SLAC-R-547
April 2003

MODEL INDEPENDENT ANALYSIS OF BEAM CENTROID DYNAMICS IN ACCELERATORS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Chun-xi Wang
September 1999

Work supported in part by Department of Energy Contract DE-AC03-76SF00515

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309, USA

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Dr. John Irwin
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Dr. Alexander W. Chao
(Principal Co-Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Dr. Helmut Wiedemann
(Applied Physics Research Co-Adviser)

Approved for the University Committee on Graduate Studies:

Abstract

Fundamental issues in Beam-Position-Monitor (BPM)-based beam dynamics observations are studied in this dissertation. The major topic is the Model-Independent Analysis (MIA) of beam centroid dynamics. Conventional beam dynamics analysis requires a certain machine model, which itself often needs to be refined by beam measurements. Instead of using any particular machine model, MIA relies on a statistical analysis of the vast amount of BPM data that often can be collected non-invasively during normal machine operation. There are two major parts in MIA. One is noise reduction and degrees-of-freedom analysis using a singular value decomposition of a BPM-data matrix, which constitutes a principal component analysis of BPM data. The other is a physical base decomposition of the BPM-data matrix based on the time structure of pulse-by-pulse beam and/or machine parameters. The combination of these two methods allows one to break the resolution limit set by individual BPMs and observe beam dynamics at more accurate levels. A physical base decomposition is particularly useful for understanding various beam dynamics issues. MIA improves observation and analysis of beam dynamics and thus leads to better understanding and control of beams in both linacs and rings. The statistical nature of MIA makes it potentially useful in other fields.

Another important topic discussed in this dissertation is the measurement of a nonlinear Poincaré section (one-turn) map in circular accelerators. The beam dynamics in a ring is intrinsically nonlinear. In fact, nonlinearities are a major factor that limits stability and influences the dynamics of halos. The Poincaré section map plays a basic role in characterizing and analyzing such a periodic nonlinear system. Although many kinds of nonlinear beam dynamics experiments have been conducted,

no direct measurement of a nonlinear map has been reported for a ring in normal operation mode. This dissertation analyzes various issues concerning map measurements and shows that it is possible to measure the Poincaré section map (in terms of Taylor series) of a circular accelerator to a surprisingly high order and accuracy based on present BPM technology. MIA can overcome the inherent limit of BPM resolution. Nonlinear map measurements will advance understanding of the beam dynamics of a ring.

Acknowledgements

Dissertation research and writing are a long process.[1] During the years I have received various kinds of help from my advisers, colleagues, and fellow students. It is hard to record all the details. Nonetheless, I would like to give it a try in order to show my thankfulness and appreciation.[2]

First of all, I would like to thank my adviser John Irwin. Without him this dissertation would simply not exist. I treasure the enjoyable experience of doing research together, especially during the time when we were trying to make some sense out of seemingly nonsense experimental data and finally established the Model-Independent Analysis (MIA) methods. From time to time, his enthusiasm and intuition sparked me to be creative; his persistence helped me to walk the extra mile when I thought I had got there (maybe because I was exhausted). I am surely grateful for his careful correction of my writing. In addition, I have the memory of his red face when he was frustrated with me for being tired of facing the hassle to obtain machine time, as well as the memory of the nice dinner at his home right after I passed my oral defense.

I would like to thank Professor A.W. Chao for his guidance during the early years of my graduate studies and his morale as well as technical support as a principal co-adviser over the years. I would also like to thank Professor H. Wiedemann for agreeing to be my department research co-adviser as required by Stanford University. As reading committee members, Professor Chao's and Wiedemann's critical reading

[1]Supported by Department of Energy contract DE-AC03-76SF00515 and Stanford University.

[2]Special thanks to those who generate and maintain the Stanford University latex thesis style files and templates (http://www-csli.Stanford.edu/Help/tex/suthesis), which makes thesis editing much easier. Let me keep the original acknowledgement in the template here:

"**I would like to thank my mother and the little green men from Mars.**"

further assures the quality of this dissertation.

I am grateful to all my five university exam committee members for agreeing to serve on a short notice. Special thanks are due to Professor Todd Smith for chairing the exam committee.

I could not postpone any longer thanking my friend Dr. William Hassenzahl. Ever since he invited me to come to Lawrence Berkeley National Laboratory more than seven years ago, he has offered me much personal help and support, including volunteering to read this dissertation and polish my writing (although the subject is outside his direct expertise). I am truly grateful to the kindness, friendship, and family-like care that he and his wife have given me. I would also like to thank him as well as Dr. B. Kincaid and Dr. K. Halbach for their recommendations to support my application to Stanford.

Thanks are also due to Dr. K.L.F. Bane who helped me to carry out the simulation studies of SLC and corrected Chapter 4 of this dissertation. I would like to thank Dr. Y. Cai and especially Dr. Y.T. Yan for helping me on numerous simulation studies and general enlightening discussions.

When conducting MIA experiments on SLC, I enjoyed assistance from many people including Dr. M. Minty, Dr. F.J. Decker, Dr. Y. Cai, Dr. L. Hendrickson, and my fellow student B. Podobedov. In particular, I would like to thank Dr. Minty for helping me to prepare some of the macros for data acquisitions and even carrying out some of the experiments with me. I am grateful to Dr. L. Hendrickson for her updating the SLC solo control program (SCP) to support my data acquisitions.

When investigating the one-turn map measurement issues at SSRL, I was indebted to J. Sebek, W.J. Corbett, and P. Tran for their cooperation. Especially for Jim's support, from whom I learned quite a few details about the BPM system he developed.

During these years, I enjoyed the company of my fellow students Dr. D.T. Palmer and Dr. Z.R. Huang as officemates. They enriched my experience as a student.

Although this is getting too long, I could not stop without mentioning my family's care and support. It is a great comfort knowing that they are always there whenever I need them and they always trust me pursuing something worthwhile.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Observation and comprehension of beam dynamics in an operating accelerator is crucial for improving machine performance. Beam Position Monitors (BPMs) are devices to measure the beam centroid along an accelerator beam line. As primary beam monitors, they provide information on beam dynamics in phase space. Usually theorists take such information for granted, while experimentalists struggle with errors due to BPM noise. In practice, one often encounters BPM resolution problems in sensitive measurements arising from nonlinearities or wakefield effects. As accelerators get larger and larger, beam control requirements become tighter and tighter, and more and more sensitive measurements become essential. To keep up with this trend and to accomplish difficult measurement/control tasks, the basic strategy is to improve individual BPM resolution. However, is there anything one can do beyond the hardware resolution limit? The easy answer is to average over time (turns in a ring or pulses in a linac), which is routinely used in beam observations when applicable. However, single pulse/turn resolution is often required. The next conceivable answer is to average over BPMs in some way. Imagine there are a large number of BPMs at the same location, averaging their readings will obviously improve the measurement accuracy. In reality, there are usually many BPMs distributed along a beam line. This dissertation shows how to obtain statistical benefits of using a large number of BPMs collectively. Remarkably, it is possible to improve the effective single-pulse individual BPM resolution by analyzing as an ensemble the readings of a large number

of BPMs for a large number of pulses.

Although many techniques have been developed to measure various beam or machine parameters, there is still a lack of systematic theoretical studies on some basic statistical issues of beam dynamics observations, especially the statistical error analysis in beam measurements. Furthermore, conventional beam dynamics measurement techniques rely on a concrete machine model, although conflicts between model and observation often leave machine physicists uncomfortable. Yet not much has been done to study beam dynamics without a machine model. This dissertation explores these gray areas and develops general data analysis methods to study beam dynamics independent of machine models. For this reason, the studies are titled Model-Independent Analysis (MIA). Such an adventure turns out to be rather fruitful and may have significant impact on beam dynamics studies in linear as well as circular accelerators. The BPM noise reduction mentioned above is a major MIA achievement. The main approach used in MIA is multivariate statistical data analysis, which is widely used in many other fields such as economics and medical research for statistical inference and so on. However, such statistical analysis has rarely been used in beam physics. So, to some extent, this dissertation adapts statistical analysis methods to beam dynamics analysis, even though it did not start that way. (In fact, we did reinvent wheels in our exploration.)

In terms of a methodology directed towards beam dynamics measurements, in addition to the statistical data analysis aspect, this dissertation focuses on the measurement of transformation matrices/maps, which are widely used in beam dynamics computations. In fact, a map approach is now a crucial aspect of beam optics designs[35]. Yet there is still a lack of direct measurements of those map coefficients, especially in rings. Usually each individual machine parameter is targeted by certain special measurement techniques. For example, instead of off-diagonal elements of a transformation matrix, linear coupling is measured by scanning horizontal vs. vertical tunes; instead of nonlinear map coefficients, chromaticity and tune-shift-with-amplitude are measured by observing tune dependency on beam energy and amplitude. In principle, one can measure transformation maps first and then extract machine parameters from the maps. In this dissertation, I discuss methods to directly

measure transformation maps, especially the nonlinear one-turn map in a ring (which contains all the global single-particle-dynamics information such as linear coupling, chromaticity, tune-shift-with-amplitude, and so on). I do not describe methods such as Lie algebraic normal-form analysis for extracting machine parameters from maps since they are well known.[70, 24, 34, 36, 35, 33]

The layout of this dissertation is as following. Chapter 2 reviews conventional BPM-based beam dynamics measurements with an emphasis on the error analysis. Chapter 3 describes various theoretical aspects of Model-Independent Analysis (MIA) techniques—the main subject of this dissertation. Chapter 4 discusses MIA applications for linacs by presenting simulation and experimental results from the Stanford Linear Collider (SLC), the first linear collider in the world. Chapter 5 detours for a moment to introduce another important subject of this dissertation: nonlinear one-turn map measurements in a ring. By evaluating various factors that may affect such measurements, I establish the possibility to measure a nonlinear one-turn map to rather high order with good accuracy. The major obstacle is BPM resolution. Chapter 6 discusses MIA applications to rings, especially how MIA can facilitate one-turn map measurements by alleviating the BPM resolution problem. Simulations for the brand new PEP-II B-factory machine are presented. Chapter 7 summarizes the major conclusions of this dissertation. Additionally, an introduction section or paragraph is given at the beginning of each chapter to further clarify the structure of this dissertation.

There are two major mathematical tools employed throughout this dissertation: least-squares regression and the singular value decomposition of matrices. Each of these well-established mathematical topics has volumes devoted to it. However, since these mathematical subjects may be unfamiliar to potential readers of this dissertation, they will be reviewed in the appendices so that this dissertation is self-contained. Appendix A covers least-squares methods and Appendix B covers the singular value decomposition. For general discussions on multivariate statistical analysis, please see the references [8, 9, 38, 68, 60, 48, 67].

Chapter 2

Review of BPM-based beam dynamics measurements and controls

I start with the basic problem of beam dynamics observations and discuss how to measure the phase-space motion of a beam, focusing on the issue of resolution and error analysis. First, the well-known linear dynamics model will be used in the discussion because linear motion is the dominant motion. Moreover, one should always choose the simplest section of a given lattice for phase-space measurements. Second, I generalize the discussion into general orbit-fitting problems. Third, I review the Response Matrix Method, which has been quite successful in accelerator model calibration. Linear transformation matrix measurements are reviewed, in which the principal components analysis has been used. In addition to these beam dynamics measurement and machine diagnostic methods, a few beam control and orbit correction methods are briefly reviewed, although this dissertation is focused on the dynamics measurements. This chapter is not critical for the understanding of the other chapters.

2.1 Measurement of beam dynamics in $1\frac{1}{2}$D phase space

Usually transverse beam dynamics is dominated by linear motions and the two transverse degrees of freedom are only weakly coupled. Thus, I start with the ideal $1\frac{1}{2}$D (one transverse degree $\{x, p_x\}$ plus energy variation δ, where p_x is the same as x' at linear order) linear beam dynamics, which is governed by:

$$\begin{cases} x &= \sqrt{2J\beta}\cos(\phi+\psi)+\eta\delta \\ p_x &= -\sqrt{\frac{2J}{\beta}}[\sin(\phi+\psi)+\alpha\cos(\phi+\psi)]+\eta'\delta \end{cases} \tag{2.1}$$

where the lattice functions β, α, η, and the phase advance ψ depend on location s but not on beam motion. (This discussion is mainly for rings, but can be adapted to linacs as well.) The action-angle variables J and ϕ as well as the relative energy deviation δ characterize a particular beam motion, independent of the location. To measure these three beam dynamics quantities, at least three BPMs are required in general. I will use $\tilde{}$ and $\bar{}$ over a position dependent quantity to indicate that it is at the 2nd and 3rd BPM locations respectively. Three BPM readings $(x, \tilde{x}, \bar{x})$ are used to determine the dynamical variables via

$$\begin{cases} x &= \sqrt{2J\beta}\cos\phi+\eta\delta\ , \\ \tilde{x} &= \sqrt{2J\tilde{\beta}}\cos(\phi+\tilde{\psi})+\tilde{\eta}\delta\ , \\ \bar{x} &= \sqrt{2J\bar{\beta}}\cos(\phi+\bar{\psi})+\bar{\eta}\delta\ . \end{cases} \tag{2.2}$$

where we have assumed $\psi = 0$ (i.e. $\tilde{\psi}$ and $\bar{\psi}$ are the phase advances from the first BPM to the 2nd and 3rd BPMs). The energy deviation is given by

$$\delta = \frac{\frac{\tilde{x}}{\sqrt{\tilde{\beta}}}\sin\bar{\psi}-\frac{\bar{x}}{\sqrt{\bar{\beta}}}\sin\tilde{\psi}-\frac{x}{\sqrt{\beta}}\sin(\bar{\psi}-\tilde{\psi})}{\frac{\tilde{\eta}}{\sqrt{\tilde{\beta}}}\sin\bar{\psi}-\frac{\bar{\eta}}{\sqrt{\bar{\beta}}}\sin\tilde{\psi}-\frac{\eta}{\sqrt{\beta}}\sin(\bar{\psi}-\tilde{\psi})}, \tag{2.3}$$

and the momentum is given by

$$p_x = \frac{\tilde{x} - \tilde{\eta}\delta - R_{11}(x - \eta\delta)}{R_{12}} + \eta'\delta, \tag{2.4}$$

where $R_{11} = \sqrt{\frac{\tilde{\beta}}{\beta}}(\cos\tilde{\psi} + \alpha\sin\tilde{\psi})$ and $R_{12} = \sqrt{\beta\tilde{\beta}}\sin\tilde{\psi}$.

Except for special cases where a denominator becomes zero, Eqs. (2.2)–(2.4) uniquely determines the dynamical quantities $\{x, p_x, \delta\}$ from three BPM readings $(x, \tilde{x}, \bar{x})$ for a given lattice. From these equations, we can clearly see many routinely-used rules ("tips"), which are listed here as a review although they are obvious.

- To measure energy, BPMs at non-zero dispersion locations are required. On the other hand, to measure betatron motion in a non-zero dispersion area, beam energy must be taken into account.

- BPMs with 180° (or multiple of) phase difference can NOT be used to measure beam momentum.

- BPMs at a phase difference of 180° are perfect for measuring the relative beam energy δ because the betatron motion can be cancelled and only two BPM readings are necessary. The expression of δ can be greatly simplified in such case. For example, if $\tilde{\psi} = \pi$,

$$\delta = \frac{\frac{\tilde{x}}{\sqrt{\tilde{\beta}}} + \frac{x}{\sqrt{\beta}}}{\frac{\tilde{\eta}}{\sqrt{\tilde{\beta}}} + \frac{\eta}{\sqrt{\beta}}}. \tag{2.5}$$

  The relative energy δ is simply proportional to the sum of the BPM readings if the two beta functions are also the same.

There are other rules of thumb for phase space measurement when taking the BPM resolution into account. I will point them out in the next section.

2.2 Resolution of phase space measurement

In this section, I consider measurement errors due to limited BPM resolutions. For simplicity, assume there is no dispersion at two BPM locations, with beta function values β and $\tilde{\beta}$ respectively, and phase advance ψ between them.[1] Then, the 1-D linear beam dynamics is given in action-angle variables by

$$\begin{cases} x &= \sqrt{2J\beta}\cos\phi \\ \tilde{x} &= \sqrt{2J\tilde{\beta}}\cos(\phi+\psi) \end{cases} \tag{2.6}$$

The deviations in coordinate variables and action-angle variables are related by

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}\tilde{x} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\beta}{2J}}\cos\phi & -\sqrt{2J\beta}\sin\phi \\ \sqrt{\frac{\tilde{\beta}}{2J}}\cos(\phi+\psi) & -\sqrt{2J\tilde{\beta}}\sin(\phi+\psi) \end{pmatrix} \begin{pmatrix} \mathrm{d}J \\ \mathrm{d}\phi \end{pmatrix} \tag{2.7}$$

and

$$\begin{pmatrix} \mathrm{d}J \\ \mathrm{d}\phi \end{pmatrix} = \frac{1}{\sin\psi} \begin{pmatrix} \sqrt{\frac{2J}{\beta}}\sin(\phi+\psi) & -\sqrt{\frac{2J}{\tilde{\beta}}}\sin\phi \\ \frac{1}{\sqrt{2J\beta}}\cos(\phi+\psi) & -\frac{1}{\sqrt{2J\tilde{\beta}}}\cos\phi \end{pmatrix} \begin{pmatrix} \mathrm{d}x \\ \mathrm{d}\tilde{x} \end{pmatrix} \tag{2.8}$$

Consider $\mathrm{d}x$ and $\mathrm{d}\tilde{x}$ as BPM measurement errors and assume they follow a Gaussian distribution. Since the noise from the two BPMs are independent, according to the general statistical theory we know that the errors $\{\mathrm{d}J, \mathrm{d}\phi\}$ are also Gaussian distributed. In the case that the two BPM noises are uncorrelated and have the same resolution (width of the distribution) Δx, we can calculate the errors in action-angle

[1] To simplify the notation, here I use ψ instead $\tilde{\psi}$ for the phase advance although $\tilde{\psi}$ is used in the last section.

variables as:

$$\begin{aligned}
\Delta J &= \frac{\sqrt{2J}}{|\sin\psi|\sqrt{\beta\tilde{\beta}}}\sqrt{\tilde{\beta}\sin^2(\phi+\psi)+\beta\sin^2\phi}\;\Delta x \\
&= \frac{\sqrt{J}}{|\sin\psi|\sqrt{\beta\tilde{\beta}}}\sqrt{\beta+\tilde{\beta}-\sqrt{(\tilde{\beta}\cos 2\psi+\beta)^2+(\tilde{\beta}\sin 2\psi)^2}\cos(2\phi+\xi)}\;\Delta x \\
&\leq \frac{\sqrt{J}}{|\sin\psi|}\sqrt{\frac{1}{\beta}+\frac{1}{\tilde{\beta}}+\sqrt{\frac{1}{\beta^2}+\frac{1}{\tilde{\beta}^2}+\frac{2}{\beta\tilde{\beta}}\cos 2\psi}}\;\Delta x \qquad (2.9)
\end{aligned}$$

and

$$\begin{aligned}
\Delta\phi &= \frac{1}{2\sqrt{J}}\frac{1}{|\sin\psi|\sqrt{\beta\tilde{\beta}}}\sqrt{\beta+\tilde{\beta}+\sqrt{(\tilde{\beta}\cos 2\psi+\beta)^2+(\tilde{\beta}\sin 2\psi)^2}\cos(2\phi+\xi)}\;\Delta x \\
&\leq \frac{1}{2\sqrt{J}}\frac{1}{|\sin\psi|}\sqrt{\frac{1}{\beta}+\frac{1}{\tilde{\beta}}+\sqrt{\frac{1}{\beta^2}+\frac{1}{\tilde{\beta}^2}+\frac{2}{\beta\tilde{\beta}}\cos 2\psi}}\;\Delta x \qquad (2.10)
\end{aligned}$$

where $\xi = \arctan\frac{\tilde{\beta}\sin 2\psi}{\tilde{\beta}\cos 2\psi+\beta}$. Since the angle variable ϕ can have any value, we should take the upper limits of the above two equations as resolution estimates. For the relative error in the amplitude, we have $\frac{\Delta\sqrt{J}}{\sqrt{J}} = \frac{1}{2}\frac{\Delta J}{J}$, which has the same limit as $\Delta\phi$ does. Therefore, the resolution in action-angle variables are

$$\left\{\frac{\Delta\sqrt{J}}{\sqrt{J}}, \Delta\phi\right\}_{\max} = F(\beta,\tilde{\beta},\psi)\frac{\Delta x}{k\sqrt{\epsilon}} = \hat{F}(\frac{\beta}{\tilde{\beta}},\psi)\frac{\Delta x}{k\,\sigma_x} \qquad (2.11)$$

where ϵ is the beam emittance and the amplitude factor $k \equiv \sqrt{\frac{2J}{\epsilon}}$ measures the amplitude of the betatron oscillation in units of beam size,

$$F(\beta,\tilde{\beta},\psi) \equiv \frac{1}{\sqrt{2}|\sin\psi|}\sqrt{\frac{1}{\beta}+\frac{1}{\tilde{\beta}}+\sqrt{\frac{1}{\beta^2}+\frac{1}{\tilde{\beta}^2}+\frac{2}{\beta\tilde{\beta}}\cos 2\psi}}, \qquad (2.12)$$

and

$$\hat{F}\left(\frac{\beta}{\tilde{\beta}},\psi\right) \equiv \frac{1}{\sqrt{2}|\sin\psi|}\sqrt{1+\frac{\beta}{\tilde{\beta}}+\sqrt{1+\left(\frac{\beta}{\tilde{\beta}}\right)^2+2\frac{\beta}{\tilde{\beta}}\cos 2\psi}}. \tag{2.13}$$

F and $\hat{F}$ are dimensionless factors that reflect the dependency of phase-space resolution on the layout of BPMs. One application of this error analysis is to determine which pair of BPMs can yield more accurate measurement of phase space motion. Since $\frac{\Delta x}{k\sqrt{\epsilon}}$ is location independent, the function $F(\beta,\tilde{\beta},\psi)$ is a figure of merit to evaluate how good a pair of BPMs are for phase space measurements.

To minimize F, we differentiate $F(\beta,\tilde{\beta},\psi)$ over ψ and get:

$$\frac{\partial F}{\partial\psi} = \frac{-1}{\sqrt{2}}\cos\psi\left[\frac{1}{\sin^2\psi}\sqrt{\cdots}+\frac{2}{\beta\tilde{\beta}\sqrt{\cdots}\sqrt{\cdots}}\right] = 0 \tag{2.14}$$

Since $[\cdots] > 0$, $\cos\psi = 0$ is the condition for optimal measurement. In other words, **for a given BPM resolution, the condition for the best phase space measurement is to have 90° phase advance between the two BPMs**, which is a physically intuitive requirement. (Although 270° or larger phase advance could satisfy Eq.(2.14), it is better to use a smaller value because a larger phase advance means more magnets between the two BPMs, and therefore more uncertainty involved in the measurement.) The optimal value of the figure of merit function F is given by

$$F_{\rm opt} = F(\beta,\tilde{\beta},\frac{\pi}{2}) = \frac{1}{\sqrt{2}}\sqrt{\frac{1}{\beta}+\frac{1}{\tilde{\beta}}+|\frac{1}{\beta}-\frac{1}{\tilde{\beta}}|} = \frac{1}{\sqrt{\min(\beta,\tilde{\beta})}} \tag{2.15}$$

This optimal value confirms another intuition: for a given BPM resolution, the larger the beta function at the measurement location, the better the phase space resolution. Moreover, it is the smaller beta function between the two BPM locations that counts. From the F function, we can draw one more observation that a phase advance between $60° \le \psi \le 120°$ is about as good as 90°. However, a value less than 30° will reduce the phase space resolution by at least a factor of 2.

In summary, a general guideline to choose BPM locations in a ring for phase space measurement is: try to have a phase advance $90° \pm 30°$ and

have $\min(\beta, \tilde{\beta})$ **as large as possible**.

Although the function $\hat{F}(\frac{\beta}{\tilde{\beta}}, \psi)$ is not as meaningful as $F(\beta, \tilde{\beta}, \psi)$ since the beam size σ_x depends on location, in some cases, $\hat{F}$ may be easier to use because it has one less variable. Under the above optimal condition $\hat{F}_{\rm opt} = 1$.

Similarly, we examine the phase space resolution in coordinate variables. The position resolution is given directly by the BPM resolution. To get the momentum resolution, we recall that

$$p = -\frac{R_{11}}{R_{12}}x + \frac{1}{R_{12}}\tilde{x} \tag{2.16}$$

where the transfer matrix elements $R_{11} = \sqrt{\frac{\tilde{\beta}}{\beta}}(\cos\psi + \alpha\sin\psi) = \sqrt{\tilde{\beta}\gamma}\cos(\psi - \arctan\alpha)$ and $R_{12} = \sqrt{\tilde{\beta}\beta}\sin\psi$. Thus, for a given BPM resolution Δx,

$$\begin{aligned} \Delta p &= \sqrt{\left(\frac{R_{11}}{R_{12}}\right)^2 + \left(\frac{1}{R_{12}}\right)^2}\,\Delta x \\ &= \frac{\sqrt{\gamma}}{|\sin\psi|}\sqrt{\frac{1}{\beta}\cos^2(\psi - \arctan\alpha) + \frac{1}{\tilde{\beta}\beta\gamma}}\,\Delta x \end{aligned} \tag{2.17}$$

Therefore, the relative errors in coordinate variables are given by

$$\left\{ \begin{array}{rcl} \frac{\Delta x}{x_{\max}} & = & \frac{\Delta x}{\sqrt{2J\beta}} = \frac{1}{\sqrt{\beta}}\frac{\Delta x}{k\sqrt{\epsilon}} \\ \frac{\Delta p}{p_{\max}} & = & \frac{\Delta p}{\sqrt{2J\gamma}} = \frac{1}{|\sin\psi|}\sqrt{\frac{1}{\beta}\cos^2(\psi - \arctan\alpha) + \frac{1}{\tilde{\beta}\beta\gamma}}\,\frac{\Delta x}{k\sqrt{\epsilon}} \end{array} \right.$$

Note that $\frac{\Delta p}{p_{\max}}$ is different from $\frac{\Delta J}{J_{\max}}$.

2.3 General orbit fitting problems

In the previous two sections I presented the explicit formulae that can be used to measure the phase space dynamics in a simple $1\frac{1}{2}$ degrees-of-freedom system, and analyzed measurement errors due to BPM noise. In this section I will give a general formula that can accommodate any number of degrees of freedom and use any number of BPMs. It covers general orbit fitting problems.

Let $\vec{b} = \{b^1, b^2, \cdots, b^M\}$ be the readings of M consecutive BPMs on the same beam pulse, then in general we have

$$\vec{b} = \vec{b}_0 + \vec{R}_{11}x_0 + \vec{R}_{12}x'_0 + \vec{R}_{13}y_0 + \vec{R}_{14}y'_0 + \vec{R}_{16}\delta + \cdots + \vec{n} \tag{2.18}$$

where $\vec{b}_0$ contains all the BPMs' offsets; $\vec{R}_{ij}$'s are the coefficients of linear transfer matrices from a certain location s_0 to the BPMs' locations s_1, s_2, $\cdots$, s_M; x_0, x'_0, y_0, y'_0, and δ are the initial beam conditions at s_0; and $\vec{n}$ contains all BPM noise. In addition to the linear terms one can have higher order terms as well as terms due to other physical effects in the expansion. We can write Eq.(2.18) in matrix form

$$\Delta\vec{b}^T \equiv \left(\vec{b} - \vec{b}_0\right)^T = F\,\vec{q}^{\,T} + \vec{n}^{\,T} \tag{2.19}$$

where superscript ^{T} indicates matrix transpose,

$$\vec{q} \equiv [\, x_0, x'_0, y_0, y'_0, \delta, \cdots \,],$$

and the matrix

$$F \equiv \left[\vec{R}_{11}^T\ \vec{R}_{12}^T\ \vec{R}_{13}^T\ \cdots\right] \equiv \begin{bmatrix} R_{11}(s_0 \to s_1) & R_{12}(s_0 \to s_1) & R_{13}(s_0 \to s_1) & \cdots \\ R_{11}(s_0 \to s_2) & R_{12}(s_0 \to s_2) & R_{13}(s_0 \to s_2) & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ R_{11}(s_0 \to s_M) & R_{12}(s_0 \to s_M) & R_{13}(s_0 \to s_M) & \cdots \end{bmatrix}$$

Eq.(2.19) links a difference orbit measurement given by $\Delta\vec{b}$ to the machine model represented by F, and is the basis for orbit fitting and phase space dynamics measurement. In the usual orbit fitting exercises, the goal is to find the orbit that satisfies a given model and best fits the measured orbit, so that one can check how much the machine deviates from its model. In phase-space dynamics measurements, the goal is to determine the beam parameters $\{x_0, x'_0, y_0, y'_0, \delta, \cdots\}$ from a measured orbit. Due to the existence of noise, one often would like to measure more beam positions than the number of parameters and make Eq.(2.19) over-determined.

The least-squares solution to Eq.(2.19) is given by

$$[\, x_0, x_0', y_0, y_0', \delta, \cdots \,]^T = \vec{q}^{\,T} = (F^T F)^{-1} F^T \Delta\vec{b}^{\,T} \tag{2.20}$$

and the fitted orbit is $F(F^T F)^{-1} F^T \Delta\vec{b}^{\,T}$, where we have assumed that BPMs have the same resolution σ_n and their noise are independent (i.e. $\left\langle \vec{n}^T \vec{n} \right\rangle = \sigma_n^2 \, I$), otherwise a more complex formula is necessary for a least-squares solution[40] (see Appendix A.4 also). Note that Eq.(2.20) covers the full rank case as well. The variance-covariance matrix of errors in $\vec{q}$ can be computed as

$$\begin{aligned} \left\langle \left[\vec{n}F(F^T F)^{-1}\right]^T \vec{n}F(F^T F)^{-1} \right\rangle_{\text{noise}} &= (F^T F)^{-1} F^T \left\langle \vec{n}^T \vec{n} \right\rangle F (F^T F)^{-1} \\ &= \sigma_n^2 (F^T F)^{-1}. \end{aligned} \tag{2.21}$$

Square roots of the diagonal terms give rms errors of the measurement due to noise, which can be easily calculated from the given F. Two general conclusions can be drawn from Eq.(2.21) without knowing the details of F: 1) $F^T F$ should be well conditioned in order to obtain accurate measurements; 2) Since $F^T F$ generally increases with M (i.e. $F^T F/M$ tends to be a constant, at least for a periodic lattice as in a beam line), the measurement error goes down as $\frac{1}{\sqrt{M}}$. Therefore one can reduce the BPM random noise effects on the measurement by using more and more BPMs. However, in addition to the random noise, the accuracy of the machine model specified in F is crucial for an accurate measurement.

Now let us consider a familiar example. For an ideal 1-D linear lattice, the beam position at the m-th BPM x_m is given by the R-matrices as $x_m = R_{11}(m)x_0 + R_{12}(m)x_0'$. Suppose we measure a pulse at M BPMs and know the transformation matrix R's from some model, and wish to fit the orbit to the model and find the

initial x_0, x_0' of the orbit. In this case, Eq.(2.19) becomes

$$[x_1, x_2, \cdots, x_M] = [x_0, x_0'] \begin{bmatrix} R_{11}(1) & R_{12}(1) \\ R_{11}(2) & R_{12}(2) \\ \vdots & \vdots \\ R_{11}(M) & R_{12}(M) \end{bmatrix}^T \tag{2.22}$$

and the transpose of Eq.(2.20) yields

$$\begin{bmatrix} x_0 \\ x_0' \end{bmatrix} = \begin{bmatrix} \vec{R}_{11}\cdot\vec{R}_{11} & \vec{R}_{11}\cdot\vec{R}_{12} \\ \vec{R}_{12}\cdot\vec{R}_{11} & \vec{R}_{12}\cdot\vec{R}_{12} \end{bmatrix}^{-1} \begin{bmatrix} \vec{R}_{11}\cdot\vec{x} \\ \vec{R}_{12}\cdot\vec{x} \end{bmatrix} \tag{2.23}$$

where $\vec{x} \equiv [x_1, x_2, \cdots, x_M]^T$ consists of the orbit, $\vec{R}_{11}$ consists of the R_{11}'s, and so on. The 2×2 matrix inversion can be done analytically and leads to the expressions seen in the literature[12]. It is obvious how to include more degrees of freedom via Eq.(2.20) or by extending Eq.(2.23) directly.

Section 2.2 showed how to optimize the errors given by Eq.(2.21) for the 1-D two BPM situation. In general, such optimization could be rather involved. However, for a given setup, Eqs. (2.20) and (2.21) provide the least-squares solution and error estimates of the general orbit fitting problems.

2.4 Accelerator model calibration via Response Matrix Method

The Response Matrix Method is a systematic method to calibrate accelerator models and has been successfully used on many machines[63, 28, 29, 61]. The basic idea is to minimize the difference, between measured and calculated (from a model), BPM responses to changes in steering magnet strengths by adjusting various model parameters such as corrector strengths, quadrupole gradients, and BPM gains. This idea can be formulated as follows. For any given magnetic steering θ_x and θ_y, responses at BPMs can be measured as well as calculated according to one's model to generate

the response matrices R^{meas} and R^{mod} so that

$$\begin{pmatrix} x \\ y \end{pmatrix} = R^{\text{meas or mod}} \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} \tag{2.24}$$

where x and y are the beam orbit response. Such a response matrix can be extended to cover all M BPMs and N steering magnets, so that there are $M \times N$ elements. The figure to be minimized can be specified by

$$\chi^2 = \sum_{i,j} \frac{\left(R_{ij}^{\text{meas}} - R_{ij}^{\text{mod}}(\xi_1, \xi_2, \cdots)\right)^2}{\sigma_i^2} \tag{2.25}$$

where i goes through all BPMs of resolution σ_i and j goes through all steering magnets. The ξ's are model parameters to be adjusted for the minimization.

In addition, BPM gains and corrector scaling factors can be introduced to correct the measured response matrix, i.e. use $BR^{\text{meas}}C$ to replace the R^{meas} in the above expression, where the diagonal matrix B contains BPM gains and C contains corrector scaling factors. Further correction factors can be added when necessary. For example, the following BPM correction scheme is used in LOCO [63] for each BPM,

$$\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \frac{1}{\sqrt{1-c^2}} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \begin{pmatrix} g_x\, x \\ g_y\, y \end{pmatrix} \tag{2.26}$$

where g is the gain, θ is the roll angle, c is a factor due to BPM pick-up electrodes layout. All such factors can be varied for the minimization.

The implementation details of the response matrix method could be quite different depending on how sophisticated one's model is. Typically there are thousands of response matrix elements and hundreds of model parameters used for the optimization. In general, the beam responses could be nonlinear and minimization over such a large parameter space could become difficult and time consuming if not senseless.

One way to solve the minimization problem is to linearize the response as

$$R_{ij}^{\mathrm{mod}} = R_{ij}^{0} + \sum_{\{\xi\}} \frac{\partial R_{ij}}{\partial \xi} \delta\xi \tag{2.27}$$

and obtain the solution using the least-squares fitting,

$$\vec{\delta\xi} = (A^T A)^{-1} A^T \left[R_{ij}^{\mathrm{meas}} - R_{ij}^{0} \right]_{M\times N,1} \tag{2.28}$$

where matrix $A = \left[\frac{\partial R_{ij}}{\partial \xi}\right]$ consists of the $M \times N$ derivatives over each parameter in a column, and the number of columns depends on the number of parameters in use. Usually, A is quite a large matrix. Here I do not include the factors such as different BPM resolutions and gains. Please see the references for such implementation details and Appendix A for more general least-squares fitting methods. Often such procedures need to be iterated due to the presence of nonlinear responses. The response matrix method has been proven successful to calibrate the linear models of many light source storage rings.

Usually, matrix A is degenerate and the above formal solution can not be used. A standard treatment is to compute the pseudo-inverse of A with the help of singular value decomposition. By removing the small singular values, the solution's sensitivity to noise is reduced, thus more reliable estimates of parameter adjustment $\delta\xi$ can be obtained. This is one of the major application of SVD in beam measurement and control. See Appendix A.3 for such treatments and general fitting error estimates.

Note that the design matrix A is generated from one's model, therefore a good model is crucial. If a sufficiently good model is available, response matrix methods provide a systematic and powerful tool to calibrate model parameters. However, response matrix methods do not work well, in case unanticipated physics affects the beam response and is not included in the model. Even worse, it will not give clues for the unknown physics. This dissertation provides novel methods to cover this gap.

2.5 R matrix measurements

Linear beam dynamics can be characterized by the R matrix that transfers beam phase-space variables from one location to another. In rings, lattice measurements are focused on phase advances and beta functions because phase advances are relatively easy to measure and the beta functions reflect the beam envelope. However R matrix measurements can provide equivalent information and could be simple also. In linacs the "lattice functions" are not uniquely defined by the lattice itself as they are in rings. Thus lattice measurements rely more on the R matrix measurement. Here I review two major techniques for R matrix measurements used for the SLC.

The first technique[17, 18] is similar to the Response Matrix method. The scheme and symbols are illustrated in Fig. 2.1. C^q and B^p are the linear transfer matrices to the q-th corrector and p-th BPM respectively. They are assumed known, from an accurate model for example. R is the matrix to be determined. M_{ij}^{pq} is the measured p-th BPM response (gradient) to the q-th corrector change, where the subscripts $i = 1$ or 3 for x or y BPM reading and $j = 2$ or 4 for x or y corrector. It is easy to see that

$$M_{ij}^{pq} = \left(B^p R \left(C^q\right)^{-1}\right)_{ij}, \tag{2.29}$$

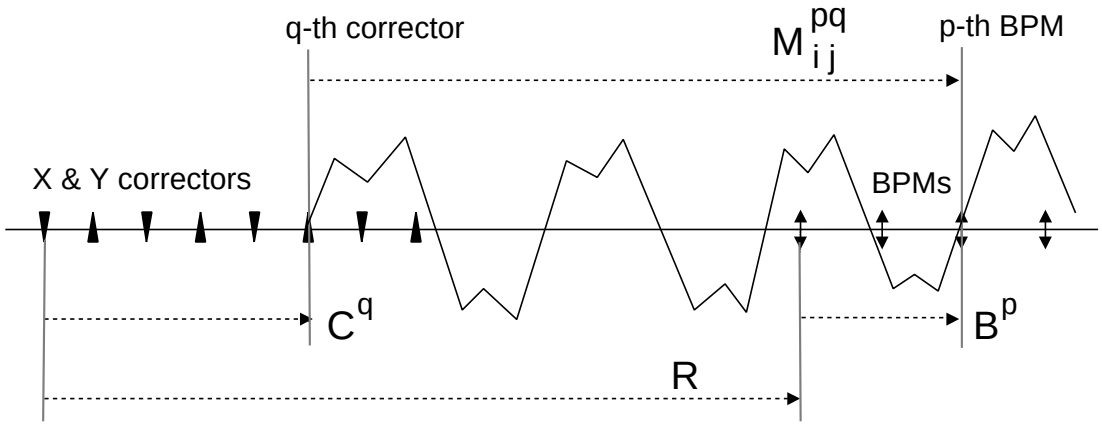

Figure 2.1: R matrix reconstruction method used by T. Barklow for SLC

from which the fitting problem can be set up to minimize

$$\chi^2 = \sum_{\{pqij\}} \left(\frac{M_{ij}^{pq} - \left(B^p R \left(C^q\right)^{-1}\right)_{ij}}{\delta M_{ij}^{pq}} \right)^2 \tag{2.30}$$

where δM_{ij}^{pq} are the raw measurement errors. $\{p, q, i, j\}$ are chosen to generate sufficient number of equations and even make the fitting overdetermined (see the reference for details). In addition, a symplectic constraint may be imposed on R, which reduces the fitting parameter space but makes the fitting problem nonlinear.

In addition to the accuracy of the C and B matrices, this technique relies on the measurement accuracy of the response matrix M_{ij}^{pq}. To achieve a good measurement, for each corrector, its strength θ_q is stepped over a number of increments, then fit them with the corresponding BPM responses according to $X^p = M_{ij}^{pq}\theta_q + \eta^p \frac{\Delta E}{E} + \text{const.}$ where the response M, dispersion η, and the constant are the parameters to be fit. A χ^2 test is performed to make sure that this response model is good. Furthermore, a global error term σ_{sys} was introduced quadratically in the denominator of Eq.(2.30) to account for the random noise and unknown systematics.

The second technique[54] is not as well documented but has a few good points. In some respects, this work came close to some techniques developed in this dissertation. The basic idea is to use measured beam trajectory parameters to reconstruct the lattice parameters by correlating each BPM reading with a full set of trajectory parameters. It correctly pointed out the significant effect of errors in the trajectory parameter measurements, which could be a pitfall for a naive least-squares fitting solution, especially when the signal-to-noise ratio is poor. The issue can be illustrated easily by a simple example used in the reference: find the parameter p of $y = p\,x + q$ from a given set of measured points $\{x_i, y_i | i = 1, \cdots, P\}$. The usual least-squares solution $\hat{p} = \sum_i x_i y_i / \sum_i x_i^2$ is inaccurate when there are errors in x's. Let R_s be the signal-to-noise ratio in x, then

$$\frac{\hat{p}}{p} = \frac{1}{1 + \frac{1}{R_s^2}}\,, \qquad (P \to \infty) \tag{2.31}$$

Therefore the least-squares solution $\hat{p}$ could yield significant errors even for modest signal-to-noise ratio.

To avoid such a pitfall, least-squares fitting was abandoned (although this may not be necessary, see Section 3.7) and a technique based on a principal components method was developed. The idea can be illustrated as follows. Consider the readings of a BPM at location s that can be expressed as

$$x(s) = R_{11}(s_0 \to s)x(s_0) + R_{12}(s_0 \to s)x'(s_0) + R_{16}(s_0 \to s)\delta_E. \tag{2.32}$$

From a measured data set consisting of $\{x(s), x(s_0), x'(s_0), \delta_E\}$ for a large number of pulses, the lattice parameters $R_{11}(s_0 \to s)$, $R_{12}(s_0 \to s)$, and $R_{16}(s_0 \to s)$ can be determined. However, due to the measurement noise, the usual least-squares fitting may be inappropriate. Suppose the beam parameters are measured by three other BPMs, say $x(1)$, $x(2)$, and $x(3)$, then the above equation can be written as

$$a\,x(s) + b\,x(1) + c\,x(2) + d\,x(3) = 0 \tag{2.33}$$

where a, b, c, and d are determined by the lattice parameters, and the beam trajectories are measured by the BPM reading x's. Due to BPM noise, the measured data points will scatter around the 3-D hyperplane (given by the above equation) in a 4-D space (defined by $\{x(s), x(1), x(2), x(3)\}$). The desired parameters $\{a, b, c, d\}$ form a vector perpendicular to this hyperplane. To find this vector direction, a principal axes technique is used to determine the axis along which the data points fluctuate the least.

The solution of this "least fluctuation direction" is given by the eigenvector corresponding to the least eigenvalue of covariance matrix $C_{ij} = \mathrm{cov}(x(i), x(j))$. (Section 3.3 will describe the principal components method in details.) This method avoids the mentioned least-squares fitting problem and thus yields much improved lattice parameter estimates. However, there is a severe limitation due to the use of the smallest eigenvalue/eigenvector. Suppose more BPMs are used for measuring the trajectory parameters in order to improve accuracy, then the system becomes overdetermined and yields many near-degenerate small eigenvalues, but none of them are suitable for determining the lattice parameters.

2.6 Orbit corrections via SVD and MICADO methods

Although this dissertation is focused on the beam dynamics measurement, I will review a couple of major orbit correction methods, especially the so called "corrector ironing" technique[77, 20, 32, 37] based on singular value decomposition (SVD). The purpose of reviewing this topic is to clarify the ways that SVD has been used in accelerator physics, so that one will not confuse the existing techniques with the methods developed in this dissertation.

The main idea of SVD-based "corrector ironing" technique can be summarized as follows.[56] Let $\Delta b = \{\Delta b_1, \cdots, \Delta b_M\}^T$ be the desired orbit correction at M BPMs, $\theta = \{\theta_1, \cdots, \theta_N\}^T$ be the strengths of a set of N correctors, and $R_{M\times N}$ be the BPM response matrix to the corrector adjustments. The required corrector strength changes should satisfy

$$\Delta b = R\,\Delta\theta. \tag{2.34}$$

This problem may have only approximate solutions that can be solved in many ways including the least-squares methods via minimization of $\|\Delta b - R\Delta\theta\|_2$. The problem is that some of the required corrector changes are often very large and fight each other, especially when the correction is made via trial-and-error approach.

There is a simple cure for this problem: solving the above least-squares via SVD of the matrix R—a common practice for mature least-squares users. Appendix A.3 contains a brief review of this subject. There are two key ingredients in this technique: 1) the solution $\Delta\hat{\theta}$ not only minimizes $\|\Delta b - R\Delta\theta\|_2$ but also has a minimal 2-norm of $\Delta\theta$ (i.e. minimal rms of corrector strength changes); 2) the near-degenerate degrees of freedom can be removed and thus corrector-fighting can be avoid (without sacrificing the correction accuracy if those correctors are redundant). For these reasons, the SVD-based orbit correction technique has been routinely used.

Another well-known orbit correction technique is referred to as the MICADO[2] method[13], which has been implemented in programs such as MAD[47, 1] and SAD[42, 2]. The basic idea is as follows: given the orbit offset to be corrected, choose the most

[2]The acronym is due to the name of the subroutine that performs the computation. It stands for "MInimisation des CArrés des Distortions d'Orbits"

effective (a given number n or using as many as necessary until the residual is sufficiently small) correctors amongst all available correctors. Note that this is not the same as using the first n singular values of the above SVD-based method. Even if only one singular value is kept, the correction may still involve all correctors. The MICADO method is accomplished via Householder QR factorization with column pivoting as described in Appendix B.3 and B.4. However, the pivoting criterion used in MACADO is different from the one described in the appendix and thus is explained here.

The goal is to minimize $\|\Delta b - R\,\Delta\theta\|_2$ by choosing appropriate corrector adjustment $\Delta\theta$'s. Since orthogonal transformations do not change the 2-norm, permutation transformation and Householder transformation are used to solve the problem. The permutation is to reorder the correctors according to their effectiveness, while the Householder transformation is to transform the response matrix into an upper triangular form so that the solution can be easily computed by backward substitution. The pivoting criterion results from the minimization requirement.

Consider a 6 by 5 response matrix R (see Fig. B.1 on page 152). Suppose only the first corrector is chosen to correct the orbit, we can apply the Householder transformation H_1 (given by the first-column r_1 of R) to zero the elements at locations indicated by the 1's in the figure and get

$$\|\Delta b - R\,\Delta\theta\|_2^2 = \|H_1\Delta b - e_1 * \|_2^2 = \|H_1\Delta b\|_2^2 - (e_1^T H_1\Delta b)^2, \tag{2.35}$$

where $* = \pm\|r_1\|\Delta\theta_1$. The first step relies on $H_1 r_1 = \pm\|r_1\|e_1$ and the second step is due to the fact that the 1-st component of $H_1\Delta b$ can be removed by choosing θ_1 while the rest are not affected by the corrector's setting. Thus the column that maximizes

$$\left(e_1^T H_1\Delta b\right)^2 = \frac{1}{\|r_1\|^2}\left[\left(H_1^T H_1 r_1\right)^T \Delta b\right]^2 = \frac{\left(r_1^T\Delta b\right)^2}{\|r_1\|^2} \tag{2.36}$$

will yield the minimum residual orbit errors and should be permuted to the leading position. Therefore, the column pivoting criterion[3] is to select the column that yields

[3]The usual pivot as described in Appendix B.4 is to choose the column with maximum 2-norm.

the maximum value of Eq. (2.36). If more correctors are used, the procedure can be repeated as in a Householder QR factorization illustrated in Fig. B.1.

Note that, in the MICADO method, the effectiveness of a corrector is measured by its ability to reduce orbit errors assuming that there is no limit to its strength. This could be a pitfall. For example, one corrector may yield a little bit smaller pivot value $|r^T \Delta b|/\|r\|$ but much larger $\|r\|$, which means that it can generate about the same level of orbit correction with much weaker strength. Thus it may be a better choice than the corrector determined by the MICADO method. Also, it is worth mention that the MICADO algorithm may not give the exact least-squares solution when more than one corrector is used.

Chapter 3

Model-Independent Analysis of beam-centroid dynamics

3.1 Introduction

In the last chapter, I reviewed some basic issues and major techniques in BPM based beam dynamics observation and control. There are two fundamental issues in BPM-based model-fitting schemes. One is the accuracy limit set by individual BPM resolution, determined by available technology and budget. The other is the accuracy of the model. This dissertation tackles these issues with a novel approach, which is titled Model-Independent Analysis (MIA) because no particular machine model is required in the analysis. This chapter will present the theoretical features and techniques of MIA.[73, 45]

In order to get more accurate beam orbits, it is a standard practice to do pulse-by-pulse averaging. Such time averages are successful in storage rings since there are stable closed orbits and the pulse repetition rate is high. However, in linacs and rings interesting beam dynamics observations often require pulse-by-pulse measurement of beam orbits. Imagine there are M BPMs at the same location, averaging their readings will obviously improve the measurement accuracy by a factor of $\frac{1}{\sqrt{M}}$. If all the M BPMs are distributed in a drift section, their BPM readings for each pulse can be fit to a straight line and obtain a similar statistical improvement. In reality,

there are usually many BPMs distributed along a beam line with magnets and devices between BPMs. Even so, if the transformation maps among the BPMs are known exactly, orbit fitting method described in Section 2.3 can be used to achieve more accurate results. However, the transformation maps are usually to be determined by BPM measurements. A contribution of this dissertation is to show that one can still improve the resolution limit by taking into account the correlations among a large number of BPM readings. To detect such correlations, BPM readings for a large number of pulses are required. The improvement in observation accuracy may then allow studies of subtle beam dynamics issues and provide better control of the beam.

Since model-fitting approaches rely on the correctness of one's model, they are more suitable when the beam dynamics is well understood, the machine is stable, and a good machine model exists. Often this is not the case. In this chapter, I develop techniques for analyzing the beam dynamics without reference to a particular machine model. Basically matrix and statistical analysis methods are used to systematically analyze the BPM readings for a large number of pulses and a large number of BPMs.

This chapter is organized as follows: Section 3.2 analyzes the BPM readings from a perturbative point of view and discusses the physical base decomposition of a BPM data matrix; Section 3.3 introduce the principal components analysis[38], a major statistical method that will be employed in MIA; Section 3.4 discusses the outcomes of a principal components analysis of BPM data via Singular Value Decomposition (SVD)[39]; Section 3.5 discusses how to improve the BPM resolution limit by using a large number of BPMs; Section 3.6 presents the degrees-of-freedom analysis of a beam line; Section 3.7 discusses how to achieve a physical base decomposition using the time structure of pulse signals; Section 3.8 discusses methods to explore and identify the unknown degrees of freedom revealed in the degrees-of-freedom analysis; Section 3.9 discusses the characteristics of the noise floor of a singular-value spectrum; Section 3.10 provides theoretical explanation of the noise floor behavior; Section 3.11 describes a kick analysis that is helpful to interpret the physical basis; Section 3.12 discusses computation and applications of the Wronskian determinant of the measured dynamical system; and finally, Section 3.13 presents analytical results of an SVD analysis of a simple system in order to better understand such analyses.

3.2 Physical base decomposition of a BPM-data matrix

The central object of a MIA is a BPM-data matrix B, which simply is the data matrix formed by the readings of P pulses on M BPMs (a matrix of P rows and M columns). Physically B contains the transverse beam centroid positions of P pulses sampled by M monitors along a beam line. Clearly, B contains all the information available from BPMs. Let us first examine the physical composition of the matrix B from a perturbative point of view. This is natural since for a short period of time all the pulses are close to an average orbit.

The transverse beam position of a pulse depends on various physical variables such as the initial incoming conditions of the beam, the settings of magnets, and rf conditions. One can Taylor expand the beam position b over all variables as

$$\begin{aligned} b &= b(\bar{x}_1, \bar{x}'_1, \bar{\delta}, \bar{\sigma}_z, \cdots) + \sum_{v \in \{x_1, x'_1, \cdots\}} \left.\frac{\partial b}{\partial v}\right|_{v=\bar{v}} \Delta v \\ &+ \frac{1}{2} \sum_{v_1, v_2 \in \{x_1, x'_1, \cdots\}} \left.\frac{\partial^2 b}{\partial v_2 \partial v_1}\right|_{\substack{v_1=\bar{v}_1 \\ v_2=\bar{v}_2}} \Delta v_1 \Delta v_2 + \cdots \end{aligned} \tag{3.1}$$

where x_1, x'_1, δ, σ_z are respectively the initial beam position, angle, relative energy, and bunch length, given as examples of possible physical variables; the over-bar indicates the expansion points; and $\Delta v = v - \bar{v}$. The zero order term may have a complicated dependency on the variables and is sensitive to unknown BPM offset errors. To eliminate this term, one may subtract the average[1] over a large ensemble

[1] In case nonlinear terms are significant, one should be aware that the ensemble average may not be the real central orbit because the nonlinear terms do not average to zeroes. In such cases, one should try to use small amplitude samples.

of pulses and study the difference

$$\begin{aligned} b - \langle b \rangle &= \sum_v \left.\frac{\partial b}{\partial v}\right|_{v=\bar{v}} (\Delta v - \langle \Delta v \rangle) \\ &\quad + \frac{1}{2} \sum_{v_1,v_2} \left.\frac{\partial^2 b}{\partial v_2 \partial v_1}\right|_{\substack{v_1=\bar{v}_1 \\ v_2=\bar{v}_2}} (\Delta v_1 \Delta v_2 - \langle \Delta v_1 \Delta v_2 \rangle) + \cdots \end{aligned} \tag{3.2}$$

where $\langle\ \rangle$ indicates the average over the ensemble of pulses. Although some second derivatives (which characterize, e.g. the chromatic dependency of the betatron motion) have been found significant at times, the third and higher order terms are generally negligible in linacs and will be dropped (one can easily include more terms when in doubt).

Concentrating for a moment on just the initial conditions and their products, a standard map formalism yields

$$\vec{z}(s) = \sum_{\vec{n}} \vec{R}_{\vec{n}}(s_0 \rightarrow s) \cdot z_0^{\vec{n}}.$$

Each $\vec{R}_{\vec{n}}(s_0 \rightarrow s)$ is a possible physical vector. Since in a ring the z_0 changes in each turn, with sufficient resolution and orbit amplitudes, one might hope to observe the $\vec{R}_{\vec{n}}(s_0 \rightarrow s)$. I will explore map measurements in later chapters.

Back on track, I treat the first and higher order terms on the same footing and rewrite Eq.(3.2) in a concise form

$$b - \langle b \rangle = \sum_{\{q\}} q\, f_q \tag{3.3}$$

where the variable $q = \frac{\Delta v - \langle \Delta v \rangle}{\mathrm{std}(\Delta v)}$ or $\frac{\Delta v_1 \Delta v_2 - \langle \Delta v_1 \Delta v_2 \rangle}{\mathrm{std}(\Delta v_1 \Delta v_2)}$ and f_q is the corresponding derivative $\left.\frac{\partial b}{\partial v}\right|_{\bar{v}} \cdot \mathrm{std}(\Delta v)$ or $\frac{1}{2} \left.\frac{\partial^2 b}{\partial v_2 \partial v_1}\right|_{\bar{v}_1,\bar{v}_2} \cdot \mathrm{std}(\Delta v_1 \Delta v_2)$. The physical variables are normalized by their standard deviations over the ensemble of pulses, so that all the q's are dimensionless and reflect the relative changes (otherwise one has to deal with different quantities such as 10^{-5} rad, 10^6 volts, etc.), while all the f's have the same dimension as the BPM readings.

Eq.(3.3) tells us that a beam orbit is a linear combination of a limited number of "basic" orbits given by the f_q's. In other words, the BPM reading pattern generated by each pulse is a superposition of certain basic patterns. This fact allows us to apply linear algebra concepts and matrix analysis techniques. According to Eq.(3.3), the BPM-data matrix B (from now on it consists of $b - \langle b \rangle$ instead of b) which is the ensemble of P pulses monitored with M BPMs can be written as

$$B = QF^T + N \tag{3.4}$$

where matrices $Q_{P\times d} = [\vec{q}_1, \cdots, \vec{q}_d]$, $F_{M\times d} = [\vec{f}_1, \cdots, \vec{f}_d]$, and $N_{P\times M}$ contains the noise associated with each BPM reading. The column vector $\vec{q}_i$ contains the P values of the i-th physical variable and $\vec{f}_i$ contains the M components of the corresponding physical pattern. The q's are referred to as temporal patterns or time structures of pulses, and the f's as spatial patterns or physical vectors. I assume without loss of generality that all the physical vectors are linearly independent, i.e. F has full column rank given by d. Neglecting BPM errors, they form a complete basis for the row space of the BPM-data matrix (i.e. range of B^T). Unlike P and M, which can be chosen at will, dimension d is determined by the dynamics. An MIA achievement (see Section 3.6) is the determination of d. Generally d is a small number and I choose P and M so that $d \ll M \ll P$ to obtain statistical benefits. Typical numbers are $d \sim 10$, $M \sim 10^2$, and $P \sim 10^3$.

One concern on the nonlinear terms in the above decomposition is that whether the temporal patterns in Q are independent. Although the nonlinear terms are derived from the linear terms, all the patterns are indeed **linearly** independent of each other. Take the x and x^2 terms as an example, one can not find a coefficient α such that $x + \alpha x^2 = 0$ for all potential x. Nonetheless, the Q matrix tends to be ill-conditioned when high order nonlinear terms involved because of near degeneracy.

The matrix F contains stationary beam-line properties such as the dispersion function. Matrices Q and N contain (often stochastic) quantities that change from one ensemble to another. However, $\frac{1}{P}Q^TQ = C_Q$, which is the sample correlation matrix of the q variables, contains statistical properties of the ensemble of pulses. Thus,

if everything is stable, C_Q contains only stationary machine properties. Similarly, $\frac{1}{P}N^TN$ characterizes BPM resolutions as well as possible correlations in BPM noise. For convenience, I normalize B, Q, and N by $\sqrt{P}$, so that the important variance-covariance matrix of BPM readings and the correlation matrix of temporal signals (q's) can be formed simply as

$$C_B = B^TB \quad \text{and} \quad C_Q = Q^TQ. \tag{3.5}$$

Note that the statistical meaning of Eq.(3.5) requires that the column means (averages) of B and Q have been taken out. Obviously

$$C_B = F\,C_Q F^T \tag{3.6}$$

if one neglects noise. This shows the mathematical relationship between C_B and C_Q.

Eq.(3.4) is called a physical base decomposition of the BPM-data matrix. Although straightforward, it is an important statement both conceptually and mathematically. In fact, it sets up the statistical model of the problem. Such a model has been widely studied by statisticians and mathematicians. It also contains the goal of MIA: to find F from measured B and Q. Physically speaking, the major goals of beam dynamics analysis are,

- identify a complete set of variables contributing to the beam motion and the physics behind them; and
- determine all the physical-basis patterns and the physics behind them.

To achieve these objectives, it is essential to isolate signals from noise. In the following sections, I will discuss how to handle N, find d, and obtain F.

Physical base decomposition is not unique in the sense that one can choose physical variables differently. Therefore the BPM-data matrix itself does not contain sufficient information to determine a physical base decomposition. Extra information about the physical variables is necessary. Although it is possible to impose certain mathematical requirements (such as orthogonality) to make the decomposition unique, such a decomposition will have limited use due to the lack of a proper physical interpretation.

Before focusing on MIA, I apply Eq.(3.4) to the conventional orbit-fitting problems discussed in Section 2.3. This will illustrate the meaning of Eq.(3.4). By setting $P = 1$, orbit-fitting problems can be accommodated easily into Eq.(3.4). F is supposedly given by the machine model and the goal is to find the corresponding beam parameters in Q. When the number of BPMs is larger than the degrees of freedom ($M > d$), least-squares fitting can be used to find the best solution:

$$Q = BF(F^T F)^{-1}. \tag{3.7}$$

This expression simply restates Eq.(2.20). Error analysis discussed in Section 2.3 indicates the potential to reduce the BPM random noise effects by using more and more BPMs. The following sections will show how to realize this potential with MIA.

3.3 Principal Components Analysis via Singular Value Decomposition

Principal component analysis is perhaps the most widely used multivariate statistical analysis technique.[38] It was initially described by Karl Pearson (1901) and further developed by Hotelling (1933). The major motivation is to account for the observed sample variance-covariance with the minimum number of variables, which might be linear combinations of a much larger number of observed variables. The basic idea is to find the first "principal axis" in the data-point space such that the sample variance of the components of all data points along this axis is maximum, then find the next such axis that is orthogonal to the other principal axes, and so on. Clearly such a procedure could determine the dimensionality of the system under investigation and minimize the number of variables needed to describe the observed variations. Principal-components analysis is a major linear data reduction technique.

From the measured BPM data B, the sample variance-covariance $B^T B$ is readily available. In principle, each BPM reading might be a variable that leads to the observed variations (e.g. there are no beam signals). However, we are interested in the situations that the BPM data manifest the underlying beam dynamics that are

responsible for the correlations. To find the leading "super" variable that accounts for most of the observed variations, we seek a linear combination of the M-BPM readings that yields the maximum variation or equivalently find a set of coefficients $v = \{v_1, \cdots, v_M\}^T$ such that var(Bv) is maximum,[2] i.e.

$$\frac{v^T B^T B v}{v^T v} = \text{max.}\,, \qquad v \in R^M \tag{3.8}$$

Naturally the constraint $v^T v = 1$ is used since we are interested in the direction of vector v, not its length. Because $B^T B$ is a symmetric matrix, we know that v must be the eigenvector with the largest eigenvalue (Courant-Fischer Minimax Theorem, Eq.(B.23)), which corresponds to the variance.[3] After finding the first principal component $v^{(1)}$, the variation due to it is taken out and then the procedure is repeated for the next largest component $v^{(2)}$, and so forth. Thus at each step the new component need to be orthogonal to the previous principal components. For a symmetric matrix, the eigenvectors of discrete eigenvalues are orthogonal to each other, i.e. the eigenvectors of the symmetric matrix $B^T B$ provide all the principal components.

Geometrically, the M BPM readings of each pulse correspond to a point in R^M, the first principal component $v^{(1)}$ gives a vector pointing to the direction that has the maximum variation; the second principal component points to the maximum-variation direction that is perpendicular to the first one; and so on. From such a geometric interpretation it is easy to understand that a principal components analysis should be very useful to separate the subspaces containing signals from the subspaces containing noise, because the former will have certain preferred directions. Furthermore, the dimensionality of the signal subspace will be revealed and an orthogonal basis of the signal subspace will be generated in the principal components analysis.

[2] As a simple example, consider pulses injected into a linear beam line with negligible initial angular deviation. Then the signal at the m-th BPM reads $R_{11}(m)x_p$ where x_p is the incoming position of the p-th pulse. In the M dimensional space, these pulses vary in the direction defined by the vector $\vec{R} \equiv \{R_{11}(1), \cdots, R_{11}(M)\}$. To find the maximum variation direction $\vec{v}$, one maximizes $\sum_p (\vec{v} \cdot (\vec{R} x_p))^2$ and would obviously find $\vec{v} \propto \vec{R}$, which is just the betatron pattern.

[3] A standard way to solve such a constrained maximization problem is to use Lagrange multiplier λ and solve $\partial_{v_i}\big[\, v^T B^T B v - \lambda(v^T v - 1)\big] = 0$, which yields $B^T B v = \lambda v$.

Using the new set of variables (coordinate system) given by the principal components, we can calculate their pulse-by-pulse values, which is given by $w^{(1)} = Bv^{(1)}$, $w^{(2)} = Bv^{(2)}$, and so on. Note that, like the "spatial vector" v's, these "temporal vector" w's are also orthogonal to each other because $(w^{(i)})^T w^{(j)} = (v^{(i)})^T B^T B v^{(j)} = \lambda^{(i)}\delta_{ij}$. Putting the spatial and temporal vectors into two matrices $V = [v^{(1)}, v^{(2)}, \cdots]$ and $W = [w^{(1)}, w^{(2)}, \cdots]$, we can decompose the BPM data matrix B into

$$B = WV^T. \tag{3.9}$$

However, it is clear that principal components analysis in general cannot yield the physical base decomposition because the physical bases are usually not orthogonal. It is unrealistic to hope for understanding everything from the BPM data only. Obviously, extra information is necessary to define the physical basis. Nonetheless, it is possible to develop beam orbit control schemes using the orthogonal basis without knowing the physical basis.

We can further separate out the variances and normalize the temporal base vectors to u's, i.e. $u^{(1)} = w^{(1)}/\sqrt{\mathrm{var}(w^{(1)})}$, $u^{(2)} = w^{(2)}/\sqrt{\mathrm{var}(w^{(2)})}$, etc. and get

$$B = USV^T \tag{3.10}$$

where $S \equiv \mathrm{diag}\{\sigma_1, \sigma_2, \cdots\} = \mathrm{diag}\{\sqrt{\mathrm{var}(w^{(1)})}, \sqrt{\mathrm{var}(w^{(2)})}, \cdots\}$. The σ's are called singular values of the matrix B and equal the square roots of the eigenvalues of the matrix $B^T B$. Such a decomposition is well known as the Singular Value Decomposition (SVD). The set of singular values is called the singular value spectrum. Therefore, when computing an SVD of B, we in fact are performing a principal components analysis of the BPM data.

3.4 Outcomes of an SVD analysis

This section will focus on the physical meaning of the SVD results in order to illustrate their usefulness and limitations for beam dynamics analysis.

The basic structure (notations) of an SVD of a matrix B is given by:

$$B = USV^T = \sum_{i=1}^{d} \sigma_i u_i v_i^T, \tag{3.11}$$

where $U_{P\times P} = [u_1, \cdots, u_P]$ and $V_{M\times M} = [v_1, \cdots, v_M]$ are orthogonal matrices, $S_{P\times M}$ is a rectangular matrix with nonnegative σ_i along the upper diagonal in non-increasing order. $d = \text{rank}(B)$ is the number of nonzero singular values. σ_i is the i-th largest singular value of B and the vector u_i (v_i) is the i-th left (right) singular vector. Often (assuming $M < P$ since we are interested in overdetermined systems only) a trimmed down version is used in which only the first M columns of U and the first M rows of S are kept. The singular values are uniquely determined and the corresponding singular vectors are determined up to a sign (assuming no degeneracy). The singular values reveal information of the matrix rank while each set of singular vectors forms an orthogonal basis of the various spaces of the matrix. These properties make the SVD extremely useful. There are direct relationships between an SVD and the eigenvalue problem of real symmetric matrices, which can be seen from

$$B^TB = VS^2V^T \quad \text{and} \quad BB^T = US^2U^T, \tag{3.12}$$

i.e. the column vectors of V (U) are eigenvectors of the real symmetric matrix B^TB (BB^T) with eigenvalues given by the corresponding diagonal term σ_i^2's.

Unlike the physical base decomposition given in Eq.(3.6), the orthogonal base decomposition in Eq.(3.12) is uniquely determined by B^TB, which is the sample covariance matrix of BPM readings and should be a stable machine property.[4] Therefore, we can conclude that both the singular values (in S) and the right singular vectors (in V) should be repeatable for different ensembles of pulses, providing that the machine is stable (i.e. all machine conditions are the same). On the other hand, the U matrix will change from one ensemble to another because BB^T does not represent a stationary statistical property of the system.[4] In fact, V reflects spatial "lattice"

[4] B^TB implies summation over pulses and reflects the sample variance-covariance of the underlying distributions (see Eq.(3.6)), which should be stable. BB^T implies summation over BPMs, thus it obviously depends on the samples, which are changing all the time.

properties of a machine and U reflects the pulse-by-pulse temporal variations of the beam motion. The word lattice is in quotes because it could contain effects due to the beam via wakefields, for example.

Now let us check out the SVD results of an experimental data set consisting of horizontal BPM readings from the SLC linac for 5000 pulses and 130 BPMs. Results from other machines have shown similar features. Fig. 3.1 plots the singular value spectrum of B. It shows that most of the singular values are small and about the same size. They are due to BPM noise. Thus the long flat part is called the noise floor. It has interesting characteristics which I will describe in Section 3.9. Above the noise floor, there are more than 10 significant singular values, which tell us the degrees of freedom of the system, valuable information for a beam dynamics analysis. For example, it indicates that the beam dynamics is much more complicated than simple linear betatron motions. As discussed in the last section, the size of a singular

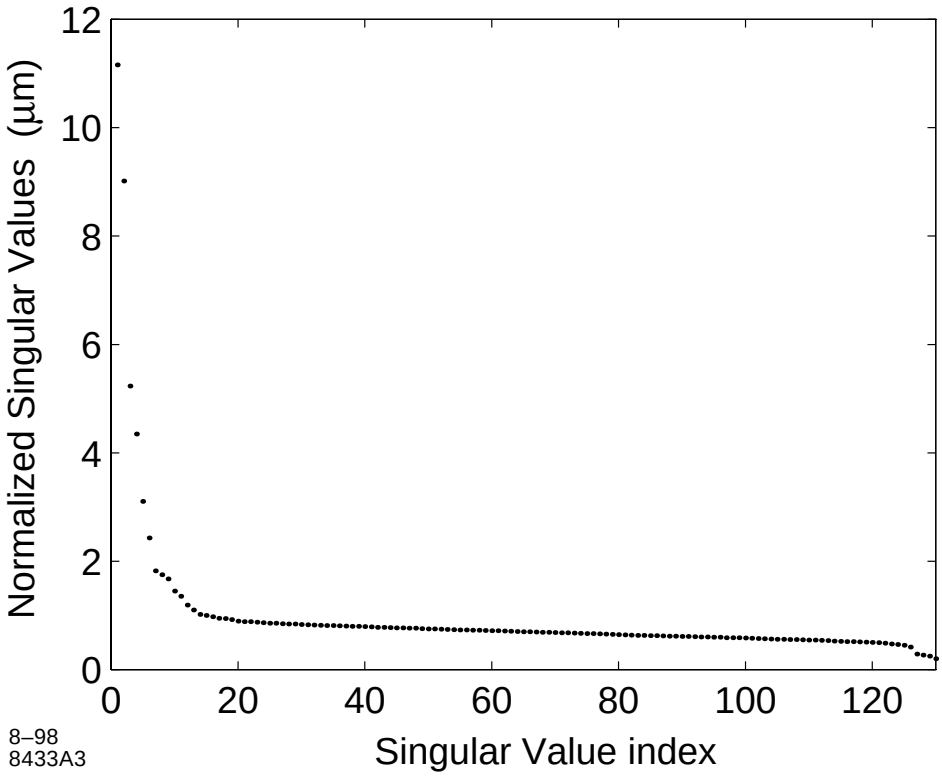

Figure 3.1: Singular-value plot

value reflects the variation magnitude (in the same unit of BPM readings) of the corresponding mode. Let us take a more detailed look of the singular values. B^TB can be written as

$$\sum_{i=1}^{d} \sigma_i^2 v_i v_i^T = \begin{bmatrix} \mathrm{var}(\mathrm{BPM}_1) & \mathrm{cov}(\mathrm{BPM}_{12}) & \cdots \\ \mathrm{cov}(\mathrm{BPM}_{21}) & \mathrm{var}(\mathrm{BPM}_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{3.13}$$

Comparing diagonal terms we have

$$\mathrm{var}(\mathrm{BPM}_k) = \sum_{i=1}^{d} \sigma_i^2 \; v_i(k)^2, \qquad k = 1, \cdots, M \tag{3.14}$$

and

$$\sum_{k=1}^{M} \mathrm{var}(\mathrm{BPM}_k) = \sum_{i=1}^{d} \sigma_i^2. \tag{3.15}$$

These equations confirm that the variance of the k-th BPM readings is the sum over the i modes with $\sigma_i^2 \, v_i(k)^2$ from each. Of course the square of a singular value is the sum of the variances of BPM readings due to the corresponding mode. Since a spatial vector v_i is normalized to unity, an evenly distributed signal $v_i^2(k) \sim \frac{1}{M}$ requires $\sigma_i^2 \sim M$. For localized BPM noise, $v_i^2(k)$ is constant and so is σ_i^2. Normalized values are used in the singular value spectrum plot since they reflect the average amplitudes of signals. Under this normalization, $\hat{\sigma}_{\mathrm{signal}}$ is roughly constant with changing M while $\hat{\sigma}_{\mathrm{noise}} \propto \frac{1}{\sqrt{M}}$. This is a very important characteristics of the singular values.

Figure 3.2 shows the first 7 singular vectors and corresponding singular values (in μm on the left labels) of the above SLC data set. We see that the singular values go down quickly from about 10μm to 1μm. Therefore at a 1μm coherent signal level, all the motions observed in B would be a linear combination of less than 10 modes. The most striking patterns are modes #5 and #6. Clearly they are due to individual BPMs. This example shows that one can easily identify problematic BPMs, valuable information for any beam control and dynamics observation. The top 2 modes are

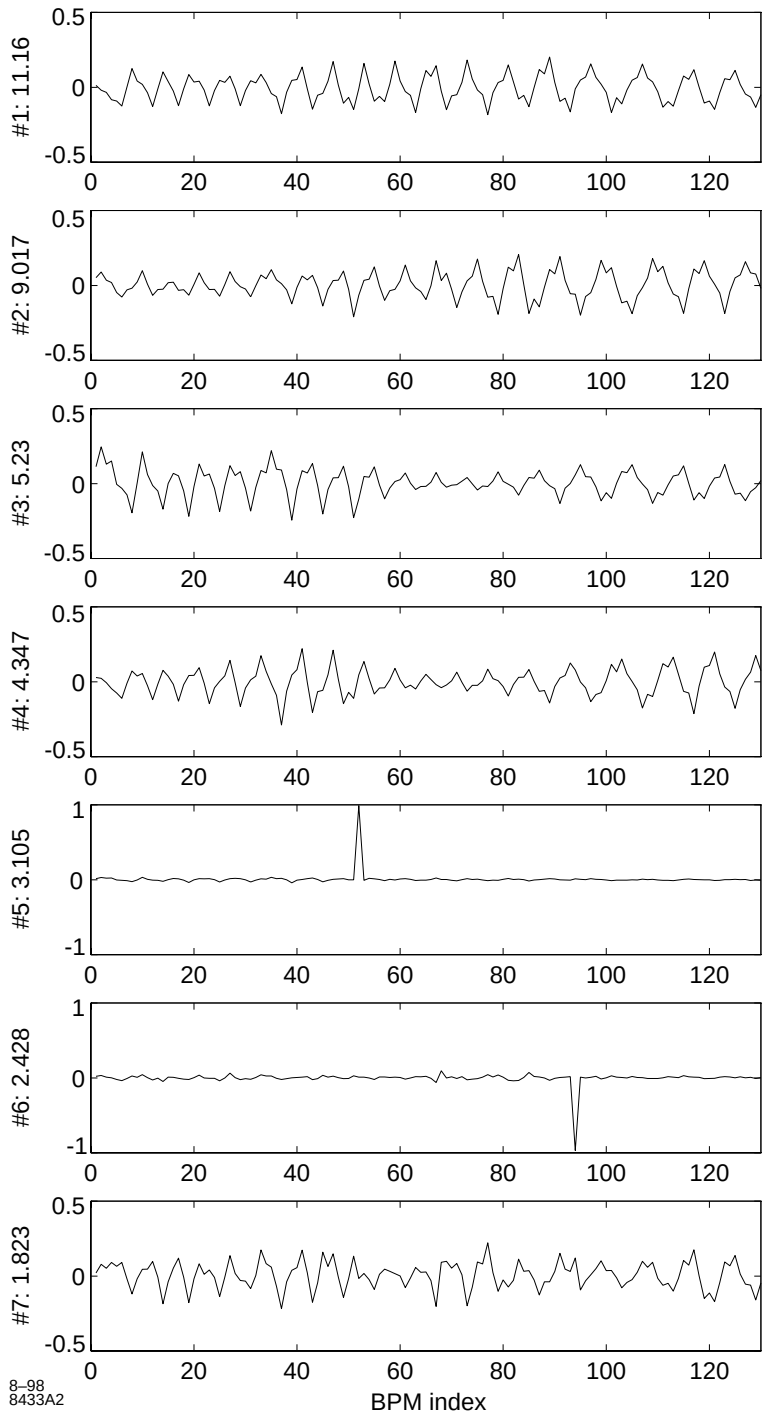

Figure 3.2: Singular-vector plot

mainly due to the 2 degrees of freedom in the horizontal betatron motion. The quasi-periodic structure is due to the focusing-defocusing optics. The vector sum of these two modes is approximately proportional to the beta function (see Eq. 3.40).

Although singular-vector plots yield valuable information such as the location of noisy BPMs, it has limited use otherwise, since the orthogonal decomposition often mixes various different physical effects. As mentioned earlier, extra information is required to determine the physical base decomposition. Nonetheless, the orthogonal base decomposition provides an important step towards physical base decomposition. Another important outcome of SVD analysis is the separation of signal subspace from the noise subspace.

To emphasize that the right (v's) and left (u's) singular vectors reflect the spatial and temporal characteristics of a mode, Fig. 3.3 shows both the v vector and u vector corresponding to the same singular value. It results from a SVD analysis of an SLC data set with one corrector dithered (i.e. varied according to certain given

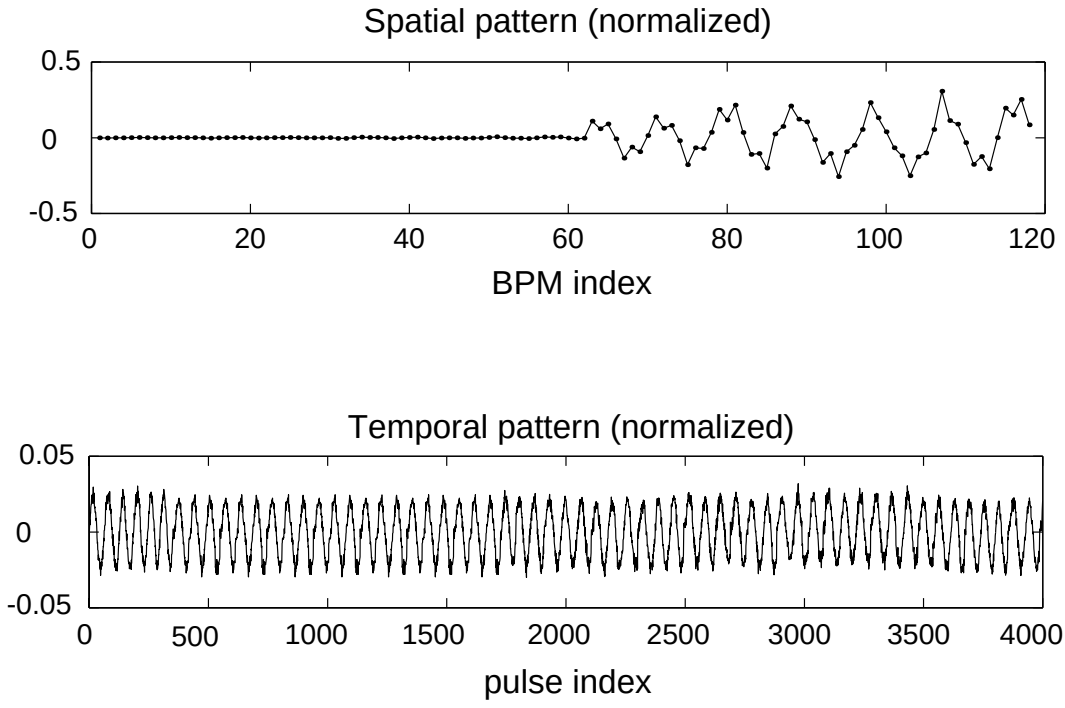

Figure 3.3: The spatial and temporal patterns of the mode due to a dithered corrector.

pattern, e.g. sinusoidally). The spatial vector clearly yields the location where the dither started and the corresponding betatron pattern. The temporal vector yields the given pulse-by-pulse dither pattern. (The data set consisted of 20 subsets with 200 pulses in each.) This example vividly demonstrates the physical meaning of the singular vectors. The singular value yields the rms amplitude of the mode.

3.5 Reduction of random noise — Breaking the pulse-by-pulse resolution limit

In beam dynamics experiments, one often encounters BPM resolution problems. Sometimes simple pulse-by-pulse averaging can improve accuracy. But often single-pulse measurements with resolution better than the BPM resolution are desirable. In this case, an option might be some sort of average over a large number of BPMs. Were there M identical BPMs at the same location to monitor the beam position, averaging these BPM readings would improve the measurement by a factor of $\frac{1}{\sqrt{M}}$. In reality, one has many BPMs distributed along a beam line. The question is: even though we do not know the exact relations among BPM readings for a pulse, can we take advantage of the potential statistical benefits of using a large number of BPMs?

This can be achieved from a SVD analysis of the BPM-data matrix. The method is fairly simple: compute the singular value decomposition of $B = USV^T$, set the singular values due to noise to zero to form the noise-cut $\underline{S}$, then recompute $U\underline{S}V^T$. The resulting matrix

$$\underline{B} = U\underline{S}V^T + O\left(\sigma_{\mathrm{BPM}}\sqrt{\frac{d}{M}}\right) \tag{3.16}$$

has a noise term reduced by $\sqrt{d/M}$ where d is the dimension of signal space. In the following sections, I will discuss the SVD in detail and show how to identify the singular values due to noise.

Figure 3.4 demonstrates the effect of the noise-cut. As a test, 5000 pulses over 125 BPMs were generated to simulate various signals in SLC (see Section 4.1 for more details). Then random noise, 1 μm for the first 7 and 10 μm for the rest BPMs, was

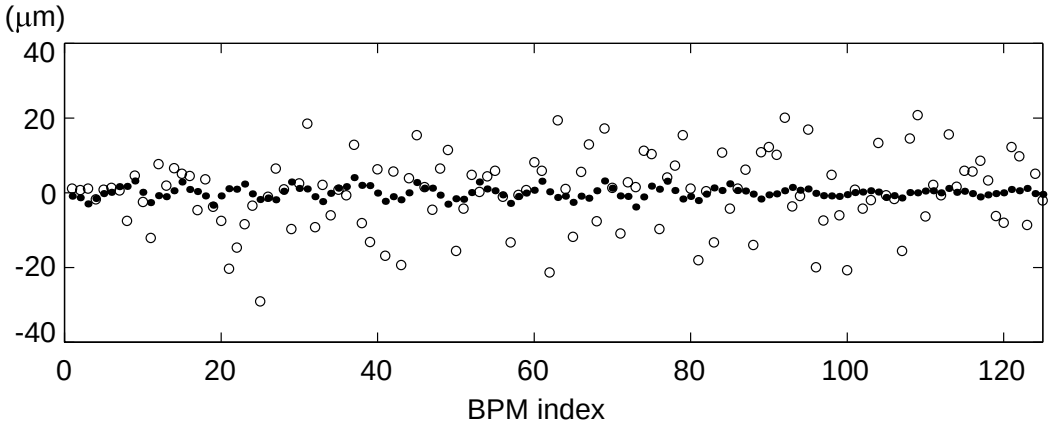

Figure 3.4: Effect of cutting noise. "○" represents the initial noise and "●" represents the residual noise.

added. After cutting the noise, the residual noise was obtained by subtracting the signals from the noise-reduced matrix. Figure 3.4 plots the original noise in circles and residual noise in dots for the first pulse. Results for all other pulses are similar. It is remarkable that this simple procedure can significantly reduce the random noise of each individual BPM reading. In other words, we can improve effective BPM resolution individually by using a large number of BPMs and sufficiently large number of pulses. Though simple and powerful, this method seems not to have been used before for beam dynamics analysis. However, a similar method (i.e. setting signal instead of noise singular values to zero) has been used for estimating BPM resolutions [66].

3.6 Analyzing degrees of freedom along a beamline

In any dynamical system, the degrees of freedom of the system offers very basic information about the dynamics. It reflects how many things are independently changing. A simple example can illustrate why it is important to analyze the degrees of freedom in a beam line. Suppose one has an "ideal" beam line in which there is no coupling between horizontal and vertical planes, no significant nonlinearities, current is low so wakefield effects are negligible, no energy variation, etc., then the

only possible motions are betatron motions excited by the initial beam position and angle. It is clear that there are only 2 degrees of freedom (usually characterized by the so called sine-like and cosine-like trajectories) available in the system. Now, suppose one of the corrector magnets in the beam line malfunctions and drifts around, it will kick the beam and excite an independent betatron motion starting from its location. Analyzing such a system, one will find 3 independent BPM patterns instead of 2. Furthermore, one can try to find where the new degree of freedom starts, and therefore locate the jitter source.

Section 3.4 showed that the singular value spectrum can reveal the degrees of freedom (d=rank(B)) of a system, after measuring the BPM-data matrix B for a sufficiently large ensemble of pulses. If there is no noise, the number of nonzero singular values gives the d value. In practice, one has to find the noise level and set up a criterion to determine which singular values are significant. This is a subtle issue which I will address in Section 3.9. Fig. 3.1 showed one typical singular-value plot. Above the noise floor, there are about 10 singular values. As shown in Fig. 3.2, some of those are due to large individual BPM noise levels instead of beam dynamics. Even so, there are still more degrees of freedom in this system than assumed by the typical on-line machine models. A traditional model-fitting approach is bound to miss important dynamics in this beam line!

It is useful to trace the increase of the number of degrees of freedom along the beam line using SVD analyses of an increasing number of BPMs, i.e. including only the first m BPM readings and perform an SVD analysis on this data set for $m \leq M$. Such systematic SVD analyses can reveal the locations where new degrees of freedom appear. These locations could be a jittering source such as a varying corrector, or a structure misalignment that shows up as a jittering source because of current jitter, and so on. Figure 3.5 is an example of such a plot (using the same data set of Fig. 3.1 including only the top 10 singular values), which is called a "degrees-of-freedom plot". Unlike Fig. 3.1, the singular values plotted here are not normalized by the number of BPMs. (thus one must divide the ending values in Fig. 3.5 by $\sqrt{130}$ in order to get the first 10 points of Fig. 3.1). There are many general features in a degrees-of-freedom plot of a normal running beam line. Modes due to random noise yield

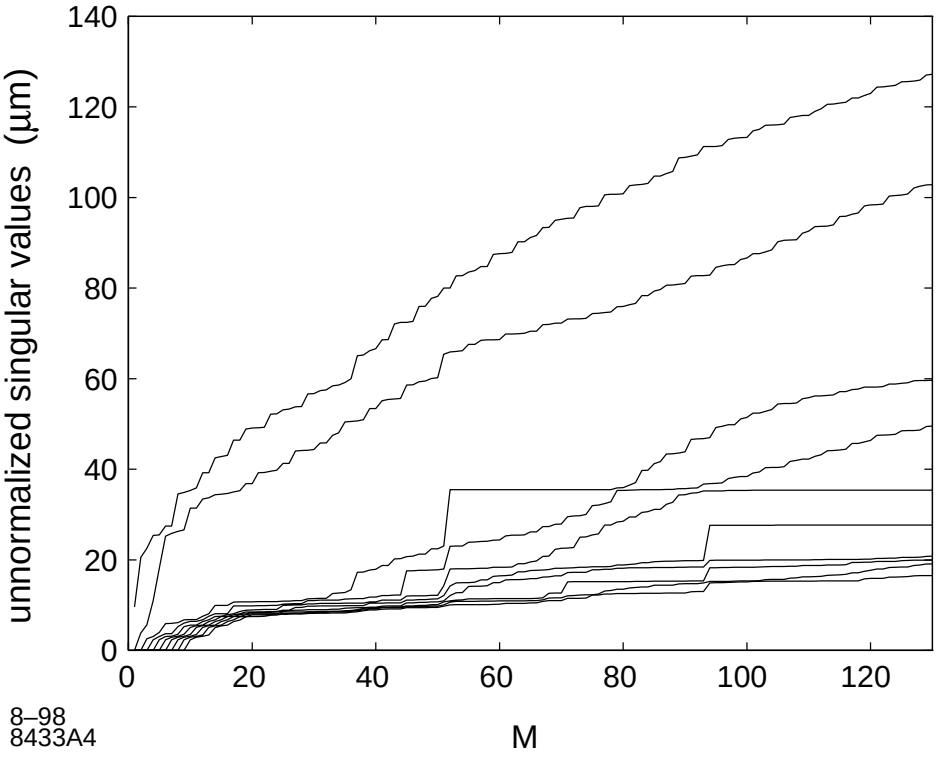

Figure 3.5: Degree-of-freedom plot

flat lines, while coherent signals grow with the number of BPMs used. The slope of a curve indicates the local strength of that signal. Usually, there are two curves on the very top of the plot well separated from the rest. They are mainly due to the 2 betatron modes. Sometimes, there is only one large singular value curve because the other mode is hardly excited due to a beam injection constraint (e.g. a negligible angular fluctuation). There are small wiggles on most of the curves, which are the result of periodic lattice function changes. The merit of a degrees-of-freedom plot is to analyze the appearance of new degrees of freedom, and the exact values are not that important since they most often do not correspond to strengths of physical modes (see discussion on orthogonal basis in the previous sections). If a mode is due to an individual BPM, we will see a curve like a step function starting at that BPM. The step level indicates the noise magnitude. There are 3 such cases clearly shown in Fig. 3.5 (all noisy BPMs are kept in order to show their effects). There is a baseline

in a degrees-of-freedom plot, which reflects the BPM resolutions. In this example, the BPM resolutions are about 10 μm. The beginning slope of eigenvalues 3–10 is special to this example, since there are several high resolution BPMs at the beginning of the linac. However, $M \geq d$ is required to establish various modes. Thus, the points within the first few BPMs are only useful to determine initial betatron amplitudes.

Before leaving this section, I mention that instead of using more and more BPMs as shown above, one can analyze various subsets (e.g. every 10 BPMs) of all BPMs and locate where new degrees of freedom appear. Unfortunately, this approach loses the advantage of using large number of BPMs, and may have more trouble detecting weak signals. Though this problem may be alleviated by cutting noise as described before, there are physical patterns which look very much like a betatron motion locally and will be degenerate in a localized degrees-of-freedom analysis. Therefore, such an approach is usually not preferred.

3.7 Physical base decomposition via time structure of pulse signals

As I mentioned earlier, the orthogonal bases obtained from SVD are often a mixture of various physical patterns and therefore hard to interpret. Extra information is necessary to achieve a physical base decomposition. In a beam line, in addition to the transverse beam positions, there are various kinds of pulse-by-pulse beam and machine parameters that can be monitored. At SLC, for example, we can monitor beam current, bunch length, incoming beam (longitudinal) phase, relative beam energy, klystron phases along the linac, and so on. This section will discuss how to take advantage of such information. If one prefers a simple example before the general description, please start from Eq.(3.22) and then come back.

Mathematically, this problem is similar to the orbit fitting problem discussed earlier. Instead of knowing F, we know Q (or a subset of it) and wish to solve for F. If we knew all the physical variables with sufficient accuracy, the corresponding

physical basis could be computed as

$$F^T = C_Q^{-1} Q^T B + O\left(\frac{\sigma_{\rm BPM}}{\sqrt{P}}\right). \tag{3.17}$$

This expression emphasizes the importance of underlying correlations among the observed variables.

Note that the accuracy of Eq.(3.17) does not rely on the number of BPMs used. It simply fits the readings of each BPM to various temporal patterns individually and ignores any correlations among BPM readings. As I discussed earlier, the BPM noise can be reduced statistically by taking into account the correlations among BPM readings. Therefore, if we cut the noise first and then apply Eq.(3.17), the noise level can potentially be reduced by a factor of $\sqrt{\frac{d}{M}}$, and we have

$$F^T = C_Q^{-1} Q^T U \underline{S} V^T + O\left(\sigma_{\rm BPM}\sqrt{\frac{d}{P\,M}}\right) \tag{3.18}$$

where USV^T is the SVD of B, and $\underline{S}$ indicates the zeroing of singular values that have been identified as due to noise. This statistical error limit (60 nm in the above SLC case) is hard to achieve due to problems such as machine instability and incomplete information in Q. Nonetheless, it indicates the inherent potential sensitivity of this method.

Usually we know only a subset of Q, say Q_s of $Q = [Q_s, Q_r]$. We can still calculate F_s according to Eq.(3.17) with Q_s, The error due to the missing part is

$$(F_s - F_s^{exact})^T = (Q_s^T Q_s)^{-1} Q_s^T Q_r F_r^T \tag{3.19}$$

Therefore, if the known subset Q_s is uncorrelated with the remaining unknown temporal patterns, i.e. $Q_s^T Q_r = 0$, then we would obtain the same results as if we had measured all Q. Otherwise, the unknown part of the physical basis (i.e. F_r) will be mixed into the measured parts. This can be a limitation of a totally non-invasive procedure. However, many known physical variables can be slightly modulated on purpose (incoming position, bunch length, and longitudinal phase, for example). In

this way the patterns due to these changeable variables can be identified and the patterns corresponding to unknown or unchangeable variables can be further clarified. Then, one will change additional suspected variables in search for these unknowns.

Often the measured temporal patterns of certain physical variables have limited accuracy due to measurement difficulties. To evaluate the effect of such errors, let us assume the measured signals are $Q + \Delta Q$ where ΔQ represents the error, then the error in F can be written as

$$\begin{aligned} \Delta F^T &\equiv C^{-1}_{Q+\Delta Q}\,(Q+\Delta Q)^T B - F^T \\ &= -\,C^{-1}_{Q+\Delta Q}\,\Delta Q^T \Delta Q\; F^T \end{aligned} \tag{3.20}$$

where I have assumed that the measurement errors ΔQ are independent of Q, i.e. $Q^T \Delta Q = 0$. Eq.(3.20) shows that the errors in temporal pattern measurement come into play mainly at the second order. Therefore, they are more tolerable than BPM errors. Furthermore, if we know the variance-covariance $\Delta Q^T \Delta Q$ of the measurement errors,[5] Eq.(3.17) can be modified to take Q errors into account via

$$F^T = \left(I - C^{-1}_{Q+\Delta Q}\Delta Q^T \Delta Q\right)^{-1} C^{-1}_{Q+\Delta Q}(Q+\Delta Q)^T B \tag{3.21}$$

Note that all the quantities in this expression are measurable.[6]

Eq.(3.17) is mathematically the same as Eq.(3.7) but they are different physically, and it turns out to be very useful in the measurement of physical-basis patterns. Eqs. (3.17), (3.18) and (3.21) seem to have never been used before, at least not in this generalized form.

Here is a simple example to illustrate the above abstract matrix formulae. Consider an ensemble of BPM readings given by

$$\{f_p = \Delta x_p f_x + \Delta y_p f_y \mid p = 1, \cdots, P\} \tag{3.22}$$

[5] Of course there is no way to know ΔQ, but statistical characteristics such as $\Delta Q^T \Delta Q$ may be obtained through equipment calibrations.

[6] The correction factor in the first parenthesis of Eq.(3.21) reduces to Eq.(2.31) for that simple case, in which $\Delta Q^T \Delta Q = \sigma^2_{\text{noise}}$ and $C_{Q+\Delta Q} = \sigma^2_{\text{signal}} + \sigma^2_{\text{noise}}$.

where f_x and f_y are two base vectors (BPM patterns) such as betatron modes. Δx_p and Δy_p are the "initial" parameters for each pulse. Suppose we can measure Δx and Δy but with certain noise given by ϵ_x and ϵ_y respectively, i.e. we can measure the temporal patterns f, $\Delta x + \epsilon_x$, and $\Delta y + \epsilon_y$. For simplicity, let us consider the case that the signals are independent of all noise and the noise is uncorrelated. To extract f_x and f_y, simply compute $\langle(\Delta x + \epsilon_x)f\rangle_p$ and $\langle(\Delta y + \epsilon_y)f\rangle_p$ and find

$$\begin{bmatrix} \langle(\Delta x + \epsilon_x)f\rangle \\ \langle(\Delta y + \epsilon_y)f\rangle \end{bmatrix} = \begin{bmatrix} \langle(\Delta x + \epsilon_x)^2\rangle - \langle\epsilon_x^2\rangle & \langle(\Delta x + \epsilon_x)(\Delta y + \epsilon_y)\rangle \\ \langle(\Delta y + \epsilon_y)(\Delta x + \epsilon_x)\rangle & \langle(\Delta y + \epsilon_y)^2\rangle - \langle\epsilon_y^2\rangle \end{bmatrix} \begin{bmatrix} f_x \\ f_y \end{bmatrix} \tag{3.23}$$

where $\langle\ \rangle$ means ensemble average. I have used $\langle\Delta x\,\epsilon_{x,y}\rangle = 0$, $\langle\Delta y\,\epsilon_{x,y}\rangle = 0$, $\langle\epsilon_x\,\epsilon_y\rangle = 0$, i.e. the signals Δx and Δy are assumed independent of noise, and the noise amplitudes are uncorrelated with each other. Inverting the matrix gives the solution

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} \langle(\Delta x + \epsilon_x)^2\rangle - \langle\epsilon_x^2\rangle & \langle(\Delta x + \epsilon_x)(\Delta y + \epsilon_y)\rangle \\ \langle(\Delta y + \epsilon_y)(\Delta x + \epsilon_x)\rangle & \langle(\Delta y + \epsilon_y)^2\rangle - \langle\epsilon_y^2\rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle(\Delta x + \epsilon_x)f\rangle \\ \langle(\Delta y + \epsilon_y)f\rangle \end{bmatrix} \tag{3.24}$$

Note that all the quantities in the right hand side are in principle measurable and an exact solution can be obtained even though there is random noise in the measured signals. We can separate the noise effect and write the solution as

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \left[I - C_{Q+\Delta Q}^{-1} \begin{pmatrix} \langle\epsilon_x^2\rangle & 0 \\ 0 & \langle\epsilon_y^2\rangle \end{pmatrix} \right]^{-1} C_{Q+\Delta Q}^{-1} \begin{bmatrix} \langle(\Delta x + \epsilon_x)f\rangle \\ \langle(\Delta y + \epsilon_y)f\rangle \end{bmatrix} \tag{3.25}$$

where

$$C_{Q+\Delta Q} = \begin{bmatrix} \langle(\Delta x + \epsilon_x)^2\rangle & \langle(\Delta x + \epsilon_x)(\Delta y + \epsilon_y)\rangle \\ \langle(\Delta y + \epsilon_y)(\Delta x + \epsilon_x)\rangle & \langle(\Delta y + \epsilon_y)^2\rangle\rangle \end{bmatrix}$$

is the covariance matrix of the measured signals (with noise in it). The first bracket in Eq.(3.25) represents the modification due to the noise, and the rest is the typical least-squares solution. This equation is a special case of Eq.(3.21).

3.8 Exploring the unknown degrees of freedom

The degrees-of-freedom analysis discussed in section 3.6 can reveal how many linearly independent sources of motion exist and where they become significant along a beam line. Usually some unknown degrees of freedom will show up during a degrees-of-freedom analysis. The ability to reveal unexpected degrees of freedom in a beam line is one of the achievements of MIA. To understand and characterize the unknown degrees of freedom is obviously desirable but much harder to achieve. In general one can not resolve all physical base patterns by analyzing only the BPM data. Extra information is necessary. However, in this section, I describe a method developed for exploring the unknown degrees of freedom.

The basic idea is to separate one-by-one the various modes according to the sequence they show along a beam line. For example, suppose that in addition to the incoming betatron motion, there is a malfunctioning corrector located at one third of the way along the beam line and a large structure (e.g. rf cavity) misalignment at two thirds of the way along the beam line. A degrees-of-freedom analysis of such a beam line will show at least four significant modes: two betatron modes starting from the beginning, one mode starting at the one-third point, and another mode starting at the two-thirds point. Clearly, the physical basis will have these spatial characteristics. However, the eigenmodes obtained from a SVD analysis can mix all modes together and may not obviously reflect the above spatial characteristics. To improve upon the SVD results, one can use the first one third of the beam line to determine the temporal patterns of the two betatron modes, project out the betatron modes of the whole beam line, and then remove the betatron modes from the data.[7] SVD analysis of the residual will show nothing until one-third down the beam line, a mode starting after that, and the location where another new mode shows up. Based on the new degrees-of-freedom analysis, one can SVD analyze the section from the beginning to right before the next mode shows up. Since in this section there exists

[7] If the betatron amplitude is correlated say with the current, then some of the wake starting at the two-third point can mix into a betatron mode determined in this way. It would be advantageous to purposely dither a corrector at the beginning of the beam line. Then the mode projected out would be guaranteed to be pure betatron motion.

only one mode, which must be due to the malfunctioning corrector, the temporal vector of this mode may be used as the temporal pattern of the corrector variation. Such a procedure can be repeated down the beam line until the temporal patterns of all degrees of freedom are determined.

The following is a mathematical description of the above procedure. We start with the BPM data matrix B of a beam line (assume noisy BPMs have been taken out and the random noise has been reduced as described in Section 3.5). Let $B^{(l)}$ denotes the sub-matrix containing the first l columns (BPMs) of B.

1. Perform a degrees-of-freedom analysis of B to identify the location, say l_1, up to where (but not beyond) only the incoming betatron motions are significant.[8]

2. Do a SVD analysis of $B^{(l_1)} = U^{(l_1)} S^{(l_1)} V^{(l_1)}$ and take say the first two (the number of betatron modes) temporal vectors $u_1^{(l_1)}$ and $u_2^{(l_1)}$ as the temporal pattern of the incoming betatron motions.

3. Use $U_{1:2}^{(l_1)} \equiv \left[u_1^{(l_1)}, u_2^{(l_1)}\right]$ to extract the spatial betatron vectors f_1 and f_2 of the whole beam line as described in Section 3.7 by computing $[f_1, f_2] = \left(U_{1:2}^{(l_1)}\right)^+ B$, where the superscript $^+$ means the pseudo-inverse of a matrix. ($A^+ = (A^T A)^{-1} A^T$ if $A^T A$ is nonsingular; see Appendix A.3 for details.) Note that $f_{1,2}$ should be accurate up to l_1. After that, other modes could be mixed in if they are correlated with the betatron motions. See Eq.(3.19) and the discussion there.

4. Take the betatron motions out of all the BPMs by computing the remainder $B_{r_1} = B - \left[u_1^{(l_1)}, u_2^{(l_1)}\right] [f_1, f_2]^T$, where the subscript r_1 indicates the remainder of first iteration.

In the second iteration, apply the same steps to B_{r_1}: identify the location l_2 up to where only one new degree of freedom exist; find the temporal pattern $u_3^{(l_2)}$ which is associated with this degree of freedom; include this new temporal pattern into the U

[8] I assume the beginning section the beam line is clean and simple, i.e. no extra degrees-of-freedom except incoming betatron motions. In a linac, as in the SLC, this is usually true. In a ring, one can always choose the simplest section as the starting location for B.

matrix as $U_{1:3}^{(l_2)} \equiv \left[u_1^{(l_1)}, u_2^{(l_1)}, u_3^{(l_2)}\right]$ and get $[f_1, f_2, f_3] = \left(U_{1:3}^{(l_2)}\right)^+ B$; take these three modes out and get the remainder $B_{r_2} = B - \left[u_1^{(l_1)}, u_2^{(l_1)}, u_3^{(l_2)}\right][f_1, f_2, f_3]^T$. Continue such iteration until the end the beam line is reached.

Section 4.1 will provide an example of the above described analysis procedure. Here only general remarks will be presented. First of all, this procedure usually will generate a new decomposition of BPM data matrix, which is likely to be closer to the physical base decomposition than the orthogonal base decomposition. Note that the number of the new bases may be larger than the initial degrees of freedom, i.e. there may be near degenerate bases. The major improvement of this new decomposition is that it reflects spatial locations of various sources, which could be informative and important. However, this analysis becomes difficult when the sources are weak and close to each other, because in such cases, Step 2 will suffer from accuracy problems. The fundamental limitation of such analysis is mentioned in Step 3. The following paragraph will furhter illustrates this issue.

Let us partition the temporal pattern matrix into $Q = [Q_1, Q_2]$, where Q_1 is known. The corresponding spatial bases partition is $F = [F_1, F_2]$ where both F_1 and F_2 are partially known up to say location l. (In fact F_2 is zero up to l.) Let $\hat{F}_1 = Q_1^+ B$, the estimated vectors for F_1 as described in Section 3.7. The residual matrix after taking out variations due to Q_1 can be written as

$$\begin{aligned} B_r &\equiv QF^T - Q_1\hat{F}_1^T = Q_1F_1^T + Q_2F_2^T - Q_1\left(F_1^T + Q_1^+Q_2F_2^T\right) \\ &= \left[I - Q_1\left(Q_1^TQ_1\right)^{-1}Q_1^T\right]Q_2F_2^T. \end{aligned} \tag{3.26}$$

Therefore the residual contains only the part of $Q_2F_2^T$ that is perpendicular to Q_1, due to the projection operator inside [ ]. Note that knowing part of F_2 does not help.

3.9 Noise floor characteristics and BPM resolution spectrum

I claimed earlier that the singular values in the flat floor (as shown in Fig. 3.1) of a singular-value plot are due to BPM noise. In this section I will discuss characteristics of the noise floor and confirm that those singular values indeed behave like noise. First of all, without the noise term in Eq.(3.4), the rank of B will be d. Since physically we do not expect a large d, most of the singular values should be zero if not due to noise. More convincing evidence comes from the statistical characteristics of the noise floor. We can examine the data and see how the noise floor behaves when changing BPM number M and pulse number P.

Figure 3.6 shows the dependency on M. In the top frame the 4 curves represent the singular values of the same 5000 pulses but with 30, 60, 90, 120 BPMs respectively. The insertion shows the first 10 singular values. In addition to the general appearance of a noise floor (i.e. a long flat part and small tails at ends), the noise levels indeed decrease with increasing M. The bottom frame plots in circles the inverse square of the median values of the noise floor part of the above 4 curves vs. M. The solid line is a linear fit. Such $\frac{1}{\sqrt{M}}$ dependency indicates that the long tail in the singular-value plot is indeed due to random noise, since coherent signals will have roughly the same singular values as M increases. (As can be seen from the insertion, the singular values of signals may vary because the variation amplitude of a signal may not be uniform across the BPMs.) Note that, were there other distributed sources of random noise (such as dark current in accelerator structures) affecting the measurement, the noise floor would not decrease as $\frac{1}{\sqrt{M}}$.

Figure 3.7 shows the noise floor dependency on P. The first frame plots respectively singular values of 7 ensembles of 200, 400, 800, 1600, 3200, 5000, 10000 pulses, with $M = 120$ for all cases. It shows that the slope of the noise floor decreases with increasing P, while the average noise levels stay approximately the same. The second frame plots the variance of the eigenvalues (excluding the first 15) vs. P. The circles correspond to 25 sample sets randomly polled from 12,000 measured pulses and the solid curve is a least-squares fit to a second order polynomial in $\frac{1}{\sqrt{P}}$, which is the

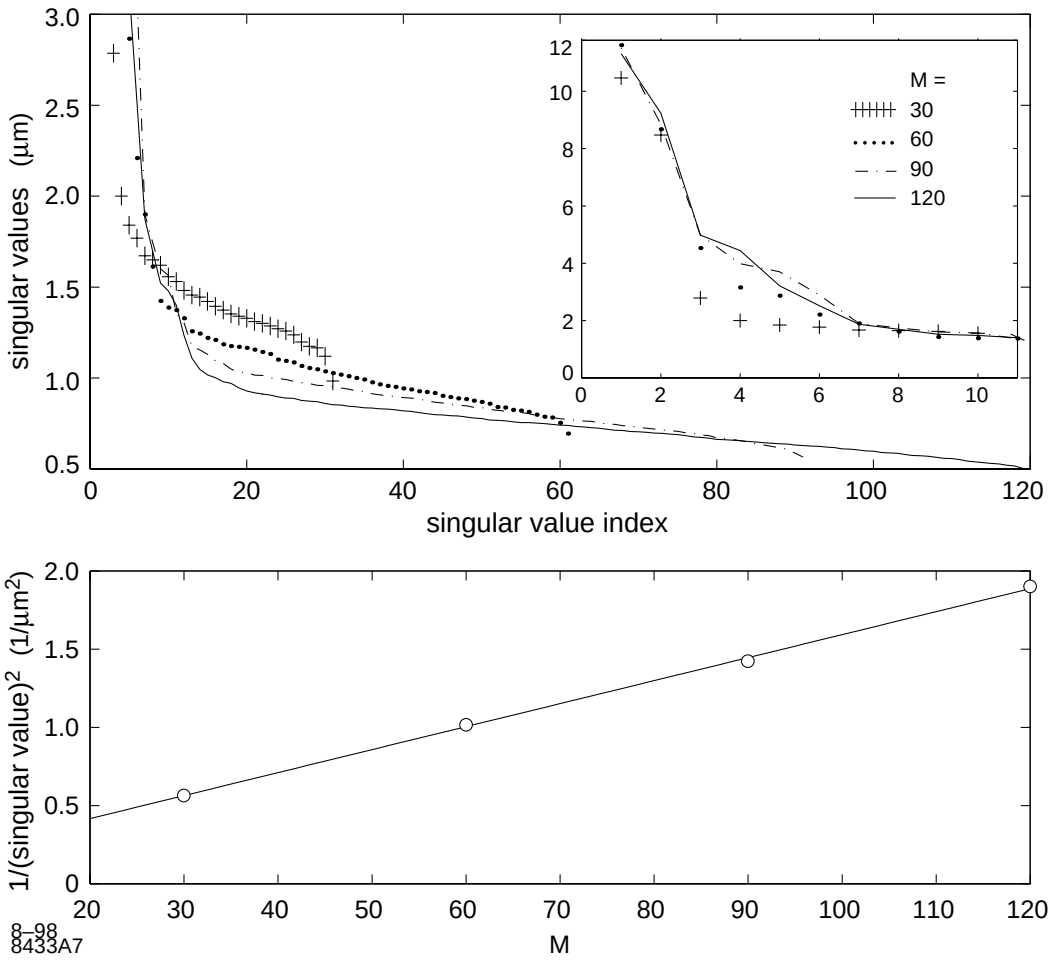

Figure 3.6: M dependency of noise floor

expected P dependency as explained in Section 3.10. We see that the measured data fit the statistical noise model very well even to $P = 10000$. This indicates that we can reasonably expect that if we were to project out a specific time pattern, as in Section 3.7, we would get a resolution improvement of 100 over the single-pulse BPM resolution. The slope reaches a limit as $P \to \infty$, which is taken to be the intrinsic resolution spread among the BPMs. Thus the noise floor of the singular-value plot reflects the BPM resolution spectrum, provided that the pulse ensemble is sufficiently large.

From the fixed point in the top frame, the average BPM resolution is about

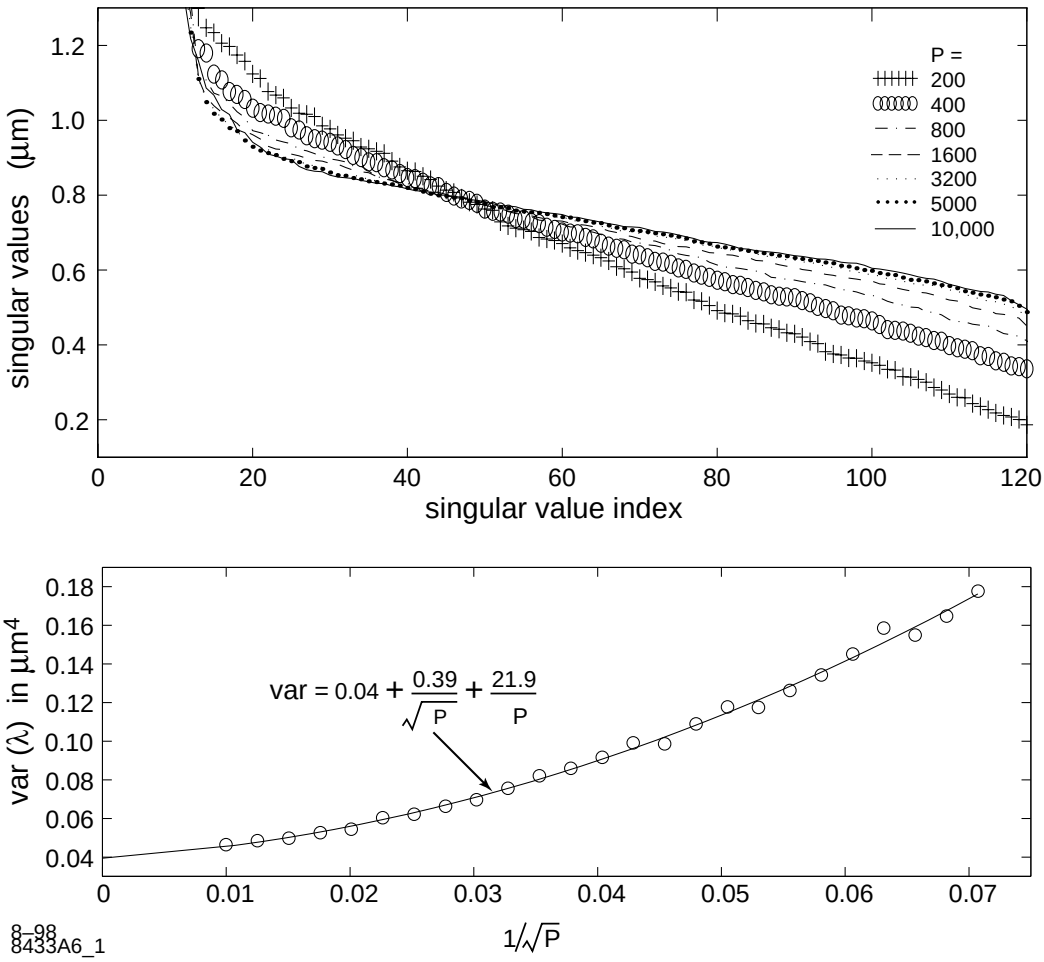

Figure 3.7: P dependency of noise floor

$\sqrt{120} * 0.8 = 9\mu m$. From the slope, the resolution spread is estimated at about 2 μm. To estimate the resolution of individual BPMs, one can use the method mentioned at the end of Section 3.5. The little tail at the very end of Fig. 3.1 arises from several high resolution BPMs at the beginning of the beam line. I deliberately removed those BPMs from the data in Figs. 3.6 and 3.7 for clarity.

Simulation studies of noise-only cases show the same M and P dependencies. Therefore the singular-value floor must be due to the BPM noise, though it is still possible, of course, that there are small coherent signals buried in this noise floor.

3.10 Understanding the P-dependency

In this section, I derive expressions for the mean and variance of the eigenvalue λ's of the variance-covariance matrix $C_B = B^T B$ in order to understand the P dependency of noise floor. At $P = \infty$ we have the decomposition $\hat{C}_B = \hat{V}\hat{\Lambda}\hat{V}^T$, where $\hat{}$ indicates the $P \to \infty$ quantities and $\hat{\Lambda} \equiv \mathrm{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_M)$. For a finite P, statistical noise will result in slightly different C_B, V, and eigenvalues. However $\hat{V}^T C_B \hat{V}$ should still be close to $\hat{\Lambda}$. Let

$$\hat{V}^T C_B \hat{V} = \hat{\Lambda} + \mathcal{E}, \tag{3.27}$$

where the symmetric matrix $\mathcal{E} \equiv \{\epsilon_{ij}\}$ represents the difference due to statistical noise. Taking the trace of both sides of Eq.(3.27) yields

$$\mathrm{tr}(C_B) = \sum_{i=1}^{M} \lambda_i = \sum_{i=1}^{M} \hat{\lambda}_i + \sum_{i=1}^{M} \epsilon_{ii} \tag{3.28}$$

Multiplying each side of Eq.(3.27) by itself and then taking the trace yields

$$\mathrm{tr}(C_B^2) = \sum_{i=1}^{M} \lambda_i^2 = \sum_{i=1}^{M} \hat{\lambda}_i^2 + 2\sum_{i=1}^{M} \hat{\lambda}_i \epsilon_{ii} + \sum_{i,j=1}^{M} \epsilon_{ij}^2 \tag{3.29}$$

Combining Eqs. (3.28) and (3.29) we have

$$\mathrm{var}(\lambda) = \mathrm{var}(\hat{\lambda}) + \frac{1}{M}\sum_{i,j=1}^{M} \epsilon_{ij}^2 + \frac{2}{M}\sum_{i=1}^{M} \Delta\hat{\lambda}_i\, \epsilon_{ii} - \left(\frac{1}{M}\sum_{i=1}^{M} \epsilon_{ii}\right)^2$$

where $\Delta\hat{\lambda} \equiv \hat{\lambda} - \langle\lambda\rangle$. The last term is obviously much smaller then the second term and can be dropped. Thus the difference in mean value and variance due to statistical noise can be expressed as

$$\begin{aligned} \langle\lambda\rangle - \langle\hat{\lambda}\rangle &= \frac{1}{M}\sum_{i=1}^{M} \epsilon_{ii} \\ \mathrm{var}(\lambda) - \mathrm{var}(\hat{\lambda}) &\simeq \frac{1}{M}\sum_{i,j=1}^{M} \epsilon_{ij}^2 + \frac{2}{M}\sum_{i=1}^{M} \Delta\hat{\lambda}_i\, \epsilon_{ii} \end{aligned} \tag{3.30}$$

Since both diagonal term ϵ_{ii} and off-diagonal term ϵ_{ij} have $\frac{1}{\sqrt{P}}$ dependency for large P, the differences vanish when $P \to \infty$ as they should. The mean value approaches the real value as $\frac{1}{\sqrt{P}}$, while the variance behavior is more complicated. For small P and $\Delta\hat{\lambda}$, the first term dominates and yields a $\frac{1}{P}$ dependency. As P becomes sufficiently large, the second term will become dominate and results in a $\frac{1}{\sqrt{P}}$ dependency, provided that the system is sufficiently stable. The summations in these expressions can greatly suppress the fluctuation due to statistical quantity ϵ's, thus we can expect a clear P dependency even for small number of pulses.

3.11 Kick analysis of base vectors

After the physical bases are obtained, most of the base vectors look like betatron oscillations because no matter what the physical sources, the resultant motion of the beam centroid is usually a sum of excited betatron oscillations. We can improve our understanding with a kick analysis. The goal is to identify the source kicks that generate the physical patterns. The basic idea has two ingredients: 1) kick representation and 2) removal of betatron response due to the lattice. The kick representation is just an equivalent representation of the same vector. Instead of giving the resultant motion, the kick representation simply shows the kicks which cause the motion. Since forces (which cause momentum changes) are much more likely to be localized along a beam line, the kick representation of a physical base usually has a simpler structure and reveals the location of the sources contributing to the motion.

There are many ways to accomplish kick analysis, including the simple method described here. Assume the betatron basis are given by u and v, and a physical pattern g is to be analyzed. For any 3 consecutive points (which form 3D vectors $\vec{u}$, $\vec{v}$ and $\vec{g}$), use the first 2 points to find a combination of the betatron basis that fits the first 2 points of g (which is always possible) and then predict what the 3rd point of g should be if it follows a betatron oscillation. This concept can be represented by the simple formula

$$\vec{g} = \alpha\vec{u} + \beta\vec{v} + \Delta z\,(0, 0, 1) \tag{3.31}$$

where α, β, and Δz are coefficients. A dot product of $\vec{u} \times \vec{v}$ to both sides yields the differences,

$$\Delta z = \frac{(\vec{u} \times \vec{v}) \cdot \vec{g}}{(\vec{u} \times \vec{v})_3}, \tag{3.32}$$

which are assigned to the corresponding BPMs as the result of a kick analysis of g. Δz gives the exact beam displacement due to potential kicks.

If kick angles are desired, one has to divide the above beam displacements by the R_{12} between the observed displacement and the location of the corresponding kick. However, such information (especially where the kick started between the two BPMs) is usually not available. A good approximation is to assume the kick is located at the previous BPM and use the distance between the BPMs as the R_{12}. Note that the above calculated displacements do not suffer such problems.

Eq.(3.32) looks cute but it is only applicable to the 3-point fittings. Sometimes it may be helpful to use more points to improve the accuracy of kick analysis. To accomplish this, one can determine the coefficients α and β first by least-squares fitting and then use them to compute the projected value for the last BPM. Assume k BPMs are used, the solution can be written as

$$\Delta z = g_k - (u_k, v_k) \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \\ \vdots & \vdots \\ u_{k-1} & v_{k-1} \end{pmatrix}^{+} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{k-1} \end{pmatrix} \tag{3.33}$$

where $^{+}$ means matrix pseudo-inversion (see Appendix A.3) to solve the least-squares fittings. When $k = 3$ this general formula is the same as Eq.(3.32), which is easy to confirm since the pseudo-inversion becomes ordinary matrix inversion. In general, using more BPMs for each subsection will improve fitting accuracy; however, when there are kicks close to each other, it may not be appropriate to use many BPMs.

Figure 3.8 shows two examples of a kick analysis in a simulation study, in which there are two 300 μm structure misalignments, 10% bunch length jitter, and 0.5° incoming beam (longitudinal) phase jitter (in addition to the betatron oscillations etc.). The top two plots are the two vectors corresponding to the bunch length

and incoming beam phase jittering respectively. The bottom two plots give the kick analysis of the two base vectors according to Eq.(3.32). Although the two base vectors look rather similar (like other betatron oscillations as well), the kick analysis yields completely different characteristics. The bunch length vector is clearly the result of two major localized wakefield kicks generated by the two structure misalignments. The kick analysis nicely shows the locations and strengths. The strength difference is due to the energy dependency of the wakefield kick. On the other hand, the incoming beam phase vector does not consist of any major kicks at all, because the wakefield kicks are not sensitive to the incoming beam phase change (the effect of beam energy change is rather weak). The apparent oscillation is due to the energy dependency of the betatron oscillation frequency. The wakefield kick merely launches the oscillation, which then grows with the increase of the accumulated betatron phase difference.

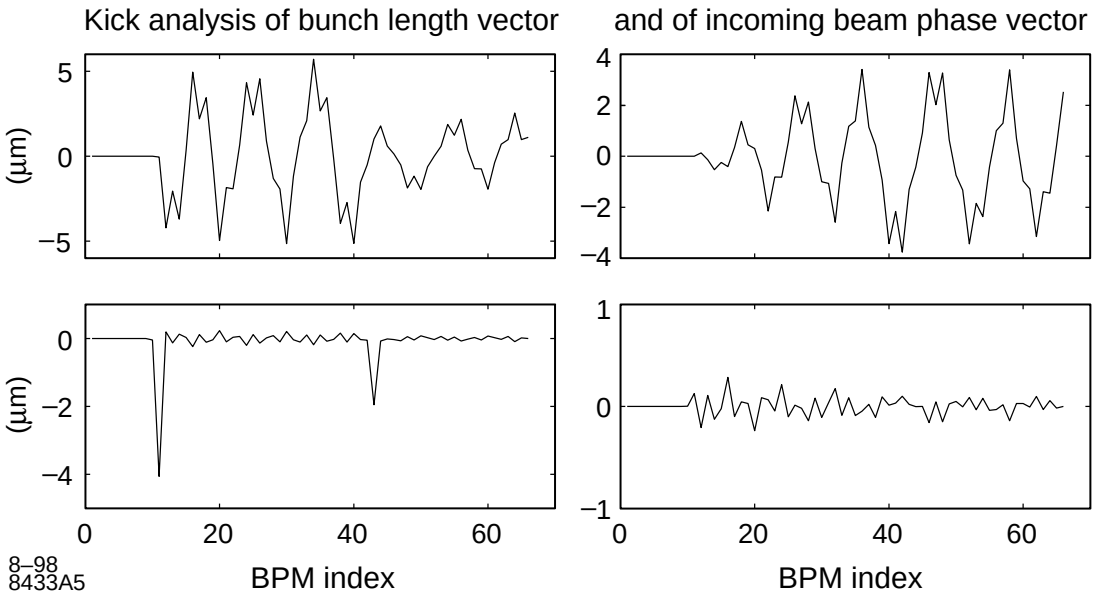

Figure 3.8: Examples of kick analysis of physical basis

3.12 Measurement and analysis of Wronskian determinant

This section discusses the determination and application of a Wronskian determinant to a beam line transverse dynamics. Since in a linac the transverse beam dynamics is affected by the longitudinal acceleration, I consider the transverse beam dynamics with/without acceleration. Let $\gamma(s)$ be a predefined energy profile along the linac, the basic transverse equation of motion for a single particle is governed by

$$\frac{d}{ds}\left(\gamma(s)\frac{d}{ds}y\right) + K(s)y = G(s). \tag{3.34}$$

If there is no acceleration, i.e. $\gamma(s) =$ constant, this equation reduces to the inhomogeneous Hill's equation $y'' + k(s)y = g(s)$ which has well known solutions.

From the general theory of ordinary differential equations, the second order homogeneous differential equation $y'' + \frac{\gamma'}{\gamma}y' + \frac{K(s)}{\gamma}y = 0$ has an important property that the Wronskian determinant can be written as

$$\begin{aligned} W(s) &\equiv \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = W(s_0)\exp\{-\int_{s_0}^{s}\frac{\gamma'}{\gamma}ds\} \\ &= W(s_0)\frac{\gamma(s_0)}{\gamma(s)} \end{aligned} \tag{3.35}$$

After obtaining the two betatron bases via MIA, one can compute the Wronskian as follows. Considering any two BPMs, from the betatron bases u and v, we have orbit values u_1, u_2 and v_1, v_2. Using the R matrix from the first to the second BPM,

we have[9]

$$W(s_1) = \begin{vmatrix} u_1 & v_1 \\ u_1' & v_1' \end{vmatrix} = \begin{vmatrix} u_1 & v_1 \\ \frac{1}{R_{12}}u_2 - \frac{R_{11}}{R_{12}}u_1 & \frac{1}{R_{12}}v_2 - \frac{R_{11}}{R_{12}}v_1 \end{vmatrix} = \frac{1}{R_{12}} \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix}, \quad (3.36)$$

thus

$$\begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} = R_{12}(s_1 \to s_2) W(s_1) \quad (3.37)$$

Note that this expression holds for any two independent betatron orbits, which are available in an MIA analysis. In fact, it appeared in the denominator of kick analysis expression Eq.(3.32). Another interesting factor that may come into this expression is the BPM gains g_1 and g_2. Taking this into account, we have

$$\begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} = g_1 g_2 R_{12}(s_1 \to s_2) W(s_1) \quad (3.38)$$

There are a few potential uses of this expression:

- If reliable information is available on the BPM calibration, beam energy profile, and the R_{12}, one can check if and where the dynamics of the system departs from Eq.(3.34) due to, for example, the wakefield effects.
- If the beam dynamics is conditioned to satisfy the Eq.(3.34), one can use Eq.(3.38) to determine the R_{12}, calibrate BPM gains, or measure energy profile from knowledge of the others. Note that although two gain factors appear, only one is actually free.

[9] In addition,

$$\begin{aligned} W(s_2) &= \begin{vmatrix} u_2 & v_2 \\ u_2' & v_2' \end{vmatrix} = \left| R(s_1 \to s_2) \begin{pmatrix} u_1 \\ u_1' \end{pmatrix}, R(s_1 \to s_2) \begin{pmatrix} v_1 \\ v_1' \end{pmatrix} \right| = |R(s_1 \to s_2)| \cdot \begin{vmatrix} u_1 & v_1 \\ u_1' & v_1' \end{vmatrix} \\ &= |R(s_1 \to s_2)| \cdot W(s_1) \end{aligned}$$

3.13 Betatron modes

This section provides an analytical SVD analysis of the betatron modes for 1-D linear optics in order to gain a more quantitative insight into the SVD analysis. This simple case can often explain the top two eigen-modes because the linear betatron motion usually dominates the beam dynamics in most accelerators, especially in rings. I will ignore BPM noise for simplicity. In such a system, the m-th BPM reading for the p-th pulse can be expressed in action-angle variables $\{J_p, \phi_p\}$ as

$$b_p^m = \sqrt{2J_p\beta_m}\cos(\phi_p + \psi_m) \tag{3.39}$$

where β_m is the beta function at the m-th BPM and ψ_m is the phase advance from a starting point to the m-th BPM. I assume the action J and angle ϕ are independent of each other, which is basically true in rings.

Without much computation, a useful expression for determining the beta function can be obtained as

$$\beta = \langle J\rangle^{-1}\left(\lambda_+ v_+^2 + \lambda_- v_-^2\right) \tag{3.40}$$

from Eqs. (3.39) and (3.14) via $\langle b^2\rangle = \langle 2J\beta\cos^2\phi\rangle = \langle J\rangle\beta = \lambda_+ v_+^2 + \lambda_- v_-^2$, where $\lambda_\pm$ and $v_\pm$ are the eigenvalues and eigenvectors of the two betatron modes. Although the betatron modes may change from one ensemble to another due to sample fluctuation, Eq.(3.40) should at least provide a good measurement of the machine beta function, provided that the BPM gains have been taken care of. Even if there are non-betatron modes, Eq.(3.40) could still yield a good estimate as long as the additional modes are weak.

Now let us go through the exercise of computing the betatron modes. For readers who do not wish to follow such an exercise, Eqs. (3.46)–(3.49) are the expressions for the two betatron modes. There could be simpler ways to get these expressions, however, the purpose of this exercise is to go through the general procedure analytically.

From the BPM data given in Eq.(3.39), the elements of the matrix $B^T B$ read

$$\begin{aligned}(B^T B)_{mn} &= \sum_{p=1}^{P} b_p^m b_p^n \\ &= \sum_{p=1}^{P} J_p \sqrt{\beta_m \beta_n}\,[\cos(\psi_m - \psi_n) + \cos(2\phi_p + \psi_m + \psi_n)] \\ &= \left(\sum_{p=1}^{P} J_p\right) \sqrt{\beta_m \beta_n}\cos(\psi_m - \psi_n). \end{aligned} \tag{3.41}$$

The last step uses the assumption that the action and angle variables are independent of each other and the angle variable ϕ is uniformly distributed over $[0, 2\pi]$. Thus, the variance-covariance matrix $C_B = \frac{B^T B}{P}$ can be written as

$$C_B = \langle J \rangle_p \begin{bmatrix} \beta_1 & \sqrt{\beta_1\beta_2}\cos\Delta\psi_{12} & \cdots & \sqrt{\beta_1\beta_M}\cos\Delta\psi_{1M} \\ \sqrt{\beta_2\beta_1}\cos\Delta\psi_{21} & \beta_2 & \cdots & \sqrt{\beta_2\beta_M}\cos\Delta\psi_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\beta_M\beta_1}\cos\Delta\psi_{M1} & \sqrt{\beta_M\beta_2}\cos\Delta\psi_{M2} & \cdots & \beta_M \end{bmatrix} \tag{3.42}$$

where $\langle J \rangle_p$ is the ensemble average of the action variable; $\Delta\psi_{mn}$ is the phase advance between the m-th and n-th BPMs.

To find eigenvalues and eigenvectors of C_B, we need to solve for ϕ_0 so that the corresponding orbit $v = \{\sqrt{2J\beta_m}\cos(\phi_0 + \psi_m), m = 1, \cdots, M\}$ satisfies the secular equation

$$C_B\, v = \lambda\, v \tag{3.43}$$

From the m-th component of the secular equation, we have

$$\begin{aligned} & \lambda\cos(\phi_0 + \psi_m) \qquad (3.44) \\ = \;& \langle J \rangle \sum_{n=1}^{M} \beta_n \cos(\psi_m - \psi_n)\cos(\phi_0 + \psi_n) \\ = \;& \langle J \rangle \left[\cos(\phi_0 + \psi_m)\sum_{n=1}^{M}\beta_n \cos^2(\phi_0 + \psi_n) + \frac{1}{2}\sin(\phi_0 + \psi_m)\sum_{n=1}^{M}\beta_n \sin 2(\phi_0 + \psi_n)\right] \end{aligned}$$

Therefore, we have the condition

$$\sum_{n=1}^{M} \beta_n \sin 2(\phi_0 + \psi_n) = 0. \tag{3.45}$$

There are two solutions,

$$\phi_0 = -\frac{1}{2} \arctan \left(\frac{\sum_n \beta_n \sin 2\psi_n}{\sum_n \beta_n \cos 2\psi_n} \right) \tag{3.46}$$

and $\phi_0 + \frac{\pi}{2}$, that lead to two different eigenvectors. The two eigenvalues are

$$\begin{aligned} \lambda_{\pm} &= \frac{1}{2}\langle J \rangle \left[\sum_{n=1}^{M} \beta_n \pm \sum_{n=1}^{M} \beta_n \cos 2(\phi_0 + \psi_n) \right] \\ &= \frac{1}{2}\langle J \rangle \left[\sum_{n=1}^{M} \beta_n \pm \sqrt{\left(\sum_n \beta_n \cos 2\psi_n \right)^2 + \left(\sum_n \beta_n \sin 2\psi_n \right)^2} \right] \end{aligned} \tag{3.47}$$

Note that the eigenvalues usually grow with the number of BPMs used. The corresponding normalized eigenvectors are

$$\begin{cases} v_+ = \frac{1}{\sqrt{\lambda_+}} \left\{ \sqrt{\langle J \rangle \beta_m} \cos(\phi_0 + \psi_m) \ , \ m = 1, \cdots, M \right\} \\ v_- = \frac{1}{\sqrt{\lambda_-}} \left\{ \sqrt{\langle J \rangle \beta_m} \sin(\phi_0 + \psi_m) \ , \ m = 1, \cdots, M \right\} \end{cases} \tag{3.48}$$

These two modes are obviously orthogonal according to Eq.(3.45). Furthermore, the beta function expression in Eq.(3.40) can be confirmed.

The normalized temporal vectors can be worked out as

$$\begin{cases} u_+ = \left\{ \sqrt{\frac{2J_p}{P\langle J \rangle}} \sin(\phi_p - \phi_0) \ , \ p = 1, \cdots, P \right\} \\ u_- = \left\{ \sqrt{\frac{2J_p}{P\langle J \rangle}} \cos(\phi_p - \phi_0) \ , \ p = 1, \cdots, P \right\} \end{cases} \tag{3.49}$$

which clearly relates to the often used normalized coordinates! Note that the orthogonality of these two vectors holds for $P \to \infty$. In other words, for finite P, these

expressions are only approximate. So is Eq. (3.41).

In summary, the SVD of the BPM data matrix is given by

$$\frac{1}{\sqrt{P}}B = \sigma_+ u_+ v_+^T + \sigma_- u_- v_-^T \tag{3.50}$$

where the singular values $\sigma_+ = \sqrt{\lambda_+}$ and $\sigma_- = \sqrt{\lambda_-}$; the spatial and temporal singular vectors are the column vectors given by Eqs. (3.48) and (3.49). Remarkably, the SVD yields the well-known normalized coordinates for this 1-D dynamical system.

3.14 Remarks

I have discussed the general theory of Model-Independent Analysis. My focus was on the spatial pattern analysis since no one has explored that before.[10] However, it is obvious that many well-established temporal (time-series) analysis techniques can be applied as well, both before and after the analysis described in this chapter. Combination of temporal and spatial analyses will extend the capability of MIA. For example, Fourier analysis can be applied to the raw BPM data in order to filter out noise at certain frequency, and/or to the resultant temporal pattern of a physical mode in order to study the frequency spectrum of that mode. Note that, it is usually necessary to collect a large amount of consecutive turn/pulse data for a large number of BPMs in order to perform both spatial and temporal pattern analysis. This is a demanding requirement for the data acquisition system.[10]

As statistical analysis methods, MIA can be applied to not only the BPM data but also various other types of data. For example, the beam position data from a button-type BPM are usually derived from the difference signals picked up at four buttons. However, the sum signal could provide the current information. Analyzing the current signals may locate where large beam loss occurs. It might be informative also to analyze the raw button signals. (However, note that, the difference signals are usually much cleaner than both the sum signals and the raw button signals.)

[10] Also, when I did the SLC studies, I could not collect a sufficient number of consecutive pulses, which is a practical reason that I did not pursue extensive temporal analyses.

Furthermore, although I have focused on the analysis of single particle dynamics, MIA can be used to study many other types of beam dynamics issues such as single (even multi) bunch collective effects, and beam-beam interactions. Of course, each MIA application may have unique features to be explored and understood. Lately researchers at KEK have started to study multi-bunch problems by using the bunch index as the BPM index and the longitudinal bunch displacements as the BPM data. Some interesting results have been obtained.[15]

Chapter 4

MIA applications to linacs—SLC example

In this chapter, I describe several MIA experiments carried out on the Stanford Linear Collider (SLC)[1] and some related simulation studies. Although the experimental results are preliminary in both quantity and quality due to the limited machine time available, they are very promising and most of them are unprecedented. The experimental results presented in this chapter are more a proof-of-principle for MIA than a systematic study of SLC.[45, 46]

4.1 Simulation studies

Many simulation studies were carried out during the development of MIA methods described in the last chapter. Some of them have been discussed in previous sections. In this section I present two more sets of simulation results obtained for SLC. One summarizes the effort to evaluate the significance of various physical effects (e.g. incoming beam position jitter, current jitter, structure misalignments) on the beam centroid dynamics. The other is an MIA done for a simulated data set based on

[1] Only the first one-third of the linac was used in my study because at that time the data acquisition system limited the number of BPMs one can use for buffered data acquisition—collecting large number of consecutive beam pulses simultaneously at selected BPMs.

the SLC linac configuration. The computer program[16] used in the simulations calculates the first order optics as well as the effects of the linac longitudinal and dipole wakefields.

To estimate the magnitude of various physical effects on the beam orbit f, the first and second order orbit deviations ($\partial_q f$ and $\partial_{q_1}\partial_{q_2} f$) are computed via the simplest difference equations

$$\begin{aligned} f_q &= \frac{1}{2}\left[f(\Delta q) - f(-\Delta q)\right] \\ f_{q_1 q_2} &= \frac{1}{4}\left[f(\Delta q_1, \Delta q_2) - f(-\Delta q_1, \Delta q_2) - f(\Delta q_1, -\Delta q_2) + f(-\Delta q_1, -\Delta q_2)\right] \end{aligned}$$

which are easy to carry out and sufficiently accurate because the perturbations are small. The perturbed beam orbits are computed numerically and these difference equations are used to compute the associated spatial patterns, which (except for those related to the incoming phase ϕ_{inj} and energy δ variations) are shown as solid lines in Fig. 4.3. (This figure is mainly for another purpose, but can be used here.) These patterns correspond to incoming beam position and angle variations $\Delta x_0 = 15\mu m$ and $\Delta x_0' = 3.3\mu rad$; 10% beam current jitter ΔI; 10% bunch length jitter $\Delta\sigma_z$; 0.5° incoming longitudinal phase jitter $\Delta\phi_{\text{inj}}$, etc. which are close to the size of the fluctuations observed in the SLC linac. The lattice used was the standard linac lattice except for two 300 μm structure misalignments at BPM 70 and 100. Table 4.1 gives the rms amplitude of the corresponding physical patterns as obtained by the computer program. The important messages from Table 4.1 are: 1) except for the betatron oscillations, all other orbit perturbations are on the order of a few μm—well below the nominal 10 μm BPM resolutions (this was a major factor that motivated MIA development); 2) the second order perturbations are also significant.

In order to understand the MIA results of the SLC, and in particular, to study the effect of structure misalignments, a system was generated to simulate certain aspects of the SLC experiments. In addition to the standard linac lattice model, the system consists of:

- two 300 μm structure misalignments at BPM 70 and 100,

- initial betatron motions of $\Delta x = 15\mu m$, $\Delta x' = 3.3\mu rad$,
- a 10% current jitter (nominal number of particles per bunch $N = 4 \times 10^{10}$),
- a 10% bunch length jitter (nominal bunch length $\sigma_z = 1.2$ mm),
- 50% correlations among x-x', x-I, I-σ_z, and
- a total of 125 BPMs with resolutions 10 μm except for the first seven 1 μm BPMs.

Using the computer program, a total of 5000 pulses are randomly generated and tracked along the linac through all the 125 BPMs.

Fig. 4.1 shows the degrees-of-freedom analysis of the simulated SLC linac system according to the method described in Section 3.6. As marked in the figure, the degrees-of-freedom analysis yields the major characteristics of the system such as the 1 and 10 μm noise floors; the continuous growing betatron modes; the location of the first structure misalignment; and the degrees of freedom due to wakefields.

The top two curves are dominated by the betatron motion and thus are marked as betatron modes. The apparent linear growth is due to 1) the increase of the beta-function along the beam line and 2) a significant mixing with other modes because of the 50% correlations between betatron motion and current variation.

The modes due to wakefields are particularly interesting. There is a weak signal arising well before the first structure misalignment and becoming significant after BPM 40. It is a mode due to 2nd order perturbations such as $\partial^2_{xI}\, b(m)$, i.e. the change of betatron motion because of current jitter. Such a mode requires both betatron and current variations: the betatron oscillation causes the beam to move off the centers

Sources	Δx_0	$\Delta x'_0$	ΔI	$\Delta\sigma_z$	$\Delta\phi_{\rm inj}$	$\Delta\delta_0$	$\Delta x_0\Delta I$, $\Delta x'_0\Delta I$	$\Delta x_0\Delta\sigma_z$, $\Delta x'_0\Delta\sigma_z$	$\Delta x_0\Delta\phi_{\rm inj}$, $\Delta x'_0\Delta\phi_{\rm inj}$
Effects	21.4	12.5	2.5	3.4	1.7	< 1	1.7	3.8	2.5

Table 4.1: Estimate of beam orbit variation magnitude (in μm) due to various physical perturbations. Done for the first one-third of SLC linac with two 300 μm structure misalignments.

of acceleration structures and thus generate wakefields, and the current jitter leads to an independent variation of the beam orbits due to the wakefields. At BPM 70, a new degree of freedom clearly appears, which is a betatron oscillation excited by the wakefield kicks of the 300 μm structure misalignment—a 1st order perturbation due to $\partial_I b(m)$. Note that there are only one degree of freedom associated with the term $\partial_I b(m)$ no matter how many structures are misaligned. However, as we will see later, individual structure misalignment can be revealed by a kick analysis of the vector $\partial_I b(m)$.

Fig. 4.2 plots the first 7 of the 125 spatial vectors of an SVD analysis of the simulated system. The singular values are given in μm on the vertical labels. There are 5 significant spatial patterns above the noise. However, unlike the degrees-of-freedom plot, these orthogonal spatial vectors do not show much of the spatial characteristics of the system, although the 6th and 7th vectors indicate much lower noise for the first 7 BPMs.

Fig. 4.3 gives the vectors extracted via all the temporal patterns (x, x', I, σ_z and their second order combinations) by using the methods described in Section 3.7. The corresponding temporal-pattern labels are shown on the left-hand-side of each spatial pattern. The solid lines are the exact spatial patterns and the dots are the extracted patterns from the simulated BPM data. The two agree well. Frames such as f_{x^2}, $f_{xx'}$, and $f_{\sigma_z^2}$ indicate the fitting errors, which are sub-micron levels. These physical bases, much more meaningful than the orthogonal basis in Fig. 4.2, clearly show the characteristics of the system and the underlying physics. In fact, some of the interpretation of the above degrees-of-freedom plot are based on the understanding of these physical bases. The physics has been discussed above. Note that the vectors depending on current and those depending on bunch length are very similar and therefore may not be well-separated in a degrees-of-freedom analysis. This is the reason that there are only 5 significant spatial patterns in Fig. 4.2.

Figs. 4.2 and 4.3 show the orthogonal and physical basis. Although the orthogonal basis is much less meaningful than the physical basis, to obtain it, one requires only the BPM data. Section 3.8 described a method to approach the physical base decomposition with BPM-data only. The idea is to require that the spatial bases reflect

the information obtained from the degrees-of-freedom analysis. Fig. 4.4 illustrates the process of applying such an idea to the simulated SLC data. The 1st sub-plot shows a degrees-of-freedom analysis of the whole beam line, which is the same as Fig. 4.1. The shaded area is up to the 7-th BPM, where only two betatron modes exist. An SVD of this subset of the BPM-data matrix was performed and the first two temporal singular vectors u_1 and u_2 were used as the betatron temporal patterns. (In fact, BPMs up to the 30-th may be used to determine the betatron temporal patterns since nothing shows up before that. However, the first 7 BPMs yield a better signal-to-noise ratio because of their high resolutions.) The residual matrix with the "betatron" motion removed can be computed via the procedure in Section 3.8. The 2nd sub-plot shows the residual degrees of freedom. The shaded area is up to BPM 45, where the leading mode is much stronger than the rest. This boundary is somewhat arbitrary. The criterion for the boundary was to contain only one degree of freedom (if there are two or more degrees of freedom that can not be separated, one has to include all of them) and use as many BPMs as possible to enhance the separation between the dominant degree of freedom and the next one. This subset (the first 45 BPMs) of the 1st residual BPM-data matrix is SVD analyzed and the first temporal singular vector is used as another temporal pattern u_3 for the whole beam line. At this stage, three temporal patterns have been chosen. Removing these 3 modes we obtain the 2nd residual matrix, whose degrees-of-freedom analysis is shown in the 3rd sub-plot, from which it is obvious that the first 70 BPMs (the shaded area) should be chosen. Again, an SVD yields another temporal pattern u_4. The next residual degrees of freedom are plotted in the last sub-plot, where only one mode left. An SVD of the whole residual matrix gives the last temporal pattern u_5.

Using these 5 temporal patterns, a new decomposition of the BPM-data matrix can be obtained. The corresponding spatial patterns are shown in Fig. 4.5. The label f_{u_i} means that the spatial pattern is corresponding to the temporal pattern u_i mentioned above. Comparing to the orthogonal basis shown in Fig. 4.2, these new spatial patterns are much closer to the physical basis, although still only the BPM data have been used. For example, the pattern f_{u_5} in Fig. 4.5 has the same characteristics as the current and bunch-length patterns (f_I and f_{σ_z}) in Fig. 4.3, but

none of the orthogonal patterns in Fig. 4.2 does. Moreover, the patterns f_{u_3} and f_{u_4} in Fig. 4.5 resemble the 2nd order physical patterns in Fig. 4.3.

However, due to the lack of information, the spatial patterns (f_u's) may not be the same as the physical patterns. The dots in the first two sub-plots of Fig. 4.5 are the results of fitting the spatial patterns to exact betatron orbits (by using the first 7 BPMs to find the best fit). As expected, they are very close. The errors are shown in Fig. 4.6. We see a small amount of mixing with other modes, because the betatron and the current variations (x-I) are correlated in the simulated system.

A kick analysis of these patterns is plotted in Fig. 4.7, in which the first two spatial patterns in Fig. 4.5 are used as the betatron bases. (They are not the exact betatron bases but very close to the exact ones and can be obtained from the BPM data alone.) The first two sub-plots of Fig. 4.7 are zeros because their spatial patterns are used as the betatron bases for the kick analysis. The kick analysis of f_{u_3} and f_{u_4} shows basically small random kicks. Some of them may be wakefield kicks and others are due to noise. In any case, they are small and there are no significant kicks.[2] The kick analysis of the spatial pattern f_{u_5} is most interesting. It shows clearly the two wakefield kicks due to the two 300 μm structure misalignments (via current and bunch-length variations). Since the beam energy is higher at the position of the 2nd structure misalignment, the effect of the misalignment is weaker than the first one. In a real situation, one may need further experiments to identify the physical sources of the measured spatial and kick patterns.

The above analysis demonstrates many features of MIA. We see that MIA can tell a great deal about an accelerator system from BPM data alone. It is even better when the temporal patterns of physical variations are available and used in combination with the BPM data. The ability to reveal the two 300 μm structure misalignments from BPM data alone is particularly remarkable.

The decomposition shown in Fig. 4.5 is based on analysis of spatial characteristics of the system. In case the temporal patterns of the system have certain known characteristics, one can take them into account and further improve the analysis via

[2]The first kick in f_{u_3} seems significant but there is no physical reason for it. This is because the denominator of Eq. (3.32) became very small and thus noise was enhanced significantly. Using Eq. (3.33) with more BPMs will reduce such problems.

Fourier analysis and so on. For example, one may perform degrees-of-freedom analysis with increasing number of pulses (instead of BPMs as shown before) and see whether new degrees of freedom appear. The point is that one should incorporate as much information as possible in MIA.

4.2 Measurement of BPM resolutions

In Chapter 3 I discussed the characteristics of the noise floor, focusing on the general properties of the singular value spectrum and noise reduction, although I did mention that the noise floor usually approaches the BPM resolution distribution and therefore provides a good measure of the BPM system. This section shows a few more experimental results from the SLC, focusing on the measurement of BPM performance.

As shown in Fig. 3.2, particularly noisy BPMs can be identified in the spatial singular vectors of an SVD analysis. The corresponding singular values (times $\sqrt{M}$) yield the resolutions of those BPMs. Usually such noisy BPMs are removed from the BPM data before further MIA analysis.

From the noise floor of the singular value spectrum in Fig. 3.1, the average and spread of BPM resolutions can be estimated, which characterize the overall performance of the BPM system for beam dynamics observations. Remember that Figs. 3.2 and 3.1 are experimental results from the SLC and represent the actual performance of SLC BPM system.

The noise floor characterizes the BPM system as a whole. Usually there is no correspondence between a noise floor singular value and a BPM resolution. If the resolution of an individual BPM is needed, one can set the signal singular values to zeros and re-multiply the SVD matrices as in the noise-cut procedure. The resultant BPM-data matrix is mainly the random noise due to the BPM resolutions. Thus computing the standard deviation of each BPM column yields an estimate of individual BPM resolution. The top frame of Fig. 4.8 plots the result of such an exercise for SLC data. There are two kinds of dots representing the BPM resolutions estimated from two different data sets. The results repeat rather well. Around BPM 50 and 95 there are three BPMs that have very small values. Those are the very noisy BPMs

whose variations have been removed when the large singular values are zeroed. Thus they do not represent those BPM resolutions. The two BPMs marked by the dash-lines yield completely different results for the two sets of data because they are at the boundary of the noise floors. Thus it happened that they were removed from the noise floor in one data set but not from the other. Such BPMs can be identified from the spatial singular vectors.

To compare the BPM resolutions with the singular value noise floor, BPM resolutions are reordered into decreasing sequence and plotted as solid dots in the bottom frame of Fig. 4.8. The circles are the singular values (not normalized by $\sqrt{M}$). As expected, the noise floor agrees with the BPM-resolution distribution fairly well except for the beginning part. Due to the existence of signals, instead of being compared with the singular value right above, the largest BPM resolution should be compared with the beginning value of the noise floor as indicated by the dash-line. (The light dots show the effect of moving the BPM-resolution distribution to match the noise floor.) Obviously, small coherent signals may cause the difference between the BPM resolutions and the noise floor. The beginning part of the noise floor tends to be larger than the actual BPM resolutions while the end part tends to be smaller due to the very nature of SVD—the larger singular values claim all linearly correlated variations even though the correlations are accidental.

The major errors in the BPM resolution estimate usually occur from the boundary area where signals and noise are hardly distinguishable. One may speculate on some improvement by using certain window functions (as is usually done in a Fourier analysis) that provide smoother weighting transitions between signals and noise. However, more research is needed on this subject.

4.3 Measurement of betatron bases and corrector jitter effects

In both high energy linacs and rings, the transverse beam dynamics is usually dominated by the betatron motions, which are characterized by the physical patterns that

I call "betatron bases". Any betatron orbit is a linear combination of the 2 betatron bases. Note that betatron bases may not be single-particle trajectories, especially when the beam current is high; instead they are the linear part of the beam orbit responses to initial transverse position and angle variations (i.e. $\partial_x \vec{b}$ and $\partial_{x'} \vec{b}$ at nominal beam conditions). In order to find the single-particle betatron bases, the beam current has to be sufficient low.

A good measurement of the betatron bases (especially the single-particle ones) is important for the understanding of beam optics. It is even more important for studying various subtle beam dynamics effects since the betatron oscillations often have to be removed from the data in order to observe the weak signals buried in the oscillation data. There are two major difficulties in an accurate measurement of the betatron bases: 1) insufficient BPM resolutions, and 2) separating betatron motions from other perturbations. MIA is meant to overcome such difficulties.

There are three methods to obtain the betatron bases with improving accuracy. The first method, the roughest one, is using the spatial vectors of the leading singular values as the betatron bases. Since the betatron oscillations dominate the motions, the singular vectors may provide a good estimate of the betatron bases. However, other spatial patterns will certainly be mixed into the "betatron patterns".

The second method is to determine the temporal patterns of the initial transverse variables such as x_0 and x_0' by using a clean linear section of the beam line, and then project out the corresponding betatron patterns via Eq. (3.18). This method is usually better than the singular vectors. However, since the measured temporal patterns may contain variations due to other perturbations, the obtained betatron patterns may still be mixed with other modes. For example, if the variations of the initial beam position x_0 is influenced by the variations of beam bunch length, then its temporal pattern is correlated with the bunch-length variations, which will mix the bunch-length vector into the betatron bases IF such correlations are not taken into account.

The first two method observe the beam passively and both are likely to suffer from having non-betatron motions mixed into the betatron bases. The third method, which is the most reliable one, excites independent betatron oscillations with weak

corrector dithering (strength modulations) and then extracts the spatial betatron response pattern by using the temporal dither pattern and Eq. (3.18). When carefully delivered, such a dither pattern will be uncorrelated with any other variations of the system, and therefore provides a betatron base without any mixing. Note that the dithering correctors must be carefully chosen in order to generate completely uncorrelated perturbations. For example, if the corrector location has non-zero dispersion, then the effective betatron perturbations could be correlated with the beam energy oscillations.

Fig. 4.9 demonstrates the determination of a betatron pattern of a section of SLC linac by using the 1st and 3rd methods above. Firstly, a large betatron oscillation ($\sim 200\,\mu m$ rms amplitude) was excited by dithering a corrector in the linac with a given temporal pattern. 8000 pulses at 125 BPMs were collected in about 20 minutes. (Due to the limitation of the SLC data acquisition system, these pulses were collected as 40 blocks of 200 consecutive pulses.) SVD of the BPM-data matrix yields a dominate mode whose spatial and temporal vectors have been shown in Fig. 3.3. The temporal pattern is the same as the corrector-dither pattern. The spatial pattern gives the betatron response pattern to the corrector kick. This spatial pattern is plotted as a solid line (the breaks are due to removal of dead and very-noisy BPMs) in each sub-plot of Fig. 4.9 in order to compare with other measurement results. The amplitude is arbitrarily scaled. Subsequently, the corrector dither strength was reduced to $\frac{1}{4}$, $\frac{1}{20}$, and $\frac{1}{50}$ of the original value and 8000 pulses were collected for each strength. Using the corrector dither pattern alone, the betatron patterns for each data set are extracted via Eq. (3.18) and plotted as dots in Fig. 4.9. For comparison, their amplitudes are scaled accordingly. We see in the bottom frame that even when the perturbation magnitude is reduced to 4 μm (200/50), the extracted betatron pattern is still very accurate. In the weak perturbation case, SVD vectors show nothing that resemble the particular betatron pattern. (See Fig. 4.12 in Section 4.5 for the first 7 vectors.) Nonetheless, as shown in Fig. 4.10, a degrees-of-freedom analysis does yields some clue about the existence of a new degree of freedom (marked by the gray dots) starting right after the dithered corrector. The strips in the figure are due to dead and noisy BPMs.

4.4 Measurement of transverse wakefield effect

A beam, when passing through a misaligned accelerator structure, will excite transverse wakefields that kick different parts of the beam by different amounts and therefore blow up the beam emittance. Such a wakefield effect can be a major source of luminosity degradation in a linear collider.[57] To counteract such effects in the SLC linac, various methods [58, 31] have been used for the detection and correction of such wakefield effects. However, it is very difficult to measure such transverse wakefield effects due to the weakness of the signals and the difficulty of separating such effects from other sources of perturbation. The current (or bunch length) dependency of the wakefield effects can be used to separate the wakefield effects from other effects. One idea is to measure the orbit at different beam currents and then calculate the difference. Unfortunately, such measurements are unlikely to succeed. One reason is the limited resolution in the orbit measurements. More importantly, when the beam current is changed, many other beam parameters (as well as the orbit) will be changed also. Similarly, bunch length changes have been used, but with limited success. To illustrate this problem, Table 4.2 shows the correlations between current (bunch length) change and the beam parameters that we are able to monitor in the linac of SLC. The data are based on 5000 electron pulses collected under normal running conditions. Clearly, such correlations have to be taken into account in order to measure the wakefield effects correctly.

MIA provides a novel approach to solve these problems because (1) it allows one

	current ΔI	bunch length σ_z
horizontal position x	-0.17	-0.52
horizontal angle x'	-0.03	-0.15
vertical position y	-0.08	0.01
vertical angle y'	0.10	0.20
long. beam phase	-0.48	-0.20
beam energy	-0.37	-0.18
bunch length	-0.05	1
beam current	1	-0.05

Table 4.2: Correlation coefficients of beam current and bunch length with other beam parameters measured in the SLC

to go below the resolution limit set by the individual BPMs and thus observe the beam dynamics at a much finer level, and (2) it takes all known signal correlations into account to generate a physical base decomposition. MIA was applied to the measurement of the transverse wakefield effect at the SLC and encouraging results were obtained.

At the SLC, in addition to the beam transverse position, beam current, bunch length, incoming beam (longitudinal) phase, and relative beam energy can be monitored on a pulse-by-pulse basis. Other signals such as klystron phases along the linac have not been used in the present analysis. As shown in Table 4.2, there are significant correlations among these signals, especially for the wakefield sensitive variables. MIA takes all known correlations into account, and therefore should provide a better measurement of the wakefield effects. To investigate this, a 5 corrector, 1.2mm, local bump in the linac of SLC was generated and its wakefield effect was measured via MIA. During the experiment, the linac was running under normal operating conditions except for the absence of the positron beam. BPMs from the beginning to about the 1/3 point of the linac (LI02–LI13) were used. Three sets of 5000 pulses were collected under the conditions: (a) before the bump was applied, (b) while it was applied, and (c) after the bump was removed. Each set of data took a few minutes to collect.

MIA was applied to all sets of data and then the vectors corresponding to the current jitter (i.e. the "current vectors") were compared. The results in the vertical measurement are shown in Fig. 4.11 (all ordinate units are in μm). The top frame shows the current vectors of conditions a and c, which are the wakefield effects due to misalignments (and corrector offsets, etc.) in the normal running machine. To test that a current vector indeed corresponds to the wakefield effects, the next two frames show the measurement of wakefield effects due to a corrector bump. The middle frame plots the differences of the averaged beam orbits. The solid line is the difference between conditions b and a, while the dots are between conditions b and c. The two curves are the same, which indicates that the beam orbit is under control. The corrector bump is clearly visible (although it is not well closed). The bottom frame shows the difference of current vectors. The dots are the difference between b

and a, while the crosses are between b and c. The solid curve shows the calculated (NOT just a fit!) wakefield effect due to the bump. We see that the agreements are fairly good, especially when considering the fact that the BPM resolution is about 10 μm, which is as large as the signals. Furthermore, no external beam perturbation was used, and the signal is based on a rather weak 1.3% natural current jitter. As far as I know, such an accurate transverse wakefield effect measurement in a linac are unprecedented. Note that the errors are on the order of a few microns, which is much larger than the statistical limit. Therefore, it is still possible to further improve both measurement and analysis. This test demonstrates that the current vectors obtained via Eq.(3.17) are correct. An immediate application of such current vectors is the detection of structure misalignments and confirmation of wakefield calculations.

A similar analysis in x was performed. However, one set of results agrees with the calculation while another has much larger deviations. The discrepancy in the horizontal cases may be due to some unknown jitter sources that are correlated with the current. These experiments were repeated several weeks later and similar results were obtained in both the vertical and horizontal planes. Unfortunately, the limited machine time prevented us from a thorough investigation, especially of the horizontal results.

The measurement results are still preliminary; nonetheless, they are very encouraging. Since such measurements need not perturb the beam, they are non-invasive to normal machine operation and in principle can be done parasitically and quickly after the method matures. On the other hand, intentionally introduced larger current variation will improve the sensitivity to the misalignments. This method can potentially become a powerful tool for finding structure misalignments.

4.5 Repeatability/stability of SLC measurements

Repeatability of MIA results for SLC measurements was checked even though there was insufficient machine time for a systematic study. Repeatability reflects the stability of the machine and provides a basic check for MIA methods. Let me summarize the previously shown experimental results before presenting a couple more.

- Fig. 3.7 shows the repeatability of the average levels of noise floors.
- Fig. 4.8 shows the repeatability of BPM resolutions measurements.
- Fig. 4.9 shows the repeatability of a corrector dither pattern measurements.
- Fig. 4.11 shows the repeatability of transverse-wakefield-effect measurements.

As discussed in Section 3.4, for a stable system, the singular values and the spatial singular vectors should be repeatable for different BPM data sets. However, the SLC linac is such a complicated system that various changes frequently occur. For example, klystrons phases are constantly changing and some klystrons might even drop off from time to time; feedback-corrector strengths are constantly changed to stabilize the beam orbit. Therefore, the singular values and vectors may or may not be stable. To check this, two BPM-data sets of 5000 pulses each are used to evaluate the repeatability of the singular values and vectors. These two data sets are the first and last data sets collected over about an hour period. Each data set took a few minutes to collect. Fig. 4.12 shows the first 7 spatial vectors of both data sets. The solid lines are for the first one and the dots for the last one. The singular values are shown accordingly on the left-hand-side labels.

From Fig. 4.12 we see that the singular values and vectors (of signals, not those of noise) are generally repeatable as expected. It is remarkable that there are signals stable to the μm level! Thus the dynamics of the linac may not (as often thought) be overwhelmed by noise (fluctuations) at this level. Sometimes, changes could occur even for the major patterns, which indicates that the system was varying. The most likely cause is the variation of the sample distributions of the initial beam conditions (i.e. $Q^T Q$). Of course, the linac configurations/parameters (i.e. the physical basis F) could also vary during data collection. However, even when there are large perturbation sources causing the leading spatial patterns to change, there are still repeatable patterns at the micron level.

4.6 Existence of unknown degrees of freedom

From the MIA of SLC data, we established the experimental fact that there are unknown degrees of freedom in the SLC linac that are significant for the beam dynamics. The most solid evidence is the degrees-of-freedom analysis such as Fig. 4.10. Clearly there are at least 8 degrees of freedom in this (first one third) linac section during the time the data was taken. Some of them we know (e.g. two due to betatron motion, one due to the weak corrector dither described at the end of Section 4.3), but most of them we do not know. At a different time, the degrees-of-freedom plot may change considerably (see Fig. 3.5 for another plot); however, there are always degrees-of-freedom whose physical sources are not clear.

To actually figure out all the degrees of freedom means that we understand the beam dynamics and the machine to the micron level (for SLC linac)! MIA provides a powerful tool to face such a challenge, yet to accomplish this difficult task requires lots of hard work and investment in machine time to investigate the perturbation sources. Unfortunately I did not have sufficient machine time to systematically investigate the SLC linac and understand its dynamics completely.

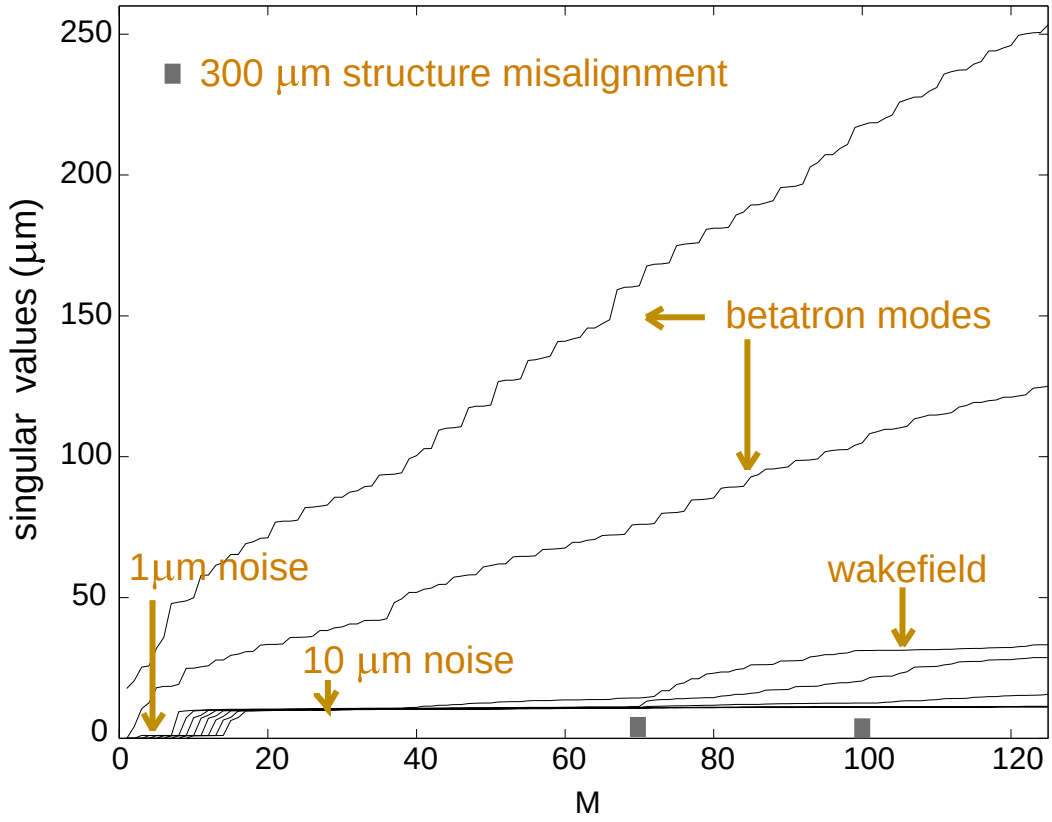

Figure 4.1: Degrees-of-freedom analysis of a simulated SLC linac system with two structure misalignments.

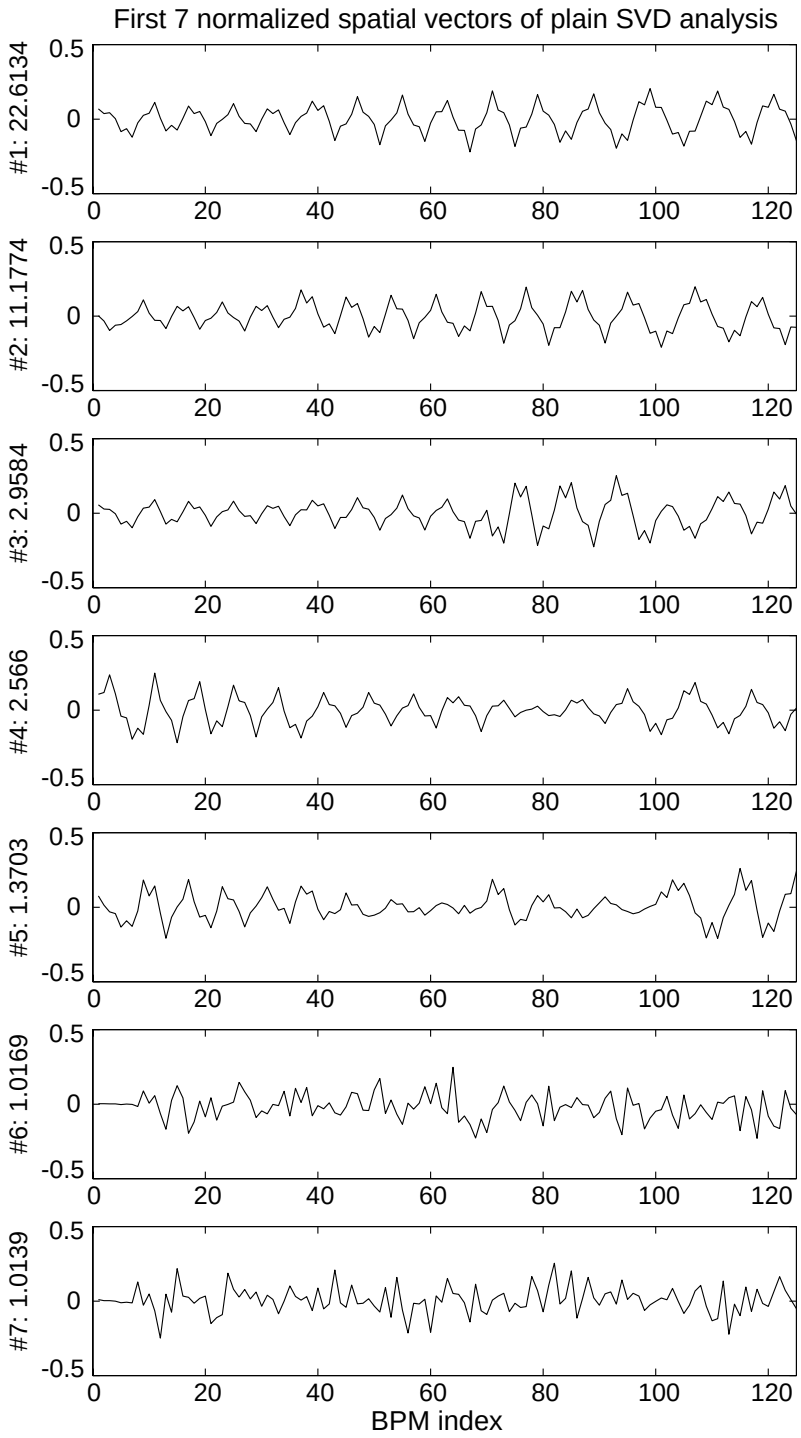

Figure 4.2: Spatial vectors of SVD analysis of a simulated SLC linac system.

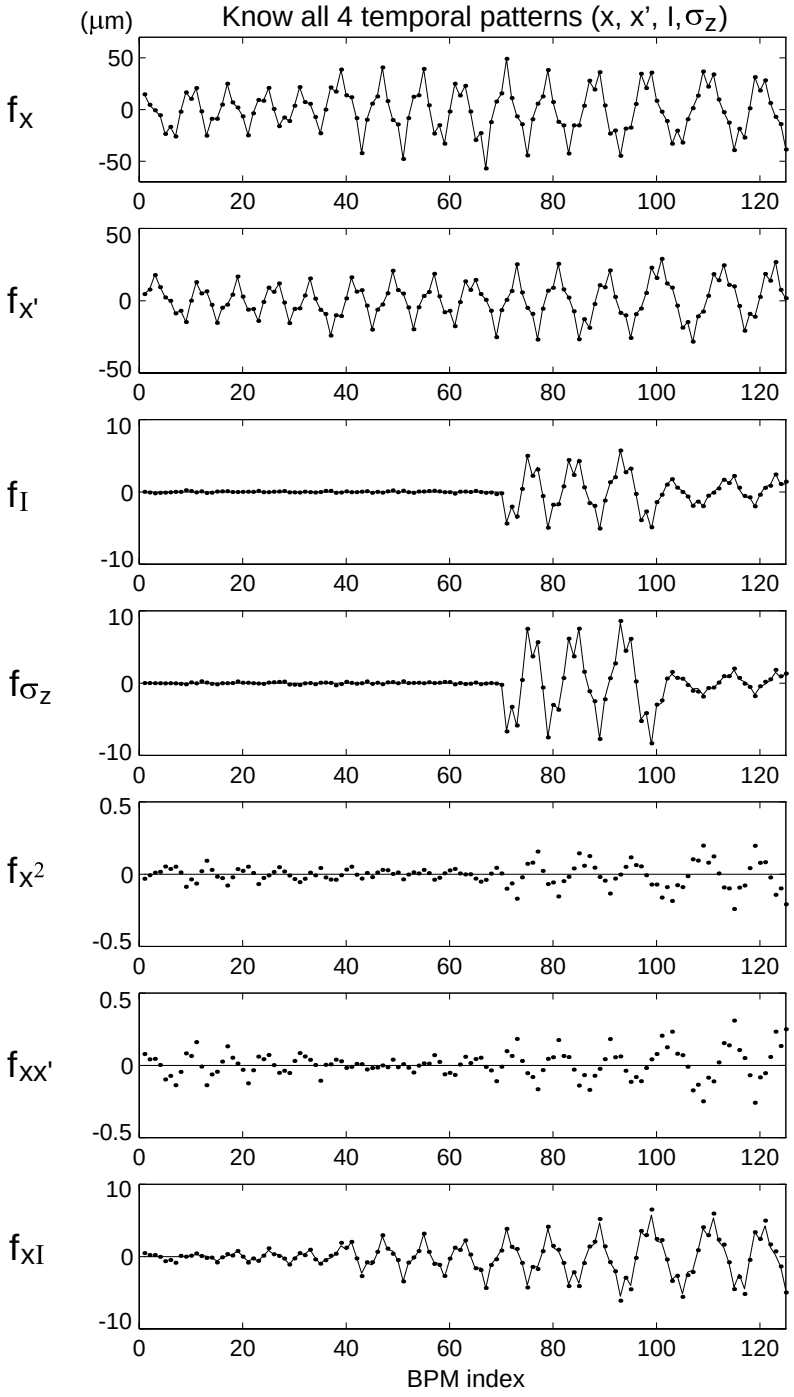

Figure 4.3: Spatial vectors extracted via temporal patterns (x, x', I, σ_z and their 2nd order combinations as labeled on the left-hand-side) of a simulated SLC linac system. (To be continued.)

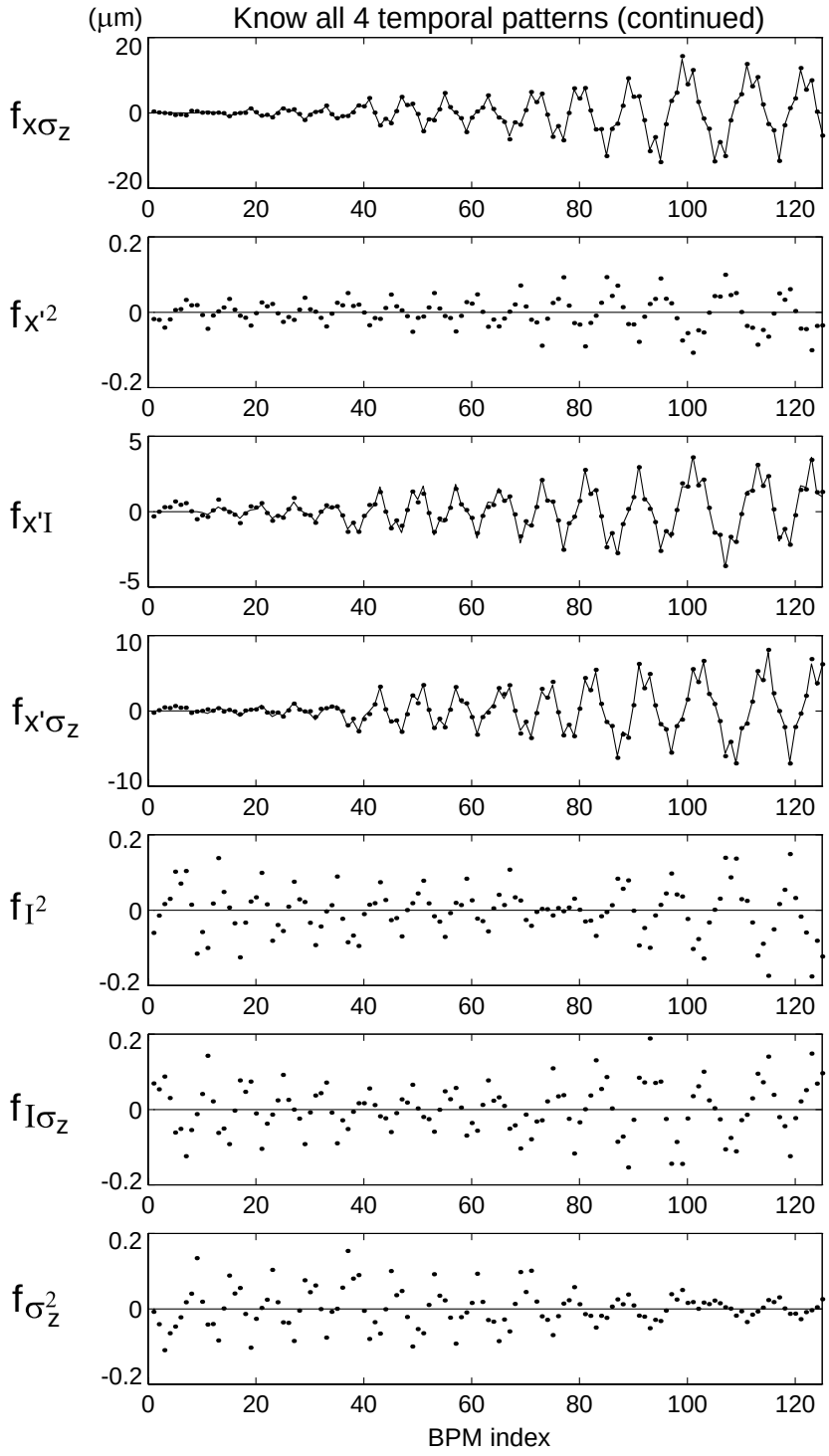

Figure 4.3: Continued.

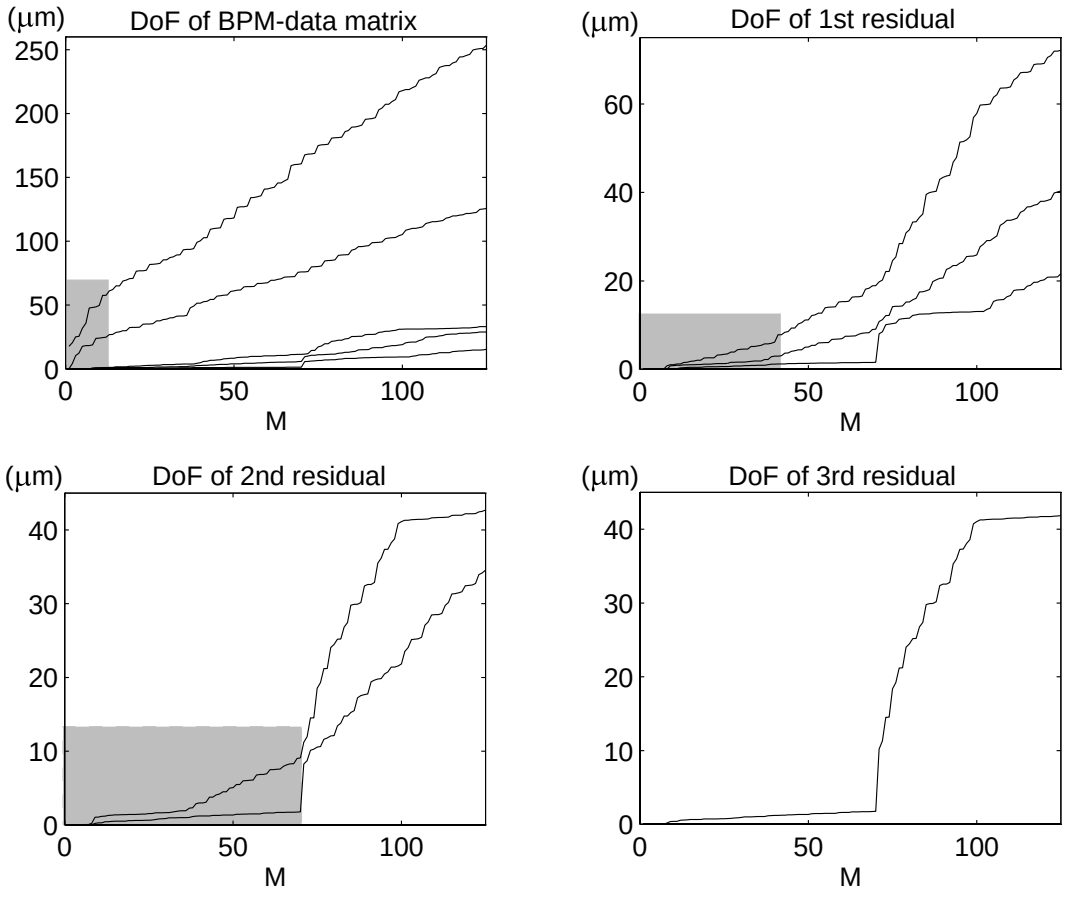

Figure 4.4: Degrees-of-freedom analysis at various stages of a spatial analysis of a simulated SLC linac system.

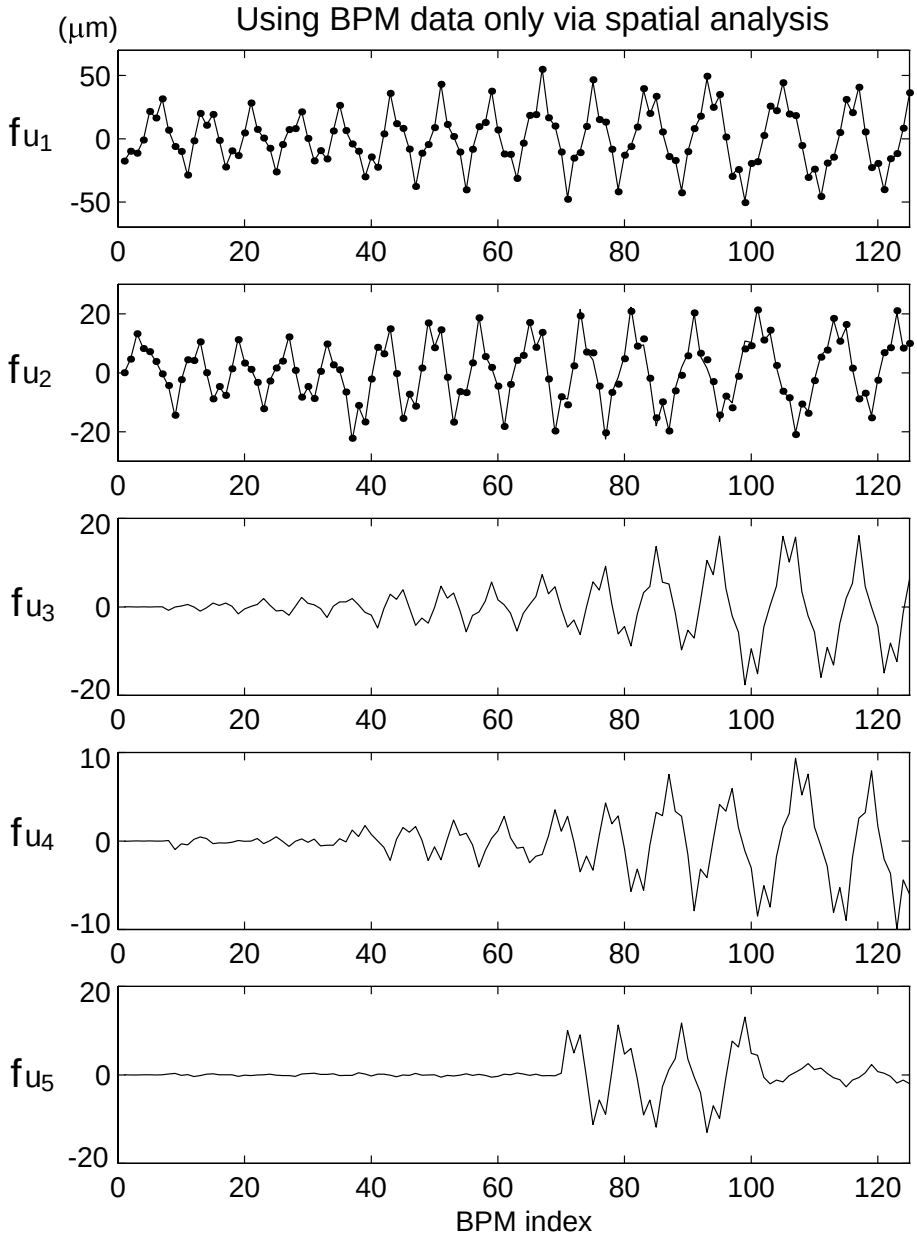

Figure 4.5: Spatial vectors extracted via temporal patterns based on spatial analysis of a simulated SLC linac system. Dots are fit to ideal betatron orbits. (See Fig. 4.6 for the difference.)

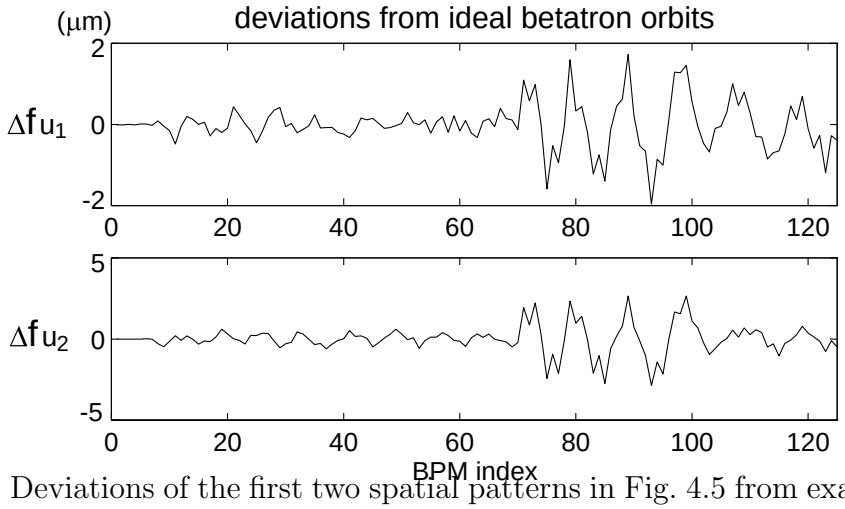

Figure 4.6: Deviations of the first two spatial patterns in Fig. 4.5 from exact betatron orbits.

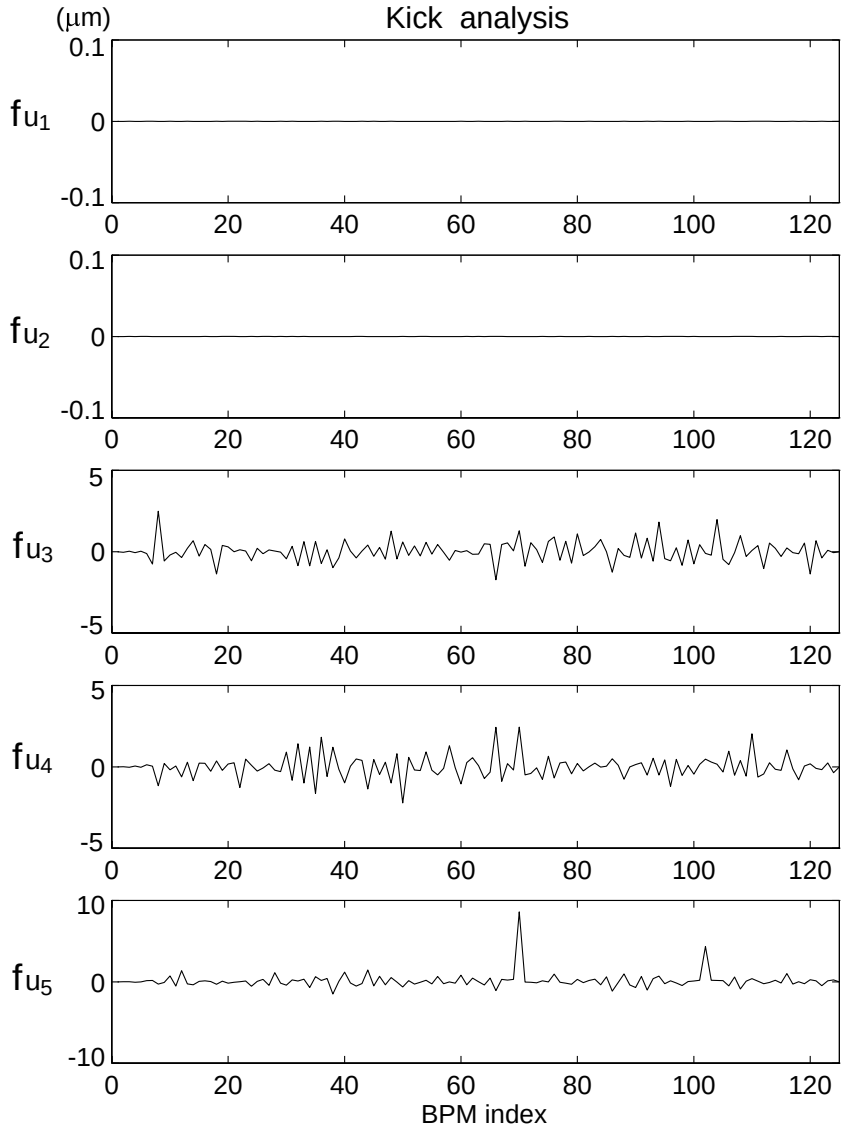

Figure 4.7: Kick analysis of the spatial patterns in Fig. 4.5. Instead of the exact betatron bases, the first two spatial patterns in Fig. 4.5 are used in this kick analysis.

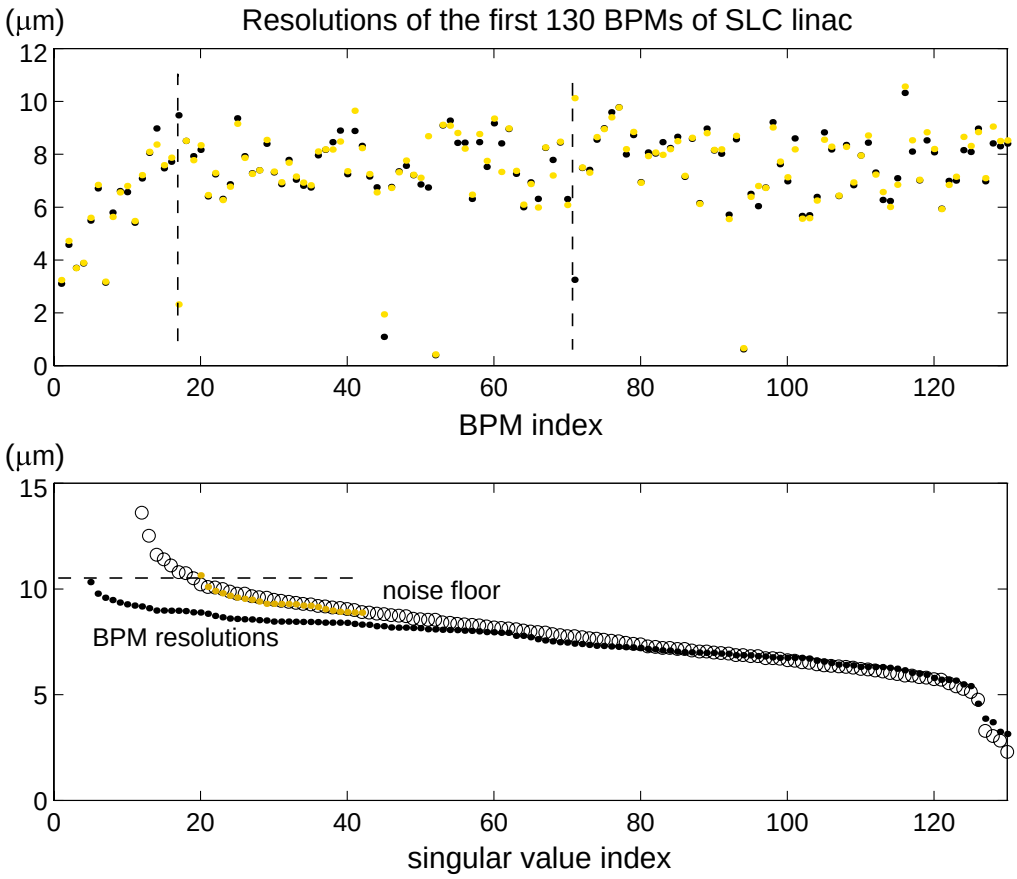

Figure 4.8: Resolutions of the first 130 BPMs of SLC linac and comparison with the singular value noise floor.

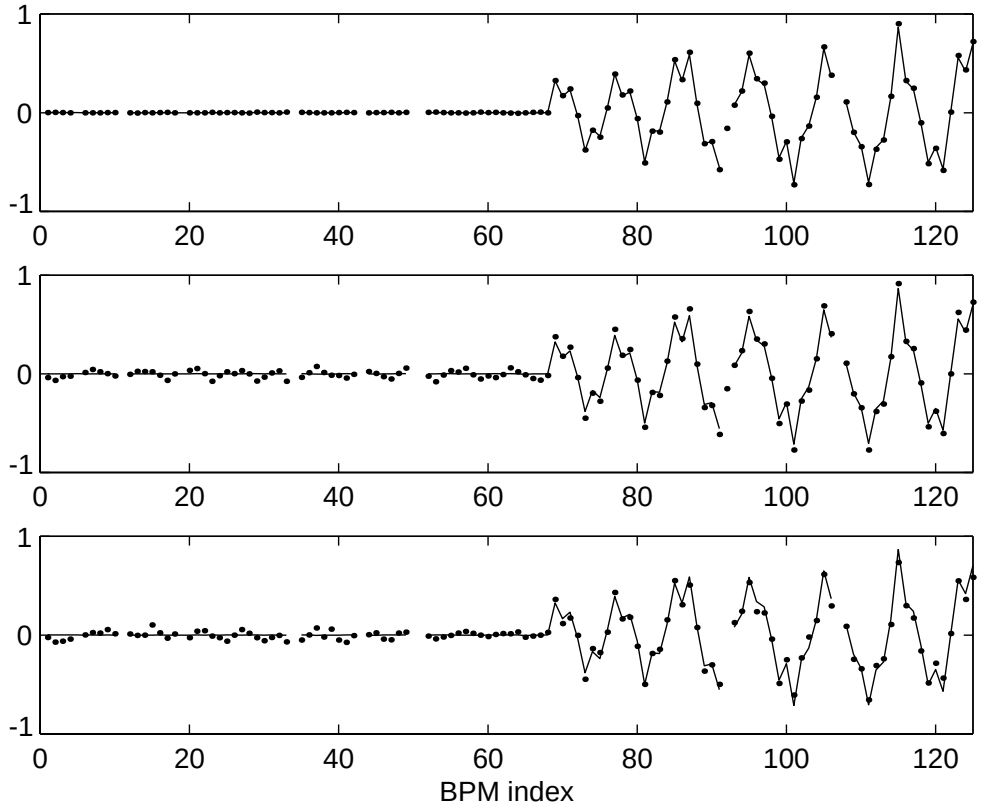

Figure 4.9: Betatron patterns measured via temporal patterns of corrector dither strength.

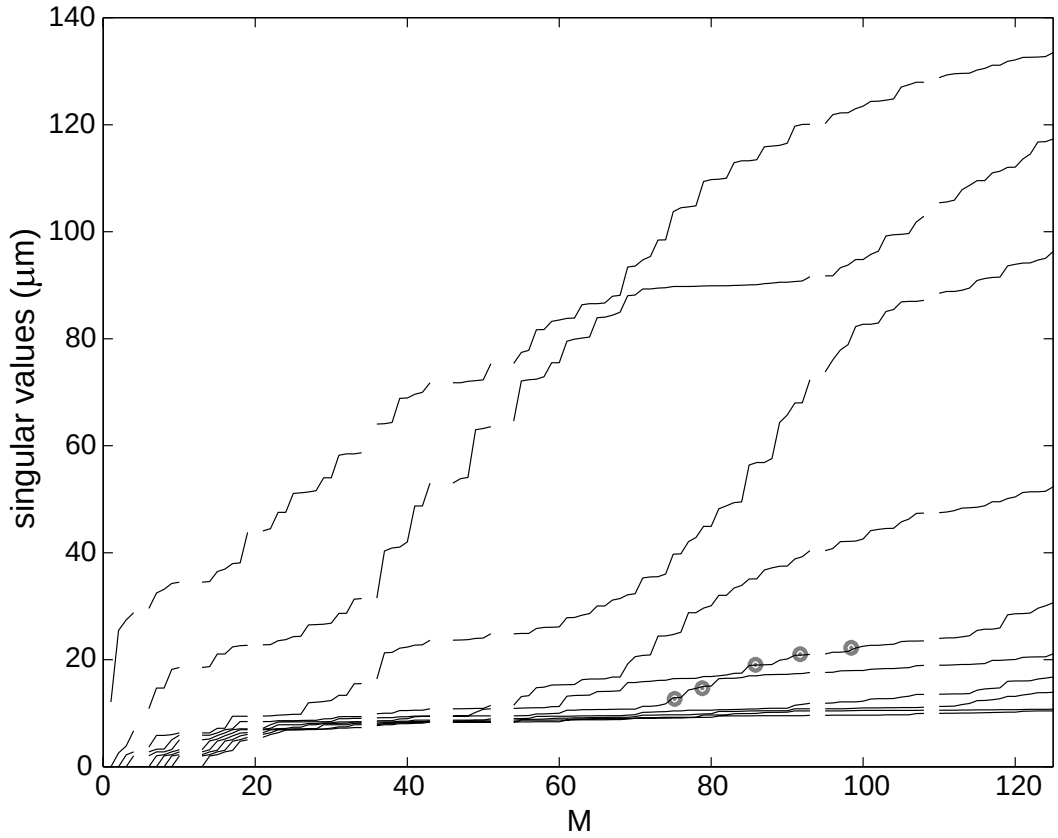

Figure 4.10: Degrees-of-freedom analysis of SLC linac with a corrector dithered, which results in a betatron oscillation of 4 μm rms amplitude.

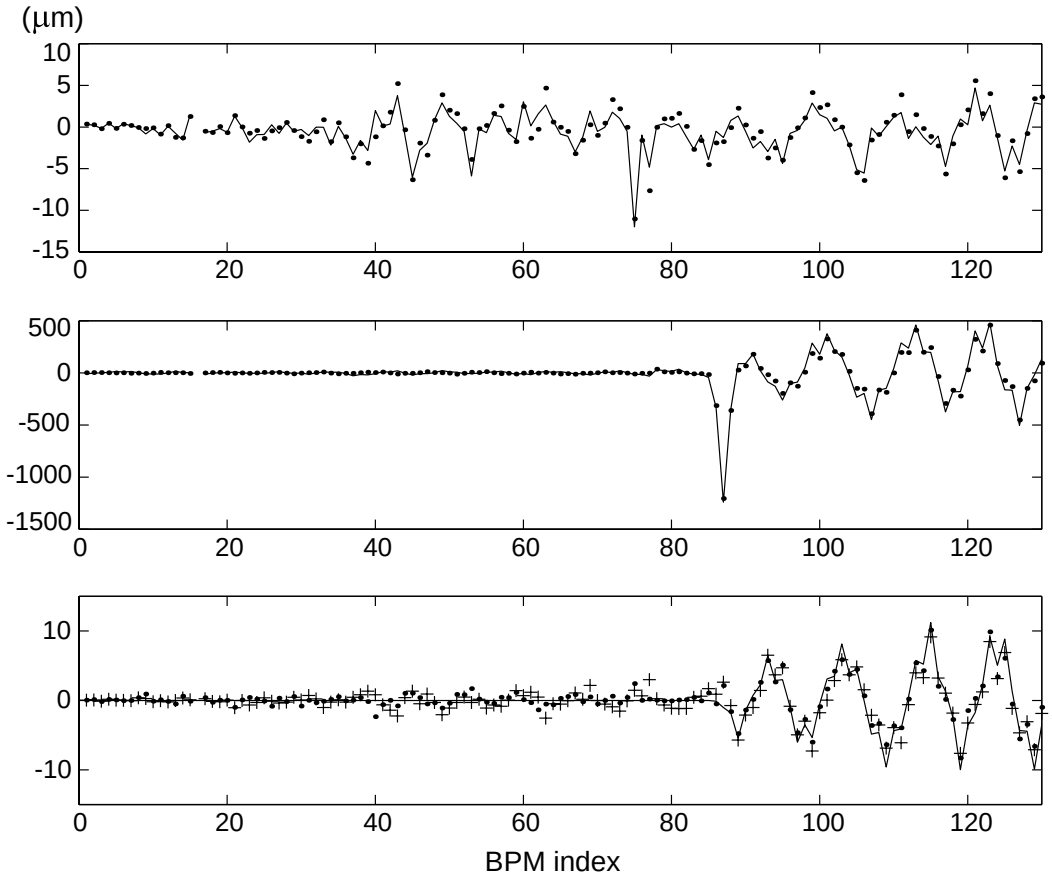

Figure 4.11: Wakefield effect measurements in vertical plane

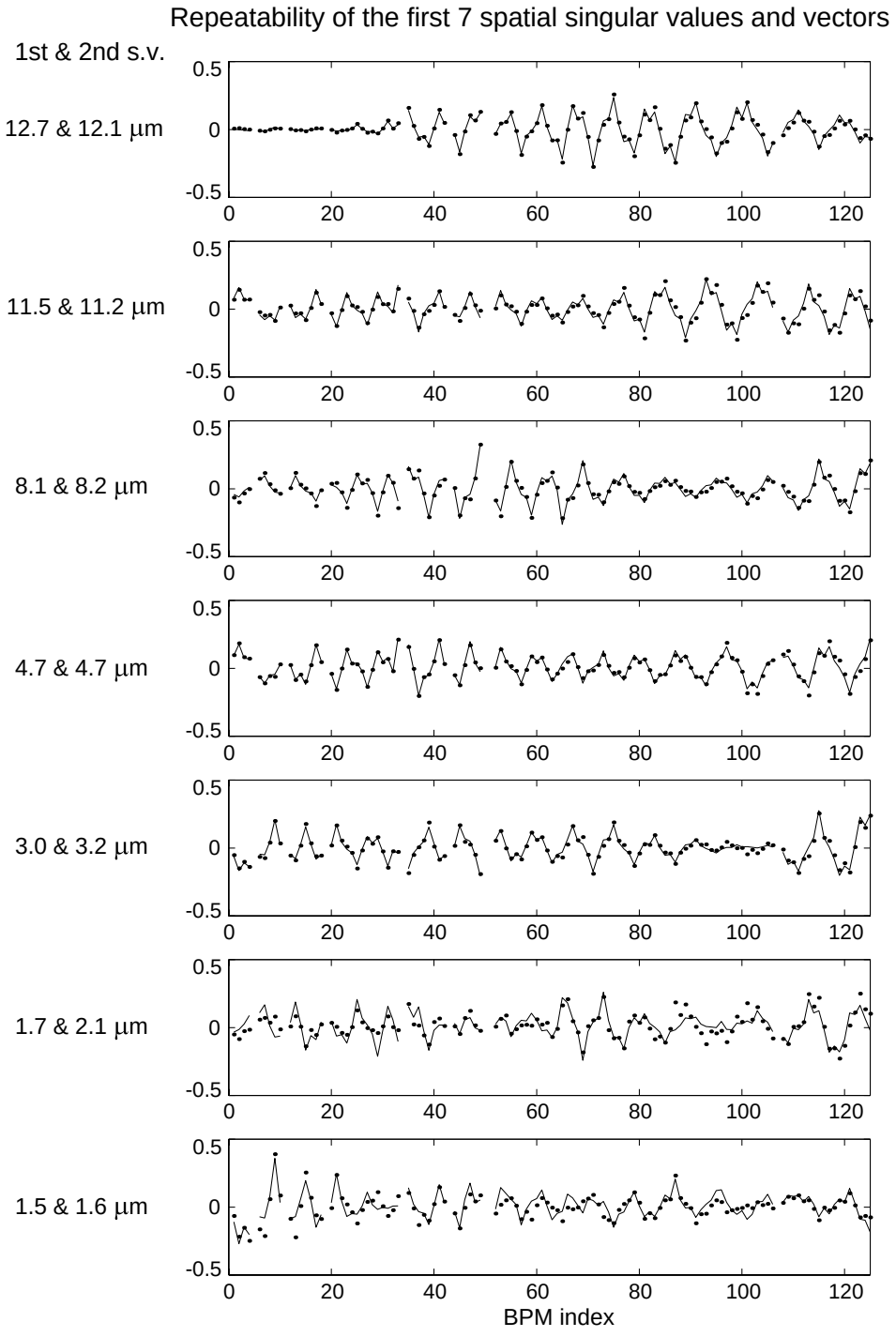

Figure 4.12: Repeatability of singular values and spatial singular vectors of two data sets.

Chapter 5

Possibility to measure nonlinear (1-turn) Taylor maps of a circular accelerator

5.1 Introduction

It is well known that the beam dynamics in accelerator systems is intrinsically nonlinear. In circular accelerators nonlinearity is one of the major factors that limits stability and influences dynamics of halos. In this chapter, I focus on the beam dynamics issues in electron storage rings. In addition to the large effort invested into the design of an accelerator to control nonlinearities, it is essential to determine nonlinearities in the "as-built" machine in order to understand and improve its performance. A standard way to characterize the single particle beam dynamics in a periodic system is via the Poincaré section map $\mathcal{M}$ in phase space. For any initial phase-space point X^i, the map $\mathcal{M}$ gives the new phase-space vector X^f after one turn.

$$\mathcal{M} : X^i \rightarrow X^f \tag{5.1}$$

Of the several ways to represent the map $\mathcal{M}$, I use the map representation in terms of Taylor series:

$$X_k^f = X_k^0 + R_{kl}X_l^i + T_{klm}X_l^iX_m^i + U_{klmp}X_l^iX_m^iX_p^i + V_{klmpq}X_l^iX_m^iX_p^iX_q^i + \cdots \quad (5.2)$$

where the summation convention on the repeated indices is assumed. X_k is the k-th component of a phase-space vector. If the closed orbit is chosen as the coordinate origin, $X_k^0 = 0$. Determination of the map is equivalent to the determination of the coefficients X^0, R, T, U, etc.

There are many kinds of single particle beam dynamics experiments, which roughly fall into two categories. One is oriented toward lattice-diagnostics. In this case, the accelerator is in a normal operation condition and the nonlinearity is weak. Only the linear transfer matrix R and its relation to the designed lattice and various errors in the system are pursued.[28, 18] No successful measurements of high order coefficients have been reported[1], although some nonlinear quantities such as tune-shift-with-amplitude coefficients have been measured.[69] The other kind of experiment is to study the nonlinear beam dynamics by introducing significant nonlinear perturbations into the system via external fields or driving the beam to certain nonlinear resonances.[26, 53, 27] In such cases, the nonlinear effects are significantly enhanced or even dominate the beam motion. Although such experiments are interesting, they provide limited information on the dynamics of an accelerator in its normal operation mode.

In this chapter, I will show that it appears possible to measure the nonlinear (1-turn) Taylor map of a normal machine to a high order and accuracy using presently available high resolution beam position monitors. There are many issues concerning the map measurement. They can be grouped as:

1. physical effects influencing the one-turn map,
    - single-particle map vs. beam-centroid map (5.3)
    - wake field effects (5.4)

[1] M. Lee informed us of his unsuccessful attempt to fit some 2nd order coefficients at SPEAR.

- radiation damping and quantum excitation (5.5)
- stability of the map in the presence of power supply ripples, temperature fluctuations, ground motion, etc. (5.5)

2. measurement of the map (5.2), and
   - number of data pairs to be measured and time required to collect (5.2)
   - number of kickers required to explore the complete phase space (5.2)
   - number of turns that can be used for data collection after kicks (5.2,3,4)
   - effect of BPM resolution and nonlinearity (5.7)
   - beam energy resolution and control of beam energy (5.2,5.7)
3. extraction of map coefficients by fitting (5.2),
   - extracting the single-particle map from the beam-centroid map (5.3)
   - map order to be used and convergence of fitting (5.6)
   - error estimate and hypothesis test (5.8)
   - advantages and drawbacks in breaking down the dimension of fitting (5.8)

Numbers in the brackets indicate the section in which the topic is addressed. I will first describe how to implement the experiment and reasons behind it, then address key issues that may affect such experiments. Simulation results will be presented.

5.2 Experiment implementation and map extraction method

The map $\mathcal{M}$ can be specified by a data set consisting of vector pairs $\{X^i, X^f\}$ which are the initial and final phase-space positions. X^i must cover the phase-space region of interest. In the approximate Taylor map representation, $\mathcal{M}$ is parameterized by its power series coefficients. We can measure the map as data pairs and then convert them into a Taylor representation by fitting. The reason to work with the Taylor map

is that the fitting problem is linear, although there is a well known drawback that the Taylor map is not necessarily symplectic in general and requires significantly more parameters than necessary. Also, as will be addressed later, we actually measure the beam-centroid map, which may not be symplectic.

The measurement is straightforward, which requires a large number (much more than the number of parameters to be determined) of data pairs of sufficient accuracy (to be addressed later). They can be collected in many ways. In the following, I describe a general approach to illustrate the major considerations. For a particular ring, one may need to work around some instrument limitations by taking advantage of certain unique features of the ring.

To sample a certain phase space volume, fast kickers are required to kick the beam to various positions in phase space. After the beam is kicked, its turn-by-turn phase-space positions are recorded. Due to the beam decoherence etc., discussed later, a limited number of turns may be used after each kick. The beam needs to be fully damped to the equilibrium orbit before the next kick to sample other phase-space points. This process can be repeated to obtain the required number of data pairs. In addition to tracking the beam centroid motion, we need to know the equilibrium emittances and energy spread in order to correct for the difference between the beam-centroid map and the single-particle map. See Section 5.3 for details.

Due to the short damping time in an electron machine, thousands of data pairs can be collected in one hour. For example, the PEP-II low-energy ring's horizontal damping time is $\tau_x = 40$ ms. If one assumes measuring each data pair takes about $10\tau_x$ (i.e. only one turn is observed for each kick), 9000 data pairs can be collected within an hour. If 10 turn data after a kick can be used, 90,000 data pairs can be collected each hour. In practice, the data collection speed may be limited by the time for the kickers to recover, or reading data from BPM digitizers.

Ideally two kickers are required in each plane to sample both coordinate and momentum. However, one can use just one kicker in each plane and use many turns of data to sample various coordinates, provided that the beam decoherence is negligible during those turns. Generally, injection kickers can provide the necessary kicks in one and only one plane. To generate kicks in the other plane may require a new kicker

in many rings. Special care is necessary to arrange the x and y kicks as well as the number of turns to be used. In some rings, it may be possible to sample the phase space by steering the injected beam. In such a case, one needs to evaluate whether the variation in the phase-space distribution of the injected bunches can yield significant errors due to decoherence. Moreover, the fluctuation of the injected currents may affect the accuracy of BPM readings.

Position and momentum can be determined with a minimum of two BPMs assuming the nonlinearity between the BPMs is negligible. This assumption should be a very good approximation because the distance between neighboring BPMs is usually short compared to the circumference of the accelerator. Moreover, one can always choose a location where the nonlinearity is not likely to be a concern. To measure the full transverse phase space, two BPMs are required in each plane. The BPMs must be sufficiently fast to measure the turn-by-turn trajectory. The accuracy of position and momentum measurements is critical and a large part of this dissertation is devoted to improving it. See Chapters 2 and 6 for more extensive discussions.

In order to study the energy dependency of a map, the beam energy can be changed by varying the RF frequency of the accelerating field[30]. The relative energy changes can be deduced via

$$\frac{\Delta\omega_{RF}}{\omega_{RF}} = -(\frac{\Delta C}{C} - \frac{\Delta v}{v}) = -(\alpha_c - \frac{1}{\gamma^2})\delta + O(\delta^2) \tag{5.3}$$

where α_c is the linear momentum compaction factor of the ring and γ is the beam energy. Often a stored beam has undamped synchrotron oscillations due to noise in the RF system, etc. In case such energy variation is not negligible, the turn-by-turn beam energy has to be measured and taken into account in map fitting. I will discuss energy resolution in Section 5.7.

The fitting procedure is the standard least-squares fitting, which will be illustrated via a 2D phase space $\{x, p\}$ example. The map to be fitted is

$$\begin{cases} x^f = x_0 + R_{11}x + R_{12}p + T_{111}x^2 + T_{112}xp + T_{122}p^2 + \cdots \\ p^f = p_0 + R_{21}x + R_{22}p + T_{211}x^2 + T_{212}xp + T_{222}p^2 + \cdots \end{cases} \tag{5.4}$$

Assume there are n terms in each series and a total of m data pairs $\{(x_i, p_i) \rightarrow (x_i^f, p_i^f),\ i = 1 \rightarrow m\}$. Then the x component of the map may be written as

$$\vec{x^f}_{m\times 1} = A_{m\times n} C_{n\times 1} \tag{5.5}$$

where

$$\vec{x^f} = [x_1^f, x_2^f, \cdots, x_m^f]^T,$$

$$A = \begin{bmatrix} 1 & x_1 & p_1 & x_1^2 & x_1p_1 & p_1^2 & \cdots \\ 1 & x_2 & p_2 & x_2^2 & x_2p_2 & p_2^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & p_m & x_m^2 & x_mp_m & p_m^2 & \cdots \end{bmatrix},$$

$$C = [x_0, R_{11}, R_{12}, T_{111}, T_{112}, T_{122}, \cdots,\ \text{n-th coefficient }]^T$$

Since the vector $\vec{x^f}$ and design matrix A are determined by the measured data, we have a set of linear equations for the map coefficients C. Because the data contain errors, which will be addressed later, it is better to have $m \gg n$ so that the equations are over-determined and look for the solution $\hat{C}$ which makes the norm $\|\vec{x^f} - A\,\hat{C}\|$ minimum. (This is just the least-squares fitting method if the Euclidean norm is used.)[40, 51] The solution reads

$$\hat{C} = A^+ \vec{x^f} = (A^T A)^{-1} A^T\ \vec{x^f} \tag{5.6}$$

where A^+ is the pseudo-inverse (Moore-Penrose inverse) of the rectangular matrix A and A^T is the transpose of A. The second expression can be used in the case $A^T A$ is not singular, otherwise more involved but standard mathematical routines are required. The same procedure holds for the other map component, i.e. $\vec{p^f}$. More discussions on fitting and error estimates will be presented in the following sections.

A very nice feature in one-turn map measurement is that all the measurement errors of transverse phase space are limited to the four BPMs, and BPMs are steadily improving in quality[14, 7]. The accuracy of the kickers do not play a role in the experiment as long as they are sufficiently fast. This is because the kickers are used to kick the beam to the neighborhood of a phase space point, while the exact beam

position is measured by the BPMs. However, besides BPM accuracy, there are many beam dynamics issues that may limit our ability to measure the single-particle map. I will address these issues in the following sections.

The analysis presented later are based on simulations which use the SLAC/PEP-II design parameters[5]. A 9-th order Taylor map of the low-energy ring (LER ring) in dynamical variables (x, P_x, y, P_y, δ) is used to model the single particle dynamics. The map has been used in the stability studies of the ring. Higher order (> 9) terms have no significant effects on beam dynamic aperture. To simulate the experiment, a large set of data pairs sampling the 5D phase space are generated randomly (or in grids). In addition, random errors are added to simulate the BPM and energy resolutions. If not specified, the number of data pairs used for fitting studies is 10 times the number of coefficients to be fit (i.e. $m/n = 10$), other ratios are studied in Section 5.8. For simplicity, all fittings are done in 5D phase space simultaneously. Fitting results presented in this chapter are only for the x component of the map. Results for other components are similar and, have the same design matrix A.

5.3 Single-particle map vs. beam-centroid map

Although the goal is to measure the single-particle map, the signals are from a bunch of particles. This may cause two kinds of problems. One is of dynamical origin–the collective effects. The other is of kinematic origin–the decoherence problem, which will exist even though all the interactions among the particles are negligible. This is because the BPM measures the centroid motion of a beam while particles in the beam may have significantly different motions. A well-known example is the decoherence of a kicked beam, in which the centroid motion can be damped to zero although each particle's motion has not yet been damped.[72, 44, 55]

Neglecting collective effects, every particle in a beam follows the same single-particle map. Therefore, the beam centroid $\bar{X}$ follows

$$
\begin{aligned}
\bar{X}_k^f &= [X_k^0 + \hat{T}_{klm}\langle\delta X_l^i \delta X_m^i\rangle + \hat{U}_{klmp}\langle\delta X_l^i \delta X_m^i \delta X_p^i\rangle + \hat{V}_{klmpq}\langle\delta X_l^i \delta X_m^i \delta X_p^i \delta X_q^i\rangle + \cdots] \\
&\quad + \bar{X}_l^i [R_{kl} + \hat{U}_{klmp}\langle\delta X_m^i \delta X_p^i\rangle + \hat{V}_{klmpq}\langle\delta X_m^i \delta X_p^i \delta X_q^i\rangle + \cdots] \\
&\quad + \bar{X}_l^i \bar{X}_m^i [T_{klm} + \hat{V}_{klmpq}\langle\delta X_p^i \delta X_q^i\rangle + \cdots] \\
&\quad + \bar{X}_l^i \bar{X}_m^i \bar{X}_p^i [U_{klmp} + \cdots] \\
&\quad + \bar{X}_l^i \bar{X}_m^i \bar{X}_p^i \bar{X}_q^i [V_{klmpq} + \cdots] \\
&\quad + \cdots \\
&\equiv \bar{X}_k^0 + \bar{R}_{kl}\bar{X}_l^i + \bar{T}_{klm}\bar{X}_l^i \bar{X}_m^i + \bar{U}_{klmp}\bar{X}_l^i \bar{X}_m^i \bar{X}_p^i + \bar{V}_{klmpq}\bar{X}_l^i \bar{X}_m^i \bar{X}_p^i \bar{X}_q^i + \cdots
\end{aligned}
\tag{5.7}
$$

where $\langle\cdots\rangle$ means average over the beam phase space distribution of X^i (assumed symmetric) and $\delta X = X - \bar{X}$ is a particle's deviation from the centroid.[2] The hatted coefficients are related to the unhatted ones by constant factors $\prod_i \binom{n(i)}{m(i)}$, where $\binom{n}{m}$ is the binomial coefficient and, $n(i)$ and $m(i)$ are the number of i among the running indices in the coefficients and in the $\langle\cdots\rangle$ terms respectively.[3] Note that permutations of the running indices are assumed not to contribute in Eqs. (5.2) and (5.7). It is clear that the beam centroid does not follow the single-particle map. Extra terms appear and depend on the beam phase-space distribution. The difference between the centroid map and single-particle map decreases with the beam emittance. The two maps are the same if the nonlinearity is negligible.

It is important to realize that, as long as the beam maintains the same distribution, the coefficients $\bar{X}^0$, $\bar{R}$, $\bar{T}$, $\bar{U}$, etc. are constant, i.e. the beam centroid follows a well-defined Taylor map also, although this centroid map is different from the single-particle map. This observation allows one to overcome the single particle vs. beam problem, but imposes a strong condition on the experiment: for each measurement

[2]For Gaussian distribution with $\langle\delta X\rangle = 0$, it is well known that $\langle\delta X_p \delta X_q\rangle = \sigma_{pq}^2 = \sigma_p \sigma_q \rho_{pq}$, where ρ_{pq} is the covariance matrix. And $\langle\delta X_p^n\rangle = \begin{cases} 0 & n\ odd \\ (n-1)!!\sigma_p^n & n\ even \end{cases}$

[3]It is not important to understand these factors. Nonetheless, here is an example to help. In 5D, the index i runs from 1 to 5. The term $\hat{U}_{1224}\langle\delta X_2^i \delta X_4^i\rangle$ in $\bar{X}_l^i \hat{U}_{1lmp}\langle\delta X_m^i \delta X_p^i\rangle$ of Eq.(5.7) should read:

$$\hat{U}_{1224}\langle\delta X_2^i \delta X_4^i\rangle = \binom{0}{0}\binom{2}{1}\binom{0}{0}\binom{1}{1}\binom{0}{0} U_{1224}\langle\delta X_2^i \delta X_4^i\rangle = 2U_{1224}\langle\delta X_2^i \delta X_4^i\rangle.$$

the beam must have the same phase-space distribution. In an electron storage ring, radiation damping can accomplish this condition, by damping the beam to the equilibrium state before each kick. The kick is assumed to move the beam centroid but does not significantly change the distributions relative to the centroid. Therefore, the same (equilibrium) phase-space distribution can be achieved for each measurement. In proton machines, it may be more difficult to control the phase-space distribution.

The centroid map is changing after the beam is kicked due to filamentation of the phase-space distribution, limiting the number of turns which can be used. However, it should be safe to assume that the centroid map will not change significantly in a few turns, because the tune spread due to tune shift with amplitude and chromaticity is usually very small. The decoherence problem could be minimized by kicking a well-damped beam and using just one-turn data. Depending on various conditions for a particular measurement, it may be possible to use many-turn data. For the PEP-II example, at least 20 turns may be used.

After the centroid map is determined, the single-particle map can be solved using the information on the equilibrium beam phase space distribution. (Thus the equilibrium emittances and energy spread information is needed.) Note that the 0-th order term may appear in the centroid map even though $X^0 = 0$, which means that the closed orbit observed with BPMs is different from the closed orbit of the single-particle map and a coordinate translation may be necessary to obtain the single-particle map. Another important issue is that the centroid map is not symplectic in general, even though the single-particle map always is. Therefore, one has to be careful when trying to use symplecticity in fitting.

Notice that the corrections to the nonlinear coefficients are due to the coefficients which are at least two orders higher, assuming symmetric distributions. Therefore, the corrections are normally small for a well damped beam, which makes it much easier to extract the single-particle map. Moreover, the leading low order terms tend to dominate the corrections—which is the reason that Table 5.1 shows approximate linear dependency on beam emittance and square of energy spread. Taking the PEP-II low energy ring as an example, the maximum corrections at each order are shown in Table 5.1. The first two columns indicate the emittance and energy spread used

emitt.	δ spread	$\bar{x}^0$	orders						
in ϵ_0	in σ_δ	in μm	1-st	2-nd	3-rd	4-th	5-th	6-th	7-th
0	1	1.03	2.60	7.3	11.9	18.9	16.6	23.4	24.5
1	1	0.66	2.66	9.9	11.7	19.4	16.9	25.2	25.0
2	1	0.25	2.72	12.4	11.6	20.0	17.3	27.3	25.5
4	1	-0.60	2.78	17.3	11.1	20.7	17.7	30.9	26.0
9	1	-2.50	3.24	29.8	10.0	22.7	19.0	39.4	27.1
0	2	4.13	9.66	25.7	41.6	68.9	60.3	93.0	97.2

Table 5.1: Maximum absolute corrections ($\times 10^3$ except for $\bar{x}^0$) of coefficients in centroid maps with increasing beam emittance and energy spread for the PEP-II low-energy ring.

for the calculation. ϵ_0 and σ_δ stand for the design emittance and energy spread of the ring. An uncorrelated 5D Gaussian distribution is used. The 0-th order terms are in μm, while others are the maximum corrections in 10^{-3} (for 10σ normalized coefficients[4]). These may be compared with typical fitting errors shown in Table 5.2.

Table 5.1 shows that, the 0-th order terms are negligible although non-zero. The corrections are fairly small in linear terms but become significant in the nonlinear terms. However, only a few terms such as δP_x, $\delta^2 x$, $\delta^3 P_x$, $\delta^4 x$, $\delta^2 P_x x^2$, etc. actually have significant changes. The corrections are dominated by the beam energy spread and not sensitive to the emittance growth. Therefore, it should not be difficult to fulfill the requirement of having exactly the same beam emittance for each kick during the measurement. Also the beam emittance and energy spread need not be known very accurately in order to get sufficiently good correction.

Before moving on to the next topic, I would like to point out that, compared to the technique of harmonic analysis of multi-turn data, the map measurement described here has the advantage of being free from problems due to beam decoherence, because the harmonic analysis requires a relatively large number of consecutive turns, while the map measurement can use just a few turns after each kick.

[4]For convenience, I normalize all dynamical variables to their 10σ values of the beam distribution, so that all map coefficients become dimensionless and their values reflect the relative importance of the corresponding terms near the border of the dynamical aperture.

5.4 Wake field effects

In addition to the dynamics described by the single-particle map, a particle in a beam experiences wake field forces due to collective effects. To measure the map, such wake field effects must be limited or procedures developed to correct its effects during map extraction. The best method to reduce collective effects is to reduce the beam current. However BPM resolution may decrease also. Therefore, beam current will need to be optimized for the best map measurement.

To estimate the wake field effects, I calculate the kicks on the beam centroid when a Gaussian bunch passes one turn of the ring. Assuming the betatron and synchrotron motions are uncoupled, the change of beam centroid due to wake forces is given by [71, 25, 41]

$$\Delta\delta = \frac{\langle\, \bar{\vec{F}}_{\parallel} \rangle}{E} \qquad \text{and} \qquad \Delta\vec{p}_{\perp} = \frac{\langle\, \bar{\vec{F}}_{\perp} \rangle}{E} \tag{5.8}$$

where E is the beam energy; $\langle \cdots \rangle$ means average over the bunch distribution; the top $\bar{}$ means integrated longitudinally over the wake structure. The average longitudinal and transverse impacts on the centroid are:

$$\begin{aligned} \langle\, \bar{\vec{F}}_{\parallel} \rangle &= -\frac{1}{N} \sum_{m=0}^{\infty} |I_m|^2 \; k_l^{(m)} \\ \langle\, \bar{\vec{F}}_{\perp} \rangle &= \frac{1}{N} \sum_{m=0}^{\infty} m(\Re[I_{m-1} I_m^*]\hat{x} - \Im[I_{m-1} I_m^*]\hat{y}) \; k_{\perp}^{(m)} \end{aligned} \tag{5.9}$$

where N is the number of particles in the bunch, I_m is the m-th moment of a Gaussian beam given by [71]

$$I_m = Ne \sum_{n=0}^{[\frac{m}{2}]} (2n-1)!!\binom{m}{2n}(x_c + i\, y_c)^{m-2n}(\sigma_x^2 - \sigma_y^2 + i\; 2\rho\sigma_x\sigma_y)^n. \tag{5.10}$$

where x_c and y_c are the beam centroid position, and ρ is the correlation coefficient of the x and y distributions. $k_l^{(m)}$ and $k_{\perp}^{(m)}$ in Eq. (5.9) are the energy loss factor and

transverse kick factor respectively for the multipole mode m, which are defined by

$$\begin{aligned} k_l^{(m)} &\equiv \int_0^\infty \frac{d\omega}{\pi}\, |\tilde{\rho}_\parallel(\omega)|^2\, \Re[Z_m^\parallel(\omega)] \\ k_\perp^{(m)} &\equiv -\int_0^\infty \frac{d\omega}{\pi}\, |\tilde{\rho}_\parallel(\omega)|^2\, \Im[Z_m^\perp(\omega)] \end{aligned} \tag{5.11}$$

where $\tilde{\rho}_\parallel$ is the Fourier transform of the longitudinal bunch distribution and Z_m is the impedance of the ring. The loss factor and kick factor are usually available for the longitudinal monopole mode and transverse dipole mode, because they are the dominating modes for a near-axis beam. The following scaling provides a useful estimate for higher modes:[25]

$$Z_m^\parallel \sim \frac{2}{b^{2m}} Z_0^\parallel \qquad \text{and} \qquad Z_m^\perp \sim \frac{1}{b^{2m-2}} Z_1^\perp, \tag{5.12}$$

where b is the pipe radius.

In map measurements, we are interested in large amplitude motions with $x_c \gg \sigma_x$ and $y_c \gg \sigma_y$. Therefore I_m is dominated by the $n = 0$ term $(x_c + i\, y_c)^m$, which means the bunch acts like a macro particle. With the impedance estimates in Eq.(5.12), we get

$$\Delta\delta \sim -\frac{N r_e}{\gamma} k_l^{(0)} \left[1 + 2\sum_{m=1}^{\infty} (\frac{r_c}{b})^{2m}\right] = -\frac{N r_e}{\gamma}\, k_l^{(0)}\, \frac{1 + (\frac{r_c}{b})^2}{1 - (\frac{r_c}{b})^2} \tag{5.13}$$

and

$$\begin{aligned} \Delta\vec{p}_\perp &\sim \frac{N r_e}{\gamma} k_\perp^{(1)} (x_c\hat{x} + y_c\hat{y}) \sum_{m=1}^{\infty} m (\frac{r_c}{b})^{2m-2} \\ &= \frac{N r_e}{\gamma} k_\perp^{(1)} (x_c\hat{x} + y_c\hat{y}) \left[1 - (\frac{r_c}{b})^2\right]^{-2} \end{aligned} \tag{5.14}$$

where r_e is electron's classical radius and γ is the beam energy. $k_l^{(0)}$ and $k_\perp^{(1)}$ are the loss factor and kick factor for the lowest multipole modes. Fig. 5.1 is a plot of the nonlinear factors in Eqs. (5.13) and (5.14). Note that, the high order contributions become significant when the beam is close to the pipe. Beam displacements within 50% of the pipe radius (i.e. $r_c = 0.5b$) are in a comfortable range, and 80% should

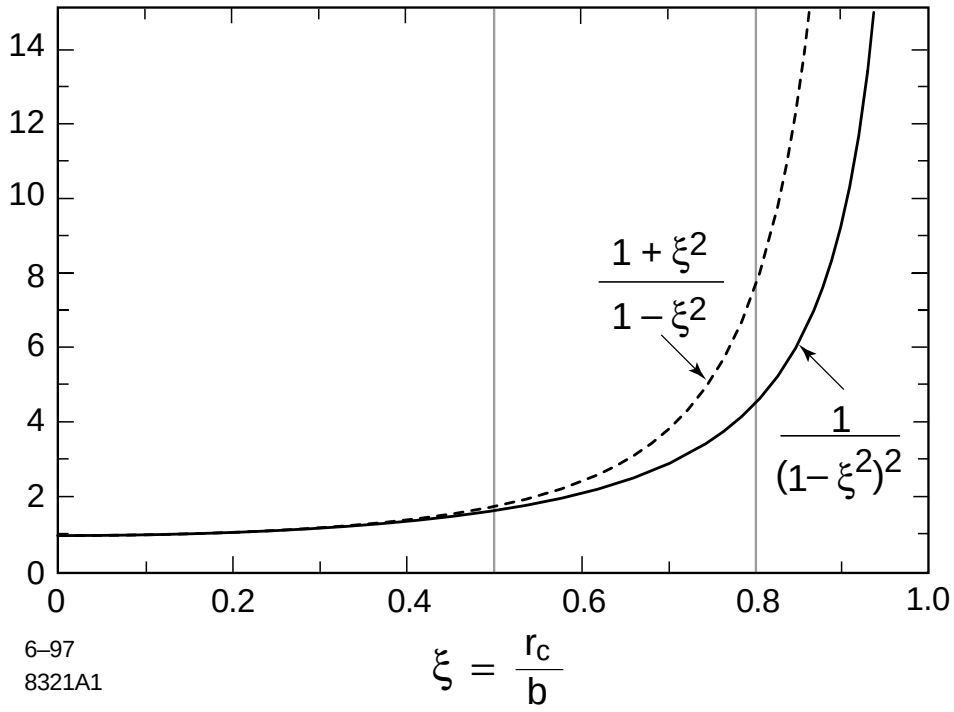

Figure 5.1: Nonlinear factors in the longitudinal and transverse (solid and dashed curves) wakefield kicks

be the upper limit. The nonlinear factors in $\Delta\delta$ and $\Delta\vec{p}_\perp$ are about the same within this range. Eq.(5.14) can be viewed as a thin kick map. Normalizing the dynamical variables to 10σ, we get roughly the same map coefficient for all orders, which reads

$$k_w \equiv \frac{Nr_e}{\gamma} k_\perp^{(1)} \, (\frac{10\sigma_x}{10\sigma_p}) = \frac{Nr_e}{\gamma} k_\perp^{(1)} \, \bar{\beta} \tag{5.15}$$

where $\bar{\beta}$ is the average beta value of the ring. k_w provides a convenient way to estimate the significance of the wake effects.

For PEP-II LER, the design wake has $k_\perp^{(1)} \simeq 200 V/pC\, m$ in c.g.s. units.[41] For a 3 GeV, 10^{10} electron bunch, and $\bar{\beta} \simeq 10m$, we get $k_w \simeq 10^{-3}$. As a nonlinear coefficient, it is negligible; but for the linear map, such a perturbation is larger than the expected linear coefficient resolution (see Table 5.3) in map extraction. Thus we should be able to see its effect on the linear map (e.g. linear tune-shift). But such small

linear perturbations will not change significantly the nonlinear map coefficients due to "feed-up" with nonlinear coefficients. For the longitudinal wake, $k_l^{(0)} \simeq 10V/pC$, hence the effect on the beam energy is negligible. Therefore, wake field effects on map measurements are tolerable for the LER. Nonetheless, in general, wake fields may be a physical factor limiting map measurements, and specially designed low-current-high-resolution BPMs may be necessary for some rings. In experiments, one can identify wake effects by making measurements at different beam currents.

Since the wake effects may be observable in map measurements, methods to extract such effects should be pursued. In general, the wake field effects can not be characterized by one-turn maps since the interaction depends on the beam phase-space distribution and the history of the beam trajectory, which are time dependent. However, in map measurements the phase-space distribution can be considered the same for each one-turn pair measurement, and there is almost no multi-pass wake effects when just one-turn data are used. Therefore it is reasonable to characterize the wake effects by one-turn kicks given above. In such a case, we can measure maps at different currents and extrapolate to zero current. Such corrections should significantly reduce wake problems. Although locally the wake force depends linearly on beam current, it may show up nonlinearly in a one-turn map because of a coupling with lattice nonlinearities. However, such coupling should be fairly small and the linear dependence on current is expected to dominate.

In summary, wake fields may limit map measurements. k_w of Eq.(5.15) and the energy loss of Eq.(5.13) can be used to estimate the significance of wake field effects. In case it is not negligible, linear extrapolation to zero current can help to extract wake effects. However, I only considered the one-turn wake force. When the impedance of a ring is large, using multi-turn data may further complicate problems through long-range wake effects.

5.5 Radiation damping, quantum excitation, and external noise

Radiation damping and quantum excitation are well-known physical processes that affect single-particle dynamics. When a particle passes through horizontal bending magnets, energy is lost due to radiation, vertical action is conserved, and horizontal action is increased. However, when the particle makes up the energy loss in the RF system, actions in both directions are damped.[64] The relative action change reads $\frac{\Delta A^2}{A^2} = 2\frac{\tau_0}{\tau_x}$, i.e. the betatron amplitude change is $\frac{\Delta A}{A} = \frac{\tau_0}{\tau_x}$, where τ_0 is the revolution period and τ_x the transverse damping time. This number is very small in general. Therefore, unless BPM resolution is better than $10\sigma_x \frac{\Delta A}{A}$ (assuming maximum kick is 10σ), radiation damping effects should be negligible. For the PEP-II LER, $\frac{\tau_0}{\tau_x} \simeq 10^{-4}$ and $\sigma_x \simeq 1\,mm$, which yields an orbit change of $1\,\mu m$. Since the radiation damping could be taken into account in the one-turn map calculation, it is in any case not a fundamental limit to map measurements.

Due to the quantum nature of the radiation process, the smooth radiation damping picture could be modified by quantum excitations. To estimate the probable magnitude of a betatron amplitude fluctuation in a turn, I use the fact that beam emittance is determined by the balance of radiation damping and quantum excitation. Thus the expected action growth in one turn is $\Delta A^2 = 2\frac{\tau_0}{\tau_x}\sigma_x^2$, which is equivalent to a rms amplitude fluctuation $\Delta A \simeq \frac{\tau_0}{\tau_x}\sigma_x$. This should also not be a concern.

Up until now, I have discussed various physical factors concerning the existence of a detectable single-particle map. Practically we still need to consider the stability of such a map under external disturbances (e.g. power supply ripples, temperature variations, ground motion) over the period of measurement. The dominant noise spectrum due to power supply ripple consists of the power line frequency and its harmonics, which are in the range of 100 – 1000 Hz. They are too slow to affect the single-turn measurement, but sufficiently fast to perturb the map from sample to sample. Due to their weak amplitude, the major effect of such high frequency noise is tune modulation. I added random linear tune kicks in our simulation to study the effect of such tune perturbations on map measurement, and found that of a tune

fluctuation up to 10^{-3} is tolerable for LER map measurements with 10 μm BPMs. A very rough estimate can also be done for power supply ripple effects by converting the map coefficient variations to beam energy error via the generic dependency on $\frac{B}{1+\delta}$ in Hamiltonians. This yields $\Delta\delta \propto \frac{\Delta B}{B} \propto \frac{\Delta I}{I}$, where B and I are the magnetic fields and driving currents. PEP-II power supplies are regulated to better than 10^{-4}. Taking into account the significant damping due to transmission lines and eddy currents in vacuum chambers[22], $\Delta\delta$ should be less than the required energy resolution. In general, power supply ripples are limited by design because they may cause emittance growth. In case they are untolerablely large, one needs to pursue methods[21] to suppress any tune modulation. In fact, the most serious high frequency noise will probably come from the RF system. Turn-by-turn beam energy measurement may help to take into account the energy variation due to noisy longitudinal beam motion. In case there is a coupling between synchrotron and betatron motions, a noisy RF system could limit the possibility of measuring the one-turn map.

Unlike the high frequency perturbations, very low-frequency disturbances may be problematic. Here I am talking about diurnal temperature change[50], ground motion due to moon tides[11], etc. whose effects has been identified both in colliders for particle physics and rings for synchrotron radiation sources. For example, SPEAR experienced several hundred μm closed-orbit oscillations because of diurnal temperature variation. Such changes may be tolerable for the machine's operation since control and feedback systems can be used to adjust machine performance. However, during map measurements, no feedback should be used and the lattice should be left alone. Therefore such very low frequency disturbances could be a problem and should be avoided. Fortunately map measurements can be done in a relatively short period of time. So the underlying map could be sufficiently stable. In general, such disturbances are machine and site dependent. I will not discuss them further. In experiments, one should check the stability of an unperturbed beam to identify possible problems, and furthermore, one should check the stability of the map measurements themselves.

5.6 Convergence of map fitting

Now consider the map fitting problem. First I discuss to which order one should fit. A more general concern is whether a sufficiently accurate one-turn Taylor map can be obtained to predict particle behavior after thousands or millions of turns. I will not look into this general question here. It suffices to point out that many numerical studies[76, 75] support the assumption that a 9-th order Taylor map is adequate for the stability studies in an electron machine. In the LER ring example, a 9-th order Taylor map has been used to generate tracking data and compare with the numerical fits. Since the 9-th order is about the highest order which may have some effect on beam dynamics, as a bottom line, a 9-th order fit map can be used. In fact, the simulation studies show that it is possible to work with a much lower order fit map and the fitted maps will converge to the correct one (the 9-th order tracked map) as the fit-map order increases. The convergence of the fitted maps is a criterion to guarantee that the correct map is obtained. (See Section 5.8 for more discussions.)

Table 5.2 shows the results of fitting simulated data to maps of increasingly higher orders. No random errors have been added to the data pairs yet. Since there are hundreds of coefficients with magnitudes spanning several orders, I have not shown the fitting results for each coefficient. Instead the maximum fitting errors ($|C_i^{fit} - C_i^{exact}|$) are tabulated for different orders in rows and for different fit-map orders in columns. For reference, the maximum coefficients of the original map at various orders are also included in the last row. Since the coefficients are for dynamical variables normalized to 10σ, they are dimensionless. For dynamical aperture concerns, values less than 10^{-3} are not significant because a particle must be lost within 10^3 turns without radiation damping, otherwise the growth will be damped.[23] This is why I present absolute fitting errors instead of relative errors. We see that the fitting is consistently improving and converges to the original map as the fit-map order increases. It also shows that to get a map to a certain order correctly, the fit-map order had better be 2 orders higher than the coefficient order (though in general this may be map dependent). The important point is that a sufficiently high order map must be used in fitting, even though only low-order terms are of interest.

fit-map order	orders					
	1-st	2-nd	3-rd	4-th	5-th	6-th
1	1.7×10^{-2}					
2	1.5×10^{-2}	4.4×10^{-2}				
3	1.1×10^{-3}	4.6×10^{-2}	2.0×10^{-2}			
4	5.1×10^{-4}	2.7×10^{-3}	2.2×10^{-2}	2.6×10^{-2}		
5	1.2×10^{-4}	1.5×10^{-3}	3.4×10^{-3}	2.9×10^{-2}	2.4×10^{-2}	
6	5.5×10^{-5}	1.0×10^{-4}	2.2×10^{-3}	2.0×10^{-3}	2.4×10^{-2}	4.1×10^{-2}
max. coeff.	0.86	0.10	0.27	0.25	0.21	0.23

Table 5.2: Maximum absolute fitting error dependency on fit-map order. The tracked map was 9-th order.

Now I will address the second apparent concern. For a high order map, there are hundreds or even thousands of coefficients. How is it possible (Does it make sense?[6]) to fit so many parameters? I have to admit that such a possibility is indeed somewhat surprising, but the answer is affirmative. At least the method works well to fit significant coefficients. Another question that is closely related to the convergence problem is the domain dependency. Here, I will not address this issue in detail. Generally speaking, the larger the phase space domain used, the higher the map order that is required. In case only low-order map coefficients are of interest, one may not want to kick the beam to 10σ. However, one should be aware that small beam amplitudes will make BPM errors more significant. In experiments, it is necessary to optimize the map order and phase space domain to be used under given conditions such as the expected (the design) map, beam aperture, BPM resolution, etc. In this chapter, I choose 8σ of the transverse distribution and 4σ energy distribution in simulations. Usually, there is larger than 12σ clearance in accelerators.

5.7 Dependency on the BPM resolution and energy resolution

Since we are trying to determine the map via a large set of data pairs, obviously the accuracy of these data pairs determines how well the map is defined and how well the data can be fit by a map. There are two concerns with BPMs. One is the linearity of the BPM. Since we are measuring small nonlinear effects, the BPM itself must be sufficiently linear or its nonlinearity must be known and extracted. Otherwise, the measured data set will consist of the nonlinearity due to BPMs as well as beam dynamics. I will not discuss details about BPM nonlinearity, which should be negligible in well-calibrated BPMs[7, 65]. Note that BPM nonlinearity which is significantly above the BPM noise level will ruin map measurements and yield false information, while low BPM resolution will limit our ability to extract the information. Limitations will be revealed during data analysis.

The other concern is BPM resolution, which is often limited by its signal digitizer. Although such BPM data error often appears as uncertainty in the least-significant bit instead of being continuously distributed, I add uniformly or Gaussian distributed random errors to the data pairs generated with the LER map to simulate BPM resolution effects, since the least-squares fitting is not very sensitive to the error distribution. To fit the data (generated with a 9-th order map), I have used a 5-th order Taylor map (instead of 9-th order) in order to show that low-order maps can be used to fit high-order maps as long as high-order terms are weak.

BPM res. in μm	orders				
	1-st	2-nd	3-rd	4-th	5-th
100	5.3×10^{-3}	4.4×10^{-3}	1.1×10^{-1}	5.0×10^{-2}	4.4×10^{-1}
10	2.2×10^{-4}	1.1×10^{-3}	5.4×10^{-3}	3.6×10^{-2}	2.8×10^{-2}
1	4.3×10^{-5}	1.6×10^{-3}	2.1×10^{-3}	3.7×10^{-2}	3.2×10^{-2}
0.1	6.7×10^{-5}	1.6×10^{-3}	1.6×10^{-3}	3.8×10^{-2}	2.9×10^{-2}

Table 5.3: Maximum absolute fitting error dependency on BPM resolution. Coefficients were normalized to 10σ. Tracked map was 9-th order.

Table 5.3 presents the simulation results with BPM resolutions of $100\mu m$, $10\mu m$, $1\mu m$, and $0.1\mu m$. We see that the possibility to extract nonlinear coefficients is dramatically improved from $100\mu m$ to $10\mu m$, while roughly constant for resolutions better than $10\mu m$. Comparing Table 5.3 to Table 5.2 shows that below $10\mu m$, BPM errors are not significant and fitting errors (due to the low order fit-map used) dominate.

Fig. 5.2 plots the fitting errors of each coefficient for the $10\,\mu m$ case, which is the hoped for PEP-II. The coefficient indices are grouped by orders with boundaries at 5, 20, 55, 125, 251 for the 1st, 2nd, etc. orders respectively. It is exciting to

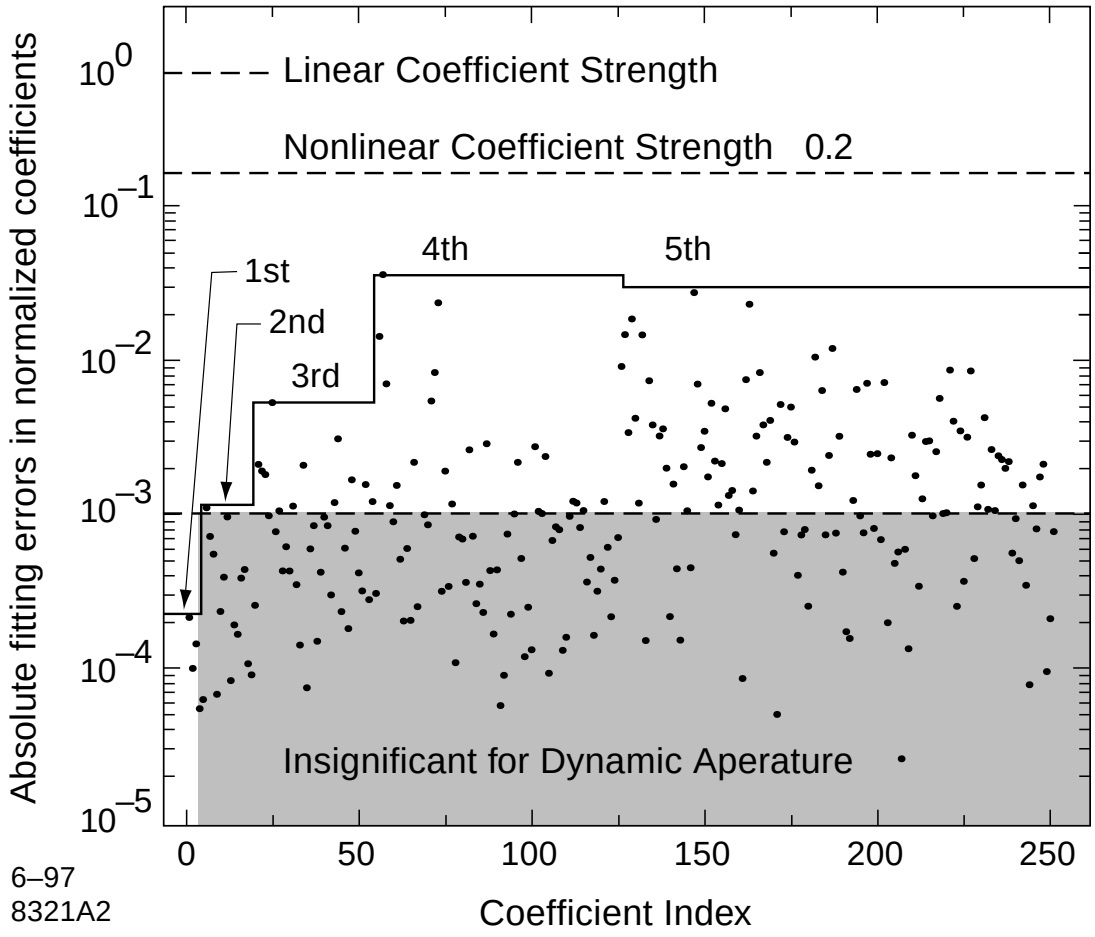

Figure 5.2: Absolute fitting errors for 10σ normalized map coefficients up to 5-th order

realize that with $10\mu m$ BPM resolution, it is possible to measure the map to an unprecedented high order and accuracy. The data in Tables 5.2 and 5.3 demonstrate that map measurements should be capable of measuring significant nonlinearities, which is very important for accelerator diagnosis.

Although present technology is able to reach $10\mu m$ (if not much better) BPM resolution, few machines offer such turn-by-turn resolution at this moment because the technology is relatively new and budget is always a concern. Therefore, for the time being, BPM resolution is the major obstacle for map measurements. In Chapter 6, I will show how to statistically improve phase-space measurements by using a large number of BPMs.

In addition to BPM resolutions, energy resolution is also important. I assumed no δ errors in Table 5.3. Dependency on beam energy resolution for the $10\mu m$ case is tabulated in Table 5.4. In order to simulate the measurements, I used 15 equally spaced grid points (instead of random sampling) to cover the $\pm 4\sigma_\delta$ range. Table 5.4 shows that, to be consistent with the $10\mu m$ BPM resolution, we need energy resolution on the order of 5×10^{-5}. This corresponds to a signal-to-noise ratio of 50, which is much less than the requirement for transverse measurement. This is because we have to extract nonlinear motion out of fairly linear betatron oscillation transversely, while the linear and nonlinear dependencies on beam energy are comparable (the LER map I am using originates in a low dispersion region). Usually it is difficult to measure the beam energy accurately. Methods using a depolarizing resonance[10] or synchrotron radiation[49] are sufficiently accurate, but need special instruments. A method based on BPM measurements at high dispersion points[52] may be used to

	orders				
δ res.	1-st	2-nd	3-rd	4-th	5-th
6×10^{-6}	1.8×10^{-4}	2.2×10^{-3}	4.9×10^{-3}	3.8×10^{-2}	5.1×10^{-2}
6×10^{-5}	4.9×10^{-4}	1.6×10^{-3}	5.7×10^{-3}	3.0×10^{-2}	3.8×10^{-2}
3×10^{-4}	9.4×10^{-4}	8.8×10^{-3}	3.0×10^{-3}	4.6×10^{-2}	2.1×10^{-1}
6×10^{-4}	3.7×10^{-3}	2.1×10^{-3}	1.0×10^{-1}	3.9×10^{-2}	5.3×10^{-1}

Table 5.4: Maximum absolute fitting error dependency on beam energy resolution.

measure the relative energy change, but special care is needed in order to reach the required resolution. Factors that are typically omitted such as nonlinear dispersion coefficients may become significant. For the LER example, the dispersion function reads $\eta_1(1 - 1.9 \times 10^2\delta - 2.4 \times 10^4\delta^2 + 6.5 \times 10^5\delta^3 + 7.1 \times 10^8\delta^4 + \cdots)$ and $4\sigma_\delta$ gives $\delta_{max} \simeq 3 \times 10^{-3}$, which could yield up to 60% error (here the nonlinear effect is exaggerated by the low linear dispersion) when using only the leading term. MIA methods developed in Chapter 3 can significantly improve turn-by-turn relative energy measurement.

Instead of measuring the beam energy, Eq.(5.3) provides a way to change the beam energy accurately. When the RF frequency is changed, the beam energy is forced to change according to Eq.(5.3) due to phase stability (assuming significant synchrotron oscillations are fully damped by radiation damping). The relative RF frequency change resolution is better than 10^{-9}, and a smooth approximation gives $\alpha_c \simeq \nu^{-2} \simeq 10^{-3}$, thus the accuracy of energy change depends on the linear and nonlinear momentum compaction factors. In general, the nonlinear momentum compaction factors are very weak. For the LER we have $\alpha_c(1+1.6\delta-45\delta^2+1.4\times10^3\delta^3+7.1\times10^5\delta^4-3.2\times10^7\delta^5-3.1\times10^9\delta^6+\cdots)$. Therefore, we can drop the nonlinear part and still get fairly good resolution. However, we have to know the linear momentum compaction factor α_c to get the energy change correctly. To get a signal-to-noise ratio of 50, the error in α_c should be less than 2%. Unfortunately, we may not know α_c sufficiently well and this may result in significant errors in the map coefficients. Nonetheless, notice that α_c is a scaling constant for the δ dimension, we still can get useful nonlinear information even though we do not know α_c exactly.

One more subtlety on beam energy is the calibration of the nominal energy. In order to compare with the design map, an absolute energy measurement is required. On the other hand, lattice magnetic field errors may dominate the difference between the design and measured maps.[5] Therefore it may not be necessary to know the beam energy better than $\Delta B_{err}/B$. However, better energy calibration will help to resolve lattice (closed orbit) errors with measured maps at various locations. Note that since

[5]One experimental way to define the nominal beam energy is to find the energy at which the closed orbit is centered at all sextupoles on average. This can be done by measuring the chromatic tune shifts at various sextupole strengths.

the closed orbits are changing with beam energy, one must use the closed orbit at the nominal energy as a reference for all BPM readings. Also, for very high accuracy measurements, the energy difference around the ring due to synchrotron radiation may need to be taken into account, which is on the order of 10^{-4} and could result in 10^{-2} errors at 5-th order coefficients.

Before moving on to the next section, I present the exact (in diamond) and fitted (in dot) map coefficients in Fig. 5.3, which is the case of 10μm BPM resolution and 6×10^{-5} energy resolution. Except for the large leading linear coefficient (which is well fit), all coefficients up to 5-th order are shown in the figure. It shows that the map fitting method is very impressive in its potential for extracting nonlinear coefficients.

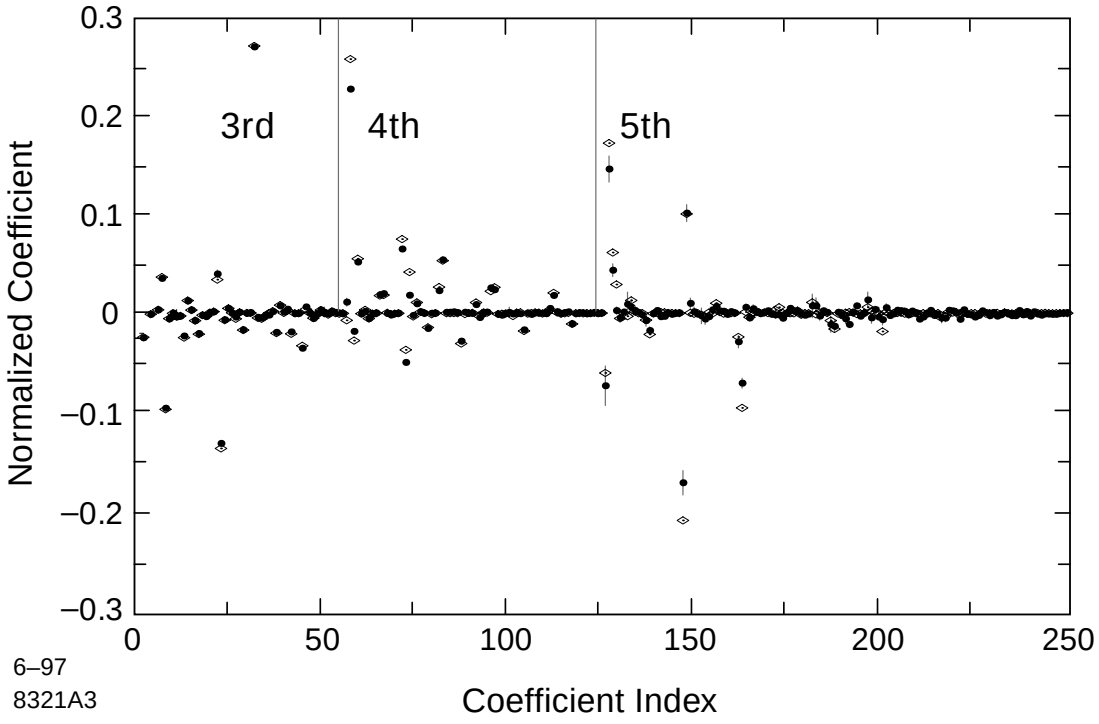

Figure 5.3: Comparison between exact (◇) and fitted (●) map coefficients

5.8 Error estimate and hypothesis test

For any measurement, it is important to obtain error estimates for the results. Besides factors affecting the existence and stability of the single-particle map, errors are mainly due to the BPM resolution and our hypothesis on the map order in the fitting procedure. BPM errors show up in the vector $\vec{x^f}$ and matrix A of Eq.(5.5). Errors in $\vec{x^f}$ are typical in least-squares fitting problems, whose effect can be estimated by the variance-covariance matrix [40]

$$\frac{V^T V}{m-n}(A^T A)^{-1} \tag{5.16}$$

where vector $V = \vec{x^f} - A\hat{C}$ (see Eq.5.5) and V^T is its transpose. Square roots of the diagonal elements of the covariance matrix yield estimated rms errors of map coefficients, which are shown in Fig. 5.3 as error bars. Due to various reasons (eg. badly conditioned matrix A, measurement errors in A, relatively low fit-map order), the error bars given by Eq.(5.16) are not very good. A more reliable error estimate is under investigation. The factor $\hat{\sigma}^2 \equiv V^T V/(m-n)$ gives an unbiased estimate of the rms errors $\hat{\sigma}$ in measurement of $\vec{x^f}$, i.e. the BPM resolution, if the design matrix A is accurate. In our case, this is not exactly true. However, significant error bars for all coefficients will indicate problems due to BPM resolution.

In general, this kind of error can be reduced statistically with the usual $\frac{1}{\sqrt{m}}$ dependency by increasing the number of data pairs m since the expectation value of $A^T A$ is proportional to m. However, due to the errors discussed below, the accuracy of map extraction does not depend on Eq.(5.16) any more for sufficiently large m. Therefore, increasing the data size may be helpful only to some limited extent. Simulations presented before use a data size which is 10 times the number of coefficients to fit. As an example, Table 5.5 shows the dependency on m for the case of $10\mu m$ BPM resolution and a 5-th order fit map. The first column lists m/n, i.e. the number of data pairs in units of the number of coefficients to fit.

The rms error estimate $\hat{\sigma}$ has another important application in our case. Since we measure a large number of data pairs over a period of time, it is possible to have

$m/250$	orders				
	1-st	2-nd	3-rd	4-th	5-th
4	1.1×10^{-3}	3.1×10^{-3}	1.4×10^{-2}	2.8×10^{-2}	9.8×10^{-2}
16	5.3×10^{-4}	1.9×10^{-3}	7.7×10^{-3}	2.9×10^{-2}	5.6×10^{-2}
64	1.3×10^{-4}	2.7×10^{-3}	2.6×10^{-3}	2.9×10^{-2}	2.1×10^{-2}
256	9.1×10^{-5}	1.7×10^{-3}	1.7×10^{-3}	3.0×10^{-2}	2.7×10^{-2}

Table 5.5: Maximum absolute fitting error dependency on number of data pairs. The 5-th order fit map has 250 terms.

some very bad data due to accidental events. Such data could significantly reduce fitting accuracy. One way to filter out such data is to check the fitting error of each data pair against $\hat{\sigma}$, and throw away data with an unreasonably large error. After such filtering, fit the map again as necessary.

Errors in the matrix A are more complicated to evaluate because BPM errors show up nonlinearly and their effects are map dependent. Using a perturbative approach, the estimate of Eq.(5.16) can be improved by adding

$$-2\,\hat{\sigma}^2 A^+ \,\mathrm{diag}[A_x \hat{C}]\,(A^+)^T \;+\; A^+ \sum_{\xi=\{x,P_x,\cdots\}} \hat{\sigma}_\xi^2 \,\mathrm{diag}[(A_\xi \hat{C})^2]\,(A^+)^T \tag{5.17}$$

where A_ξ characterizes the linear dependence of A on errors in ξ and can be computed from measured data. However, I will not go into the technical details of Eq.(5.17) here. Another possible, although primitive, way to improve the error estimate is to choose many (they may overlap) subsets of the original data set, and obtain different fittings of the coefficients, which indicate the possible error. In this way one gets a rough estimate of the maximum fitting errors for the $10\mu m$ BPM resolution and 6×10^{-5} energy resolution case. We find $3.9(\pm 2.3) \times 10^{-4}$, $1.6(\pm 0.6) \times 10^{-3}$, $8.3(\pm 5.7) \times 10^{-3}$, $3.7(\pm 0.4) \times 10^{-2}$, and $4.8(\pm 2.1) \times 10^{-2}$ for the 1-st thorough 5-th orders respectively. This method can also be used to check the consistency of the data. For example, the data collected at the beginning and end of the whole measurement may yield significantly different maps if there is slow (adiabatic) variation of the underlying map due to external perturbations.

Errors due to map-order hypothesis have been shown in Section 5.6. They are map dependent and hard to estimate. Checking the convergence of different order fit maps is necessary. Another potentially useful method is singular value decomposition (SVD). By checking the magnitudes of singular values, one may determine at which order the fitting is not sensitive to its coefficients anymore. One may attempt to use a statistical hypothesis test also. Note that the (quasi-) symplecticity of the map implies nonlinear constraints among the coefficients, which makes usual hypothesis test routines inappropriate.

In the above sections, the 5D phase space was treated equally and simultaneously. It is possible to break down the problem into lower dimensions, especially for the δ degree of freedom since it is not changed by the map. The advantage is that the parameter space for fitting can be reduced to many fittings in subspaces instead of a single step fitting in 5D. This could be important since one may need to fit thousands (2000 for 9-th order) of coefficients, which requires handling huge matrices. It is actually more natural to work with 4+1 dimensions if synchrotron oscillation is negligible. However, it is not expected to be significantly different in terms of fitting results. Further studies are required to settle this issue. The possible drawback to breaking down the dimensions is that the process of separating subspaces is likely to be an additional error source for fitting due to unavoidable coupling, even though it may be very weak in well tuned machines.

Chapter 6

MIA applications to rings—PEP-II example

In this chapter I will consider the applications of MIA in rings focusing on the measurement of nonlinear transformation maps. In the last chapter I argued that all the potential physical effects such as decoherence, wakefields, etc. can be controlled and it appears possible to measure nonlinear transformation maps with good accuracy. The major difficulty arises from limited BPM resolutions. However, Chapter 3 showed that MIA can reduce the BPM random noise significantly. Therefore MIA can facilitate/accomodate nonlinear map measurements. In the following I will explore this possibility and show the details of nonlinear map measurements via MIA. First I discuss how to accommodate general nonlinear map measurements in an MIA analysis by viewing the map coefficients as a physical basis; then I discuss how to significantly improve measurement accuracy of the phase-space dynamical variables, which is crucial for map measurements; finally I present some simulation results for the PEP-II high energy ring.[74, 73, 45]

6.1 Application of MIA to nonlinear (one-turn) map measurements

The single particle beam dynamics can be represented by the transformation map $\mathcal{M}_{a\to b}$ that maps any initial phase space point X^a at location a to a phase space point X^b at location b. Using a Taylor map representation, $\mathcal{M}_{a\to b}$ may be written as

$$X_k^b = C_k^b + R_{kl}^{a\to b} X_l^a + T_{klm}^{a\to b} X_l^a X_m^a + U_{klmp}^{a\to b} X_l^a X_m^a X_p^a + V_{klmpq}^{a\to b} X_l^a X_m^a X_p^a X_q^a + \cdots \quad (6.1)$$

where the summation convention on the repeated indices is assumed. X_k is the k-th component of a phase-space vector. C_k^b is the zero-order term of the k-th component and should vanish if the closed orbit is chosen as a reference. $R^{a\to b}$, $T^{a\to b}$, $U^{a\to b}$, and $V^{a\to b}$ are the usual TRANSPORT notations for the 1st, 2nd, 3rd, and 4th order map coefficients.

Now consider BPM readings for P pulses (turns) at M BPM locations b_1, b_2, $\cdots$, b_M in a ring. Obviously we can cast the BPM-data matrix B for the horizontal plane into a physical base decomposition via Eq.(6.1),

$$B = QF^T \quad (6.2)$$

where F consists of all coefficients in the x components of the maps as the physical basis (see Fig. 6.9 for some examples) and Q contains the corresponding initial conditions, such as

$$F = \begin{bmatrix} C_1^{b_1} & R_{11}^{a\to b_1} & R_{12}^{a\to b_1} & \cdots & T_{111}^{a\to b_1} & \cdots \\ C_1^{b_2} & R_{11}^{a\to b_2} & R_{12}^{a\to b_2} & \cdots & T_{111}^{a\to b_2} & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \\ C_1^{b_M} & R_{11}^{a\to b_M} & R_{12}^{a\to b_M} & \cdots & T_{111}^{a\to b_M} & \cdots \end{bmatrix}$$

and

$$Q = \begin{bmatrix} 1 & x_1^a & p_{x_1}^a & \cdots & (x_1^a)^2 & \cdots \\ 1 & x_2^a & p_{x_2}^a & \cdots & (x_2^a)^2 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \\ 1 & x_P^a & p_{x_P}^a & \cdots & (x_P^a)^2 & \cdots \end{bmatrix}.$$

Now the constant terms, the C_1^b's, may contain BPM offsets also. The best way to take out such terms is to use the measured closed orbit as the reference and use the difference orbits to construct B. It is better not to use the average orbit as the reference because the high order terms may not average to zero and yield significant errors, when the beam is excited to the large amplitudes necessary to measure the high order map coefficients. On the other hand, it is possible to get a very accurate closed orbit in a ring by averaging over a large number of turns of the unperturbed beam. As discussed in Section 3.2, all the column vectors (linear and nonlinear terms) in Q are linearly independent of each other.

Similarly one can construct a physical base decomposition for the vertical BPM readings with the y components of the maps. However, to get the x' and y' (p_x and p_y) components of the maps, this information would be required at each BPM. Note that the Q matrix is the same for all components. In fact, one can stack the BPM readings for both planes into one matrix and extend the physical base decomposition to cover both planes.

One particularly interesting nonlinear map is the one-turn map of a ring. To accommodate this into Eq.(6.2), F should contain the one-turn map coefficients and B should contain the measured phase-space variables one-turn after the initial values used in Q. Section 5.2 presents a concrete example. Note that in this case, F does not contain any BPM patterns as in the usual MIA application. However, the Q matrix is still the same.

Standard least-squares fitting can be used to solve Eq.(6.2) for the map coefficients. The difficulty is to get a sufficiently accurate B and Q. The next section will focus on the measurement of phase-space variables.

6.2 Phase-space variable measurement and noise reduction

To measure the weak nonlinearity present in the phase-space dynamics of a ring, the background has to be sufficiently clean. MIA provides a nice way to check this requirement. First obtain a BPM-data matrix $B_{P\times M}$ by recording a large number (e.g. $P = 5000$) of turn-by-turn data at all available BPMs (e.g. $M = 150$) with the stored beam unperturbed. Then check the singular value spectrum of B. Ideally it should contain only the BPM noise floor, since all physical motions should be well damped. Noisy BPMs can be easily identified at this stage. Any other significant modes indicate systematic BPM errors and/or physical sources exciting the beam. Such problems need to be fixed in order to pursue nonlinear map measurements. I assume the singular value spectrum is clean. We should take the average orbit as the reference orbit and identify the noise level for later use.

Now measure another BPM-data matrix with large (e.g. 8σ) betatron oscillations excited by fast kickers for example. Subtract the mentioned reference orbit from each measured orbit in order to get rid of BPM offsets and define the expansion points of the measured maps. At this stage, two MIA procedures can be employed to improve the phase-space dynamics measurement: SVD noise reduction (Section 3.5) and degrees-of-freedom analysis (Section 3.6).

To reduce the random noise, compute a Singular Value Decomposition (SVD) as $B = USV^T$, and identify the noise floor or use the noise level mentioned above. Set the corresponding noise singular values to zeroes, and then re-multiply these matrices to construct a noise-cut matrix $\underline{B}$. This simple procedure can reduce the BPM random noise by a factor of $\sqrt{\frac{d}{M}}$, where d is the number of remaining singular values above the noise floor. Depending on the situation, such noise reduction could be rather significant, especially when the BPM resolution is poor and one is struggling to measure a few leading nonlinear coefficients.

Further improvement of phase-space-variable measurement is still possible by using a well-chosen BPM subset. (This is the BPMs used to compute the phase-space variables in Q after the above noise-cut, not the BPMs for the noise-cut.) A seemingly

enticing thought is to use all available BPMs instead of only a few ones in a linear section of the ring. The idea is to obtain accurate linear maps first by using small amplitude oscillations, and then use them to fit each trajectory including those with large oscillations to get the corresponding beam parameters. This should have the statistical benefit of $\frac{1}{\sqrt{M}}$ while the nonlinear effects were hoped to be negligible in this process. Unfortunately, such a scheme will not work well because the nonlinear effects are indeed significant. There is a way to salvage this scheme: instead of using all the BPMs, one can select those BPMs that suffer only small nonlinear effects. However, careful selection of BPMs is necessary in order to gain the statistical benefit and avoid the nonlinearity problem at the same time. It is best to choose a beam line section that has many BPMs and very weak nonlinearity. Simulation studies can be helpful in choosing the right BPMs.

Although the design lattice can be used as a guide to avoid major nonlinearities, judgement based on observation is important. How can one tell where the nonlinearity is negligible before the nonlinear coefficients are measured? The degrees-of-freedom analysis can help locate the best linear section in a ring for phase-space-variable measurements. The criteria can be imposed that only the two betatron degrees of freedom should exist and the other modes be below the noise floor. Note that in a usual degrees-of-freedom plot, the degrees of freedom always increase with the number of BPMs used. If a linear section exists after some nonlinear sections instead of at the beginning of the beam line, it would appear that there were not just betatron modes in the linear section. However, the singular values of all the non-betatron modes will not increase in a linear section. One has to be careful when examining a degrees-of-freedom plot, and sometimes it is better to look at the increments of singular values.

The measurement process can be separated into two parts: 1) measure a full set of independent orbits to the best possible accuracy; 2) determine the desired phase-space variables x, x', etc. from the measured independent orbits. In another word, one is to establish a basis (such as the orthogonal basis from SVD) so that the dynamics can be accurately described, the other is to connect that basis to the desired "physical" basis (such as x and x') so that the dynamics can be better understood and compared

with models. Step 1 can be achieved without any model, while Step 2 has to rely on some extra model dependent knowledge. The noise reduction methods discussed above work for the first step. Conventionally these two steps are mixed together, thus errors due to a poor model and errors due to noise are confused. However, it is useful to separate these two types of errors, because the second one is often somewhat artificial. In principle, one can use the orthogonal linear modes as the phase-space variables. However, it is probably better to use a model (e.g. the design R matrix between two BPMs) of the mentioned linear section to define the phase-space variables from orbit measurements. Note that even if the linear machine model might not be sufficiently accurate, it will not affect the sensitivity of nonlinear map measurement. It is like using a pair of glasses which is perfectly clear but may generate a little bit distortion.

I have outlined the procedure to measure beam phase-space dynamics via MIA and methods to compute various nonlinear maps of a ring. The following section presents some simulation studies.

6.3 Simulations for PEP-II high energy ring

Simulations for the PEP-II high energy ring have been carried out to investigate the feasibility of nonlinear map measurements. 5000 turn data at all BPMs (147 for each plane) were generated by tracking 200 turns of 25 randomly chosen x and y initial conditions within a 10σ range. (For simplicity, there is no energy change; however the procedure and results should be similar.) Figure 6.1 shows the major 2D subspaces of the 4D phase-space samples used in the simulations. In addition, various levels of random noise were added in order to test the sensitivity of the map measurement and the effects of noise reduction described in Section 6.2. BPM resolutions are randomly selected from the specified ranges.

Fig. 6.2 shows the singular value spectra of the simulated system. The insertions show the full scale for both planes. Linear coupling and nonlinear modes are orders of magnitude weaker than the two dominant betatron modes—the reason nonlinear map measurements are challenging. The main frame shows the tail part of the x spectra

in detail. Circles are signals only, dots are $100 \pm 20\mu m$ BPM noise only, and crosses are signals with noise. Note that the noise floor is about 10 μm, much lower than the individual BPM resolutions—the statistical benefit. Otherwise, all the nonlinear signals would be below the 100 μm noise level. The arrow indicates where to cut off the noise floor.

Before focusing on the measurements of phase-space variables and nonlinear maps, let us check out the spatial vectors of SVD analysis. The results for the x BPMs are plotted in Fig. 6.3, where the first 14 vectors are plotted. The labels on the left show the corresponding singular values in μm. These vectors should be compared with the map coefficients shown in Fig. 6.9. We can see that some of the dominant features show up in the SVD analysis. However, as we addressed in Chapter 3, the orthogonal basis obtained from SVD is not meant to be the same as physical basis.

Table 6.1 shows the rms errors of dynamical variable measurements for various BPM resolutions and the effects of noise-cut. Figure 6.4 plots these data with linear fits. Despite noticeable fluctuations, the accuracy of phase-space measurements is significantly improved. Here only two BPMs are used to compute the phase-space variables from the SVD noise-cut data.

To investigate the possibility of improving measurement accuracy by using more BPMs in fitting, Fig. 6.5 shows the errors due to nonlinearity. In this study, there is no BPM noise. The top frame plots the initial x values of the 5000 pulses used in the simulation. The other frames plot the errors of the fitted x obtained by using the exact linear transfer maps and various number of BPMs as indicated in the figure caption. The errors clearly correspond to the amplitudes of the oscillations and are

BPM resolutions (μm)	**without noise-cut**				**with noise-cut**			
	$x(\mu m)$	$x'(\mu r)$	y	y'	x	x'	y	y'
100±20	102	7.1	156	14.	36.	2.6	25.	2.1
80±16	74.	6.4	126	11.	17.	2.1	24.	2.0
60±12	55.	4.9	103	9.3	12.	2.4	20.	1.8
40±8	30.	2.3	56.	5.1	8.2	1.2	14.	1.3
20±4	20.	1.7	30.	3.4	10.	1.1	11.	1.3

Table 6.1: rms errors of phase-space-variable measurements for various BPM resolutions with and without SVD noist-cut.

not symmetrically distributed. Except the cases using BPMs within the first 10, the rms magnitude of the nonlinear errors is over 20 μm, which is quite significant. The amplitude dependency of the errors is probably more detrimental than the same amplitude random noise. In addition to the cases shown, many other BPM selections are tested and the results are similar to using all BPMs. This study demonstrates the shortcoming of using a large number of BPMs to compute the phase-space variables. In general, errors due to nonlinearity are not negligible.

Fig. 6.6 repeats the above study but with $100 \pm 20 \mu m$ BPM noise. As hoped, the random noise levels are considerably reduced when a large number of BPMs are used. However, the nonlinear errors prevent us from taking advantage of such statistical benefits. The last case is an exception, where the first 10 BPMs are used and there is not much nonlinearity. In this case, a factor of $\sqrt{10}$ can be gained. Fig. 6.7 continues this study by applying a noise-cut. The improvement is obvious. The highlights of the above noise reduction methods are shown in Table 6.2, which clearly demonstrates the statistical ability to reduce the random noise, which can make nonlinear map measurements feasible without stringent BPM resolution requirements, especially in rings with a very large number of BPMs. Let us summarize what can be learned from the results shown in Figs. 6.5–6.7:

- SVD noise reduction is generally applicable and effective as long as the number of BPMs is much larger than the number of significant nonlinear modes.

- **If** the exact linear transfer maps are known, the improvement of using a large number of BPMs to fit the phase-space variables is significant, even after SVD noise reduction. However, nonlinearity and the accuracy of the linear maps may limit the applicability of this method.

Method used	no reduction	SVD only	many BPMs only	both methods
Shown in frame	Fig. 6.6 , #2	Fig. 6.5 , #2	Fig. 6.6 , #6	Fig. 6.5 , #6
rms noise level	102 μm	36 μm	30 μm	17 μm

Table 6.2: Summary of phase-space measurement improvement via noise reduction.

- Nonlinear errors are generally not negligible. Nonetheless, it may be possible to take advantage of both noise-reduction methods and still avoid nonlinear errors. For this purpose, a subset of BPMs must be carefully selected.

As shown above, it is important to choose a linear section of a ring for the phase-space-variable measurements, especially when wishing to improve resolution and avoid the nonlinear errors at the same time. To find a good linear lattice section for phase-space measurement, a degrees-of-freedom analysis can be employed. Fig. 6.8 shows such analysis for the simulated data with 100 μm BPM noise. The two betatron modes clearly dominate in both planes. The big jump of the singular values at the center is due to the large beta function close to the interaction point. The blow-ups in the bottom frames show the modes due to nonlinearity. We see that the beginning 10 BPMs are in a fairly linear region, where nothing shows up above the 100 μm noise floor except the two betatron modes. That is why this section can be used to improve measurement accuracy as shown above. The section around BPM 50 is fairly linear in x but not in y.

Fig. 6.9 plots the linear and a few nonlinear map coefficients along all the BPMs. They are normalized to the 10σ of phase-space variables, thus reflecting the strength of each nonlinear term near the border of dynamical aperture. The apparent non-sinusoidal (amplitude $\sqrt{1+\alpha_0^2}$) patterns in R_{11} and R_{12} are due to the uneven BPM locations. The peaks in the 2nd order coefficients are due to the main sextupoles around the interaction point at the center. Such spatial patterns form the physical bases for the BPM readings. Simulation results are also shown for the case of 100 μm BPM resolutions, with and without noise-cut, on top of the exact solid curves. Although the main features can be obtained even without a noise-cut, the accuracy is significantly improved by the noise-cut, which is crucial. Note that the linear coupling terms R_{13} and R_{14} can be obtained rather accurately. Such information can be used to calculate the global linear coupling coefficient and furthermore help to localize the coupling sources. The errors in T_{113} are due to the weakness of this nonlinear coupling term.

Fig. 6.10 plots the results of simulated one-turn map measurement with 100 μm BPM resolution. Again the effect of noise-cut is obvious. The accuracy is sufficient

to reveal useful nonlinearity information. For example, a few percent of the main sextupole strength error should be detectable according to the simulations.

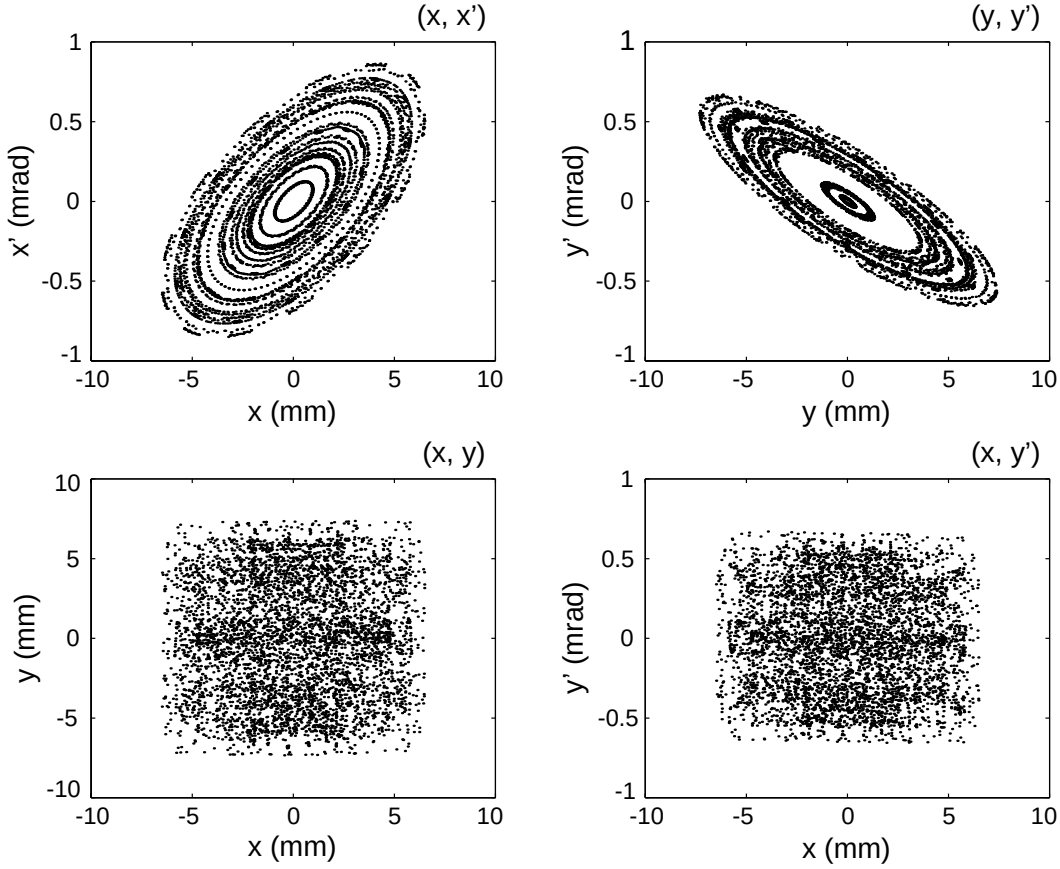

Figure 6.1: Subspaces of the 4D phase-space samples used in the simulations

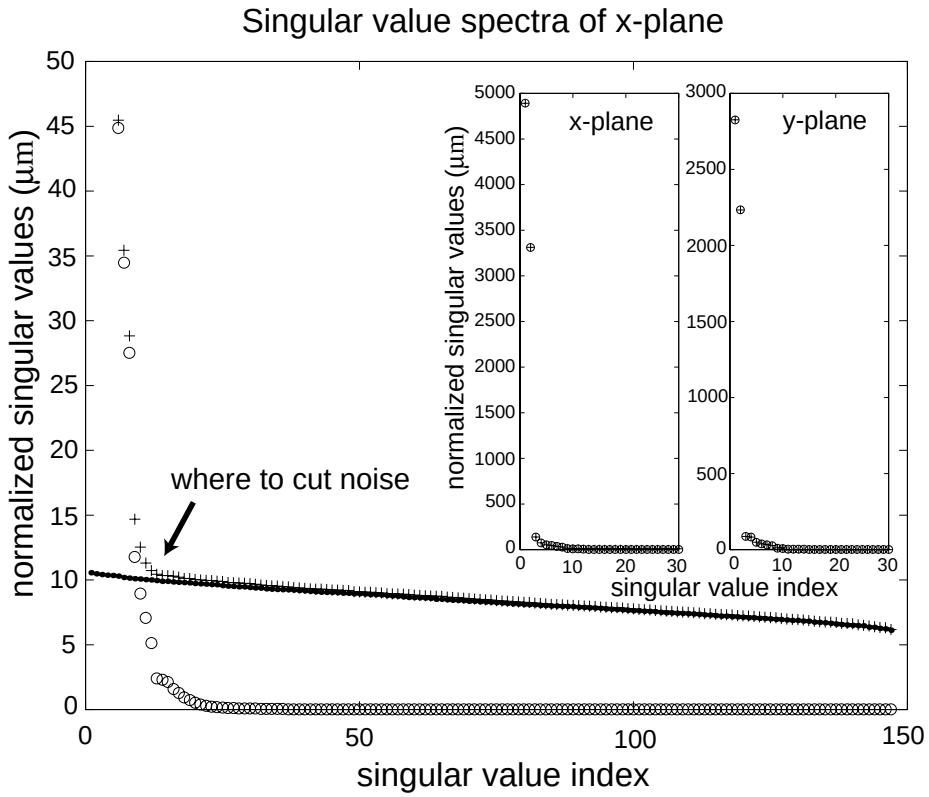

Figure 6.2: Singular value spectra of simulated data with "$\circ$" for signal only, "$\cdot$" for $100 \pm 20 \mu m$ BPM noise only, and "+" for both. Insertions show the full vertical scale.

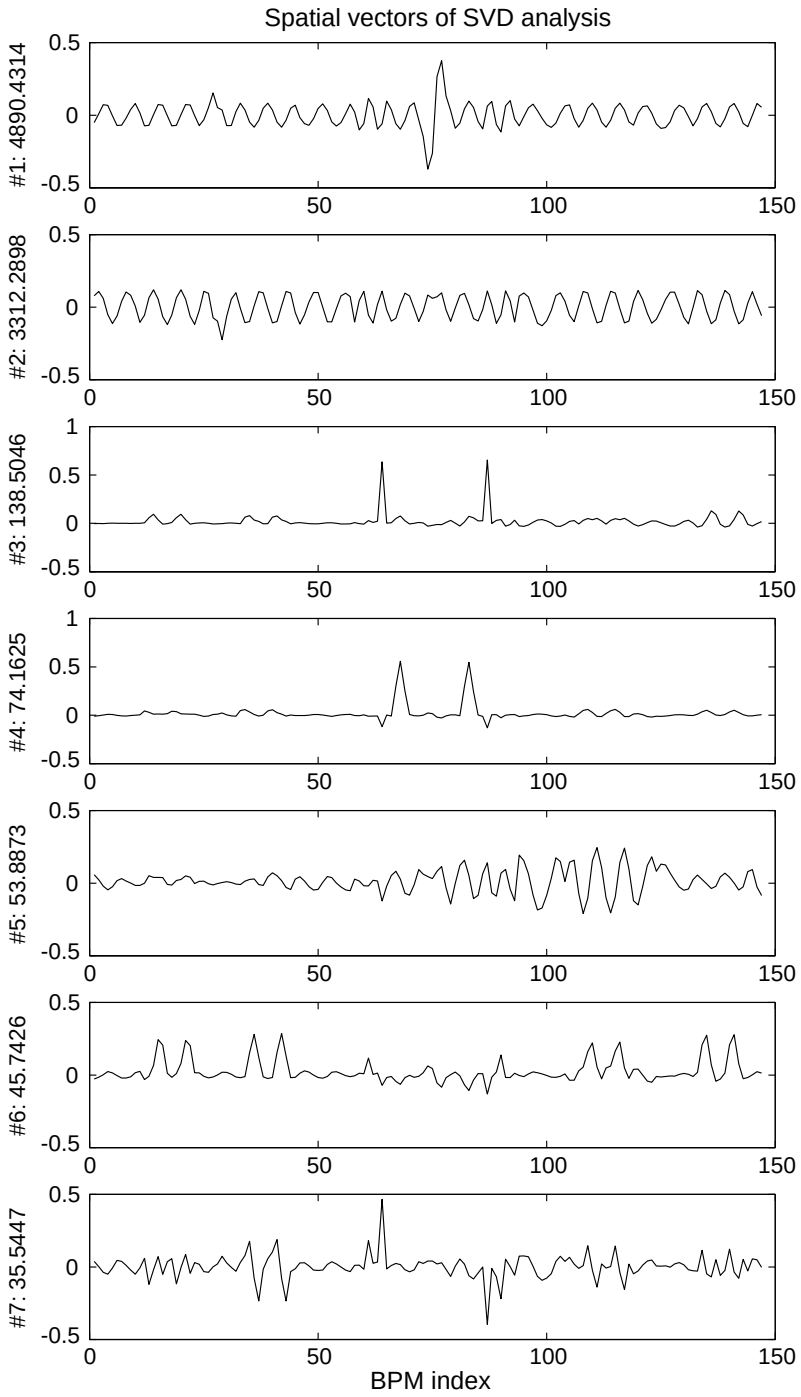

Figure 6.3: Spatial singular vectors of simulated data with $100 \pm 20\mu m$ BPM noise. Vertical labels are the singular value indices and the singular values in μm. (To be continued.)

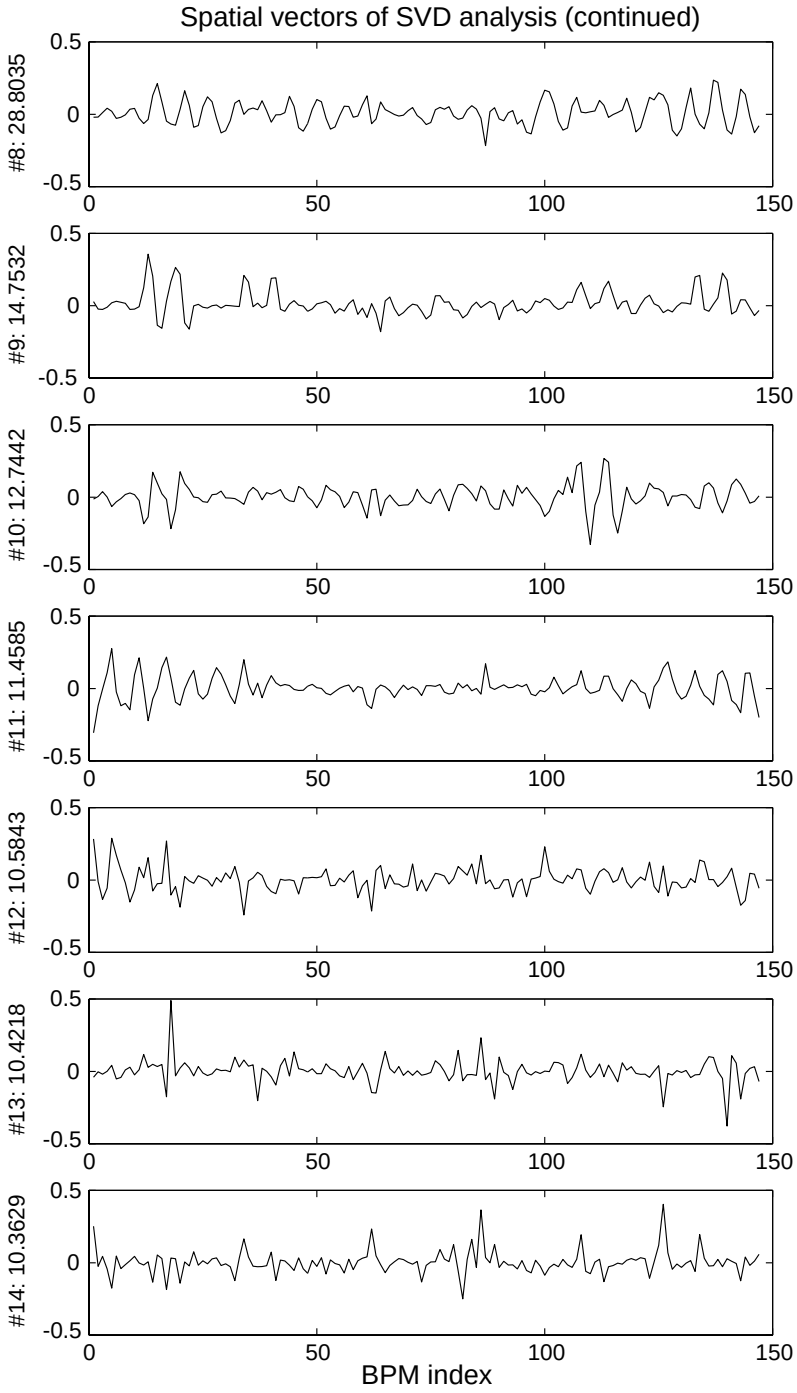

Figure 6.3: Continued.

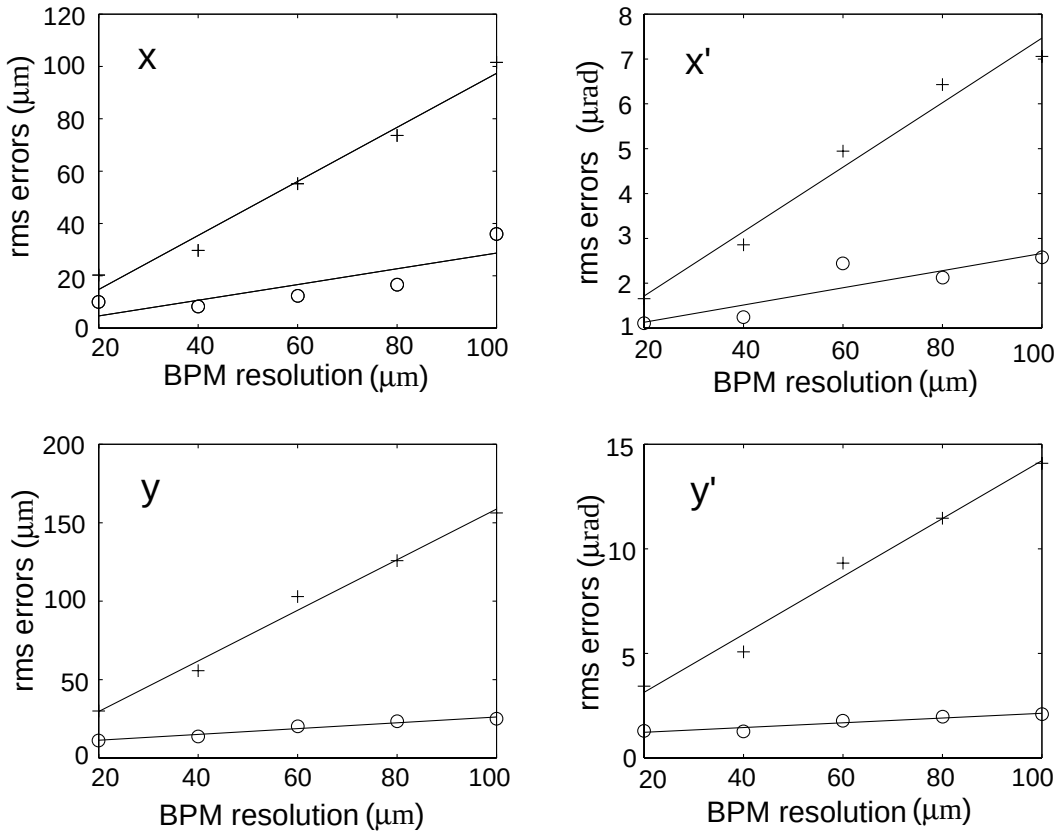

Figure 6.4: rms errors of phase-space-variable measurements with various BPM resolutions. "+" and "○" are without and with noise-cut respectively. Lines are linear fits.

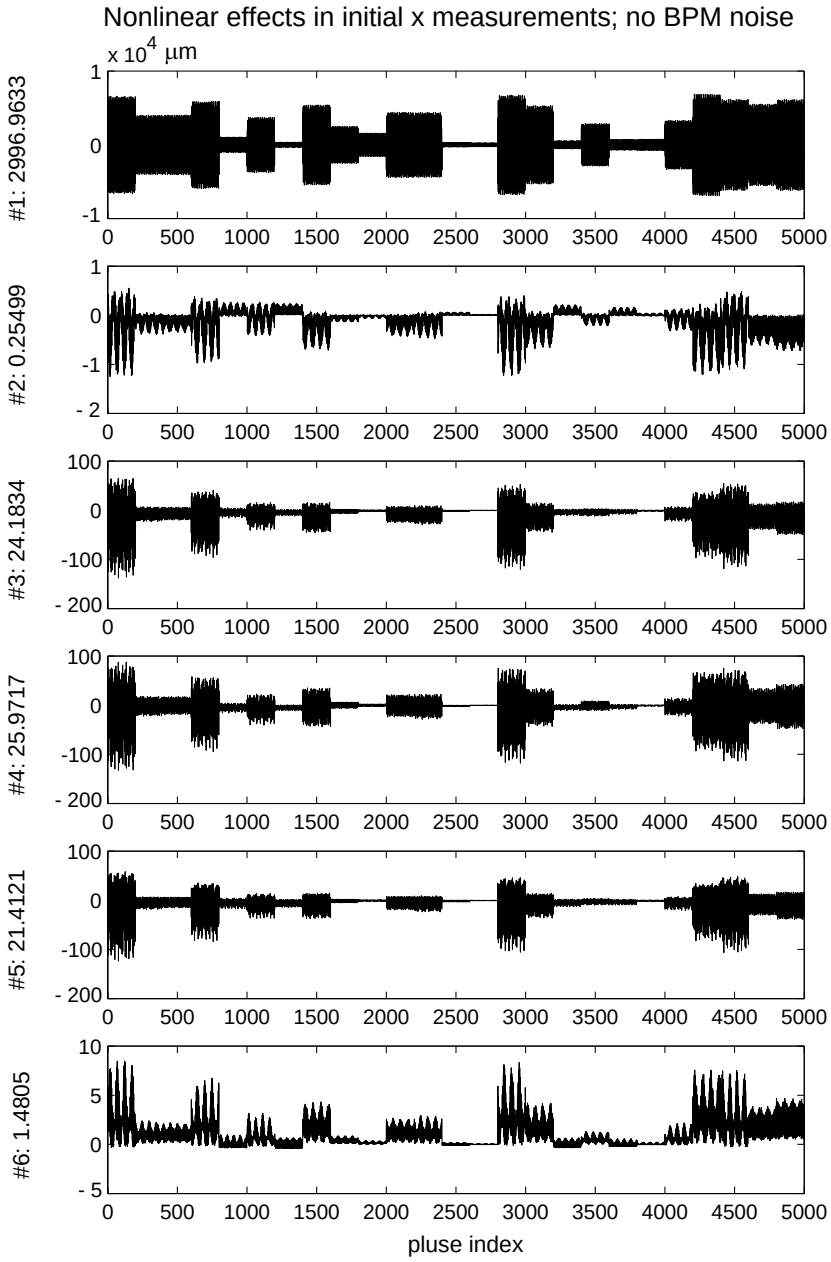

Figure 6.5: Nonlinear effect in phase-space-variable measurement. No BPM noise and exact linear maps are used. All vertical units are in μm. Frame #1 is the exact initial x values; #2 is the error of using the first 2 BPMs to fit x; #3 using all BPMs; #4 using all BPMs except 34 and 87 which seat on the peaks of the 2nd order coefficients; #5 using 63 BPMs that have small 2nd order coefficients; #6 using the first 10 BPMs, where the nonlinearity is rather weak. The rms values in μm for each case are listed on the vertical labels. The nonlinear effect is clear and significant. (See Fig. 6.9 for the distribution of major nonlinearity.)

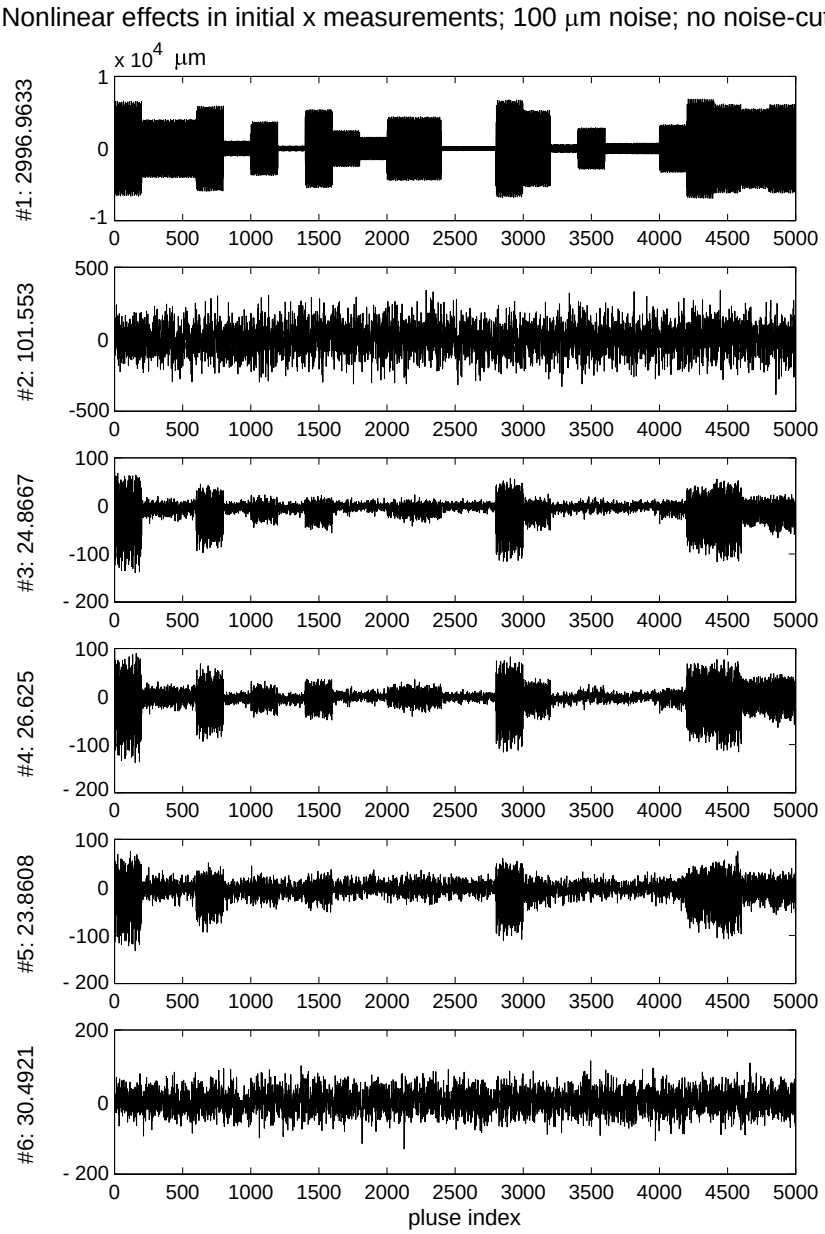

Figure 6.6: Nonlinear effect in phase-space-variable measurement with BPM noise. Similar to Fig. 6.5 except 100±20 μm BPM noise are added and no noise-cut is used. Note that when a large number of BPMs is used, the random noise are dramatically reduced and the nonlinearity errors become dominate.

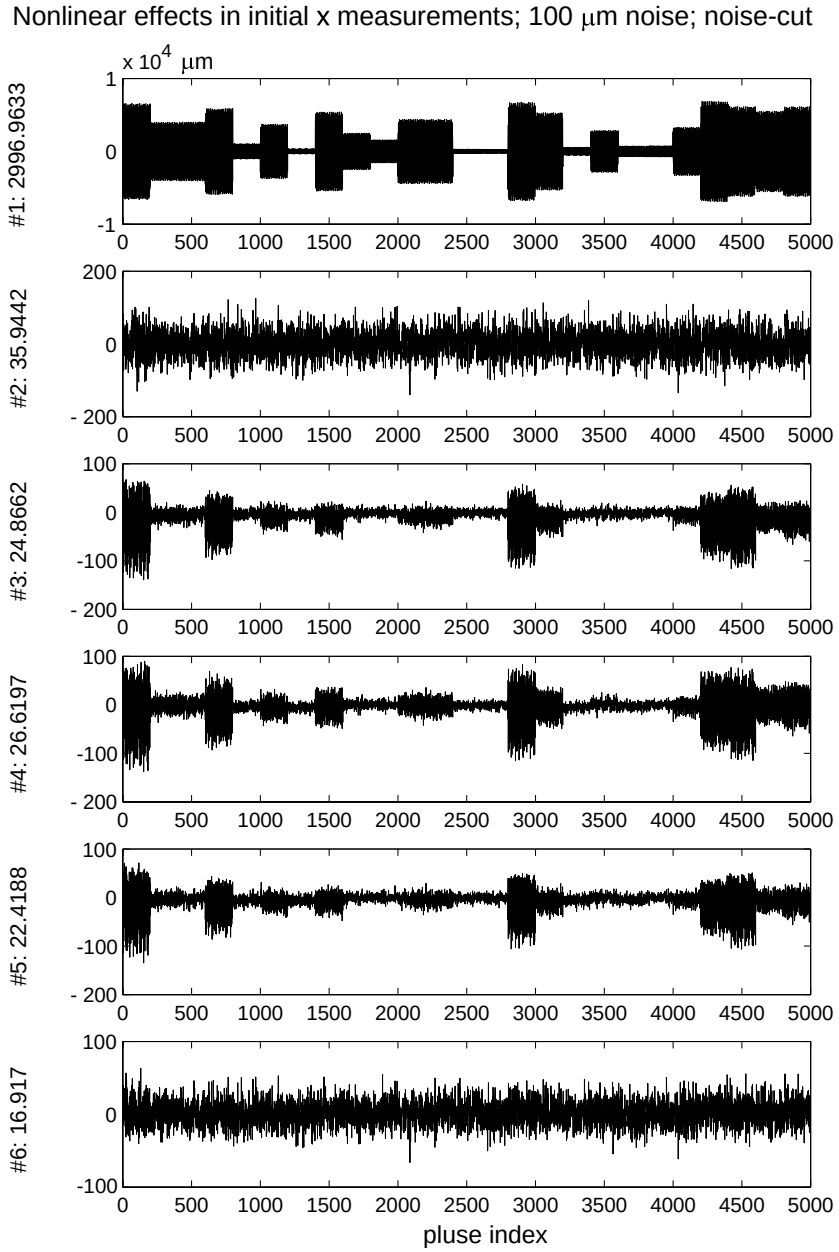

Figure 6.7: Various effects in phase-space-variable measurement. Similar to Fig. 6.5 except 100±20 μm BPM noise are added and noise-cut is used. Note that, in the last frame, the combination of cut-noise and using many BPMs in the fitting can reduce the BPM random noise effect by a factor of 6 (a factor of 2 better than the cut-noise only case in Frame #2) while the nonlinear effect is still small. However, careful section of BPMs is clearly necessary to avoid nonlinear errors.

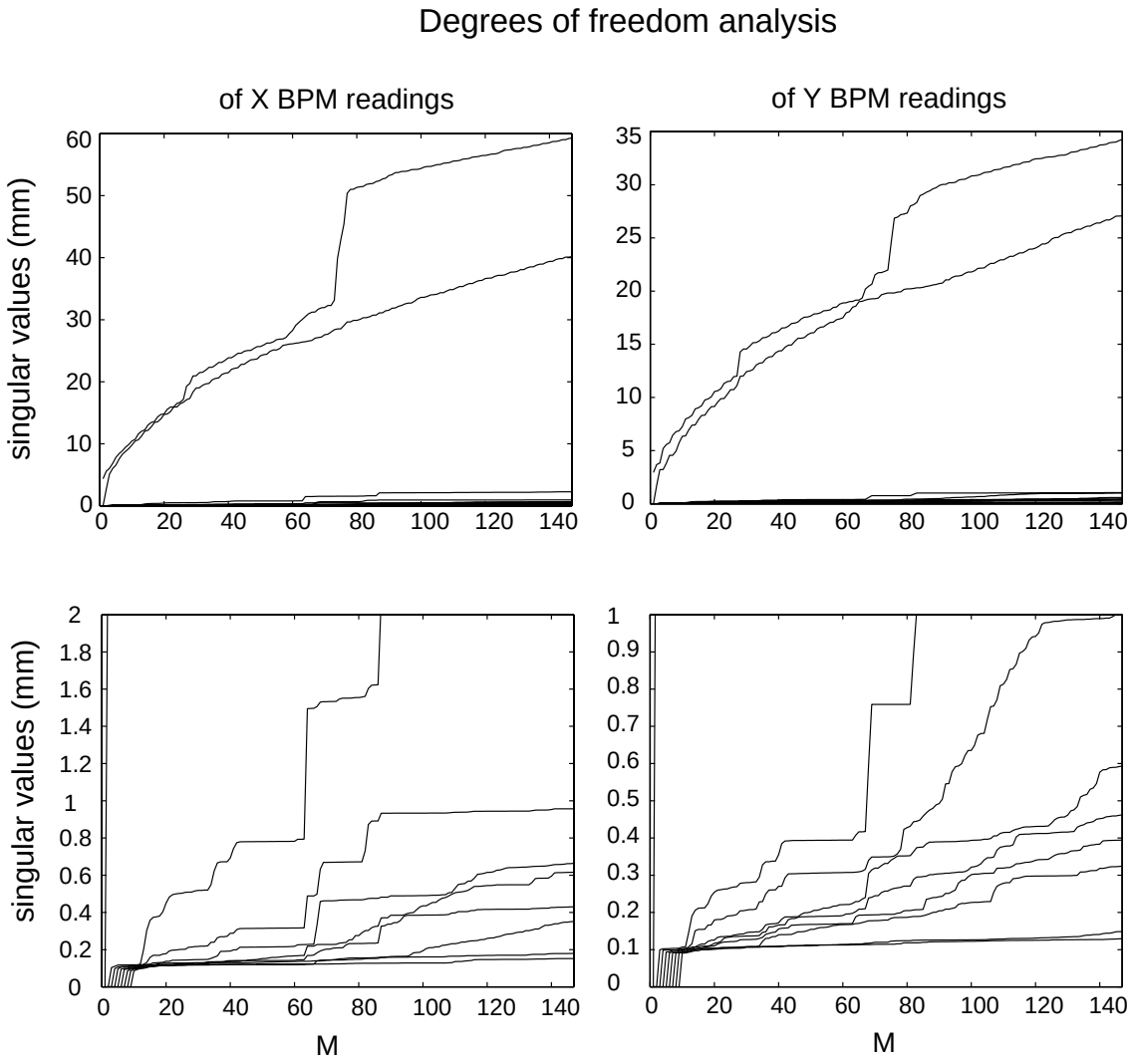

Figure 6.8: Degrees-of-freedom analysis of simulated 5000 turn x and y BPM data with $100 \pm 20 \mu m$ noise. The top frames show the dominating two betatron modes in each plane; the bottom frames are blow-ups to show the other degrees of freedom.

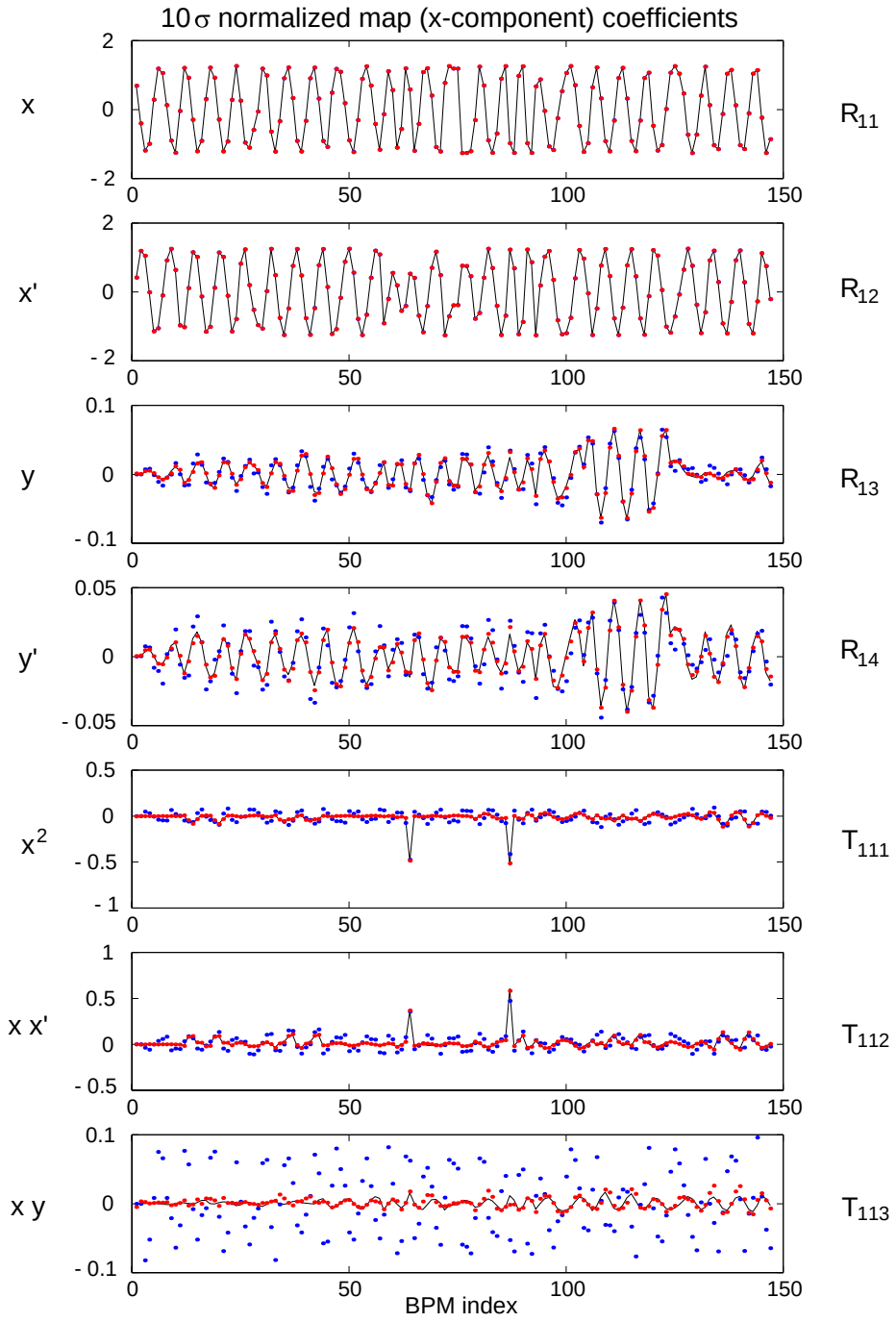

Figure 6.9: Map coefficients as physical basis. Solid lines are the exact values and dots are the simulated measurements with 100μm BPM resolutions. The lighter/darker (orange/blue if in color) dots are with/without noise reduction. Up to the 2nd order terms are shown. (To be continued.)

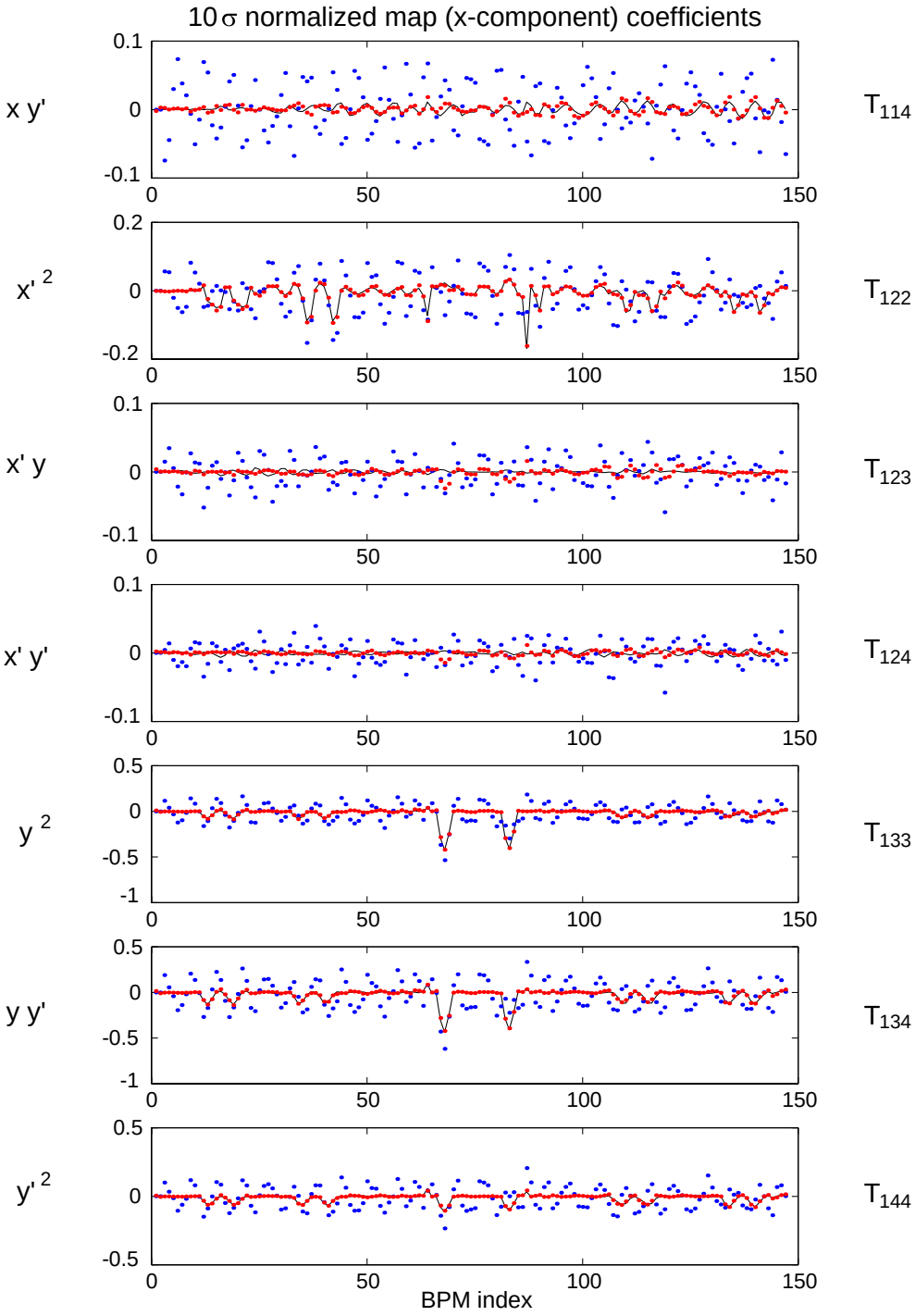

Figure 6.9: Continued

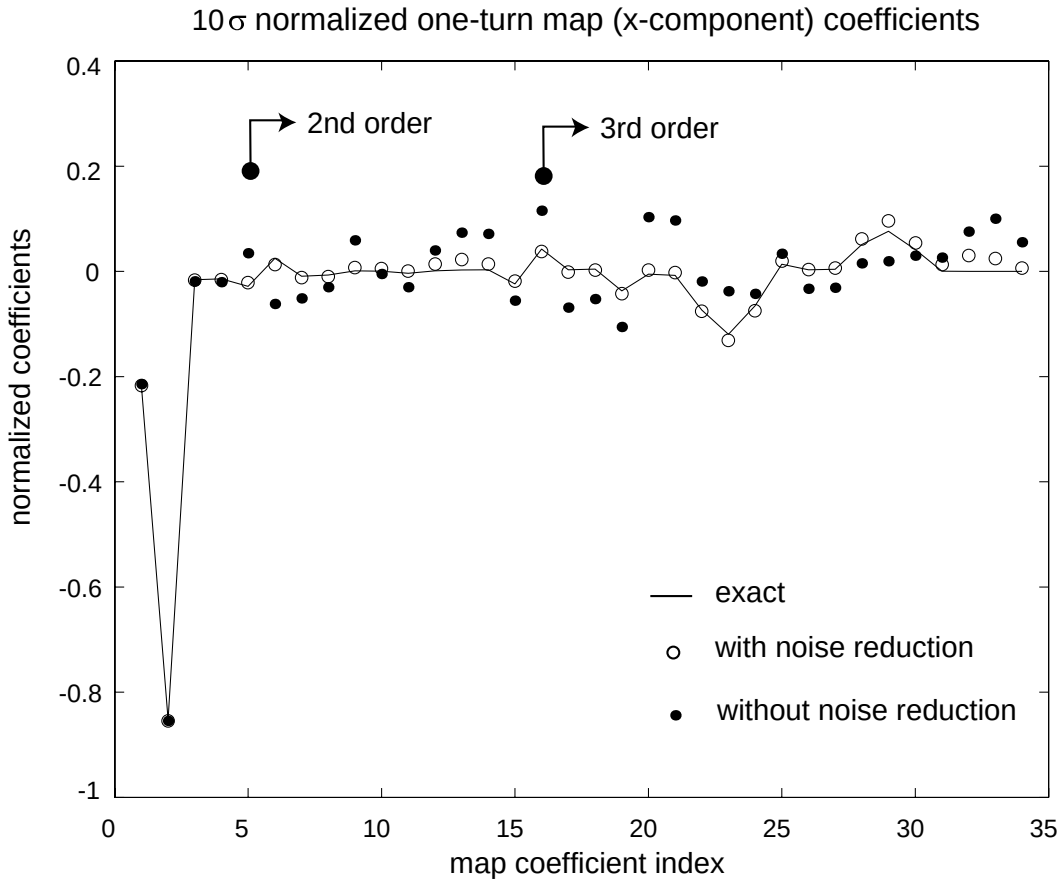

Figure 6.10: Normalized 1-turn map coefficients up to the 3rd order (x-component). "○" and "●" are for $100 \pm 20\mu m$ BPM noise, with and without noise-cut respectively.

Chapter 7

Conclusions

Novel beam dynamics analysis methods (under the general title Model-Independent Analysis–MIA) have been developed in this dissertation. This approach has been shown to be applicable to both linacs and rings. MIA provides a method to analyze beam dynamics without resorting to any particular machine model. The main feature of MIA is a systematic statistical analysis of a BPM-data matrix. By taking advantage of correlations among a large number of BPM readings, one can easily identify problematic BPMs and significantly reduce the effects of BPM random noise. The degrees-of-freedom analysis of a beam line provides valuable information about potential jitter sources that may be due to unknown physics or malfunctioning machine components. The physical base decomposition of a noise-reduced data matrix via measurable physical variables can be used to extract various physical patterns with greatly enhanced sensitivity. In some cases this information is impossible to obtain with other known techniques. Further analysis (such as kick analysis) of the spatial physical patterns can facilitate the interpretation of the results. Conventional spectrum analysis (such as Fourier analysis) can be applied to the temporal physical patterns as usual and further enhance the understanding of beam dynamics. Therefore we believe that MIA can significantly advance beam observation and dynamics analysis and lead to better control of a beam.

In summary, my vision is that the foundation of MIA (and nonlinear map measurements) has been firmly established and preliminary experimental applications

were successful. The described methods will mature as more experimental studies are completed. Eventually they will be used in routine machine operations. Because it is simple and operationally non-invasive, MIA (or parts of) will be widely used for beam dynamics analysis and machine control, even though its title could be forgotten.[1] The SVD noise reduction is particularly important (could be essential) for addressing the beam observation and control challenges arising in future machines (such as the Next Generation Linear Collider–NLC[4], the Linac-based Coherent Light Source–LCLS[3]), where hundreds even thousands of BPMs are available and beam control requirements are extremely tight. Furthermore, MIA can be very useful in the situation that a machine changes so often that a good machine model is hard to maintain. Many new interesting applications, especially for the collective effects and the beam-beam interaction, are bound to appear because of the improvement in the observation sensitivity and the unique power of examining physical modes spatially and temporally at the same time. In terms of further development of MIA, three obvious directions can be pursued: 1) general statistical analysis studies; 2) development of systematic methods to build a machine model based on MIA results and other available information; 3) development of model-independent control schemes.

Nonlinear map measurements are feasible, especially with the help of MIA, and can yield valuable local and global nonlinearity information for built machines. Since few other options are available to measure nonlinearity (except for tune-based global nonlinearity measurements such as the tune-shift-with-amplitude measurements and the "frequency-map" measurements, which offer few clues on the sources of nonlinearities), nonlinear Taylor map measurements should be explored. Special high resolution phase-space measurement systems (basically several high resolution BPMs and a couple of strong kickers) can be built and will considerably improve the quality of nonlinear map measurements. For high-cost, giant machines, such specially built monitoring system could be a wise investment since it can significantly reduce the amount of time to commission and maintain a machine.

[1] As the acronym MIA usually means–Missing In Action :-(

Appendix A

Least-squares methods

Incorporating the creativity of three great mathematicians/physicists (C.F. Gauss,[1] A.M. Legendre and P.S. Laplace), least-squares methods are now widely used in many branches of sciences. Although the simplest least-squares fitting case is well known, the general methods may not be, and its statistical implications and subtlety could be a pitfall for careless users. This appendix provides a brief review of the least-squares methods, which have been frequently used in this dissertation. Reference [59] is one of the nice books on this topic.

A.1 Simple least-squares

The simplest least-squares fitting case is to fit a straight line $y = \beta_0 + \beta_1 x$ through a given set of points $\{ (x_i, y_i) \,|i = 1, \cdots, n\}$ so that the square-sum of the deviations between each y_i and the line is minimized, i.e.

$$\sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = \|\Delta y\|_2^2 = \min. \tag{A.1}$$

where $\hat{\beta}$ means an estimate of the parameter β. The solution is easy. Requiring both $\partial_{\hat{\beta}_0}(\cdots) = 0$ and $\partial_{\hat{\beta}_1}(\cdots) = 0$ generates two equations in two unknowns called the

[1] Here is an anecdote I felt exciting. By inventing the least-squares methods, Gauss magically found the lost star that astronomers were hunting for.

normal equations:

$$\begin{aligned} n\,\hat{\beta}_0 + (\sum x_i)\hat{\beta}_1 &= \sum y_i \\ (\sum x_i)\hat{\beta}_0 + (\sum x_i^2)\hat{\beta}_1 &= \sum x_i y_i \end{aligned} \tag{A.2}$$

whose solution is given by

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{A.3}$$

where $\bar{x}$ and $\bar{y}$ represent average values.

This simple case illustrates the basic technique of least-squares fitting. The essence of the least-squares requirement in Eq. (A.1) is rooted in statistics. The assumption is that y_i is a measurement of $y(x_i)$ with random error ϵ_i, i.e. $y_i - (\beta_0 + \beta_1 x_i) = \epsilon_i$. Usually errors are independent, normally distributed with constant variance. Thus the possibility of having the data pair (x_i, y_i) is $\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\epsilon_i^2/(2\sigma_i^2)}$ and the joint possibility of having the given data set is

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\epsilon_i^2}{2\sigma_i^2}} = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n \epsilon_i^2} = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2} \tag{A.4}$$

Therefore, the possibility of having the particular given data set is maximized by minimizing the square-sum in the exponential, i.e. the least-squares estimates coincide with maximum likelihood estimates. However, the normality of errors is not necessary for the least-squares estimators to have the smallest variance among all linear unbiased estimators. There are many interesting issues I can not touch in such an appendix.

Note that the roles of x and y are not quite the same in the above fitting problem. x is considered the independent (control) variable and y the dependent variable. Usually the independent variables are assumed error-free. The attempt to discern the relationship between a dependent variable and one or more independent variables is named **regression analysis**.[2] When the dependency is linear, it is called **linear**

[2]Regression analysis can be traced to Sir Francis Galton (1822–1911) who observed that children's heights tended to "revert" to the average height of the population rather than diverting from it. That is, the future generations of off-spring who are taller than average are not progressively taller

regression.

In Eq. (A.4), the ϵ_i's are assumed to be independent and have the same variance σ^2. Least-squares problems satisfying these properties are termed **simple least-squares**. In Section A.4 I will describe how to handle more general cases where the variances are not the same and even correlations among the errors may exist.

A.2 Multiple linear regression

The above one independent variable case can be straightforwardly generalized to accommodate multiple independent variables. The general statistical model of multiple linear regression can be expressed as

$$Y = X\beta + \epsilon \tag{A.5}$$

or more explicitly

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nm} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{A.6}$$

where Y and ϵ are random vectors that consist of random variables. The matrix X is considered to be a matrix of known constants, which are the observations on the independent variables. The β vector consists of parameters to be estimated.

The normal equation is given by

$$X^T X\hat{\beta} = X^T Y \tag{A.7}$$

where $\hat{\beta}$ is the least-squares estimate of the parameter β. If $X^T X$ has an inverse, the

than their respective parents, and parents who are shorter than average do not beget successively smaller children.[62] For another account, see Ref. [19].

normal equations have an unique solution given by

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y. \tag{A.8}$$

In matrix notation, the least-squares solution can be expressed by such a simple form! which is easy and worth to remember. The matrix $\left(X^T X\right)^{-1} X^T$ is called pseudo-inverse of the rectangular matrix X, and is often denoted as A^+. Apparently the purpose of multiplying X^T is to make a square matrix out of the rectangular matrix X so that it can be inversed. However, as in the case of one independent variable, the least-squares requirement

$$\sum \epsilon_i^2 = \epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta) = \min \tag{A.9}$$

is the statistical reason behind the above normal equation.

A.3 Error estimate, degeneracy, and application of SVD

The major statistical properties of the least-squares solution $\hat{\beta}$ are its mean and variance. If the linear model is correct, the mean value is equal to the real parameter β, i.e. the least-squares estimator is an unbiased estimator. The variance of $\hat{\beta}$ reflects the error bar of the least-squares estimate, which can be computed via

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \left[(X^T X)^{-1} X^T\right] \mathrm{Var}(Y) \left[(X^T X)^{-1} X^T\right]^T \\ &= \left[(X^T X)^{-1} X^T\right] \mathrm{Var}(\epsilon) \left[(X^T X)^{-1} X^T\right]^T \\ &= \left[(X^T X)^{-1} X^T\right] I\sigma^2 \left[(X^T X)^{-1} X^T\right]^T \\ &= \left(X^T X\right)^{-1} \sigma^2 \end{aligned} \tag{A.10}$$

The square roots of the diagonal terms of $\mathrm{Var}(\hat{\beta})$ give the error bars for each estimated parameter. Note that it is valid only when the linear model is correct.

We see that the accuracy of the estimated parameters are determined by the

inversion of X^TX. When there is near degeneracy in X (i.e. some independent variables are almost linearly dependent of each others), $\left(X^TX\right)^{-1}$ could be very large and even close to singular. In such cases, the estimated parameters may have huge error bars and thus are not reliable anymore. It may be necessary to remove those degenerated variables in order to make each estimated parameter sufficiently accurate. This is particularly important when the results are used in feedback controls.

Singular value decomposition (SVD) provides a general way to solve the least-squares problem (including the above rank-deficient case). Let the SVD of $X = USV^T = \sum_i \sigma_i u_i v_i^T$, then

$$\begin{aligned} \|Y - X\beta\|_2^2 &= \|U^TY - (U^TXV)(V^T\hat{\beta})\|_2^2 = \|S(V^T\hat{\beta}) - U^TY\|_2^2 \\ &= \sum_{i=1}^{r}(\sigma_i\alpha_i - u_i^TY)^2 + \sum_{i=r}^{n}(u_i^TY)^2 \end{aligned} \tag{A.11}$$

where $r = \text{rank}(X)$ and the column vector $\alpha = V^T\hat{\beta}$. In the first step, the orthogonality of U and V has been used. Clearly, $\alpha_i = u_i^TY/\sigma_i$ for $i = 1, \cdots, r$ gives the least-squares solution. Furthermore, if we set $\alpha_i = 0$ for $i = r+1, \cdots, m$, then the solution $\hat{\beta} = V\alpha = X^+Y$ clearly has minimal 2-norm $\|\hat{\beta}\|_2$. Therefore the general pseudo-inverse can be written as

$$X^+ = \sum_i \frac{1}{\sigma_i} v_i u_i^T = V \,\text{diag}\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_r}, 0, \cdots, 0\right) U^T. \tag{A.12}$$

If X is full-rank, then $X^+ = \left(X^TX\right)^{-1}X^T$. Otherwise, removing the close-to-zero singular values could make the pseudo-inverse well behaved and yields more reliable least-squares solution $\hat{\beta} = X^+Y$. This is the most commonly used feature of SVD in accelerator community before MIA were developed.

The SVD basically generates a set of orthogonal variables from the original independent ones and the corresponding singular values reflect the significance of each orthogonal variable. By setting the insignificant singular values to zeros, the degenerated degrees of freedom are effectively removed.

Note the other very important feature of the least-squares solution $\hat{\beta} = X^+Y$

given by Eq. (A.12)—it has the smallest 2-norm of all minimizers that give the least squares. (When X is rank-deficient, there are infinite number of solutions to the least-squares requirement.) This feature has been widely used to reduce the rms corrector strengths for orbit corrections.

A.4 Generalized least-squares

In the simple least-squares method discussed above, the random errors associated with each measurement of the dependent variable are assumed to be independent and have the same variance, i.e. $\mathrm{Var}(\epsilon) = \sigma^2 \mathbf{I}$. However, many applications do not satisfy this condition. For example, each BPM may have different resolution, thus $\mathrm{Var}(\epsilon)$ is diagonal of unequal variances. In some cases, noise from different BPMs may even be correlated, which results in a general covariance matrix $\mathrm{Var}(\epsilon) = \sigma^2 \mathbf{V}$ that is symmetric positive definite but have non-zero off-diagonal entries. Convention labels the first case **weighted least squares** and the second, more general case **generalized least squares.**

The idea behind both is the same: transform the general least-squares problem into a simple least-squares problem. To accomplish this, the covariance matrix is decomposed as

$$\mathbf{V} \equiv \frac{1}{\sigma^2}\mathrm{Var}(\epsilon) = W\, W^T \tag{A.13}$$

which is always possible for a symmetric positive definite matrix. For example Cholesky decomposition can be used, which yields a unique upper-triangular W.

Since V is positive definite, W is nonsingular. Pre-multiplying the general least-squares model by W^{-1} gives

$$\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon} \tag{A.14}$$

where $\tilde{Y} = W^{-1}Y$, $\tilde{X} = W^{-1}X$, and $\tilde{\epsilon} = W^{-1}\epsilon$. With this transformation,

$$\mathrm{Var}(\tilde{\epsilon}) = W^{-1}V(W^{-1})^T\sigma^2 = \sigma^2 \mathbf{I}. \tag{A.15}$$

Therefore, the simple least-squares method can be applied to Eq.(A.14) to obtain an

estimate $\hat{\beta}$ as

$$\hat{\beta} = (\tilde{X}^T\tilde{X})^{-1}(\tilde{X})^T\tilde{Y} = (X^TV^{-1}X)^{-1}X^TV^{-1}Y \tag{A.16}$$

Note that the result does not depends on W, which may not be unique. The variance of the estimate $\hat{\beta}$ is given by

$$\mathrm{Var}(\hat{\beta}) = (\tilde{X}^T\tilde{X})^{-1}\sigma^2 = (XV^{-1}X)^{-1}\sigma^2 \tag{A.17}$$

Note that this solution minimized $\mathrm{Var}(\tilde{\epsilon}) = \langle\epsilon^TV^{-1}\epsilon\rangle$ instead of $\langle\epsilon^T\epsilon\rangle$.

Thanks to the matrix notation, which provides a concise expression for the general multiple regression problems, without it, even the simplest least-squares problem appears complicated.

Appendix B

Singular Value Decomposition (SVD)

A major computation method used in this dissertation is the singular value decomposition of a matrix. It is extremely useful in matrix theory from both theoretical and computational point of views. As a matrix analysis method, it clearly reveals the structure of a matrix and the various linear spaces associated with the matrix. This appendix will briefly review SVD, mainly for the purpose of understanding this powerful tool, although some computational aspects are discussed. Section B.1 summarizes the definition and major properties of SVD. Section B.2 describes a method to compute SVD by solving the eigenvalue problem of a symmetric matrix. The rest sections describe basic orthogonal transformations, via which various useful matrix decompositions, especially the SVD, can be constructed. The important statistical aspects of SVD are discussed in Section 3.3. Most of the materials in this appendix come from the reference [39].

B.1 Definition, existence, and properties of SVD

If $A_{m\times n}$ is a real matrix, then there exist real orthogonal matrices $U_{m\times m} = [u_1, \cdots, u_m]$ and $V_{n\times n} = [v_1, \cdots, v_n]$ such that $U^T A V = \text{diag}(\sigma_1, \cdots, \sigma_p) \equiv S \in R^{m\times n}$, or

$$A = USV^T = \sum_i \sigma_i \, u_i \, v_i^T \tag{B.1}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ and $p = \min(m, n)$. This decomposition of a matrix A is called **Singular Value Decomposition (SVD)** and the expansion is called SVD expansion. The σ's are called **singular values** and the u's (v's) are called left (right) **singular vectors**. From orthogonality of u's and v's, it is easy to see that

$$Av_i = \sigma_i u_i \quad \text{and} \quad A^T u_i = \sigma_i v_i \qquad (i = 1, \cdots, p). \tag{B.2}$$

When $m > n$ a trimmed down version is often used, where only the first n columns of $U_{m\times m}$ and the first n rows of $S_{m\times n}$ are kept, thus $A = U_1 S_1 V^T$, which is called the thin SVD.

SVD is extremely useful because it reveals a great deal about the structure of a matrix, especially the linear spaces associated with the matrix. There are two important subspaces associated with a matrix $A_{m\times n}$. The **range** of A is defined by $\text{ran}(A) \equiv \{y \in R^m | y = Ax, \text{ for some } x \in R^n\}$, and the **null space** of A is defined by $\text{null}(A) \equiv \{x \in R^n | Ax = 0\}$. If $A_{m\times n} = [a_1, \cdots, a_n]$ is a column partitioning, then $\text{ran}(A) = \text{span}\{a_1, \cdots, a_n\} \equiv \{\sum_{j=1}^n \beta_j a_j | \beta_j \in R\}$. The **rank** of a matrix A is defined by $\text{rank}(A) \equiv \dim(\text{ran}(A))$. Note that $\text{rank}(A) = \text{rank}(A^T)$. $A_{m\times n}$ is **rank deficient** if $\text{rank}(A) < \min(m, n)$ and is **full rank** otherwise.

If the number of nonzero singular values is r, then

$$\begin{aligned}
\text{rank}(A) &= \text{rank}(A^T) = r && \text{(B.3)}\\
\text{ran}(A) &= \text{span}\{u_1, \cdots, u_r\} && \text{(B.4)}\\
\text{ran}(A^T) &= \text{span}\{v_1, \cdots, v_r\} && \text{(B.5)}\\
\text{null}(A) &= \text{span}\{v_{r+1}, \cdots, v_n\} && \text{(B.6)}\\
\text{null}(A^T) &= \text{span}\{u_{r+1}, \cdots, u_m\} && \text{(B.7)}
\end{aligned}$$

Thus the SVD reveals the dimension of the associated subspaces and the singular vectors provide a set of orthogonal basis for those subspaces. By specifying the threshold for a singular value to be nonzero, SVD provides a practical method to numerically define a matrix rank.

The importance of SVD can also be appreciated from the orthogonal projection

point of view. Let $S \subseteq R^n$ be a subspace. $P \in R^{n\times n}$ is an **orthogonal projection** onto S if $\text{ran}(P) = S$, $P^2 = P$, and $P^T = P$. The usefulness of orthogonal projection P is that if $x \in R^n$, then $Px \in S$ and $(I - P)x \in S^\perp$. From SVD result $A = USV^T$, we have the partitionings $U = [U_r, \tilde{U}_r]$ and $V = [V_r, \tilde{V}_r]$, the following important orthogonal projections can be constructed,

$$V_r V_r^T = \text{projection on to null}(A)^\perp = \text{ran}(A^T) \tag{B.8}$$

$$\tilde{V}_r \tilde{V}_r^T = \text{projection on to null}(A) \tag{B.9}$$

$$U_r U_r^T = \text{projection on to ran}(A) \tag{B.10}$$

$$\tilde{U}_r \tilde{U}_r^T = \text{projection on to ran}(A)^\perp = \text{null}(A^T) \tag{B.11}$$

The singular values have close connections to various matrix norms, especially the 2-norm $\|A\|_2$ and the Frobenius norm $\|A\|_F$.

$$\|A\|_2 \equiv \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1 \qquad \text{(the max. singular value)} \tag{B.12}$$

$$\|A\|_F^2 \equiv \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{p} \sigma_i^2 \tag{B.13}$$

$$\min_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_n \qquad (m \geq n). \tag{B.14}$$

where the vector 2-norm is defined by $\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1} x_i^2}$.

From the matrix approximation point of view [59], the SVD expansion in Eq.(B.1) expresses a matrix A as a sum of p rank-1 matrices. The first of these matrices is the "best" rank-1 approximation to A, the sum of the first two matrices is the "best" rank-2 approximation of A, and so forth. These are "best" approximations in the least-squares sense; that is, no other matrix of the same rank will give a better agreement with the original matrix A as measured by the sum of squared differences between the corresponding elements of A and the approximating matrix [43]. The goodness of fit of the approximation in each case is given by the ratio of the sum of the eigenvalues used in the approximation to the sum of all eigenvalues. This property can be understood from Eq.(B.13).

The geometrical meaning of SVD is as follows: consider a unit hypersphere defined by $\{x \,|\, x \in R^n, \|x\|_2 = 1\}$, the matrix $A_{m\times n}$ specifies a linear transformation that transform the hypersphere into a hyperellipsoid defined by $\{Ax \,|\, x \in R^n, \|x\|_2 = 1\}$. The singular values give the lengths of the semi-axes and the v-vectors give the directions of the semi-axes.

The statistical meaning of SVD is that it accomplishes the principal components analysis of a data matrix. See Section 3.3 for detail descriptions.

The applications of SVD in the least-squares problem and matrix pseudo-inverse problem are discussed in the Appendix A.3. This probably is the most familiar encounter with SVD in the accelerator physics community.

B.2 Computing SVD decomposition of a matrix

There are many algorithms that can accomplish the SVD of a matrix with different efficiency and accuracy. In this section, I will outline one SVD algorithm in order to illustrate important relations involved and to understand such decomposition better. Numerical efficiency and accuracy are not a concern here (in fact, the algorithm described here is among the worst in these regard).

Given a matrix $A \in R^{m\times n}$ with $m \geq n$, SVD gives $A = USV^T$ where matrices U and V are orthogonal and matrix S is diagonal with positive elements decreasing along the diagonal. It is easy to see that

$$(A^TA)V = (VSU^T)(USV^T)V = VS^2 \tag{B.15}$$

So the column vectors of V are eigenvectors of A^TA with eigenvalues given by the corresponding diagonal terms in S^2. Since A^TA is symmetric and semi-positive definite, its eigenvalues and eigenvectors are unique (eigenvectors up to a scaling factor) unless there are degeneracy.

Equation(B.15) suggests us to form the matrix A^TA and compute its eigenvalues and eigenvectors. Then take the positive roots of all eigenvalues and order them in descending order to form the diagonal matrix S. Meanwhile, use the corresponding

eigenvectors as the column vectors to form the V matrix. Having found S and V, the matrix U can be computed from $U = AVS^{-1}$, whose orthogonality follows from Equation(B.15) and the orthogonality of V. Note that the result is a "thin" SVD of the matrix A. The following sections discuss basic orthogonal transformations and better ways to compute SVD of a matrix. Skipping these sections will not affect the understanding of this dissertation.

B.3 Basic orthogonal transformations

Since orthogonal transformations conserve vector 2-norm and matrix 2-norm (as well as Frobenius norm of a matrix), they have an important role to play in least-squares fitting and eigenvalue computations. Except the basic permutation matrices which will be reviewed first, there are two more elementary orthogonal transformations that are often used as building blocks for more complicated orthogonal transformations: Householder reflections and Givens rotations. They are generalizations of our everyday concepts of reflection and rotation (in Euclidean space).

Permutation matrix Π is just the identity matrix with its rows re-ordered. It represents operation that transforms a matrix A by permuting its rows (ΠA) or columns ($A\,\Pi$). For example, if

$$\Pi = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then ΠA will switch the 1st and 3rd rows of A and $A\,\Pi$ will permute the 1st and 3rd columns of A. Obviously, permutation matrices are orthogonal and a product of permutation matrices is a permutation matrix.

Householder reflection is specified by the matrix

$$H = I - \frac{2}{v^T v} v v^T \tag{B.16}$$

where v is a nonzero column vector and I is identity matrix of proper dimension. It is clear that H is symmetric and orthogonal. H can be viewed geometrically as a reflection in the subspace that orthogonal to a vector v (in the case of mirror reflection in 3D, v is the normal vector of the mirror and the orthogonal subspace is the 2D mirror surface).

For any nonzero vector x, one can generate a Householder matrix to zero all components of x except the first one, which equals to the length of the vector with an optional sign. Mathematically speaking, $Hx = \mp\|x\|_2\, e_1$ where H is generated by the Householder vector $v = x \pm \|x\|_2\, e_1$ and $e_1 = [1\, 0 \cdots 0]^T$. Geometrically speaking, this property means that given a vector in certain orthogonal basis (v specifies its components), one can transform to another set of orthogonal basis (via orthogonal transformation H) that has the first base vector along the given vector in either direction (the optional sign, the length of the vector remains the same because the transformation H is orthogonal). This property and the simple determination of the Householder vector make the Householder reflection very useful. It was used by A. S. Householder (1958) in discussing certain eigenvalue problems.

Givens rotation (Givens 1954) is specified by the matrix

$$G(i,k,\theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{matrix} \\ \\ i \\ \\ k \\ \\ \\ \end{matrix} \qquad \text{(B.17)}$$

$$\begin{matrix} & & i & & k & & \end{matrix}$$

where $c = \cos\theta$ and $s = \sin\theta$. Clearly G is orthogonal. The geometrical meaning is obvious also: G generates a rotation of angle θ in the plane given by the i-th and k-th base vectors.

Householder routine : $H = H_5 H_4 \cdots H_1$, Givens routine : $G = G_{15} G_{14} \cdots G_1$

$$\begin{bmatrix} * & * & * & * & * \\ 1 & * & * & * & * \\ 1 & 2 & \underline{*} & \bar{*} & \bar{*} \\ 1 & 2 & \underline{3} & \bar{*} & \bar{*} \\ 1 & 2 & \underline{3} & \bar{4} & \bar{*} \\ 1 & 2 & \underline{3} & \bar{4} & \bar{5} \end{bmatrix} \qquad \begin{bmatrix} * & * & * & * & * \\ 5 & * & * & * & * \\ 4 & 9 & \underline{*} & \bar{*} & \bar{*} \\ 3 & 8 & \underline{12} & \bar{*} & \bar{*} \\ 2 & 7 & 11 & 14 & * \\ 1 & 6 & 10 & 13 & 15 \end{bmatrix}$$

Figure B.1: Illustration of QR factorization processes via a series of Householder transformations or Givens rotations. The numbers indicate at which step that element is zeroed.

The usefulness of Householder reflection and Givens rotation in matrix computations is that they can transform orthogonally a matrix into simpler form by zeroing certain elements. Householder reflection is used to zero a block of elements in a column vector v, while Givens rotation can be used to zero one element of a vector v. Figure B.1 illustrates how Householder reflections and Givens rotations can be used to transform a matrix $A_{6\times5}$ into an upper triangular form. The numbers in the matrix indicate at which step that element is zeroed. For example, H_3 is obtained by putting the 4×4 Householder matrix $\tilde{H}_3$ for the underlined elements into the lower diagonal block of a 6×6 identity matrix, i.e. $H_3 = \text{diag}(I_2, \tilde{H}_3)$; the other affected elements are marked with a $\bar{}$. On the other hand, $G_{12} = G(3, 4, -\arctan(\underline{12}/\underline{*}))$.

We see that a series of orthogonal transformations can bring a matrix A into an upper triangular form R (which equals HA in Householder routine and GA in Givens routine. HA and GA may or may not be equal but both are upper triangular matrices). This leads to an important matrix factorization called **QR factorization**.

B.4 QR factorization of a matrix

For any matrix $A_{m\times n}$, there exists orthogonal matrix Q such that

$$A = Q_{m\times m} R_{m\times n} \tag{B.18}$$

where R is upper triangular matrix.

If $m \geq n$ and $A_{m\times n}$ has full column rank (i.e. all columns of A are linearly independent and $rank(A) = n$), QR factorization has the property that the first n columns of Q form an orthonormal basis for ran(A). In fact, a stronger statement can be made:

$$\text{span}\{a_1, \cdots, a_k\} = \text{span}\{q_1, \cdots, q_k\}, \qquad k = 1, \cdots, n \tag{B.19}$$

where a_k and q_k are column vectors of A and Q matrices. Q and R matrices can even be trimmed down (keep only first n columns of Q and first n rows of R) to get a thin QR factorization

$$A = Q_{m\times n}R_{n\times n} \equiv Q_1 R_1. \tag{B.20}$$

This factorization is unique when R_1 uses positive diagonal entries. Moreover, R_1^T is the lower triangular Cholesky factor of the symmetric matrix $A^T A$.

If $m \geq n$ and $A_{m\times n}$ is rank deficient (i.e. some columns of A are a linear combination of the other columns), the above properties may not hold anymore. Especially Q does not necessarily produce an orthonormal basis for ran(A). A simple cure of this is called **QR factorization with Column Pivoting**. The basic idea is to group dependent and independent columns into separate blocks via column permutations. The Householder QR factorization can be modified to include column pivoting: before each step in Householder QR factorization, the first column with the largest 2-norm (of the columns in the submatrix being considered) should be switched to the lead position. QR factorization with column pivoting yields

$$Q^T A\,\Pi = \begin{bmatrix} R_{11} & R_{12} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} \tag{B.21}$$

where $r = \text{rank}(A)$, Q is orthogonal, R_{11} is upper triangular and non-singular, and Π is a permutation. The first r columns of Q form orthonormal basis for ran(A).

B.5 Symmetric Schur Decomposition

For any real symmetric matrix A of order n, there exists a real orthogonal Q such that

$$Q^T AQ = \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n). \tag{B.22}$$

Moreover, the k-th column of Q is the eigenvector of A that has eigenvalue λ_k. Eq.(B.22) is called **symmetric Schur decomposition**.

The eigenvalues of a symmetric matrix have the "minimax" property that

$$\lambda_k(A) = \max_{\dim(S)=k} \min_{0 \neq y \in S} \frac{y^T Ay}{y^T y}, \qquad (k = 1, \cdots, n) \tag{B.23}$$

This is referred to as **Courant-Fischer Minimax Theorem**.

Bibliography

[1] MAD home page: http://wwwslap.cern.ch/~fci/mad/mad_home.html.

[2] SAD home page: http://www-acc-theory.kek.jp/SAD/sad.html.

[3] Linac Coherent Light Source (LCLS) design study report. SLAC-R-521, 1999.

[4] NLC home page. http://www-project.slac.stanford.edu/nlc/home.html.

[5] PEP-II an Asymmetric B Factory Conceptual Design Report. SLAC-418.

[6] F. S. Acton. *Numerical Methods that Work*. Mathematical Association of America, 1990.

[7] G. R. Aiello and M. R. Mills. Log-ratio technique for beam position monitor systems. *AIP Conf. Proc. no.281*, page 301, 1992.

[8] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc., 2nd edition, 1984.

[9] T.W. Anderson. Inference in linear models. Multivariate Analysis and Its Applications, IMS Lecture Notes–Monograph Series Volume 24, edited by T.W. Anderson, K.T. Fang, and I. Olkin, 1994.

[10] L. Arnaudon, et al. Measurement of LEP beam energy by resonant spin depolarization. *Phys. Lett. B 284, no.3-4*, page 431, 1992.

[11] L. Arnaudon, et al. Effects of terrestrial tides on the LEP beam energy. *Nucl. Instrum. Meth. in Phys. Res.*, A357:249, 1995.

[12] R. Assmann, et al. LIAR–a computer program for the modeling and simulation of high performance linacs. SLAC internal report, AP-103.

[13] B. Autin and Y. Marti. Closed orbit correction of alternating gradient machines using a small number of magnets. CERN/ISR-MA/73-17, 1973.

[14] V. Balakin. Beam position monitor with nanometer resolution for linear collider. *EPAC94. Forth European Part. Acc. Conf.*, page 1539, 1994.

[15] K.L.F. Bane. private communication.

[16] K.L.F. Bane. Landau damping in the SLAC linac. SLAC-PUB-3670, 1985.

[17] T. Barklow. A technique for measuring the 4×4 transfer matrix for beamline sections with cross-plane coupling. SLAC PUB-4969, 1989.

[18] T. Barklow, P. Emma, P. Krejcik, and N. Walker. Review of lattice measurement techniques at the SLC. *AIP Conf. Proc.*, 255:347–54, 1992. SLAC-PUB-5695.

[19] David Birkes and Yadolah Dodge. *Alternative Methods of Regression.* John Wiley & Sons, Inc., New York, 1993.

[20] E. Bozoki and A. Friedmann. *Proceedings of Particle Accelerator Conference 1993*, 1993.

[21] O. S. Bruning, et at. Reduction of proton losses in HERA by compensating tune ripple due to power supplies. *Phys. Rev. Lett.*, v.76, no.20:3719, 1996.

[22] P. Burla, et al. Power supply ripple study at the SPS. *AIP Conf. Proc. no.326*, page 167, 1992.

[23] Y. Cai and at al. Low energy ring lattice of the PEP-II asymmetric B-factory. *Proceedings of the 1995 Particle Accelerator Conference*, 1995.

[24] A. W. Chao. Notes on nonlinear dynamics. unpublished.

[25] A. W. Chao. *Physics of Collective Beam Instabilities in High Energy Accelerators.* John Wiley & Sons Inc., 1993.

[26] A. W. Chao, et al. *Phys. Rev. Lett.*, v.61, no.24:2752, 1988.

[27] T. Chen, et al. *Phys. Rev. Lett.*, v.68, no.1:33, 1991.

[28] W. J. Corbett, M. J. Lee, and V. Ziemann. A fast model-calibration procedure for storage rings. *Proceedings of Particle Accelerator Conference 1993*, 1993.

[29] W.J. Corbett, M.J. Lee, and V. Ziemann. Model calibration and beam control systems for storage rings. *SPIE's 1993 International Symposium on Optics, Imaging and Instrumentation, San Diego, CA, 11-16 Jul 1993.*, 1993.

[30] G. A. Decker. Operational aspects of experimental accelerator physics. *AIP Conf. Proc. no.249*, page 637, 1991.

[31] F.-J. Decker, et al. Super-ASSET: a technique for measuring and correcting accelerator structure misalignments at the SLC. *Proceedings of PAC97.*

[32] M.H.R. Donald, C.A. Blocker, A.W. Chao, R.J. Hollebeek, Martin J. Lee, E. Linstadt, J. Siegrist, and N.C. Spencer. Some schemes for online correction of the closed orbit, dispersion and beta functions in PEP. *Proc. IEEE 1981 Particle Accelerator Conference, Washington, D.C.*, 1981.

[33] A. Dragt. Lectures on nonlinear orbit dynamics. *AIP Conf. Proc. No.87, Phys. High Energy Part. Accel.*, page 147, 1982.

[34] E. Forest. Normal form algorithm on non-linear symplectic maps. SSC-29, 1985.

[35] E. Forest. *Beam Dynamics: A New Attitute and Framework.* Gordon & Breach, 1998.

[36] E. Forest, David Douglas, and Beat Leemnann. Study of the aberrations of periodic arc using the lie algebraic techniques. *Nucl. Instr. Meth. in Phys. Res. A258*, pages 355–363, 1987.

[37] A. Friedmann and E. Bozoki. Use of eigenvectors in understanding and correcting storage ring orbits. *Nucl. Instrum. Meth. A344*, pages 269–277, 1994.

[38] R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley & Sons, Inc., New York, 2nd edition, 1997.

[39] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[40] Walter C. Hamilton. *Statistics in Physical Science*. The Ronald Press Company, 1964.

[41] S. Heifets, et al. Impedance study for the PEP-II B-factory. SLAC/AP-99 (1995).

[42] K. Hirata. An introduction to SAD. Second Advanced ICFA Beam Dynamics Workshop, CERN 88-04, 1988.

[43] A. S. Householder and G. Young. Matrix approximation and latent roots. *American Mathematical Monthly*, pages 165–171, 1938.

[44] I. C. HSU. *Particle Accelerator*, v.34:43, 1990.

[45] J. Irwin, C.X. Wang, Y.T.Yan, K.L.F. Bane, Y. Cai, F.-J. Decker, M.G. Minty, G.V. Stupakov, and F. Zimmermann. Model-Independent Beam Dynamics Analysis. *Phys. Rev. Lett.*, 82(8):1684–1687, 1999.

[46] John Irwin, Chun-xi Wang, Karl Bane, Yiton Yan, Michiko Minty, Franz-J. Decker, and Gennady Stupakov. Transverse wakefield effect measurement via Model-Independent Analysis. *Proceedings of the XIX international LINAC conference, vol.1*, page 478, 1998.

[47] F. Christoph Iselin. The MAD Program. CERN/SL/98-XX(AP), 1998.

[48] J. E. Jackson. *A User's Guide to Principal Components*. Wiley, New York, 1991.

[49] I. P. Karabekov. High accuracy beam energy measurement for linear and circular accelerators. *Nucl. Instr. Meth. in Phys. Res. A*, 321, no.1-2:18, 1992.

[50] T. Katsura, et al. Variation of the circumference in a storage ring and expansion of the building. *EPAC92. Third European Part. Acc. Conf.*, page 1631, 1992.

[51] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Inc., 1974.

[52] S. Y. Lee. Beam dynamics experiments at IUCF. *AIP Conf. Proc. no.326*, page 18, 1992.

[53] S. Y. Lee, et al. *Phys. Rev. Lett.*, v.67, no.27:3768, 1991.

[54] T. Lohse and P. Emma. Linear fitting of bpm orbits and lattice parameters. SLAC-CN-371 (Februrary 1989).

[55] R. E. Meller, et al. SSC report SSC-N360, 1987.

[56] M. Minty and F. Zimmermann. Beam technique–beam control and manipulation. Lecture note of US particle accelerator school, Jun, 1999.

[57] T.O. Raubenheimer. The generation and acceleration of low emittance flat beams for future linear colliders. SLAC-Report-387, 1991.

[58] T.O. Raubenheimer and K. Kubo. A technique for measuring and correcting emittance dilutions due to accelerator structure misalignments. *Nucl. Instr. Meth. in Phys. Res. A*, 370:303–311.

[59] John O. Rawlings, Sastry G. Pantula, and David A. Dickey. *Applied Regression Analysis–A Research Tool*. Springer, 2nd edition, 1998.

[60] Richard Reyment and K.G. Jöreskog. *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press, 1996.

[61] D. Robin, J. Safranek, G. Portmann, and H. Nishimura. Model calibration and symmetry restoration of the advanced light source. *Proceedings of 1996 European Particle Accelerator Conference*, 1996.

[62] Thomas P. Ryan. *Modern Regression Methods*. John Wiley & Sons, Inc., New York, 1997.

[63] J. Safranek. Experimental determination of storage ring optics using orbit response measurements. *Nucl. Instr. and Meth. A*, 388:27–36, 1997.

[64] M. Sands. the physics of electron storage rings. *SLAC report-121*, 1970.

[65] R. E. Shafer. Beam position monitoring. *AIP Conf. Proc. no.249*, page 601.

[66] Steve Smith. (unpublished) private communication, 1998.

[67] M. D. Springer. *The algebra of random variables*. Wiley, New York, 1979.

[68] Barbara G. Tabachnick and Linda S. Fidell. *Using Multivariate Statistics*. HarperCollins College Publishers, 3rd edition, 1996.

[69] P. Tran, et al. *Proc. of the 1995 Part. Accel. Conf.*, page 3323.

[70] Chun-xi Wang and Alex Chao. Notes on Lie algebraic analysis of achromats. SLAC internal report AP-100 (1995).

[71] Chun-xi Wang and John Irwin. Decoherence and wakefield effects in one-turn map measurement. *AIP Conf. Proc. no.405*, page 99, 1996.

[72] Chun-xi Wang and John Irwin. Centroid and emittance of a kicked beam in rings. *Proc. of 1997 Part. Acc. Conf.*, pages 360–2, 1997.

[73] Chun-xi Wang, John Irwin, Karl L.F. Bane, Yunhai Cai, Michiko Minty, Franz-J. Decker, and Yiton T. Yan. Model-Independent Analysis of beam dynamics in accelerators. *submited to Phys. Rev. E.*, 1999.

[74] Chun-xi Wang, John Irwin, and Yiton T. Yan. Computation of nonlinear one-turn maps from measurement with Model-Independent Analysis. *Proceedings of the 1999 Particle Accelerator Conference*, 1999.

[75] Y. T. Yan. Application of differential algebra to single-particle dynamics in storage rings. *Physics of Particle Accelerators, AIP Conf. Proc. no.249, chapter 4*, page 380, 1992.

[76] Y. T. Yan, J. Irwin, and T. Chen. Nonlinear analysis of storage ring lattice using one-turn map. *Proceedings of the 1995 Particle Accelerator Conference*, page 2765, 1995.

[77] V. Ziemann. "corrector ironing". SLAC-CN-393 (1992).