadrupole magnets in the accelerator, beam motion can be simplified to Hill’s equations:

\[ x'' + K_x(s)x = 0, \quad K_x = 1/\rho^2 \mp K_1(s) \]  \hspace{1cm} (1.14)
\[ z'' + K_z(s)z = 0, \quad K_z = \pm K_1(s) \]  \hspace{1cm} (1.15)

where \( K_1(s) = B_1(s)/B \rho \) is defined as the effective focusing function. \( B_1 = \frac{\partial B}{\partial z} \) is the quadrupole gradient function.

II Floquet Transformation

Generalize the Hill’ equations, we have

\[ y'' + K(s)y = 0 \]  \hspace{1cm} (1.16)

where the focusing coefficient \( K(s) \) is periodic so that \( K(s + L) = K(s) \). \( L \) is the length of a period. The solution of this equation can be obtained by Floquet’s theorem which express the solution in terms of the amplitude and phase function.

\[ y(s) = w(s)e^{j\phi(s)} \]  \hspace{1cm} (1.17)

which satisfies the boundary conditions:

\[ w(s) = w(s + L), \quad \phi(s + L) - \phi(s) = 2\pi\nu \]  \hspace{1cm} (1.18)

\( \nu \) is the phase advance in one period. This leads to the differential equations:

\[ w'' + K(s)w - \frac{1}{w^3} = 0 \]  \hspace{1cm} (1.19)

\[ \phi' = \frac{1}{w^2} \]  \hspace{1cm} (1.20)

Conventionally, the Courant-Snyder parameters can be defined as:

\[ \beta = w^2, \quad \alpha = -ww', \quad \gamma = \frac{1 + \alpha^2}{\beta} \]  \hspace{1cm} (1.21)

Hence, the betatron amplitude function and phase function can be written in forms of:

\[ \frac{1}{2} \beta'' + K(s)\beta - \frac{1}{\beta}[1 + (\frac{\beta'}{\beta})^2] = 0 \]  \hspace{1cm} (1.22)
\[ \phi(s) = \int_{0}^{s} \frac{ds}{\beta(s)} \]  

(1.23)

A general solution of the Hill’s equation is:

\[ y(s) = \sqrt{\epsilon \beta(s)} \sin (\phi(s) + \chi) \]  

(1.24)

where \( \epsilon \) is the emittance of the beam whereas \( \chi \) can be determined by the initial phase. Based upon Eq. (1.23), the betatron tune which is defined as the number of betatron motion in one revolution can be expressed as:

\[ \nu_y = \frac{1}{2\pi} \int_{s}^{s+C} \frac{ds}{\beta(s)} \]  

(1.25)

\( C \) is the circumference of an accelerator. For one revolution, particle phase advances \( 2\pi \nu_y \).

### III Transfer Matrix

The beam phase space coordinates \((y, y')\) can be traced by a transfer map from point to point in an accelerator. For linear system, the phase space coordinates at one particular location can be obtained by propagating from a previous location.

\[
\begin{pmatrix}
    y_2 \\
    y'_2
\end{pmatrix} = M(s_2, s_1) \begin{pmatrix}
    y_1 \\
    y'_1
\end{pmatrix}
\]  

(1.26)

\( M(s_2, s_1) \) is the transfer matrix from \( s_1 \) to \( s_2 \) in any beam transport line which can be expressed as:

\[
M(s_2, s_1) = \begin{pmatrix}
\sqrt{\beta_2} & 0 \\
\frac{\alpha_1}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}}
\end{pmatrix} \begin{pmatrix}
\cos \phi & \sin \phi \\
-\sin \phi & \cos \phi
\end{pmatrix} \begin{pmatrix}
\frac{1}{\sqrt{\beta_1}} & 0 \\
-\frac{\alpha_2}{\sqrt{\beta_1}} & \sqrt{\beta_1}
\end{pmatrix}
\]

where the subscripts 1, 2 indicate the position of the Courant-Snyder parameters. \( \phi = \phi_2 - \phi_1 \) is the phase advance between two position 1 and 2.
For one complete revolution, the transfer matrix can be simplified to:

\[
M = \begin{pmatrix}
\cos \Phi + \alpha \sin \Phi & \beta \sin \Phi \\
-\gamma \sin \Phi & \cos \Phi - \alpha \sin \Phi
\end{pmatrix}
\] (1.27)

where \( \Phi \) represents the phase advance for one complete revolution.

IV Beam Position Monitor

Beam transverse motion signals are picked up by the beam position monitors (BPMs). BPMs are usually composed of two or four conductor plates or button like structures. Fig. 1.3 shows a stripe-line BPM with four conducting metal stripes. As beam passes by, the induced image charges on the stripes can be transmitted to a low impedance circuit. Or the induced voltage can be measured. In general, the beam position can be determined by [8]:

\[
y \approx \frac{w U_+ - U_-}{2 U_+ + U_-} = \frac{w \Delta}{2 \Sigma}
\] (1.28)

where \( U_+ \) and \( U_- \) are either the current or the voltage signals from the right (up) and left (down) plates. \( \Delta = U_+ - U_- \) is named the difference signal. \( \Sigma = U_+ + U_- \) is the sum signal, and \( \frac{w}{2} \) is the effective width of the pickup electrodes. Nowadays, BPM signals are digitized so that by measuring the beam centroid turn by turn, beam transverse motion can be attained. Typically, the structure showed in Fig. 1.3 is used in proton synchrotron, while in electron machines, where the bunch length is much smaller, the button-like structure is usually employed.
Figure 1.3: Schematic drawing of strip beam position monitor [9]
1. Introduction
Chapter 2

Introduction to Independent Component Analysis (ICA)

Nowadays, people are suffering from the so called "information overload" or in another word, "data overload". In almost every scientific discipline, we have access to great amounts of data. However, only a relatively small amount of useful information can be extracted from these vast data sets. Therefore data mining becomes an important technique for uncovering crucial information in complex systems. Independent component analysis (ICA) is essentially a method for extracting useful information from observed data. It is of great interest to a wide variety of scientists and engineers. Because it can reveal the hidden driving forces which underlie the observed phenomena.

ICA belongs to a class of blind source separation (BSS) methods for separating data into underlying components. However, ICA is more powerful than the other BSS methods like principle component analysis (PCA) and factor analysis (FA). This is because what lies at the core of ICA is the "statistical independence". At an intuitive level, if two random variables are independent, then the value of one variable provides
absolutely no information about the value of the other variable. But this is not the case for PCA and FA. They can only find a set of signals which are uncorrelated with each other. Notice that lack of correlation is a weaker property than independence. Independence implies a lack of correlation, but a lack of correlation does not imply independence. This subtle distinction has far reaching consequences for the power of ICA relative to PCA and FA. For example, if both PCA and ICA are applied on a set of microphone outputs which blends several people’s voices together. Then most likely, the signals extracted by PCA would be a new set of voice mixtures whereas the signals coming out of ICA would be a set of signal voices.

This chapter is dedicated to introduce the basic concepts of independent component analysis (ICA). We will start with the basic model and assumption of ICA in Sec. 2.1 then discuss in detail about two types of ICA methods, ICA methods using non-gaussianity (Sec. 2.2) and ICA methods using time structures (Sec. 2.3).

## 2.1 The Basics of ICA

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random signals. ICA assumes that the observed data variables are linear or nonlinear mixtures of some unknown latent variables, and the mixing mechanism is also unknown. The latent variables are assumed mutually independent. They are called the independent components of the observed data. These independent components or hidden sources can be identified by ICA.

ICA can be seen as an extension to principle component analysis (PCA) and factor analysis. It is a much more powerful technique, capable of finding the underlying factors or sources when these classical multivariate statistical methods fail completely.
2.1 The Basics of ICA

2.1.1 ICA Model

We observe $m$ variables $x_1, x_2, \ldots, x_m$, which are modeled as linear combinations of $n$ variables $s_1, s_2, \ldots, s_n$:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{in}s_n$$  \hspace{1cm} (2.1)

for all $i = 1, 2, \ldots, n$, where $a_{ij}$ are some real mixing coefficients. The $s_i$ are statistically mutually independent. The matrix representation of Eq. (2.1) can be expressed as:

$$x = As,$$  \hspace{1cm} (2.2)

where $x$ is the observed data matrix. $A$ is the mixing matrix and $s$ is the source signal matrix. Basic ICA model describes how the observed data are generated by a process of mixing the independent components (ICs) $s_i$. The independent components $s_i$ are the latent variables which means that they are not observable directly. Also the mixing coefficients $a_{ij}$ are unknown. ICA uses solely the observed data $x_i$ to estimate both the ICs $s_i$ and the mixing coefficients $a_{ij}$.

2.1.2 The basic Assumption of ICA

The basic assumption of ICA is that source signals are statistically independent. Statistically, random variables $y_1, y_2, \ldots, y_n$ are said to be independent if the joint probability density function (pdf) $p(y_1, y_2, \ldots, y_n)$ and the marginal pdf of $y_i$ which is denoted by $p(y_i)$ have the following relation.

$$p(y_1, y_2, \ldots, y_n) = p_1(y_1)p_2(y_2)\ldots p_n(y_n).$$ \hspace{1cm} (2.3)

Independence is a stronger constraint than the "uncorrelatedness" which is assumed by principle component analysis (PCA). Two random variables $y_1$ and $y_2$ are uncorrelated, if their covariance is zero.

$$\text{cov}(y_1, y_2) = E\{y_1, y_2\} - E\{y_1\}E\{y_2\} = 0$$ \hspace{1cm} (2.4)
However, if two random variables are independent, the covariance of any function of these two variables is zero, even for the nonlinear functions.

\[ E\{f_1(y_1)f_2(y_2)\} = E\{f_1(y_1)\}E\{f_2(y_2)\} \quad (2.5) \]

Uncorrelatedness is actually only a special case of independence when \( f_1(y_1) = y_1, f_2(y_2) = y_2 \). Based upon the independent source signal assumption, ICA is usually superior to PCA and some other methods that only assume the uncorrelatedness of the source signals. But PCA can serve as a perfect preprocessing method for ICA, since it can effectively reduce the redundancy of the raw data and facilitate the rest of the ICA processing.

Data in science are classified either in ensembles or in time sequences. For data in ensembles, ICA assume that independent components should be non-Gaussian while for data with time structures, ICA relate the independence of the source signals with the diagonality of their auto-covariance matrices. For our purpose, since BPM readings are time dependent signals, the latter scenario is applicable.

### 2.1.3 Application and History of ICA

Due to the generality of ICA method, the data analyzed by ICA could originate from many different kinds of application fields, including digital images, economic indicators as well as psychometric measurements and telecommunication signals. In many cases, the measurements are given as a set of parallel signals or time series. Typical examples are mixtures of simultaneous speech signals that have been picked up by several microphones, brain waves recorded by multiple sensors, interfering radio signals arriving at a mobile phone, or parallel time series obtained from some industrial process.

The technique of ICA is a relatively new invention. It was for the first time introduced in early 1980s in the context of neural network modeling [10]. In mid-1990s,
some highly successful new algorithms were introduced by several research groups [11, 12, 13, 14, 15, 16, 17, 3]. ICA became one of the exciting new topics, both in the field of neural networks, especially unsupervised learning and more generally in advanced statistics and signal processing. Now ICA has been widely applied on biomedical signal processing, audio signal separation, telecommunication, fault diagnosis, feature extraction, financial time series analysis and data mining.

### 2.2 ICA by Maximization of Non-gaussianity

For most of the ICA methods without considering the time structures, non-gaussianity is of paramount importance. ICA assumes that independent components should be non-gaussian. This can be intuitively explained by the central limit theorem which states that a sum of independent random variables tends to a Gaussian distribution. Generally speaking, a sum of two independent random variables has a distribution that is closer to Gaussian than any of the two original random variables. Therefore, in ICA, the observed mixture signals are more Gaussian than any of the single source components. The process to extract these source components is equivalent to finding the components with the least gaussianity. For Gaussian independent components, ICA can not get us further than PCA. Because one important property of Gaussian distribution is that uncorrelated Gaussian are also independent. This can be seen from the probability density function (pdf) of the bi-variate Gaussian distribution:

$$ f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2(1 - \rho^2)} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y} \right) \right) \quad (2.6) $$

\( x \) and \( y \) are two Gaussian variables with zero covariance. \( \rho = \frac{\text{cov}(x, y)}{\sigma_x\sigma_y} \) is the correlation between \( x \) and \( y \). If the covariance of \( x \) and \( y \), \( \text{cov}(x, y) \) is zero, then the pdf of the
bi-variate Gaussian distribution can be simplified to:

\[ f(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} \exp \left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) = f(x)f(y) \quad (2.7) \]

Eq. (2.7) proves the independence of the two random variables. In this case, ICA can only decorrelate components, just like what PCA can do. No independent Gaussian components can be separated from each other by ICA.

To maximizing non-gaussianity in ICA estimation, we must have a quantitative measure of non-gaussianity of a random variable. One classic measure of non-gaussianity is kurtosis, the fourth-order cumulant of a random variable.

### 2.2.1 Measuring Non-gaussianity by Kurtosis

The kurtosis of a random variable \( y \) is defined as:

\[
kurt(y) = E\{y^4\} - 3E\{y^2\}^2
\]

(2.8)

If \( y \) is normalized, \( E\{y^2\} = 1 \), then the kurtosis can be simplified to

\[
kurt(y) = E\{y^4\} - 3
\]

(2.9)

As a measure of non-gaussianity, the kurtosis is zero for Gaussian random variables, negative for sub-gaussians and positive for super-gaussians. The sub-gaussians have flatter probability density function than the Gaussian’, whereas the super-gaussians have sharper peaks and longer tails.

To illustrate the scheme of the kurtosis maximization, we look at a simple two dimensional model. \( x = Ms \), where the independent components \( s_1, s_2 \) both have non-zero kurtosis. \( M \) is the mixing matrix. The goal of ICA is to search for an independent component \( y = A^Tx = A^TMs = b^Ts = b_1s_1 + b_2s_2 \) which has the maximum kurtosis. Here \( A \) is the demixing matrix.

\[
kurt(y) = kurt(b_1s_1) + kurt(b_2s_2) = b_1^4 kurt(s_1) + b_2^4 kurt(s_2)
\]

(2.10)
Make the constrain that $y$ has unit variance. This imposes an constraint $b_1^2 + b_2^2 = 1$. Now the problem of finding independent components has been transformed into the problem of how to find a vector $b = (b_1, b_2)$ which maximizes $kurt(y)$ under the condition of $||b|| = 1$. In practice, we usually use the gradient ascent algorithm to maximize the value of kurtosis. Start from whitened data $z$, $y = w^Tz$ is the component whose kurtosis needs to be maximized. Adjust the direction of matrix $w$ along the direction of the gradient of the absolute value of kurtosis, we can finally find the optimal $w$ which gives the largest kurtosis.

$$\frac{\partial |kurt(w^Tz)|}{\partial w} = 4\text{sign}(kurt(w^Tz))[E\{z(w^Tz)^3\} - 3w||w||^2]$$ (2.11)

Since only the direction of $w$ is of interest to us and the last term in brackets in Eq. 2.11 only affects the norm of $w$, but not its direction, we can simply ignore that term. We obtain the following algorithm:

$$\Delta w \propto \text{sign}(kurt(w^Tz))E\{z(w^Tz)^3\}$$ (2.12)

$$w \leftarrow w/||w||$$ (2.13)

More advanced algorithms, such as the FastICA algorithm, are also developed in order to speed up the convergence and be more reliable.

### 2.2.2 Measuring non-gaussianity by Neg-entropy

The drawback of using kurtosis to measure non-gaussianity is that kurtosis is very sensitive to outliers. A few outliers can drastically change the value of kurtosis. Therefore, people seek for more robust measure of non-gaussianity. Neg-entropy measure is sometimes a better choice. It can be robust but also computationally complicated. The entropy $H$ of a random variable $z$ with density function $p_z(z)$ is defined as

$$H(z) = -\int p_z(z) \log p_z(z) dz$$ (2.14)
According to information theory, under the condition of equal variance, Gaussian variables have the largest entropy, therefore, can be considered as the most random distribution. This provides us a way to measure the non-gaussianity by the measurement of entropy. Define neg-entropy $J$ of an arbitrary random variable $z$ as

$$J(z) = H(z_{\text{gaussian}}) - H(z)$$  \hspace{1cm} (2.15)

Neg-entropy is always non-negative, and is zero if $z$ is Gaussian distribution. In practice, it is really difficult to compute the integral in Eq. (2.14), because the estimation of the probability density function inside could be rather complicated. Therefore approximation of the integral calculation is needed. The classic method of approximating neg-entropy is using the higher-order cumulants. This is base upon the idea that the probability density function (p.d.f.) of a random variable is close to a Gaussian distribution, thus we can do a Taylor-like expansion of the pdf around the Gaussian density. This simple approximation gives us

$$J(z) \approx \frac{1}{12} E\{z^3\}^2 + \frac{1}{48} \text{kurt}(z)^2$$  \hspace{1cm} (2.16)

However, cumulant-based methods sometimes provide a rather poor approximation of entropy, essentially due to the sensitivity to the outliers and the insensitivity to the center structure. So approximation based upon non-polynomial functions are introduced. If we use only on non-quadratic function $G(x)$, the approximation becomes

$$J(z) \approx [E\{G(z)\} - E\{G(\nu)\}]^2$$  \hspace{1cm} (2.17)

where $\nu$ is a Gaussian variable. Similar to the kurtosis method, we can develop a gradient ascent algorithm like this.

$$\Delta w \propto E\{G(w^T z) - E\{G(\nu)\}\}E\{z g(w^T z)\}$$  \hspace{1cm} (2.18)

$$w \leftarrow w / ||w||$$  \hspace{1cm} (2.19)

Aside from the methods aforementioned, ICA can also be implemented by maximum likelihood estimation, minimization of mutual information and so on.
2.3 ICA Using Time Structure

Basic ICA model is based on the assumption that the source signals or the so called independent components are non-gaussian. But when it comes to time signals which may contain much more structure than simple random variables, this assumption may no longer be valid. Instead, we should assume that the ICs have different auto-covariance. Auto-covariance is the covariance between the values of the signal at different time points. For one signal, the auto-covariance is \( \text{cov}(x_i(t)x_i(t - \tau)) \), where \( \tau \) is some lag Constant, \( \tau = 1, 2, 3, \ldots \). If the data has time-dependencies, the auto-covariances are usually different from zero. For two different signals, the auto-covariances is \( \text{cov}(x_i(t)x_j(t - \tau)), i \neq j \). All these covariances can be grouped into a matrix

\[
C^x_\tau = E\{x(t)x(t - \tau)^T\} \tag{2.20}
\]

Due to the independence of the source signal \( s_i(t) \), both the instantaneous and the time-lagged covariance matrix should be diagonal. In other words,

\[
E\{s_i(t)s_j(t - \tau)\} = 0, \quad i \neq j, \quad \tau = 0, 1, 2, 3, \ldots \tag{2.21}
\]

2.3.1 Using One Time Lag and The AMUSE Algorithm

In the simplest case we can use only one time lag, denoted by \( \tau \). This method is called the Algorithm for Multiple Unknown Signals (AMUSE). Consider whitened data \( z \). Then we have the orthogonal separating matrix \( W \):

\[
Wz(t) = s(t) \tag{2.22}
\]
\[
Wz(t - \tau) = s(t - \tau) \tag{2.23}
\]

Now instead of the simple time-lagged covariance matrix, let’s consider a slightly modified version

\[
\bar{C}^z_\tau = \frac{1}{2}[C^z_\tau + (C^z_\tau)^T] \tag{2.24}
\]

Due to the independence of the source signal \( s_i(t) \), both the instantaneous and the time-lagged covariance matrix should be diagonal. In other words,
The new time-lagged covariance matrix is symmetric. Therefore the eigenvalue decomposition is well defined and easy to compute.

\[
\bar{C}_z = \frac{1}{2} W^T \left[ E \{ s(t) s(t - \tau)^T + E \{ s(t - \tau) s(t)^T \} \} \right] W = W^T \bar{C}_s W
\]  

(2.25)

As we have already noticed, \( \bar{C}_s \) is diagonal. Let ’s denote it as D. Then clearly we can have

\[
\bar{C}_z = W^T D W
\]  

(2.26)

What this equation shows is that, we can obtain the demixing matrix \( W \) from the eigenvalue decomposition of \( \bar{C}_z \). Once \( W \) is obtained, getting the source signals is straightforward. But this only works when the eigenvalues of the matrix \( \bar{C}_z \) are all distinct. Therefore, the corresponding eigenvectors are all uniquely defined. Otherwise, the ICs or the source signals we get from this method are not unique. This considerably restricts the applicability of this method. To fix this problem, one can search for a suitable time lag \( \tau \) so that the eigenvalues are distinct. However, this is only possible when signals do not have the same power spectrum or say, identical auto-covariance. To summarize, the procedures of AMUSE algorithm follows these steps: (1) Whiten the data \( x \) to obtain \( z(t) \); (2) Get the time-lagged covariance matrix \( \bar{C}_z \); (3) Do the eigenvalue decomposition to \( \bar{C}_z \) and get \( W \). The rows of \( W \) are given by the eigenvectors.

### 2.3.2 Using Several Time Lags and the SOBI algorithm

As an improvement to the AMUSE algorithm, we start to consider several time lags instead of a single one. Thus the choice of an individual time lag \( \tau \) becomes a less serious problem. One method of this kind is the so called second order blind identification (SOBI). We employ this method in our BPM data analysis.

The general SOBI algorithm is defined by the following implementation:
2.3 ICA Using Time Structure

1. Use the original data \(x(t)\) to estimate the equal time sample covariance matrix \(\hat{C}(0)\). Denote by \(\lambda_1, ..., \lambda_n\) the \(n\) largest eigenvalues and \(h_1, ..., h_n\) the corresponding eigenvectors of \(\hat{C}(0)\).

2. Under the white noise assumption, estimate the noise variance \(\hat{\sigma}^2\) by the average of the smallest eigenvalues of \(\hat{C}(0)\). The whitened signals are \(z(t) = [z_1(t), ..., z_n(t)]^T\), which are computed by \(z_i(t) = (\lambda_i - \hat{\sigma}^2)^{-\frac{1}{2}} h_i^* x(t)\), for \(1 \leq x \leq n\). This is equivalent to forming a whitening matrix \(\hat{V} = [(\lambda_1 - \hat{\sigma}^2)^{-\frac{1}{2}} h_1, ..., (\lambda_1 - \hat{\sigma}^2)^{-\frac{1}{2}} h_n]^H\).

3. Estimates \(\hat{C}(\tau)\) by computing the sample auto-covariance matrices of the whitened data \(z(t)\) for a fixed set of time lags \(\tau \in \{\tau_j | j = 1, ..., K\}\).

4. A unitary matrix \(W\) is then obtained as the joint diagonalizer of the set \(\{\hat{C}(\tau_j) | j = 1, ..., K\}\).

5. The source signals are estimated as \(\hat{s}(t) = \hat{W}^H \hat{V} x(t)\) and the mixing matrix \(A\) is estimated as \(\hat{A} = \hat{V}^{-1} \hat{W}\).

In the 4th step of SOBI algorithm, several auto-covariance matrices are constructed for different time lags, we need to figure out a way to get the joint diagonalizer \(W\) for all the lagged auto-covariance matrices. Mathematically, there is no analytical method to joint diagonalize a series of matrices. However, the joint-diagonalization can be accomplished numerically. First, consider the way to numerically diagonalize one single matrix is to minimize the off-diagonal elements. Here as a measurement of the diagonality of a matrix \(M\), we define an “off” function:

\[
off(M) = \sum_{i \neq j} m_{ij}^2
\]  

which gives the sum of squares of the off-diagonal elements of matrix \(M\). Minimizing \(off(M)\) is equivalent to diagonalizing \(M\). In our case, with the presence of several matrices, we are actually minimizing the following function, under the constraint that
W is orthogonal.

\[ j(W) = \sum_{\tau \in S} \text{off}(W \tilde{C}_\tau W^T) \] (2.28)

The total non-diagonality of \( W \tilde{C}_\tau W^T \) at different time lags can be measured by the sum of the “off” functions for all the time lags we are using. Several schemes can be used to minimize \( j(W) \). One possible way is to use a property of positive definite matrix which states that for any positive definite matrix, we have

\[ \sum_i \log m_{ii} \geq \log |\det M| \] (2.29)

The equality holds only for diagonal matrices. Thus, we can quantify the non-diagonality of matrix \( M \) by a function

\[ F(M) = \sum_i \log m_{ii} - \log |\det M| \] (2.30)

One can define an objective function \( j_1(W) \)

\[ j_1(W) = \frac{1}{2} \sum_{\tau \in S} F(W \tilde{C}_\tau W^T) \]

\[ = \sum_{\tau \in S} \sum_i \frac{1}{2} \log(w_i^T \tilde{C}_\tau w_i) - \log |\det W| - \frac{1}{2} \log |\det \tilde{C}_\tau| \]

\[ = \sum_{\tau \in S} \sum_i \frac{1}{2} \log(w_i^T \tilde{C}_\tau w_i) + \text{const.} \] (2.31)

The minimization of this objective function can be performed by gradient descent.

An alternative approach to minimizing \( j(W) \) in Eq. 2.28 is to express the sum of the squares of the off-diagonal terms as the difference between the total sum of squares of all the elements of a matrix and the sum of squares of the diagonal entries. Hence, minimizing the following function \( j_2 \) is equivalent to minimizing \( j \)

\[ j_2(W) = \sum_{\tau \in S} \sum_{i,j} (w_i^T \tilde{C}_\tau w_j)^2 - \sum_{\tau \in S} \sum_i (w_i^T \tilde{C}_\tau w_i)^2 \] (2.32)
For any orthogonal transformation $W$, the sum of the squares of the elements of matrix $WMW^T$ is constant. Because

$$\sum_{i,j} (WMW^T)_{ij} = \text{Trace}(WMW^T(WMW^T)^T)$$

$$= \text{Trace}(WMM^TW^T) = \text{Trace}(W^TWM^T) = \text{Trace}(MM^T)$$

(2.33)

Therefore the first term in $j_2$ is constant. We can simply maximizing the second term of $j_2$.

In our study, we adopt the second joint-diagonalization method, the so called “Jacobi-like” joint diagonalization algorithm. We now look at this method in more details. This method continuously minimizes the $j(W)$ function by successive given rotations. Suppose we take out a 2 by 2 sub-matrix of the auto-covariance matrix $\tilde{C}^z_\tau$.

$$\tilde{C}^z_\tau = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}.$$  

where $a_{ij}$ denotes the element of $\tilde{C}^z_\tau$ on the $i$th row and $j$th column. Define a new matrix $\tilde{C}^z_\tau'$ as

$$\tilde{C}^z_\tau' = W^H \tilde{C}^z_\tau W (\tau = 1, ..., K)$$

(2.34)

Then a unitary matrix $W$ is sought such that $j(W)$ is minimized. The unitary transformation $W$ is given by a complex rotation:

$$W = \begin{bmatrix} \cos \theta & e^{j\phi} \sin \theta \\ -e^{-j\phi} \sin \theta & \cos \theta \end{bmatrix}.$$  

Let $a'_{ij}$ denote the (i,j)-th entry of matrix $\tilde{C}^z_\tau'$. As we know, orthogonal transformation and unitary transformation preserve the sum of squares of all the elements of a matrix.

$$off(\tilde{C}^z_\tau') + |a'_{ii}|^2 + |a'_{jj}|^2 = off(\tilde{C}^z_\tau) + |a_{ii}|^2 + |a_{jj}|^2$$

(2.35)
Hence, the minimization of \( \text{off}(\bar{C}_z) \) is equivalent to the maximization of \(|a'_{ii}|^2 + |a'_{jj}|^2\). The latter, in turn, indicate the maximization of \(|a'_{ii} - a'_{jj}|^2\), because
\[
2(|a'_{ii}|^2 + |a'_{jj}|^2) = |a'_{ii} + a'_{jj}|^2 + |a'_{ii} - a'_{jj}|^2
\]
and the invariance of the trace under unitary transforms.

Now we define a quantity \( Q \):
\[
Q = \sum_{\tau=1..K} |a'_{\tau ii} - a'_{\tau jj}|^2.
\] (2.36)
which is the sum of \(|a'_{ii}|^2 + |a'_{jj}|^2\) for all the auto-covariance matrices we have constructed. Now we finally transform the ultimate goal of diagonalizing matrix \( \bar{C}_z' \) into the maximization of \( Q \). It is easy to check that
\[
a'_{\tau ii} - a'_{\tau jj} = (a_{\tau ii} - a_{\tau ii}) \cos 2\theta - (a_{\tau ij} + a_{\tau ji}) \sin 2\theta \cos \phi - j(a_{\tau ji} - a_{\tau ij}) \sin 2\theta \sin \phi
\] (2.37)
And there are totally \( K \) such equations since we have used \( K \) different time lags.

Define the vectors
\[
u^T = \begin{bmatrix} a'_{1 ii} - a'_{1 jj}, & .., & a'_{K ii} - a'_{K jj} \end{bmatrix}.
\]
\[w^T = \begin{bmatrix} \cos 2\theta, & -\sin 2\theta \cos \phi, & -\sin 2\theta \sin \phi \end{bmatrix}.
\]
\[g^T = \begin{bmatrix} a_{\tau ii} - a_{\tau jj}, & a_{\tau ij} + a_{\tau ji}, & j(a_{\tau ji} - a_{\tau ij}) \end{bmatrix}.
\]
The \( K \) equations represented by Eq. (2.37) can then be written in the form \( u = Gw \) where \( G^T = [g_1, ..., g_K] \). So \( Q \) can be expressed as:
\[
Q = u^H u = w^T G^H G w = w^T \Re(G^H G) w
\] (2.38)
Here we have used the fact that \( G^H G \) is hermitian. So its imaginary part contributes nothing to the above quadratic form. The last step is to maximize \( Q \) under the
condition that $w^T w = 1$. And the solution is given by the unit norm eigenvector of the $3 \times 3$ matrix $\Re(G^H G)$ associated with the largest eigenvalue. Notice that, in our case, all the auto-covariance matrices are real symmetric matrices. Therefore the rotation matrices are also real. The last component of each $g_k^T$ is zero. And $G^H G$ can be reduced to a $2 \times 2$ matrix.

2.4 Summary

In this chapter, we have introduced the basic concepts of independent component analysis method. We discuss two different kinds of ICA methods, the ICA method by maximizing the non-gaussianity and the ICA method using time structures. For our purpose, since BPM signals are time-correlated data, we use the ICA method with time structure to analyze beam dynamics. The simplest time-related ICA method is the AMUSE algorithm which we have discussed in Sec. 2.3.1. However, this method would easily encounter difficulty as an unfortunate choice of one particular time lag is made. As a remedy, a more advanced method using a series of time lags are introduced in Sec. 2.3.2. In our study, the SOBI algorithm is employed.
2. Introduction to Independent Component Analysis (ICA)
Chapter 3

Independent Component Analysis (ICA) for Linear Motion

The application of ICA to beam diagnosis involves three phases: data acquisition and pre-processing, source signal separation and beam motion identification.

In accelerators, beam position monitors (BPMs) are used to detect the beam transverse motion. There are usually several BPMs around an accelerator, with one or more in a superperiod. Nowadays, BPMs of many accelerators are able to digitize and record their readings on a turn-by-turn base. They can provide massive information about beam transverse motion which could potentially affect the performance of the machine. Usually during the data acquisition process, beam needs to undergo coherent transverse motion while turn-by-turn data are taken. A pinger or RF resonant excitation kicker can be fired once or periodically to excite the beam. BPM gains may be applied to correct the BPM calibration error if necessary.

After the data acquisition, original BPM data have to go through a whitening data process to reduce the redundancy and noise preliminarily. This step can be achieved by principle component analysis (PCA). The choice of the cut-off threshold eigenvalue
\( \lambda \) may play an important role in determining the accuracy of ICA.

Post-PCA steps include constructing unequal time auto-covariance matrices, joint-diagonalizing autocovariance matrices and recovering the source signals using demixing matrix.

Beam transverse motion is usually a linear combination of betatron motion, synchrotron motion and higher order nonlinear motion, etc. ICA can decouple these modes from the original BPM data. Linear betatron motion can be identified by its signature frequency, the betatron tune. The fast Fourier transform (FFT) of the source signals reveal the tune of the betatron motion, while the spatial patterns contain information about the \( \beta \) functions of the accelerator lattice.

In this text, we describe the application of ICA method on linear betatron motion analysis. We will start with the BPM data arrangement and details of ICA algorithm in Sec. 3.1 and then apply the algorithm to simple sinusoidal motion and betatron motion in Sec. 3.2. Betatron amplitude function and phase advance can be obtained. Finally, the effects of noises and errors are discussed in Sec. 3.3.

### 3.1 BPM Data Arrangement and ICA

As reflection of the beam transverse motion, turn-by-turn BPM signals are usually composed of signals of betatron motion, synchrotron motion, higher order nonlinear motions and all kinds of noises. The observation vector can be expressed in terms of linear superposition of the \( n \)-dimensional source signal vector \( s(t) \) and the noise vector \( N(t) \):

\[
X(t) = As(t) + N(t) \tag{3.1}
\]

where \( A \in \mathbb{R}^{m \times n} \) is the mixing matrix with \( m \geq n \) (\( n \) is unknown \( \text{a priori} \)) and \( N(t) \) is the noise vector, assumed to be stationary, zero mean, temporally white and statistically independent of the source signal \( s(t) \). ICA method can simultaneously
3.1 BPM Data Arrangement and ICA determine both the mixing matrix \( \mathbf{A} \) and the source signals \( \mathbf{s}(t) \) from the sampled signal \( \mathbf{X}(t) \), measured at \( t = 0, T_0, 2T_0, \ldots \) for a constant time interval \( T_0 \). Since the source signals in accelerator physics are usually harmonic oscillations with different tunes, the auto-covariances between them are zeros. This is consistent with the basic assumption of the time-related ICA which states that the auto-covariance matrix of independent components should be diagonal. Therefore, it is completely legal to employ this robust method in BPM data analysis. The BPM data matrix holds all the data sampled by \( M \) BPMs in \( N \) turns can be arranged as

\[
x(t) = \begin{pmatrix}
x_1(1) & x_1(2) & \ldots & x_1(N) \\
x_2(1) & x_2(2) & \ldots & x_2(N) \\
\vdots & \vdots & \ddots & \vdots \\
x_M(1) & x_M(2) & \ldots & x_M(N)
\end{pmatrix}
\]

(3.2)

Each row records the BPM readings for one particular BPM at different turns while each column holds the signals picked up by all the BPMs at one particular turn. So the rows display the temporal variation of the beam motion while the columns depict the spatial property of the beam signals. ICA method is based on the assumption that the source signals are independent, therefore

\[
\langle s_i(t)s_j(t - \tau) \rangle = 0, i \neq j, \tau = 0, T_0, 2T_0, 3T_0, \ldots, \quad (3.3)
\]

where \( \langle \cdots \rangle \) is the expectation value or the time average, and \( \tau \) is the time lag. The covariances of vectors can be grouped into matrices:

\[
\mathbf{C}_X(\tau) = \langle \mathbf{X}(t)\mathbf{X}(t - \tau)^T \rangle, \quad (3.4)
\]

\[
\mathbf{C}_S(\tau) = \langle \mathbf{s}(t)\mathbf{s}(t - \tau)^T \rangle = \mathbf{D}(\tau), \quad (3.5)
\]

where the definition of independent component implies that the matrix \( \mathbf{D} \) is diagonal.
The detailed implementation of SOBI (Second Order Blind Identification) ICA algorithm follows these steps:

1. Whitening

The goal of whitening is to reduce the redundancy of the original data and eliminate some of the noises. This is achieved essentially by singular value decomposition (SVD).

First, each row of the BPM data has to be centered at zero by subtracting the mean over the temporal variation. Then compute the $M \times M$ equal-time covariance matrix

$$C_x(0) = <X(t)X(t)^T>$$  \hspace{1cm} (3.6)

Perform Singular Value Decomposition (SVD) to it to obtain:

$$C_x(0) = \left( \begin{array}{c} U_1 \\ U_2 \end{array} \right) \left( \begin{array}{cc} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{array} \right) \left( \begin{array}{c} U_1^T \\ U_2^T \end{array} \right)$$  \hspace{1cm} (3.7)

where \( \Lambda_1, \Lambda_2 \) are diagonal matrices with \( \min(\text{diag}[\Lambda_1]) \geq \lambda_c \geq \max(\text{diag}[\Lambda_2]) \geq 0 \). \( \lambda_c \) is a cutoff threshold set to remove the singularity of the matrix. Whitened data can be obtained by:

$$\xi = \Lambda_1^{-\frac{1}{2}} U_1^T X$$  \hspace{1cm} (3.8)

Since \( XX^T = U_1 \Lambda_1 U_1^T \), it is straightforward to check that \( \xi \) is white.

$$<\xi \xi^T> = <\Lambda_1^{-\frac{1}{2}} U_1^T XX^T U_1 \Lambda_1^\frac{1}{2} > = <\Lambda_1^{-\frac{1}{2}} U_1^T (U_1 \Lambda_1 U_1^T) U_1 \Lambda_1^\frac{1}{2} > = I$$  \hspace{1cm} (3.9)

This step decorrelates and normalizes the original data to facilitate the next ICA step.

2. Joint diagonalization
For a selected set of time lags \( \{ \tau_k | k = 1, 2, ..., K \} \), we first compute the time-lagged auto-covariance matrices

\[
C_{\xi}(\tau_k) = \langle \xi(t)\xi(t + \tau_k)^T \rangle. \quad (3.10)
\]

Instead of using these auto-covariance matrices directly, we construct slightly modified auto-covariance matrices

\[
\bar{C}_{\xi}(\tau_k) = \frac{C_{\xi}(\tau_k) + C_{\xi}(\tau_k)^T}{2} \quad (3.11)
\]

These matrices are real and symmetric, therefore the singular value decomposition (SVD) calculations are well defined and easy to compute. Next, the Jacobi-like joint diagonalization algorithm is applied to find out a unitary matrix \( W \) which is a joint diagonalizer for all the auto-covariance matrices \( \bar{C}_{\xi}(\tau_k) \) of this set.

\[
\bar{C}_{\xi}(\tau_k) = WD_kW^T \quad (3.12)
\]

Here \( D_k \) is a diagonal matrix. The source signals and the mixing matrix can be obtained by

\[
s = W^T V X \quad (3.13)
\]

and

\[
A = V^{-1}W \quad (3.14)
\]

where \( V = \Lambda_1^{1/2} U_1^T \).

### 3.2 Linear Beam Motion Analysis

#### 3.2.1 PCA on Simple Sinusoidal Motion

We consider a simple sinusoidal motion in one dimension:

\[
X(s) = A \sin(\nu_s \phi(s)) \quad (3.15)
\]
where \( A \) is the amplitude, and \( \nu_x \) is the tune and \( \phi \) is the phase angle variable. With \( M \) BPMs in the accelerator and \( N \) turns of tracking, the \((i,j)\) element of the turn-by-turn BPM data matrix \( X(s) \) is

\[
x_{ij} = A \sin(\nu_x \phi_i + 2\pi \nu_x (j - 1))
\]

(3.16)

where, \( i = 1, 2, ..., M \), and \( j = 1, 2, ..., N \). Apply singular value decomposition (SVD) on matrix \( X(s) \) and obtain:

\[
X = U \Lambda V^T,
\]

(3.17)

where

\[
U = \begin{pmatrix}
\sqrt{\frac{2}{M}} \sin(\nu_x \phi_1) & \sqrt{\frac{2}{M}} \cos(\nu_x \phi_1) & 0 & \cdots \\
\sqrt{\frac{2}{M}} \sin(\nu_x \phi_2) & \sqrt{\frac{2}{M}} \cos(\nu_x \phi_2) & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

\[
\lambda = \begin{pmatrix}
A \frac{\sqrt{MN}}{2} & 0 & 0 & \cdots \\
0 & A \frac{\sqrt{MN}}{2} & 0 & \cdots \\
0 & 0 & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

\[
V^T = \begin{pmatrix}
\sqrt{\frac{2}{N}} \cos(2\pi \nu_x \cdot 0) & \sqrt{\frac{2}{N}} \cos(2\pi \nu_x \cdot 1) & \cdots & \sqrt{\frac{2}{N}} \cos(2\pi \nu_x \cdot N) \\
\sqrt{\frac{2}{N}} \sin(2\pi \nu_x \cdot 0) & \sqrt{\frac{2}{N}} \sin(2\pi \nu_x \cdot 1) & \cdots & \sqrt{\frac{2}{N}} \sin(2\pi \nu_x \cdot N) \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

From this simple study, we conclude that for every sinusoidal motion, we need two modes to fully describe it. The normalized matrices \( U \) and \( V \) describe respectively the spatial and temporal properties of the particle motion. Every singular value is a half of the product of the amplitude of the sinusoidal motion and the square root of the BPM number \( M \) and the tracking turn number \( N \). The zero diagonal elements
of the $\lambda$ matrix will not be zero with the presence of noises, finite sampling error and floating point error, etc. If both $M$ and $N$ increase, the errors become less important.

### 3.2.2 Betatron Motion

#### I PCA on Betatron Motion

Now we consider the betatron motion:

$$X(s) = \sqrt{2}\beta_x(s)J \sin(\nu_x \phi(s))$$ \hspace{1cm} (3.18)

where $\beta_x(s)$ is the betatron amplitude function and $J$ is the action. We can set up the data matrix $X(s)$ as described previously, and SVD becomes:

$$U = \begin{pmatrix}
P \sqrt{\frac{2\beta_{x1}}{M}} \sin(\nu_x \phi_1) & P \sqrt{\frac{2\beta_{x1}}{M}} \cos(\nu_x \phi_1) & 0 & \cdots \\
P \sqrt{\frac{2\beta_{x2}}{M}} \sin(\nu_x \phi_2) & P \sqrt{\frac{2\beta_{x2}}{M}} \cos(\nu_x \phi_2) & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$

$$\lambda = \begin{pmatrix}
\frac{\sqrt{2}JMN}{2P} & 0 & 0 & \cdots \\
0 & \frac{\sqrt{2}JMN}{2P} & 0 & \cdots \\
0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$

$$V^T = \begin{pmatrix}
\frac{2}{N} \cos(2\pi \nu_x \cdot 0) & \frac{2}{N} \cos(2\pi \nu_x \cdot 1) & \cdots & \frac{2}{N} \cos(2\pi \nu_x \cdot N) \\
\frac{2}{N} \sin(2\pi \nu_x \cdot 0) & \frac{2}{N} \sin(2\pi \nu_x \cdot 1) & \cdots & \frac{2}{N} \sin(2\pi \nu_x \cdot N) \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots
\end{pmatrix}$$

where $\beta_{xi}$ and $\phi_i$ are the betatron amplitude function and phase at the $ith$ BPM. The coefficient $P$ is introduced to normalize the matrix $U$. Since the action variable $J$ is invariant for linear betatron motion, we find that the eigenvalues of two betatron modes are equal or nearly equal. They depend on the amplitude of the particle.
motion, the number of BPMs and the number of tracking turns. The $\beta$ functions can be attained by summing up the squares of the first two columns.

$$U_1^2 + U_2^2 = \frac{2\beta_x(s)P^2}{M} \propto \beta_x(s) \quad (3.19)$$

And the phase advance can be deduced from

$$\phi(s) = \tan^{-1}\left(\frac{U_1(s)}{U_2(s)}\right) \quad (3.20)$$

In order to compare the betatron amplitude function and phase advance obtained by using Eq. (3.19) and Eq. (3.20) with the model $\beta_x$ function and phase from MAD tracking, we consider a simple lattice with 12 superperiods of FODO cells which resembles the combined function alternating gradient synchrotron (AGS) at the Brookhaven National Laboratory (BNL) (see Fig.3.1). Figure 3.2 shows the comparison between Eq. (3.19) and the $\beta_x(s)$ function obtained from MAD tracking. It can be observed that without any noise or error, Eq. (3.19) and Eq. (3.20) give very good estimation of the $\beta_x(s)$ function and phase $\phi(s)$. However, in real-time data with heavy noises from all kinds of sources, the post-PCA steps become indispensable in order to get accurate estimations.
3.2 Linear Beam Motion Analysis

Figure 3.1: The 7th superperiod of AGS lattice, which contains one sextupole at location 420.37m. The AGS lattice consists of bending dipoles with various gradients.
Figure 3.2: Compare the sum of squares of the first two columns of the $U$ matrix (Eq. (3.19)) with the real $\beta_x$ function from MAD. The initial tracking position is $x_0 = 1mm$. 
3.2 Linear Beam Motion Analysis

Figure 3.3: Compare the phase obtained by Eq. (3.20) with the real phase advance \( \phi \) from MAD. The initial tracking position is \( x_0 = 1mm \).
II ICA on Betatron Motion

Now we apply ICA algorithm to BPM data $X(t)$ to extract the mixing matrix $A$ and the source signals $s$.

$$X = \begin{pmatrix}
    s_1 \\
    s_2 \\
    s_3 \\
    s_4 \\
    \vdots
\end{pmatrix}$$

(3.21)

The betatron oscillation signals which have different phases at different BPMs show two modes with nearly identical frequency spectra. Each mode has one source signal $s_i$ and its corresponding spatial distribution $A_i$, where $A_i$ is the $ith$ column of mixing matrix $A$. The physical origin of each mode can be identified by its spatial vector $A_i$ and the temporal vector $s_i$. Use the AGS-like lattice with a few numbers of sextupoles and 84 BPMs around the ring, we carried out MAD tracking for 1000 turns. Figure 3.4 shows the fast fourier transform (FFT) results of six source signals. It is clear that the first two modes have the frequencies of betatron tune, therefore correspond to betatron motions. The other modes are the results of higher order nonlinear motions caused by sextupoles which we will discuss in detail in the next chapter. Let $X_\beta$ denote the linear betatron motion component of the transverse motion. Then we have

$$X_\beta = A_1 s_1(t) + A_2 s_2(t)$$

(3.22)

Figure 3.5 displays the spatial and the temporal distributions of the betatron modes. Eq. (3.22) can also be expressed as:

$$X_\beta = A_1 s(t) + A_2 c(t)$$

(3.23)
Figure 3.4: The fast Fourier transform (FFT) of source signals for 6 modes. The first two modes correspond to betatron motion. The 3rd and 4th modes are $2\nu_x$ modes which are caused by sextupole kicks. The 5th and 6th are $3\nu_x$ modes. They are the result of the combination of betatron motion and $2\nu_x$ motions. With 1mm initial tracking displacement, the singular values (SVs or eigenvalues) of the first two modes are about 600 times larger than the SV of the 3rd mode, $2 \times 10^5$ times of the 4th SV, $3 \times 10^7$ times of the 5th SV and $10^8$ times of the SV of the 6th mode. Notice that PCA or SVD ranks its modes by the magnitude of the corresponding eigenvalues.
where $s(t)$ and $c(t)$ are temporal functions of the sine-like and cosine-like modes. The betatron amplitude function and phase advance can be obtained.

$$\beta_i = (A_{1,i}^2 + A_{2,i}^2) \times \text{const}$$  \hspace{1cm} (3.24)

$$\phi_i = \tan^{-1}(\frac{A_{1,i}}{A_{2,i}})$$  \hspace{1cm} (3.25)

Figure 3.6 and 3.7 compare the $\beta_x$ and phase derived from Eq. (3.24) and (3.25) with the real $\beta_x$ and phase obtained from MAD. One can conclude that the post-PCA process gains better accuracy for the estimation of $\beta$ function and phase advance by comparing Fig. 3.6 and Fig. 3.7 with Fig. 3.2 and Fig. 3.3.

3.3 Error Analysis

ICA method can successfully decouple different modes from raw BPM data. The linear betatron modes can be identified by their tunes. The betatron amplitude function and phase can be derived from the spatial patterns of the betatron modes. However, the accuracy of the ICA result is subject to sampling error, numerical error and BPM noise.

3.3.1 Effect of Sampling Error

Sampling error or estimation error is caused by observing a sample instead of the whole population. In our analysis, BPM data are taken turn by turn. Information between different turns can not be acquired. This will cause the variation of the calculation accuracy with different sample size. The likely size of the sampling error can generally be controlled by taking a large enough random sample from the population. In our case, more tracking turns can help to ease the sampling error, although the cost of doing this may be expensive. Define $(\frac{\Delta \beta}{\beta})_{\text{rms}}$ as the r.m.s. percentage difference.
Figure 3.5: Left (from top to bottom): normalized spatial function of the first mode $A_1$ (first column of the mixing matrix $A$), temporal function of the first mode $s_1$ (first source signal), FFT of the temporal function of the first mode. Right (from top to bottom): normalized spatial function of the second mode $A_2$ (second column of the mixing matrix $A$), temporal function of the second mode $s_2$ (second source signal), FFT of the temporal function of the second mode.
Figure 3.6: Compare $\beta_x$ from ICA with the real $\beta_x$ at every BPM from MAD. The initial tracking position is $x_0 = 1\text{mm}$.
Figure 3.7: Compare $\phi$ from ICA with the phase advance at every BPM from MAD. The initial tracking position is $x_0 = 1mm$. 
between ICA or PCA calculated $\beta$ function and the real $\beta$ function from MAD.

$$\left( \frac{\Delta \beta}{\beta} \right)_{\text{rms}} = \sqrt{\sum_{i=1}^{M} \frac{(\beta_{\text{ICA/PCA}} - \beta_{\text{MAD}})^2}{\beta_{\text{MAD}}^2 M}}$$

(3.26)

Due to the spurious behavior of the quantity $\left( \Delta \phi \right)_{\text{rms}}$ at small phase values, we quantify the phase measurement accuracy by $\left( \Delta \phi \right)_{\text{rms}}$ which is the r.m.s difference between the ICA or PCA calculated phase and the real phase from MAD.

$$\left( \Delta \phi \right)_{\text{rms}} = \sqrt{\sum_{i=1}^{M} \frac{(\phi_{\text{ICA}} - \phi)^2}{4\pi^2 M}}$$

(3.27)

where $M$ is the number of BPMs in the accelerator. Both $\left( \Delta \beta \right)_{\text{rms}}$ and $\left( \Delta \phi \right)_{\text{rms}}$ are used to define the accuracy of the linear betatron motion calculation. The less the values, the better the accuracy. We found that the accuracy of both ICA and PCA calculation improves as tracking turn number increases, although ICA tends to be more sensitive to the sampling error. Fig. 3.8 shows that ICA calculation can yield better result for every 20 turns of tracking, while for PCA, as shown in Fig. 3.9, every 40 turns of tracking can produce better accuracy. By calculating the sampling errors for different tunes, we have proved that this phenomenon is directly related to the betatron tune of the lattice, i.e. in Fig. 3.9, 40 is the smallest integer that multiplies 8.825 to yield an integer.

Finally, in Fig. 3.10, we compare the sampling error of ICA and PCA calculations and conclude that even for the worst case, ICA can still get about 10 times better accuracy than PCA.

### 3.3.2 Effect of the cut-off threshold $\lambda_c$ and the number of time lags $\tau_k$

An important step of preprocessing for ICA is to whiten the raw data. In the whitening process, singularities can be removed by eliminating the modes with eigenvalues
3.3 Error Analysis

Figure 3.8: $\Delta \beta / \beta$ of ICA calculation for the first two betatron modes vs. different tracking turn numbers. $\nu_x = 8.825$.

less than a threshold value $\lambda_c$. This process reduces the redundancy of the original data matrix and also preliminarily cut off some of the noises, therefore, facilitates the rest ICA procedures. However, the choice of the threshold value is critical for ICA. It determines how many modes are left, or what kind of information is worth keeping for the rest of the ICA steps. The threshold value $\lambda_c$ needs to be small enough to keep all the significant modes which really have physical meanings yet large enough to eliminate the singularities.

One easier way to set the threshold value $\lambda_c$ is to decide how many modes one needs to keep. From Fig. 3.11, one can observe that if we keep more than eight eigen-modes in the whitening step, which means that $\lambda_c$ is set to be the eigenvalue of the eighth eigenmode, the accuracy of ICA deteriorates noticeably as the mode number increases. In this simulation, since 84 BPMs are used to detect the beam motion.
Figure 3.9: $\Delta \beta / \beta$ of PCA calculation for the first two betatron modes vs. different tracking turn numbers. $\nu_x = 8.825$.

Totally 84 modes would appear if no cut-off action is carried out. But as we have shown in Fig. 3.4, among the 84 eigenmodes, only six modes with meaningful physical origins emerge. All the other modes are the decomposed closed orbit which do not bear any signature frequencies. PCA preprocessing step preparatorily disentangles different motions. However, this disentanglement is preliminary. There are still mixings between different physical motions and the closed orbit modes. The more modes other than the six identifiable modes are kept in the preprocessing step, the more couplings are introduced into the post-PCA steps, therefore the less accurate the final ICA result would be.

The joint diagonalization of several auto-covariance matrices is intended to reduce the possibility that an unfortunate choice of one particular time lag $\tau$ results in degeneracy of the auto-covariance matrix of Eq. 3.12. In that case, no uniquely defined
demixing matrix $W$ can be found. One can imagine that the more auto-covariance matrices are used, the more constraints are applied to the joint diagonalizer $W$, therefore possibly the better accuracy can be achieved. However, joint diagonalizing a too large set of auto-covariance matrices will not bring any benefit. Instead as shown in Fig. 3.11, it will decrease the accuracy of ICA. Because the way to numerically joint-diagonalize auto-covariance matrices is to tune the rotation matrix $W$ until we finally minimize the sum of the square of the off-diagonal terms of all the auto-covariance matrices. The more matrices we have the more difficulties we would encounter in adjusting $W$ in order to diagonalize all the matrices at the same time.

In the presence of BPM noises as we will see in the next section, the choice of number of time lags will not affect the ICA results as much. (See Fig. 3.12)
Figure 3.11: Top: $\langle \Delta \beta / \beta \rangle_{\text{rms}}$ vs. number of time lags for different number of eigenmode kept in whitening process. Tracking initial position is $x_0 = 1 \text{mm}$. Bottom: $\langle \Delta \phi \rangle_{\text{rms}}$ vs. number of time lags for different number of eigenmode kept in whitening process.
Figure 3.12: Top: $(\Delta \beta / \beta)_{\text{rms}}$ vs. number of time lags for different BPM noise levels. Tracking initial position is $x_0 = 1mm$. Six eigenmodes are kept in the whitening process. Bottom: $(\Delta \phi)_{\text{rms}}$ vs. number of time lags for different BPM noise levels. Tracking initial position is $x_0 = 1mm$. Six eigenmodes are kept.
3.3.3 Effect of BPM noise

BPM data inevitably contain noises which will affect the ICA estimation accuracy. Usually in electron accelerators, the BPM noise level is about 100µm, while in proton machines the noise is usually about 200µm. In our simulation, we add uniform white noise to MAD tracking data to simulate the effect of BPM noise. Fig. 3.13 and 3.14 show the estimated β function and phase for noise level 100µm and 200µm respectively using one random seed. For this particular random seed, ICA can keep the estimation accuracy of βx within 1% for 100µm noise and 2% for 200µm noise. The estimation error of the phase is kept with 0.001 for 100µm noise, and 0.0015 for 200µm noise level. With only one random seed being considered, this estimation may not be true for the other cases. In order to generally quantify the estimation accuracy of both the β function and phase under different tracking conditions and noise levels, a systematic investigation has been carried out as shown in Fig. 3.15 and 3.16. 20 random seeds are used. The r.m.s. percentage differences \( (\Delta \beta)_{rms} \) and \( (\Delta \phi)_{rms} \) are calculated for BPM noise level ranging from 0µm up to 500µm and initial tracking displacement at 1mm, 2mm, 5mm and 10mm. First, the results indicate that ICA can achieve higher precision than PCA in any case. Second, for both ICA and PCA, the estimation error grows as the noise level increases. However, even with noise level as high as 500µm and initial tracking displacement as small as 1mm, the error of the ICA estimation of the β function measured by the r.m.s. percentage differences would most likely be around 1.5%. And the ICA phase estimation error is kept below 0.0015 for all the cases. Figure 3.16 shows the scaling property of the ICA error estimation. It implies that the larger the initial tracking displacement, the less the error and the error is inversely proportional to the initial displacement \( x_0 \). But when \( x_0 \) is as large as 10mm, this scaling property will on longer hold.
3.3 Error Analysis

Figure 3.13: In electron machine, the BPM noise level is about 100µm. The estimation error for β function and phase using one random seed. The large error in the phase calculation at small phase angle is derived from the numerical error of dividing by a very small number.
Figure 3.14: In proton machine, the BPM noise level is about 200µm. The estimation error for β function and phase using one random seed.
3.3 Error Analysis

Figure 3.15: The ICA and PCA errors of $\beta$ function and phase advance vs. different BPM noise levels for various initial positions of tracking. The solid lines correspond to ICA results. Dashed lines are for the PCA results. The blue, red, green and black lines correspond to initial tracking position $x_0 = 1mm, 2mm, 5mm, 10mm$ respectively.
Figure 3.16: The scaled errors of $\beta$ function and phase advance vs. different BPM noise levels for various initial positions of tracking. When $x_0$ is too large the scaling property would no longer be valid.
3.4 Summary

In this chapter, we have thoroughly detailed the application of the time-related ICA method on linear transverse beam motion. We first start with the BPM data arrangement and the implementation of the ICA algorithm in Sec. 3.1. We then move on to the study of the betatron motion in Sec. 3.2. Both PCA and ICA analysis on betatron motion have been carried out. We conclude that ICA can decouple different eigenmodes further more than PCA, therefore can achieve better accuracy. In the last section, the influences of sampling error, cut-off threshold $\lambda_c$, the number of time lags $\tau_k$ and BPM noise are discussed. In the presence of BPM noises, the choice of $\lambda_c$ and $\tau_k$ play less important roles in the analysis accuracy. In order to suppress the influence of large noises, we can use a reasonably large initial tracking displacement. Like shown in Fig. 3.15, with 10mm initial displacement, the ICA error for $\beta$ function estimation can be limited within 0.2%. For phase estimation, error can be as small as $2 \times 10^{-4}$. 
3. Independent Component Analysis (ICA) for Linear Motion
Chapter 4

Independent Component Analysis (ICA) for Nonlinear Motion

This chapter presents the ICA study of the nonlinear beam dynamics. Beam motion can be treated as a combination of both linear and nonlinear motions. In the previous chapter, the linear part of beam transverse motion was discussed. As we have seen, principle component analysis (PCA), although can not compete with ICA on better accuracy, is actually capable of extracting betatron amplitude functions. However, for nonlinear beam motions, it is often found that PCA is not sufficient to recover the source signals, therefore make the nonlinear motion difficult to analyze. But with ICA, which promises to extract independent components, we could accurately single out the nonlinear modes and their corresponding spatial wave functions.

This chapter investigates the feasibility of using ICA to disentangle the source of non-linearity in an accelerator. Sources of non-linearity can be sextupoles and higher order magnetic multipoles. Nonlinear magnetic field plays an important role in the quality of particle beam in accelerators. There have been many efforts to measure the nonlinear magnetic fields [5, 6, 7]. In Ref. [5], phase space ellipses were
measured and resonance strengths were determined by fitting phase space ellipses with a nonlinear Hamiltonian. To determine all nonlinear multipoles in an accelerator, one needs to measure Hamiltonian in the entire betatron tune space, which needs many measurements. Similarly, Benedikt et al in Ref. [6] employed the longitudinal variation of resonance Fourier amplitude to extract resonance strengths; and Bartolini et al in Ref. [7] added parametric fitting to fit the extracted resonance widths with the modeled sextupole strengths. As shown in Refs. [6, 7], the extracted resonance strengths do not have enough precision.

Since ICA can provide much better precision compared with PCA, ICA analysis is a natural candidate in the effort to extract sextupole strengths. In our study, we consider the combined function alternating gradient synchrotron (AGS) lattice at the Brookhaven National Laboratory (BNL) which consists of twelve superperiods of FODO cells. We add sextupoles in the lattice. Particle tracking is carried out using MAD[18]. To simplify our study, we will only consider the horizontal motion of the beam. The vertical initial tracking position is set to be zero. Ignoring the coupling effect, no vertical motion will be induced through out our study. In Sec. 4.1, the analytical calculation of the $x_{2\nu}$ mode is discussed. During the data analysis, PCA is first employed in order to roughly separate out distinct modes. The singular values of the nonlinear motion are carefully studied in Sec. 4.2. With the introduction of ICA, the $x_{2\nu}$ component of betatron motion is discussed in Sec. 4.3. Based upon the study of the $x_{2\nu}$ motion, in Sec. 4.4, we finally propose two beam-based methods to identify the strength of sextupoles in an accelerator.
4.1 Equation of Motion for $2\nu_x$ mode

The $2\nu_x$ modes can arise from non-linearity in the Hill’s equation:

$$x'' + K_x(s)x = \frac{B_2(s)}{2B\rho}(x^2 - z^2). \quad (4.1)$$

where $x$ and $z$ are the horizontal and vertical betatron coordinates, the “time” variable $s$ is the distance along the reference orbit, $x' = dx/ds$ and $x'' = d^2x/ds^2$, $B_2 = \frac{\partial^2 B_z(s)}{\partial s^2}$ is the sextupole strength located at $s = s_{\text{sext}}$. For a short sextupole, we can use a localized kick: $(B_2/B\rho) = \sum K_2 L \delta(s - s_{\text{sext}} - nC)$, where $C$ is the circumference, $n$ is integer. We consider the case of pure horizontal motion, so the initial vertical position is zero.

We carry out Floquet transformation to the Hill’s equation by using the coordinate $\xi = \frac{x}{\sqrt{\beta_x}}$, and changing the “time” variable $s$ to $\phi = \frac{1}{\nu_x} \int \frac{ds}{\beta_x}$. The Hill’s equation becomes

$$\ddot{\xi} + \nu^2 x \xi = \nu_x^2 \beta_x^{3/2} \frac{B_x}{2B\rho} \sum_n \delta(s - s_{\text{sext}} - 2n\pi)$$

$$= \nu_x^2 \beta_x^{3/2} \frac{B_x}{2B\rho} (\sum_n \delta(\phi - \phi_{\text{sext}} - 2n\pi) \frac{1}{\nu_x \beta_x})$$

$$= \frac{1}{2} \nu_x \beta_x^{3/2} K_2 L [\cos(2\nu_x \phi + 2\chi) + 1] \sum_n \delta(\phi - \phi_{\text{sext}} - 2n\pi) \quad (4.2)$$

where $\ddot{\xi} = d^2\xi/d\phi^2$, $\nu_x$ is the betatron tune, and $K_2 L = \int B_2(s)ds/B\rho$ is integrated sextupole strength. The solution of the linearized Hill’s equation $x_0 = \sqrt{2/\beta_x \int \cos(\nu_x \phi + \chi)}$ was substituted to the right hand side Eq. (4.1).

The solution of Eq. (4.2) is $\xi = \xi_0 + \xi_1$, where $\xi_0 = x_0/\sqrt{\beta_x}$, the normalized betatron motion, is the solution of the homogeneous equation, and $x_1 = \sqrt{\beta_x} \xi_1$. Here $\xi_1$ is the particular solution of Eq. (4.2). To obtain the particular solution, we use
the Fourier expansion on the right hand side of Eq. (4.2):

\[
\frac{1}{2} \beta_3 \beta_s \sum_n \delta(\phi - \phi_{\text{sext}} - 2n\pi) = \sum_k f_k e^{ik\phi}
\]

\[
f_k = \frac{1}{4\pi} \sum_{\text{sext}} \beta_3 \beta_s K_2 L e^{-ik\phi_{\text{sext}}}
\]

(4.3)

to obtain

\[
\ddot{\xi} + \nu_x^2 \xi = \nu_x J_x \sum_k f_k e^{ik\phi} [\cos(2\nu_x \phi + 2\chi) + 1]
\]

\[
= \frac{\sum_k f_k \nu_x J_x}{2} e^{i[(2\nu_x + k)\phi + 2\chi]} + \sum_k \frac{f_k \nu_x J_x}{2} e^{-i[(2\nu_x - k)\phi + 2\chi]}
\]

\[
+ \sum_k f_k \nu_x J_x e^{ik\phi}
\]

(4.4)

By solving this equation, one can get the particular solution \(\xi_1\).

\[
\xi_1 = \sum_k \frac{f_k \nu_x J_x}{2[\nu_x^2 - (2\nu_x + k)^2]} e^{i[(2\nu_x + k)\phi + 2\chi]} + \sum_k \frac{f_k \nu_x J_x}{2[\nu_x^2 - (2\nu_x - k)^2]} e^{-i[(2\nu_x - k)\phi + 2\chi]}
\]

\[
+ \sum_k \frac{f_k \nu_x J_x}{\nu_x^2 - k^2} e^{ik\phi}
\]

(4.5)

If we define \(f_k = |f_k| e^{i\phi_k}\), and \(f_{-k} = f_k^*\). Eq. (4.2) and (4.5) can be simplified to

\[
\ddot{\xi} + \nu_x^2 \xi = \frac{1}{2} \nu_x J_x \beta_3 \beta_s K_2 L [\cos(2\nu_x \phi + 2\chi) + 1] \sum_n \delta(\phi - \phi_{\text{sext}} - 2n\pi)
\]

(4.6)

and

\[
\xi_1 = \sum_k \frac{|f_k| \nu_x J_x}{\nu_x^2 - (2\nu_x + k)^2} \cos[(2\nu_x + k)\phi + 2\chi + \varphi_k] + \sum_k \frac{|f_k| \nu_x J_x}{\nu_x^2 - k^2} \cos[k\phi + \varphi_k]
\]

(4.7)

where \(J_x = \frac{x_0^2}{2\beta_s}\) is the action of the betatron motion. Notice that after the beam passes through a sextupole, the total beam action will change. So here \(J_x\) only denotes the action of linear betatron motion, or in another word, the action at the beginning of the tracking when no sextupole has been encountered. The last term in Eq. (4.7) is the closed orbit induced solely by sextupole kick.
4.2 Singular Values of the Nonlinear Modes

Singular values are obtained in ICA preprocessing, which is essentially the singular value decomposition (SVD). Totally six distinct modes are found as shown in Fig. 4.1. The first two highest-ranked singular values correspond to the betatron motion. All the other higher order nonlinear modes appear due to the existence of sextupoles. The 3rd and the 4th modes are introduced solely by sextupole kick, while the 5th and 6th modes are the results of the combination of both linear betatron motion and sextupole kick. By looking at the source term on the right hand side of the Hill’s equation Eq. (4.1) and assuming that beam horizontal motion is a simple superposition of the betatron motion and $x_{2\nu_x}$ motion, one can find that both $2\nu_x$ and $3\nu_x$ motions can be introduced. On the right hand side of Eq. (4.1)

$$\frac{B_2(s)}{2B\rho}x^2 = \frac{B_2(s)}{2B\rho}(x_\beta + x_{2\nu_x})^2$$

$$= \frac{B_2(s)}{2B\rho}(x_\beta^2 + 2x_\beta x_{2\nu_x} + x_{2\nu_x}^2) \quad (4.8)$$

Since betatron motion $x_\beta$ bear the tune of $\nu_x$, the first term in Eq. (4.8) introduces the $2\nu_x$ modes whereas the second term causes $3\nu_x$ motion. The $4\nu_x$ modes which appear as the third term in Eq. (4.8) are buried under noises. In this section, we will just focus on the 3rd and 4th modes at the tune of $2\nu_x$. In the ICA preprocessing step, the singular values from SVD contains information about the amplitude of different beam motions. We need to look into the details of the singular values to start the study of the beam motion. By varying the number of sextupoles, sextupole strengths and positions, we systematically study the 3rd and 4th singular values.

4.2.1 Analysis of Singular Value with One Sextupole

When only one sextupole is placed in the lattice, the singular value of the 3rd mode is linearly proportional to the sextupole strength and the square of the initial coordinate
Figure 4.1: The fast fourier transform (FFT) of the first 6 columns of temporal matrix $V$. All other modes due to truncation errors, noises are small. The first two modes are betatron modes. The 3rd and 4th are $2\nu_x$ modes. The 5th and 6th are $3\nu_x$ modes. Only 6 modes are relevant in this example.
of tracking $x_2^2$. This observation is consistent with what has been shown in Eq. (4.5) and Eq. (4.7). The amplitude of the particular solution $\xi_1$ is proportional to the factor $f_k$ which in turn depends on the integrated sextupole strength $K_2L$ (see Eq. (4.3)), and also the action variable $J_x$ which is $\frac{x_2^2}{2\beta_x}$ at the start point of beam tracking. Although the 4th mode contains essential information about the non-linearity of particle motion, its singular value (SV) is much smaller than the SV of the 3rd mode. However, both of them are proportional to the square of the initial tracking amplitude. Fig. 4.2 and 4.3 show the SVs of the 3rd and 4th modes normalized by $x_1^2$ vs. the integrated sextupole strength $K_2L$ for an ideal AGS lattice with a single sextupole.

In both Fig. 4.2 and 4.3, singular values are non-zero with zero sextupole strengths. This is especially apparent for the 4th mode’s SVs due to the smaller values. Since no sextupoles are turned on in the lattice, they must come from the intrinsic non-linearity of the lattice. Once the sextupoles are turned on, the 4th SVs experience sudden jumps in value.

4.2.2 Analysis of Singular Values with two Sextupoles

As we have shown in the previous section, with only one sextupole in the lattice, the 3rd singular value (SV) is more than 1000 times larger than the 4th singular value (SV). That means the 3rd mode contains most of the information about $2\nu_x$ motion, although the 4th mode is also indispensable to obtain an accurate picture of the motion. However, with two or more sextupoles in the lattice, the role played by the 4th mode becomes more and more important. The ratio between the 4th SV and the 3rd SV is dramatically enhanced from about 1 over 1000 to 1 over 5, if two sextupoles are identical, and 1 over 2 when they have opposite signs. (see Figure 4.4 and 4.5). Notice that the way the SVD (or PCA) arranges its modes is ranking by the values of SVs from high to low, therefore the 4th SV is always smaller than the
Figure 4.2: The SVs of the 3rd and 4th modes vs the sextupole strength $K_2L$ for a single sextupole in an ideal AGS lattice with horizontal tune 8.825. The sextupole is located at the 7th superperiod of the ring. The $\beta$ function at that location is 15.217m. Note that the SV of the 3rd mode is much larger than that of the 4th mode. However, if there are more than one sextupoles with alternating signs of field strengths. The 4th mode will also be proportional to sextupole strength.
Figure 4.3: The SVs of the 3rd and 4th modes vs the sextupole strength $K_2L$ for a single sextupole in an ideal AGS lattice with horizontal tune 8.501. The sextupole is located at the 7th superperiod of the ring. The $\beta$ function at that location is 15.804 m. Because the tune is close to half integer, higher sextupole strength may result in larger SV, thereby distorts the linear dependence of the SV on sextupole strength $K_2L$. This is more apparent for the 4th mode whose SV is much smaller therefore easier to see.
3rd SV. Figure 4.4 and 4.5 also show that different signs in sextupoles can result in the larger 4th SV and also the larger ratio between the 3rd and the 4th SVs. That indicates a stronger need for the 4th mode in order to fully represent the complicated perturbation excited by the sextupoles with opposite signs.

**Figure 4.4:** The SVs of the 3rd and 4th modes vs. the sextupole strength $K_2L$ when two identical sextupoles are placed in an ideal AGS lattice with horizontal tune $8.825$. The sextupoles are located respectively at the 3rd and 7th superperiod of the ring. The $\beta$ function at the sextupole locations are $15.217m$. Compare to single sextupole cases, the 4th mode SV is relatively larger.
Figure 4.5: The SVs of the 3rd and 4th modes vs the sextupole strength $K_2L$ when two sextupoles with opposite signs are placed in an ideal AGS lattice with horizontal tune 8.825. The sextupoles are located respectively at the 3rd and 7th superperiod of the ring. The $\beta$ function at the sextupole locations are 15.271m. The alternating signs of the sextupoles enhance SVs of the 4th mode.
4.3 Compare ICA with the analytical calculation

ICA decomposes BPM data matrix into source signals $s_i$’s and mixing coefficients $A_i$’s. $s_i$’s can be any kind of beam motion, i.e. betatron motion, synchrotron motion, higher order nonlinear motion etc.

$$X = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \begin{pmatrix} A_1 & A_2 & A_3 & A_4 & \ldots \end{pmatrix}$$

BPM data matrix $X$ is a mixture of all modes. For $2\nu_x$ mode, only part of the data which is denoted by $x_{2\nu}$ is of interest:

$$x_{2\nu} = \begin{pmatrix} . \\ . \\ s_3 \\ s_4 \\ \vdots \\ \vdots \end{pmatrix}$$

or in another form,

$$x_{2\nu}(s, t) = A_3(s)s_3(t) + A_4(s)s_4(t). \tag{4.9}$$

We consider the AGS lattice with twelve superperiods (Fig. 3.1). One sextupole is placed in the ring at location $420.37m$ (the 7th superperiod). The sextupole strength is $K_2L = 1.0(m^{-2})$. Figure 4.6 shows that the 3rd and 4th columns of mixing matrix
4.3 Compare ICA with the analytical calculation

A, the 3rd and 4th rows of source signal $s$ and also the resulting $x_{2\nu}$. From Fig. 4.7, one can find very good agreement of the $x_{2\nu}$ from ICA calculation with the analytical result obtained by the first two terms of Eq. (4.7). Figure 4.8 demonstrates the same agreement as Fig. 4.7 for two consecutive turns. The amplitude of the particle motion changes when particles encounter sextupole kicks. Figure 4.9 and 4.10 further display that with more than one sextupole, i.e. two and ten sextupoles in the lattice, $x_{2\nu}$ still agrees with analytical result with high precision. In Fig. 4.11, we compare the closed orbit term (the last term) of Eq. (4.7) caused solely by the sextupole kick with the closed orbit extracted by ICA. In ICA, the closed orbit is calculated by subtracting $x_{\nu}, x_{2\nu}$ and $x_{3\nu}$ from the original BPM data $X(s, t)$.

We have shown that ICA can disentangle the $x_{2\nu}$ mode very accurately. This usually is difficult for PCA to achieve. Figure 4.12 displays the comparison of the spatial wave functions and the recovered $x_{2\nu}$ by using both PCA and ICA. For the PCA modes, the original data have $X(s, t) = U \times D \times V^T$, where $D$ is a diagonal matrix. $x_{2\nu}$ can be obtained by combining both the spatial and temporal vectors corresponding to the $2\nu_x$ motion.

$$x_{2\nu} = \begin{pmatrix} u_3 & u_4 \end{pmatrix} \begin{pmatrix} d_3 & 0 \\ 0 & d_4 \end{pmatrix} \begin{pmatrix} v_3 \\ v_4 \end{pmatrix}$$

Clearly, the spatial wave functions (spatial mixing coefficients) $A_3, A_4$ extracted by ICA obey the linear betatron motion outside the sextupole position, while the PCA spatial wave function $u_4$ does not. PCA-post process is very important in order to obtain proper spatial wave function and $x_{2\nu}$ motion. However, it is still essential to keep PCA as a preprocessing procedure for ICA.

Figure 4.13 to 4.16 investigate the dependency of the accuracies of the $x_{2\nu}$ motion and the closed orbit recovered by ICA upon the choice of the horizontal tune. $\Delta x_{2\nu}$ and $\Delta x_{\text{closed orbit}}$ are calculated by comparing the differences between $x_{2\nu}$ and $x_{\text{closed orbit}}$. 
Figure 4.6: AGS lattice with single sextupole at 420.37m and strength $K_2 L = 1$. From top to bottom, left to right: 3rd column of the mixing matrix $A$ divided by $\sqrt{\beta_x}$, 4th column of the mixing matrix $A$ divided by $\sqrt{\beta_x}$, 3rd row of temporal function $s$, 4th row of temporal function $s$, FFT of $s_3$, FFT of $s_4$, $x_{2\nu}$ divided by $\sqrt{\beta_x}$ which follows Eq. (4.9). The sextupole location was marked by a red line.
4.3 Compare ICA with the analytical calculation

Figure 4.7: AGS lattice with single sextupole located at 420.37m with sextupole strength $K_2L = 1$. The vertical line marks the location of the sextupole.
Figure 4.8: AGS lattice with single sextupole located at 420.37 m with sextupole strength $K_2 L = 1$. Results for two tracking turns are compared.
Figure 4.9: AGS lattice with two sextupoles located at 185m and 420.37m with strengths $K_2L = 1m^{-2}$ and $K_2L = -1.5m^{-2}$ respectively. Green lines indicate the locations of the two sextupoles.
Figure 4.10: AGS lattice with ten sextupoles located at various locations around the ring. Green lines indicate the locations of the sextupoles.
Figure 4.11: Compare the last term of Eq. (4.7) with the normalized closed orbit spatial wave function obtained by ICA. Same lattice layout as that of Fig. 4.7.
**Figure 4.12:** AGS lattice with single sextupole at 420.37 m and strength $K_2L = 1$. The sextupole location is indicated by a red dashed line. PCA can not extract proper $x_{2\nu}$ while ICA can. Notice that $u_4$ does not display any betatron motion like feature outside of the sextupole location.
recovered by ICA with the first and second terms of Eq. (4.7). It is obvious that
the accuracies are affected by resonances. When the horizontal tune is close to an
integer, or one third, one fourth, three fourths of an integer, the accuracy of our
estimation deteriorates. This is simply because that Eq. (4.7) is obtained using small
perturbation method which no longer holds true when resonances are encountered.
However, from Eq. (4.7), one can also see that when tune is close to an integer or one
third of an integer the $x_{2\nu}$ motion should actually be amplified and therefore should
be easier to be extracted. Obviously, Fig. 4.13 to 4.16 prove that this influence is
less important than the large perturbation introduced by resonances.

4.4 Beam-based measurement of sextupole strength

Figure 4.7 conveys two important messages: (1) Sextupole kick can change the ampli-
tude of the $x_{2\nu}$ motion, therefore the amplitude of the total motion would also change.
(2) Before and after the sextupole kick, $x_{2\nu}$ motion follows the pattern of simple be-
tatron oscillation. This indicates that outside of sextupole location, transverse beam
motion which is composed of both linear betatron motion and $x_{2\nu}$ motion, can be
expressed in term of simple sinusoidal functions. Inspired by these two messages, we
propose two different methods for beam-based measurement of sextupole strengths.

4.4.1 Action change of $x_{2\nu}$ and sextupole strength

As showed in Fig. 4.7, after every sextupole kick, the amplitude of the $x_{2\nu}(s,t)$
changes. Since the amplitude of the betatron motion can not be perturbed by small
nonlinear element kicks, the amplitude change of $x_{2\nu}$ will directly result in the change
of the action variable $J$. One can find a relation between the change of action and
the sextupole strength. At the location of a short sextupole, the change of action of
4. Independent Component Analysis (ICA) for Nonlinear Motion

Figure 4.13: AGS lattice with one sextupole. The initial tracking position is 1mm. Top: differences between the $x_{2\nu}$ from ICA with the first term of Eq. (4.7) for different tunes. Middle: differences between the closed orbit from ICA with the second term of Eq. (4.7) for different tunes. Bottom: ratio of the singular value for the third mode (the mode corresponds to $x_{2\nu}$ motion) over the average of the first two singular values (correspond to linear betatron motion).
Figure 4.14: Zoom into the tune range around the third order resonance for Fig. 4.13
Figure 4.15: Same as Fig. 4.13 except that the initial tracking position is 10mm.
Figure 4.16: Zoom into the tune range around the third order resonance for Fig. 4.15
Independent Component Analysis (ICA) for Nonlinear Motion

Particle motion is given by

\[ \Delta J = \frac{1}{2\beta} \Delta \left[ x^2 + (\alpha x + \beta x')^2 \right] \]

\[ = (\alpha x + \beta x') \Delta x' \]

\[ = \frac{1}{2} K_2 L x^2 (\alpha x + \beta x'). \quad (4.10) \]

The minute change of action is embedded in the second harmonic of the betatron motion. It is normally difficult to measure this change of action in tracking data. Using beam position monitors (BPMs), we can only measure particle transverse displacement \( x \), but not the \( x' \). Since ICA can effectively separate out the first, second harmonic and also the closed orbit term, as shown in Fig. 4.7 to Fig. 4.11, we can restore the particle motion \( x \) by ICA and then express the \( x' \) and action of the beam motion at one particular location in terms of coordinate \( x' \)'s at that location and the adjacent position [8]. Using the transfer matrix form \( s_1 \) to \( s_2 \) in beam transport line:

\[
\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\beta_2}{\beta_1}} (\cos \psi_{21} + \alpha \sin \psi_{21}) & \sqrt{\frac{\beta_1 \beta_2}{\beta_1}} \sin \psi_{21} \\ -\frac{1+\alpha_1}{\sqrt{\beta_1 \beta_2}} \sin \psi_{21} + \frac{\alpha_1-\alpha_2}{\sqrt{\beta_1 \beta_2}} \cos \psi_{21} & \sqrt{\frac{\beta_1}{\beta_2}} (\cos \psi_{21} - \alpha_2 \sin \psi_{21}) \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix}
\]

\[
x_1' = -\frac{x_1}{\beta_1} (\cot \psi_{21} + \alpha_1) + \frac{x_2}{\sqrt{\beta_1 \beta_2}} \csc \psi_{21} \quad (4.11)
\]

\[ J = \frac{1}{2\beta_{x_1}} \left[ x_1^2 + \left( x_2 \sqrt{\beta_{x_1}/\beta_{x_2}} \csc \psi_{21} - x_1 \cot \psi_{21} \right)^2 \right]. \quad (4.12) \]

where \( \psi_{21} = \psi_2 - \psi_1 \) is the phase advance. Although in reality, \( x' \) is difficult to measure directly, MAD simulation can generate accurate \( x' \). Figure 4.17 compares the \( x' \) obtained from Eq. (4.11) with the \( x' \) from MAD tracking. Figure 4.18 and 4.19 show the comparison between the real action obtained directly from MAD tracking data and the deduced action from ICA (Eq. (4.12)). In Fig. 4.18, we can see that since we are using two \( x' \)'s at adjacent locations to deduce \( x' \), action calculated by ICA at sextupole location is different from that obtained by MAD tracking. But this
would not affect the difference between the actions before and after the sextupole kick which is really the quantity of interest to us.

![Graph showing comparison between x' of Eq. (4.11) and the real x' obtained from the MAD tracking data.](image)

**Figure 4.17:** Compare $x'$ of Eq. (4.11) with the real $x'$ directly obtained from the MAD tracking data. In this calculation, one sextupole with strength $K_2L = 1 \text{ m}^{-2}$ is placed in the lattice at 420.37 m (circumference of the ring is 807.1104 m) and the $\beta_{sz} = 15.217 \text{ m}$.

Figure 4.20 compares the action changes from the MAD tracking data with those from the analytic calculation, while Fig. 4.21 compares the change of actions from ICA with those from MAD tracking. Together, they demonstrate that by using this method, we can identify the action change $\Delta J$ very accurately.
Figure 4.18: Compare $J$ of Eq. (4.12) for about one tracking turn with the real $J$ directly obtained from the MAD tracking data.

In this calculation, one sextupole with strength $K_2L = 1 \text{ m}^{-2}$ is placed in the lattice at 420.37 m (circumference of the ring is 807.1104 m) and the $\beta_{sx} = 15.217 \text{ m}$. 
Figure 4.19: Compare $J$ of Eq. (4.12) for about four tracking turns with the real $J$ directly obtained from the MAD tracking data. In this calculation, one sextupole with strength $K_2L = 1 \text{ m}^{-2}$ is placed in the lattice at 420.37$\text{m}$ (circumference of the ring is 807.1104$\text{m}$) and the $\beta_{sx} = 15.217 \text{ m}$.
Figure 4.20: Compare $\Delta J$ of Eq. (4.10) with the real action change directly obtained from the tracking data of $x_{2\nu}(s,t)$. In this calculation, one sextupole with strength $K_2L = 0.1$ m$^{-2}$ is placed in the lattice and the $\beta_{sx} = 10.285$ m.
Figure 4.21: Compare $\Delta J$ from ICA with the real action change directly obtained from the MAD tracking data. In this calculation, one sextupole with strength $K_2L = 0.1 \text{ m}^{-2}$ is placed in the lattice and the $\beta_{sx} = 10.285 \text{ m}$. 
The sextupole strength can be calculated by using Eq. (4.10). Figure 4.22 compares the predicted sextupole strength vs the input sextupole strength. We find that the method can be practically used to measure the sextupole strength. We find similar agreements of sextupole strengths for the cases with 2 sextupoles, and also 12 sextupoles in an accelerator. The only requirement is that there are at least two BPMs between any two sextupoles, so that the action of \( x_{2\nu}(s,t) \) can be accurately obtained.

![Figure 4.22: Sextupole deduced from \( x_{2\nu} \) vs the “input” sextupole strength of the lattice. The agreement provides a useful method to deduce nonlinear elements in accelerators.](image)

Figure 4.22: Sextupole deduced from \( x_{2\nu} \) vs the “input” sextupole strength of the lattice. The agreement provides a useful method to deduce nonlinear elements in accelerators.
4.4.2 Error Analysis

The components of nonlinear motion, shown in Figure 4.7, are of the order of microns. Typical noise level in BPM turn-by-turn data is of the order of 100 $\mu$m. The nonlinear signals will be buried in the noise. ICA may be able to enhance the narrow-band signal provided that the signal is visible. Since the signal strength is proportional to the product of $x_0^2$ and $K_2L$, we explore the necessary conditions for the determination of $K_2L$.

Two modes are required to establish $x_{2\nu}(s,t)$. As mentioned earlier in Sec. 4.2.2, in order to obtain two significant $2\nu_x$ modes, a necessary condition is to have more than one sextupole in the lattice. Especially if both focusing and defocusing sextupoles are present in the lattice, the second $2\nu_x$ mode would be greatly enhanced. Normally, we always need two families of sextupoles for chromatic correction, and thus it is natural to have both focusing and defocusing sextupoles. With two families of sextupoles, both two singular value are proportional to sextupole strengths.

As we have explained earlier, BPM data are obtained from particle tracking code. Random noise is then added to each BPM signal at the level of 100 $\mu$m, or 200 $\mu$m in uniform distribution. Furthermore, sampling error plays an important role [3, 4], and we carry out calculations from 1000 turns to 5000 turns for every 20 turns. Because of the betatron tunes, the sampling error has a maximum and minimum in every 20 turns.

Instead of using an initial amplitude of $x_i = 1$ mm, we use the initial amplitude $x_i = x_0 = 5$mm or 10mm in order to excite the $2\nu_x$ modes. We define the merit function $\Delta x_{2\nu}$ as the r.m.s. error of the mode function $x_{2\nu}$, i.e.

$$\langle (\Delta x_{2\nu})^2 \rangle = \frac{1}{MN} \sum (x_{2\nu}(s_i, t_j) - x_{2\nu}^{\text{theory}})^2$$

(4.13)

where $M$ and $N$ are the number of BPMs and tracking turns respectively. Fig. 4.23 shows the histogram of the rms error $\Delta x_{2\nu}$ for many random seeds and different
number of tracking turns. Since \( x_{2\nu}(s, t) \) is proportional to \( x_0^2 \), we plot the histogram as a function of \( \Delta x_{2\nu}/x_0^2 \). Because of the sampling error and random BPM errors with many seeds, the histogram has no sample at \( \Delta x_{2\nu} = 0 \). The vertical line at \( \Delta x_{2\nu}/x_0^2 = 0.2 \) marks the boundary that the sextupoles can be determined to within \( \pm 5\% \) error. When the sextupole strength is reduced and when the noise level is increased, there are many more samples outside the boundary. When the noise level is reduced or the sextupole strength is increased, more samples fall within the boundary. This constraint can be derived by

\[
\Delta x_{2\nu} \leq \frac{1}{20}|x_{2\nu}| \sim \frac{1}{20}|x_{\text{noise}}| \tag{4.14}
\]

where the amplitude of the \( 2\nu_x \) mode must have at least the noise amplitude in order for the PCA-ICA to work.

We also study the effect of sampling error vs the tracking turns. Fig. 4.24 shows the ensemble average and r.m.s. width of \( \Delta x_{2\nu} \) vs the tracking turn number. At each tracking turn number, 10 random seeds are used to generate random errors in the BPMs. We have also carried out tracking with an initial amplitude of 5\( mm \). Our numerical sampling studies shows that Eq. (4.14) is satisfied when the nonlinear amplitude is larger than the noise amplitude, i.e. \( |x_{2\nu}| > |x_{\text{noise}}| \).

### 4.4.3 Determine sextupole strength using Less BPMs

In order to calculate the action variable, the method stated in Sec. 4.4.1 requires that there are at least two BPMs in between the sextupoles. But sometimes, many accelerators have only one sextupole in between two sextupoles. A new method is in need to overcome this problem. This method will use only the displacements and phases at BPMs but not the action variable. Suppose BPM1 and BPM2 are located in an accelerator segment where no nonlinear magnetic elements are present as shown
Figure 4.23: Histogram of $\Delta x_{2\nu}/x_0^2$ for various sextupole strengths and various level of BPM random errors.
Figure 4.24: The ensemble average and the r.m.s. width of $\Delta x_{2u}/x_0^2$ vs the tracking turn numbers. At each turn number, BPM errors with 10 random seeds are used to obtain the ensemble average and r.m.s. width.
in Fig. 4.25. And only one BPM, BPM3 follows the sextupole to be measured. The measured coordinates at three BPM locations are given by:

\[
\begin{align*}
    x_1 & = \sqrt{2\beta_{x1}} J_n \sin[\phi_{1,n} + 2\pi \nu_x (n - 1)] \\
    x_2 & = \sqrt{2\beta_{x2}} J_n \sin[\phi_{2,n} + 2\pi \nu_x (n - 1)] \\
    x_3 & = \sqrt{2\beta_{x3}} J_n \sin[\phi_{3,n} + 2\pi \nu_x (n - 1)] \\
    & + \sqrt{\beta_{xs}} \beta_{x3} \sin[\nu_x (\phi_{3,n} - \phi_{s,n})] \Delta x_s' 
\end{align*}
\]

\(\nu_x\) is the betatron tune, \(\beta_{xi}\) is the betatron function at the \(i\)th BPM location. \(J_n\) is the action variable at the \(n\)th revolution. \(J_n\) does not vary in any nonlinear-element-free section. \(\phi_{i,n}\) is the betatron phase at the \(i\)th BPM and \(n\)th revolution. Combining data from the first two BPMs in the nonlinear-element-free region, one can recover the action by

\[
J(n) = \frac{1}{4} \left[ \frac{1}{\cos^2 \left( \frac{\Delta \phi_{21}}{2} \right)} \left( \frac{x_1}{\sqrt{2\beta_{x1}}} + \frac{x_2}{\sqrt{2\beta_{x2}}} \right)^2 + \frac{1}{\sin^2 \left( \frac{\Delta \phi_{21}}{2} \right)} \left( \frac{x_1}{\sqrt{2\beta_{x1}}} - \frac{x_2}{\sqrt{2\beta_{x2}}} \right)^2 \right] 
\]

where \(\Delta \phi_{21} = \phi_{2,n} - \phi_{1,n}\) denotes the phase advance between the first two BPMs. The transverse coordinate at BPM3 location is composed of two parts: the linear betatron motion \(x_{30}\) and the motion introduced by and propagated from the sextupole. The linear motion at BPM3 location can be predicted very accurately by ICA.

\[
\frac{x_{30}}{\sqrt{2\beta_{x3}}} = \frac{1}{2 \cos \left( \frac{\Delta \phi_{21}}{2} \right)} \left( \frac{x_1}{\sqrt{2\beta_{x1}}} + \frac{x_2}{\sqrt{2\beta_{x2}}} \right) \cos(\Delta \phi_3) - \frac{1}{2 \sin \left( \frac{\Delta \phi_{21}}{2} \right)} \left( \frac{x_1}{\sqrt{2\beta_{x1}}} - \frac{x_2}{\sqrt{2\beta_{x2}}} \right) \sin(\Delta \phi_3) 
\]

where \(\Delta \phi_3 = \phi_{3,n} - \frac{\phi_{1,n} + \phi_{2,n}}{2}\). The nonlinear kick \(\Delta x'_s\) due to sextupole causes the beam an coordinate offset from the value predicted by linear optics.

\[
\Delta x_3 = x_3 - x_{30} = \beta_{xs} \beta_{x3} \sin \Delta \phi_{3s} \Delta x_s' 
\]

This offset is proportional to the sextupole kick angle which is given by:

\[
\Delta x'_s = \frac{1}{2} K_2 L x_s^2 
\]
where $K_2L$ is the integrated sextupole strength and $x_s$ is the transverse coordinate at the sextupole location which can be predicted by ICA.

$$x_s = \sqrt{2\beta_{xs}}J_n[\phi_{s,n} + 2\pi\nu_s(n - 1)] \quad (4.20)$$

Due to the proportionality of the offset $\Delta x_3$ to the integrated sextupole strength $K_2L$, one can deduce the $K_2L$ by:

$$K_2L = \frac{2\Delta x_3}{\sqrt{\beta_{xs}\beta_{x3}} \sin \Delta \phi_{s3} x_s^2} \quad (4.21)$$

where $\Delta \phi_{s3} = \phi_s - \phi_3$ is the phase advance from the target sextupole to BPM3.

Figure 4.25: Schematic drawing of one accelerator section. BPM1 and BPM2 are placed upstream in the nonlinear-element-free region. BPM3 follows the target sextupole.

According to Eq. (4.21), we can obtain the integrated sextupole strength $K_2L$ using $x_s$ predicted by linear motion every turn for every sextupole. For example, if only one sextupole is present in the lattice, 1000 turns of tracking would provide 1000 estimations of the sextupole strength. Figure 4.26 shows the relation between $2\Delta x_3/\sqrt{\beta_{xs}\beta_{x3}} \sin \Delta \phi_{s3}$ and $x_s^2$, where the slope gives the integrated sextupole strength. Depending on BPM noise level, data have spreads proportional to the noise level. The sextupole strength can then be determined by looking for the slope of the line connecting the centroids of every bin of $x_s^2$. Similarly, one can plot
4.4 Beam-based measurement of sextupole strength

\[ 2\Delta x_3/(\sqrt{\beta_{xs}\beta_{x3}} \sin \Delta \phi_{s3}) \] vs. \( x_s \) shown as quadratic curve in Fig. 4.27. For a pure sextupole, the curve is symmetric with respect to \( x_s \). In the presence of higher order multipoles, the curve will be asymmetric, and the data can be used to fit multipole strengths. Figure 4.28 shows the relation between \( 2\Delta x_3/(\sqrt{\beta_{xs}\beta_{x3}} \sin \Delta \phi_{s3}) \) and \( x_{oct} \), \( x_{oct}^3 \) when there is a pure octupole in the lattice.

Since Eq. (4.21) is very sensitive to the value of \( x_s \) at sextupole location, which can become small in betatron oscillation. The \( K_2L \) distribution will have long tail when \( |x_s| \) is small. The centroid of the distribution is the integrated sextupole strength.

Keeping the \( K_2L \) distribution within twice Full Width at Half Maximum (FWHM), we can calculate the centroid and the standard deviation of the \( K_2L \) distribution. (see Fig. 4.29) The top of Fig. 4.30 shows the "integrated sextupole strengths" (centroid) of 50 numerical experiments with different random seeds. The mean and r.m.s. error of these 50 centroids are plotted in the bottom plot in Fig. 4.30. The mean of the centroids can be deemed as the prediction value of the sextupole strength. The estimations of the standard deviation for these 50 experiments are displayed in Fig. 4.31. The bottom plot of Fig. 4.31 shows that the standard deviation of \( K_2L \) distribution is proportional to the BPM noise level.

In Fig. 4.32 to 4.34, twelve sextupoles are placed in the AGS lattice and their integrated strengths are estimated for different noise levels.

If multiple sextupoles or higher order multipoles are present in the lattice, one can use several downstream BPMs to solve for the strengths of individual sextupoles or multipoles. One example is the SPEAR3 (SLAC) lattice. It is a typical DBA configuration as shown in Fig. 4.35. There are six BPMs and four sextupoles in a cell. Only one BPM is located in between the sextupoles. The sextupoles of the same kind are identical and are powered in series. In this case, BPM 2 and 3 can be used to derive the linear betatron coordinates. The four downstream BPMs 4,5,6 and BPM 1 of the next cell can be used to deduce the strength of the sextupole kick.
Figure 4.26: According to Eq. (4.21), sextupole strength can be obtained by finding the slope of the centroid of each bin of $x_{sx}^2$. The width of each bin is proportional to its noise level.
Figure 4.27: According to Eq. (4.21), the sextupole strength can be obtained by using quadratic fit. Asymmetry of the curve indicates the existence of other multipoles.
Figure 4.28: Similar to Eq. (4.21), the relation between
\[ 2\Delta x_3 / (\sqrt{\beta_{oct}} \beta_3 \sin \Delta \phi_{oct3}) \] vs. \[ x_{oct} \] can be found. \[ x_{oct} \] is the coordinate at the octupole location which can be derived by ICA.
Figure 4.29: Distribution of integrated sextupole strength $K_2L$ for different noise levels. Although different noise levels result in different standard deviations, the centroids of the distribution are always around the real sextupole strength. The standard deviation of the distribution should be proportional to the noise levels.
Figure 4.30: Top: The predicted sextupole strength obtained by estimating the center of the 50 sextupole strengths obtained from 50 experiments.

Bottom: The "measured" integrated sextupole strength vs. BPM noise levels.
4.4 Beam-based measurement of sextupole strength

Figure 4.31: Top: The standard deviation (rms noise level) obtained by estimating the standard deviation of the distribution in Fig. 4.29 for 50 numerical experiments.
Bottom: The standard deviation of $K_2L$ distribution is found to be proportional to the BPM noise level.
Figure 4.32: Predicted sextupole strengths for 12 sextupoles in the lattice, when no noise is present.
Figure 4.33: Predicted sextupole strengths for 12 sextupoles in the lattice, with noise.
Figure 4.34: Predicted sextupole strengths for 12 sextupole in the lattice with noise.
Figure 4.35: Standard cell of SPEAR3
4.5 Summary

In this chapter, we have developed the ICA method for nonlinear betatron motion study. Our result shows that \( x_{2\nu}(s, t) = A_3(\phi)s_3(t) + A_4(\phi)s_4(t) \) of \( 2\nu_x \) modes from ICA calculation agree very well with the analytic solution of the Hill's equation for \( 2\nu_x \) mode. This agreement inspired us to propose a method to extract sextupole strength by using the relation between the action change and sextupole strength as demonstrated by Eq. 4.10. Fig. 4.22 shows that the resulting sextupole strength can be measured with high precision.

Since \( x_{2\nu}(s, t) \)'s at two BPMs are required to determine the action, our method requires two BPMs between each sextupole element to determine the change of nonlinear action before and after a sextupole. The same is true for the method proposed in Refs. [6, 7], where normalized phase space coordinates are used to extract nonlinear amplitudes.

Many light source storage rings have only one BPM between sextupoles or BPMs following a few sextupoles. We found a novel beam-based method for sextupole strength measurement. We carry out numerical simulations to demonstrate its applicability. Since ICA can determine the linear motion in high precision, both Eq. (4.10) and Eq. (4.21) can be used to determine the sextupole strength.

Our method can easily be extended to measure other high order multipoles. Since it can extract sextupole strength from turn-by-turn data, it is less prone to coherence in beam experiment.
In this thesis, the feasibility of applying independent component analysis (ICA) on both linear and nonlinear beam transverse motion diagnoses has been examined. ICA is an advanced multivariate statistical method for blind source separation. It can help us to uncover the hidden sources of a system or a physical process. It has been widely used in several branches of science like in neural science, telecommunication and economics etc.. However, to accelerator physics research, ICA is still a relatively new method. This dissertation is dedicated to explain how ICA actually works to separate out the linear betatron motion and explore the possibility of using ICA on the nonlinear motion analysis.

In Chapter 2, the fundamental ideas and algorithms of ICA are introduced. ICA methods can be grouped into two major categories: (1) ICA methods for systems without time structure, (2) ICA methods using time structures. For the first kind of ICA method. The common way to find the independent components is to bridge the “independence” of the variables with the “non-gaussianity”. For our purpose, the second kind of ICA is of interest. Data acquired by BPMs are time-related signals. We have to utilize the time sequence feature to extract useful information
about the transverse beam motion. This kind of ICA methods are built upon the
fact that independent components or the source signals should have diagonal auto-
covariance matrix. The simplest ICA method using time structure is the so call
AMUSE (Algorithm for Multiple Unknown Signals) algorithm, which can potentially
cause problems, for an unfortunate choice of time lag may lead to no unique solu-
tion. As a more robust method, SOBI (second order blind identification) algorithm
is introduced later. Instead of using only one time lag, it use several time lags to
eliminate the negative effect of an unfortunate choice. However, SOBI requires joint-
diagonalization of several auto-covariance matrices simultaneously. This is eventually
achieved by a “Jacobi-like” joint diagonalization algorithm.

After the discussion of the ICA algorithm, Chapter 3 showed details about how
ICA methods have been applied on linear betatron motion diagnosis. The PCA
preprocessing has been carefully examined. Results of the study show that for any
single motion, two eigenmodes are needed to fully represent it. PCA preprocessing can
preliminarily decouple different eigenmodes and reduce some of the noises. Then the
post-PCA steps further disentangle the modes to achieve better accuracy. Our results
indicate that ICA can extract out the $\beta$ function and phase of the linear betatron
motion very accurately. And under all circumstances, ICA method can achieve better
precision that PCA.

Since we have successfully applied ICA on linear betatron motion analysis, we then
further extend this method onto the nonlinear beam motion study in Chapter 4. This
is of fundamental importance to the storage ring performance. Because the dynamical
apertures of low emittance storage rings depend sensitively on sextupole families.
However, no effective method has been discovered to detect and analyze the nonlinear
elements. In this chapter, we developed ICA for nonlinear beam motion, in our
case the $x_{2\nu}$ motion excited by sextupole kicks, and finally proposed two methods of
beam-base measurement of the sextupole strengths. We found a very good agreement
between the $x_{2\nu}$ motion from ICA with our analytical solution of the Hill’s equation. And the study of the $x_{2\nu}$ motion reveals two important features of the $x_{2\nu}$ motion. First, sextupole kick can cause action change. Second, before and after any sextupole, the $x_{2\nu}$ motion would just look like simple betatron oscillation. Inspired by these two important features. We found two methods for beam-based measurement of the sextupoles strengths. One method can determine the integrated sextupole strength by calculating the action change before and after the sextupole kick. However, this method requires at least two BPMs in between any sextupoles. In some machine, like the SPEAR3 in SLAC, there may be only one BPM between some sextupoles. In order to overcome this problem, by using the second feature of the $x_{2\nu}$ motion we proposed another method to measure the sextupole strength. This method using two BPMs upstream and one BPM downstream of the target sextupole to get enough information calculate the kick strength. Our simulation results show great accuracy in identifying both the locations and the strengths of the sextupoles.
5. Conclusions
Principle component analysis (PCA) is a classical technique in statistical data analysis, feature extraction and data compression. Given a set of multivariate measurement, the goal of PCA is to find a smaller set of variables with less redundancy, that would provide as good a representation as possible. These variables are called the principle components (PCs). They may represent the important hidden factors underlying the system. Therefore the ultimate goal of PCA is to uncover these factors or PCs. The main difference between PCA and ICA is that ICA finds independent source signals, whereas PCA finds source signals which are merely uncorrelated. This subtle distinction has far reaching consequences for the power of ICA methods relative to PCA methods. However, PCA can serve as an important preprocessing step for ICA.
A. Principle component analysis (PCA)

A.1 Principle Component

Suppose the original data set consists of some measurements on $p$ variables. In order to reproduce the total system variability, eventually, $p$ principle components are attained from PCA. However, often much of this variability can be accounted for by a smaller number of $k$ of the principal components ($k \leq p$). This means that there is as much information in the $k$ PCs as there is in the original $p$ PCs. The $k$ principal components can then replace the initial $p$ PCs, thereby reduced the dimension of the original data set.

Algebraically, principal components $Y_i$’s are linear combinations of the $p$ measured variables $x_1, x_2, ..., x_p$.

\[
Y_1 = w_1^T x = w_{11}x_1 + w_{12}x_2 + ... + w_{1p}x_p \\
Y_2 = w_2^T x = w_{21}x_1 + w_{22}x_2 + ... + w_{2p}x_p \\
\vdots \\
Y_p = w_p^T x = w_{p1}x_1 + w_{p2}x_2 + ... + w_{pp}x_p
\]  

(A.1)

The $w_{i,j}$’s are the elements of the demixing matrix $W$. Although here we list $p$ of the PCs, there may be only a few essential PCs which are really needed to represent the whole system. Geometrically, the principal components represent the selection of a new coordinate system attained by rotating the original coordinate system with $x_1, x_2, ..., x_p$ as the axes. Using Eq.( A.1), we obtain

\[
Var(Y_i) = W_i^T \Sigma W_i \quad i = 1, 2, ..., p \\
Cov(Y_i, Y_j) = W_i^T \Sigma W_j \quad i, j = 1, 2, ..., p
\]  

(A.2)

where $W_i$ denotes the $i$th row of the demixing matrix $W$. $\Sigma$ is the covariance matrix
of \( x_1, x_2, ..., x_p \) with definition as follows:

\[
\Sigma = \text{Cov}(\mathbf{x}) = \mathbb{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]
\]

\[
= \mathbb{E} \left[ \begin{pmatrix}
    x_1 - \mu_1 \\
    x_2 - \mu_2 \\
    \vdots \\
    x_p - \mu_p
\end{pmatrix}
\begin{pmatrix}
    x_1 - \mu_1 \\
    x_2 - \mu_2 \\
    \vdots \\
    x_p - \mu_p
\end{pmatrix}^T \right]
\]

\[
= \begin{pmatrix}
    \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\
    \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}
\end{pmatrix}
\quad (A.3)
\]

\( \mu_i \) is the mean of the vector \( x_i \). \( \sigma_{ij} \) is the correlation of two vectors \( x_i \) and \( x_j \). The principal components are those uncorrelated linear combination \( Y_1, Y_2, ..., Y_p \) whose variance in Eq.(A.2) are as large as possible under the constraint that the coefficient vectors \( a_i \)'s are of unit length. The first principal component is the linear combination with maximum variance.

First principal component

\[
= W_1^T X \text{ that maximizes Var}(W_1^T X)
\]

subject to \( W_1^T W_1 = 1 \)

The \( i \)th principal component

\[
= W_i^T X \text{ that maximizes Var}(W_i^T X)
\]

subject to \( W_i^T W_i = 1 \)

and \( \text{Cov}(W_i^T x, W_k^T x) = 0 \) for \( k < i \)

Actually, PCA orders the extracted signals according to their variances (variances can be equated with power or amplitude), so that signals associated with high variance are deemed more important that those with low variance. As the principal components with the lowest variances are discarded, this usually ensures that the retained principal
components capture the main statistical structure of the original signal mixtures. The ordering of the extracted signals is often used to reduce the size of a given data. For example, with four mixtures, we were supposed to find four principal components. However, if the variance of the first two extracted signals are high, and the variance of the last two are close to zero. Then the last two extracted signals can be ignored. This effectively halves the size of the data set while retaining most of the variance associated with the original data set.

**A.2 Singular Value Decomposition**

A very general and useful form of PCA is singular value decomposition (SVD). Given a set of signal mixtures $x = (x_1, x_2, ..., x_N)^T$ in the form of an $M \times N$ matrix. SVD provides a decomposition of the form:

$$x = UDV^T \quad (A.4)$$

where $U = (U_1, U_2, ..., U_M)$ is an $M \times M$ matrix of $M$ spatial column eigenvectors while $V = (V_1, V_2, ..., V_N)$ an $N \times N$ array of $N$ temporal column eigenvectors. $D$ is an $M \times N$ diagonal matrix of $M$ (if $M < N$, vice versa) ordered singular values. Each singular value is equal to $\lambda^{1/2}$, where $\lambda$ is an eigenvalue of one eigenvector in $U$ and $V$. More formally, the columns in $U$ are the left singular vectors and the columns in $V$ are the right singular vectors of $x$. $U$ can be found by diagonalizing matrix $xx^T$ and $V$ can be solved by finding the eigenvectors of matrix $x^Tx$. The eigenvectors in the columns of $U$ and $V$ are orthogonal and are therefore uncorrelated.

Each eigenvalue specifies the amount of variance associated with the direction defined by a corresponding eigenvector in $U$ and $V$. The eigenvectors with very small eigenvalues can be discarded without losing any essential information contained by the original data set.
Bibliography


Wang, Ph.D. thesis, unpublished (Stanford University, 1999); Chun-·xi Wang,
Vadim Sajaev, and Chih-Yuan Yao, Phys. Rev. ST Accel. Beams 6, 104001

[3] X. Huang, S.Y. Lee, E. Prebys, R. Tomlin, PRST AB 8, 064001 (2005); X. Huang,

unpublished (Indiana University, 2008).


and Beams, 10, 034002 (2007).

and Beams, 11, 104002 (2008).


